IMPROVED METHODS AND COMPOSITIONS FOR MODULATING A GENOME

RELATED APPLICATIONS

This application claims priority to U.S. Serial No.: 62/985,264 filed March 4, 2020 and U.S. Serial No.: 63/035,674 filed June 5, 2020, the entire contents of each of which is incorporated herein by reference.

BACKGROUND

Integration of a nucleic acid of interest into a genome occurs at low frequency and with little site specificity, in the absence of a specialized protein to promote the insertion event. Some existing approaches, like CRISPR/Cas9, are more suited for small edits and are less effective at integrating longer sequences. Other existing approaches, like Cre/loxP, require a first step of inserting a loxP site into the genome and then a second step of inserting a sequence of interest into the loxP site. There is a need in the art for improved proteins for inserting sequences of interest into a genome.

SUMMARY OF THE INVENTION

This disclosure relates to novel compositions, systems and methods for altering a genome at one or more locations in a host cell, tissue or subject, *in vivo* or *in vitro*. In particular, the invention features compositions, systems and methods for the introduction of exogenous genetic elements into a host genome. The disclosure also provides systems for altering a genomic DNA sequence of interest, e.g., by inserting, deleting, or substituting one or more nucleotides into/from the sequence of interest.

Features of the compositions or methods can include one or more of the following enumerated embodiments.

1. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table Z1 or Table X) and (ii) an endonuclease domain, wherein one or both of (i) or (ii) have an amino acid sequence encoded by a nucleic acid sequence of an element of

Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

2. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table Z1 or Table X) and (ii) an endonuclease domain, wherein one or both of (i) or (ii) have an amino acid sequence encoded by a nucleic acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

3. The system ofany of the preceding embodiments, wherein the heterologous object sequence encodes a therapeutic polypeptide or that encodes a mammalian (e.g., human) polypeptide, or a fragment or variant thereof.

4. The system of any of the preceding embodiments, wherein the heterologous object sequence encodes a therapeutic non-coding RNA (e.g., a miRNA).

5. The system of any of the preceding embodiments, wherein the heterologous object sequence comprises a regulatory sequence (e.g., a promoter, an enhancer, a binding site for an endogenous regulatory component, e.g., a miRNA binding site), e.g., which alters the expression of an endogenous gene or non-coding RNA.

6. The system of any of the preceding embodiments, wherein the regulatory sequence results in the upregulation of an endogenous gene or non-coding RNA.

7. The system of any of the preceding embodiments wherein the regulatory sequence results in the downregulation of an endogenous gene or non-coding RNA.

8. A system for modifying DNA comprising :

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a second strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target homology domain;

wherein:

(i) the polypeptide comprises a heterologous targeting domain (e.g., in the DBD or the endonuclease domain) that binds specifically to a sequence comprised in the target site; and/or

(ii) the template RNA comprises a heterologous homology sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence comprised in a target site.

9. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table Z1 or Table X) and (ii) an endonuclease domain, wherein one or both of (i) or (ii) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

10. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, or Table Z1 or Table X) and (ii) an endonuclease domain, wherein one or

both of (i) or (ii) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

11. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table Z1 or Table X) and (ii) a target DNA binding domain, wherein one or both of (i) or (ii) have an amino acid sequence encoded by a nucleic acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

12. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table Z1 or Table X) and (ii) a target DNA binding domain, wherein one or both of (i) or (ii) have an amino acid sequence encoded by a nucleic acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

13. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain from a protein other than a retrotransposase, e.g., from a retrovirus, e.g., reverse transcriptase domain as listed in Table Z1

or Z2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and (ii) an endonuclease domain and/or a target DNA binding domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

14. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain from a protein other than a retrotransposase, e.g., from a retrovirus, e.g., reverse transcriptase domain as listed in Table Z1 or Z2, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and (ii) an endonuclease domain and/or a target DNA binding domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

15. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table Z1 or Z2) and (ii) an endonuclease domain and/or a target DNA binding domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence;

wherein the sequence of the template RNA that binds the polypeptide has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 5' UTR or a 3' UTR of a sequence of an element of Table 10 or Table X.

16. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table Z1 or Z2) and (ii) an endonuclease domain and/or a target DNA binding domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence;

wherein the sequence of the template RNA that binds the polypeptide has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to: (i) the nucleotides located 5' relative to a start codon of a sequence of an element of Table 10 or Table X (e.g., comprising a retrotransposase-binding region), or (ii) the nucleotides located 3' relative to a stop codon of a sequence of an element of Table 10 or Table X (e.g., comprising a retrotransposase-binding region).

17. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table Z1, or Z2 or Table X) and (ii) an endonuclease domain and/or a target DNA binding domain;

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence; and

(c) an intein.

18. The system of any of the preceding embodiments, wherein the polypeptide comprises the intein.

19. The system of any of the preceding embodiments, wherein the intein is a split intein.

20. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence (e.g., a CRISPR spacer) that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain.

21. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the RT domain has an amino acid sequence of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

22. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (etRNA) (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the system is capable of producing an insertion into the target site of at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides.

23. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the heterologous object sequence is at least 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 120, 140, 160, 180, 200, 500, or 1,000 nts in length.

24. The system of any of the preceding embodiments, wherein one or more of: the RT domain is heterologous to the DBD; the DBD is heterologous to the endonuclease domain; or the RT domain is heterologous to the endonuclease domain.

25. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the system is capable of producing a deletion into the target site of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides.

26. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein (a)(ii) and/or (a)(iii) comprises a TALE molecule; a zinc finger molecule; or a CRISPR/Cas molecule chosen from Table 1 or a functional variant (e.g., mutant) thereof.

27. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence (e.g., a CRISPR spacer) that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the endonuclease domain, e.g., nickase domain, cuts both strands of the target site DNA, and wherein the cuts are separated from one another by at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or 30 nucleotides.

28. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) a sequence that specifically binds the RT domain, (iii) a heterologous object sequence, and (iv) a 3' homology domain.

29. The system of any of the preceding embodiments, wherein the template RNA further comprises a sequence that binds (a)(ii) and/or (a)(iii).

30. A system for modifying DNA comprising:

(a) a first polypeptide or a nucleic acid encoding the first polypeptide, wherein the first polypeptide comprises (i) a reverse transcriptase (RT) domain and (ii) optionally a DNA-binding domain,

(b) a second polypeptide or a nucleic acid encoding the second polypeptide, wherein the second polypeptide comprises (i) a DNA-binding domain (DBD); (ii) an endonuclease domain, e.g., a nickase domain; and

(c) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds the second polypeptide (e.g., that binds (b)(i) and/or (b)(ii)), (ii) optionally a sequence that binds the first polypeptide (e.g., that specifically binds the RT domain), (iii) a heterologous object sequence, and (iv) a 3' homology domain.

31. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, and (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain;

(b) a first template RNA (or DNA encoding the RNA) comprising (e.g., from 5' to 3') (i) a sequence that binds the polypeptide (e.g., that binds (a)(ii) and/or (a)(iii)) and (ii) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (e.g., wherein the first RNA comprises a gRNA);

(c) a second template RNA (or DNA encoding the RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds the polypeptide (e.g., that specifically binds the RT domain), (ii) a heterologous object sequence, and (iii) a 3' homology domain.

32. The system of any of the preceding embodiments, wherein the second template RNA comprises (i).

33. The system of any of the preceding embodiments, wherein the first template RNA comprises a first conjugating domain and the second template RNA comprises a second conjugating domain.

34. The system of any of the preceding embodiments, wherein the first and second conjugating domains are capable of hybridizing to one another, e.g., under stringent conditions.

35. The system of any of the preceding embodiments, wherein association of the first conjugating domain and the second conjugating domain colocalizes the first template RNA and the second template RNA.

36. The system of any of the preceding embodiments, wherein the template RNA comprises (i).

37. The system of any of the preceding embodiments, wherein the template RNA comprises (ii).

38. The system of any of the preceding embodiments, wherein the template RNA comprises (i) and (ii).

39. A system for modifying DNA, comprising:

(a) a first polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises a reverse transcriptase (RT) domain, wherein the RT domain has a sequence of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and optionally a DNA-binding domain (DBD) (e.g., a first DBD); and

(b) a second polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a DBD (e.g., a second DBD); and (ii) an endonuclease domain, e.g., a nickase domain.

40. The system of any of the preceding embodiments, wherein the nucleic acid encoding the first polypeptide and the nucleic acid encoding the second polypeptide are two separate nucleic acids.

41. The system of any of the preceding embodiments, wherein the nucleic acid encoding the first polypeptide and the nucleic acid encoding the second polypeptide are part of the same nucleic acid molecule, e.g., are present on the same vector.

42. The system of any of the preceding embodiments, having one or more (e.g., 1, 2, 3, 4, 5, 6, or all) of the following characteristics:

i. the heterologous object sequence encodes a protein, e.g. an enzyme (e.g., a lysosomal enzyme) or a blood factor (e.g., Factor I, II, V, VII, X, XI, XII or XIII);

ii. the heterologous object sequence comprises a tissue specific promoter or enhancer;

iii. the heterologous object sequence encodes a polypeptide of greater than 50, 100, 150, 200, 250, 300, 400, 500, or 1,000 amino acids, and optionally up to 7,500 amino acids;

iv. the heterologous object sequence encodes a fragment of a mammalian gene but does not encode the full mammalian gene, e.g., encodes one or more exons but does not encode a full-length protein;

v. the heterologous object sequence encodes one or more introns;

vi. the heterologous object sequence is other than a GFP, e.g., is other than a fluorescent protein or is other than a reporter protein; or

vii. the heterologous object sequence comprises only non-coding sequences, e.g., regulatory elements.

43. The system of any of the preceding embodiments, wherein the polypeptide further comprises a target DNA binding domain, e.g., having an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

44. The system of any of the preceding embodiments, wherein the polypeptide further comprises a target DNA binding domain, e.g., having an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

45. The system of any of the preceding embodiments, wherein the polypeptide further comprises a target DNA binding domain, e.g., having an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

46. The system of any of the preceding embodiments, wherein the polypeptide further comprises a target DNA binding domain, e.g., having an amino acid sequence of an element of

Table 3B, Table 10, Table 11, or Table X, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

47. The system of any of the preceding embodiments, wherein the polypeptide has an activity at 37°C that is no less than 70%, 75%, 80%, 85%, 90%, or 95% of its activity at 25°C under otherwise similar conditions.

48. The system of any of the preceding embodiments, wherein the polypeptide is derived from a homeothermic organism, e.g., a bird or a mammal.

49. The system of any of the preceding embodiments, wherein the polypeptide is derived from one or more of: a CRE retrotransposase, an NeSL retrotransposase, an R4 retrotransposase, an R2 retrotransposase, a Hero retrotransposase, an L1 retrotransposase, an RTE retrotransposase, an I retrotransposase, a Jockey retrotransposase, a CR1 retrotransposase, a Rex1 retrotransposase, an RandI/Dualen retrotransposase, a Penelope or Penelope-like retrotransposase, a Tx1 retrotransposase, an RTEX retrotransposase, a Crack retrotransposase, a Nimb retrotransposase, a Proto1 retrotransposase, a Proto2 retrotransposase, an RTETP retrotransposase, an L2 retrotransposase, a Tad1 retrotransposase, a Loa retrotransposase, an Ingi retrotransposase, an Outcast retrotransposase, an R1 retrotransposase, a Daphne retrotransposase, an L2A retrotransposase, an L2B retrotransposase, an Ambal retrotransposase, a Vingi retrotransposase, and/or a Kiri retrotransposase.

50. The system of any of the preceding embodiments, wherein the polypeptide comprises an endonuclease domain from a transposable element, e.g., a restriction-like endonuclease (RLE), an apurininc/apyrimidinic endonuclease-like endonuclease (APE), a GIY-YIG endonuclease.

51. The system of any of the preceding embodiments, wherein the endonuclease domain is intact.

52. The system of any of the preceding embodiments, wherein the endonuclease domain is inactivated.

53. The system of any of the preceding embodiments, wherein the endonuclease nicks DNA.

54. The system of any of the preceding embodiments, wherein the endonuclease makes a double stranded break.

55. The system of any of the preceding embodiments, wherein the template RNA comprises a sequence of Table 3A or 3B or 10 (e.g., one or both of a 5' untranslated region of column 6 of Table 3A or 3B and a 3' untranslated region of column 7 of Table 3A or 3B), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

56. The system of any of the preceding embodiments, wherein the template RNA comprises a sequence of Table 3A or 3B or 11 (e.g., one or both of a 5' untranslated region of column 6 of Table 3A or 3B and a 3' untranslated region of column 7 of Table 3A or 3B), or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

57. The system of any of the preceding embodiments, wherein the system has one or more (e.g., 1, 2, 3, or all) of the following characteristics:

i. the nucleic acid encoding the polypeptide and the template RNA or a nucleic acid encoding the template RNA are separate nucleic acids;

ii. the template RNA does not encode an active reverse transcriptase, e.g., comprises an inactivated mutant reverse transcriptase, e.g., as described in Examples 1-2, or does not comprise a reverse transcriptase sequence;

iii. the template RNA does not encode an active endonuclease, e.g., comprises an inactivated endonuclease or does not comprise an endonuclease; or

iv. the template RNA comprises one or more chemical modifications.

58. The system of any of the preceding embodiments, wherein the template RNA (or DNA encoding the template RNA) comprises (i) a 5' UTR sequence that binds the polypeptide, (ii) a 3' UTR sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a promoter operably linked to the heterologous object sequence,

wherein the promoter is disposed between the 5' untranslated sequence that binds the polypeptide and the heterologous sequence, or

wherein the promoter is disposed between the 3' untranslated sequence that binds the polypeptide and the heterologous sequence.

59. The system of any of the preceding embodiments, wherein the template RNA (or DNA encoding the template RNA) comprises (i) a 5' UTR sequence that binds the polypeptide, (ii) a 3' UTR sequence that binds the polypeptide, and (iii) a heterologous object sequence, and

wherein the heterologous object sequence comprises an open reading frame (or the reverse complement thereof) in a 5' to 3' orientation on the template RNA; or

wherein the heterologous object sequence comprises an open reading frame (or the reverse complement thereof) in a 3' to 5' orientation on the template RNA.

60. The system of any of the preceding embodiments, wherein the 5' UTR sequence of the template RNA (or DNA encoding the template RNA) has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 5' UTR sequence of a sequence of an element of Table 3B, Table 10, or Table X.

61. The system of any of the preceding embodiments, wherein the 5' UTR sequence of the template RNA (or DNA encoding the template RNA) has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to nucleotides located 5' relative to a start codon of a sequence of an element of Table 3B, Table 10, or Table X (e.g., comprising a retrotransposase-binding region).

62. The system of any of the preceding embodiments, wherein the 5' UTR sequence of the template RNA (or DNA encoding the template RNA) has substantial structural similarity (e.g., substantial secondary structural similarity) to a 5' UTR sequence of a sequence of an element of Table 3B, Table 10, or Table X.

63. The system of any of the preceding embodiments, wherein the 5' UTR sequence of the template RNA (or DNA encoding the template RNA) has at least 70%, 75%, 80%, 85%, 90%,

95%, 96%, 97%, 98%, 99%, or 100% identity to a 5' UTR sequence of column 6 of Table 3A or 3B or 10.

64. The system of any of the preceding embodiments, wherein the 3' UTR sequence of the template RNA (or DNA encoding the template RNA) has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 3' UTR sequence of a sequence of an element of Table 3B, Table 10, or Table X.

65. The system of any of the preceding embodiments, wherein the 3' UTR sequence of the template RNA (or DNA encoding the template RNA) has substantial structural similarity (e.g., substantial secondary structural similarity) to a 3' UTR sequence of a sequence of an element of Table 3B, Table 10, or Table X.

66. The system of any of any of the preceding embodiments, wherein the 3' UTR sequence of the template RNA (or DNA encoding the template RNA) has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 3' UTR sequence of Table 10 or column 7 of Table 3A or 3B.

67. The system of any of any of the preceding embodiments, wherein the 3' UTR sequence of the template RNA (or DNA encoding the template RNA) has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to the nucleotides located 3' relative to a stop codon of a sequence of an element of Table 3B, Table 10, or Table X (e.g., comprising a retrotransposase-binding region).

68. The system of any of any of the preceding embodiments, wherein the 3' UTR of the template RNA (or DNA encoding the template RNA) is flanked by a homology domain, e.g., as described herein, e.g., a homology domain having at least 5, 10, 20, 50, or 100 bases of at least 80% (e.g., at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 100%) identity to a target DNA strand, wherein optionally the homology domain comprises a sequence according to a 3' homology arm of Table 11, or a sequence having at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

69. The system of any of any of the preceding embodiments, wherein the 5' UTR is flanked by a homology domain, e.g., as described herein, e.g., a homology domain having at least 10, 20, 50, or 100 bases of at least 80% (e.g., at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 100%) identity to a target DNA strand, wherein optionally the homology domain comprises a sequence according to a 5' homology arm of Table 11, or a sequence having at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

70. The system of any of the preceding embodiments, wherein at least one of the reverse transcriptase domain, the endonuclease domain, or the target DNA binding domain is heterologous, e.g., relative to the other domains.

71. The system of any of the preceding embodiments, wherein the endonuclease domain is heterologous relative to the reverse transcriptase domain and/or the target DNA binding domain.

72. The system of any of the preceding embodiments, wherein the polypeptide comprises (i) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of an apurinic/apyrimidinic endonuclease (APE)-type non-LTR retrotransposon and (ii) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to an endonuclease domain of an APE-type non-LTR retrotransposon.

73. The system of any of the preceding embodiments, wherein the polypeptide comprises (i) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of a Penelope-like element (PLE)-type retrotransposon and (ii) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to an endonuclease domain of a PLE-type retrotransposon, e.g., wherein the PLE-type retrotransposon comprises a GIY-YIG endonuclease.

74. The system of any of the preceding embodiments, wherein the PLE-type retrotransposon does not comprise a functional endonuclease domain.

75. The system of any of the preceding embodiments, wherein the PLE-type retrotransposase comprises a Penelope-like element that naturally lacks an endonuclease domain (e.g., an Athena element, e.g., as described in Gladyshev and Arkhipova PNAS 104, 9352-9357 (2007)).

76. The system of any of the preceding embodiments, wherein the PLE-type retrotransposase lacking a functional endonuclease domain is fused to Cas9, e.g., wherein the PLE-type retrotransposase fused to Cas9 has DBD and/or endonuclease (e.g., nickase) function.

77. The system of any of the preceding embodiments, wherein the polypeptide comprises (i) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of a restriction enzyme-like endonuclease (RLE)-type non-LTR retrotransposon, (ii) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to an endonuclease domain of a RLE-type non-LTR retrotransposon, and (iii) a target DNA binding domain heterologous to (i) and/or (ii) (e.g., a heterologous zinc-finger DNA binding domain).

78. The system of any of the preceding embodiments, wherein the polypeptide comprises (i) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of a restriction enzyme-like endonuclease (RLE)-type non-LTR retrotransposon, and (ii) a heterologous endonuclease domain and/or a heterologous DNA binding domain.

79. The system of any of the preceding embodiments, wherein the polypeptide comprises (i) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of an apurinic/apyrimidinic endonuclease (APE)-type non-LTR retrotransposon, (ii) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to an endonuclease domain of an APE-type non-LTR retrotransposon, and (iii) a target DNA binding domain heterologous to (i) and/or (ii) (e.g., a heterologous zinc-finger DNA binding domain).

80. The system of any of the preceding embodiments, wherein the polypeptide comprises (i) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of an apurinic/apyrimidinic endonuclease (APE)-type non-LTR retrotransposon, and (ii) a heterologous endonuclease domain and/or a heterologous DNA binding domain.

81. The system of any of the preceding embodiments, wherein the polypeptide comprises (i) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of a Penelope-like element (PLE)-type retrotransposon, (ii) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to an endonuclease domain of a PLE-type retrotransposon, and (iii) a target DNA binding domain heterologous to (i) and/or (ii) (e.g., a heterologous zinc-finger DNA binding domain).

82. The system of any of the preceding embodiments, wherein the polypeptide comprises (i) a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a reverse transcriptase domain of a Penelope-like element (PLE)-type retrotransposon, and (ii) a heterologous endonuclease domain and/or a heterologous DNA binding domain.

83. The system of any of the preceding embodiments, wherein the template RNA comprises (iii) a promoter operably linked to the heterologous object sequence.

84. The system of any of the preceding embodiments, wherein the polypeptide further comprises (iii) a DNA-binding domain.

85. The system of any of the preceding embodiments, wherein the DNA binding domain has endonuclease activity.

86. The system of any of the preceding embodiments, wherein the endonuclease domain or endonuclease activity forms a double stranded break in DNA.

87. The system of any of the preceding embodiments, wherein the endonuclease domain or endonuclease activity nicks DNA.

88. The system of any of the preceding embodiments, wherein the polypeptide comprises a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to a sequence in column 7 of Table 3A or 3B.

89. The system of any of the preceding embodiments, wherein the polypeptide comprises a sequence at least 80% identical (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identical) to the sequence of an element listed in Table 10, Table 11, or Table X.

90. The system of any of the preceding embodiments, wherein the nucleic acid encoding the polypeptide and the template RNA or the nucleic acid encoding the template RNA are covalently linked, e.g., are part of a fusion nucleic acid.

91. The system of any of the preceding embodiments, wherein the fusion nucleic acid comprises RNA.

92. The system of any of the preceding embodiments, wherein the fusion nucleic acid comprises DNA.

93. The system of any of the preceding embodiments, wherein (b) comprises template RNA.

94. The system of any of the preceding embodiments, wherein the template RNA further comprises a nuclear localization signal.

95. The system of any of the preceding embodiments, wherein (a) comprises RNA encoding the polypeptide.

96. The system of any of the preceding embodiments, wherein the RNA of (a) and the RNA of (b) are separate RNA molecules.

97. The system of any of the preceding embodiments, wherein the RNA of (a) and the RNA of (b) are present at a ratio of between 100:1 and 10:1, 10:1 and 5:1, 5:1 and 2:1, 2:1 and 1:1, 1:1 and 1:2, 1:2 and 1:5, 1:5 and 1:10, or 1:10 and 1:100.

98. The system of any of the preceding embodiments, wherein the RNA of (a) does not comprise a nuclear localization signal.

99. The system of any of the preceding embodiments, wherein the polypeptide further comprises a nuclear localization signal and/or a nucleolar localization signal.

100. The system of any of the preceding embodiments, wherein (a) comprises an RNA that encodes: (i) the polypeptide and (ii) a nuclear localization signal and/or a nucleolar localization signal.

101. The system of any of the preceding embodiments, wherein the RNA comprises a pseudoknot sequence, e.g., 5' of the heterologous object sequence.

102. The system of any of the preceding embodiments, wherein the RNA comprises a stem-loop sequence or a helix, 5' of the pseudoknot sequence.

103. The system of any of the preceding embodiments, wherein the RNA comprises one or more (e.g., 2, 3, or more) stem-loop sequences or helices 3' of the pseudoknot sequence, e.g. 3' of the pseudoknot sequence and 5' of the heterologous object sequence.

104. The system of any of the preceding embodiments, wherein the template RNA comprising the pseudoknot has catalytic activity, e.g., RNA-cleaving activity, e.g, cis-RNA-cleaving activity.

105. The system of any of the preceding embodiments, wherein the RNA comprises at least one stem-loop sequence or helix, e.g., 3' of the heterologous object sequence, e.g. 1, 2, 3, 4, 5 or more stem-loop sequences, hairpins or helices sequences.

106. The system of any of the preceding embodiments, wherein the reverse transcriptase domain has an amino acid sequence of a reverse transcriptase domain of an element listed in Figure 10, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

107. The system of any of the preceding embodiments, wherein the endonuclease domain has an amino acid sequence of an endonuclease domain of an element listed in Figure 10, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

108. The system of any of the preceding embodiments, wherein the reverse transcriptase domain has an amino acid sequence of a reverse transcriptase domain of R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-54_AAe, Tad1-65B_BG, TART-1_DWi, or RTE_Ele2 as provided herein, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

109. The system of any of the preceding embodiments, wherein the endonuclease domain has an amino acid sequence of an endonuclease domain of R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-54_AAe, Tad1-65B_BG, TART-1_DWi, or RTE_Ele2 as provided herein, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

110. The system of any of the preceding embodiments, wherein the polypeptide has an amino acid sequence of R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-54_AAe, Tad1-65B_BG, TART-1_DWi, or RTE_Ele2 as provided herein, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

111. The system of any of the preceding embodiments, wherein the sequence of the template RNA that binds the polypeptide has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 5' UTR or R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-54_AAe, Tad1-65B_BG, TART-1_DWi, or RTE_Ele2 as provided herein.

112. The system of any of the preceding embodiments, wherein the sequence of the template RNA that binds the polypeptide has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 3' UTR or R2NS-1_CSi, R4-1_PH, RTE-1_MD, L2-18_ACar, L2-2_DRe, Dong-1_MMa, CR1_AC_1, Vingi-1_Acar, SR2, R4-1_AC, DongAa, CR1-1_PH, RTE-2_LMi, CR1-10_AMi, RTE-2_OL, L1-1_Cho, BovB, Line1-2_ZM, RTE-1_Aip, CR1-3_IS, CR1-54_AAe, Tad1-65B_BG, TART-1_DWi, or RTE_Ele2 as provided herein.

113. The system of any of the preceding embodiments, wherein the nucleic encoding the polypeptide comprises a coding sequence that is codon-optimized for expression in human cells.

114. The system of any of the preceding embodiments, wherein the template RNA comprises a coding sequence that is codon-optimized for expression in human cells.

115. The system of any of the preceding embodiments, wherein the system comprises one or more circular RNA molecules (circRNAs).

116. The system of any of the preceding embodiments, wherein the circRNA encodes the Gene Writer polypeptide.

117. The system of any of the preceding embodiments, wherein the circRNA comprises a template RNA.

118. The system of any of the preceding embodiments, wherein circRNA is delivered to a host cell.

119. The system of any of the preceding embodiments, wherein the circRNA is capable of being linearized, e.g., in a host cell, e.g., in the nucleus of the host cell.

120. The system of any of the preceding embodiments, wherein the circRNA comprises a cleavage site.

121. The system of any of the preceding embodiments, wherein the circRNA further comprises a second cleavage site.

122. The system of any of the preceding embodiments, wherein the cleavage site can be cleaved by a ribozyme, e.g., a ribozyme comprised in the circRNA (e.g., by autocleavage).

123. The system of any of the preceding embodiments, wherein the circRNA comprises a ribozyme sequence.

124. The system of any of the preceding embodiments, wherein the ribozyme sequence is capable of autocleavage, e.g., in a host cell, e.g., in the nucleus of the host cell.

125. The system of any of the preceding embodiments, wherein the ribozyme is an inducible ribozyme.

126. The system of any of the preceding embodiments, wherein the ribozyme is a protein-responsive ribozyme, e.g., a ribozyme responsive to a nuclear protein, e.g., a genome-interacting protein, e.g., an epigenetic modifier, e.g., EZH2.

127. The system of any of the preceding embodiments, wherein the ribozyme is a nucleic acid-responsive ribozyme.

128. The system of any of the preceding embodiments, wherein the catalytic activity (e.g., autocatalytic activity) of the ribozyme is activated in the presence of a target nucleic acid molecule (e.g., an RNA molecule, e.g., an mRNA, miRNA, ncRNA, lncRNA, tRNA, snRNA, or mtRNA).

129. The system of any of the preceding embodiments, wherein the ribozyme is responsive to a target protein (e.g., an MS2 coat protein).

130. The system of any of the preceding embodiments, wherein the target protein localized to the cytoplasm or localized to the nucleus (e.g., an epigenetic modifier or a transcription factor).

131. The system of any of the preceding embodiments, wherein the ribozyme comprises the ribozyme sequence of a B2 or ALU retrotransposon, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

132. The system of any of the preceding embodiments, wherein the ribozyme comprises the sequence of a tobacco ringspot virus hammerhead ribozyme, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

133. The system of any of the preceding embodiments, wherein the ribozyme comprises the sequence of a hepatitis delta virus (HDV) ribozyme, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

134. The system of any of the preceding embodiments, wherein the ribozyme is activated by a moiety expressed in a target cell or target tissue.

135. The system of any of the preceding embodiments, wherein the ribozyme is activated by a moiety expressed in a target subcellular compartment (e.g., a nucleus, nucleolus, cytoplasm, or mitochondria).

136. The system of any of the preceding embodiments, wherein the ribozyme is comprised in a circular RNA or a linear RNA.

137. A system comprising a first circular RNA encoding the polypeptide of a Gene Writing system; and

a second circular RNA comprising the template RNA of a Gene Writing system.

138. The system of any of the preceding embodiments, wherein the template RNA, e.g., the 5' UTR, comprises a ribozyme which cleaves the template RNA (e.g., in the 5' UTR).

139. The system of any of the preceding embodiments, wherein the template RNA comprises a ribozyme that is heterologous to (a)(i) (the a reverse transcriptase domain), (a)(ii) (the endonuclease domain), (b)(i) (a sequence of the template RNA that binds the polypeptide), or a combination thereof.

140. The system of any of the preceding embodiments, wherein the heterologous ribozyme is capable of cleaving RNA comprising the ribozyme, e.g., 5' of the ribozyme, 3' of the ribozyme, or within the ribozyme.

141. A lipid nanoparticle (LNP) comprising the system, polypeptide (or RNA encoding the same), nucleic acid molecule, or DNA encoding the system or polypeptide, of any of the preceding embodiments.

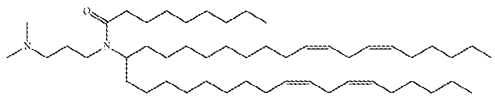
142. A system comprising a first lipid nanoparticle comprising the polypeptide (or DNA or RNA encoding the same) of a Gene Writing system (e.g., as described herein); and

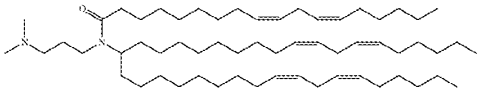
a second lipid nanoparticle comprising a nucleic acid molecule of a Gene Writing System (e.g., as described herein).

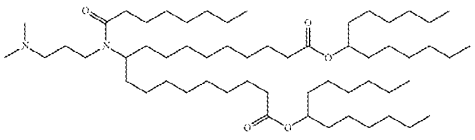
143. The system or polypeptide, of any of the preceding embodiments, wherein the system, nucleic acid molecule, polypeptide, and/or DNA encoding the same, is formulated as a lipid nanoparticle (LNP).

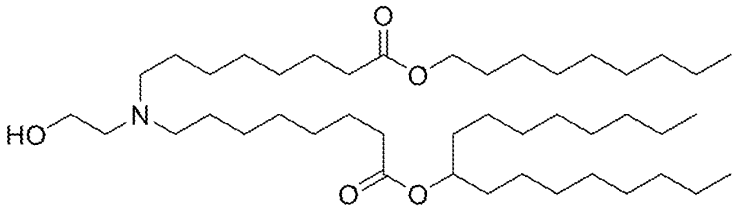
144. The LNP of any of the preceding embodiments, comprising a cationic lipid.

145. The LNP of any of the preceding embodiments, wherein the cationic lipid having a following structure:

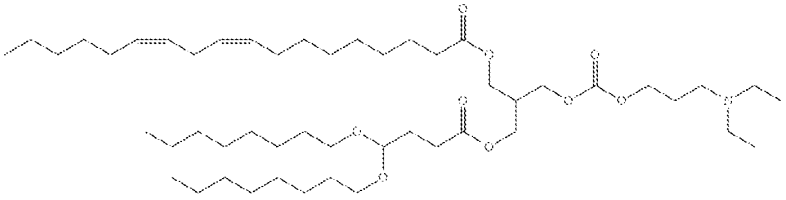

(i),


(ii),


(iii),



(vii), or


(ix).

146. The LNP of any of the preceding embodiments, further comprising one or more neutral lipid, e.g., DSPC, DPPC, DMPC, DOPC, POPC, DOPE, SM, a steroid, e.g., cholesterol, and/or

one or more polymer conjugated lipid, e.g., a pegylated lipid, e.g., PEG-DAG, PEG-PE, PEG-S-DAG, PEG-cer or a PEG dialkyoxypropylcarbamate.

147. The system, kit, or polypeptide, of any of the preceding embodiments, wherein the system, polypeptide, and/or DNA encoding the same, is formulated as a lipid nanoparticle (LNP).

148. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle (or a formulation comprising a plurality of the lipid nanoparticles) lacks reactive impurities (e.g., aldehydes), or comprises less than a preselected level of reactive impurities (e.g., aldehydes).

149. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle (or a formulation comprising a plurality of the lipid nanoparticles) lacks aldehydes, or comprises less than a preselected level of aldehydes.

150. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle is comprised in a formulation comprising a plurality of the lipid nanoparticles.

151. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagents comprising less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content.

152. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagents comprising less than 3% total reactive impurity (e.g., aldehyde) content.

153. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagents comprising less than 5%,

4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

154. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagent comprising less than 0.3% of any single reactive impurity (e.g., aldehyde) species.

155. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation is produced using one or more lipid reagents comprising less than 0.1% of any single reactive impurity (e.g., aldehyde) species.

156. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content.

157. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 3% total reactive impurity (e.g., aldehyde) content.

158. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

159. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 0.3% of any single reactive impurity (e.g., aldehyde) species.

160. The system, kit, or polypeptide of any of the preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 0.1% of any single reactive impurity (e.g., aldehyde) species.

161. The system, kit, or polypeptide of any of the preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content.

162. The system, kit, or polypeptide of any of the preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 3% total reactive impurity (e.g., aldehyde) content.

163. The system, kit, or polypeptide of any of the preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

164. The system, kit, or polypeptide of any of the preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 0.3% of any single reactive impurity (e.g., aldehyde) species.

165. The system, kit, or polypeptide of any of the preceding embodiments, wherein one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 0.1% of any single reactive impurity (e.g., aldehyde) species.

166. The system, kit, or polypeptide of any of the preceding embodiments, wherein the total aldehyde content and/or quantity of any single reactive impurity (e.g., aldehyde) species is determined by liquid chromatography (LC), e.g., coupled with tandem mass spectrometry (MS/MS), e.g., according to the method described in Example 7.

167. The system, kit, or polypeptide of any of the preceding embodiments, wherein the total aldehyde content and/or quantity of reactive impurity (e.g., aldehyde) species is determined by

detecting one or more chemical modifications of a nucleic acid molecule (e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents.

168. The system, kit, or polypeptide of any of the preceding embodiments, wherein the total aldehyde content and/or quantity of aldehyde species is determined by detecting one or more chemical modifications of a nucleotide or nucleoside (e.g., a ribonucleotide or ribonucleoside, e.g., comprised in or isolated from a nucleic acid molecule, e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents, e.g., as described in Example 8.

169. The system, kit, or polypeptide of any of the preceding embodiments, wherein the chemical modifications of a nucleic acid molecule, nucleotide, or nucleoside are detected by determining the presence of one or more modified nucleotides or nucleosides, e.g., using LC-MS/MS analysis, e.g., as described in Example 8.

170. Any above-numbered system, wherein the polypeptide comprises a sequence of at least 50 amino acids (e.g., at least 100, 150, 200, 300, 500 amino acids) having at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a sequence of a polypeptide encoded by a sequence of an element of Table 10, Table 11, or Table X, or a reverse transcriptase domain or endonuclease domain thereof.

171. Any above-numbered system, wherein the polypeptide comprises a sequence of at least 50 amino acids (e.g., at least 100, 150, 200, 300, 500 amino acids) having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 amino acid differences to a sequence of a polypeptide encoded by a sequence of an element of Table 10, Table 11, or Table X, or a reverse transcriptase domain or endonuclease domain thereof.

172. Any above-numbered system, wherein the polypeptide comprises a sequence of at least 50 amino acids (e.g., at least 100, 150, 200, 300, 500 amino acids) having at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a sequence of a polypeptide

listed in Table 3A or 3B or a reverse transcriptase domain, endonuclease domain, or DNA binding domain thereof.

173. Any above-numbered system, wherein the polypeptide comprises a sequence of at least 50 amino acids (e.g., at least 100, 150, 200, 300, 500 amino acids) having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 amino acid differences to a sequence of a polypeptide listed in Table 3A or 3B or a reverse transcriptase domain, endonuclease domain, or DNA binding domain thereof.

174. Any above-numbered system, wherein the polypeptide comprises a sequence of at least 50 amino acids (e.g., at least 100, 150, 200, 300, 500 amino acids) having at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to the amino acid sequence of column 7 of Table 3A or 3B, or a reverse transcriptase domain, endonuclease domain, or DNA binding domain thereof.

175. Any above-numbered system, wherein the polypeptide comprises a sequence of at least 50 amino acids (e.g., at least 100, 150, 200, 300, 500 amino acids) having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 amino acid differences to the amino acid sequence of column 7 of Table 3A or 3B, or a reverse transcriptase domain, endonuclease domain, or DNA binding domain thereof.

176. Any above-numbered system, wherein the template RNA comprises a sequence of an element of Table 10 or Table X (e.g., one or both of a 5' UTR of Table X or 10 and a 3' UTR of Table X or 10), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

177. Any above-numbered system, wherein the template RNA comprises a sequence of an element of Table 10 or Table X (e.g., one or both of a 5' UTR of Table X or 10 and a 3' UTR of Table X or 10), or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

178. Any above-numbered system, wherein the template RNA comprises a sequence of Table 3A or 3B (e.g., one or both of a 5' UTR of column 5 of Table 3A or 3B and a 3' UTR of column 6 of Table 3A or 3B), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

179. Any above-numbered system, wherein the template RNA comprises a sequence of Table 3A or 3B (e.g., one or both of a 5' UTR of column 5 of Table 3A or 3B and a 3' UTR of column 6 of Table 3A or 3B), or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

180. Any above-numbered system, wherein the template RNA comprises a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to: (i) the nucleotides located 5' relative to a start codon of a sequence of an element of Table 10 or Table X (e.g., comprising a retrotransposase-binding region).

181. Any above-numbered system, wherein the template RNA comprises a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to: (i) the nucleotides located 3' relative to a stop codon of a sequence of an element of Table 10 or Table X (e.g., comprising a retrotransposase-binding region).

182. The system of any of the preceding embodiments, wherein the template RNA comprises a sequence of about 100-125 bp from a 3' UTR of Table 10 or 11 or column 6 of Table 3A or 3B, e.g., wherein the sequence comprises nucleotides 1-100, 101-200, or 201-325 of the 3' UTR of column 6 of Table 3A or 3B, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

183. The system of any of the preceding embodiments, wherein the template RNA comprises a sequence of about 100-125 bp from a 3' UTR of Table 10 or 11 or column 6 of Table 3A or 3B, e.g., wherein the sequence comprises nucleotides 1-100, 101-200, or 201-325 of the 3' UTR of column 6 of Table 3A or 3B, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

184. Any above-numbered system, wherein (a) comprises RNA and (b) comprises RNA.

185. Any above-numbered system, which comprises only RNA, or which comprises more RNA than DNA by an RNA:DNA ratio of at least 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, or 100:1.

186. Any above-numbered system, which does not comprise DNA, or which does not comprise more than 10%, 5%, 4%, 3%, 2%, or 1% DNA by mass or by molar amount.

187. Any above-numbered system, which is capable of modifying DNA by insertion of the heterologous object sequence without an intervening DNA-dependent RNA polymerization of (b).

188. Any above-numbered system, which is capable of modifying DNA by insertion of the heterologous object sequence via target primed reverse transcription.

189. Any above-numbered system, which is capable of modifying DNA by insertion of a heterologous object sequence in the presence of an inhibitor of a DNA repair pathway (e.g., SCR7, a PARP inhibitor), or in a cell line deficient for a DNA repair pathway (e.g., a cell line deficient for the nucleotide excision repair pathway or the homology-directed repair pathway).

190. Any above-numbered system, which does not cause formation of a detectable level of double stranded breaks in a target cell.

191. Any above-numbered system, which is capable of modifying DNA using reverse transcriptase activity, and optionally in the absence of homologous recombination activity.

192. Any above-numbered system, wherein the template RNA has been treated to reduce secondary structure, e.g., was heated, e.g., to a temperature that reduces secondary structure, e.g., to at least 70, 75, 80, 85, 90, or 95°C.

193. The system of any of the preceding embodiments, wherein the template RNA was subsequently cooled, e.g., to a temperature that allows for secondary structure, e.g, to less than or equal to 37, 30, 25, or 20°C.

194. The system of any of the preceding embodiments, wherein the template RNA (or DNA encoding the template RNA) comprises a first homology domain having at least 5 or at least 10 bases of 100% identity to a target DNA strand, at the 5' end of the template RNA, and a second homology domain having at least 5 or at least 10 bases of 100% identity to a target DNA strand, at the 3' end of the template RNA.

195. The system of any of the preceding embodiments, wherein (a) and (b) are part of the same nucleic acid.

196. The system of any of the preceding embodiments, wherein (a) and (b) are separate nucleic acids.

197. The system of any of the preceding embodiments, wherein the template RNA comprises at least 5 or at least 10 bases of 100% identity to a target DNA strand (e.g., wherein the target DNA strand is a human DNA sequence), at the 5' end of the template RNA.

198. The system of any of the preceding embodiments, wherein the template RNA comprises at least 5 or at least 10 bases of 100% identity to a target DNA strand (e.g., wherein the target DNA strand is a human DNA sequence), at the 3' end of the template RNA.

199. The system of any of the preceding embodiments, wherein the polypeptide comprises an active RNase H domain.

200. The system of any of the preceding embodiments, wherein the polypeptide does not comprise an active RNase H domain.

201. The system of any of the preceding embodiments, wherein an endogenous RNase H domain of the polypeptide is inactivated.

202. The system of any of the preceding embodiments, wherein the transposase polypeptide comprises a mutation inactivating and/or deleting a nucleolar localization signal.

203. The system of any of the preceding embodiments, wherein the polypeptide does not comprise a functional nucleolar localization signal, e.g., does not comprise a nucleolar localization signal.

204. The system of any of the preceding embodiments, wherein activity of the nucleolar localization signal is reduced by at least 50%, 60%, 70%, 80%, 90%, 95%, or 99%.

205. The system of any of the preceding embodiments, wherein the polypeptide comprises a nuclear localization signal (NLS), e.g., an endogenous NLS or an exogenous NLS.

206. The system of any of the preceding embodiments, wherein:

the polypeptide comprises (i) a first target DNA binding domain, e.g., comprising a first Zn finger domain, (ii) a reverse transcriptase domain, (iii) an endonuclease domain, and (iv) a second target DNA binding domain, e.g., comprising a second Zn finger domain, heterologous to the first target DNA binding domain; and

wherein (a) binds to a smaller number of target DNA sequences in a target cell than a similar polypeptide that comprises only the first target DNA binding domain, e.g., wherein the presence of the second target DNA binding domain in the polypeptide with the first DNA binding domain refines the target sequence specificity of the polypeptide relative to the polypeptide target sequence specificity of the polypeptide comprising only the first target DNA binding domain.

207. The system of any of the preceding embodiments, wherein (iii) comprises (iv).

208. The system of any of the preceding embodiments, wherein the second target DNA binding domain binds to a genomic DNA sequence that is less than 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 nucleotides away from a genomic sequence to which the first target DNA binding domain binds.

209. The system of any of the preceding embodiments, wherein the second target DNA binding domain binds to a genomic DNA sequence that is 1-100, 1-90, 1-80, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10, 1-5, 5-100, 5-90, 5-80, 5-70, 5-60, 5-50, 5-40, 5-30, 5-20, 5-10, 10-100, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, 10-20, 20-100, 20-90, 20-80, 20-70, 20-60, 20-50, 20-40, 20-30, 30-100, 30-90, 30-80, 30-70, 30-60, 30-50, 30-40, 40-100, 40-90, 40-80, 40-70, 40-60, 40-50, 50-100, 50-90, 50-80, 50-70, 50-60, 60-100, 60-90, 60-80, 60-70, 70-100, 70-90, 70-80, 80-100, 80-90, or 90-100 nucleotides away from a genomic sequence to which the first target DNA binding domain binds.

210. The system of any of the preceding embodiments, wherein the first or second target DNA binding domain comprises a CRISPR/Cas protein, a TAL Effector domain, a Zn finger domain, or a meganuclease domain.

211. The system of any of the preceding embodiments, wherein the system is capable of cutting the first strand of the target DNA at least twice (e.g., twice), and

optionally wherein the cuts are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or 200 nucleotides away one another (and optionally no more than 500, 400, 300, 200, or 100 nucleotides away from one another).

212. The system of any of the preceding embodiments, wherein the system is capable of cutting the first strand and the second strand of the target DNA, and

wherein the distance between the cuts is the same as the distance between cuts made by the reverse transcriptase domain, e.g., the reverse transcriptase domain when situated in its endogenous polypeptide.

213. The system of any of the preceding embodiments, wherein the cuts are 1-500, 1-400, 1-300, 1-200, 1-100, 1-90, 1-80, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10, 1-5, 5-500, 5-400, 5-300, 5-200, 5-100, 5-90, 5-80, 5-70, 5-60, 5-50, 5-40, 5-30, 5-20, 5-10, 10-500, 10-400, 10-300, 10-200, 10-100, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, 10-20, 20-500, 20-400, 20-300, 20-200, 20-100, 20-90, 20-80, 20-70, 20-60, 20-50, 20-40, 20-30, 30-500, 30-400, 30-300, 30-200, 30-100, 30-90, 30-80, 30-70, 30-60, 30-50, 30-40, 40-500, 40-400, 40-300, 40-200, 40-100, 40-90, 40-80, 40-70, 40-60, 40-50, 50-500, 50-400, 50-300, 50-200, 50-100, 50-90, 50-80, 50-70, 50-60, 60-500, 60-400, 60-300, 60-200, 60-100, 60-90, 60-80, 60-70, 70-500, 70-400, 70-300, 70-200, 70-100, 70-90, 70-80, 80-500, 80-400, 80-300, 80-200, 80-100, 80-90, 90-500, 90-400, 90-300, 90-200, 90-100, 100-500, 100-400, 100-300, 100-200, 200-500, 200-400, 200-300, 300-500, 300-400, or 400-500 nucleotides away from one another.

214. The system of any of the preceding embodiments, wherein the distance between the cuts is the same as the distance between cuts made by the reverse transcriptase domain, e.g., the reverse transcriptase domain when situated in its endogenous polypeptide.

215. The system of any of the preceding embodiments, wherein the two cuts are both made by the same endonuclease domain (e.g., a CRISPR/Cas protein, e.g., directed by a plurality of gRNAs, e.g., disposed in the template RNA).

216. The system of any of the preceding embodiments, wherein the polypeptide further comprises a second endonuclease domain.

217. The system of any of the preceding embodiments, wherein:

i) the first endonuclease domain (e.g., nickase) cuts the to-be-edited strand of the target DNA and the second endonuclease domain (e.g., nickase) cuts the non-edited strand of the target DNA, or

ii) the first endonuclease domain (e.g., nickase) makes one of the two cuts to the to-be-edited strand of the target DNA and the second endonuclease domain (e.g., nickase) makes the other cut to the to-be-edited strand of the target DNA.

218. The system of any of the preceding embodiments, wherein (a), (b), or (a) and (b) further comprises a 5' UTR and/or 3' UTR operably linked to the sequence encoding the polypeptide, the heterologous object sequence (e.g., a coding sequence contained in the heterologous object sequence), or both.

219. The system of any of the preceding embodiments, wherein the 5' UTR and/or 3' UTR increase expression of the operably linked sequence(s) by at least 10%, 20%, 30%, 40%, 50%, 70%, 70%, 80%, 90%, or 100% relative to an otherwise similar nucleic acid comprising the endogenous UTR(s) associated with the heterologous object sequence or a minimal 5' UTR and a minimal 3' UTR.

220. The system of any of the preceding embodiments, wherein the template RNA (or DNA encoding the template RNA) comprises (i) a sequence that binds the polypeptide, (ii) a heterologous object sequence, and (iii) a ribozyme that is heterologous to (a)(i), (a)(ii), (b)(i), or a combination thereof.

221. The system of any of the preceding embodiments, wherein (a), (b), or (a) and (b) comprise an intron that increases the expression of the polypeptide, the heterologous object sequence (e.g., a coding sequence situated in the heterologous object sequence), or both.

222. A host cell (e.g., a mammalian cell, e.g., a human cell) comprising any preceding numbered system.

223. A method of modifying a target DNA strand in a cell, tissue or subject, comprising administering any preceding numbered system to the cell, tissue or subject, wherein the system reverse transcribes the template RNA sequence into the target DNA strand, thereby modifying the target DNA strand.

224. The method of any of the preceding embodiments, wherein the cell, tissue or subject is a mammalian (e.g., human) cell, tissue or subject.

225. The method of any of the preceding embodiments, wherein the tissue is liver, lung, skin, blood, immune, or muscle tissue.

226. The method of any of the preceding embodiments, wherein the cell is a fibroblast.

227. The method of any of the preceding embodiments, wherein the cell is a primary cell.

228. The method of any of the preceding embodiments, where in the cell is not immortalized.

229. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

230. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence encoded by a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

231. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

232. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

233. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence;

wherein the sequence of the template RNA that binds the polypeptide has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 5' UTR sequence or a 3' UTR sequence of a sequence of an element of Table 3B, Table 10, or Table X.

234. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence;

wherein the sequence of the template RNA that binds the polypeptide has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to: (i) the nucleotides located 5' relative to a start codon of a sequence of an element of Table 3B, Table 10, or Table X (e.g., comprising a retrotransposase-binding region), or (ii) the nucleotides located 3' relative to a stop codon of a sequence of an element of Table 3B, Table 10, or Table X (e.g., comprising a retrotransposase-binding region).

235. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain from a protein other than a retrotransposase, e.g., from a retrovirus, e.g., reverse transcriptase domain as listed in Table Z1 or Z2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

236. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain from a protein other than a retrotransposase, e.g., from a retrovirus, e.g., reverse transcriptase domain as listed in Table Z1 or Z2, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and (ii) an endonuclease domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

237. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain;

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence; and

(c) an intein.

238. The method of any of the preceding embodiments, wherein the polypeptide comprises the intein.

239. The system of any of the preceding embodiments, wherein the intein is a split intein.

240. The method of any of the preceding embodiments, wherein the polypeptide does not comprise a target DNA binding domain.

241. The method of any of the preceding embodiments, wherein the polypeptide is derived from an APE-type retrotransposon reverse transcriptase.

242. The method of any of the preceding embodiments, wherein the polypeptide further comprises a target DNA binding domain, e.g., having an amino acid sequence encoded by a

sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

243. The method of any of the preceding embodiments, wherein the polypeptide further comprises a target DNA binding domain, e.g., having an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

244. The method of any of the preceding embodiments, wherein the polypeptide further comprises a target DNA binding domain, e.g., having an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

245. The method of any of the preceding embodiments, wherein the polypeptide further comprises a target DNA binding domain, e.g., having an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

246. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) an RNA encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid.

247. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) an RNA encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid.

248. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) an RNA encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid.

249. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) an RNA encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid.

250. The method of any of the preceding embodiments, which results in the addition of at least 1, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of exogenous DNA sequence to the genome of the mammalian cell.

251. The method of any of the preceding embodiments, which results in the deletion of at least 1, 2, 3, 4, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA sequence from the genome of the mammalian cell.

252. The method of any of the preceding embodiments, which results in the alteration of at least 1, 2, 3, 4, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA sequence from the genome of the mammalian cell.

253. The method of any of the preceding embodiments, which results in the addition of a protein coding sequence to the genome of the mammalian cell.

254. The method of any of the preceding embodiments, which results in the deletion of a protein coding sequence to the genome of the mammalian cell.

255. The method of any of the preceding embodiments, which results in the alteration of a protein coding sequence to the genome of the mammalian cell.

256. The method of any of the preceding embodiments, which results in the addition of a non-coding sequence to the genome of the mammalian cell, e.g., encoding a non-coding RNA, e.g., a miRNA.

257. The method of any of the preceding embodiments, which results in the deletion of a non-coding sequence to the genome of the mammalian cell, e.g., encoding a non-coding RNA, e.g., a miRNA.

258. The method of any of the preceding embodiments, which results in the alteration of a non-coding sequence to the genome of the mammalian cell, e.g., encoding a non-coding RNA, e.g., a miRNA.

259. The method of any of the preceding embodiments, which results in the addition of a regulatory sequence to the genome of the mammalian cell, e.g., a promoter, an enhancer, a miRNA binding site.

260. The method of any of the preceding embodiments, which results in the deletion of a regulatory sequence to the genome of the mammalian cell, e.g., a promoter, an enhancer, a miRNA binding site.

261. The method of any of the preceding embodiments, which results in the alteration of a regulatory sequence to the genome of the mammalian cell, e.g., a promoter, an enhancer, a miRNA binding site.

262. The method of any of the preceding embodiments, wherein the addition, deletion, or alteration of the regulatory sequence to the genome of the mammalian cell results in increased expression of a coding or non-coding sequence in the genome of the mammalian cell.

263. The method of any of the preceding embodiments, wherein the addition, deletion, or alteration of the regulatory sequence to the genome of the mammalian cell results in decreased expression of a coding or non-coding sequence in the genome of the mammalian cell.

264. A method of inserting DNA into the genome of a mammalian cell, comprising contacting the cell with an RNA composition, wherein the RNA composition comprises:

(a) a first RNA that directs insertion of a template RNA into the genome, and

(b) a template RNA comprising a heterologous sequence,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid,

wherein the method results in the addition of at least 1, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA (e.g., exogenous DNA) sequence to the genome of the mammalian cell; and

wherein the first RNA encodes a polypeptide encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, wherein the polypeptide directs insertion of the template RNA into the genome.

265. A method of inserting DNA into the genome of a mammalian cell, comprising contacting the cell with an RNA composition, wherein the RNA composition comprises:

(a) a first RNA that directs insertion of a template RNA into the genome, and

(b) a template RNA comprising a heterologous sequence,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid,

wherein the method results in the addition of at least 1, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA (e.g., exogenous DNA) sequence to the genome of the mammalian cell; and

wherein the first RNA encodes a polypeptide encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having no more than 10, 20, 30, 40, 50,

60, 70, 80, 90, or 100 nucleotide differences therefrom, wherein the polypeptide directs insertion of the template RNA into the genome.

266. A method of inserting DNA into the genome of a mammalian cell, comprising contacting the cell with an RNA composition, wherein the RNA composition comprises:

(a) a first RNA that directs insertion of a template RNA into the genome, and

(b) a template RNA comprising a heterologous sequence,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid,

wherein the method results in the addition of at least 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA (e.g., exogenous DNA) sequence to the genome of the mammalian cell; and

wherein the first RNA encodes a polypeptide of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, wherein the polypeptide directs insertion of the template RNA into the genome.

267. A method of inserting DNA into the genome of a mammalian cell, comprising contacting the cell with an RNA composition, wherein the RNA composition comprises:

(a) a first RNA that directs insertion of a template RNA into the genome, and

(b) a template RNA comprising a heterologous sequence,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid,

wherein the method results in the addition of at least 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA (e.g., exogenous DNA) sequence to the genome of the mammalian cell; and

wherein the first RNA encodes a polypeptide of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100

nucleotide differences therefrom, wherein the polypeptide directs insertion of the template RNA into the genome.

268. A method of inserting DNA into the genome of a mammalian cell, comprising contacting the cell with an RNA composition, wherein the RNA composition comprises:

(a) a first RNA that directs insertion of a template RNA into the genome, and

(b) a template RNA comprising a heterologous sequence and (i) a flanking sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 5' UTR sequence of a sequence of an element of Table 3B, Table 10, or Table X, and/or (ii) a flanking sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 3' UTR sequence of a sequence of an element of Table 3B, Table 10, or Table X,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid,

wherein the method results in the addition of at least 1, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA (e.g., exogenous DNA) sequence to the genome of the mammalian cell.

269. A method of inserting DNA into the genome of a mammalian cell, comprising contacting the cell with an RNA composition, wherein the RNA composition comprises:

(a) a first RNA that directs insertion of a template RNA into the genome, and

(b) a template RNA comprising a heterologous sequence and (i) a flanking sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to the nucleotides located 5' relative to a start codon of a sequence of an element of Table 3B, Table 10, or Table X (e.g., comprising a retrotransposase-binding region), and/or (ii) a flanking sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to the nucleotides located 3' relative to a stop codon of a sequence of an element of Table 3B, Table 10, or Table X (e.g., comprising a retrotransposase-binding region),

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid,

wherein the method results in the addition of at least 1, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA (e.g., exogenous DNA) sequence to the genome of the mammalian cell.

270. The method of any of the preceding embodiments, wherein the template RNA further comprises a sequence that binds the polypeptide; optionally wherein the sequence that binds the polypeptide has:

(a) at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 5' UTR sequence or a 3' UTR sequence of a sequence of an element of Table 3B, Table 10, or Table X; or

(b) at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to: (i) the nucleotides located 5' relative to a start codon of a sequence of an element of Table 3B, Table 10, or Table X (e.g., comprising a retrotransposase-binding region), or (ii) the nucleotides located 3' relative to a stop codon of a sequence of an element of Table X (e.g., comprising a retrotransposase-binding region).

271. The method of any of the preceding embodiments, wherein at least 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 bp of exogenous DNA are added to the genome of a mammalian cell, without delivery of DNA to the cell.

272. The method of any of the preceding embodiments, wherein at least 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 bp of exogenous DNA are added to the genome of a mammalian cell,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the method comprises contacting the mammalian cell with a composition comprising less than 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, or 0.01% DNA by mass or by molar amount of nucleic acid.

273. The method of any of the preceding embodiments, wherein at least 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 bp of exogenous DNA are added to the genome of a mammalian cell, wherein the only RNA is delivered to the mammalian cell.

274. The method of any of the preceding embodiments, wherein at least at least 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 1000 bp of exogenous DNA are added to the genome of a mammalian cell, wherein RNA and protein are delivered to the mammalian cell.

275. The method of any of the preceding embodiments, wherein the template RNA serves as the template for insertion of the exogenous DNA.

276. The method of any of the preceding embodiments, which does not comprise DNA-dependent RNA polymerization of exogenous DNA.

277. The method of any of the preceding embodiments, which results in the addition of at least 1, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000 base pairs of DNA to the genome of the mammalian cell.

278. The methods of any of the preceding embodiments, wherein the RNA of (a) and the RNA of (b) are covalently linked, e.g., are part of the same transcript.

279. The methods of any of the preceding embodiments, wherein the RNA of (a) and the RNA of (b) are separate RNAs.

280. The method of any of the preceding embodiments, which does not comprise contacting the mammalian cell with a template DNA.

281. A method of modifying the genome of a human cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the method results in insertion of the heterologous object sequence into the human cell's genome,

wherein the human cell does not show upregulation of any DNA repair genes and/or tumor suppressor genes, or wherein no DNA repair gene and/or tumor suppressor gene is upregulated by more than 50%, 25%, 10%, 5%, 2%, or 1%, e.g., wherein upregulation is measured by RNA-seq.

282. A method of modifying the genome of a human cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the method results in insertion of the heterologous object sequence into the human cell's genome,

wherein the human cell does not show upregulation of any DNA repair genes and/or tumor suppressor genes, or wherein no DNA repair gene and/or tumor suppressor gene is upregulated by more than 50%, 25%, 10%, 5%, 2%, or 1%, e.g., wherein upregulation is measured by RNA-seq, e.g., as described in Example 14 of PCT/US2019/048607, herein incorporated by reference in its entirety.

283. A method of modifying the genome of a human cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the method results in insertion of the heterologous object sequence into the human cell's genome,

wherein the human cell does not show upregulation of any DNA repair genes and/or tumor suppressor genes, or wherein no DNA repair gene and/or tumor suppressor gene is upregulated by more than 50%, 25%, 10%, 5%, 2%, or 1%, e.g., wherein upregulation is measured by RNA-seq.

284. A method of modifying the genome of a human cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table Table 3B, Table 10, Table 11, Table X, or Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; wherein one, two, or three of (i), (ii), and/or (iii) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence,

wherein the method results in insertion of the heterologous object sequence into the human cell's genome,

wherein the human cell does not show upregulation of any DNA repair genes and/or tumor suppressor genes, or wherein no DNA repair gene and/or tumor suppressor gene is

upregulated by more than 50%, 25%, 10%, 5%, 2%, or 1%, e.g., wherein upregulation is measured by RNA-seq.

285. A method of modifying the genome of a human cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 5' UTR sequence or a 3' UTR sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X, and (ii) a heterologous object sequence,

wherein the method results in insertion of the heterologous object sequence into the human cell's genome,

wherein the human cell does not show upregulation of any DNA repair genes and/or tumor suppressor genes, or wherein no DNA repair gene and/or tumor suppressor gene is upregulated by more than 50%, 25%, 10%, 5%, 2%, or 1%, e.g., wherein upregulation is measured by RNA-seq.

286. A method of modifying the genome of a human cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain (e.g., as listed in Table 3B, Table 10, Table 11, Table X, or Table Z1 or Z2), (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to: (i) the portion of a sequence of an element of Table 3B, Table 10, or Table X consisting of the nucleotides located 5' relative to the start codon, or (ii) the portion of a sequence of an element of Table 3B, Table 10, or Table X consisting of the nucleotides located 3' relative to the stop codon,

wherein the method results in insertion of the heterologous object sequence into the human cell's genome,

wherein the human cell does not show upregulation of any DNA repair genes and/or tumor suppressor genes, or wherein no DNA repair gene and/or tumor suppressor gene is upregulated by more than 50%, 25%, 10%, 5%, 2%, or 1%, e.g., wherein upregulation is measured by RNA-seq.

287. A method of adding an exogenous coding region to the genome of a cell (e.g., a mammalian cell), comprising contacting the cell with: (i) an RNA comprising the non-coding strand of the exogenous coding region, wherein optionally the RNA does not comprise a coding strand of the exogenous coding region, and (ii) a polypeptide comprising a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, wherein optionally the delivery comprises non-viral delivery.

288. A method of adding an exogenous coding region to the genome of a cell (e.g., a mammalian cell), comprising contacting the cell with: (i) an RNA comprising the non-coding strand of the exogenous coding region, wherein optionally the RNA does not comprise a coding strand of the exogenous coding region, and (ii) a polypeptide comprising a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, wherein optionally the delivery comprises non-viral delivery.

289. A method of expressing a polypeptide of interest in a cell (e.g., a mammalian cell), comprising contacting the cell with: (i) an RNA, wherein the RNA comprises a non-coding strand that is the reverse complement of a sequence that would encoding the polypeptide of interest, wherein optionally the RNA does not comprise a coding strand encoding the polypeptide of interest, and (ii) and a retrotransposase polypeptide comprising an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, wherein optionally the delivery comprises non-viral delivery.

290. A method of expressing a polypeptide of interest in a cell (e.g., a mammalian cell), comprising contacting the cell with: (i) an RNA, wherein the RNA comprises a non-coding strand that is the reverse complement of a sequence that would encoding the polypeptide of interest, wherein optionally the RNA does not comprise a coding strand encoding the polypeptide of interest, and (ii) and a retrotransposase polypeptide comprising an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, wherein optionally the delivery comprises non-viral delivery.

291. A method of expressing a polypeptide of interest in a cell (e.g., a mammalian cell), comprising contacting the cell with: (i) an RNA, wherein the RNA comprises a non-coding strand that is the reverse complement of a sequence that would encoding the polypeptide of interest, wherein optionally the RNA does not comprise a coding strand encoding the polypeptide of interest, and (ii) a retrotransposase polypeptide comprising an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, wherein optionally the delivery comprises non-viral delivery.

292. A method of expressing a polypeptide of interest in a cell (e.g., a mammalian cell), comprising contacting the cell with: (i) an RNA, wherein the RNA comprises a non-coding strand that is the reverse complement of a sequence that would encoding the polypeptide of interest, wherein optionally the RNA does not comprise a coding strand encoding the polypeptide of interest, and (ii) a retrotransposase polypeptide comprising an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X, or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, wherein optionally the delivery comprises non-viral delivery.

293. The method of any of the preceding embodiments, wherein the sequence that is inserted into the mammalian genome is a sequence that is exogenous to the mammalian genome.

294. The method of any of the preceding embodiments, wherein the exogenous sequence inserted into the mammalian genome does not naturally occur elsewhere in the mammalian genome.

295. The method of any of the preceding embodiments, wherein the exogenous sequence inserted into the mammalian genome naturally occurs elsewhere in the mammalian genome.

296. The method of any of the preceding embodiments, which operates independently of a DNA template.

297. The method of any of the preceding embodiments, wherein the cell is part of a tissue.

298. The method of any of the preceding embodiments, wherein the mammalian cell is euploid, is not immortalized, is part of an organism, is a primary cell, is non-dividing, is a hepatocyte, or is from a subject having a genetic disease.

299. The method of any of the preceding embodiments, wherein the mammalian cell is in a subject does not have a disease, e.g., for supplementing the genome of the subject.

300. The method of any of the preceding embodiments, wherein the contacting comprises contacting the cell with a plasmid, virus, viral-like particle, virosome, liposome, vesicle, exosome, fusosome, or lipid nanoparticle.

301. The method of any of the preceding embodiments, wherein the contacting comprises using non-viral delivery.

302. The method of any of the preceding embodiments, which comprises comprising contacting the cell with the template RNA (or DNA encoding the template RNA), wherein the template RNA comprises the non-coding strand of an exogenous coding region, wherein optionally the template RNA does not comprise a coding strand of the exogenous coding region,

wherein optionally the delivery comprises non-viral delivery, thereby adding the exogenous coding region to the genome of the cell.

303. The method of any of the preceding embodiments, which comprises contacting the cell with the template RNA (or DNA encoding the template RNA), wherein the template RNA comprises a non-coding strand that is the reverse complement of a sequence encoding the polypeptide, wherein optionally the template RNA does not comprise a coding strand encoding the polypeptide, wherein optionally the delivery comprises non-viral delivery, thereby adding the exogenous coding region to the genome of the cell and expressing the polypeptide in the cell.

304. The method of any of the preceding embodiments, wherein the contacting comprises administering (a) and (b) to a subject, e.g., intravenously.

305. The method of any of the preceding embodiments, wherein the contacting comprises administering a dose of (a) and (b) to a subject at least twice.

306. The method of any of the preceding embodiments, wherein the polypeptide reverse transcribes the template RNA sequence into the target DNA strand, thereby modifying the target DNA strand.

307. The method of any of the preceding embodiments, wherein (a) and (b) are administered separately.

308. The method of any of the preceding embodiments, wherein (a) and (b) are administered together.

309. The method of any of the preceding embodiments, wherein the nucleic acid of (a) is not integrated into the genome of the host cell.

310. The method of any of the preceding embodiments, wherein the tissue is liver, lung, skin, muscle tissue (e.g., skeletal muscle), eye or ocular tissue, or central nervous system.

311. The method of any of the preceding embodiments, wherein the cell is a hematopoietic stem cell (HSC), a T-cell, or a Natural Killer (NK) cell.

312. Any preceding numbered method, wherein the sequence that binds the polypeptide has one or more of the following characteristics:

(a) is at the 3' end of the template RNA;

(b) is at the 5' end of the template RNA;

(b) is a non-coding sequence;

(c) is a structured RNA;

(d) forms at least 1 hairpin loop structures; and/or

(e) is a guide RNA.

313. Any preceding numbered method, wherein the template RNA further comprises a sequence comprising at least 20 nucleotides of at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a target DNA strand.

314. Any preceding numbered method, wherein the template RNA further comprises a sequence comprising at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 nucleotides of at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a target DNA strand.

315. Any preceding numbered method, wherein the sequence comprising at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 nucleotides, or about: 2-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 10-100, or 2-100 nucleotides, of at least 80% identity to a target DNA strand is at the 3' end of the template RNA.

316. Any preceding numbered method, wherein the template RNA further comprises a sequence comprising at least 100 nucleotides of at least 80% identity (e.g., at least 85%, 90%,

95%, 97%, 98%, 99%, 100% identity) to a target DNA strand, e.g., at the 3' end of the template RNA.

317. The method of any of the preceding embodiments, wherein the site in the target DNA strand to which the sequence comprises at least 80% identity is proximal to (e.g., within about: 0-10, 10-20, 20-30, 30-50, or 50-100 nucleotides of) a target site on the target DNA strand that is recognized (e.g., bound and/or cleaved) by the polypeptide comprising the endonuclease.

318. Any preceding numbered method, wherein the target RNA comprises a homology domain that comprises a sequence according to a 3' homology arm of Table 11, or a sequence having at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

319. Any preceding numbered method, wherein the target RNA comprises a homology domain that comprises a sequence according to a 5' homology arm of Table 11, or a sequence having at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

320. Any preceding numbered method, wherein the sequence comprising at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 nucleotides, or about: 2-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 10-100, or 2-100 nucleotides, of at least 80% identity to a target DNA strand is at the 3' end of the template RNA;

optionally wherein the site in the target DNA strand to which the sequence comprises at least 80% identity is proximal to (e.g., within about: 0-10, 10-20, or 20-30 nucleotides of) a target site on the target DNA strand that is recognized (e.g., bound and/or cleaved) by the polypeptide comprising the endonuclease.

321. The method of any of the preceding embodiments, wherein the target site is the site in the human genome that has the closest identity to a native target site of the polypeptide comprising the endonuclease, e.g., wherein the target site in the human genome has at least about: 16, 17, 18, 19, or 20 nucleotides identical to the native target site.

322. Any preceding numbered method, wherein the template RNA has at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand.

323. Any preceding numbered method, wherein the at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand are at the 3' end of the template RNA.

324. Any preceding numbered method, wherein the at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand are at the 5' end of the template RNA.

325. Any preceding numbered method, wherein the template RNA comprises at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand at the 5' end of the template RNA and at least 3, 4, 5, 6, 7, 8, 9, or 10 bases of 100% identity to the target DNA strand at the 3' end of the template RNA.

326. Any preceding numbered method, wherein the heterologous object sequence is between 50-50,000 base pairs (e.g., between 50-40,000 bp, between 500-30,000 bp between 500-20,000 bp, between 100-15,000 bp, between 500-10,000 bp, between 50-10,000 bp, between 50-5,000 bp).

327. Any preceding numbered method, wherein the heterologous object sequence is at least 1, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, or 700 bp.

328. Any preceding numbered method, wherein the heterologous object sequence is at least 715, 750, 800, 950, 1,000, 2,000, 3,000, or 4,000 bp.

329. Any preceding numbered method, wherein the heterologous object sequence is less than 5,000, 10,000, 15,000, 20,000, 30,000, or 40,000 bp.

330. Any preceding numbered method, wherein the heterologous object sequence is less than 700, 600, 500, 400, 300, 200, 150, or 100 bp.

331. Any preceding numbered method, wherein the heterologous object sequence comprises one or more of:

(a) an open reading frame, e.g., a sequence encoding a polypeptide, e.g., an enzyme (e.g., a lysosomal enzyme), a membrane protein, a blood factor, an exon, an intracellular protein (e.g., a cytoplasmic protein, a nuclear protein, an organellar protein such as a mitochondrial protein or lysosomal protein), an extracellular protein, a structural protein, a signaling protein, a regulatory protein, a transport protein, a sensory protein, a motor protein, a defense protein, a storage protein, or an immune receptor protein (e.g., a chimeric antigen receptor (CAR) protein, a T cell receptor, a B cell receptor), or an antibody;

(b) a non-coding and/or regulatory sequence, e.g., a sequence that binds a transcriptional modulator, e.g., a promoter, an enhancer, an insulator;

(c) a splice acceptor site;

(d) a polyA site;

(e) an epigenetic modification site; or

(f) a gene expression unit.

332. Any preceding numbered method, wherein the target DNA is a genomic safe harbor (GSH) site.

333. Any preceding numbered method, wherein the target DNA is a genomic Natural Harbor™ site.

334. Any preceding numbered method, which results in insertion of the heterologous object sequence into the genome at an average copy number of at least 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.5, 3, 4, or 5 copies per genome.

335. Any preceding numbered method, which results in about 25-100%, 50-100%, 60-100%, 70-100%, 75-95%, 80%-90%, of integrants into a target site in the genome being non-truncated, as measured by an assay described herein, e.g., an assay of Example 6 of PCT Application No. PCT/US2019/048607.

336. Any preceding numbered method, which results in insertion of the heterologous object sequence only at one target site in the genome of the cell.

337. Any preceding numbered method, which results in insertion of the heterologous object sequence into a target site in a cell, wherein the inserted heterologous sequence comprises less than 10%, 5%, 2%, 1%, 0.5%, 0.2%, or 0.1% mutations (e.g., SNPs or one or more deletions, e.g., truncations or internal deletions) relative to the heterologous sequence prior to insertion, e.g., as measured by an assay of Example 12 of PCT Application No. PCT/US2019/048607.

338. Any preceding numbered method, which results in insertion of the heterologous object sequence into a target site in a plurality of cells, wherein less than 10%, 5%, 2%, or 1% of copies of the inserted heterologous sequence comprise a mutation (e.g., a SNP or a deletion, e.g., a truncation or an internal deletion), e.g., as measured by an assay of Example 12 of PCT Application No. PCT/US2019/048607.

339. Any preceding numbered method, which results in insertion of the heterologous object sequence into a target cell genome, and wherein the target cell does not show upregulation of p53, or shows upregulation of p53 by less than 50%, 25%, 10%, 5%, 2%, or 1%, e.g., wherein upregulation of p53 is measured by p53 protein level, e.g., according to the method described in Example 30, or by the level of p53 phosphorylated at Ser15 and Ser20.

340. Any preceding numbered method, which results in insertion of the heterologous object sequence into a target cell genome, and wherein the target cell does not show upregulation of any DNA repair genes and/or tumor suppressor genes, or wherein no DNA repair gene and/or tumor suppressor gene is upregulated by more than 50%, 25%, 10%, 5%, 2%, or 1%, e.g., wherein upregulation is measured by RNA-seq.

341. Any preceding numbered method, which results in insertion of the heterologous object sequence into the target site (e.g., at a copy number of 1 insertion or more than one insertion) in about 1-80% of cells in a population of cells contacted with the system, e.g., about: 1-10%, 10-

20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, or 70-80% of cells, e.g., as measured using single cell ddPCR, e.g., as described in Example 17.

342. Any preceding numbered method, which results in insertion of the heterologous object sequence into the target site at a copy number of 1 insertion in about 1-80% of cells in a population of cells contacted with the system, e.g., about: 1-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, or 70-80% of cells, e.g., as measured using colony isolation and ddPCR, e.g., as described in Example 18.

343. Any preceding numbered method, which results in insertion of the heterologous object sequence into the target site (on-target insertions) at a higher rate that insertion into a non-target site (off-target insertions) in a population of cells, wherein the ratio of on-target insertions to off-target insertions is greater than 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1. 90:1, 100:1, 200:1, 500:1, or 1,000:1, e.g., using an assay of Example 11 of PCT Application No. PCT/US2019/048607.

344. Any above-numbered method, which results in insertion of a heterologous object sequence in the presence of an inhibitor of a DNA repair pathway (e.g., SCR7, a PARP inhibitor), or in a cell line deficient for a DNA repair pathway (e.g., a cell line deficient for the nucleotide excision repair pathway or the homology-directed repair pathway).

345. The method of any of the preceding embodiments, wherein the cell has decreased Rad51 repair pathway activity, decreased expression of Rad51 or a component of the Rad51 repair pathway, or does not comprise a functional Rad51 repair pathway, e.g., does not comprise a functional Rad51 gene, e.g., comprises a mutation (e.g., deletion) inactivating one or both copies of the Rad51 gene or another gene in the Rad51 repair pathway.

346. Any preceding numbered system, formulated as a pharmaceutical composition.

347. Any preceding numbered system, disposed in a pharmaceutically acceptable carrier (e.g., a vesicle, a liposome, a natural or synthetic lipid bilayer, a lipid nanoparticle, an exosome).

348. Any preceding numbered method, which results in a plurality of (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, or 10) insertions of a heterologous object sequence into a target cell genome.

349. The method of any of the preceding embodiments, wherein the plurality of insertions occur simultaneously or sequentially.

350. Any preceding numbered method, which results in a plurality of (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, or 10) deletions of a heterologous object sequence into a target cell genome.

351. The method of any of the preceding embodiments, wherein the plurality of deletions occur simultaneously or sequentially.

352. Any preceding numbered method, which results in a plurality of (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, or 10) base changes of a heterologous object sequence into a target cell genome.

353. The method of any of the preceding embodiments, wherein the plurality of base changes occur simultaneously or sequentially.

354. Any preceding numbered method, comprising contacting the cell with a plurality of distinct template RNAs each comprising a heterologous object sequence.

355. The method ofany of the preceding embodiments, wherein the distinct template RNAs comprise at least two (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, or 10) distinct heterologous object sequences.

356. The method of any of the preceding embodiments, wherein the at least two distinct heterologous object sequences each comprise a distinct payload.

357. The method of any of the preceding embodiments, wherein the at least two distinct heterologous object sequences each comprise the same payload.

358. Any preceding numbered embodiment, in which the target cell of a Gene Writing system has been previously modified at one or more loci.

359. Any preceding numbered embodiment, wherein the previously edited cell is a T-cell.

360. Any preceeding numbered embodiment, wherein the one or more previous modifications are selected from gene knockouts, e.g., of an endogenous TCR (e.g., TRAC, TRBC), HLA Class I (B2M), PD1, CD52, CTLA-4, TIM-3, LAG-3, or DGK.

361. Any preceding numbered embodiment, wherein the heterologous object sequence comprises a TCR or a CAR.

362. A method of making a system for modifying the genome of a mammalian cell, comprising:

a) providing a template RNA comprising (i) a sequence that binds a polypeptide comprising a reverse transcriptase domain and an endonuclease domain, the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a 5' UTR sequence or a 3' UTR sequence of a sequence of an element of Table X; and (ii) a heterologous object sequence.

363. The method of any of the preceding embodiments, further comprising:

b) treating the template RNA to reduce secondary structure, e.g., heating the template RNA, e.g., to at least 70, 75, 80, 85, 90, or 95 C, and/orc) subsequently cooling the template RNA, e.g., to a temperature that allows for secondary structure, e.g, to less than or equal to 37, 30, 25, or 20 C.

364. A method of making a system for modifying DNA (e.g., as described herein), the method comprising:

(a) providing a template nucleic acid (e.g., a template RNA or DNA) comprising a heterologous homology sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence comprised in a target DNA molecule, and/or
(b) providing a polypeptide of the system (e.g., comprising a DNA-binding domain (DBD) and/or an endonuclease domain) comprising a heterologous targeting domain that binds specifically to a sequence comprised in the target DNA molecule.

365. The method of any of the preceding embodiments, wherein:
(a) comprises introducing into the template nucleic acid (e.g., a template RNA or DNA) a heterologous homology sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to the sequence comprised in a target DNA molecule, and/or
(b) comprises introducing into the polypeptide of the system (e.g., comprising a DNA-binding domain (DBD) and/or an endonuclease domain) the heterologous targeting domain that binds specifically to a sequence comprised in the target DNA molecule.

366. The method of any of the preceding embodiments, wherein the introducing of (a) comprises inserting the homology sequence into the template nucleic acid.

367. The method of any of the preceding embodiments, wherein the introducing of (a) comprises replacing a segment of the template nucleic acid with the homology sequence.

368. The method of any of the preceding embodiments, wherein the introducing of (a) comprises mutating one or more nucleotides (e.g., at least 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, or 100 nucleotides) of the template nucleic acid, thereby producing a segment of the template nucleic acid having the sequence of the homology sequence.

369. The method of any of the preceding embodiments, wherein the introducing of (b) comprises inserting the amino acid sequence of the targeting domain into the amino acid sequence of the polypeptide.

370. The method of any of the preceding embodiments, wherein the introducing of (b) comprises inserting a nucleic acid sequence encoding the targeting domain into a coding sequence of the polypeptide comprised in a nucleic acid molecule.

371. The method of any of the preceding embodiments, wherein the introducing of (b) comprises replacing at least a portion of the polypeptide with the targeting domain.

372. The method of any of the preceding embodiments, wherein the introducing of (a) comprises mutating one or more amino acids (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, or more amino acids) of the polypeptide.

373. A method for modifying a target site in genomic DNA in a cell, the method comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds the target site (e.g., a second strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target homology domain,

wherein:

(i) the polypeptide comprises a heterologous targeting domain (e.g., in the DBD or the endonuclease domain) that binds specifically to a sequence comprised in or adjacent to the target site of the genomic DNA; and/or

(ii) the template RNA comprises a heterologous homology sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence comprised in or adjacent to the target site of the genomic DNA;

thereby modifying the target site in genomic DNA in a cell.

374. A method of making a system for modifying the genome of a mammalian cell, comprising:

a) providing a template RNA comprising (i) a sequence that binds a polypeptide comprising a reverse transcriptase domain and an endonuclease domain, the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to: (i) the nucleotides located 5' relative to a start codon of a sequence of an element of Table 3B, Table 10, or Table X (e.g., comprising a retrotransposase-binding region), or (ii) the nucleotides located 3' relative to a stop codon of a sequence of an element of Table 3B, Table 10, or Table X (e.g., comprising a retrotransposase-binding region)

375. The method of any of the preceding embodiments, further comprising:

b) treating the template RNA to reduce secondary structure, e.g., heating the template RNA, e.g., to at least 70, 75, 80, 85, 90, or 95 C, and/or

c) subsequently cooling the template RNA, e.g., to a temperature that allows for secondary structure, e.g, to less than or equal to 37, 30, 25, or 20 C.

376. The method of any of the preceding embodiments, wherein the system is the system of any of the preceding embodiments.

377. The method of any of the preceding embodiments, which further comprises contacting the template RNA with a polypeptide that comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, or with a nucleic acid (e.g., RNA) encoding the polypeptide.

378. The method of any of the preceding embodiments, which further comprises contacting the template RNA with a cell.

379. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes a therapeutic polypeptide.

380. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes a mammalian (e.g., human) polypeptide, or a fragment or variant thereof.

381. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes an enzyme (e.g., a lysosomal enzyme), a blood factor (e.g., Factor I, II, V, VII, X, XI, XII or XIII), a membrane protein, an exon, an intracellular protein (e.g., a cytoplasmic protein, a nuclear protein, an organellar protein such as a mitochondrial protein or lysosomal protein), an extracellular protein, a structural protein, a signaling protein, a regulatory protein, a transport protein, a sensory protein, a motor protein, a defense protein, a storage protein, an immune receptor protein (e.g., a chimeric antigen receptor (CAR) protein, a T cell receptor, a B cell receptor), or an antibody.

382. The system or method of any of the preceding embodiments, wherein the heterologous object sequence comprises a tissue specific promoter or enhancer.

383. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes a polypeptide of greater than 250, 300, 400, 500, or 1,000 amino acids, and optionally up to 1300 amino acids.

384. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes a fragment of a mammalian gene but does not encode the full mammalian gene, e.g., encodes one or more exons but does not encode a full-length protein.

385. The system or method of any of the preceding embodiments, wherein the heterologous object sequence encodes one or more introns.

386. The system or method of any of the preceding embodiments, wherein the heterologous object sequence is other than a GFP, e.g., is other than a fluorescent protein or is other than a reporter protein.

387. The system or method of any of the preceding embodiments, wherein the polypeptide has an activity at 37°C that is no less than 70%, 75%, 80%, 85%, 90%, or 95% of its activity at 25°C under otherwise similar conditions.

388. The system or method of any of the preceding embodiments, wherein the nucleic acid encoding the polypeptide and the template RNA or a nucleic acid encoding the template RNA are separate nucleic acids.

389. The system or method of any of the preceding embodiments, wherein the template RNA does not encode an active reverse transcriptase, e.g., comprises an inactivated mutant reverse transcriptase, e.g., as described in Example 1 or 2, or does not comprise a reverse transcriptase sequence.

390. The system or method of any of the preceding embodiments, wherein the template RNA comprises one or more chemical modifications.

391. The system or method of any of the preceding embodiments, wherein the heterologous object sequence is disposed between the promoter and the sequence that binds the polypeptide.

392. The system or method of any of the preceding embodiments, wherein the promoter is disposed between the heterologous object sequence and the sequence that binds the polypeptide.

393. The system or method of any of the preceding embodiments, wherein the heterologous object sequence comprises an open reading frame (or the reverse complement thereof) in a 5' to 3' orientation on the template RNA.

394. The system or method of any of the preceding embodiments, wherein the heterologous object sequence comprises an open reading frame (or the reverse complement thereof) in a 3' to 5' orientation on the template RNA.

395. The system or method of any of the preceding embodiments, wherein the polypeptide comprises (a) a reverse transcriptase domain and (b) an endonuclease domain, wherein at least one of (a) or (b) is heterologous.

396. The system or method of any of the preceding embodiments, wherein the polypeptide comprises (a) a target DNA binding domain, (b) a reverse transcriptase domain and (c) an endonuclease domain, wherein at least one of (a), (b) or (c) is heterologous.

397. A substantially pure polypeptide comprising (a) a reverse transcriptase domain and (b) a heterologous endonuclease domain; wherein one or both of (a) or (b) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

398. A substantially pure polypeptide comprising (a) a reverse transcriptase domain and (b) a heterologous endonuclease domain; wherein one or both of (a) or (b) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

399. A substantially pure polypeptide comprising (a) a reverse transcriptase domain and (b) a heterologous endonuclease domain; wherein one or both of (a) or (b) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

400. A substantially pure polypeptide comprising (a) a reverse transcriptase domain and (b) a heterologous endonuclease domain; wherein one or both of (a) or (b) have an amino acid sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

401. A substantially pure polypeptide comprising (a) a reverse transcriptase domain encoded by a first sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and (b) a endonuclease domain encoded by a second sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; wherein the first sequence and the second sequences are selected from elements of different rows of Table 3B, Table 10, Table 11, or Table X.

402. A substantially pure polypeptide comprising (a) a reverse transcriptase domain encoded by a first sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, and (b) a endonuclease domain encoded by a second sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; wherein the first sequence and the second sequences are selected from elements of different rows of Table 3B, Table 10, Table 11, or Table X.

403. A substantially pure polypeptide comprising (a) a reverse transcriptase domain of a first sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and (b) a endonuclease domain of a second sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; wherein the first sequence and the second sequences are selected from elements of different rows of Table 3B, Table 10, Table 11, or Table X.

404. A substantially pure polypeptide comprising (a) a reverse transcriptase domain of a first sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, and (b) a endonuclease domain of a second sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100

nucleotide differences therefrom; wherein the first sequence and the second sequences are selected from elements of different rows of Table 3B, Table 10, Table 11, or Table X.

405. The substantially pure polypeptide of any of the preceding embodiments, further comprising a target DNA binding domain encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

406. The substantially pure polypeptide of any of the preceding embodiments, further comprising a target DNA binding domain encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

407. The substantially pure polypeptide of any of the preceding embodiments, further comprising a target DNA binding domain of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

408. The substantially pure polypeptide of any of the preceding embodiments, further comprising a target DNA binding domain of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

409. A substantially pure polypeptide comprising (a) a reverse transcriptase domain and (b) a heterologous endonuclease domain; wherein (a) has an amino acid sequence listed in Table 3B, Table 10, Table 11, or Table Z1 or Z2 or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, optionally wherein (b) has an amino acid sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

410. A substantially pure polypeptide comprising (a) a reverse transcriptase domain and (b) a heterologous endonuclease domain; wherein (a) has an amino acid sequence listed in Table 3B, Table 10, Table 11, or Table Z1 or Z2 or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, optionally wherein (b) has an amino acid sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

411. The substantially pure polypeptide of any of the preceding embodiments, further comprising a target DNA binding domain comprising an amino acid sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

412. The substantially pure polypeptide of any of the preceding embodiments, further comprising a target DNA binding domain comprising an amino acid sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

413. A substantially pure polypeptide comprising (a) a target DNA binding domain encoded by a first sequence listed in Table 3B, Table 10, Table 11, or Table X, (b) a reverse transcriptase domain encoded by a second sequence listed in Table 3B, Table 10, Table 11, or Table X, and (c) an endonuclease domain encoded by a third sequence listed in Table 3B, Table 10, Table 11, or Table X; wherein:

(i) the first sequence and the second sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X;

(ii) the first sequence and the third sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X;

(iii) the second sequence and the third sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X; or

(iv) the first sequence, second sequence, and third sequence are each selected from different rows of Table 3B, Table 10, Table 11, or Table X.

414. A substantially pure polypeptide comprising (a) a target DNA binding domain of a first sequence listed in Table 3B, Table 10, Table 11, or Table X, (b) a reverse transcriptase domain of a second sequence listed in Table 3B, Table 10, Table 11, or Table X, and (c) an endonuclease domain of a third sequence listed in Table 3B, Table 10, Table 11, or Table X; wherein:

(i) the first sequence and the second sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X;

(ii) the first sequence and the third sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X;

(iii) the second sequence and the third sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X; or

(iv) the first sequence, second sequence, and third sequence are each selected from different rows of Table 3B, Table 10, Table 11, or Table X.

415. A substantially pure polypeptide comprising (a) a target DNA binding domain, e.g., comprising a first amino acid sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, (b) a reverse transcriptase domain comprising a second amino acid sequence listed in Table Z1 or Z2 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and (c) an endonuclease domain comprising a third amino acid sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto;
optionally wherein the elements of the first amino acid sequence and the third amino acid sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X.

416. A substantially pure polypeptide comprising (a) a target DNA binding domain, e.g., comprising a first amino acid sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, (b) a reverse transcriptase domain comprising a second amino acid

sequence listed in Table Z1 or Z2 or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, and (c) an endonuclease domain comprising a third amino acid sequence of an element listed in Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom;

optionally wherein the elements of the first amino acid sequence and the third amino acid sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X.

417. A substantially pure polypeptide comprising (a) a target DNA binding domain comprising a first amino acid sequence, (b) a reverse transcriptase domain comprising a second amino acid sequence, and (c) an endonuclease domain comprising a third amino acid sequence; wherein the first, second, and third amino acid sequences are each encoded by a sequence listed in Table 3B, Table 10, Table 11, or Table X or comprise an amino acid sequence listed in Table Z1 or an amino acid sequence of a domain listed in Table Z2.

418. The substantially pure polypeptide of any of the preceding embodiments, wherein the first amino acid sequence is encoded by a sequence listed in Table 3B, Table 10, Table 11, or Table X.

419. The substantially pure polypeptide of any of the preceding embodiments, wherein the first amino acid sequence comprises an amino acid sequence listed in Table Z1 or Z2.

420. The substantially pure polypeptide of any of the preceding embodiments, wherein the second amino acid sequence is encoded by a sequence listed in Table 3B, Table 10, Table 11, or Table X.

421. The substantially pure polypeptide of any of the preceding embodiments, wherein the second amino acid sequence comprises an amino acid sequence listed in Table Z1 or Z2.

422. The substantially pure polypeptide of any of the preceding embodiments, wherein the third amino acid sequence is encoded by a sequence listed in Table 3B, Table 10, Table 11, or Table X.

423. The substantially pure polypeptide of any of the preceding embodiments, wherein the third amino acid sequence comprises an amino acid sequence listed in Table Z1 or Z2.

424. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (a) a reverse transcriptase domain and (b) an endonuclease domain, wherein one or both of (a) or (b) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and wherein at least one of (a) or (b) is heterologous to the other.

425. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (a) a reverse transcriptase domain and (b) an endonuclease domain, wherein one or both of (a) or (b) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, and wherein at least one of (a) or (b) is heterologous to the other.

426. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide is a fusion protein comprising (a) a reverse transcriptase domain encoded by a first sequence of an element listed in Table 3B, Table 10, Table 11, or Table X, and (b) a endonuclease domain encoded by a second sequence of an element listed in Table 3B, Table 10, Table 11, or Table X; wherein the first sequence and the second sequences are selected from elements of different rows of Table 3B, Table 10, Table 11, or Table X.

427. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (a) a target DNA binding domain, (b) a reverse transcriptase domain and (c) an endonuclease domain, wherein at least one (e.g., 1, 2, or all) of (a), (b) or (c) comprises an amino

acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and wherein at least one of (a), (b) or (c) is heterologous to the other.

428. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (a) a target DNA binding domain, (b) a reverse transcriptase domain and (c) an endonuclease domain, wherein at least one (e.g., 1, 2, or all) of (a), (b) or (c) comprises an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, and wherein at least one of (a), (b) or (c) is heterologous to the other.

429. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide is a fusion protein comprising (a) a target DNA binding domain encoded by a first sequence listed in Table 3B, Table 10, Table 11, or Table X, (b) a reverse transcriptase domain encoded by a second sequence listed in Table 3B, Table 10, Table 11, or Table X, and (c) an endonuclease domain encoded by a third sequence listed in Table 3B, Table 10, Table 11, or Table X; wherein:

(i) the first sequence and the second sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X;

(ii) the first sequence and the third sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X;

(iii) the second sequence and the third sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X; or

(iv) the first sequence, second sequence, and third sequence are each selected from different rows of Table 3B, Table 10, Table 11, or Table X.

430. Any polypeptide of any of the preceding embodiments, wherein the reverse transcriptase domain has at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a reverse transcriptase domain encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

431. Any polypeptide of any of the preceding embodiments, wherein the endonuclease domain has at least 80% identity e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity, to a endonuclease domain encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

432. Any polypeptide or method of any of the preceding embodiments, wherein the DNA binding domain has at least 80% identity e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity, to a DNA binding domain encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

433. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (a) a reverse transcriptase domain and (b) an endonuclease domain, wherein one or both of (a) or (b) have an amino acid sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and wherein at least one of (a) or (b) is heterologous to the other.

434. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (a) a reverse transcriptase domain and (b) an endonuclease domain, wherein one or both of (a) or (b) have an amino acid sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, and wherein at least one of (a) or (b) is heterologous to the other.

435. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide is a fusion protein comprising (a) a reverse transcriptase domain of a first sequence listed in Table 3B, Table 10, Table 11, or Table X, and (b) an endonuclease domain of a second sequence listed in Table 3B, Table 10, Table 11, or Table X; wherein the first sequence and the second sequences are selected from different rows of Table 3B, Table 10, Table 11, or Table X.

436. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (a) a target DNA binding domain, (b) a reverse transcriptase domain and (c) an endonuclease domain, wherein at least one (e.g., 1, 2, or all) of (a), (b) or (c) comprises an amino acid sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and wherein at least one of (a), (b) or (c) is heterologous to the other.

437. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (a) a target DNA binding domain, (b) a reverse transcriptase domain and (c) an endonuclease domain, wherein at least one (e.g., 1, 2, or all) of (a), (b) or (c) comprises an amino acid sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom, and wherein at least one of (a), (b) or (c) is heterologous to the other.

438. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide is a fusion protein comprising (a) a target DNA binding domain of a first sequence listed in Table 3B, Table 10, Table 11, or Table X, (b) a reverse transcriptase domain of a second sequence listed in Table 3B, Table 10, Table 11, or Table X, and (c) an endonuclease domain of a third sequence listed in Table 3B, Table 10, Table 11, or Table X; wherein:

(i) the first sequence and the second sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X;

(ii) the first sequence and the third sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X;

(iii) the second sequence and the third sequence are selected from different rows of Table 3B, Table 10, Table 11, or Table X; or

(iv) the first sequence, second sequence, and third sequence are each selected from different rows of Table 3B, Table 10, Table 11, or Table X.

439. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain, wherein the RT domain has a sequence of Table

3B, Table 10, Table 11, or Table X, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

440. Any polypeptide of any of the preceding embodiments, wherein the reverse transcriptase domain has at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a reverse transcriptase domain of a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

441. Any polypeptide of any of the preceding embodiments, wherein the endonuclease domain has at least 80% identity e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity, to an endonuclease domain of a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

442. Any polypeptide or method of any of the preceding embodiments, wherein the DNA binding domain has at least 80% identity e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity, to a DNA binding domain of a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

443. A nucleic acid encoding the polypeptide of any of the preceding embodiments.

444. A vector comprising the nucleic acid of any of the preceding embodiments.

445. A host cell comprising the nucleic acid of any of the preceding embodiments.

446. A host cell comprising the polypeptide of any of the preceding embodiments.

447. A host cell comprising the vector of any of the preceding embodiments.

448. A host cell (e.g., a human cell) comprising:

a heterologous object sequence (e.g., a sequence encoding a therapeutic polypeptide) at a target site in a chromosome, and one or both of:

(a) one or both of an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of Table 10 or column 5 of Table 3A or 3B) on one side (e.g., upstream) of the heterologous object sequence, and an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of Table 10 or column 6 of Table 3A or 3B) on the other side (e.g., downstream) of the heterologous object sequence; and/or

(b) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, wherein one or both of (i) or (ii) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

449. A host cell (e.g., a human cell) comprising:

a heterologous object sequence (e.g., a sequence encoding a therapeutic polypeptide) at a target site in a chromosome, and one or both of:

(a) one or both of an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of Table 10 or column 5 of Table 3A or 3B) on one side (e.g., upstream) of the heterologous object sequence, and an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of Table 10 or column 6 of Table 3A or 3B) on the other side (e.g., downstream) of the heterologous object sequence; and/or

(b) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, wherein one or both of (i) or (ii) have an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

450. A host cell (e.g., a human cell) comprising:

a heterologous object sequence (e.g., a sequence encoding a therapeutic polypeptide) at a target site in a chromosome, and one or both of:

(a) one or both of an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of Table 10 or column 5 of Table 3A or 3B) on one side (e.g., upstream) of the heterologous object sequence, and an untranslated region (e.g., a retrotransposon untranslated

sequence, e.g., a sequence of Table 10 or column 6 of Table 3A or 3B) on the other side (e.g., downstream) of the heterologous object sequence; and/or

(b) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, wherein one or both of (i) or (ii) have an amino acid sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

451. A host cell (e.g., a human cell) comprising:

a heterologous object sequence (e.g., a sequence encoding a therapeutic polypeptide) at a target site in a chromosome, and one or both of:

(a) one or both of an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of column 5 of Table 10 or Table 3A or 3B) on one side (e.g., upstream) of the heterologous object sequence, and an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of column 6 of Table 10 or Table 3A or 3B) on the other side (e.g., downstream) of the heterologous object sequence; and/or

(b) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, wherein one or both of (i) or (ii) have an amino acid sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom.

452. The host cell of any of the preceding embodiments, comprising: (i) a heterologous object sequence (e.g., a sequence encoding a therapeutic polypeptide) at a target site in a chromosome, wherein the target locus is a Natural Harbor™ site, e.g., a site of Table 4 herein.

453. The host cell of any of the preceding embodiments, which further comprises (ii) one or both of an untranslated region 5' of the heterologous object sequence, and an untranslated region 3' of the heterologous object sequence.

454. The host cell of any of the preceding embodiments, which further comprises (ii) one or both of an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of Table 10 or column 5 of Table 3A or 3B) on one side (e.g., upstream) of the heterologous object sequence, and an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of Table 10 or column 6 of Table 3A or 3B) on the other side (e.g., downstream) of the heterologous object sequence.

455. The host cell of any of the preceding embodiments, which comprises a heterologous object sequence at only the target site.

456. A pharmaceutical composition, comprising any preceding numbered system, nucleic acid, polypeptide, or vector; and a pharmaceutically acceptable excipient or carrier.

457. The pharmaceutical composition of any of the preceding embodiments, wherein the pharmaceutically acceptable excipient or carrier is selected from a vector (e.g., a viral or plasmid vector), a vesicle (e.g., a liposome, an exosome, a natural or synthetic lipid bilayer), a fusosome, a lipid nanoparticle.

458. A polypeptide of any of the preceding embodiments, wherein the polypeptide further comprises a nuclear localization sequence.

459. A template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain.

460. The template RNA of any of the preceding embodiments, wherein the template RNA comprises (i).

461. The template RNA of any of the preceding embodiments, wherein the template RNA comprises (ii).

462. A template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) a sequence that specifically binds an RT domain of a polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain.

463. The template RNA of any of the preceding embodiments, wherein the RT domain comprises a sequence selected of Table 3B, Table 10, Table 11, or Table X or a sequence that has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

464. The template RNA of any of the preceding embodiments, further comprising (v) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide (e.g., the same polypeptide comprising the RT domain).

465. The template RNA of any of the preceding embodiments, wherein the sequence of (ii) specifically binds an RT domain of Table 3B, Table 10, Table 11, or Table X or an RT domain sequence that has at least 70, 75, 80, 85, 90, 95, or 99% identity thereto.

466. The template RNA of any of the preceding embodiments, wherein the sequence that specifically binds the RT domain is a sequence of Table 10, Table 11, or Table X, 3A, or 3B, or a sequence having at least 70, 75, 80, 85, 90, 95, or 99% identity thereto.

467. A template RNA (or DNA encoding the template RNA) comprising from 5' to 3': (ii) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide, (i) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (iii) a heterologous object sequence, and (iv) a 3' homology domain.

468. A template RNA (or DNA encoding the template RNA) comprising from 5' to 3': (iii) a heterologous object sequence, (iv) a 3' homology domain, (i) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), and (ii) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide.

469. The system or template RNA of any of the preceding embodiments, wherein the template RNA, first template RNA, or second template RNA comprises a sequence that specifically binds the RT domain.

470. The system or template RNA of any of the preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (i) and (ii).

471. The system or template RNA of any of the preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (ii) and (iii).

472. The system or template RNA of any of the preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (iii) and (iv).

473. The system or template RNA of any of the preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (iv) and (i).

474. The system or template RNA of any of the preceding embodiments, wherein the sequence that specifically binds the RT domain is disposed between (i) and (iii).

475. A system for modifying DNA, comprising:

(a) a first template RNA (or DNA encoding the first template RNA) comprising (i) sequence that binds an endonuclease domain, e.g., a nickase domain, and/or a DNA-binding domain (DBD) of a polypeptide, and (ii) a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (e.g., wherein the first RNA comprises a gRNA);

(b) a second template RNA (or DNA encoding the second template RNA) comprising (i) a sequence that specifically binds a reverse transcriptase (RT) domain of a polypeptide (e.g., the polypeptide of (a)), (ii) a target site binding sequence (TSBS), and (iii) an RT template sequence.

476. The system of any of the preceding embodiments, wherein the nucleic acid encoding the first template RNA and the nucleic acid encoding the second template RNA are two separate nucleic acids.

477. The system of any of the preceding embodiments, wherein the nucleic acid encoding the first template RNA and the nucleic acid encoding the second template RNA are part of the same nucleic acid molecule, e.g., are present on the same vector.

478. A method of modifying a target DNA strand in a cell, tissue or subject, comprising administering any preceding numbered system to the cell, tissue or subject, thereby modifying the target DNA strand.

479. Any preceding numbered embodiment, wherein the sequence of an element of Table X is selected from Vingi-1 EE, BovB, AviRTE_Brh, Penelope_SM, and Utopia_Dyak retrotransposases.

480. Any preceding numbered embodiment, wherein the template RNA comprises the 5' UTR and 3' UTR of the same sequence of an element of Table 3B, Table 10, or Table X.

481. Any preceeding numbered embodiment, wherein the sequence of an element of Table X comprises Vingi-1 EE retrotransposase and wherein the template RNA comprises the 5' UTR and 3' UTR of Vingi-1EE.

482. Any preceding numbered embodiment, wherein the sequence of an element of Table X belongs to the restriction-endonuclease-like (RLE) clade.

483. Any preceding numbered embodiment, wherein the sequence of an element of Table X belongs to the apurinic-endonuclease-like (APE) clade.

484. Any preceding numbered embodiment, wherein the sequence of an element of Table X belongs to the Penelope-like element (PLE) clade, optionally wherein the sequence of the element of Table X comprises a GIY-YIG domain (e.g., a GIY-YIG endonuclease domain).

485. Any preceding numbered embodiment, wherein the sequence of an element of Table X belongs to a clade selected from the CRE, NeSL, R4, R2, Hero, L1, RTE (e.g., AviRTE_Brh or BovB), I, Jockey, CR1, Rex1, RandI/Dualen, Penelope (e.g., Penelope_SM), Tx1, RTEX, Crack, Nimb, Proto1, Proto2, RTETP, L2, Tad1, Loa, Ingi, Outcast, R1, Daphne, L2A, L2B, Ambal, Vingi (e.g., Vingi-1_EE), and Kiri clades.

486. Any preceding numbered embodiment, wherein the target DNA binding domain is heterologous relative to one or more other domains of the polypeptide (e.g., a reverse transcriptase domain and/or an endonuclease domain).

487. Any preceeding numbered embodiment, wherein the heterologous target DNA binding domain comprises a Cas9, Cas9 nickase, dCas9, zinc finger, or TAL domain.

488. Any preceeding numbered embodiment, wherein the heterologous target DNA binding domain comprises a Cas domain according to Table 9 or Table 37, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

489. Any preceeding numbered embodiment, wherein the heterologous target DNA binding domain comprises an N-terminal dCas9 domain.

490. Any preceding numbered embodiment, wherein the system further comprises a guide RNA (e.g., a U6-driven gRNA) comprising at least 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides of homology to a target DNA sequence.

491. Any preceding numbered embodiment, wherein the template RNA further comprises a gRNA region comprising at least 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides of homology to a target DNA sequence.

492. Any preceeding numbered embodiment, wherein the gRNA sequence is at the 5' end of the template.

493. Any preceeding numbered embodiment, wherein the gRNA sequence is at the 3' end of the template.

494. Any preceeding numbered embodiment, wherein the gRNA sequence comprises a scaffold capable of recruiting Cas9.

495. Any preceeding numbered embodiment, wherein the gRNA sequence comprises a homology domain, e.g., as described herein.

496. Any preceding numbered embodiment, wherein the endonuclease domain is heterologous relative to one or more other domains of the polypeptide (e.g., a reverse transcriptase domain and/or a target DNA binding domain).

497. Any preceding numbered embodiment, wherein the heterologous endonuclease domain comprises a Cas9, Cas9 nickase, or FokI domain.

498. Any preceding numbered embodiment, wherein the polypeptide comprises an RNase H domain.

499. Any preceding numbered embodiment, wherein the polypeptide does not comprise an RNase H domain or comprises an inactivated RNase H domain.

500. Any preceding numbered embodiment, wherein the nucleic acid encoding the polypeptide further comprises a second open reading frame.

501. Any preceding numbered embodiment, wherein the nucleic acid encoding the polypeptide comprises a 2A sequence, e.g., positioned between an ORF1 sequence and an ORF2 sequence,

optionally wherein the 2A sequence is selected from T2A (EGRGSLLTCGDVEENPGP), P2A (ATNFSLLKQAGDVEENPGP), E2A (QCTNYALLKLAGDVESNPGP), or F2A (VKQTLNFDLLKLAGDVESNPGP).

502. Any preceding numbered embodiment, wherein the polypeptide comprises an intein.

503. Any preceding numbered embodiment, wherein the system comprises an intein (e.g., comprised in a second polypeptide).

504. Any preceding numbered embodiment, wherein the polypeptide is encoded by two or more separate open reading frames each encoding a polypeptide fragment.

505. Any preceding numbered embodiment, wherein the intein (e.g., a trans-splicing intein) joins the two or more polypeptide fragments to form the polypeptide.

506. Any preceding numbered embodiment, wherein the system comprises: (i) a first polypeptide fragment comprising at least one of a reverse transcriptase domain, an endonuclease domain, and a target DNA binding domain, and (ii) a second polypeptide fragment comprising at least one of a reverse transcriptase domain, an endonuclease domain, and a target DNA binding domain, wherein the first polypeptide fragment does not comprise the same type of domain as the second polypeptide fragment.

507. Any preceding numbered embodiment, wherein:

(a) the first polypeptide fragment comprises a reverse transcriptase domain and the second polypeptide fragment comprises an endonuclease domain, optionally wherein the first polypeptide fragment further comprises a target DNA binding domain or wherein the second polypeptide fragment further comprises a target DNA binding domain;

(b) the first polypeptide fragment comprises a reverse transcriptase domain and the second polypeptide fragment comprises a target DNA binding domain, optionally wherein the first polypeptide fragment further comprises an endonuclease domain or wherein the second polypeptide fragment further comprises an endonuclease domain; or

(a) the first polypeptide fragment comprises an endonuclease domain and the second polypeptide fragment comprises a target DNA binding domain, optionally wherein the first polypeptide fragment further comprises a reverse transcriptase domain or wherein the second polypeptide fragment further comprises a reverse transcriptase domain.

508. Any preceding numbered embodiment, wherein the intein joins the first polypeptide fragment to the second polypeptide to form the polypeptide.

509. Any preceding numbered embodiment, wherein the intein induces fusion of:
(i) a reverse transcriptase domain to an endonuclease domain,
(ii) a reverse transcriptase domain to a target DNA binding domain, or
(iii) an endonuclease domain to a target DNA binding domain.

510. Any preceding numbered embodiment, wherein the intein is heterologous relative to one or more (e.g., 1, 2, or all) of the reverse transcriptase domain, the endonuclease domain, and the target DNA binding domain.

511. Any preceding numbered embodiment, wherein the intein is a split intein.

512. Any of the preceding embodiments, wherein the DNA encoding the polypeptide comprises a plasmid, minicircle, a Doggybone DNA (dbDNA), or a ceDNA.

513. Any of the preceding embodiments, wherein the RNA encoding the polypeptide comprises one or more of the following: a cap region, a poly-A tail, and/or a chemical modification, e.g., one or more chemically modified nucleotides.

514. Any of the preceding embodiments, wherein the RNA encoding the polypeptide comprises a circRNA.

515. Any of the preceding embodiments, wherein the nucleic acid encoding the polypeptide is comprised within a virus (e.g., AAV, adenovirus, or lentivirus, e.g., integration-deficient lentivirus).

516. Any of the preceding embodiments, wherein the nucleic acid encoding the polypeptide is comprised within a nanoparticle (e.g., a lipid nanoparticle), vesicle, or fusosome.

517. Any preceding numbered embodiment, wherein the polypeptide comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an amino acid sequence encoded by a sequence of an element of Table X.

518. Any preceding numbered embodiment, wherein the reverse transcriptase domain comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the reverse transcriptase domain of an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

519. Any preceding numbered embodiment, wherein the retrotransposase comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an amino acid sequence encoded by a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

520. Any preceding numbered embodiment, wherein the polypeptide comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an amino acid sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

521. Any preceding numbered embodiment, wherein the reverse transcriptase domain comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the reverse transcriptase domain of an amino acid sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

522. Any preceding numbered embodiment, wherein the retrotransposase comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an amino acid sequence of a sequence of an element of Table 3B, Table 10, Table 11, or Table X.

523. Any preceding numbered embodiment, wherein the polypeptide comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

524. Any preceding numbered embodiment, wherein the reverse transcriptase domain comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

525. Any preceding numbered embodiment, wherein the retrotransposase comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

526. Any preceding numbered embodiment, wherein the polypeptide, reverse transcriptase domain, or retrotransposase comprises a linker comprising an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

527. Any preceding numbered embodiment, wherein the polypeptide comprises a DNA binding domain covalently attached to the remainder of the polypeptide by a linker, e.g., a linker comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200, 300, 400, or 500 amino acids.

528. Any preceding numbered embodiment, wherein the linker is attached to the remainder of the polypeptide at a position in the DNA binding domain, RNA binding domain, reverse transcriptase domain, or endonuclease domain (e.g., as shown in any of Figures 17A-17F of PCT Application No. PCT/US2019/048607).

529. Any preceding numbered embodiment, wherein the linker is attached to the remainder of the polypeptide at a position in the N-terminal side of an alpha helical region of the polypeptide, e.g., at a position corresponding to version v1 as described in Example 26 of PCT Application No. PCT/US2019/048607.

530. Any preceding numbered embodiment, wherein the linker is attached to the remainder of the polypeptide at a position in the C-terminal side of an alpha helical region of the polypeptide, e.g., preceding an RNA binding motif (e.g., a -1 RNA binding motif), e.g., at a position corresponding to version v2 as described in Example 26 of PCT Application No. PCT/US2019/048607.

531. Any preceding numbered embodiment, wherein the linker is attached to the remainder of the polypeptide at a position in the C-terminal side of a random coil region of the polypeptide, e.g., N-terminal relative to a DNA binding motif (e.g., a c-myb DNA binding motif), e.g., at a position corresponding to version v3 as described in Example 26 of PCT Application No. PCT/US2019/048607.

532. Any preceding numbered embodiment, wherein the linker comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 1023) or GGGS (SEQ ID NO: 1024).

533. Any preceding numbered embodiment, wherein a polynucleotide sequence comprising at least about 500, 1000, 2000, 3000, 3500, 3600, 3700, 3800, 3900, or 4000 contiguous nucleotides from the 5' end of the template RNA sequence are integrated into a target cell genome.

534. Any preceding numbered embodiment, wherein a polynucleotide sequence comprising at least about 500, 1000, 2000, 2500, 2600, 2700, 2800, 2900, or 3000 contiguous nucleotides from the 3' end of the template RNA sequence are integrated into a target cell genome.

535. Any preceding numbered embodiment, wherein the nucleic acid sequence of the template RNA, or a portion thereof (e.g., a portion comprising at least about 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500, or 4000 nucleotides) integrates into the genomes of a population of target cells at a copy number of at least about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1.0 integrants/genome.

536. Any preceding numbered embodiment, wherein the nucleic acid sequence of the template RNA, or a portion thereof (e.g., a portion comprising at least about 100, 200, 300, 400, 500, 1000, 2000, 2500, 3000, 3500, or 4000 nucleotides) integrates into the genomes of a population of target cells at a copy number of at least about 0.01, 0.02, 0.03, 0.04, 0.05, 0.75, or 0.1 integrants/genome.

537. Any preceding numbered embodiment, wherein the polypeptide comprises a functional endonuclease domain (e.g., wherein the endonuclease domain does not comprise a mutation that abolishes endonuclease activity, e.g., as described herein).

538. Any preceding numbered embodiment, wherein introduction of the system into a target cell does not result in alteration (e.g., upregulation) of p53 and/or p21 protein levels, H2AX phosphorylation (e.g., gamma H2AX), ATM phosphorylation, ATR phosphorylation, Chk1 phosphorylation, Chk2 phosphorylation, and/or p53 phosphorylation.

539. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of p53 protein level in the target cell to a level that is less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, or 90% of the p53 protein level

induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

540. Any preceding numbered embodiment, wherein the p53 protein level is determined according to the method described in Example 30.

541. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of p53 phosphorylation level in the target cell to a level that is less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, or 90% of the p53 phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

542. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of p21 protein level in the target cell to a level that is less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, or 90% of the p53 protein level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

543. Any preceding numbered embodiment, wherein the p21 protein level is determined according to the method described in Example 30.

544. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of H2AX phosphorylation level in the target cell to a level that is less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, or 90% of the H2AX phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

545. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of ATM phosphorylation level in the target cell to a level that is less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, or 90% of the ATM phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

546. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of ATR phosphorylation level in the target cell to a level that is less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, or 90% of the ATR phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

547. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of Chk1 phosphorylation level in the target cell to a level that is less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, or 90% of the Chk1 phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

548. Any preceding numbered embodiment, wherein introduction of the system into a target cell results in upregulation of Chk2 phosphorylation level in the target cell to a level that is less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, or 90% of the Chk2 phosphorylation level induced by introducing a site-specific nuclease, e.g., Cas9, that targets the same genomic site as said system.

549. Any preceding numbered embodiment, wherein the target DNA binding domain recognizes a specific target DNA sequence.

550. Any preceding numbered embodiment, wherein the target DNA binding domain binds to a plurality of (e.g., random) target DNA sequences.

551. Any preceding numbered embodiment, wherein the template RNA comprises a guide RNA (e.g., a U6-driven gRNA).

552. Any preceding numbered embodiment, wherein the target DNA binding domain comprises one or more of a DNA binding domain of a retrotransposase as described herein (e.g., a retrotransposase of an element of Table X, 10, 11, 3A, or 3B), Cas9, nickase Cas9, dCas9, a zinc finger, a TAL, a meganuclease, and/or a transcription factor.

553. Any preceding numbered embodiment, wherein the reverse transcriptase domain comprises a reverse transcriptase domain of a retrotransposase as described herein (e.g., a retrotransposase of an element of Table X, 10, 11, Z1, Z2, 3A, or 3B).

554. Any preceding numbered embodiment, wherein the endonuclease domain comprises an endonuclease domain of of a retrotransposase as described herein (e.g., a retrotransposase of an element of Table X, 10, 11, 3A, or 3B), Cas9, nickase Cas9, a type II restriction enzyme (e.g., FokI), a Holliday junction resolvase, an RLE endonuclease domain, an APE endonuclease domain, or a GIY-YIG endonuclease domain.

555. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain; wherein the DBD and/or the endonuclease domain comprise a heterologous targeting domain that binds specifically to a sequence comprised in a target DNA molecule (e.g., a genomic DNA).

556. A template RNA (or DNA encoding the template RNA) comprising a targeting domain (e.g., a heterologous targeting domain) that binds specifically to a sequence comprised in the target DNA molecule (e.g., a genomic DNA), a sequence that specifically binds an RT domain of a polypeptide, and a heterologous object sequence.

557. The system, method, or template RNA of any of the preceding embodiments, wherein the polypeptide comprises a heterologous targeting domain that binds specifically to a sequence comprised in the target DNA molecule (e.g., a genomic DNA).

558. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous targeting domain binds to a different nucleic acid sequence than the unmodified polypeptide.

559. The system, method, or template RNA of any of the preceding embodiments, wherein the polypeptide does not comprise a functional endogenous targeting domain (e.g., wherein the polypeptide does not comprise an endogenous targeting domain).

560. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous targeting domain comprises a zinc finger (e.g., a zinc finger that binds specifically to the sequence comprised in the target DNA molecule).

561. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous targeting domain comprises a Cas domain (e.g., a Cas9 domain, or a mutant or variant thereof, e.g., a Cas9 domain that binds specifically to the sequence comprised in the target DNA molecule).

562. The system, method, or template RNA of any of the preceding embodiments, wherein the Cas domain is associated with a guide RNA (gRNA).

563. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous targeting domain comprises an endonuclease domain (e.g., a heterologous endonuclease domain).

564. The system, method, or template RNA of any of the preceding embodiments, wherein the endonuclease domain comprises a Cas domain (e.g., a Cas9 or a mutant or variant thereof).

565. The system, method, or template RNA of any of the preceding embodiments, wherein the Cas domain is associated with a guide RNA (gRNA).

566. The system, method, or template RNA of any of the preceding embodiments, wherein the endonuclease domain comprises a Fok1 domain.

567. The system, method, or template RNA of any of the preceding embodiments, wherein the template nucleic acid molecule comprises at least one (e.g., one or two) heterologous homology sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence comprised in a target DNA molecule (e.g., a genomic DNA).

568. The system, method, or template RNA of any of the preceding embodiments, wherein one of the at least one heterologous homology sequences is positioned at or within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of the 5' end of the template nucleic acid molecule.

569. The system, method, or template RNA of any of the preceding embodiments, wherein one of the at least one heterologous homology sequences is positioned at or within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of the 3' end of the template nucleic acid molecule.

570. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous homology sequence binds within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nick site (e.g., produced by a nickase, e.g., an endonuclease domain, e.g., as described herein) in the target DNA molecule.

571. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous homology sequence has less than 50%, 40%, 30%, 20%, 10%, 5%, 4%, 3%, 2%, or 1% sequence identity with a nucleic acid sequence complementary to an endogenous homology sequence of an unmodified form of the template RNA.

572. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous homology sequence has having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence of the target DNA molecule that is different the sequence bound by an endogenous homology sequence (e.g., replaced by the heterologous homology sequence).

573. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous homology sequence comprises a sequence (e.g., at its 3' end) having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence positioned 5' to a nick site of the target DNA molecule (e.g., a site nicked by a nickase, e.g., an endonuclease domain as described herein).

574. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous homology sequence comprises a sequence (e.g., at its 5' end) suitable for priming target-primed reverse transcription (TPRT) initiation.

575. The system, method, or template RNA of any of the preceding embodiments, wherein the heterologous homology sequence has at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a sequence positioned within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of (e.g., 3' relative to) a target insertion site, e.g., for a heterologous object sequence (e.g., as described herein), in the target DNA molecule.

576. The system, method, or template RNA of any of the preceding embodiments, wherein the template nucleic acid molecule comprises a guide RNA (gRNA), e.g., as described herein.

577. The system, method, or template RNA of any of the preceding embodiments, wherein the template nucleic acid molecule comprises a gRNA spacer sequence (e.g., at or within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides of its 5' end).

578. A template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) a sequence that binds a target site (e.g., a second strand of a site in a target genome), (ii) a sequence that specifically binds an RT domain of a polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target homology domain.

579. The template RNA of any of the preceding embodiments, further comprising (v) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide (e.g., the same polypeptide comprising the RT domain).

580. The template RNA of any of the preceding embodiments, wherein the RT domain comprises a sequence selected of Table 3B, 10, 11, or X or a sequence that has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

581. The template RNA of any of the preceding embodiments, wherein the RT domain comprises a sequence selected of Table 3B, 10, 11, or X , wherein the RT domain further comprises a number of substitutions relative to the natural sequence, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions.

582. The template RNA of any of the preceding embodiments, wherein the sequence of (ii) specifically binds the RT domain.

583. The template RNA of any of the preceding embodiments, wherein the sequence that specifically binds the RT domain is a sequence, e.g., a UTR sequence, of Table 3B or 10, or a sequence having at least 70, 75, 80, 85, 90, 95, or 99% identity thereto.

584. A template RNA (or DNA encoding the template RNA) comprising from 5' to 3': (ii) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide, (i) a sequence that binds a target site (e.g., a second strand of a site in a target genome), (iii) a heterologous object sequence, and (iv) a 3' target homology domain.

585. A template RNA (or DNA encoding the template RNA) comprising from 5' to 3': (iii) a heterologous object sequence, (iv) a 3' target homology domain, (i) a sequence that binds a target site (e.g., a second strand of a site in a target genome), and (ii) a sequence that binds an endonuclease and/or a DNA-binding domain of a polypeptide.

586. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein an RNA of the system (e.g., template RNA, the RNA encoding the polypeptide of (a), or an RNA expressed from a heterologous object sequence integrated into a target DNA) comprises a microRNA binding site, e.g., in a 3' UTR.

587. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodimentswherein the microRNA binding site is recognized by a miRNA that is present in a non-target cell type, but that is not present (or is present at a reduced level relative to the non-target cell) in a target cell type.

588. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the miRNA is miR-142, and/or wherein the non-target cell is a Kupffer cell or a blood cell, e.g., an immune cell.

144. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the miRNA is miR-182 or miR-183, and/or wherein the non-target cell is a dorsal root ganglion neuron.

588. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system comprises a first miRNA binding site that is recognized by a first miRNA (e.g., miR-142) and the system further comprises a second miRNA binding site that is recognized by a second miRNA (e.g., miR-182 or miR-183), wherein the first miRNA binding site and the second miRNA binding site are situated on the same RNA or on different RNAs of the system.

589. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the template RNA comprises at least 2, 3, or 4 miRNA binding sites, e.g., wherein the miRNA binding sites are recognized by the same or different miRNAs.

590. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the RNA encoding the polypeptide of (a) comprises at least 2, 3, or 4 miRNA binding sites, e.g., wherein the miRNA binding sites are recognized by the same or different miRNAs.

591. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the RNA expressed from a heterologous object sequence integrated into a target DNA comprises at least 2, 3, or 4 miRNA binding sites, e.g., wherein the miRNA binding sites are recognized by the same or different miRNAs.

Definitions

Domain: The term "domain" as used herein refers to a structure of a biomolecule that contributes to a specified function of the biomolecule. A domain may comprise a contiguous region (e.g., a contiguous sequence) or distinct, non-contiguous regions (e.g., non-contiguous sequences) of a biomolecule. Examples of protein domains include, but are not limited to, an endonuclease domain, a DNA binding domain, a reverse transcription domain; an example of a domain of a nucleic acid is a regulatory domain, such as a transcription factor binding domain.

Exogenous: As used herein, the term exogenous, when used with reference to a biomolecule (such as a nucleic acid sequence or polypeptide) means that the biomolecule was introduced into a host genome, cell or organism by the hand of man. For example, a nucleic acid that is as added into an existing genome, cell, tissue or subject using recombinant DNA techniques or other methods is exogenous to the existing nucleic acid sequence, cell, tissue or subject.

Genomic safe harbor site (GSH site): A genomic safe harbor site is a site in a host genome that is able to accommodate the integration of new genetic material, e.g., such that the inserted genetic element does not cause significant alterations of the host genome posing a risk to

the host cell or organism. A GSH site generally meets 1, 2, 3, 4, 5, 6, 7, 8 or 9 of the following criteria: (i) is located >300kb from a cancer-related gene; (ii) is >300kb from a miRNA/other functional small RNA; (iii) is >50kb from a 5' gene end; (iv) is >50kb from a replication origin; (v) is >50kb away from any ultraconservered element; (vi) has low transcriptional activity (i.e. no mRNA +/- 25 kb); (vii) is not in copy number variable region; (viii) is in open chromatin; and/or (ix) is unique, with 1 copy in the human genome. Examples of GSH sites in the human genome that meet some or all of these criteria include (i) the adeno-associated virus site 1 (AAVS1), a naturally occurring site of integration of AAV virus on chromosome 19; (ii) the chemokine (C-C motif) receptor 5 (CCR5) gene, a chemokine receptor gene known as an HIV-1 coreceptor; (iii) the human ortholog of the mouse Rosa26 locus; (iv) the rDNA locus. Additional GSH sites are known and described, e.g., in Pellenz et al. epub August 20, 2018 (https://doi.org/10.1101/396390).

Heterologous: The term heterologous, when used to describe a first element in reference to a second element means that the first element and second element do not exist in nature disposed as described. For example, a heterologous polypeptide, nucleic acid molecule, construct or sequence refers to (a) a polypeptide, nucleic acid molecule or portion of a polypeptide or nucleic acid molecule sequence that is not native to a cell in which it is expressed, (b) a polypeptide or nucleic acid molecule or portion of a polypeptide or nucleic acid molecule that has been altered or mutated relative to its native state, or (c) a polypeptide or nucleic acid molecule with an altered expression as compared to the native expression levels under similar conditions. For example, a heterologous regulatory sequence (e.g., promoter, enhancer) may be used to regulate expression of a gene or a nucleic acid molecule in a way that is different than the gene or a nucleic acid molecule is normally expressed in nature. In another example, a heterologous domain of a polypeptide or nucleic acid sequence (e.g., a DNA binding domain of a polypeptide or nucleic acid encoding a DNA binding domain of a polypeptide) may be disposed relative to other domains or may be a different sequence or from a different source, relative to other domains or portions of a polypeptide or its encoding nucleic acid. In certain embodiments, a heterologous nucleic acid molecule may exist in a native host cell genome, but may have an altered expression level or have a different sequence or both. In other embodiments, heterologous nucleic acid molecules may not be endogenous to a host cell or host genome but instead may have been introduced into a host cell by transformation (e.g., transfection,

electroporation), wherein the added molecule may integrate into the host genome or can exist as extra-chromosomal genetic material either transiently (e.g., mRNA) or semi-stably for more than one generation (e.g., episomal viral vector, plasmid or other self-replicating vector). In some embodiments, a domain is heterologous relative to another domain, if the first domain is not naturally comprised in the same polypeptide as the other domain (e.g., a fusion between two domains of different proteins from the same organism).

Mutation or Mutated: The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference (e.g., native) nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. A nucleic acid sequence may be mutated by any method known in the art. In some embodiments a mutation occurs naturally. In some embodiments a desired mutation can be produced by a system described herein.

Nucleic acid molecule: Nucleic acid molecule refers to both RNA and DNA molecules including, without limitation, cDNA, genomic DNA and mRNA, and also includes synthetic nucleic acid molecules, such as those that are chemically synthesized or recombinantly produced, such as RNA templates, as described herein. The nucleic acid molecule can be double-stranded or single-stranded, circular or linear. If single-stranded, the nucleic acid molecule can be the sense strand or the antisense strand. Unless otherwise indicated, and as an example for all sequences described herein under the general format "SEQ. ID NO:," "nucleic acid comprising SEQ. ID NO:1" refers to a nucleic acid, at least a portion which has either (i) the sequence of SEQ. ID NO:1, or (ii) a sequence complimentary to SEQ. ID NO:1. The choice between the two is dictated by the context in which SEQ. ID NO:1 is used. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complimentary to the desired target. Nucleic acid sequences of the present disclosure may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more naturally occurring nucleotides with an analog, inter-nucleotide modifications such as uncharged linkages (for example, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (for

example, phosphorothioates, phosphorodithioates, etc.), pendant moieties, (for example, polypeptides), intercalators (for example, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (for example, alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of a molecule. Other modifications can include, for example, analogs in which the ribose ring contains a bridging moiety or other structure such as modifications found in "locked" nucleic acids.

Gene expression unit: a gene expression unit is a nucleic acid sequence comprising at least one regulatory nucleic acid sequence operably linked to at least one effector sequence. A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if the promoter or enhancer affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be contiguous or non-contiguous. Where necessary to join two protein-coding regions, operably linked sequences may be in the same reading frame.

Host: The terms host genome or host cell, as used herein, refer to a cell and/or its genome into which protein and/or genetic material has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell and/or genome, but to the progeny of such a cell and/or the genome of the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A host genome or host cell may be an isolated cell or cell line grown in culture, or genomic material isolated from such a cell or cell line, or may be a host cell or host genome which composing living tissue or an organism. In some instances, a host cell may be an animal cell or a plant cell, e.g., as described herein. In certain instances, a host cell may be a bovine cell, horse cell, pig cell, goat cell, sheep cell, chicken cell, or turkey cell. In certain instances, a host cell may be a corn cell, soy cell, wheat cell, or rice cell.

Pseudoknot: A "pseudoknot sequence" sequence, as used herein, refers to a nucleic acid (e.g., RNA) having a sequence with suitable self-complementarity to form a pseudoknot structure, e.g., having: a first segment, a second segment between the first segment and a third segment, wherein the third segment is complementary to the first segment, and a fourth segment, wherein the fourth segment is complementary to the second segment. The pseudoknot may optionally have additional secondary structure, e.g., a stem loop disposed in the second segment, a stem-loop disposed between the second segment and third segment, sequence before the first segment, or sequence after the fourth segment. The pseudoknot may have additional sequence between the first and second segments, between the second and third segments, or between the third and fourth segments. In some embodiments, the segments are arranged, from 5' to 3': first, second, third, and fourth. In some embodiments, the first and third segments comprise five base pairs of perfect complementarity. In some embodiments, the second and fourth segments comprise 10 base pairs, optionally with one or more (e.g., two) bulges. In some embodiments, the second segment comprises one or more unpaired nucleotides, e.g., forming a loop. In some embodiments, the third segment comprises one or more unpaired nucleotides, e.g., forming a loop.

Stem-loop sequence: As used herein, a "stem-loop sequence" refers to a nucleic acid sequence (e.g., RNA sequence) with sufficient self-complementarity to form a stem-loop, e.g., having a stem comprising at least two (e.g., 3, 4, 5, 6, 7, 8, 9, or 10) base pairs, and a loop with at least three (e.g., four) base pairs. The stem may comprise mismatches or bulges.

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIG. 1 is a schematic of the Gene Writing™ genome editing system.

FIG. 2 is a schematic of the structure of the Gene Writer™ genome editor polypeptide.

FIG. 3 is a schematic of the structure of exemplary Gene Writer™ template RNAs.

FIGS. 4A and 4B are a series of diagrams showing examples of configurations of Gene Writers using domains derived from a variety of sources. Gene Writers as described herein may or may not comprise all domains depicted. For example, a GeneWrite may, in some instances,

lack an RNA-binding domain, or may have single domains that fulfill the functions of multiple domains, e.g., a Cas9 domain for DNA binding and endonuclease activity. Exemplary domains that can be included in a GeneWriter polypeptide include DNA binding domains (e.g., comprising a DNA binding domain of an element of a sequence listed in any of Tables X, Y, Z1, Z2, 3A, or 3B; a zinc finger; a TAL domain; Cas9; dCas9; nickase Cas9; a transcription factor, or a meganuclease), RNA binding domains (e.g., comprising an RNA binding domain of B-box protein, MS2 coat protein, dCas, or an element of a sequence listed in any of Tables X, Y, Z1, Z2, 3A, or 3B), reverse transcriptase domains (e.g., comprising a reverse transcriptase domain of an element of a sequence listed in any of Tables X, Y, 3A, or 3B; other retrotransposases (e.g., as listed in Table Z1); a peptide containing a reverse transctipase domain (e.g., as listed in Table Z2)), and/or an endonuclease domain (e.g., comprising an endonuclease domain of an element of a sequence listed in any of Tables X, Y, 3A, or 3B; Cas9; nickase Cas9; a restriction enzyme (e.g., a type II restriction enzyme, e.g., FokI); a meganuclease; a Holliday junction resolvase; an RLE retrotranspase; an APE retrotransposase; or a GIY-YIG retrotransposase). Exemplary GeneWriter polypeptides comprising exemplary combinations of such domains are shown in the bottom panel.

FIG. 5 is a diagram showing the modules of an exemplary GeneWriter RNA template. Individual modules of the exemplary template can be combined, re-arranged, and/or omitted, e.g., to produce a Gene Writer template. A = 5' homology arm; B = Ribozyme; C = 5' UTR; D = heterologous object sequence; E = 3' UTR; F = 3' homology arm.

FIG. 6 is a table listing the modules of an exemplary Gene Writer RNA template. Individual modules can be combined, re-arranged, and/or omitted, e.g., to produce a Gene Writer template. A = 5' homology arm; B = Ribozyme; C = 5' UTR; D = heterologous object sequence; E = 3' UTR; F = 3' homology arm.

FIGS. 7A and 7B are diagrams showing an exemplary second strand nicking process. (A) A Cas9 nickase is fused to a Gene Writer protein. The Gene Writer protein introduces a nick in a DNA strand through its EN domain (shown as *), and the fused Cas9 nickase introduces a nicks on either top or bottom DNA strands (shown as X). (B) A Gene Writer is targeted to DNA through its DNA biding domain and introduces a DNA nick with its EN domain (*). A Cas9 nickase is then used the generate a second nick (X) at the top or bottom strand, upstream or downstream of the EN introduced nick.

FIG. 8 represents screening construct design for retrotransposon-mediated integration in human cells. A driver plasmid comprising a retrotransposase (Driver) expression cassette is transfected together with a template plasmid comprising a retrotransposon-dependent reporter cassette. Whereas expression from the template plasmid results in a non-functional GFP because of an interrupting antisense intron, transcription of the template molecule from the template plasmid results in the generation of an RNA with the intron removed by splicing that can then be reverse transcribed and integrated by the system. Expression of the reporter cassette will thus only occur from the integrated reporter cassette (Integrated gDNA, bottom) and not from the template plasmid. HA=homology arm, where applicable; CMV=mammalian CMV promoter; HiBit=HiBit tag for quantification of protein expression;T7=T7 RNA polymerase promoter; UTR=untranslated sequence, e.g., native retrotransposon UTRs; pA=poly A signal; SD-SA is used to indicate the splice-donor and splice-acceptor sites of an antisense intron in the GFP coding sequence. HA=homology arm, where applicable (e.g., see Example 6); CMV=mammalian CMV promoter; HiBit=HiBit tag for quantification of protein expression;T7=T7 RNA polymerase promoter; UTR=untranslated sequence, e.g., native retrotransposon UTRs; pA=poly A signal; SD-SA is used to indicate the splice-donor and splice-acceptor sites of an antisense intron in the GFP coding sequence

FIG.9 discloses screening of candidate retrotransposons and identifies 25 candidates that integrate a trans payload in human cells. A total of 163 retrotransposon systems were assayed for activity in human cells as described in Example 4. Integration as measured by ddPCR is shown as copies/genome for each retrotransposon driver/template system. The height of each bar indicates the average value of two replicates. After further optimization in Examples 4 and 5, the constructs with higher activity are further highlighted in Figure 10.

FIG.10 : High-activity Gene Writing configurations based on retrotransposon hits from Table 3B and further improved in Examples 7,8 , and 9. Where multiple configurations of a given system, e.g., alternate coding sequence of the retrotransposase (Example 8) or addition of homology arms (Example 9) were tested, only the highest performing configuration is shown. For systems improved beyond the initial configuration set forth in Table 3B and evaluated in Example 7, the improvements described in Example 5 (Figure 11) and Example 6(Figure 12) are detailed in Table 11.

FIG. 11 Retrotransposon consensus sequence can rescue or improve trans integration activity in human cells. Integration efficiency measured by copies/genome via ddPCR as described in Example 5 is shown for each retrotransposon driver/template system. The height of each bar indicates the average value of two replicates. Empty circles indicate each replicate and bars represent the original sequences (light gray) or the novel consensus-generated protein sequence (dark gray).

FIG.12 shows that consensus motif-generated homology arm sequences can rescue or improve retrotransposon trans integration activity in human cells. Integration efficiency as measured by copies/genome via ddPCR (Example 6) is shown for each Gene Writer driver/template system. The height of each bar indicates the average value of two replicates. Filled circles and white bars represent template sequence without homology arms and unfilled circles and hashed bars indicate designs comprising homology arms.

FIG.13 A & B show luciferase activity assay for primary cells. LNPs formulated as according to Example 11 were analyzed for delivery of cargo to primary human (A) and mouse (B) hepatocytes, as according to Example 12. The luciferase assay revealed dose-responsive luciferase activity from cell lysates, indicating successful delivery of RNA to the cells and expression of Firefly luciferase from the mRNA cargo.

FIG.14 discloses LNP-mediated delivery of RNA cargo to the murine liver. Firefly luciferase mRNA-containing LNPs were formulated and delivered to mice by iv, and liver samples were harvested and assayed for luciferase activity at 6, 24, and 48 hours post administration. Reporter activity by the various formulations followed the ranking LIPIDV005>LIPIDV004>LIPIDV003. RNA expression was transient and enzyme levels returned near vehicle background by 48 hours. Post-administration.

DETAILED DESCRIPTION

This disclosure relates to compositions, systems and methods for targeting, editing, modifying or manipulating a DNA sequence (e.g., inserting a heterologous object DNA sequence into a target site of a mammalian genome) at one or more locations in a DNA sequence in a cell, tissue or subject, e.g., *in vivo* or *in vitro*. The object DNA sequence may include, e.g., a coding sequence, a regulatory sequence, a gene expression unit.

More specifically, the disclosure provides retrotransposon-based systems for inserting a sequence of interest into the genome. This disclosure is based, in part, on a bioinformatic analysis to identify retrotransposase sequences and the associated 5' UTR and 3' UTR from a variety of organisms (see Tables 3B, 10, 11, and X). Additional examples of retrotransposon elements are listed, e.g., in Tables 1 and 2 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety.

In some embodiments, systems described herein can have a number of advantages relative to various earlier systems. For instance, the disclosure describes retrotransposases capable of inserting long sequences of heterologous nucleic acid into a genome. In addition, retrotransposases described herein can insert heterologous nucleic acid in an endogenous site in the genome, such as the rDNA locus. This is in contrast to Cre/loxP systems which require a first step of inserting an exogenous loxP site before a second step of inserting a sequence of interest into the loxP site.

Gene-writer™ genome editors

Non-long terminal repeat (LTR) retrotransposons are a type of mobile genetic elements that are widespread in eukaryotic genomes. They include, for example, the apurinic/apyrimidinic endonuclease (APE)-type, the restriction enzyme-like endonuclease (RLE)-type, and the Penelope-like element (PLE)-type. The APE class retrotransposons are comprised of two functional domains: an endonuclease/DNA binding domain, and a reverse transcriptase domain. Examples of APE-class retrotransposons can be found, for example, in Table 1 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety, including the sequence listing and sequences referred to in Table 1 therein. The RLE class are comprised of three functional domains: a DNA binding domain, a reverse transcription domain, and an endonuclease domain. Examples of RLE-class retrotransposons can be found, for example, in Table 2 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety, including the sequence listing and sequences referred to in Table 2 therein. The reverse transcriptase domain of non-LTR retrotransposon functions by binding an RNA sequence template and reverse transcribing it into the host genome's target DNA. The RNA sequence template has a 3' untranslated region which is specifically bound to the retrotransposase, and a variable 5' region generally having Open Reading Frame(s) ("ORF") encoding retrotransposase

proteins. The RNA sequence template may also comprise a 5' untranslated region which specifically binds the retrotransposase. Penelope-like elements (PLEs) are distinct from both LTR and non-LTR retrotransposons. PLEs generally comprise a reverse transcriptase domain distinct from that of APE and RLE elements, but similar to that of telomerases and Group II introns, and an optional GIY-YIG endonuclease domain.

Other exemplary classes of retrotransposon include, without limitation, CRE, NeSL, R4, R2, Hero, L1, RTE (e.g., AviRTE_Brh or BovB), I, Jockey, CR1, Rex1, RandI/Dualen, Penelope or Penelope-like (PLE) (e.g., Penelope_SM), Tx1, RTEX, Crack, Nimb, Proto1, Proto2, RTETP, L2, Tad1, Loa, Ingi, Outcast, R1, Daphne, L2A, L2B, Ambal, Vingi (e.g., Vingi-1_EE), and Kiri retrotransposons.

As described herein, the elements of such retrotransposons can be functionally modularized and/or modified to target, edit, modify or manipulate a target DNA sequence, e.g., to insert an object (e.g., heterologous) nucleic acid sequence into a target genome, e.g., a mammalian genome, by reverse transcription. Such modularized and modified nucleic acids, polypeptide compositions and systems are described herein and are referred to as Gene Writer™ gene editors. A Gene Writer™ gene editor system comprises: (A) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain, and either (x) an endonuclease domain that contains DNA binding functionality or (y) an endonuclease domain and separate DNA binding domain; and (B) a template RNA comprising (i) a sequence that binds the polypeptide and (ii) a heterologous insert sequence. For example, the Gene Writer genome editor protein may comprise a DNA-binding domain, a reverse transcriptase domain, and an endonuclease domain. In other embodiments, the Gene Writer genome editor protein may comprise a reverse transcriptase domain and an endonuclease domain. In certain embodiments, the elements of the Gene Writer™ gene editor polypeptide can be derived from sequences of retrotransposons, e.g., APE-type, RLE-type, or PLE-type retrotransposons or portions or domains thereof. In some embodiments the RLE-type non-LTR retrotransposon is from the R2, NeSL, HERO, R4, or CRE clade. In some embodiments the Gene Writer genome editor is derived from R4 element X4_Line, which is found in the human genome. In some embodiments the APE-type non-LTR retrotransposon is from the R1, or Tx1 clade. In some embodiments the Gene Writer genome editor is derived from Tx1 element Mare6, which is found in the human genome. The RNA template element of a Gene Writer™ gene

editor system is typically heterologous to the polypeptide element and provides an object sequence to be inserted (reverse transcribed) into the host genome. In some embodiments the Gene Writer genome editor protein is capable of target primed reverse transcription.

In some embodiments the Gene Writer genome editor is combined with a second polypeptide. In some embodiments the second polypeptide is derived from an APE-type non-LTR retrotransposon. In some embodiments the second polypeptide has a zinc knuckle-like motif. In some embodiments the second polypeptide is a homolog of Gag proteins.

In some embodiments, the Gene Writer genome editor comprises a retrotransposase sequence of an element listed in Table X. Table X provides a series of nucleic acid and amino acid sequences (listed by Repbase gene name and species), with associated GenBank Accession numbers where available. The Repbase nucleic acid and amino acid sequences of elements listed in Table X are incorporated herein by reference in their entireties. The GenBank sequences of elements listed in Table X are also incorporated herein by reference in their entireties. The nucleic acid sequences of Table X include, in some instances, a sequence encoding a polypeptide (e.g., an open reading frame) (e.g., the protein-coding sequence of the corresponding Repbase entry, incorporated herein by reference). The nucleic acid sequences of Table X include, in some instances, a 5' UTR sequence (e.g., the 5' UTR sequence of the corresponding Repbase entry, incorporated herein by reference). The nucleic acid sequences of Table X include, in some instances, a 3' UTR sequence (e.g., the 3' UTR sequence of the corresponding Repbase entry, incorporated herein by reference).

In some embodiments, an open reading frame (ORF) for the amino acid sequence is annotated in Repbase. In some embodiments, an ORF for the amino acid sequence is not annotated in Repbase. When the ORF is not annotated, it can be identified by one of skill in the art, e.g., by performing one or more translations (e.g., all-frame translations), and optionally comparing the translation to a reverse transcriptase sequence or consensus motif. In some embodiments, an amino acid sequence of Table X, e.g., as used herein, is an amino acid sequence listed in the corresponding Repbase entry, or an amino acid sequence encoded by the nucleic acid sequence in the corresponding Repbase entry. In some embodiments, an amino acid sequence encoded by an element of Table X is an amino acid sequence encoded by the full length sequence of an element listed in Table X, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the full

length sequence of an element listed in Table X may comprise one or more (e.g., all of) of a 5' UTR, polypeptide-encoding sequence, or 3' UTR of a retrotransposon as described herein. In some embodiments, an amino acid sequence of Table X is an amino acid sequence encoded by the full length sequence of an element listed in Table X, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, a 5' UTR of an element of Table X comprises a 5' UTR of the full length sequence of an element listed in Table X, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, a 3' UTR of an element of Table X comprises a 3' UTR of the full length sequence of an element listed in Table X, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

Also indicated in Table X are the host organisms from which the nucleic acid sequences were obtained and a listing of domains present within the polypeptide encoded by the open reading frame of the nucleic acid sequence. The domains listed in Table X are indicated as domain identifiers, which correspond to InterPro domain entries as listed in Table Y. Thus, a Repbase sequence listed in Table X can encode a polypeptide including one or more domains, indicated in Table X by their domain identifiers. The specific domain or domain type associated with each domain identifier is described in greater detail in Table Y. In some embodiments, a domain of interest (e.g., as listed in Table Y) can be identified in a nucleic acid sequence (e.g., as listed in Table X) by performing all-frame translations and identifying amino acid sequences encoding the domain of interest.

Retrotransposon discovery tools

As the result of repeated mobilization over time, transposable elements in genomic DNA often exist as tandem or interspersed repeats (Jurka Curr Opin Struct Biol 8, 333-337 (1998)). Tools capable of recognizing such repeats can be used to identify new elements from genomic DNA and for populating databases, e.g., Repbase (Jurka et al Cytogenet Genome Res 110, 462-467 (2005)). One such tool for identifying repeats that may comprise transposable elements is RepeatFinder (Volfovsky et al Genome Biol 2 (2001)), which analyzes the repetitive structure of genomic sequences. Repeats can further be collected and analyzed using additional tools, e.g., Censor (Kohany et al BMC Bioinformatics 7, 474 (2006)). The Censor package takes genomic

repeats and annotates them using various BLAST approaches against known transposable elements. An all-frames translation can be used to generate the ORF(s) for comparison.

Other exemplary methods for identification of transposable elements include RepeatModeler2, which automates the discovery and annotation of transposable elements in genome sequences (Flynn et al bioRxiv (2019)). In addition to accomplishing this via available packages like Censor, one can perform an all-frames translation of a given genome or sequence and annotate with a protein domain tool like InterProScan, which tags the domains of a given amino acid sequence using the InterPro database (Mitchell et al. Nucleic Acids Res 47, D351-360 (2019)), allowing the identification of potential proteins comprising domains associated with known transposable elements (e.g., domains of elements listed in Table X).

Retrotransposons can be further classified according to the reverse transcriptase domain using a tool such as RTclass1 (Kapitonov et al Gene 448, 207-213 (2009)).

Polypeptide component of Gene Writer gene editor system

RT domain:

In certain aspects of the present invention, the reverse transcriptase domain of the Gene Writer system is based on a reverse transcriptase domain of an APE-type or RLE-type non-LTR retrotransposon, or of a PLE-type retrotransposon. A wild-type reverse transcriptase domain of an APE-type, RLE-type, or PLE-type retrotransposon can be used in a Gene Writer system or can be modified (e.g., by insertion, deletion, or substitution of one or more residues) to alter the reverse transcriptase activity for target DNA sequences. In some embodiments the reverse transcriptase is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some embodiments the reverse transcriptase domain is a heterologous reverse transcriptase from a different retrovirus, retron, diversity-generating retroelement, retroplasmid, Group II intron, LTR-retrotransposon, non-LTR retrotransposon, or other source, e.g., as exemplified in Table Z1 or as comprising a domain listed in Table Z2. In certain embodiments, a Gene Writer system includes a polypeptide that comprises a reverse transcriptase domain of an RLE-type non-LTR retrotransposon from the R2, NeSL, HERO, R4, or CRE clade, of an APE-type non-LTR retrotransposon from the L1, RTE, I, Jockey, CR1, Rex1, RandI/Dualen, T1, RTEX, Crack, Nimb, Proto1, Proto2, RTETP, L2, Tad1, Loa, Ingi, Outcast, R1, Daphne, L2A, L2B, Ambal, Vingi, or Kiri clade, or of a PLE-type retrotransposon. In certain embodiments, a

Gene Writer system includes a polypeptide that comprises a reverse transcriptase domain of a retrotransposon listed in Table 10, Table 11, Table X, Table Z1, Table Z2, or Table 3A or 3B. In embodiments, the amino acid sequence of the reverse transcriptase domain of a Gene Writer system is at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identical to the amino acid sequence of a reverse transcriptase domain of a retrotransposon whose DNA sequence is referenced in Table 10, Table 11, Table X, Table Z1, Table Z2, or Table 3A or 3B. Reverse transcription domains can be identified, for example, based upon homology to other known reverse transcription domains using routine tools as Basic Local Alignment Search Tool (BLAST). In some embodiments, reverse transcriptase domains are modified, for example by site-specific mutation. In some embodiments, the reverse transcriptase domain is engineered to bind a heterologous template RNA.

In some embodiments, a polypeptide (e.g., RT domain) comprises an RNA-binding domain, e.g., that specifically binds to an RNA sequence. In some embodiments, a template RNA comprises an RNA sequence that is specifically bound by the RNA-binding domain.

In some embodiments, the RT domain exhibits enhanced stringency of target-primed reverse transcription (TPRT) initiation, e.g., relative to an endogenous RT domain. In some embodiments, the RT domain initiates TPRT when the 3 nt in the target site immediately upstream of the first strand nick, e.g., the genomic DNA priming the RNA template, have at least 66% or 100% complementarity to the 3 nt of homology in the RNA template. In some embodiments, the RT domain initiates TPRT when there are less than 5 nt mismatched (e.g., less than 1, 2, 3, 4, or 5 nt mismatched) between the template RNA homology and the target DNA priming reverse transcription. In some embodiments, the RT domain is modified such that the stringency for mismatches in priming the TPRT reaction is increased, e.g., wherein the RT domain does not tolerate any mismatches or tolerates fewer mismatches in the priming region relative to a wild-type (e.g., unmodified) RT domain. In some embodiments, the RT domain comprises a HIV-1 RT domain. In embodiments, the HIV-1 RT domain initiates lower levels of synthesis even with three nucleotide mismatches relative to an alternative RT domain (e.g., as

described by Jamburuthugoda and Eickbush J Mol Biol 407(5):661-672 (2011); incorporated herein by reference in its entirety).

In some embodiments, the RT domain forms a dimer (e.g., a heterodimer or homodimer). In some embodiments, the RT domain is monomeric. In some embodiments, an RT domain, e.g., a retroviral RT domain, naturally functions as a monomer or as a dimer (e.g., heterodimer or homodimer). In some embodiments, an RT domain naturally functions as a monomer, e.g., is derived from a virus wherein it functions as a monomer. Exemplary monomeric RT domains, their viral sources, and the RT signatures associated with them can be found in Table 30 with descriptions of domain signatures in Table 32. In some embodiments, the RT domain of a system described herein comprises an amino acid sequence of Table 30, or a functional fragment or variant thereof, or a sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto. In embodiments, the RT domain is selected from an RT domain from murine leukemia virus (MLV; sometimes referred to as MoMLV) (e.g., P03355), porcine endogenous retrovirus (PERV) (e.g., UniProt Q4VFZ2), mouse mammary tumor virus (MMTV) (e.g., UniProt P03365), Mason-Pfizer monkey virus (MPMV) (e.g., UniProt P07572), bovine leukemia virus (BLV) (e.g., UniProt P03361), human T-cell leukemia virus-1 (HTLV-1) (e.g., UniProt P03362), human foamy virus (HFV) (e.g., UniProt P14350), simian foamy virus (SFV) (e.g., UniProt P23074), or bovine foamy/syncytial virus (BFV/BSV) (e.g., UniProt O41894), or a functional fragment or variant thereof (e.g., an amino acid sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto). In some embodiments, an RT domain is dimeric in its natural functioning. Exemplary dimeric RT domains, their viral sources, and the RT signatures associated with them can be found in Table 31 with descriptions of domain signatures in Table 32. In some embodiments, the RT domain of a system described herein comprises an amino acid sequence of Table 31, or a functional fragment or variant thereof, or a sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain is derived from a virus wherein it functions as a dimer. In embodiments, the RT domain is selected from an RT domain from avian sarcoma/leukemia virus (ASLV) (e.g., UniProt A0A142BKH1), Rous sarcoma virus (RSV) (e.g., UniProt P03354), avian myeloblastosis virus (AMV) (e.g., UniProt Q83133), human immunodeficiency virus type I (HIV-1) (e.g., UniProt P03369), human immunodeficiency virus type II (HIV-2) (e.g., UniProt P15833), simian immunodeficiency virus (SIV) (e.g., UniProt P05896), bovine immunodeficiency virus (BIV) (e.g., UniProt P19560), equine infectious

anemia virus (EIAV) (e.g., UniProt P03371), or feline immunodeficiency virus (FIV) (e.g., UniProt P16088) (Herschhorn and Hizi Cell Mol Life Sci 67(16):2717-2747 (2010)), or a functional fragment or variant thereof (e.g., an amino acid sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto). Naturally heterodimeric RT domains may, in some embodiments, also be functional as homodimers. In some embodiments, dimeric RT domains are expressed as fusion proteins, e.g., as homodimeric fusion proteins or heterodimeric fusion proteins. In some embodiments, the RT function of the system is fulfilled by multiple RT domains (e.g., as described herein). In further embodiments, the multiple RT domains are fused or separate, e.g., may be on the same polypeptide or on different polypeptides.

In some embodiment, a GeneWriter described herein comprises an integrase domain, e.g., wherein the integrase domain may be part of the RT domain. In some embodiments, an RT domain (e.g., as described herein) comprises an integrase domain. In some embodiments, an RT domain (e.g., as described herein) lacks an integrase domain, or comprises an integrase domain that has been inactivated by mutation or deleted. In some embodiment, a GeneWriter described herein comprises an RNase H domain, e.g., wherein the RNase H domain may be part of the RT domain. In some embodiments, an RT domain (e.g., as described herein) comprises an RNase H domain, e.g., an endogenous RNAse H domain or a heterologous RNase H domain. In some embodiments, an RT domain (e.g., as described herein) lacks an RNase H domain. In some embodiments, an RT domain (e.g., as described herein) comprises an RNase H domain that has been added, deleted, mutated, or swapped for a heterologous RNase H domain. In some embodiments, mutation of an RNase H domain yields a polypeptide exhibiting lower RNase activity, e.g., as determined by the methods described in Kotewicz et al. Nucleic Acids Res 16(1):265-277 (1988) (incorporated herein by reference in its entirety), e.g., lower by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% compared to an otherwise similar domain without the mutation. In some embodiments, RNase H activity is abolished.

In some embodiments, an RT domain is mutated to increase fidelity compared to to an otherwise similar domain without the mutation. For instance, in some embodiments, a YADD or YMDD motif in an RT domain (e.g., in a reverse transcriptase) is replaced with YVDD. In embodiments, replacement of the YADD or YMDD or YVDD results in higher fidelity in retroviral reverse transcriptase activity (e.g., as described in Jamburuthugoda and Eickbush J Mol Biol 2011; incorporated herein by reference in its entirety).

The diversity of reverse transcriptases (e.g., comprising RT domains) has been described in, but not limited to, those used by prokaryotes (Zimmerly et al. *Microbiol Spectr* 3(2):MDNA3-0058-2014 (2015); Lampson B.C. (2007) Prokaryotic Reverse Transcriptases. In: Polaina J., MacCabe A.P. (eds) Industrial Enzymes. Springer, Dordrecht), viruses (Herschhorn et al. *Cell Mol Life Sci* 67(16):2717-2747 (2010); Menéndez-Arias et al. *Virus Res* 234:153-176 (2017)), and mobile elements (Eickbush et al. *Virus Res* 134(1-2):221-234 (2008); Craig et al. *Mobile DNA III* 3rd Ed. DOI:10.1128/9781555819217 (2015)), each of which is incorporated herein by reference.

Table 30: Exemplary monomeric retroviral reverse transcriptases and their RT domain signatures

Name	Accession	Organism	Sequence	RT Signatures

Q4VFZ2_9GAMR - residues only	Q4VFZ2	Porcine endogenous retrovirus	MGATGQQQYPWTTRRTVDLGVGRVT HSFLVIPECPAPLLGRDLLTKMGAQISF EQGKPEVSANNKPITVLTLQLDDEYRL YSPLVKPDQNIQFWLEQFPQAWAETA GMGLAKQVPPQVIQLKASATPVSVRQ YPLSKEAQEGIRPHVQRLIQQGILVPVQ SPWNTPLLPVRKPGTNDYRPVQDLRE VNKRVQDIHPTVPNPYNLLCALPPQRS WYTVLDLKDAFFCLRLHPTSQPLFAFE WRDPGTGRTGQLTWTRLPQGFKNSPTI FDEALHRDLANFRIQHPQVTLLQYVDD LLLAGATKQDCLEGTKALLLELSDLGY RASAKKAQICRREVTYLGYSLRDGQR WLTEARKKTVVQIPAPTTAKQVREFLG TAGFCRLWIPGFATLAAPLYPLTKEKG EFSWAPEHQKAFDAIKKALLSAPALAL PDVTKPFTLYVDERKGVARGVLTQTL GPWRRPVAYLSKKLDPVASGWPVCLK AIAAVAILVKDADKLTLGQNITVIAPH ALENIVRQPPDRWMTNARMTHYQSLL LTERVTFAPPAALNPATLLPEETDEPVT HDCHQLLIEETGVRKDLTDIPLTGEVLT WFTDGSSYVVEGKRMAGAAVVDGTR TIWASSLPEGTSAQKAELMALTQALRL AEGKSINIYTDSRYAFATAHVHGAIYK QRGLLTSAGREIKNKEEILSLLEALHLP KRLAIIHCPGHQKAKDPISRGNQMADR VAKQAAQGVNLLPMIETPKAPEPGRQ YTLEDWQEIKKIDQFSETPEGTCYTSD GKEILPHKEGLEYVQQIHRLTHLGTKH LQQLVRTSPYHVLRLPGVADSVVKHC VPCQLVNANPSRIPPGKRLRGSHPGAH WEVDFTEVKPAKYGNKYLLVFVDTFS GWVEAYPTKKETSTVVAKKILEEIFPR FGIPKVIGSDNGPAFVAQVSQGLAKILG IDWKLHCAYRPQSSGQVERMNRTIKET LTKLTAETGVNDWIALLPFVLFRVRNT PGQFGLTPYELLYGGPPPLVEIASVHSA DVLLSQPLFSRLKALEWVRQRAWRQL REAYSGGGDLQIPHRFQVGDSVYVRR HRAGNLETRWKGPYHVLLTTPTAVKV EGISTWIHASHVKPAPPPDSGWKAEKT ENPLKLRLHRVVPYSVNNFSS	IPR043502, SSF56672, IPR000477, PF00078, cd03715

POL_SFV1-residues only	P23074	Simian foamy virus type 1	MDPLQLLQPLEAEIKGTKLKAHWDSG ATITCVPEAFLEDERPIQTMLIKTIHGEK QQDVYYLTFKVQGRKVEAEVLASPYD YILLNPSDVPWLMKKPLQLTVLVPLHE YQERLLQQTALPKEQKELLQKLFLKY DALWQHWENQVGHRRIKPHNIATGTL APRPQKQYPINPKAKPSIQIVIDDLLKQ GVLIQQNSTMNTPVYPVPKPDGKWRM VLDYREVNKTIPLIAAQNQHSAGILSSI YRGKYKTTLDLTNGFWAHPITPESYW LTAFTWQGKQYCWTRLPQGFLNSPAL FTADVVDLLKEIPNVQAYVDDIYISHD DPQEHLEQLEKIFSILLNAGYVVSLKKS EIAQREVEFLGFNITKEGRGLTDTFKQK LLNITPPKDLKQLQSILGLLNFARNFIPN YSELVKPLYTIVANANGKFISWTEDNS NQLQHIISVLNQADNLEERNPETRLIIK VNSSPSAGYIRYYNEGSKRPIMYVNYIF SKAEAKFTQTEKLLTTMHKGLIKAMD LAMGQEILVYSPIVSMTKIQRTPLPERK ALPVRWITWMTYLEDPRIQFHYDKSLP ELQQIPNVTEDVIAKTKHPSEFAMVFY TDGSAIKHPDVNKSHSAGMGIAQVQFI PEYKIVHQWSIPLGDHTAQLAEIAAVE FACKKALKISGPVLIVTDSFYVAESAN KELPYWKSNGFLNNKKKPLRHVSKW KSIAECLQLKPDIIIMHEKGHQQPMTTL HTEGNNLADKLATQGSYVVHCNTTPS LDAELDQLLQGHYPPGYPKQYKYTLE ENKLIVERPNGIRIVPPKADREKIISTAH NIAHTGRDATFLKVSSKYWWPNLRKD VVKSIRQCKQCLVTNATNLTSPPILRPV KPLKPFDKFYIDYIGPLPPSNGYLHVLV VVDSMTGFVWLYPTKAPSTSATVKAL NMLTSIAIPKVLHSDQGAAFTSSTFAD WAKEKGIQLEFSTPYHPQSSGKVERKN SDIKRLLTKLLIGRPAKWYDLLPVVQL ALNNSYSPSSKYTPHQLLFGVDSNTPF ANSDTLDLSREEELSLLQEIRSSLHQPT SPPASSRSWSPSVGQLVQERVARPASL RPRWHKPTAILEVVNPRTVIILDHLGN RRTVSVDNLKLTAYQDNGTSNDSGTM ALMEEDESSTSST	IPR043502, SSF56672, IPR000477, PF00078

POL_MPMV - residues only	P07572	Mason-Pfizer monkey virus	MGQELSQHERYVEQLKQALKTRGVK VKYADLLKFFDFVKDTCPWFPQEGTID IKRWRRVGDCFQDYYNTFGPEKVPVT AFSYWNLIKELIDKKEVNPQVMAAVA QTEEILKSNSQTDLTKTSQNPDLDLISL DSDDEGAKSSSLQDKGLSSTKKPKRFP VLLTAQTSKDPEDPNPSEVDWDGLED EAAKYHNPDWPPFLTRPPPYNKATPSA PTVMAVVNPKEELKEKIAQLEEQIKLE ELHQALISKLQKLKTGNETVTHPDTAG GLSRTPHWPGQHIPKGKCCASREKEEQ IPKDIFPVTETVDGQGQAWRHHNGFDF AVIKELKTAASQYGATAPYTLAIVESV ADNWLTPTDWNTLVRAVLSGGDHLL WKSEFFENCRDTAKRNQQAGNGWDF DMLTGSGNYSSTDAQMQYDPGLFAQI QAAATKAWRKLPVKGDPGASLTGVK QGPDEPFADFVHRLITTAGRIFGSAEAG VDYVKQLAYENANPACQAAIRPYRKK TDLTGYIRLCSDIGPSYQQGLAMAAAF SGQTVKDFLNNKNKEKGGCCFKCGKK GHFAKNCHEHAHNNAEPKVPGLCPRC KRGKHWANECKSKTDNQGNPIPPHQG NRVEGPAPGPETSLWGSQLCSSQQKQP ISKLTRATPGSAGLDLCSTSHTVLTPEM GPQALSTGIYGPLPPNTFGLILGRSSITM KGLQVYPGVIDNDYTGEIKIMAKAVN NIVTVSQGNRIAQLILLPLIETDNKVQQ PYRGQGSFGSSDIYWVQPITCQKPSLTL WLDDKMFTGLIDTGADVTIIKLEDWPP NWPITDTLTNLRGIGQSNNPKQSSKYL TWRDKENNSGLIKPFVIPNLPVNLWGR DLLSQMKIMMCSPNDIVTAQMLAQGY SPGKGLGKKENGILHPIPNQGQSNKKG FGNFLTAAIDILAPQQCAEPITWKSDEP VWVDQWPLTNDKLAAAQQLVQEQLE AGHITESSSPWNTPIFVIKKKSGKWRLL QDLRAVNATMVLMGALQPGLPSPVAI PQGYLKIIIDLKDCFFSIPLHPSDQKRFA FSLPSTNFKEPMQRFQWKVLPQGMAN SPTLCQKYVATAIHKVRHAWKQMYII HYMDDILIAGKDGQQVLQCFDQLKQE LTAAGLHIAPEKVQLQDPYTYLGFELN GPKITNQKAVIRKDKLQTLNDFQKLLG DINWLRPYLKLTTGDLKPLFDTLKGDS DPNSHRSLSKEALASLEKVETAIAEQF	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			VTHINYSLPLIFLIFNTALTPTGLFWQD NPIMWIHLPASPKKVLLPYYDAIADLII LGRDHSKKYFGIEPSTIIQPYSKSQIDW LMQNTEMWPIACASFVGILDNHYPPN KLIQFCKLHTFVFPQIISKTPLNNALLVF TDGSSTGMAAYTLTDTTIKFQTNLNSA QLVELQALIAVLSAFPNQPLNIYTDSAY LAHSIPLLETVAQIKHISETAKLFLQCQ QLIYNRSIPFYIGHVRAHSGLPGPIAQG NQRADLATKIVASNINTNLESAQNAHT LHHLNAQTLRLMFNIPREQARQIVKQC PICVTYLPVPHLGVNPRGLFPNMIWQM DVTHYSEFGNLKYIHVSIDTFSGFLLAT LQTGETTKHVITHLLHCFSIIGLPKQIKT DNGPGYTSKNFQEFCSTLQIKHITGIPY NPQGQGIVERAHLSLKTTIEKIKKGEW YPRKGTPRNILNHALFILNFLNLDDQN KSAADRFWHNNPKKQFAMVKWKDPL DNTWHGPDPVLIWGRGSVCVYSQTYD AARWLPERLVRQVSNNNQSRE	

POL_MMTVB - residues only	P03365	Mouse mammary tumor virus	MGVSGSKGQKLFVSVLQRLLSERGLH VKESSAIEFYQFLIKVSPWFPEEGGLNL QDWKRVGREMKRYAAEHGTDSIPKQ AYPIWLQLREILTEQSDLVLLSAEAKSV TEEELEEGLTGLLSTSSQEKTYGTRGT AYAEIDTEVDKLSEHIYDEPYEEKEKA DKNEEKDHVRKIKKVVQRKENSEGKR KEKDSKAFLATDWNDDDLSPEDWDD LEEQAAHYHDDDELILPVKRKVVKKK PQALRRKPLPPVGFAGAMAEAREKGD LTFTFPVVFMGESDEDDTPVWEPLPLK TLKELQSAVRTMGPSAPYTLQVVDMV ASQWLTPSDWHQTARATLSPGDYVL WRTEYEEKSKEMVQKAAGKRKGKVS LDMLLGTGQFLSPSSQIKLSKDVLKDV TTNAVLAWRAIPPPGVKKTVLAGLKQ GNEESYETFISRLEEAVYRMMPRGEGS DILIKQLAWENANSLCQDLIRPIRKTGT IQDYIRACLDASPAVVQGMAYAAAMR GQKYSTFVKQTYGGGKGGQGAEGPV CFSCGKTGHIRKDCKDEKGSKRAPPGL CPRCKKGYHWKSECKSKFDKDGNPLP PLETNAENSKNLVKGQSPSPAQKGDG VKGSGLNPEAPPFTIHDLPRGTPGSAGL DLSSQKDLILSLEDGVSLVPTLVKGTLP EGTTGLIIGRSSNYKKGLEVLPGVIDSD FQGEIKVMVKAAKNAVIIHKGERIAQL LLLPYLKLPNPVIKEERGSEGFGSTSHV HWVQEISDSRPMLHIYLNGRRFLGLLD TGADKTCIAGRDWPANWPIHQTESSLQ GLGMACGVARSSQPLRWQHEDKSGII HPFVIPTLPFTLWGRDIMKDIKVRLMT DSPDDSQDLMIGAIESNLFADQISWKS DQPVWLNQWPLKQEKLQALQQLVTE QLQLGHLEESNSPWNTPVFVIKKKSGK WRLLQDLRAVNATMHDMGALQPGLP SPVAVPKGWEIIIIDLQDCFFNIKLHPED CKRFAFSVPSPNFKRPYQRFQWKVLPQ GMKNSPTLCQKFVDKAILTVRDKYQD SYIVHYMDDILLAHPSRSIVDEILTSMI QALNKHGLVVSTEKIQKYDNLKYLGT HIQGDSVSYQKLQIRTDKLRTLNDFQK LLGNINWIRPFLKLTTGELKPLFEILNG DSNPISTRKLTPEACKALQLMNERLST ARVKRLDLSQPWSLCILKTEYTPTACL WQDGVVEWIHLPHISPKVITPYDIFCTQ	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			LIIKGRHRSKELFSKDPDYIVVPYTKVQ FDLLLQEKEDWPISLLGFLGEVHFHLP KDPLLTFTLQTAIIFPHMTSTTPLEKGIV IFTDGSANGRSVTYIQGREPIIKENTQN TAQQAEIVAVITAFEEVSQPFNLYTDSK YVTGLFPEIETATLSPRTKIYTELKHLQ RLIHKRQEKFYIGHIRGHTGLPGPLAQG NAYADSLTRILTALESAQESHALHHQN AAALRFQFHITREQAREIVKLCPNCPD WGHAPQLGVNPRGLKPRVLWQMDVT HVSEFGKLKYVHVTVDTYSHFTFATA RTGEATKDVLQHLAQSFAYMGIPQKIK TDNAPAYVSRSIQEFLARWKISHVTGIP YNPQGQAIVERTHQNIKAQLNKLQKA GKYYTPHHLLAHALFVLNHVNMDNQ GHTAAERHWGPISADPKPMVMWKDL LTGSWKGPDVLITAGRGYACVFPQDA ETPIWVPDRFIRPFTERKEATPTPGTAE KTPPRDEKDQQESPKNESSPHQREDGL ATSAGVDLRSGGGP	

POL_MLVMS - residues only	P03355	Moloney murine leukemia virus	MGQTVTTPLSLTLGHWKDVERIAHNQ SVDVKKRRWVTFCSAEWPTFNVGWP RDGTFNRDLITQVKIKVFSPGPHGHPD QVPYIVTWEALAFDPPPWVKPFVHPKP PPPLPPSAPSLPLEPPRSTPPRSSLYPALT PSLGAKPKPQVLSDSGGPLIDLLTEDPP PYRDPRPPPSDRDGNGGEATPAGEAPD PSPMASRLRGRREPPVADSTTSQAFPL RAGGNGQLQYWPFSSSDLYNWKNNN PSFSEDPGKLTALIESVLITHQPTWDDC QQLLGTLLTGEEKQRVLLEARKAVRG DDGRPTQLPNEVDAAFPLERPDWDYT TQAGRNHLVHYRQLLLAGLQNAGRSP TNLAKVKGITQGPNESPSAFLERLKEA YRRYTPYDPEDPGQETNVSMSFIWQSA PDIGRKLERLEDLKNKTLGDLVREAEK IFNKRETPEEREERIRRETEEKEERRRTE DEQKEKERDRRRHREMSKLLATVVSG QKQDRQGGERRRSQLDRDQCAYCKE KGHWAKDCPKKPRGPRGPRPQTSLLT LDDQGGQGQEPPPEPRITLKVGGQPVT FLVDTGAQHSVLTQNPGPLSDKSAWV QGATGGKRYRWTTDRKVHLATGKVT HSFLHVPDCPYPLLGRDLLTKLKAQIH FEGSGAQVMGPMGQPLQVLTLNIEDE HRLHETSKEPDVSLGSTWLSDFPQAW AETGGMGLAVRQAPLIIPLKATSTPVSI KQYPMSQEARLGIKPHIQRLLDQGILV PCQSPWNTPLLPVKKPGTNDYRPVQD LREVNKRVEDIHPTVPNPYNLLSGLPPS HQWYTVLDLKDAFFCLRLHPTSQPLFA FEWRDPEMGISGQLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYV DDLLLAATSELDCQQGTRALLQTLGN LGYRASAKKAQICQKQVKYLGYLLKE GQRWLTEARKETVMGQPTPKTPRQLR EFLGTAGFCRLWIPGFAEMAAPLYPLT KTGTLFNWGPDQQKAYQEIKQALLTA PALGLPDLTKPFELFVDEKQGYAKGVL TQKLGPWRRPVAYLSKKLDPVAAGWP PCLRMVAAIAVLTKDAGKLTMGQPLV ILAPHAVEALVKQPPDRWLSNARMTH YQALLLDTDRVQFGPVVALNPATLLPL PEEGLQHNCLDILAEAHGTRPDLTDQP LPDADHTWYTDGSSLLQEGQRKAGAA VTTETEVIWAKALPAGTSAQRAELIAL	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			TQALKMAEGKKLNVYTDSRYAFATA HIHGEIYRRRGLLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEAR GNRMADQAARKAAITETPDTSTLLIEN SSPYTSEHFHYTVTDIKDLTKLGAIYDK TKKYWVYQGKPVMPDQFTFELLDFLH QLTHLSFSKMKALLERSHSPYYMLNR DRTLKNITETCKACAQVNASKSAVKQ GTRVRGHRPGTHWEIDFTEIKPGLYGY KYLLVFIDTFSGWIEAFPTKKETAKVV TKKLLEEIFPRFGMPQVLGTDNGPAFV SKVSQTVADLLGIDWKLHCAYRPQSS GQVERMNRTIKETLTKLTLATGSRDW VLLLPLALYRARNTPGPHGLTPYEILY GAPPPLVNFPDPDMTRVTNSPSLQAHL QALYLVQHEVWRPLAAAYQEQLDRP VVPHPYRVGDTVWVRRHQTKNLEPR WKGPYTVLLTTPTALKVDGIAAWIHA AHVKAADPGGGPSSRLTWRVQRSQNP LKIRLTREAP	

POL_HTL1A - residues only	P03362	Human T-cell leukemia virus 1	MGQIFSRSASPIPRPPRGLAAHHWLNFL QAAYRLEPGPSSYDFHQLKKFLKIALE TPARICPINYSLLASLLPKGYPGRVNEIL HILIQTQAQIPSRPAPPPPSSPTHDPPDS DPQIPPPYVEPTAPQVLPVMHPHGAPP NHRPWQMKDLQAIKQEVSQAAPGSPQ FMQTIRLAVQQFDPTAKDLQDLLQYL CSSLVASLHHQQLDSLISEAETRGITGY NPLAGPLRVQANNPQQQGLRREYQQL WLAAFAALPGSAKDPSWASILQGLEEP YHAFVERLNIALDNGLPEGTPKDPILRS LAYSNANKECQKLLQARGHTNSPLGD MLRACQTWTPKDKTKVLVVQPKKPPP NQPCFRCGKAGHWSRDCTQPRPPPGP CPLCQDPTHWKRDCPRLKPTIPEPEPEE DALLLDLPADIPHPKNLHRGGGLTSPP TLQQVLPNQDPASILPVIPLDPARRPVI KAQVDTQTSHPKTIEALLDTGADMTV LPIALFSSNTPLKNTSVLGAGGQTQDH FKLTSLPVLIRLPFRTTPIVLTSCLVDTK NNWAIIGRDALQQCQGVLYLPEAKRPP VILPIQAPAVLGLEHLPRPPQISQFPLNP ERLQALQHLVRKALEAGHIEPYTGPGN NPVFPVKKANGTWRFIHDLRATNSLTI DLSSSSPGPPDLSSLPTTLAHLQTIDLRD AFFQIPLPKQFQPYFAFTVPQQCNYGP GTRYAWKVLPQGFKNSPTLFEMQLAH ILQPIRQAFPQCTILQYMDDILLASPSHE DLLLLSEATMASLISHGLPVSENKTQQ TPGTIKFLGQIISPNHLTYDAVPTVPIRS RWALPELQALLGEIQWVSKGTPTLRQP LHSLYCALQRHTDPRDQIYLNPSQVQS LVQLRQALSQNCRSRLVQTLPLLGAIM LTLTGTTTVVFQSKEQWPLVWLHAPL PHTSQCPWGQLLASAVLLLDKYTLQS YGLLCQTIHHNISTQTFNQFIQTSDHPS VPILLHHSHRFKNLGAQTGELWNTFLK TAAPLAPVKALMPVFTLSPVIINTAPCL FSDGSTSRAAYILWDKQILSQRSFPLPP PHKSAQRAELLGLLHGLSSARSWRCL NIFLDSKYLYHYLRTLALGTFQGRSSQ APFQALLPRLLSRKVVYLHHVRSHTNL PDPISRLNALTDALLITPVLQLSPAELHS FTHCGQTALTLQGATTTEASNILRSCH ACRGGNPQHQMPRGHIRRGLLPNHIW QGDITHFKYKNTLYRLHVWVDTFSGAI	IPR043502, SSF56672, IPR000477, PF00078

			SATQKRKETSSEAISSLLQAIAHLGKPS YINTDNGPAYISQDFLNMCTSLAIRHTT HVPYNPTSSGLVERSNGILKTLLYKYF TDKPDLPMDNALSIALWTINHLNVLTN CHKTRWQLHHSPRLQPIPETRSLSNKQ THWYYFKLPGLNSRQWKGPQEALQEA AGAALIPVSASSAQWIPWRLLKRAACP RPVGGPADPKEKDLQHHG	

POL_FOAMV - residues only	P14350	Human spumaretrovirus	MNPLQLLQPLPAEIKGTKLLAHWDSG ATITCIPESFLEDEQPIKKTLIKTIHGEK QQNVYYVTFKVKGRKVEAEVIASPYE YILLSPTDVPWLTQQPLQLTILVPLQEY QEKILSKTALPEDQKQQLKTLFVKYDN LWQHWENQVGHRKIRPHNIATGDYPP RPQKQYPINPKAKPSIQIVIDDLLKQGV LTPQNSTMNTPVYPVPKPDGRWRMVL DYREVNKTIPLTAAQNQHSAGILATIV RQKYKTTLDLANGFWAHPITPESYWL TAFTWQGKQYCWTRLPQGFLNSPALF TADVVDLLKEIPNVQVYVDDIYLSHDD PKEHVQQLEKVFQILLQAGYVVSLKKS EIGQKTVEFLGFNITKEGRGLTDTFKTK LLNITPPKDLKQLQSILGLLNFARNFIPN FAELVQPLYNLIASAKGKYIEWSEENT KQLNMVIEALNTASNLEERLPEQRLVI KVNTSPSAGYVRYYNETGKKPIMYLN YVFSKAELKFSMLEKLLTTMHKALIKA MDLAMGQEILVYSPIVSMTKIQKTPLP ERKALPIRWITWMTYLEDPRIQFHYDK TLPELKHIPDVYTSSQSPVKHPSQYEGV FYTDGSAIKSPDPTKSNNAGMGIVHAT YKPEYQVLNQWSIPLGNHTAQMAEIA AVEFACKKALKIPGPVLVITDSFYVAES ANKELPYWKSNGFVNNKKKPLKHISK WKSIAECLSMKPDITIQHEKGISLQIPVF ILKGNALADKLATQGSYVVNCNTKKP NLDAELDQLLQGHYIKGYPKQYTYFL EDGKVKVSRPEGVKIIPPQSDRQKIVLQ AHNLAHTGREATLLKIANLYWWPNM RKDVVKQLGRCQQCLITNASNKASGPI LRPDRPQKPFDKFFIDYIGPLPPSQGYL YVLVVVDGMTGFTWLYPTKAPSTSAT VKSLNVLTSIAIPKVIHSDQGAAFTSST FAEWAKERGIHLEFSTPYHPQSGSKVE RKNSDIKRLLTKLLVGRPTKWYDLLPV VQLALNNTYSPVLKYTPHQLLFGIDSN TPFANQDTLDLTREEELSLLQEIRTSLY HPSTPPASSRSWSPVVGQLVQERVARP ASLRPRWHKPSTVLKVLNPRTVVILDH LGNNRTVSIDNLKPTSHQNGTTNDTAT MDHLEKNE	IPR043502, SSF56672, IPR000477, PF00078

POL_BLVJ - residues only	P03361	Bovine leukemia virus	MGNSPSYNPPAGISPSDWLNLLQSAQR LNPRPSPSDFTDLKNYIHWFHKTQKKP WTFTSGGPTSCPPGRFGRVPLVLATLN EVLSNEGGAPGASAPEEQPPPYDPPAIL PIISEGNRNRHRAWALRELQDIKKEIEN KAPGSQVWIQTLRLAILQADPTPADLE QLCQYIASPVDQTAHMTSLTAAIAAAE AANTLQGFNPKTGTLTQQSAQPNAGD LRSQYQNLWLQAGKNLPTRPSAPWSTI VQGPAESSVEFVNRLQISLADNLPDGV PKEPIIDSLSYANANRECQQILQGRGPV AAVGQKLQACAQWAPKNKQPALLVH TPGPKMPGPRQPAPKRPPPGPCYRCLK EGHWARDCPTKATGPPPGPCPICKDPS HWKRDCPTLKSKNKLIEGGLSAPQTIT PITDSLSEAELECLLSIPLARSRPSVAVY LSGPWLQPSQNQALMLVDTGAENTVL PQNWLVRDYPRIPAAVLGAGGVSRNR YNWLQGPLTLALKPEGPFITIPKILVDT SDKWQILGRDVPSRLQASISIPEEVRPP VVGVLDTPPSHIGLEHLPPPPEVPQFPL NLERLQALQDLVHRSLEAGYISPWDGP GNNPVFPVRKPNGAWRFVHDLRATNA LTKPIPALSPGPPDLTAIPTHPPHIICLDL KDAFFQIPVEDRFRFYLSFTLPSPGGLQ PHRRFAWRVLPQGFINSPALFERALQE PLRQVSAAFSQSLLVSYMDDILYASPT EEQRSQCYQALAARLRDLGFQVASEK TSQTPSPVPFLGQMVHEQIVTYQSLPTL QISSPISLHQLQAVLGDLQWVSRGTPTT RRPLQLLYSSLKRHHDPRAIIQLSPEQL QGIAELRQALSHNARSRYNEQEPLLAY VHLTRAGSTLVLFQKGAQFPLAYFQTP LTDNQASPWGLLLLLGCQYLQTQALS SYAKPILKYYHNLPKTSLDNWIQSSED PRVQELLQLWPQISSQGIQPPGPWKTLI TRAEVFLTPQFSPDPIPAALCLFSDGAT GRGAYCLWKDHLLDFQAVPAPESAQK GELAGLLAGLAAAPPEPVNIWVDSKY LYSLLRTLVLGAWLQPDPVPSYALLYK SLLRHPAIVVGHVRSHSSASHPIASLNN YVDQLLPLETPEQWHKLTHCNSRALS RWPNPRISAWDPRSPATLCETCQKLNP TGGGKMRTIQRGWAPNHIWQADITHY KYKQFTYALHVFVDTYSGATHASAKR GLTTQTTIEGLLEAIVHLGRPKKLNTD	IPR043502, SSF56672, IPR000477, PF00078

			QGANYTSKTFVRFCQQFGVSLSHHVP YNPTSSGLDERTNGLLKLLLSKYHLDE PHLPMTQALSRALWTHNQINLLPILKT RWELHHSPPLAVISEGGETPKGSDKLF LYLLPGQNNRRWLGPLPALVEASGGA LLATDPPVWVPWRLLKAFKCLKNDGP EDAHNRSSDG	

O41894_9RETR - residues only	O41894	Bovine foamy virus	MPALRPLQVEIKGNHLKGYWDSGAEI TCVPAIYIIEEQPVGKKLITTIHNEKEHD VYYVEMKIEKRKVQCEVIATALDYVL VAPVDIPWYKPGPLELTIKIDVESQKHT LITESTLSPQGQMRLKKLLDQYQALW QCWENQVGHRRIEPHKIATGALKPRPQ KQYHINPRAKADIQIVIDDLLRQGVLR QQNSEMNTPVYPVPKADGRWRMVLD YREVNKVTPLVATQNCHSASILNTLYR GPYKSTLDLANGFWAHPIKPEDYWITA FTWGGKTYCWTVLPQGFLNSPALFTA DVVDILKDIPNVQVYVDDVYVSSATE QEHLDILETIFNRLSTAGYIVSLKKSKL AKETVEFLGFSISQNGRGLTDSYKQKL MDLQPPTTLRQLQSILGLINFARNFLPN FAELVAPLYQLIPKAKGQCIPWTMDHT TQLKTIIQALNSTENLEERRPDVDLIMK VHISNTAGYIRFYNHGGQKPIAYNNAL FTSTELKFTPTEKIMATIHKGLLKALDL SLGKEIHVYSAIASMTKLQKTPLSERK ALSIRWLKWQTYFEDPRIKFHHDATLP DLQNLPVPQQDTGKEMTILPLLHYEAI FYTDGSAIRSPKPNKTHSAGMGIIQAKF EPDFRIVHLWSFPLGDHTAQYAEIAAF EFAIRRATGIRGPVLIVTDSNYVAKSYN EELPYWESNGFVNNKKKTLKHISKWK AIAECKNLKADIHVIHEPGHQPAEASP HAQGNALADKQAVSGSYKVFSNELKP SLDAELEQVLSTGRPNPQGYPNKYEYK LVNGLCYVDRRGEEGLKIIPPKADRVK LCQLAHDGPGSAHLGRSALLLKLQQK YWWPRMHIDASRIVLNCTVCAQTNST NQKPRPPLVIPHDTKPFQVWYMDYIGP LPPSNGYQHALVIVDAGTGFTWIYPTK AQTANATVKALTHLTGTAVPKVLHSD QGPAFTSSILADWAKDRGIQLEHSAPY HPQSSGKVERKNSEIKRLLTKLLAGRP TKWYPLIPIVQLALNNTPNTRQKYTPH QLMYGADCNLPFENLDTLDLTREEQL AVLKEVRDGLLDLYPSPSQTTARSWTP SPGLLVQERVARPAQLRPKWRKPTPIK KVLNERTVIIDHLGQDKVVSIDNLKPA AHQKLAQTPDSAEICPSATPCPPNTSL WYDLDTGTWTCQRCGYQCPDKYHQP QCTWSCEDRCGHRWKECGNCIPQDGS SDDASAVAAVEI	IPR043502, SSF56672, IPR000477, PF00078

POL_MLVBM - residues only	Q7SVK7	Murine leukemia virus	MGQTVTTPLSLTLEHWGDVQRIASNQ SVGVKKRRWVTFCSAEWPTFGVGWP QDGTFNLDIILQVKSKVFSPGPHGHPD QVPYIVTWEAIAYEPPPWVKPFVSPKL SLSPTAPILPSGPSTQPPPRSALYPAFTP SIKPRPSKPQVLSDDGGPLIDLLTEDPPP YGEQGPSSPDGDGDREEATSTSEIPAPS PMVSRLRGKRDPPAADSTTSRAFPLRL GGNGQLQYWPFSSSDLYNWKNNNPSF SEDPGKLTALIESVLTTHQPTWDDCQQ LLGTLLTGEEKQRVLLEARKAVRGND GRPTQLPNEVNSAFPLERPDWDYTTPE GRNHLVLYRQLLLAGLQNAGRSPTNL AKVKGITQGPNESPSAFLERLKEAYRR YTPYDPEDPGQETNVSMSFIWQSAPAI GRKLERLEDLKSKTLGDLVREAEKIFN KRETPEEREERIRRETEEKEERRRAGDE QREKERDRRRQREMSKLLATVVTGQR QDRQGGERRRPQLDKDQCAYCKEKG HWAKDCPKKPRGPRGPRPQTSLLTLD DQGGQGQEPPPEPRITLTVGGQPVTFL VDTGAQHSVLTQNPGPLSDRSAWVQG ATGGKRYRWTTDRKVHLATGKVTHSF LHVPDCPYPLLGRDLLTKLKAQIHFEG SGAQVVGPKGQPLQVLTLGIEDEYRLH ETSTEPDVSLGSTWLSDFPQAWAETGG MGLAVRQAPLIIPLKATSTPVSIQQYPM SHEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNK RVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRD PGMGISGQLTWTRLPQGFKNSPTLFDE ALHRDLADFRIQHPDLILLQYVDDILLA ATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLGYLLREGQRWLT EARKETVMGQPVPKTPRQLREFLGTA GFCRLWIPGFAEMAAPLYPLTKTGTLF SWGPDQQKAYQEIKQALLTAPALGLP DLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRM VAAIAVLTKDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNARMTHYQAM LLDTDRVQFGPVVALNPATLLPLPEEG APHDCLEILAETHGTRPDLTDQPIPDAD HTWYTDGSSFLQEGQRKAGAAVTTET EVIWAGALPAGTSAQRAELIALTQALK	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			MAEGKRLNVYTDSRYAFATAHIHGEI YRRRGLLTSEGREIKNKSEILALLKALF LPKRLSIIHCLGHQKGDSAEARGNRLA DQAAREAAIKTPPDTSTLLIEDSTPYTP AYFHYTETDLKKLRDLGATYNQSKGY WVFQGKPVMPDQFVFELLDSLHRLTH LGYQKMKALLDRGESPYYMLNRDKT LQYVADSCTVCAQVNASKAKIGAGVR VRGHRPGTHWEIDFTEVKPGLYGYKY LLVFVDTFSGWVEAFPTKRETARVVSK KLLEEIFPRFGMPQVLGSDNGPAFTSQ VSQSVADLLGIDWKLHCAYRPQSSGQ VERINRTIKETLTKLTLAAGTRDWVLL LPLALYRARNTPGPHGLTPYEILYGAPP PLVNFHDPDMSELTNSPSLQAHLQALQ TVQREIWKPLAEAYRDRLDQPVIPHPF RIGDSVWVRRHQTKNLEPRWKGPYTV LLTTPTALKVDGISAWIHAAHVKAATT PPIKPSWRVQRSQNPLKIRLTRGAP	

Table 31: Exemplary dimeric retroviral reverse transcriptases and their RT domain signatures

Name	Accession	Organism	Sequence	RT Signatures

Q8313 3_AVI MA	Q83133	Avian myeloblastosis-associated virus type 1	RATVLTVALHLAIPLKWKPNHTPVWID QWPLPEGKLVALTQLVEKELQLGHIEP SLSCWNTPVFVIRKASGSYRLLHDLRA VNAKLVPFGAVQQGAPVLSALPRGWP LMVLDLKDCFFSIPLAEQDREAFAFTLP SVNNQAPARRFQWKVLPQGMTCSPTI CQLIVGQILEPLRLKHPSLRMLHYMDD LLLAASSHDGLEAAGEEVISTLERAGF TISPDKVQREPGVQYLGYKLGSTYVAP VGLVAEPRIATLWDVQKLVGSLQSVR PALGIPPRLMGPFYEQLRGSDPNEARE WNLDMKMAWREIVQLSTTAALERWD PALPLEGAVARCEQGAIGVLGQGLSTH PRPCLWLFSTQPTKAFTAWLEVLTLLIT KLRASAVRTFGKEVDILLLPACFREDL PLPEGILLALRGFAGKIRSSDTPSIFDIA RPLHVSLKVRVTDHPVPGPTVFTDASS STHKGVVVWREGPRWEIKEIADLGAS VQQLEARAVAMALLLWPTTPTNVVTD SAFVAKMLLKMGQEGVPSTAAAFILE DALSQRSAMAAVLHVRSHSEVPGFFTE GNDVADSQATFQAYPLREAKDLHTAL HIGPRALSKACNISMQQAREVVQTCPH CNSAPALEAGVNPRGLGPLQIWQTDFT LEPRMAPRSWLAVTVDTASSAIVVTQ HGRVTSVAAQHHWATAIAVLGRPKAI KTDNGSCFTSKSTREWLARWGIAHTT GIPGNSQGQAMVERANRLLKDKIRVL AEGDGFMKRIPTSKQGELLAKAMYAL NHFERGENTKTPIQKHWRPTVLTEGPP VKIRIETGEWEKGWNVLVWGRGYAA VKNRDTDKVIWVPSRKVKPDITQKDE VTKKDEASPLFAGISDWAPWEGEQEG LQEETASNKQERPGEDTPAANES	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

POL_SIVM1	P05896	Simian immunodeficiency virus	MGARNSVLSGKKADELEKIRLRPGGK KKYMLKHVVWAANELDRFGLAESLL ENKEGCQKILSVLAPLVPTGSENLKSL YNTVCVIWCIHAEEKVKHTEEAKQIVQ RHLVMETGTAETMPKTSRPTAPFSGRG GNYPVQQIGGNYTHLPLSPRTLNAWV KLIEEKKFGAEVVSGFQALSEGCLPYDI NQMLNCVGDHQAAMQIIRDIINEEAAD WDLQHPQQAPQQGQLREPSGSDIAGT TSTVEEQIQWMYRQQNPIPVGNIYRRW IQLGLQKCVRMYNPTNILDVKQGPKEP FQSYVDRFYKSLRAEQTDPAVKNWMT QTLLIQNANPDCKLVLKGLGTNPTLEE MLTACQGVGGPGQKARLMAEALKEA LAPAPIPFAAAQQKGPRKPIKCWNCGK EGHSARQCRAPRRQGCWKCGKMDHV MAKCPNRQAGFFRPWPLGKEAPQFPH GSSASGADANCSPRRTSCGSAKELHAL GQAAERKQREALQGGDRGFAAPQFSL WRRPVVTAHIEGQPVEVLLDTGADDSI VTGIELGPHYTPKIVGGIGGFINTKEYK NVEIEVLGKRIKGTIMTGDTPINIFGRN LLTALGMSLNLPIAKVEPVKSPLKPGK DGPKLKQWPLSKEKIVALREICEKMEK DGQLEEAPPTNPYNTPTFAIKKKDKNK WRMLIDFRELNRVTQDFTEVQLGIPHP AGLAKRKRITVLDIGDAYFSIPLDEEFR QYTAFTLPSVNNAEPGKRYIYKVLPQG WKGSPAIFQYTMRHVLEPFRKANPDV TLVQYMDDILIASDRTDLEHDRVVLQL KELLNSIGFSSPEEKFQKDPPFQWMGY ELWPTKWKLQKIELPQRETWTVNDIQ KLVGVLNWAAQIYPGIKTKHLCRLIRG KMTLTEEVQWTEMAEAEYEENKIILSQ EQEGCYYQESKPLEATVIKSQDNQWS YKIHQEDKILKVGKFAKIKNTHTNGVR LLAHVIQKIGKEAIVIWGQVPKFHLPVE KDVWEQWWTDYWQVTWIPEWDFIST PPLVRLVFNLVKDPIEGEETYYVDGSC SKQSKEGKAGYITDRGKDKVKVLEQT TNQQAELEAFLMALTDSGPKANIIVDS QYVMGIITGCPTESESRLVNQIIEEMIK KTEIYVAWVPAHKGIGGNQEIDHLVSQ GIRQVLFLEKIEPAQEEHSKYHSNIKEL VFKFGLPRLVAKQIVDTCDKCHQKGE AIHGQVNSDLGTWQMDCTHLEGKIVI	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			VAVHVASGFIEAEVIPQETGRQTALFLL KLASRWPITHLHTDNGANFASQEVKM VAWWAGIEHTFGVPYNPQSQGVVEA MNHHLKNQIDRIREQANSVETIVLMAV HCMNFKRRGGIGDMTPAERLINMITTE QEIQFQQSKNSKFKNFRVYYREGRDQL WKGPGELLWKGEGAVILKVGTDIKVV PRRKAKIIKDYGGGKEMDSSSHMEDT GEAREVA	

POL_ RSVP	P03354	Rous sarcoma virus	MEAVIKVISSACKTYCGKTSPSKKEIGA MLSLLQKEGLLMSPSDLYSPGSWDPIT AALSQRAMILGKSGELKTWGLVLGAL KAAREEQVTSEQAKFWLGLGGGRVSP PGPECIEKPATERRIDKGEEVGETTVQR DAKMAPEETATPKTVGTSCYHCGTAI GCNCATASAPPPPYVGSGLYPSLAGVG EQQGQGGDTPPGAEQSRAEPGHAGQA PGPALTDWARVREELASTGPPVVAMP VVIKTEGPAWTPLEPKLITRLADTVRT KGLRSPITMAEVEALMSSPLLPHDVTN LMRVILGPAPYALWMDAWGVQLQTVI AAATRDPRHPANGQGRGERTNLNRLK GLADGMVGNPQGQAALLRPGELVAIT ASALQAFREVARLAEPAGPWADIMQG PSESFVDFANRLIKAVEGSDLPPSARAP VIIDCFRQKSQPDIQQLIRTAPSTLTTPG EIIKYVLDRQKTAPLTDQGIAAAMSSAI QPLIMAVVNRERDGQTGSGGRARGLC YTCGSPGHYQAQCPKKRKSGNSRERC QLCNGMGHNAKQCRKRDGNQGQRPG KGLSSGPWPGPEPPAVSLAMTMEHKD RPLVRVILTNTGSHPVKQRSVYITALLD SGADITIISEEDWPTDWPVMEAANPQI HGIGGGIPMRKSRDMIELGVINRDGSL ERPLLLFPAVAMVRGSILGRDCLQGLG LRLTNLIGRATVLTVALHLAIPLKWKP DHTPVWIDQWPLPEGKLVALTQLVEK ELQLGHIEPSLSCWNTPVFVIRKASGSY RLLHDLRAVNAKLVPFGAVQQGAPVL SALPRGWPLMVLDLKDCFFSIPLAEQD REAFAFTLPSVNNQAPARRFQWKVLP QGMTCSPTICQLVVGQVLEPLRLKHPS LCMLHYMDDLLLAASSHDGLEAAGEE VISTLERAGFTISPDKVQREPGVQYLGY KLGSTYVAPVGLVAEPRIATLWDVQK LVGSLQWLRPALGIPPRLMGPFYEQLR GSDPNEAREWNLDMKMAWREIVRLST TAALERWDPALPLEGAVARCEQGAIG VLGQGLSTHPRPCLWLFSTQPTKAFTA WLEVLTLLITKLRASAVRTFGKEVDIL LLPACFREDLPLPEGILLALKGFAGKIR SSDTPSIFDIARPLHVSLKVRVTDHPVP GPTVFTDASSSTHKGVVVWREGPRWE IKEIADLGASVQQLEARAVAMALLLW PTTPTNVVTDSAFVAKMLLKMGQEGV	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

			PSTAAAFILEDALSQRSAMAAVLHVRS HSEVPGFFTEGNDVADSQATFQAYPLR EAKDLHTALHIGPRALSKACNISMQQA REVVQTCPHCNSAPALEAGVNPRGLG PLQIWQTDFTLEPRMAPRSWLAVTVD TASSAIVVTQHGRVTSVAVQHHWATA IAVLGRPKAIKTDNGSCFTSKSTREWL ARWGIAHTTGIPGNSQGQAMVERANR LLKDRIRVLAEGDGFMKRIPTSKQGEL LAKAMYALNHFERGENTKTPIQKHWR PTVLTEGPPVKIRIETGEWEKGWNVLV WGRGYAAVKNRDTDKVIWVPSRKVK PDITQKDEVTKKDEASPLFAGISDWIP WEDEQEGLQGETASNKQERPGEDTLA ANES	

POL_HV2D2	P15833	Human immunodeficiency virus type 2	MGARGSVLSGKKTDELEKVRLRPGGK KKYMLKHVVWAVNELDRFGLAESLL ESKEGCQKILKVLAPLVPTGSENLKSLF NIVCVIFCLHAEEKVKDTEEAKKIAQR HLAADTEKMPATNKPTAPPSGGNYPV QQLAGNYVHLPLSPRTLNAWVKLVEE KKFGAEVVPGFQALSEGCTPYDINQML NCVGEHQAAMQIIREIINEEAADWDQQ HPSPGPMPAGQLRDPRGSDIAGTTSTV EEQIQWMYRAQNPVPVGNIYRRWIQL GLQKCVRMYNPTNILDIKQGPKEPFQS YVDRFYKSLRAEQTDPAVKNWMTQT LLIQNANPDCKLVLKGLGMNPTLEEM LTACQGIGGPGQKARLMAEALKEALT PAPIPFAAVQQKAGKRGTVTCWNCGK QGHTARQCRAPRRQGCWKCGKTGHI MSKCPERQAGFLRVRTLGKEASQLPH DPSASGSDTICTPDEPSRGHDTSGGDTI CAPCRSSSGDAEKLHADGETTEREPRE TLQGGDRGFAAPQFSLWRRPVVKACIE GQSVEVLLDTGVDDSIVAGIELGSNYT PKIVGGIGGFINTKEYKDVEIEVVGKRV RATIMTGDTPINIFGRNILNTLGMTLNF PVAKVEPVKVELKPGKDGPKIRQWPLS REKILALKEICEKMEKEGQLEEAPPTNP YNTPTFAIKKKDKNKWRMLIDFRELN KVTQDFTEVNWVFPTRQVAEKRRITVI DVGDAYFSIPLDPNFRQYTAFTLPSVN NAEPGKRYIYKVLPQGWKGSQSICQYS MRKVLDPFRKANSDVIIIQYMDDILIAS DRSDLEHDRVVSQLKELLNDMGFSTPE EKFQKDPPFKWMGYELWPKKWKLQK IQLPEKEVWTVNAIQKLVGVLNWAAQ LFPGIKTRHICKLIRGKMTLTEEVQWTE LAEAELQENKIILEQEQEGSYYKERVPL EATVQKNLANQWTYKIHQGNKVLKV GKYAKVKNTHTNGVRLLAHVVQKIG KEALVIWGEIPVFHLPVERETWDQWW TDYWQVTWIPEWDFVSTPPLIRLAYNL VKDPLEGRETYYTDGSCNRTSKEGKA GYVTDRGKDKVKVLEQTTNQQAELEA FALALTDSEPQVNIIVDSQYVMGIIAAQ PTETESPIVAKIIEEMIKKEAVYVGWVP AHKGLGGNQEVDHLVSQGIRQVLFLE KIEPAQEEHEKYHGNVKELVHKFGIPQ LVAKQIVNSCDKCQQKGEAIHGQVNA	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

			DLGTWQMDCTHLEGKIIIVAVHVASGF IEAEVIPQETGRQTALFLLKLASRWPIT HLHTDNGANFTSPSVKMVAWWVGIE QTFGVPYNPQSQGVVEAMNHHLKNQI DRLRDQAVSIETVVLMATHCMNFKRR GGIGDMTPAERLVNMITTEQEIQFFQA KNLKFQNFQVYYREGRDQLWKGPGEL LWKGEGAVIIKVGTEIKVVPRRKAKIIR HYGGGKGLDCSADMEDTRQAREMAQ SD	

POL_HV1A2	P03369	Human immunodeficiency virus type 1	MGARASVLSGGELDKWEKIRLRPGGK KKYKLKHIVWASRELERFAVNPGLLET SEGCRQILGQLQPSLQTGSEELRSLYNT VATLYCVHQRIDVKDTKEALEKIEEEQ NKSKKKAQQAAAAAGTGNSSQVSQN YPIVQNLQGQMVHQAISPRTLNAWVK VVEEKAFSPEVIPMFSALSEGATPQDL NTMLNTVGGHQAAMQMLKETINEEA AEWDRVHPVHAGPIAPGQMREPRGSD IAGTTSTLQEQIGWMTNNPPIPVGEIYK RWIILGLNKIVRMYSPTSILDIRQGPKEP FRDYVDRFYKTLRAEQASQDVKNWM TETLLVQNANPDCKTILKALGPAATLE EMMTACQGVGGPGHKARVLAEAMSQ VTNPANIMMQRGNFRNQRKTVKCFNC GKEGHIAKNCRAPRKKGCWRCGREGH QMKDCTERQANFLREDLAFLQGKARE FSSEQTRANSPTRRELQVWGGENNSLS EAGADRQGTVSFNFPQITLWQRPLVTI RIGGQLKEALLDTGADDTVLEEMNLP GKWKPKMIGGIGGFIKVRQYDQIPVEI CGHKAIGTVLVGPTPVNIIGRNLLTQIG CTLNFPISPIETVPVKLKPGMDGPKVKQ WPLTEEKIKALVEICTEMEKEGKISKIG PENPYNTPVFAIKKKDSTKWRKLVDFR ELNKRTQDFWEVQLGIPHPAGLKKKK SVTVLDVGDAYFSVPLDKDFRKYTAF TIPSINNETPGIRYQYNVLPQGWKGSPA IFQSSMTKILEPFRKQNPDIVIYQYMDD LYVGSDLEIGQHRTKIEELRQHLLRWG FTTPDKKHQKEPPFLWMGYELHPDKW TVQPIMLPEKDSWTVNDIQKLVGKLN WASQIYAGIKVKQLCKLLRGTKALTE VIPLTEEAELELAENREILKEPVHEVYY DPSKDLVAEIQKQGQGQWTYQIYQEPF KNLKTGKYARMRGAHTNDVKQLTEA VQKVSTESIVIWGKIPKFKLPIQKETWE AWWMEYWQATWIPEWEFVNTPPLVK LWYQLEKEPIVGAETFYVDGAANRET KLGKAGYVTDRGRQKVVSIADTTNQK TELQAIHLALQDSGLEVNIVTDSQYAL GIIQAQPDKSESELVSQIIEQLIKKEKVY LAWVPAHKGIGGNEQVDKLVSAGIRK VLFLNGIDKAQEEHEKYHSNWRAMAS DFNLPPVVAKEIVASCDKCQLKGEAM HGQVDCSPGIWQLDCTHLEGKIILVAV	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661, PF06815, IPR010659

			HVASGYIEAEVIPAETGQETAYFLLKL AGRWPVKTIHTDNGSNFTSTTVKAAC WWAGIKQEFGIPYNPQSQGVVESMNN ELKKIIGQVRDQAEHLKTAVQMAVFIH NFKRKGGIGGYSAGERIVDIIATDIQTK ELQKQITKIQNFRVYYRDNKDPLWKG PAKLLWKGEGAVVIQDNSDIKVVPRR KAKIIRDYGKQMAGDDCVASRQDED	

POL_FIVPE	P16088	Feline immunodeficiency virus	KEFGKLEGGASCSPSESNAASSNAICTS NGGETIGFVNYNKVGTTTTLEKRPEILI FVNGYPIKFLLDTGADITILNRRDFQVK NSIENGRQNMIGVGGGKRGTNYINVH LEIRDENYKTQCIFGNVCVLEDNSLIQP LLGRDNMIKFNIRLVMAQISDKIPVVK VKMKDPNKGPQIKQWPLTNEKIEALTE IVERLEKEGKVKRADSNNPWNTPVFAI KKKSGKWRMLIDFRELNKLTEKGAEV QLGLPHPAGLQIKKQVTVLDIGDAYFT IPLDPDYAPYTAFTLPRKNNAGPGRRF VWCSLPQGWILSPLIYQSTLDNIIQPFIR QNPQLDIYQYMDDIYIGSNLSKKEHKE KVEELRKLLLWWGFETPEDKLQEEPP YTWMGYELHPLTWTIQQKQLDIPEQPT LNELQKLAGKINWASQAIPDLSIKALT NMMRGNQNLNSTRQWTKEARLEVQK AKKAIEEQVQLGYYDPSKELYAKLSLV GPHQISYQVYQKDPEKILWYGKMSRQ KKKAENTCDIALRACYKIREESIIRIGK EPRYEIPTSREAWESNLINSPYLKAPPP EVEYIHAALNIKRALSMIKDAPIPGAET WYIDGGRKLGKAAKAAYWTDTGKW RVMDLEGSNQKAEIQALLLALKAGSE EMNIITDSQYVINIILQQPDMMEGIWQE VLEELEKKTAIFIDWVPGHKGIPGNEE VDKLCQTMMIIEGDGILDKRSEDAGYD LLAAKEIHLLPGEVKVIPTGVKLMLPK GYWGLIIGKSSIGSKGLDVLGGVIDEG YRGEIGVIMINVSRKSITLMERQKIAQL IILPCKHEVLEQGKVVMDSERGDNGY GSTGVFSSWVDRIEEAEINHEKFHSDP QYLRTEFNLPKMVAEEIRRKCPVCRIIG EQVGGQLKIGPGIWQMDCTHFDGKIIL VGIHVESGYIWAQIISQETADCTVKAV LQLLSAHNVTELQTDNGPNFKNQKME GVLNYMGVKHKFGIPGNPQSQALVEN VNHTLKVWIQKFLPETTSLDNALSLAV HSLNFKRRGRIGGMAPYELLAQQESLR IQDYFSAIPQKLQAQWIYYKDQKDKK WKGPMRVEYWGQGSVLLKDEEKGYF LIPRRHIRRVPEPCALPEGDE	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

POL_EIAVY	P03371	Equine infectious anemia virus	TAWTFLKAMQKCSKKREARGSREAPE TNFPDTTEESAQQICCTRDSSDSKSVPR SERNKKGIQCQGEGSSRGSQPGQFVGV TYNLEKRPTTIVLINDTPLNVLLDTGAD TSVLTTAHYNRLKYRGRKYQGTGIIGV GGNVETFSTPVTIKKKGRHIKTRMLVA DIPVTILGRDILQDLGAKLVLAQLSKEI KFRKIELKEGTMGPKIPQWPLTKEKLE GAKETVQRLLSEGKISEASDNNPYNSPI FVIKKRSGKWRLLQDLRELNKTVQVG TEISRGLPHPGGLIKCKHMTVLDIGDA YFTIPLDPEFRPYTAFTIPSINHQEPDKR YVWKCLPQGFVLSPYIYQKTLQEILQP FRERYPEVQLYQYMDDLFVGSNGSKK QHKELIIELRAILQKGFETPDDKLQEVP PYSWLGYQLCPENWKVQKMQLDMVK NPTLNDVQKLMGNITWMSSGVPGLTV KHIAATTKGCLELNQKVIWTEEAQKEL EENNEKIKNAQGLQYYNPEEEMLCEV EITKNYEATYVIKQSQGILWAGKKIMK ANKGWSTVKNLMLLLQHVATESITRV GKCPTFKVPFTKEQVMWEMQKGWYY SWLPEIVYTHQVVHDDWRMKLVEEPT SGITIYTDGGKQNGEGIAAYVTSNGRT KQKRLGPVTHQVAERMAIQMALEDTR DKQVNIVTDSYYCWKNITEGLGLEGP QNPWWPIIQNIREKEIVYFAWVPGHKG IYGNQLADEAAKIKEEIMLAYQGTQIK EKRDEDAGFDLCVPYDIMIPVSDTKIIP TDVKIQVPPNSFGWVTGKSSMAKQGL LINGGIIDEGYTGEIQVICTNIGKSNIKLI EGQKFAQLIILQHHSNSRQPWDENKIS QRGDKGFGSTGVFWVENIQEAQDEHE NWHTSPKILARNYKIPLTVAKQITQECP HCTKQGSGPAGCVMRSPNHWQADCT HLDNKIILHFVESNSGYIHATLLSKENA LCTSLAILEWARLFSPKSLHTDNGTNF VAEPVVNLLKFLKIAHTTGIPYHPESQG IVERANRTLKEKIQSHRDNTQTLEAAL QLALITCNKGRESMGGQTPWEVFITNQ AQVIHEKLLLQQAQSSKKFCFYKIPGE HDWKGPTRVLWKGDGAVVVNDEGK GIIAVPLTRTKLLIKPN	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659

POL_ BIV29	P19560	Bovine immunodeficiency virus	MKRRELEKKLRKVRVTPQQDKYYTIG NLQWAIRMINLMGIKCVCDEECSAAE VALIITQFSALDLENSPIRGKEEVAIKNT LKVFWSLLAGYKPESTETALGYWEAF TYREREARADKEGEIKSIYPSLTQNTQ NKKQTSNQTNTQSLPAITTQDGTPRFD PDLMKQLKIWSDATERNGVDLHAVNI LGVITANLVQEEIKLLLNSTPKWRLDV QLIESKVREKENAHRTWKQHHPEAPK TDEIIGKGLSSAEQATLISVECRETFRQ WVLQAAMEVAQAKHATPGPINIHQGP KEPYTDFINRLVAALEGMAAPETTKEY LLQHLSIDHANEDCQSILRPLGPNTPME KKLEACRVVGSQKSKMQFLVAAMKE MGIQSPIPAVLPHTPEAYASQTSGPEDG RRCYGCGKTGHLKRNCKQQKCYHCG KPGHQARNCRSKNREVLLCPLWAEEP TTEQFSPEQHEFCDPICTPSYIRLDKQPF IKVFIGGRWVKGLVDTGADEVVLKNI HWDRIKGYPGTPIKQIGVNGVNVAKR KTHVEWRFKDKTGIIDVLFSDTPVNLF GRSLLRSIVTCFTLLVHTEKIEPLPVKV RGPGPKVPQWPLTKEKYQALKEIVKD LLAEGKISEAAWDNPYNTPVFVIKKKG TGRWRMLMDFRELNKITVKGQEFSTG LPYPPGIKECEHLTAIDIKDAYFTIPLHE DFRPFTAFSVVPVNREGPIERFQWNVL PQGWVCSPAIYQTTTQKIIENIKKSHPD VMLYQYMDDLLIGSNRDDHKQIVQEI RDKLGSYGFKTPDEKVQEERVKWIGF ELTPKKWRFQPRQLKIKNPLTVNELQQ LVGNCVWVQPEVKIPLYPLTDLLRDKT NLQEKIQLTPEAIKCVEEFNLKLKDPE WKDRIREGAELVIKIQMVPRGIVFDLL QDGNPIWGGVKGLNYDHSNKIKKILRT MNELNRTVVIMTGREASFLLPGSSEDW EAALQKEESLTQIFPVKFYRHSCRWTSI CGPVRENLTTYYTDGGKKGKTAAAVY WCEGRTKSKVFPGTNQQAELKAICMA LLDGPPKMNIITDSRYAYEGMREEPET WAREGIWLEIAKILPFKQYVGVGWVP AHKGIGGNTEADEGVKKALEQMAPCS PPEAILLKPGEKQNLETGIYMQGLRPQS FLPRADLPVAITGTMVDSELQLQLLNI GTEHIRIQKDEVFMTCFLENIPSATEDH ERWHTSPDILVRQFHLPKRIAKEIVARC	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661

			QECKRTTTSPVRGTNPRGRFLWQMDN THWNKTIIWVAVETNSGLVEAQVIPEE TALQVALCILQLIQRYTVLHLHSDNGP CFTAHRIENLCKYLGITKTTGIPYNPQS QGVVERAHRDLKDRLAAYQGDCETV EAALSLALVSLNKKRGGIGGHTPYEIY LESEHTKYQDQLEQQFSKQKIEKWCY VRNRRKEWKGPYKVLWDGDGAAVIE EEGKTALYPHRHMRFIPPPDSDIQDGSS	

			TVALHLAIPLKWKPDHTPVWIDQWPL PEGKLVALTQLVEKELQLGHIEPSLSC WNTPVFVIRKASGSYRLLHDLRAVNA KLVPFGAVQQGAPVLSALPRGWPLMV LDLKDCFFSIPLAEQDREAFAFTLPSVN NQAPARRFQWKVLPQGMTCSPTICQL VVGQVLEPLRLKHPSLRMLHYMDDLL LAASSHDGLEAAGEEVISTLERAGFTIS PDKIQREPGVQYLGYKLGSTYVAPVGL VAEPRIATLWDVQKLVGSLQWLRPAL GIPPRLMGPFYEQLRGSDPNEAREWNL DMKMAWREIVQLSTTAALERWDPALP LEGAVARCEQGAIGVLGQGLSTHPRPC LWLFSTQPTKAFTAWLEVLTLLITKLR ASAVRTFGKEVDVLLLPACFREDLPLP EGILLALRGFAGKIRSSDTPSIFDIARPL HVSLKVRVTDHPVPGPTVFTDASSSTH KGVVVWREGPRWEIKEIADLGASVQQ LEARAVAMALLLWPTTPTNVVTDSAF VAKMLLKMGQEGVPSTAAAFILEDAL SQRSAMAAVLHVRSHSEVPGFFTEGN DVADSQATFQAYPLREAKDLHTALHI GPRALSKACNISMQQAREVVQTCPHC NSAPALEAGVNPRGLGPLQIWQTDFTL EPRMAPRSWLAVTVATASSAIVVTQH GRVTSVAARHHWATAIAVLGRPKAIK	
A0A142BKH1_ALV	A0A142BKH1	Avian leukosis and sarcoma virus	TDNGSCFTSKSTREWLARWGIAHTTGI PGNSQGQAMVERANRLLKDKIRVLAE GDGFMKRIPTGKQGELLAKAMYALNH FERGENTKTPIQKHWRPTVLTEGPPVKI RIETGEWEKGWNVLVWGRGYAAVKN RDTDKIIWVPSRKVKPDITQKDELTKK DEASPLFAGISDWAPWKGEQEGL	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661

Table 32: InterPro descriptions of signatures present in reverse transcriptases in Table 30 (monomeric viral RTs) and Table 31 (dimeric viral RTs).

Signature	Database	Short Name	Description

cd01645	CDD	*RT_Rtv*	RT_Rtv: Reverse transcriptases (RTs) from retroviruses (Rtvs). RTs catalyze the conversion of single-stranded RNA into double-stranded viral DNA for integration into host chromosomes. Proteins in this subfamily contain long terminal repeats (LTRs) and are multifunctional enzymes with RNA-directed DNA polymerase, DNA directed DNA polymerase, and ribonuclease hybrid (RNase H) activities. The viral RNA genome enters the cytoplasm as part of a nucleoprotein complex, and the process of reverse transcription generates in the cytoplasm forming a linear DNA duplex via an intricate series of steps. This duplex DNA is colinear with its RNA template, but contains terminal duplications known as LTRs that are not present in viral RNA. It has been proposed that two specialized template switches, known as strand-transfer reactions or "jumps", are required to generate the LTRs. [PMID: 9831551, PMID: 15107837, PMID: 11080630, PMID: 10799511, PMID: 7523679, PMID: 7540934, PMID: 8648598, PMID: 1698615]

cd03715	CDD	*RT_ZFREV_like*	RT_ZFREV_like: A subfamily of reverse transcriptases (RTs) found in sequences similar to the intact endogenous retrovirus ZFERV from zebrafish and to Moloney murine leukemia virus RT. An RT gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. RTs occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. These elements can be divided into two major groups. One group contains retroviruses and DNA viruses whose propagation involves an RNA intermediate. They are grouped together with transposable elements containing long terminal repeats (LTRs). The other group, also called poly(A)-type retrotransposons, contain fungal mitochondrial introns and transposable elements that lack LTRs. Phylogenetic analysis suggests that ZFERV belongs to a distinct group of retroviruses. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]
PF00078	Pfam	*RVT_1*	A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. [PMID: 1698615]

IPR000477	InterPro	*RT_dom*	The use of an RNA template to produce DNA, for integration into the host genome and exploitation of a host cell, is a strategy employed in the replication of retroid elements, such as the retroviruses and bacterial retrons. The enzyme catalysing polymerisation is an RNA-directed DNA-polymerase, or reverse trancriptase (RT) (2.7.7.49). Reverse transcriptase occurs in a variety of mobile elements, including retrotransposons, retroviruses, group II introns [PMID: 12758069], bacterial msDNAs, hepadnaviruses, and caulimoviruses. Retroviral reverse transcriptase is synthesised as part of the POL polyprotein that contains; an aspartyl protease, a reverse transcriptase, RNase H and integrase. POL polyprotein undergoes specific enzymatic cleavage to yield the mature proteins. The discovery of retroelements in the prokaryotes raises intriguing questions concerning their roles in bacteria and the origin and evolution of reverse transcriptases and whether the bacterial reverse transcriptases are older than eukaryotic reverse transcriptases [PMID: 8828137]. Several crystal structures of the reverse transcriptase (RT) domain have been determined [PMID: 1377403].
IPR043502	InterPro	*DNA/RNA polymerase superfamily*	This entry represents the DNA/RNA polymerase superfamily, which includes DNA polymerase I, reverse transcriptase, T7 RNA polymerase, lesion bypass DNA polymerase (Y-family), RNA-dependent RNA-polymerase and dsRNA phage RNA-dependent RNA-polymerase. These enzymes share a similar protein fold at their active site, which resembles the palm subdomain of the right-hand-shaped polymerases. [PMID: 26931141]
SSF56672	Superfamily	*DNA/RNA polymerases*	This superfamily comprises DNA polymerases and RNA polymerases
PF06817	Pfam	*RVT_thumb*	This domain is known as the thumb domain. It is composed of a four helix bundle [PMID:1377403].

IPR010661	InterPro	*RVT_thumb*	This domain is known as the thumb domain. It is composed of a four helix bundle. Reverse transcriptase converts the viral RNA genome into double-stranded viral DNA. Reverse transcriptase often occurs in a polyprotein; with integrase, ribonuclease H and/or protease, which is cleaved before the enzyme takes action. The impact of antiretroviral treatment on the first 400 amino acids of HIV reverse transcriptase is good. Little is known, however, of the antiretroviral drug impact on the C-terminal domains of Pol, which includes the thumb, connection and RNase H. Evidence suggests that these might be well conserved domains. [PMID:1377403, PMID:18335052]
PF06815	Pfam	*RVT_connect*	This domain is known as the connection domain. This domain lies between the thumb and palm domains [PMID:1377403].
IPR010659	InterPro	*RVT_connect*	This domain is known as the connection domain. This domain lies between the thumb and palm domains [PMID:1377403].
cd03715	CDD	*RT_ZFREV_like*	RT_ZFREV_like: A subfamily of reverse transcriptases (RTs) found in sequences similar to the intact endogenous retrovirus ZFERV from zebrafish and to Moloney murine leukemia virus RT. An RT gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. RTs occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. These elements can be divided into two major groups. One group contains retroviruses and DNA viruses whose propagation involves an RNA intermediate. They are grouped together with transposable elements containing long terminal repeats (LTRs). The other group, also called poly(A)-type retrotransposons, contain fungal mitochondrial introns and transposable elements that lack LTRs. Phylogenetic analysis suggests that ZFERV belongs to a distinct group of retroviruses. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]

Endonuclease domain:

In some embodiments, the polypeptide comprises an endonuclease domain (e.g., a heterologous endonuclease domain). In certain embodiments, the endonuclease/DNA binding domain of an APE-type retrotransposon, the endonuclease domain of an RLE-type retrotransposon, or the endonuclease domain of a PLE-type retrotransposon can be used or can be modified (e.g., by insertion, deletion, or substitution of one or more residues) in a Gene Writer system described herein. In some embodiments the endonuclease domain or endonuclease/DNA binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some embodiments the endonuclease element is a heterologous endonuclease element, such as Fok1 nuclease, Cas9, Cas9 nickase, a type-II restriction enzyme like endonuclease (RLE-type nuclease), or another RLE-type endonuclease (also known as REL). In some embodiments, the heterologous endonuclease domain cleaves both DNA strands and forms double-stranded breaks. In some embodiments, the heterologous endonuclease activity has nickase activity and does not form double stranded breaks. The amino acid sequence of an endonuclease domain of a Gene Writer system described herein may be at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identical to the amino acid sequence of an endonuclease domain of a retrotransposon whose DNA sequence is referenced in Table X, Z1, Z2, 3A, or 3B. Endounclease domains can be identified, for example, based upon homology to other known endonuclease domains using tools as Basic Local Alignment Search Tool (BLAST). In certain embodiments, the heterologous endonuclease is Cas9 or Cas9 nickase or a functional fragment thereof. In certain embodiments, the heterologous endonuclease is Fok1 or a functional fragment thereof. In certain embodiments, the heterologous endonuclease is a Holliday junction resolvase or homolog thereof, such as the Holliday junction resolving enzyme from Sulfolobus solfataricus—Ssol Hje (Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). In certain embodiments, the heterologous endonuclease is the endonuclease of the large fragment of a spliceosomal protein, such as Prp8 (Mahbub et al., *Mobile DNA* 8:16, 2017). For example, a Gene Writer polypeptide described herein may comprise a reverse transcriptase domain from an APE-, RLE-, or PLE-type retrotransposon and an endonuclease domain that comprises Fok1 or a

functional fragment thereof. In still other embodiments, homologous endonuclease domains are modified, for example by site-specific mutation, to alter DNA endonuclease activity. In still other embodiments, endonuclease domains are modified to remove any latent DNA-sequence specificity.

In some embodiments, a Gene Writer polypeptide possesses the function of DNA target site cleavage via an endonuclease domain. In some embodiments, the endonuclease domain is also a DNA-binding domain. In some embodiments, the endonuclease domain is also a template nucleic acid (e.g., template RNA) binding domain. For example, in some embodiments a polypeptide comprises a CRISPR-associated endonuclease domain that binds a template RNA comprising a gRNA, binds a target DNA sequence (e.g., with complementarity to a portion of the gRNA), and cuts the target DNA sequence. In certain embodiments, the endonuclease/DNA binding domain of an APE-type retrotransposon or the endonuclease domain of an RLE-type retrotransposon can be used or can be modified (e.g., by insertion, deletion, or substitution of one or more residues) in a Gene Writer system described herein. In some embodiments the endonuclease domain or endonuclease/DNA binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some embodiments the endonuclease element is a heterologous endonuclease element, such as Fok1 nuclease, a type-II restriction l-like endonuclease (RLE-type nuclease), or another RLE-type endonuclease (also known as REL). In some embodiments the heterologous endonuclease activity has nickase activity and does not form double stranded breaks. The amino acid sequence of an endonuclease domain of a Gene Writer system described herein may be at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identical to the amino acid sequence of an endonuclease domain of a retrotransposon whose DNA sequence is referenced in Table 1, 2, 3A, or 3B. Endonuclease domains can be identified, for example, based upon homology to other known endonuclease domains using tools such as Basic Local Alignment Search Tool (BLAST). In certain embodiments, the heterologous endonuclease is Fok1 or a functional fragment thereof. In certain embodiments, the heterologous endonuclease is a Holliday junction resolvase or homolog thereof, such as the Holliday junction resolving enzyme from Sulfolobus solfataricus—Ssol Hje (Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). In certain embodiments, the heterologous endonuclease is the endonuclease of the

large fragment of a spliceosomal protein, such as Prp8 (Mahbub et al., *Mobile DNA* 8:16, 2017). In certain embodiments, the heterologous endonuclease is derived from a CRISPR-associated protein, e.g., Cas9. In certain embodiments, the heterologous endonuclease is engineered to have only ssDNA cleavage activity, e.g., only nickase activity, e.g., be a Cas9 nickase. For example, a Gene Writer polypeptide described herein may comprise a reverse transcriptase domain from an APE- or RLE-type retrotransposon and an endonuclease domain that comprises Fok1 or a functional fragment thereof. In still other embodiments, homologous endonuclease domains are modified, for example by site-specific mutation, to alter DNA endonuclease activity. In still other embodiments, endonuclease domains are modified to remove any latent DNA-sequence specificity.

In some embodiments the endonuclease domain has nickase activity and does not form double stranded breaks. In some embodiments, the endonuclease domain forms single stranded breaks at a higher frequency than double stranded breaks, e.g., at least 90%, 95%, 96%, 97%, 98%, or 99% of the breaks are single stranded breaks, or less than 10%, 5%, 4%, 3%, 2%, or 1% of the breaks are double stranded breaks. In some embodiments, the endonuclease forms substantially no double stranded breaks. In some embodiments, the enonuclease does not form detectable levels of double stranded breaks.

In some embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the to-be-edited strand; e.g., in some embodiments, the endonuclease domain cuts the genomic DNA of the target site near to the site of alteration on the strand that will be extended by the writing domain. In some embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the to-be-edited strand and does not nick the target site DNA of the non-edited strand. For example, when a polypeptide comprises a CRISPR-associated endonuclease domain having nickase activity and that does not form double stranded breaks, in some embodiments said CRISPR-associated endonuclease domain nicks the target site DNA strand containing the PAM site (e.g., and does not nick the target site DNA strand that does not contain the PAM site).

In some other embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the to-be-edited strand and the non-edited strand. Without wishing to be bound by theory, after a writing domain (e.g., RT domain) of a polypeptide described herein polymerizes (e.g., reverse transcribes) from the heterologous object sequence of a template

nucleic acid (e.g., template RNA), the cellular DNA repair machinery must repair the nick on the to-be-edited DNA strand. The target site DNA now contains two different sequences for the to-be-edited DNA strand: one corresponding to the original genomic DNA and a second corresponding to that polymerized from the heterologous object sequence. It is thought that the two different sequences equilibrate with one another, first one hybridizing the non-edited strand, then the other, and which the cellular DNA repair apparatus incorporates into its repaired target site is thought to be random. Without wishing to be bound by theory, introducing an additional nick to the non-edited strand may bias the cellular DNA repair machinery to adopt the heterologous object sequence-based sequence more frequently than the original genomic sequence. In some embodiments, the additional nick is positioned at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, or 150 nucleotides 5' or 3' of the target site modification (e.g., the insertion, deletion, or substitution) or to the nick on the to-be-edited strand.

Alternatively or additionally, without wishing to be bound by theory, an additional nick to the non-edited strand may promote second strand synthesis. In some embodiments, where the Gene Writer has inserted or substituted a portion of the edited strand, synthesis of a new sequence corresponding to the insertion/substitution in the non-edited strand is necessary.

In some embodiments, the polypeptide comprises a single domain having endonuclease activity (e.g., a single endonuclease domain) and said domain nicks both the to-be-edited strand and the non-edited strand. For example, in such an embodiment the endonuclease domain may be a CRISPR-associated endonuclease domain, and the template nucleic acid (e.g., template RNA) comprises a gRNA that directs nicking of the to-be-edited strand and an additional gRNA that directs nicking of the non-edited strand. In some embodiments, the polypeptide comprises a plurality of domains having endonuclease activity, and a first endonuclease domain nicks the to-be-edited strand and a second endonuclease domain nicks the non-edited strand (optionally, the first endonuclease domain does not (e.g., cannot) nick the non-edited strand and the second endonuclease domain does not (e.g., cannot) nick the to-be-edited strand).

In some embodiments, the endonuclease domain is capable of nicking a first strand and a second strand. In some embodiments, the first and second strand nicks occur at the same position in the target site but on opposite strands. In some embodiments, the second strand nick occurs in a staggered location, e.g., upstream or downstream, from the first nick. In some embodiments,

the endonuclease domain generates a target site deletion if the second strand nick is upstream of the first strand nick. In some embodiments, the endonuclease domain generates a target site duplication if the second strand nick is downstream of the first strand nick. In some embodiments, the endonuclease domain generates no duplication and/or deletion if the first and second strand nicks occur in the same position of the target site (e.g., as described in Gladyshev and Arkhipova Gene 2009, incorporated by reference herein in its entirety). In some embodiments, the endonuclease domain has altered activity depending on protein conformation or RNA-binding status, e.g., which promotes the nicking of the first or second strand (e.g., as described in Christensen et al. PNAS 2006; incorporated by reference herein in its entirety).

In some embodiments, the endonuclease domain comprises a meganuclease, or a functional fragment thereof. In some embodiments, the endonuclease domain comprises a homing endonuclease, or a functional fragment thereof. In some embodiments, the endonuclease domain comprises a meganuclease from the LAGLIDADG, GIY-YIG, HNH, His-Cys Box, or PD-(D/E) XK families, or a functional fragment or variant thereof, e.g., which possess conserved amino acid motifs, e.g., as indicated in the family names. In some embodiments, the endonuclease domain comprises a meganuclease, or fragment thereof, chosen from, e.g., I-SmaMI (Uniprot F7WD42), I-SceI (Uniprot P03882), I-AniI (Uniprot P03880), I-DmoI (Uniprot P21505), I-CreI (Uniprot P05725), I-TevI (Uniprot P13299), I-OnuI (Uniprot Q4VWW5), or I-BmoI (Uniprot Q9ANR6). In some embodiments, the meganuclease is naturally monomeric, e.g., I-SceI, I-TevI, or dimeric, e.g., I-CreI, in its functional form. For example, the LAGLIDADG meganucleases with a single copy of the LAGLIDADG motif generally form homodimers, whereas members with two copies of the LAGLIDADG motif are generally found as monomers. In some embodiments, a meganuclease that normally forms as a dimer is expressed as a fusion, e.g., the two subunits are expressed as a single ORF and, optionally, connected by a linker, e.g., an I-CreI dimer fusion (Rodriguez-Fornes et al. Gene Therapy 2020; incorporated by reference herein in its entirety). In some embodiments, a meganuclease, or a functional fragment thereof, is altered to favor nickase activity for one strand of a double-stranded DNA molecule, e.g., I-SceI (K122I and/or K223I) (Niu et al. J Mol Biol 2008), I-AniI (K227M) (McConnell Smith et al. PNAS 2009), I-DmoI (Q42A and/or K120M) (Molina et al. J Biol Chem 2015). In some embodiments, a meganuclease or functional fragment thereof possessing this preference for single-strand cleavage is used as an endonuclease domain, e.g.,

with nickase activity. In some embodiments, an endonuclease domain comprises a meganuclease, or a functional fragment thereof, which naturally targets or is engineered to target a safe harbor site, e.g., an I-CreI targeting SH6 site (Rodriguez-Fornes et al., *supra*). In some embodiments, an endonuclease domain comprises a meganuclease, or a functional fragment thereof, with a sequence tolerant catalytic domain, e.g., I-TevI recognizing the minimal motif CNNNG (Kleinstiver et al. PNAS 2012). In some embodiments, a target sequence tolerant catalytic domain is fused to a DNA binding domain, e.g., to direct activity, e.g., by fusing I-TevI to: (i) zinc fingers to create Tev-ZFEs (Kleinstiver et al. PNAS 2012), (ii) other meganucleases to create MegaTevs (Wolfs et al. Nucleic Acids Res 2014), and/or (iii) Cas9 to create TevCas9 (Wolfs et al. PNAS 2016).

In some embodiments, the endonuclease domain comprises a restriction enzyme, e.g., a Type IIS or Type IIP restriction enzyme. In some embodiments, the endonuclease domain comprises a Type IIS restriction enzyme, e.g., FokI, or a fragment or variant thereof. In some embodiments, the endonuclease domain comprises a Type IIP restriction enzyme, e.g., PvuII, or a fragment or variant thereof. In some embodiments, a dimeric restriction enzyme is expressed as a fusion such that it functions as a single chain, e.g., a FokI dimer fusion (Minczuk et al. Nucleic Acids Res 36(12):3926-3938 (2008)).

The use of additional endonuclease domains is described, for example, in Guha and Edgell Int J Mol Sci 18(22):2565 (2017), which is incorporated herein by reference in its entirety.

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises a *Streptococcus pyogenes* Cas9 (SpCas9) or a functional fragment or variant thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises a modified SpCas9. In embodiments, the modified SpCas9 comprises a modification that alters protospacer-adjacent motif (PAM) specificity. In embodiments, the PAM has specificity for the nucleic acid sequence 5'-NGT-3'. In embodiments, the modified SpCas9 comprises one or more amino acid substitutions, e.g., at one or more of positions L1111, D1135, G1218, E1219, A1322, of R1335, e.g., selected from L1111R, D1135V, G1218R, E1219F, A1322R, R1335V. In embodiments, the modified SpCas9 comprises the amino acid substitution T1337R and one or more additional amino acid substitutions, e.g., selected from L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K,

D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereto. In embodiments, the modified SpCas9 comprises: (i) one or more amino acid substitutions selected from D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337; and (ii) one or more amino acid substitutions selected from L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereto.

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises a Cas domain, e.g., a Cas9 domain. In embodiments, the endonuclease domain or DNA binding domain comprises a nuclease-active Cas domain, a Cas nickase (nCas) domain, or a nuclease-inactive Cas (dCas) domain. In embodiments, the endonuclease domain or DNA binding domain comprises a nuclease-active Cas9 domain, a Cas9 nickase (nCas9) domain, or a nuclease-inactive Cas9 (dCas9) domain. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 domain of Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises an S. pyogenes or an S. thermophilus Cas9, or a functional fragment thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 sequence, e.g., as described in Chylinski, Rhun, and Charpentier (2013) RNA Biology 10:5, 726-737; incorporated herein by reference. In some embodiments, the endonuclease domain or DNA binding domain comprises the HNH nuclease subdomain and/or the RuvC1 subdomain of a Cas, e.g., Cas9, e.g., as described herein, or a variant thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas polypeptide (e.g., enzyme), or a functional fragment thereof. In embodiments, the Cas polypeptide (e.g., enzyme) is selected from Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (e.g., Csn1 or

Csx12), Cas10, Cas10d, Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Type II Cas effector proteins, Type V Cas effector proteins, Type VI Cas effector proteins, CARF, DinG, Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12b/C2c1, Cas12c/C2c3, SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, hyper accurate Cas9 variant (HypaCas9), homologues thereof, modified or engineered versions thereof, and/or functional fragments thereof. In embodiments, the Cas9 comprises one or more substiutions, e.g., selected from H840A, D10A, P475A, W476A, N477A, D1125A, W1126A, and D1127A. In embodiments, the Cas9 comprises one or more mutations at positions selected from: D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987, e.g., one or more substitutions selected from D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas (e.g., Cas9) sequence from Corynebacterium ulcerans, Corynebacterium diphtheria, Spiroplasma syrphidicola, Prevotella intermedia, Spiroplasma taiwanense, Streptococcus iniae, Belliella baltica, Psychroflexus torquis, Streptococcus thermophilus, Listeria innocua, Campylobacter jejuni, Neisseria meningitidis, Streptococcus pyogenes, or Staphylococcus aureus, or a fragment or variant thereof.

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises a Cpf1 domain, e.g., comprising one or more substitutions, e.g., at position D917, E1006A, D1255 or any combination thereof, e.g., selected from D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, and D917A/E1006A/D1255A.

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises spCas9, spCas9-VRQR, spCas9- VRER, xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER, spCas9-LRKIQK, or spCas9- LRVSQL.

In some embodiments, an endonuclease domain or DNA binding domain (e.g., as described herein) comprises an amino acid sequence as listed in Table 37 below, or an amino acid sequence having at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the endonuclease domain or DNA-binding domain comprises an

amino acid sequence that has no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 differences (e.g., mutations) relative to any of the amino acid sequences described herein.

5

Table 37. Each of the Reference Sequences are incorporated by reference in their entirety.

Name	**Amino Acid Sequence or Reference Sequence**
Streptococcus pyogenes Cas9	
Exemplary Linker	SGSETPGTSESATPES
Exemplary Linker Motif	$(SGGS)_n$
Exemplary Linker Motif	$(GGGS)_n$
Exemplary Linker Motif	$(GGGGS)_n$
Exemplary Linker Motif	$(G)_n$
Exemplary Linker Motif	$(EAAAK)_n$
Exemplary Linker Motif	$(GGS)_n$
Exemplary Linker Motif	$(XP)_n$
Cas9 from Streptococcus pyogenes	NCBI Reference Sequence: NC_002737.2 and Uniprot Reference Sequence: Q99ZW2
Cas9 from Corynebacterium ulcerans	NCBI Refs: NC_015683.1, NC_017317.1
Cas9 from Corynebacterium diphtheria	NCBI Refs: NC_016782.1, NC_016786.1

Cas9 from Spiroplasma syrphidicola	NCBI Ref: NC_021284.1
Cas9 from Prevotella intermedia	NCBI Ref: NC_017861.1
Cas9 from Spiroplasma taiwanense	NCBI Ref: NC_021846.1
Cas9 from Streptococcus iniae	NCBI Ref: NC_021314.1
Cas9 from Belliella baltica	NCBI Ref: NC_018010.1
Cas9 from Psychroflexus torquisI	NCBI Ref: NC_018721.1
Cas9 from Streptococcus thermophilus	NCBI Ref: YP_820832.1
Cas9 from Listeria innocua	NCBI Ref: NP_472073.1
Cas9 from Campylobacter jejuni	NCBI Ref: YP_002344900.1
Cas9 from Neisseria meningitidis	NCBI Ref: YP_002342100.1
dCas9 (D10A and H840A)	
Catalytically inactive Cas9 (dCas9)	
Cas9 nickase (nCas9)	
Catalytically active Cas9	

CasY	((ncbi.nlm.nih.gov/protein/APG80656.1) >APG80656.1 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium])
CasX	uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53
CasX	>tr|F0NH53|F0NH53_SULIR CRISPR associated protein, Casx OS = Sulfolobus islandicus (strain REY15A) GN=SiRe_0771 PE=4 SV=1
Deltaproteobacteria CasX	
Cas12b/C2c1	((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG CRISPR- associated endonuclease C2c1 OS = Alicyclobacillus acido-terrestris (strain ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1)
BhCas12b (Bacillus hisashii)	NCBI Reference Sequence: WP_095142515
BvCas12b (Bacillus sp. V3-13)	NCBI Reference Sequence: WP_101661451.1
Wild-type Francisella novicida Cpf1	
Francisella novicida Cpf1 D917A	
Francisella novicida Cpf1 E1006A	

Francisella novicida Cpf1 D1255A	
Francisella novicida Cpf1 D917A/E1006A	
Francisella novicida Cpf1 D917A/D1255A	
Francisella novicida Cpf1 E1006A/D1255A	
Francisella novicida Cpf1 D917A/E1006A	
SaCas9	
SaCas9n	
PAM-binding SpCas9	
PAM-binding SpCas9n	
PAM-binding SpEQR Cas9	
PAM-binding SpVQR Cas9	
PAM-binding SpVRER Cas9	
PAM-binding SpVRQR Cas9	
SpyMacCas9	

In some embodiments, a Gene Writing polypeptide has an endonuclease domain comprising a Cas9 nickase, e.g., Cas9 H840A. In embodiments, the Cas9 H840A has the following amino acid sequence:

Cas9 nickase (H840A):

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS
GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE
RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV
DAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED
NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

In some embodiments, a Gene Writing polypeptide comprises the RT domain from a retroviral reverse transcriptase, e.g., a wild-type M-MLV RT, e.g., comprising the following sequence:

M-MLV (WT):
TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
GQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAP
ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRY
AFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
GNRMADQAARKAAITETPDTSTLLI

In some embodiments, a Gene Writing polypeptide comprises the RT domain from a retroviral reverse transcriptase, e.g., an M-MLV RT, e.g., comprising the following sequence:
TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
GQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAP
ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRY

AFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
GNRMADQAARKAAITETPDTSTLL

In some embodiments, a Gene Writing polypeptide comprises the RT domain from a retroviral reverse transcriptase comprising the sequence of amino acids 659-1329 of NP_057933. In embodiments, the Gene Writing polypeptide further comprises one additional amino acid at the N-terminus of the sequence of amino acids 659-1329 of NP_057933, e.g., as shown below:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP**VKKPGTNDYRPVQDLREVN**
KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP
EMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAAT
SELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKE
TVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAY
QEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPV
AAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTH
YQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDAD
HTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK
KLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPG
HQKGHSAEARGNRMADQAARKAA

Core RT (bold), annotated per above
RNAseH (underlined), annotated per above

In embodiments, the Gene Writing polypeptide further comprises one additional amino acid at the C-terminus of the sequence of amino acids 659-1329 of NP_057933. In embodiments, the Gene Writing polypeptide comprises an RNaseH1 domain (e.g., amino acids 1178-1318 of NP_057933).

In some embodiments, a retroviral reverse transcriptase domain, e.g., M-MLV RT, may comprise one or more mutations from a wild-type sequence that may improve features of the RT, e.g., thermostability, processivity, and/or template binding. In some embodiments, an M-MLV RT domain comprises, relative to the M-MLV (WT) sequence above, one or more mutations, e.g., selected from D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, K103L, e.g., a combination of mutations, such as D200N, L603W, and T330P, optionally further including T306K and W313F. In some embodiments, an M-MLV RT used herein comprises the

mutations D200N, L603W, T330P, T306K and W313F. In embodiments, the mutant M-MLV RT comprises the following amino acid sequence:

M-MLV (PE2):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
GQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAP
ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRY
AFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
GNRMADQAARKAAITETPDTSTLLI

In some embodiments, a Gene Writer polypeptide may comprise a linker, e.g., a peptide linker, e.g., a linker as described in Table 7. In some embodiments, a Gene Writer polypeptide comprises a flexible linker between the endonuclease and the RT domain, e.g., a linker comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS. In some embodiments, an RT domain of a Gene Writer polypeptide may be located C-terminal to the endonuclease domain. In some embodiments, an RT domain of a Gene Writer polypeptide may be located N-terminal to the endonuclease domain.

Table 7 Exemplary linker sequences

Amino Acid Sequence
GGS
GGSGGS
GGSGGSGGS
GGSGGSGGSGGS

GGSGGSGGSGGSGGS
GGSGGSGGSGGSGGSGGS
GGGGS
GGGGSGGGGS
GGGGSGGGGSGGGGS
GGGGSGGGGSGGGGSGGGGS
GGGGSGGGGSGGGGSGGGGSGGGGS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS
GGG
GGGG
GGGGG
GGGGGG
GGGGGGG
GGGGGGGG
GSS
GSSGSS
GSSGSSGSS
GSSGSSGSSGSS
GSSGSSGSSGSSGSS
GSSGSSGSSGSSGSSGSS
EAAAK
EAAAKEAAAK
EAAAKEAAAKEAAAK
EAAAKEAAAKEAAAKEAAAK
EAAAKEAAAKEAAAKEAAAKEAAAK
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK
PAP
PAPAP
PAPAPAP
PAPAPAPAP
PAPAPAPAPAP
PAPAPAPAPAPAP
GGSGGG
GGGGGS
GGSGSS
GSSGGS
GGSEAAAK
EAAAKGGS
GGSPAP

PAPGGS
GGGGSS
GSSGGG
GGGEAAAK
EAAAKGGG
GGGPAP
PAPGGG
GSSEAAAK
EAAAKGSS
GSSPAP
PAPGSS
EAAAKPAP
PAPEAAAK
GGSGGGGSS
GGSGSSGGG
GGGGGSGSS
GGGGSSGGS
GSSGGSGGG
GSSGGGGGS
GGSGGGEAAAK
GGSEAAAKGGG
GGGGGSEAAAK
GGGEAAAKGGS
EAAAKGGSGGG
EAAAKGGGGGS
GGSGGGPAP
GGSPAPGGG
GGGGGSPAP
GGGPAPGGS
PAPGGSGGG
PAPGGGGGS
GGSGSSEAAAK
GGSEAAAKGSS
GSSGGSEAAAK
GSSEAAAKGGS
EAAAKGGSGSS
EAAAKGSSGGS
GGSGSSPAP
GGSPAPGSS

GSSGGSPAP
GSSPAPGGS
PAPGGSGSS
PAPGSSGGS
GGSEAAAKPAP
GGSPAPEAAAK
EAAAKGGSPAP
EAAAKPAPGGS
PAPGGSEAAAK
PAPEAAAKGGS
GGGGSSEAAAK
GGGEAAAKGSS
GSSGGGEAAAK
GSSEAAAKGGG
EAAAKGGGGSS
EAAAKGSSGGG
GGGGSSPAP
GGGPAPGSS
GSSGGGPAP
GSSPAPGGG
PAPGGGGSS
PAPGSSGGG
GGGEAAAKPAP
GGGPAPEAAAK
EAAAKGGGPAP
EAAAKPAPGGG
PAPGGGEAAAK
PAPEAAAKGGG
GSSEAAAKPAP
GSSPAPEAAAK
EAAAKGSSPAP
EAAAKPAPGSS
PAPGSSEAAAK
PAPEAAAKGSS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA
GGGGSEAAAKGGGGS
EAAAKGGGGSEAAAK
SGSETPGTSESATPES
GSAGSAAGSGEF

SGGSSGGSSGSETPGTSESATPESSGGSSGGSS

In some embodiments, a Gene Writer polypeptide comprises a dCas9 sequence comprising a D10A and/or H840A mutation, e.g., the following sequence:

SMDKKYSIGLAIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

In some embodiments, a template RNA molecule for use in the system comprises, from 5' to 3' (1) a gRNA spacer; (2) a gRNA scaffold; (3) heterologous object sequence (4) 3' homology domain. In some embodiments:

(1) Is a Cas9 spacer of ~18-22 nt, e.g., is 20 nt

(2) Is a gRNA scaffold comprising one or more hairpin loops, e.g., 1, 2, of 3 loopd for associating the template with a nickase Cas9 domain. In some embodiments, the gRNA scaffold carries the sequence, from 5' to 3',
GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT
GAAAAAGTGGGACCGAGTCGGTCC.

(3) In some embodiments, the heterologous object sequence is, e.g., 7-74, e.g., 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, or 70-80 nt or, 80-90 nt in length. In some embodiments, the first (most 5') base of the sequence is not C.

(4) In some embodiments, the 3' homology domain that binds the target priming sequence after nicking occurs is e.g., 3-20 nt, e.g., 7-15 nt, e.g., 12-14 nt. In some embodiments, the 3' homology domain has 40-60% GC content.

A second gRNA associated with the system may help drive complete integration. In some embodiments, the second gRNA may target a location that is 0-200 nt away from the first-strand nick, e.g., 0-50, 50-100, 100-200 nt away from the first-strand nick. In some embodiments, the second gRNA can only bind its target sequence after the edit is made, e.g., the gRNA binds a sequence present in the heterologous object sequence, but not in the initial target sequence.

In some embodiments, a Gene Writing system described herein is used to make an edit in HEK293, K562, U2OS, or HeLa cells. In some embodiment, a Gene Writing system is used to make an edit in primary cells, e.g., primary cortical neurons from E18.5 mice.

In some embodiments, a reverse transcriptase or RT domain (e.g., as described herein) comprises a MoMLV RT sequence or variant thereof. In embodiments, the MoMLV RT sequence comprises one or more mutations selected from D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, and K103L. In embodiments, the MoMLV RT sequence comprises a combination of mutations, such as D200N, L603W, and T330P, optionally further including T306K and/or W313F.

In some embodiments, an endonuclease domain (e.g., as described herein) comprises nCAS9, e.g., comprising the H840A mutation.

In some embodiments, the heterologous object sequence (e.g., of a system as described herein) is about 1-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, or more, nucleotides in length.

In some embodiments, the RT and endonuclease domains are joined by a flexible linker, e.g., comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS.

In some embodiments, the endonuclease domain is N-terminal relative to the RT domain. In some embodiments, the endonuclease domain is C-terminal relative to the RT domain.

In some embodiments, the system incorporates a heterologous object sequence into a target site by TPRT, e.g., as described herein.

In some embodiments, a system or method described herein involves a CRISPR DNA targeting enzyme or system described in US Pat. App. Pub. No. 20200063126, 20190002889, or 20190002875 (each of which is incorporated by reference herein in its entirety) or a functional fragment or variant thereof. For instance, in some embodiments, a GeneWriter polypeptide or Cas endonuclease described herein comprises a polypeptide sequence of any of the applications mentioned in this paragraph, and in some embodiments a template RNA or guide RNA comprises a nucleic acid sequence of any of the applications mentioned in this paragraph.

Template nucleic acid binding domain:

A Gene Writer polypeptide typically contains regions capable of associating with the Gene Writer template nucleic acid (e.g., template RNA). In some embodiments, the template nucleic acid binding domain is an RNA binding domain. In some embodiments, the RNA binding domain is a modular domain that can associate with RNA molecules containing specific signatures, e.g., structural motifs, e.g., secondary structures present in the 3' UTR in non-LTR retrotransposons. In other embodiments, the template nucleic acid binding domain (e.g., RNA binding domain) RNA binding domain is contained within the reverse transcription domain, e.g., the reverse transcriptase-derived component has a known signature for RNA preference, e.g., secondary structures present in the 3' UTR in non-LTR retrotransposons. In other embodiments, the template nucleic acid binding domain (e.g., RNA binding domain) is contained within the DNA binding domain. For example, in some embodiments, the DNA binding domain is a CRISPR-associated protein that recognizes the structure of a template nucleic acid (e.g., template RNA) comprising a gRNA. In some embodiments, the gRNA is a short synthetic RNA composed of a scaffold sequence that participates in CRISPR-associated protein binding and a user-defined ~20 nucleotide targeting sequence for a genomic target. The structure of a complete gRNA was described by Nishimasu et al. Cell 156, P935-949 (2014). The gRNA (also referred to as sgRNA for single-guide RNA) consists of crRNA- and tracrRNA-derived sequences connected by an artificial tetraloop. The crRNA sequence can be divided into guide (20 nt) and repeat (12 nt) regions, whereas the tracrRNA sequence can be divided into anti-repeat (14 nt) and

three tracrRNA stem loops (Nishimasu et al. Cell 156, P935-949 (2014)). In practice, guide RNA sequences are generally designed to have a length of between 17 – 24 nucleotides (e.g., 19, 20, or 21 nucleotides) and be complementary to a targeted nucleic acid sequence. Custom gRNA generators and algorithms are available commercially for use in the design of effective guide RNAs. In some embodiments, the gRNA comprises two RNA components from the native CRISPR system, e.g. crRNA and tracrRNA. As is well known in the art, the gRNA may also comprise a chimeric, single guide RNA (sgRNA) containing sequence from both a tracrRNA (for binding the nuclease) and at least one crRNA (to guide the nuclease to the sequence targeted for editing/binding). Chemically modified sgRNAs have also been demonstrated to be effective for use with CRISPR-associated proteins; see, for example, Hendel et al. (2015) Nature Biotechnol., 985 – 991. In some embodiments, a gRNA comprises a nucleic acid sequence that is complementary to a DNA sequence associated with a target gene. In some embodiments, a polypeptide comprises a DNA-binding domain comprising a CRISPR-associated protein that associates with a gRNA that allows the DNA-binding domain to bind a target genomic DNA sequence. In some embodiments, the gRNA is comprised within the template nucleic acid (e.g., template RNA), thus the DNA-binding domain is also the template nucleic acid binding domain. In some embodiments, the polypeptide possesses RNA binding function in multiple domains, e.g., can bind a gRNA structure in a CRISPR-associated DNA binding domain and a 3' UTR structure in a non-LTR retrotransposon derived reverse transcription domain.

In some embodiments, a template nucleic acid (e.g., template RNA) comprises a 3' target homology domain. In some embodiments, a 3' target homology domain is disposed 3' of the heterologous object sequence and is complementary to a sequence adjacent to a site to be modified by a system described herein, or comprises no more than 1, 2, 3, 4, or 5 mismatches to a sequence complementary to the sequence adjacent to a site to be modified by the system/Gene Writer™. In some embodiments, the 3' target homology domain anneals to the target site, which provides a binding site and the 3' hydroxyl for the initiation of TPRT by a Gene Writer polypeptide. In some embodiments, the 3' target homology domain is 3-5, 5-10, 10-30, 10-25, 10-20, 10-19, 10-18, 10-17, 10-16, 10-15, 10-14, 10-13, 10-12, 10-11, 11-30, 11-25, 11-20, 11-19, 11-18, 11-17, 11-16, 11-15, 11-14, 11-13, 11-12, 12-30, 12-25, 12-20, 12-19, 12-18, 12-17, 12-16, 12-15, 12-14, 12-13, 13-30, 13-25, 13-20, 13-19, 13-18, 13-17, 13-16, 13-15, 13-14, 14-30, 14-25, 14-20, 14-19, 14-18, 14-17, 14-16, 14-15, 15-30, 15-25, 15-20, 15-19, 15-18, 15-17,

15-16, 16-30, 16-25, 16-20, 16-19, 16-18, 16-17, 17-30, 17-25, 17-20, 17-19, 17-18, 18-30, 18-25, 18-20, 18-19, 19-30, 19-25, 19-20, 20-30, 20-25, or 25-30 nt in length, e.g., 10-17, 12-16, or 12-14 nt in length.

In some embodiments, the template nucleic acid, e.g., template RNA, may comprise a gRNA (e.g., pegRNA). In some embodiments, the template nucleic acid, e.g., template RNA, may bind to the Gene Writer™ polypeptide by interaction of a gRNA portion of the template nucleic acid with a template nucleic acid binding domain, e.g., a RNA binding domain (e.g., a heterologous RNA binding domain). In some embodiments, the heterologous RNA binding domain is a CRISPR/Cas protein, e.g., Cas9.

In some embodiments, the region of the template nucleic acid, e.g., template RNA, comprising the gRNA adopts an underwound ribbon-like structure of gRNA bound to target DNA (e.g., as described in Mulepati et al. Science 19 Sep 2014:Vol. 345, Issue 6203, pp. 1479-1484). Without wishing to be bound by theory, this non-canonical structure is thought to be facilitated by rotation of every sixth nucleotide out of the RNA-DNA hybrid. Thus, in some embodiments, the region of the template nucleic acid, e.g., template RNA, comprising the gRNA may tolerate increased mismatching with the target site at some interval, e.g., every sixth base. In some embodiments, the region of the template nucleic acid, e.g., template RNA, comprising the gRNA comprising homology to the target site may possess wobble positions at a regular interval, e.g., every sixth base, that do not need to base pair with the target site.

gRNAs with inducible activity

In some embodiments, a template nucleic acid, e.g., template RNA, comprises a guide RNA (gRNA) with inducible activity. Inducible activity may be achieved by the template nucleic acid, e.g., template RNA, further comprising (in addition to the gRNA) a blocking domain, wherein the sequence of a portion of or all of the blocking domain is at least partially complementary to a portion or all of the gRNA. The blocking domain is thus capable of hybridizing or substantially hybridizing to a portion of or all of the gRNA. In some embodiments, the blocking domain and inducibly active gRNA are disposed on the template nucleic acid, e.g., template RNA, such that the gRNA can adopt a first conformation where the blocking domain is hybridized or substantially hybridized to the gRNA, and a second conformationwhere the blocking domain is not hybridized or or not substantially hybridized to

the gRNA. In some embodiments, in the first conformationthe gRNA is unable to bind to the Gene Writer polypeptide (e.g., the template nucleic acid binding domain, DNA binding domain, or endonuclease domain (e.g., a CRISPR/Cas protein)) or binds with substantially decreased affinity compared to an otherwise similar template RNA lacking the blocking domain. In some embodiments, in the second conformationthe gRNA is able to bind to the Gene Writer polypeptide (e.g., the template nucleic acid binding domain, DNA binding domain, or endonuclease domain (e.g., a CRISPR/Cas protein)). In some embodiments, whether the gRNA is in the first or second conformationcan influence whether the DNA binding or endonuclease activities of the Gene Writer polypeptide (e.g., of the CRISPR/Cas protein the Gene Writer polypeptide comprises) are active. In some embodiments, hybridization of the gRNA to the blocking domain can be disrupted using an opener molecule. In some embodiments, an opener molecule comprises an agent that binds to a portion or all of the gRNA or blocking domain and inhibits hybridization of the gRNA to the blocking domain. In some embodiments, the opener molecule comprises a nucleic acid, e.g., comprising a sequence that is partially or wholly complementary to the gRNA, blocking domain, or both. By choosing or designing an appropriate opener molecule, providing the opener molecule can promote a change in the conformation of the gRNA such that it can associate with a CRISPR/Cas protein and provide the associated functions of the CRISPR/Cas protein (e.g., DNA binding and/or endonuclease activity). Without wishing to be bound by theory, providing the opener molecule at a selected time and/or location may allow for spatial and temporal control of the activity of the gRNA, CRISPR/Cas protein, or Gene Writer system comprising the same. In some embodiments, a Gene Writer may comprise a Cas protein as listed in Table 9 or Table 37 or a functional fragment thereof, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto.

Table .9 CRISPR/Cas Proteins, Species, and Mutations

Variant	Parental Host	Protein Sequence	Nickase Mutation
Nme2Cas9	Neisseria meningitidis	MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLAR SVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVL LHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGHIRNQR GDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFE PAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGL EDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIGTAFSLFKT DEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKN TEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEE NRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDHAL PFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQ KFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAEN DRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQE VMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRS	N611A

		AKRFVKHNEKISVKRVWLTEIKLADLENMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKDN PFYKKGGQLVKAVRVEKTQESGVLLNKKNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYA WQVAENILPDIDCKGYRIDDSYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHD KGSKEQQFRISTQNLVLIQKYQVNELGKEIRPCRLKKRPPVR	
PpnCas9	Pasteurella pneumotropica	MQNNPLNYILGLDLGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVAKTGESLALSRRLARSSRR LIKRRAERLKKAKRLLKAEKILHSIDEKLPINVWQLRVKGLKEKLERQEWAAVLLHLSKHRGYLS QRKNEGKSDNKELGALLSGIASNHQMLQSSEYRTPAEIAVKKFQVEEGHIRNQRGSYTHTFSRLDL LAEMELLFQRQAELGNSYTSTTLLENLTALLMWQKPALAGDAILKMLGKCTFEPSEYKAAKNSY SAERFVWLTKLNNLRILENGTERALNDNERFALLEQPYEKSKLTYAQVRAMLALSDNAIFKGVRY LGEDKKTVESKTTLIEMKFYHQIRKTLGSAELKKEWNELKGNSDLLDEIGTAFSLYKTDDDICRYL EGKLPERVLNALLENLNFDKFIQLSLKALHQILPLMLQGGRYDEAVSAIYGDHYGKKSTETTRLLP TIPADEIRNPVVLRTLTQARKVINAVVRLYGSPARIHIETAREVGKSYQDRKKLEKQQEDNRKQRE SAVKKFKEMFPHFVGEPKGKDILKMRLYELQQAKCLYSGKSLELHRLLEKGYVEVDHALPFSRT WDDSFNNKVLVLANENQNKGNLTPYEWLDGKNNSERWQHFVVRVQTSGFSYAKKQRILNHKLD EKGFIERNLNDTRYVARFLCNFIADNMLLVGKGKRNVFASNGQITALLRHRWGLQKVREQNDRH HALDAVVVACSTVAMQQKITRFVRYNEGNVFSGERIDRETGEIIPLHFPSPWAFFKENVEIRIFSEN PKLELENRLPDYPQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSAKRLNEGLSVLKVPLTQLKL SDLERMVNRDREIALYESLKARLEQFGNDPAKAFAEPFYKKGGALVKAVRLEQTQKSGVLVRDG NGVADNASMVRVDVFTKGGKYFLVPIYTWQVAKGILPNRAATQGKDENDWDIMDEMATFQFSL CQNDLIKLVTKKKTIFGYFNGLNRATSNINIKEHDLDKSKGKLGIYLEVGVKLAISLEKYQVDELG	N605A
SauCas9	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGN TLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETG NYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFV TVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
SauCas9-KKH	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNR IEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
SauriCas9	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNRRSKRGARRLKRRRI HRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPLTKEEFAIALLHIAKRRGLHNISVSMGD EEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQ YHNIDDQFIQQYIDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKS GKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQDEISIKKALDQLPELLTESEKSQIAQL TGYTGTHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFI QSIKVINAVINRFGLPEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIE KIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQSENSKKGNRTP YQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEERDINKFEVQKEFINRNLVDTRYATR ELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTH KALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRQLIN DTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLMTILNQYAEAK NPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSLKSFRFDI YKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYEAEKQKKKIKESDLFVGSFYYNDLIMYEDELFR VIGVNSDINNLVELNMVDITYKDFCEVNNVTGEKRIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKP QLIFKRGEL	N588A

SauriCas9-KKH	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNRRSKRGARRLKRRRI HRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPLTKEEFAIALLHIAKRRGLHNISVSMGD EEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQ YHNIDDQFIQQYIDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKS GKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQDEISIKKALDQLPELLTESEKSQIAQL TGYTGTHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFI QSIKVINAVINRFGLPEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIE KIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQSENSKKGNRTP YQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEERDINKFEVQKEFINRNLVDTRYATR ELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTH KALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRKLIN DTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLMTILNQYAEAK NPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSLKSFRFDI YKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYEAEKQKKKIKESDLFVGSFYKNDLIMYEDELFR VIGVNSDINNLVELNMVDITYKDFCEVNNVTGEKHIKKTIGKRVVLIEKYTTDILGNLYKTPLPKK PQLIFKRGEL	N588A
ScaCas9-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRN YPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQL SKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQD LALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRSGKLATEEEFYKFIKPILEKMDGA EELLAKLNREDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGP LARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFT VYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL KRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQG DSLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRI EEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDD SIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKA GFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINN YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMN FFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKES ILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSY EKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQHLVRLLYYTQ NISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLL KYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A
SpyCas9-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL	N863A

		AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPRAFKYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-SpRY	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAERTRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDK LIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT RLGAPRAFKYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A
St1Cas9	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY SKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADL QFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQK HYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPK LDF	N622A
BlatCas9	Brevibacillus laterosporus	MAYTMGIDVGIASCGWAIVDLERQRIIDIGVRTFEKAENPKNGEALAVPRREARSSRRRLRRKKHR IERLKHMFVRNGLAVDIQHLEQTLRSQNEIDVWQLRVDGLDRMLTQKEWLRVLIHLAQRRGFQS NRKTDGSSEDGQVLVNVTENDRLMEEKDYRTVAEMMVKDEKFSDHKRNKNGNYHGVVSRSSL LVEIHTLFETQRQHHNSLASKDFELEYVNIWSAQRPVATKDQIEKMIGTCTFLPKEKRAPKASWHF QYFMLLQTINHIRITNVQGTRSLNKEEIEQVVNMALTKSKVSYHDTRKILDLSEEYQFVGLDYGKE DEKKKVESKETIIKLDDYHKLNKIFNEVELAKGETWEADDYDTVAYALTFFKDDEDIRDYLQNKY KDSKNRLVKNLANKEYTNELIGKVSTLSFRKVGHLSLKALRKIIPFLEQGMTYDKACQAAGFDFQ GISKKKRSVVLPVIDQISNPVVNRALTQTRKVINALIKKYGSPETIHIETARELSKTFDERKNITKDY KENRDKNEHAKKHLSELGIINPTGLDIVKYKLWCEQQGRCMYSNQPISFERLKESGYTEVDHIIPY SRSMNDSYNNRVLVMTRENREKGNQTPFEYMGNDTQRWYEFEQRVTTNPQIKKEKRQNLLLKG FTNRRELEMLERNLNDTRYITKYLSHFISTNLEFSPSDKKKKVVNTSGRITSHLRSRWGLEKNRGQ NDLHHAMDAIVIAVTSDSFIQQVTNYYKKRERELNGDDKFPLPWKFFREEVIARLSPNPKEQIEA LPNHFYSEDELADLQPIFVSRMPKRSITGEAHQAQFRRVVGKTKEGKNITAKKTALVDISYDKNGD FNMYGRETDPATYEAIKERYLEFGGNVKKAFSTDLHKPKKDGTKGPLIKSVRIMENKTLVHPVNK GKGVVYNSSIVRTDVFQRKEKYYLLPVYVTDVTKGKLPNKVIVAKKGYHDWIEVDDSFTFLFSLY PNDLIFIRQNPKKKISLKKRIESHSISDSKEVQEIHAYYKGVDSSTAAIEFIIHDGSYYAKGVGVQNL DCFEKYQVDILGNYFKVKGEKRLELETSDSNHKGKDVNSIKSTSR	N607A
cCas9-v16	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN	N580A

		DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNSDKNNL IEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
cCas9-v17	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNSTRNI VELNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
cCas9-v21	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNSDDRNII ELNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
cCas9-v42	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVN NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNNRLNK IELNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
CdiCas9	Corynebacterium diphtheriae	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVTRLASSGIARRTRRL YRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVRAELAASYIADEKERGEKLSVALRHIAR HRGWRNPYAKVSSLYLPDGPSDAFKAIREEIKRASGQPVPETATVGQMVTLCELGTLKLRGEGGV LSARLQQSDYAREIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKASD AFQRYRIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIAEILGIDRGQLIGTAT MTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKALSNAEVDDFDSPEGAK VQAFFADLDDDVHAKLDSLHLPVGRAAYSEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPT PRIGEPVGNPAVDRVLKTVSRWLESATKTWGAPERVIIEHVREGFVTEKRAREMDGDMRRRAAR NAKLFQEMQEKLNVQGKPSRADLWRYQSVQRQNCQCAYCGSPITFSNSEMDHIVPRAGQGSTNT RENLVAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAVERTRHWVTDTGMRSTDFKKFTKAVVE	H573A (Alternate)

		RFQRATMDEEIDARSMESVAWMANELRSRVAQHFASHGTTVRVYRGSLTAEARRASGISGKLKF FDGVGKSRLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQEAPQWREFTGKDAEHRAA WRVWCQKMEKLSALLTEDLRDDRVVVMSNVRLRLGNGSAHKETIGKLSKVKLSSQLSVSDIDKA SSEALWCALTREPGFDPKEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAELGSSFH HARVYKITSGKKPAFAMLRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKLRDALATGNAEYL GWLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVDGFFSPSKLRLRPLQMSKEGIKKESAPEL SKIIDRPGWLPAVNKLFSDGNVTVVRRDSLGRVRLESTAHLPVTWKVQ	
CjeCas9	Campylobacter jejuni	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLN HLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYD DIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQ SFLKDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFM FVALTRIINLLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFI EFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNIS FKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRK VLNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILK LRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFE AFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVLNYTKDYLD FLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSI VKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALH EETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMD FALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLI VSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK	N582A
GeoCas9	Geobacillus stearothermophilus	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLARSARRRLRRRKHR LERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDRKLNNDELARVLLHLAKRRGFKSNRK SERSNKENSTMLKHIEENRAILSSYRTVGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFS KQREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHIN KLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYDRGESRKQNENIRF LELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLA NKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLLPN IPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIR QLMEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDSYTNK VLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEETEFKNRN LNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIV ACTTPSDIAKVTAFYQRREQNKELAKKTEPHFPQPWPHFADELRARLSKHPKESIKALNLGNYDD QKLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDASGHFPMYGK ESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIIDTKNQVIPLNDGKTVAYN SNIVRVDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIE LPREKTVKTAAGEEINVKDVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLG NIYKVRGEKRVGLASSAHSKPGKTIRPLQSTRD	N605A
iSpyMac Cas9	Streptococcus spp.	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLKRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPL EVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQ VGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNH NQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGV KLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGEDSGGSGGSKRTADGSEFES	N863A
NmeCas9	Neisseria meningitidis	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLAR SVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVL LHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQR SDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEP AEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLE DTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTD EDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTE EKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENR KDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFS RTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKF	N611A

		DEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAEND RHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEV MIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVK SAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKYDK AGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGIL PDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTGNINIRIHDLDH KIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR	
ScaCas9	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRN YPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQL SKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQD LALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGIGIKHRKRTTKLATQEEFYKFIKPILEKMDGAE ELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPL ARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTV YNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDR FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGD SLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEE GIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSI DNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAG FIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNY HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNF FKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESI LSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYE KDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQHLVRLLYYTQN ISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLK YTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A
ScaCas9-HiFi-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRN YPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQL SKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQD LALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRSGKLATEEEFYKFIKPILEKMDGA EELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGP LARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFT VYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL KRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNANFMQLIHDDSLTFKEEIEKAQVSGQ GDSLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKR IEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKD DSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADK AGFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDIN NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIM NFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKE SILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSY EKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQHLVRLLYYTQ NISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLL KYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A
SpyCas9-3var-NRRH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTAAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIGFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLHKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGVPAAFKYFDTTIDKKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A

SpyCas9-3var-NRTH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIGFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASASVLHKGNELALPSKYVNFLYLASHYEKLKGSS EDNKQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGASAAFKYFDTTIGRKLYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A
SpyCas9-3var-NRCH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETRIDLSQLGGD	N863A
SpyCas9-HF1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A
SpyCas9-QQR1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW	N863A

		MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAQLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTFKQKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-SpG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A
SpyCas9-VQR	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A
SpyCas9-VRER	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL	N863A

		VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
SpyCas9-xCas	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A
SpyCas9-xCas-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPRAFKYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A
St1Cas9-CNRZ1066	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRK YSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKATGKYEILGLKYA DLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTLPKQ KHYVELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLAKSNISIYKVRTDVLGNQHIIKNEGDKP KLDF	N622A
St1Cas9-LMG1831	Streptococcus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT	N622A

	thermophi lus	QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRK YSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYA DLQFEKKTGTYKISQEKYNGIMKEEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPN VKYYVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPK LDF	
St1Cas9-MTH17CL396	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY SKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDM QFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHY VELKPYNRQKFEGSEYLIKSLGTVAKGGQCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A
St1Cas9-TH1477	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVG DYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFR AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY SKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDM QFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHY VELKPYNRQKFEGSEYLIKSLGTVVKGGRCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A

Table 9B provides parameters to define the necessary components for designing gRNA and/or Template RNAs to apply Cas variants listed in Table 3A for Gene Writing. Tier indicates preferred Cas variants if they are available for use at a given locus. The cut site indicates the validated or predicted protospacer adjacent motif (PAM) requirements, validated or predicted location of cut site (relative to the most upstream base of the PAM site). The gRNA for a given enzyme can be assembled by concatenating the crRNA, Tetraloop, and tracrRNA sequences, and further adding a 5' spacer of a length within Spacer (min) and Spacer (max) that matches a protospacer at a target site. Further, the predicted location of the ssDNA nick at the target is

important for designing the 3' region of a Template RNA that needs to anneal to the sequence immediately 5' of the nick in order to initiate target primed reverse transcription.

Table 9B parameters to define the necessary components for designing gRNA and/or Template RNAs to apply Cas variants listed in Table 9A for Gene Writing

Variant	PAM(s)	Cut	Tier	Spacer (min)	Spacer (max)	crRNA	Tetraloop	tracrRNA
Nme2Cas9	NNNNCC	-3	1	22	24	GTTGTAGCTCCCTTTCTCATTTCG	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
PpnCas9	NNNNRTT		1	21	24	GTTGTAGCTCCCTTTTTCATTTCGC	GAAA	GCGAAATGAAAAACGTTGTTACAATAAGAGATGAATTTCTCGCAAAGCTCTGCCTCTTGAAATTTCGGTTTCAAGAGGCATCTTTTT
SauCas9	NNGRR;NNGRRT	-3	1	21	23	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SauCas9-KKH	NNNRR;NNNRRT	-3	1	21	21	GTTTTAGTACTCTGTAAT	GAAA	ATTACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SauriCas9	NNGG	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTT
SauriCas9-KKH	NNRG	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTT
ScaCas9-Sc++	NNG	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9	NGG	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTT

								ATCAACTTGAAAAA GTGGCACCGAGTCG GTGC
SpyCas9 -NG	NG (NGG= NGA=N GT>NG C)	-3	1	20	20	GTTTAAGAGCT ATGCTG	GAAA	CAGCATAGCAAGTT TAAATAAGGCTAGT CCGTTATCAACTTGA AAAAGTGGCACCGA GTCGGTGC
SpyCas9 -SpRY	NRN>N YN	-3	1	20	20	GTTTTAGAGCT A	GAAA	TAGCAAGTTAAAAT AAGGCTAGTCCGTT ATCAACTTGAAAAA GTGGCACCGAGTCG GTGC
St1Cas9	NNAGA AW>NN AGGA W=NNG GAAW	-3	1	20	20	GTCTTTGTACT CTG	GTAC	CAGAAGCTACAAAG ATAAGGCTTCATGC CGAAATCAACACCC TGTCATTTTATGGCA GGGTGTTTT
BlatCas 9	NNNNC NAA>N NNNCN DD>NN NNC	-3	1	19	23	GCTATAGTTCC TTACT	GAAA	GGTAAGTTGCTATA GTAAGGGCAACAGA CCCGAGGCGTTGGG GATCGCCTAGCCCG TGTTTACGGGCTCTC CCCATATTCAAAAT AATGACAGACGAGC ACCTTGGAGCATTTA TCTCCGAGGTGCT
cCas9- v16	NNVAC T;NNV ATGM; NNVAT T;NNV GCT;N NVGTG ;NNVG TT	-3	2	21	21	GUCUUAGUAC UCUG	GAAA	CAGAAUCUACUAAG ACAAGGCAAAAUGC CGUGUUUAUCUCGU CAACUUGUUGGCGA GAUUUUUUU
cCas9- v17	NNVRR N	-3	2	21	21	GUCUUAGUAC UCUG	GAAA	CAGAAUCUACUAAG ACAAGGCAAAAUGC CGUGUUUAUCUCGU CAACUUGUUGGCGA GAUUUUUUU
cCas9- v21	NNVAC T;NNV ATGM; NNVAT T;NNV GCT;N NVGTG ;NNVG TT	-3	2	21	21	GUCUUAGUAC UCUG	GAAA	CAGAAUCUACUAAG ACAAGGCAAAAUGC CGUGUUUAUCUCGU CAACUUGUUGGCGA GAUUUUUUU
cCas9- v42	NNVRR N	-3	2	21	21	GUCUUAGUAC UCUG	GAAA	CAGAAUCUACUAAG ACAAGGCAAAAUGC CGUGUUUAUCUCGU

								CAACUUGUUGGCGA GAUUUUUUU
CdiCas9	NNRHH HY;NN RAAAY		2	22	22	ACUGGGGUUC AG	GAAA	CUGAACCUCAGUAA GCAUUGGCUCGUUU CCAAUGUUGAUUGC UCCGCCGGUGCUCC UUAUUUUUAAGGGC GCCGGC
CjeCas9	NNNNR YAC	-3	2	21	23	GTTTTAGTCCC T	GAAA	AGGGACTAAAATAA AGAGTTTGCGGGAC TCTGCGGGGTTACA ATCCCCTAAAACCG CTTTTTT
GeoCas 9	NNNNC RAA		2	21	23	GUCAUAGUUC CCCUGA	GAAA	UCAGGGUUACUAUG AUAAGGGCUUUCUG CCUAAGGCAGACUG ACCCGCGGCGUUGG GGAUCGCCUGUCGC CCGCUUUUGGCGGG CAUUCCCCAUCCUU
iSpyMac Cas9	NAAN	-3	2	19	21	GTTTTAGAGCT A	GAAA	TAGCAAGTTAAAAT AAGGCTAGTCCGTT ATCAACTTGAAAAA GTGGCACCGAGTCG GTGC
NmeCas 9	NNNNG AYT;N NNNGY TT;NNN NGAYA ;NNNN GTCT	-3	2	20	24	GTTGTAGCTCC CTTTCTCATTT CG	GAAA	CGAAATGAGAACCG TTGCTACAATAAGG CCGTCTGAAAAGAT GTGCCGCAACGCTC TGCCCCTTAAAGCTT CTGCTTTAAGGGGC ATCGTTTA
ScaCas9	NNG	-3	2	20	20	GTTTTAGAGCT A	GAAA	TAGCAAGTTAAAAT AAGGCTAGTCCGTT ATCAACTTGAAAAA GTGGCACCGAGTCG GTGC
ScaCas9 -HiFi-Sc++	NNG	-3	2	20	20	GTTTTAGAGCT A	GAAA	TAGCAAGTTAAAAT AAGGCTAGTCCGTT ATCAACTTGAAAAA GTGGCACCGAGTCG GTGC
SpyCas9 -3var-NRRH	NRRH	-3	2	20	20	GTTTAAGAGCT ATGCTG	GAAA	CAGCATAGCAAGTT TAAATAAGGCTAGT CCGTTATCAACTTGA AAAAGTGGCACCGA GTCGGTGC
SpyCas9 -3var-NRTH	NRTH	-3	2	20	20	GTTTAAGAGCT ATGCTG	GAAA	CAGCATAGCAAGTT TAAATAAGGCTAGT CCGTTATCAACTTGA AAAAGTGGCACCGA GTCGGTGC

SpyCas9-3var-NRCH	NRCH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-HF1	NGG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-QQR1	NAAG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-SpG	NGN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VQR	NGAN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VRER	NGCG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas	NG;GAA;GAT	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas-NG	NG	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
St1Cas9-CNRZ1066	NNACAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-LMG1831	NNGCAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-	NNAAAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCC

MTH17 CL396								TGTCATTTTATGGCA GGGTGTTTT
St1Cas9 - TH1477	NNGAA A	-3	2	20	20	GTCTTTGTACT CTG	GTAC	CAGAAGCTACAAAG ATAAGGCTTCATGC CGAAATCAACACCC TGTCATTTTATGGCA GGGTGTTTT

In some embodiments, the opener molecule is exogenous to the cell comprising the Gene Writer polypeptide and or template nucleic acid. In some embodiments, the opener molecule comprises an endogenous agent (e.g., endogenous to the cell comprising the Gene Writer polypeptide and or template nucleic acid comprising the gRNA and blocking domain). For example, an inducible gRNA, blocking domain, and opener molecule may be chosen such that the opener molecule is an endogenous agent expressed in a target cell or tissue, e.g., thereby ensuring activity of a Gene Writer system in the target cell or tissue. As a further example, an inducible gRNA, blocking domain, and opener molecule may be chosen such that the opener molecule is absent or not substantially expressed in one or more non-target cells or tissues, e.g., thereby ensuring that activity of a Gene Writer system does not occur or substantially occur in the one or more non-target cells or tissues, or occurs at a reduced level compared to a target cell or tissue. Exemplary blocking domains, opener molecules, and uses thereof are described in PCT App. Publication WO2020044039A1, which is incorporated herein by reference in its entirety.In some embodiments, the template nucleic acid, e.g., template RNA, may comprise one or more UTRs (e.g. from an R2-type retrotransposon) and a gRNA. In some embodiments, the UTR facilitates interaction of the template nucleic acid (e.g., template RNA) with the writing domain, e.g., reverse transcriptase domain, of the Gene Writer polypeptide. In some embodiments, the gRNA facilitates interaction with the template nucleic acid binding domain (e.g., RNA binding domain) of the polypeptide. In some embodiments, the gRNA directs the polypeptide to the matching target sequence, e.g., in a target cell genome. In some embodiments, the template nucleic acid may contain only the reverse transcriptase binding motif (e.g. 3' UTR from R2) and the gRNA may be provided as a second nucleic acid molecule (e.g., second RNA molecule) for target site recognition. In some embodiments, the template nucleic acid containing the RT-binding motif may exist on the same molecule as the gRNA, but be processed into two RNA molecules by cleavage activity (e.g. ribozyme).

In some embodiments, a template RNA may be customized to correct a given mutation in the genomic DNA of a target cell (e.g., ex vivo or in vivo, e.g., in a target tissue or organ, e.g., in a subject). For example, the mutation may be a disease-associated mutation relative to the wild-type sequence. Without wishing to be bound by theory, sets of empirical parameters help ensure optimal initial in silico designs of template RNAs or portions thereof. As a non-limiting illustrative example, for a selected mutation, the following design parameters may be employed. In some embodiments, design is initiated by acquiring approximately 500 bp (e.g., up to 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, or 700 bp, and optionally at least 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, or 650 bp) flanking sequence on either side of the mutation to serve as the target region. In some embodiments, a template nucleic acid comprises a gRNA. Methodology for designing gRNAs is known to those of skill in the art. In some embodiments, a gRNA comprises a sequence (e.g., a CRISPR spacer) that binds a target site . In some embodiments, the sequence (e.g., a CRISPR spacer) that binds a target site for use in targeting a template nucleic acid to a target region is selected by considering the particular Gene Writer polypeptide (e.g., endonuclease domain or writing domain, e.g., comprising a CRISPR/Cas domain) being used (e.g., for Cas9, a protospacer-adjacent motif (PAM) of NGG immediately 3' of a 20 nt gRNA binding region). In some embodiments, the CRISPR spacer is selected by ranking first by whether the PAM will be disrupted by the Gene Writing induced edit. In some embodiments, disruption of the PAM may increase edit efficiency. In some embodiments, the PAM can be disrupted by also introducing (e.g., as part of or in addition to another modification to a target site in genomic DNA) a silent mutation (e.g., a mutation that does not alter an amino acid residue encoded by the target nucleic acid sequence, if any) in the target site during Gene Writing. In some embodiments, the CRISPR spacer is selected by ranking sequences by the proximity of their corresponding genomic site to the desired edit location. In some embodiments, the gRNA comprises a gRNA scaffold. In some embodiments, the gRNA scaffold used may be a standard scaffold (e.g., for Cas9, 5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC-3'), or may contain one or more nucleotide substitutions. In some embodiments, the heterologous object sequence has at least 90% identity, e.g., at least 90%, 95%, 98%, 99%, or 100% identity, or comprises no more than 1, 2, 3, 4, or 5 positions of non-identity to the target site 3' of the first strand nick (e.g., immediately 3' of the first strand

nick or up to 1, 2, 3, 4, or 5 nucleotides 3' of the first strand nick), with the exception of any insertion, substitution, or deletion that may be written into the target site by the Gene Writer.In some embodiments, the 3' target homology domain contains at least 90% identity, e.g., at least 90%, 95%, 98%, 99%, or 100% identity, or comprises no more than 1, 2, 3, 4, or 5 positions of non-identity to the target site 5' of the first strand nick (e.g., immediately 5' of the first strand nick or up to 1, 2, 3, 4, or 5 nucleotides 3' of the first strand nick).

In some embodiments, a template RNA can comprise a gRNA sequence, e.g., to direct the GeneWriter to a target site of interest. In some embodiments, a template RNA comprises (e.g., from 5' to 3') (i) optionally a sequence (e.g., a CRISPR spacer) that binds a target site (e.g., a second strand of a site in a target genome), (ii) optionally a sequence that binds a polypeptide described herein (e.g., a GeneWriter or a Cas polypeptide), (iii) a heterologous object sequence, and (iv) 5' homology domain and/or a 3' target homology domain.

In some embodiments, the template nucleic acid molecule comprises a 5' homology domain and/or a 3' homology domain. In some embodiments, the 5' homology domain comprises a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with a nucleic acid sequence comprised in a target nucleic acid molecule. In embodiments, the nucleic acid sequence in the target nucleic acid molecule is within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of (e.g., 5' relative to) a target insertion site, e.g., for a heterologous object sequence, e.g., comprised in the template nucleic acid molecule.

In some embodiments, the 3' homology domain comprises a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with a nucleic acid sequence comprised in a target nucleic acid molecule. In embodiments, the nucleic acid sequence in the target nucleic acid molecule is within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides of (e.g., 3' relative to) a target insertion site, e.g., for a heterologous object sequence, e.g., comprised in the template nucleic acid molecule. In some embodiments, the 5' homology domain is heterologous to the remainder of the template nucleic acid molecule. In some embodiments, the 3' homology domain is heterologous to the remainder of the template nucleic acid molecule.

In some embodiments, a template nucleic acid (e.g., template RNA) comprises a 3' target homology domain. In some embodiments, a 3' target homology domain is disposed 3' of the

heterologous object sequence and is complementary to a sequence adjacent to a site to be modified by a system described herein, or comprises no more than 1, 2, 3, 4, or 5 mismatches to a sequence complementary to the sequence adjacent to a site to be modified by the system/Gene Writer™. In some embodiments, the 3' homology domain binds within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nick site in the target nucleic acid molecule. In some embodiments, binding of the 3' homology domain to the target nucleic acid molecule permits initiation of target-primed reverse transcription (TPRT), e.g., with the 3' homology domain acting as a primer for TPRT.

Methods and Compositions for Modified RNA (e.g., gRNA or template RNA)

In some embodiments, an RNA component of the system (e.g., a template RNA or a gRNA, e.g., as described herein) comprises one or more nucleotide modifications. In some embodiments, the modification pattern of a gRNA can significantly affect *in vivo* activity compared to unmodified or end-modified guides (e.g., as shown in Figure 1D from Finn et al. *Cell Rep* 22(9):2227-2235 (2018); incorporated herein by reference in its entirety). Without wishing to be bound by theory, this process may be due, at least in part, to a stabilization of the RNA conferred by the modifications. Non-limiting examples of such modifications may include 2'-O-methyl (2'-O-Me), 2'-0-(2-methoxyethyl) (2'-0-MOE), 2'- fluoro (2'-F), phosphorothioate (PS) bond between nucleotides, G-C substitutions, and inverted abasic linkages between nucleotides and equivalents thereof.

In some embodiments, the template RNA (e.g., at the portion thereof that binds a target site) or the guide RNA comprises a 5' terminus region. In some embodiments, the template RNA or the guide RNA does not comprise a 5' terminus region. In some embodiments, the 5' terminus region comprises a CRISPR spacer region, e.g., as described with respect to sgRNA in Briner AE et al, Molecular Cell 56: 333-339 (2014) (incorporated herein by reference in its entirety; applicable herein, e.g., to all guide RNAs). In some embodiments, the 5' terminus region comprises a 5' end modification. In some embodiments, a 5' terminus region with or without a spacer region may be associated with a crRNA, trRNA, sgRNA and/or dgRNA. The CRISPR spacer region can, in some instances, comprise a guide region, guide domain, or targeting domain. In some embodiments, a target domain or target sequence may comprise a sequence of nucleic acid to which the guide region/domain directs a nuclease for cleavage. In some

embodiments, a spyCas9 protein may be directed by a guide region/domain to a target sequence of a target nucleic acid molecule by the nucleotides present in the CRISPR spacer region.

In some embodiments, the template RNA (e.g., at the portion thereof that binds a target site) or guide RNA, e.g., as described herein, comprises any of the sequences shown in Table 4 of WO2018107028A1, incorporated herein by reference in its entirety. In some embodiments, where a sequence shows a guide and/or spacer region, the composition may comprise this region or not. In some embodiments, a guide RNA comprises one or more of the modifications of any of the sequences shown in Table 4 of WO2018107028A1, e.g., as identified therein by a SEQ ID NO. In embodiments, the nucleotides may be the same or different, and/or the modification pattern shown may be the same or similar to a modification pattern of a guide sequence as shown in Table 4 of WO2018107028A1. In some embodiments, a modification pattern includes the relative position and identity of modifications of the gRNA or a region of the gRNA (e.g. 5' terminus region, lower stem region, bulge region, upper stem region, nexus region, hairpin 1 region, hairpin 2 region, 3' terminus region). In some embodiments, the modification pattern contains at least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the modifications of any one of the sequences shown in the sequence column of Table 4 of WO2018107028A1, and/or over one or more regions of the sequence. In some embodiments, the modification pattern is at least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the modification pattern of any one of the sequences shown in the sequence column of Table 4 of WO2018107028A1. In some embodiments, the modification pattern is at least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over one or more regions of the sequence shown in Table 4 of WO2018107028A1, e.g., in a 5 ' terminus region, lower stem region, bulge region, upper stem region, nexus region, hairpin 1 region, hairpin 2 region, and/or 3' terminus region. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the modification pattern of a sequence over the 5 ' terminus region. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the lower stem. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the bulge. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%,

98%, 99%, or 100% identical over the upper stem. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the nexus. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the hairpin 1. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the hairpin 2. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the 3 ' terminus. In some embodiments, the modification pattern differs from the modification pattern of a sequence of Table 4 of WO2018107028A1, or a region (e.g. 5' terminus, lower stem, bulge, upper stem, nexus, hairpin 1, hairpin 2, 3' terminus) of such a sequence, e.g., at 0, 1, 2, 3, 4, 5, 6, or more nucleotides. In some embodiments, the gRNA comprises modifications that differ from the modifications of a sequence of Table 4 of WO2018107028A1, e.g., at 0, 1, 2, 3, 4, 5, 6, or more nucleotides. In some embodiments, the gRNA comprises modifications that differ from modifications of a region (e.g. 5 ' terminus, lower stem, bulge, upper stem, nexus, hairpin 1, hairpin 2, 3' terminus) of a sequence of Table 4 of WO2018107028A1, e.g., at 0, 1, 2, 3, 4, 5, 6, or more nucleotides.

In some embodiments, the template RNA (e.g., at the portion thereof that binds a target site) or the gRNA comprises a 2'-O-methyl (2'-O-Me) modified nucleotide. In some embodiments, the gRNA comprises a 2'-O-(2-methoxy ethyl) (2'-O-moe) modified nucleotide. In some embodiments, the gRNA comprises a 2'-fluoro (2'- F) modified nucleotide. In some embodiments, the gRNA comprises a phosphorothioate (PS) bond between nucleotides. In some embodiments, the gRNA comprises a 5' end modification, a 3' end modification, or 5' and 3' end modifications. In some embodiments, the 5' end modification comprises a phosphorothioate (PS) bond between nucleotides. In some embodiments, the 5' end modification comprises a 2'-O-methyl (2'-O-Me), 2'-O-(2-methoxy ethyl) (2'-O-MOE), and/or 2'-fluoro (2'-F) modified nucleotide. In some embodiments, the 5' end modification comprises at least one phosphorothioate (PS) bond and one or more of a 2'-O-methyl (2'-O- Me), 2'-O-(2-methoxyethyl) (2'-O-MOE), and/or 2'-fluoro (2'-F) modified nucleotide. The end modification may comprise a phosphorothioate (PS), 2'-O-methyl (2'-O-Me) , 2'-O-(2- methoxyethyl) (2'-O-MOE), and/or 2'-fluoro (2'-F) modification. Equivalent end modifications are also encompassed by embodiments described herein. In some embodiments, the template RNA or gRNA comprises

an end modification in combination with a modification of one or more regions of the template RNA or gRNA. Additional exemplary modifications and methods for protecting RNA, e.g., gRNA, and formulae thereof, are described in WO2018126176A1, which is incorporated herein by reference in its entirety.

In some embodiments, structure-guided and systematic approaches are used to introduce modifications (e.g., 2′-OMe-RNA, 2′-F-RNA, and PS modifications) to a template RNA or guide RNA, for example, as described in Mir et al. *Nat Commun* 9:2641 (2018) (incorporated by reference herein in its entirety). In some embodiments, the incorporation of 2′-F-RNAs increases thermal and nuclease stability of RNA:RNA or RNA:DNA duplexes, e.g., while minimally interfering with C3′-endo sugar puckering. In some embodiments, 2′-F may be better tolerated than 2′-OMe at positions where the 2′-OH is important for RNA:DNA duplex stability. In some embodiments, a crRNA comprises one or more modifications that do not reduce Cas9 activity, e.g., C10, C20, or C21 (fully modified), e.g., as dscribed in Supplementary Table 1 of Mir et al. *Nat Commun* 9:2641 (2018), incorporated herein by reference in its entirety. In some embodiments, a tracrRNA comprises one or more modifications that do not reduce Cas9 activity, e.g., T2, T6, T7, or T8 (fully modified) of Supplementary Table 1 of Mir et al. *Nat Commun* 9:2641 (2018). In some embodiments, a crRNA comprises one or more modifications (e.g., as described herein) may be paired with a tracrRNA comprising one or more modifications, e.g., C20 and T2. In some embodiments, a gRNA comprises a chimera, e.g., of a crRNA and a tracrRNA (e.g., Jinek et al. *Science* 337(6096):816-821 (2012)). In embodiments, modifications from the crRNA and tracrRNA are mapped onto the single-guide chimera, e.g., to produce a modified gRNA with enhanced stability.

In some embodiments, gRNA molecules may be modified by the addition or subtraction of the naturally occurring structural components, e.g., hairpins. In some embodiments, a gRNA may comprise a gRNA with one or more 3’ hairpin elements deleted, e.g., as described in WO2018106727, incorporated herein by reference in its entirety. In some embodiments, a gRNA may contain an added hairpin structure, e.g., an added hairpin structure in the spacer region, which was shown to increase specificity of a CRISPR-Cas system in the teachings of Kocak et al. *Nat Biotechnol* 37(6):657-666 (2019). Additional modifications, including examples of shortened gRNA and specific modifications improving in vivo activity, can be found in US20190316121, incorporated herein by reference in its entirety.

In some embodiments, structure-guided and systematic approaches (e.g., as described in Mir et al. *Nat Commun* 9:2641 (2018); incorporated herein by reference in its entirety) are employed to find modifications for the template RNA. In embodiments, the modifications are identified with the inclusion or exclusion of a guide region of the template RNA. In some embodiments, a structure of polypeptide bound to template RNA is used to determine non-protein-contacted nucleotides of the RNA that may then be selected for modifications, e.g., with lower risk of disrupting the association of the RNA with the polypeptide. Secondary structures in a template RNA can also be predicted in silico by software tools, e.g., the RNAstructure tool available at rna.urmc.rochester.edu/RNAstructureWeb (Bellaousov et al. *Nucleic Acids Res* 41:W471-W474 (2013); incorporated by reference herein in its entirety), e.g., to determine secondary structures for selecting modifications, e.g., hairpins, stems, and/or bulges.

It is contemplated that it may be useful to employ circular and/or linear RNA states during the formulation, delivery, or Gene Writing reaction within the target cell. Thus, in some embodiments of any of the aspects described herein, a Gene Writing system comprises one or more circular RNAs (circRNAs). In some embodiments of any of the aspects described herein, a Gene Writing system comprises one or more linear RNAs. In some embodiments, a nucleic acid as described herein (e.g., a template nucleic acid, a nucleic acid molecule encoding a Gene Writer polypeptide, or both) is a circRNA. In some embodiments, a circular RNA molecule encodes the Gene Writer polypeptide. In some embodiments, the circRNA molecule encoding the Gene Writer polypeptide is delivered to a host cell. In some embodiments, a circular RNA molecule encodes a recombinase, e.g., as described herein. In some embodiments, the circRNA molecule encoding the recombinase is delivered to a host cell. In some embodiments, the circRNA molecule encoding the Gene Writer polypeptide is linearized (e.g., in the host cell, e.g., in the nucleus of the host cell) prior to translation.

Circular RNAs (circRNAs) have been found to occur naturally in cells and have been found to have diverse functions, including both non-coding and protein coding roles in human cells. It has been shown that a circRNA can be engineered by incorporating a self-splicing intron into an RNA molecule (or DNA encoding the RNA molecule) that results in circularization of the RNA, and that an engineered circRNA can have enhanced protein production and stability (Wesselhoeft et al. *Nature Communications* 2018). In some embodiments, the Gene Writer™

polypeptide is encoded as circRNA. In certain embodiments, the template nucleic acid is a DNA, such as a dsDNA or ssDNA.

In some embodiments, the circRNA comprises one or more ribozyme sequences. In some embodiments, the ribozyme sequence is activated for autocleavage, e.g., in a host cell, e.g., thereby resulting in linearization of the circRNA. In some embodiments, the ribozyme is activated when the concentration of magnesium reaches a sufficient level for cleavage, e.g., in a host cell. In some embodiments the circRNA is maintained in a low magnesium environment prior to delivery to the host cell. In some embodiments, the ribozyme is a protein-responsive ribozyme. In some embodiments, the ribozyme is a nucleic acid-responsive ribozyme. In some embodiments, the circRNA comprises a cleavage site. In some embodiments, the circRNA comprises a second cleavage site.

In some embodiments, the circRNA is linearized in the nucleus of a target cell. In some embodiments, linearization of a circRNA in the nucleus of a cell involves components present in the nucleus of the cell, e.g., to activate a cleavage event. For example, the B2 and ALU retrotransposons contain self-cleaving ribozymes whose activity is enhanced by interaction with the Polycomb protein, EZH2 (Hernandez et al. PNAS 117(1):415-425 (2020)). Thus, in some embodiments, a ribozyme, e.g., a ribozyme from a B2 or ALU element, that is responsive to a nuclear element, e.g., a nuclear protein, e.g., a genome-interacting protein, e.g., an epigenetic modifier, e.g., EZH2, is incorporated into a circRNA, e.g., of a Gene Writing system. In some embodiments, nuclear localization of the circRNA results in an increase in autocatalytic activity of the ribozyme and linearization of the circRNA.

In some embodiments, the ribozyme is heterologous to one or more of the other components of the Gene Writing system. In some embodiments, an inducible ribozyme (e.g., in a circRNA as described herein) is created synthetically, for example, by utilizing a protein ligand-responsive aptamer design. A system for utilizing the satellite RNA of tobacco ringspot virus hammerhead ribozyme with an MS2 coat protein aptamer has been described (Kennedy et al. Nucleic Acids Res 42(19):12306-12321 (2014), incorporated herein by reference in its entirety) that results in activation of the ribozyme activity in the presence of the MS2 coat protein. In embodiments, such a system responds to protein ligand localized to the cytoplasm or the nucleus. In some embodiments the protein ligand is not MS2. Methods for generating RNA aptamers to target ligands have been described, for example, based on the systematic evolution of ligands by

exponential enrichment (SELEX) (Tuerk and Gold, Science 249(4968):505-510 (1990); Ellington and Szostak, Nature 346(6287):818-822 (1990); the methods of each of which are incorporated herein by reference) and have, in some instances, been aided by in silico design (Bell et al. PNAS 117(15):8486-8493, the methods of which are incorporated herein by reference). Thus, in some embodiments, an aptamer for a target ligand is generated and incorporated into a synthetic ribozyme system, e.g., to trigger ribozyme-mediated cleavage and circRNA linearization, e.g., in the presence of the protein ligand. In some embodiments, circRNA linearization is triggered in the cytoplasm, e.g., using an aptamer that associates with a ligand in the cytoplasm. In some embodiments, circRNA linearization is triggered in the nucleus, e.g., using an aptamer that associates with a ligand in the nucleus. In embodiments, the ligand in the nucleus comprises an epigenetic modifier or a transcription factor. In some embodiments the ligand that triggers linearization is present at higher levels in on-target cells than off-target cells.

It is further contemplated that a nucleic acid-responsive ribozyme system can be employed for circRNA linearization. For example, biosensors that sense defined target nucleic acid molecules to trigger ribozyme activation are described, e.g., in Penchovsky (Biotechnology Advances 32(5):1015-1027 (2014), incorporated herein by reference). By these methods, a ribozyme naturally folds into an inactive state and is only activated in the presence of a defined target nucleic acid molecule (e.g., an RNA molecule). In some embodiments, a circRNA of a Gene Writing system comprises a nucleic acid-responsive ribozyme that is activated in the presence of a defined target nucleic acid, e.g., an RNA, e.g., an mRNA, miRNA, guide RNA, gRNA, sgRNA, ncRNA, lncRNA, tRNA, snRNA, or mtRNA. In some embodiments the nucleic acid that triggers linearization is present at higher levels in on-target cells than off-target cells.

In some embodiments of any of the aspects herein, a Gene Writing system incorporates one or more ribozymes with inducible specificity to a target tissue or target cell of interest, e.g., a ribozyme that is activated by a ligand or nucleic acid present at higher levels in a target tissue or target cell of interest. In some embodiments, the Gene Writing system incorporates a ribozyme with inducible specificity to a subcellular compartment, e.g., the nucleus, nucleolus, cytoplasm, or mitochondria. In some embodiments, the ribozyme that is activated by a ligand or nucleic acid present at higher levels in the target subcellular compartment. In some embodiments, an RNA component of a Gene Writing system is provided as circRNA, e.g., that is activated by linearization. In some embodiments, linearization of a circRNA encoding a Gene Writing

polypeptide activates the molecule for translation. In some embodiments, a signal that activates a circRNA component of a Gene Writing system is present at higher levels in on-target cells or tissues, e.g., such that the system is specifically activated in these cells.

In some embodiments, an RNA component of a Gene Writing system is provided as a circRNA that is inactivated by linearization. In some embodiments, a circRNA encoding the Gene Writer polypeptide is inactivated by cleavage and degradation. In some embodiments, a circRNA encoding the Gene Writing polypeptide is inactivated by cleavage that separates a translation signal from the coding sequence of the polypeptide. In some embodiments, a signal that inactivates a circRNA component of a Gene Writing system is present at higher levels in off-target cells or tissues, such that the system is specifically inactivated in these cells.

Further included here are compositions and methods for the assembly of full or partial template RNA molecules (e.g., Gene Writing template RNA molecules optionally comprising a gRNA, or separate gRNA molecules). In some embodiments, RNA molecules may be assembled by the connection of two or more (e.g., two, three, four, five, six, seven, eight, nine, ten, or more) RNA segments with each other. In an aspect, the disclosure provides methods for producing nucleic acid molecules, the methods comprising contacting two or more linear RNA segments with each other under conditions that allow for the 5′ terminus of a first RNA segment to be covalently linked with the 3′ terminus of a second RNA segment. In some embodiments, the joined molecule may be contacted with a third RNA segment under conditions that allow for the 5’ terminus of the joined molecule to be covalently linked with the 3’ terminus of the third RNA segment. In embodiments, the method further comprises joining a fourth, fifth, or additional RNA segments to the elongated molecule. This form of assembly may, in some instances, allow for rapid and efficient assembly of RNA molecules.

The present disclosure also provides compositions and methods for the connection (e.g., covalent connection) of crRNA molecules and tracrRNA molecules. In some embodiments, guide RNA molecules with specificity for different target sites can be generated using a single tracrRNA molecule/segment connected to a target site specific crRNA molecule/segment (e.g., as shown in FIG. 10 of US20160102322A1; incorporated herein by reference in its entirety). For example, FIG. 10 of US20160102322A1 shows four tubes with different crRNA molecules with crRNA molecule 3 being connected to a tracrRNA molecule to form a guide RNA molecule,

thereby depicting an exemplary connection of two RNA segments to form a product RNA molecule.

The disclosure also provides compositions and methods for the production of template RNA molecules with specificity for a Gene Writer polypeptide and/or a genomic target site. In an aspect, the method comprises: (1) identification of the target site and desired modification thereto, (2) production of RNA segments including an upstream homology segment, a heterologous object sequence segment, a Gene Writer polypeptide binding motif, and a gRNA segment, and/or (3) connection of the four or more segments into at least one molecule, e.g., into a single RNA molecule. In some embodiments, some or all of the template RNA segments comprised in (2) are assembled into a template RNA molecule, e.g., one, two, three, or four of the listed components. In some embodiments, the segments comprised in (2) may be produced in further segmented molecules, e.g., split into at least 2, at least 3, at least 4, or at least 5 or more sub-segments, e.g., that are subsequently assembled, e.g., by one or more methods described herein.

In some embodiments, RNA segments may be produced by chemical synthesis. In some embodiments, RNA segments may be produced by in vitro transcription of a nucleic acid template, e.g., by providing an RNA polymerase to act on a cognate promoter of a DNA template to produce an RNA transcript. In some embodiments, in vitro transcription is performed using, e.g., a T7, T3, or SP6 RNA polymerase, or a derivative thereof, acting on a DNA, e.g., dsDNA, ssDNA, linear DNA, plasmid DNA, linear DNA amplicon, linearized plasmid DNA, e.g., encoding the RNA segment, e.g., under transcriptional control of a cognate promoter, e.g., a T7, T3, or SP6 promoter. In some embodiments, a combination of chemical synthesis and in vitro transcription is used to generate the RNA segments for assembly. In embodiments, the gRNA, upstream target homology, and Gene Writer polypeptide binding segments are produced by chemical synthesis and the heterologous object sequence segment is produced by in vitro transcription. Without wishing to be bound by theory, in vitro transcription may be better suited for the production of longer RNA molecules. In some embodiments, reaction temperature for in vitro transcription may be lowered, e.g., be less than 37°C (e.g., between 0-10C, 10-20C, or 20-30C), to result in a higher proportion of full-length transcripts (Krieg *Nucleic Acids Res* 18:6463 (1990)). In some embodiments, a protocol for improved synthesis of long transcripts is employed to synthesize a long template RNA, e.g., a template RNA greater than 5 kb, such as the use of

e.g., T7 RiboMAX Express, which can generate 27 kb transcripts in vitro (Thiel et al. J Gen Virol 82(6):1273-1281 (2001)). In some embodiments, modifications to RNA molecules as described herein may be incorporated during synthesis of RNA segments (e.g., through the inclusion of modified nucleotides or alternative binding chemistries), following synthesis of RNA segments through chemical or enzymatic processes, following assembly of one or more RNA segments, or a combination thereof.

In some embodiments, an mRNA of the system (e.g., an mRNA encoding a Gene Writer polypeptide) is synthesized in vitro using T7 polymerase-mediated DNA-dependent RNA transcription from a linearized DNA template, where UTP is optionally substituted with 1-methylpseudoUTP. In some embodiments, the transcript incorporates 5′ and 3′ UTRs, e.g., GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC and UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA, or functional fragments or variants thereof, and optionally includes a poly-A tail, which can be encoded in the DNA template or added enzymatically following transcription. In some embodiments, a donor methyl group, e.g., S-adenosylmethionine, is added to a methylated capped RNA with cap 0 structure to yield a cap 1 structure that increases mRNA translation efficiency (Richner et al. *Cell* 168(6): P1114-1125 (2017)).

In some embodiments, the transcript from a T7 promoter starts with a GGG motif. In some embodiments, a transcript from a T7 promoter does not start with a GGG motif. It has been shown that a GGG motif at the transcriptional start, despite providing superior yield, may lead to T7 RNAP synthesizing a ladder of poly(G) products as a result of slippage of the transcript on the three C residues in the template strand from +1 to +3 (Imburgio et al. Biochemistry 39(34):10419-10430 (2000). For tuning transcription levels and altering the transcription start site nucleotides to fit alternative 5' UTRs, the teachings of Davidson et al. *Pac Symp Biocomput* 433-443 (2010) describe T7 promoter variants, and the methods of discovery thereof, that fulfill both of these traits.

In some embodiments, RNA segments may be connected to each other by covalent coupling. In some embodiments, an RNA ligase, e.g., T4 RNA ligase, may be used to connect two or more RNA segments to each other. When a reagent such as an RNA ligase is used, a 5′ terminus is typically linked to a 3′ terminus. In some embodiments, if two segments are

connected, then there are two possible linear constructs that can be formed (i.e., (1) 5′-Segment 1-Segment 2-3′ and (2) 5′-Segment 2-Segment 1-3′). In some embodiments, intramolecular circularization can also occur. Both of these issues can be addressed, for example, by blocking one 5′ terminus or one 3′ terminus so that RNA ligase cannot ligate the terminus to another terminus. In embodiments, if a construct of 5′-Segment 1-Segment 2-3′ is desired, then placing a blocking group on either the 5′ end of Segment 1 or the 3′ end of Segment 2 may result in the formation of only the correct linear ligation product and/or prevent intramolecular circularization. Compositions and methods for the covalent connection of two nucleic acid (e.g., RNA) segments are disclosed, for example, in US20160102322A1 (incorporated herein by reference in its entirety), along with methods including the use of an RNA ligase to directionally ligate two single-stranded RNA segments to each other.

One example of an end blocker that may be used in conjunction with, for example, T4 RNA ligase, is a dideoxy terminator. T4 RNA ligase typically catalyzes the ATP-dependent ligation of phosphodiester bonds between 5′-phosphate and 3′-hydroxyl termini. In some embodiments, when T4 RNA ligase is used, suitable termini must be present on the termini being ligated. One means for blocking T4 RNA ligase on a terminus comprises failing to have the correct terminus format. Generally, termini of RNA segments with a 5-hydroxyl or a 3′-phosphate will not act as substrates for T4 RNA ligase.

Additional exemplary methods that may be used to connect RNA segments is by click chemistry (e.g., as described in U.S. Patent Nos. 7,375,234 and 7,070,941, and US Patent Publication No. 2013/0046084, the entire disclosures of which are incorporated herein by reference). For example, one exemplary click chemistry reaction is between an alkyne group and an azide group (see FIG. 11 of US20160102322A1, which is incorporated herein by reference in its entirety). Any click reaction may potentially be used to link RNA segments (e.g., Cu-azide-alkyne, strain-promoted-azide-alkyne, staudinger ligation, tetrazine ligation, photo-induced tetrazole-alkene, thiol-ene, NHS esters, epoxides, isocyanates, and aldehyde-aminooxy). In some embodiments, ligation of RNA molecules using a click chemistry reaction is advantageous because click chemistry reactions are fast, modular, efficient, often do not produce toxic waste products, can be done with water as a solvent, and/or can be set up to be stereospecific.

In some embodiments, RNA segments may be connected using an Azide-Alkyne Huisgen Cycloaddition. reaction, which is typically a 1,3-dipolar cycloaddition between an azide and a

terminal or internal alkyne to give a 1,2,3-triazole for the ligation of RNA segments. Without wishing to be bound by theory, one advantage of this ligation method may be that this reaction can initiated by the addition of required Cu(I) ions. Other exemplary mechanisms by which RNA segments may be connected include, without limitatoin, the use of halogens (F—, Br—, I—)/alkynes addition reactions, carbonyls/sulfhydryls/maleimide, and carboxyl/amine linkages. For example, one RNA molecule may be modified with thiol at 3′ (using disulfide amidite and universal support or disulfide modified support), and the other RNA molecule may be modified with acrydite at 5′ (using acrylic phosphoramidite), then the two RNA molecules can be connected by a Michael addition reaction. This strategy can also be applied to connecting multiple RNA molecules stepwise. Also provided are methods for linking more than two (e.g., three, four, five, six, etc.) RNA molecules to each other. Without wishing to be bound by theory, this may be useful when a desired RNA molecule is longer than about 40 nucleotides, e.g., such that chemical synthesis efficiency degrades, e.g., as noted in US20160102322A1 (incorporated herein by reference in its entirety).

By way of illustratoin, a tracrRNA is typically around 80 nucleotides in length. Such RNA molecules may be produced, for example, by processes such as in vitro transcription or chemical synthesis. In some embodiments, when chemical synthesis is used to produce such RNA molecules, they may be produced as a single synthesis product or by linking two or more synthesized RNA segments to each other. In embodiments, when three or more RNA segments are connected to each other, different methods may be used to link the individual segments together. Also, the RNA segments may be connected to each other in one pot (e.g., a container, vessel, well, tube, plate, or other receptacle), all at the same time, or in one pot at different times or in different pots at different times. In a non-limiting example, to assemble RNA Segments 1, 2 and 3 in numerical order, RNA Segments 1 and 2 may first be connected, 5′ to 3′, to each other. The reaction product may then be purified for reaction mixture components (e.g., by chromatography), then placed in a second pot, for connection of the 3′ terminus with the 5′ terminus of RNA Segment 3. The final reaction product may then be connected to the 5′ terminus of RNA Segment 3.

In another non-limiting example, RNA Segment 1 (about 30 nucleotides) is the target locus recognition sequence of a crRNA and a portion of Hairpin Region 1. RNA Segment 2 (about 35 nucleotides) contains the remainder of Hairpin Region 1 and some of the linear

tracrRNA between Hairpin Region 1 and Hairpin Region 2. RNA Segment 3 (about 35 nucleotides) contains the remainder of the linear tracrRNA between Hairpin Region 1 and Hairpin Region 2 and all of Hairpin Region 2. In this example, RNA Segments 2 and 3 are linked, 5′ to 3′, using click chemistry. Further, the 5′ and 3′ end termini of the reaction product are both phosphorylated. The reaction product is then contacted with RNA Segment 1, having a 3′ terminal hydroxyl group, and T4 RNA ligase to produce a guide RNA molecule.

A number of additional linking chemistries may be used to connect RNA segments according to method of the invention. Some of these chemistries are set out in Table 6 of US20160102322A1, which is incorporated herein by reference in its entirety.

DNA binding domain:

In certain aspects, the DNA-binding domain of a Gene Writer polypeptide described herein is selected, designed, or constructed for binding to a desired host DNA target sequence. In certain embodiments, the DNA-binding domain of the engineered retrotransposon is a heterologous DNA-binding protein or domain relative to a native retrotransposon sequence. In some embodiments, the heterologous DNA binding element is a zinc-finger element or a TAL effector element, e.g., a zinc-finger or TAL polypeptide or functional fragment thereof. In some embodiments, the heterologous DNA binding element is a sequence-guided DNA binding element, such as Cas9, Cpf1, or other CRISPR-related protein that has been altered to have no endonuclease activity. In some embodiments the heterologous DNA binding element retains endonuclease activity. In some embodiments the heterologous DNA binding element replaces the endonuclease element of the polypeptide. In specific embodiments, the heterologous DNA-binding domain can be any one or more of Cas9 (e.g., Cas9, Cas9 nickase, dCas9), TAL domain, zinc finger (ZF) domain, Myb domain, combinations thereof, or multiples thereof. In certain embodiments, the heterologous DNA-binding domain is a DNA binding domain of a retrotransposon described in Table X, Table Z1, Table Z2, or Table 3A or 3B. In some embodiments, DNA binding domains can be identified based upon homology to other known DNA binding domains using tools as Basic Local Alignment Search Tool (BLAST). In still other embodiments, DNA-binding domains are modified, for example by site-specific mutation, increasing or decreasing DNA-binding elements (for example, number and/or specificity of zinc fingers), etc., to alter DNA-binding specificity and affinity. In some embodiments the DNA

binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells.

In some embodiments, the DNA binding domain comprises a meganuclease domain (e.g., as described herein, e.g., in the endonuclease domain section), or a functional fragment thereof. In some embodiments, the meganuclease domain possesses endonuclease activity, e.g., double-strand cleavage and/or nickase activity. In other embodiments, the meganuclease domain has reduced activity, e.g., lacks endonuclease activity, e.g., the meganuclease is catalytically inactive. In some embodiments, a catalytically inactive meganuclease is used as a DNA binding domain, e.g., as described in Fonfara et al. Nucleic Acids Res 40(2):847-860 (2012), incorporated herein by reference in its entirety. In embodiments, the DNA binding domain comprises one or more modifications relative to a wild-type DNA binding domain, e.g., a modification via directed evolution, e.g., phage-assisted continuous evolution (PACE).

In certain aspects of the present invention, the host DNA-binding site integrated into by the Gene Writer system can be in a gene, in an intron, in an exon, an ORF, outside of a coding region of any gene, in a regulatory region of a gene, or outside of a regulatory region of a gene. In other aspects, the engineered retrotransposon may bind to one or more than one host DNA sequence. In other aspects, the engineered retrotransposon may have low sequence specificity, e.g., bind to multiple sequences or lack sequence preference.

In some embodiments, a Gene Writing system is used to edit a target locus in multiple alleles. In some embodiments, a Gene Writing system is designed to edit a specific allele. For example, a Gene Writing polypeptide may be directed to a specific sequence that is only present on one allele, e.g., comprises a template RNA with homology to a target allele, e.g., a gRNA or annealing domain, but not to a second cognate allele. In some embodiments, a Gene Writing system can alter a haplotype-specific allele. In some embodiments, a Gene Writing system that targets a specific allele preferentially targets that allele, e.g., has at least a 2, 4, 6, 8, or 10-fold preference for a target allele.

In certain embodiments, a Gene Writer™ gene editor system RNA further comprises an intracellular localization sequence, e.g., a nuclear localization sequence. The nuclear localization sequence may be an RNA sequence that promotes the import of the RNA into the nucleus. In certain embodiments the nuclear localization signal is located on the template RNA. In certain embodiments, the retrotransposase polypeptide is encoded on a first RNA, and the template RNA

is a second, separate, RNA, and the nuclear localization signal is located on the template RNA and not on an RNA encoding the retrotransposase polypeptide. While not wishing to be bound by theory, in some embodiments, the RNA encoding the retrotransposase is targeted primarily to the cytoplasm to promote its translation, while the template RNA is targeted primarily to the nucleus to promote its retrotransposition into the genome. In some embodiments the nuclear localization signal is at the 3' end, 5' end, or in an internal region of the template RNA. In some embodiments the nuclear localization signal is 3' of the heterologous sequence (e.g., is directly 3' of the heterologous sequence) or is 5' of the heterologous sequence (e.g., is directly 5' of the heterologous sequence). In some embodiments the nuclear localization signal is placed outside of the 5' UTR or outside of the 3' UTR of the template RNA. In some embodiments the nuclear localization signal is placed between the 5' UTR and the 3' UTR, wherein optionally the nuclear localization signal is not transcribed with the transgene (e.g., the nuclear localization signal is an anti-sense orientation or is downstream of a transcriptional termination signal or polyadenylation signal). In some embodiments the nuclear localization sequence is situated inside of an intron. In some embodiments a plurality of the same or different nuclear localization signals are in the RNA, e.g., in the template RNA. In some embodiments the nuclear localization signal is less than 5, 10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or 1000 bp in legnth. Various RNA nuclear localization sequences can be used. For example, Lubelsky and Ulitsky, *Nature* 555 (107-111), 2018 describe RNA sequences which drive RNA localization into the nucleus. In some embodiments, the nuclear localization signal is a SINE-derived nuclear RNA localization (SIRLOIN) signal. In some embodiments the nuclear localization signal binds a nuclear-enriched protein. In some embodiments the nuclear localization signal binds the HNRNPK protein. In some embodiments the nuclear localization signal is rich in pyrimidines, e.g., is a C/T rich, C/U rich, C rich, T rich, or U rich region. In some embodiments the nuclear localization signal is derived from a long non-coding RNA. In some embodiments the nuclear localization signal is derived from MALAT1 long non-coding RNA or is the 600 nucleotide M region of MALAT1 (described in Miyagawa et al., *RNA* 18, (738-751), 2012). In some embodiments the nuclear localization signal is derived from BORG long non-coding RNA or is a AGCCC motif (described in Zhang et al., *Molecular and Cellular Biology* 34, 2318-2329 (2014). In some embodiments the nuclear localization sequence is described in Shukla et al., *The EMBO*

Journal e98452 (2018). In some embodiments the nuclear localization signal is derived from a non-LTR retrotransposon, an LTR retrotransposon, retrovirus, or an endogenous retrovirus.

In some embodiments, a polypeptide described herein comprises one or more (e.g., 2, 3, 4, 5) nuclear targeting sequences, for example, a nuclear localization sequence (NLS), e.g., as described above. In some embodiments, the NLS is a bipartite NLS. In some embodiments, an NLS facilitates the import of a protein comprising an NLS into the cell nucleus. In some embodiments, the NLS is fused to the N-terminus of a Gene Writer described herein. In some embodiments, the NLS is fused to the C-terminus of the Gene Writer. In some embodiments, the NLS is fused to the N-terminus or the C-terminus of a Cas domain. In some embodiments, a linker sequence is disposed between the NLS and theneighboring domain of the Gene Writer.

In some embodiments, an NLS comprises the amino acid sequence MDSLLMNRRKFLYQFKNVRWAKGRRETYLC, PKKRKVEGADKRTADGSEFESPKKKRKV, RKSGKIAAIWKRPRKPKKKRKV KRTADGSEFESPKKKRKV, KKTELQTTNAENKTKKL, or KRGINDRNFWRGENGRKTR, KRPAATKKAGQAKKKK, or a functional fragment or variant thereof. Exemplary NLS sequences are also described in PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, an NLS comprises an amino acid sequence as disclosed in Table 8. An NLS of this table may be utilized with one or more copies in a polypeptide in one or more locations in a polypeptide, e.g., 1, 2, 3 or more copies of an NLS in an N-terminal domain, between peptide domains, in a C-terminal domain, or in a combination of locations, in order to improve subcellular localization to the nucleus. Multiple unique sequences may be used within a single polypeptide. Sequences may be naturally monopartite or bipartite, e.g., having one or two stretches of basic amino acids, or may be used as chimeric bipartite sequences. Sequence references correspond to UniProt accession numbers, except where indicated as SeqNLS for sequences mined using a subcellular localization prediction algorithm (Lin et al BMC Bioinformat 13:157 (2012), incorporated herein by reference in its entirety).

Table 8. Exemplary nuclear localization signals for use in Gene Writing systems

Sequence	Sequence References
AHFKISGEKRPSTDPGKKAK NPKKKKKKDP	Q76IQ7

AHRAKKMSKTHA	P21827
ASPEYVNLPINGNG	SeqNLS
CTKRPRW	O88622, Q86W56, Q9QYM2, O02776
DKAKRVSRNKSEKKRR	O15516, Q5RAK8, Q91YB2, Q91YB0, Q8QGQ6, O08785, Q9WVS9, Q6YGZ4
EELRLKEELLKGIYA	Q9QY16, Q9UHL0, Q2TBP1, Q9QY15
EEQLRRRKNSRLNNTG	G5EFF5
EVLKVIRTGKRKKKAWKR MVTKVC	SeqNLS
HHHHHHHHHHHHQPH	Q63934, G3V7L5, Q12837
HKKKHPDASVNFSEFSK	P10103, Q4R844, P12682, B0CM99, A9RA84, Q6YKA4, P09429, P63159, Q08IE6, P63158, Q9YH06, B1MTB0
HKRTKK	Q2R2D5
IINGRKLKLKKSRRRSSQTS NNSFTSRRS	SeqNLS
KAEQERRK	Q8LH59
KEKRKRREELFIEQKKRK	SeqNLS
KKGKDEWFSRGKKP	P30999
KKGPSVQKRKKT	Q6ZN17
KKKTVINDLLHYKKEK	SeqNLS, P32354
KKNGGKGKNKPSAKIKK	SeqNLS
KKPKWDDFKKKKK	Q15397, Q8BKS9, Q562C7
KKRKKD	SeqNLS, Q91Z62, Q1A730, Q969P5, Q2KHT6, Q9CPU7
KKRRKRRRK	SeqNLS
KKRRRRARK	Q9UMS6, D4A702, Q91YE8
KKSKRGR	Q9UBS0
KKSRKRGS	B4FG96
KKSTALSRELGKIMRRR	SeqNLS, P32354
KKSYQDPEIIAHSRPRK	Q9U7C9
KKTGKNRKLKSKRVKTR	Q9Z301, O54943, Q8K3T2
KKVSIAGQSGKLWRWKR	Q6YUL8
KKYENVVIKRSPRKRGRPR K	SeqNLS
KNKKRK	SeqNLS
KPKKKR	SeqNLS
KRAMKDDSHGNSTSPKRRK	Q0E671
KRANSNLVAAYEKAKKK	P23508
KRASEDTTSGSPPKKSSAGP KR	Q9BZZ5, Q5R644
KRFKRRWMVRKMKTKK	SeqNLS
KRGLNSSFETSPKKVK	Q8IV63

KRGNSSIGPNDLSKRKQRKK	SeqNLS
KRIHSVSLSQSQIDPSKKVKRAK	SeqNLS
KRKGKLKNKGSKRKK	O15381
KRRRRRREKRKR	Q96GM8
KRSNDRTYSPEEEKQRRA	Q91ZF2
KRTVATNGDASGAHRAKKMSK	SeqNLS
KRVYNKGEDEQEHLPKGKKR	SeqNLS
KSGKAPRRRAVSMDNSNK	Q9WVH4, O43524
KVNFLDMSLDDIIIYKELE	Q9P127
KVQHRIAKKTTRRRR	Q9DXE6
LSPSLSPL	Q9Y261, P32182, P35583
MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	Q9GZX7
MPQNEYIELHRKRYGYRLDYHEKKRKKESREAHERSKKAKKMIGLKAKLYHK	SeqNLS
MVQLRPRASR	SeqNLS
NNKLLAKRRKGGASPKDDPMDDIK	Q965G5
NYKRPMDGTYGPPAKRHEGE	O14497, A2BH40
PDTKRAKLDSSETTMVKKK	SeqNLS
PEKRTKI	SeqNLS
PGGRGKKK	Q719N1, Q9UBP0, A2VDN5
PGKMDKGEHRQERRDRPY	Q01844, Q61545
PKKGDKYDKTD	Q45FA5
PKKKSRK	O35914, Q01954
PKKNKPE	Q22663
PKKRAKV	P04295, P89438
PKPKKLKVE	P55263, P55262, P55264, Q64640
PKRGRGR	Q9FYS5, Q43386
PKRRLVDDA	P0C797
PKRRRTY	SeqNLS
PLFKRR	A8X6H4, Q9TXJ0
PLRKAKR	Q86WB0, Q5R8V9
PPAKRKCIF	Q6AZ28, O75928, Q8C5D8
PPARRRRL	Q8NAG6
PPKKKRKV	Q3L6L5, P03070, P14999, P03071
PPNKRMKVKH	Q8BN78

PPRIYPQLPSAPT	P0C799
PQRSPFPKSSVKR	SeqNLS
PRPRKVPR	P0C799
PRRRVQRKR	SeqNLS, Q5R448, Q5TAQ9
PRRVRLK	Q58DJ0, P56477, Q13568
PSRKRPR	Q62315, Q5F363, Q92833
PSSKKRKV	SeqNLS
PTKKRVK	P07664
QRPGPYDRP	SeqNLS
RGKGGKGLGKGGAKRHRK	SeqNLS
RKAGKGGGGHKTTKKRSAKDEKVP	B4FG96
RKIKLKRAK	A1L3G9
RKIKRKRAK	B9X187
RKKEAPGPREELRSRGR	O35126, P54258, Q5IS70, P54259
RKKRKGK	SeqNLS, Q29243, Q62165, Q28685, O18738, Q9TSZ6, Q14118
RKKRRQRRR	P04326, P69697, P69698, P05907, P20879, P04613, P19553, P0C1J9, P20893, P12506, P04612, Q73370, P0C1K0, P05906, P35965, P04609, P04610, P04614, P04608, P05905
RKKSIPLSIKNLKRKHKRKKNKITR	Q9C0C9
RKLVKPKNTKMKTKLRTNPY	Q14190
RKRLILSDKGQLDWKK	SeqNLS, Q91Z62, Q1A730, Q2KHT6, Q9CPU7
RKRLKSK	Q13309
RKRRVRDNM	Q8QPH4, Q809M7, A8C8X1, Q2VNC5, Q38SQ0, O89749, Q6DNQ9, Q809L9, Q0A429, Q20NV3, P16509, P16505, Q6DNQ5, P16506, Q6XT06, P26118, Q2ICQ2, Q2RCG8, Q0A2D0, Q0A2H9, Q9IQ46, Q809M3, Q6J847, Q6J856, B4URE4, A4GCM7, Q0A440, P26120, P16511,
RKRSPKDKKEKDLDGAGKRRKT	Q7RTP6
RKRTPRVDGQTGENDMNKRRRK	O94851
RLPVRRRRRR	P04499, P12541, P03269, P48313, P03270
RLRFRKPKSK	P69469
RQQRKR	Q14980
RRDLNSSFETSPKKVK	Q8K3G5

RRDRAKLR	Q9SLB8
RRGDGRRR	Q80WE1, Q5R9B4, Q06787, P35922
RRGRKRKAEKQ	Q812D1, Q5XXA9, Q99JF8, Q8MJG1, Q66T72, O75475
RRKKRR	Q0VD86, Q58DS6, Q5R6G2, Q9ERI5, Q6AYK2, Q6NYC1
RRKRSKSEDMDSVESKRRR	Q7TT18
RRKRSR	Q99PU7, D3ZHS6, Q92560, A2VDM8
RRPKGKTLQKRKPK	Q6ZN17
RRRGFERFGPDNMGRKRK	Q63014, Q9DBR0
RRRGKNKVAAQNCRK	SeqNLS
RRRKRR	Q5FVH8, Q6MZT1, Q08DH5, Q8BQP9
RRRQKQKGGASRRR	SeqNLS
RRRREGPRARRRR	P08313, P10231
RRTIRLKLVYDKCDRSCKIQKKNRNKCQYCRFHKCLSVGMSHNAIRFGRMPRSEKAKLKAE	SeqNLS
RRVPQRKEVSRCRKCRK	Q5RJN4, Q32L09, Q8CAK3, Q9NUL5
RVGGRRQAVECIEDLLNEPGQPLDLSCKRPRP	P03255
RVVKLRIAP	P52639, Q8JMN0
RVVRRR	P70278
SKRKTKISRKTR	Q5RAY1, O00443
SYVKTVPNRTRTYIKL	P21935
TGKNEAKKRKIA	P52739, Q8K3J5, Q5RAU9
TLSPASSPSSVSCPVIPASTDESPGSALNI	SeqNLS
VSKKQRTGKKIH	P52739, Q8K3J5, Q5RAU9
SPKKKRKVE	
KRTAD GSEFE SPKKKRKVE	
PAAKRVKLD	
PKKKRKV	
MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	
SPKKKRKVEAS	
MAPKKKRKVGIHRGVP	

In some embodiments, the NLS is a bipartite NLS. A bipartite NLS typically comprises two basic amino acid clusters separated by a spacer sequence (which may be, e.g., about 10

amino acids in length). A monopartite NLS typically lacks a spacer. An example of a bipartite NLS is the nucleoplasmin NLS, having the sequence KR[PAATKKAGQA]KKKK, wherein the spacer is bracketed. Another exemplary bipartite NLS has the sequence PKKKRKVEGADKRTADGSEFESPKKKRKV. Exemplary NLSs are described in International Application WO2020051561, which is herein incorporated by reference in its entirety, including for its disclosures regarding nuclear localization sequences.

In certain embodiments, a Gene Writer™ gene editor system polypeptide further comprises an intracellular localization sequence, e.g., a nuclear localization sequence and/or a nucleolar localization sequence. The nuclear localization sequence and/or nucleolar localization sequence may be amino acid sequences that promote the import of the protein into the nucleus and/or nucleolus, where it can promote integration of heterologous sequyence into the genome. In certain embodiments, a Gene Writer gene editor system polypeptide (e.g., a retrotransposase, e.g., a polypeptide according to any of Tables 1, 2, 3A, or 3B herein) further comprises a nucleolar localization sequence. In certain embodiments, the retrotransposase polypeptide is encoded on a first RNA, and the template RNA is a second, separate, RNA, and the nucleolar localization signal is encoded on the RNA encoding the retrotransposase polypeptide and not on the template RNA. In some embodiments, the nucleolar localization signal is located at the N-terminus, C-terminus, or in an internal region of the polypeptide. In some embodiments, a plurality of the same or different nucleolar localization signals are used. In some embodiments, the nuclear localization signal is less than 5, 10, 25, 50, 75, or 100 amino acids in length. Various polypeptide nucleolar localization signals can be used. For example, Yang et al., *Journal of Biomedical Science* 22, 33 (2015), describe a nuclear localization signal that also functions as a nucleolar localization signal. In some embodiments, the nucleolar localization signal may also be a nuclear localization signal. In some embodiments, the nucleolar localization signal may overlap with a nuclear localization signal. In some embodiments, the nucleolar localization signal may comprise a stretch of basic residues. In some embodiments, the nucleolar localization signal may be rich in arginine and lysine residues. In some embodiments, the nucleolar localization signal may be derived from a protein that is enriched in the nucleolus. In some embodiments, the nucleolar localization signal may be derived from a protein enriched at ribosomal RNA loci. In some embodiments, the nucleolar localization signal may be derived from a protein that binds rRNA. In some embodiments, the nucleolar localization signal may be derived from MSP58. In

some embodiments, the nucleolar localization signal may be a monopartite motif. In some embodiments, the nucleolar localization signal may be a bipartite motif. In some embodiments, the nucleolar localization signal may consist of a multiple monopartite or bipartite motifs. In some embodiments, the nucleolar localization signal may consist of a mix of monopartite and bipartite motifs. In some embodiments, the nucleolar localization signal may be a dual bipartite motif. In some embodiments, the nucleolar localization motif may be a KRASSQALGTIPKRRSSSRFIKRKK (SEQ ID NO: 1530). In some embodiments, the nucleolar localization signal may be derived from nuclear factor-κB-inducing kinase. In some embodiments, the nucleolar localization signal may be an RKKRKKK motif (SEQ ID NO: 1531) (described in Birbach et al., Journal of Cell Science, 117 (3615-3624), 2004).

In some embodiments, a nucleic acid described herein (e.g., an RNA encoding a GeneWriter polypeptide, or a DNA encoding the RNA) comprises a microRNA binding site. In some embodiments, the microRNA binding site is used to increase the target-cell specificity of a GeneWriter system. For instance, the microRNA binding site can be chosen on the basis that is is recognized by a miRNA that is present in a non-target cell type, but that is not present (or is present at a reduced level relative to the non-target cell) in a target cell type. Thus, when the RNA encoding the GeneWriter polypeptide is present in a non-target cell, it would be bound by the miRNA, and when the RNA encoding the GeneWriter polypeptide is present in a target cell, it would not be bound by the miRNA (or bound but at reduced levels relative to the non-target cell). While not wishing to be bound by theory, binding of the miRNA to the RNA encoding the GeneWriter polypeptide may reduce production of the GeneWriter polypeptide, e.g., by degrading the mRNA encoding the polypeptide or by interfering with translation. Accordingly, the heterologous object sequence would be inserted into the genome of target cells more efficiently than into the genome of non-target cells. A system having a microRNA binding site in the RNA encoding the GeneWriter polypeptide (or encoded in the DNA encoding the RNA) may also be used in combination with a template RNA that is regulated by a second microRNA binding site, e.g., as described herein in the section entitled "Template RNA component of Gene Writer™ gene editor system."

Based on the Accession numbers provided in Table X, Z1, Z2, 3A, or 3B, the nucleic acid and corresponding polypeptide sequences of each retrotransposon and domains thereof can be determined, e.g., by using routine sequence analysis tools as Basic Local Alignment Search

Tool (BLAST) or CD-Search for conserved domain analysis. Other sequence analysis tools are known and can be found, e.g., at https://molbiol-tools.ca, for example, at https://molbiol-tools.ca/Motifs.htm. SEQ ID NOs 113-1015 align with the first 903 rows of Table 2.

Tables X, 3A, and 3B herein provide the sequences of exemplary transposons, including the amino acid sequence of the retrotransposase, and sequences of 5' and 3' untranslated regions to allow the retrotransposase to bind the template RNA. In some embodiments, a 5' UTR of any of Table X, Z1, Z2, 3A, or 3B allows the retrotransposase to bind the template RNA. In some embodiments, a 3' UTR of any of Tables X, 3A, and 3B allows the retrotransposase to bind the template RNA. Thus, in some embodiments, a polypeptide for use in any of the systems described herein can be a polypeptide of any of Tables X, 3A, and 3B herein, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the system further comprises one or both of a 5' or 3' untranslated region of any of Tables X, 3A, and 3B herein (or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto), e.g., from the same retrotransposon as the polypeptide referred to in the preceding sentence, as indicated in the same row of the same table. In some embodiments, the system comprises one or both of a 5' or 3' untranslated region of any of Tables X, 3A, and 3B herein, e.g., a segment of the full retrotransposon sequence that encodes an RNA that is capable of binding a retrotransposase, and/or the sub-sequence provided in the column entitled Predicted 5' UTR or Predicted 3' UTR.

In some embodiments, a polypeptide for use in any of the systems described herein can be a molecular reconstruction or ancestral reconstruction based upon the aligned polypeptide sequence of multiple retrotransposons. In some embodiments, a 5' or 3' untranslated region for use in any of the systems described herein can be a molecular reconstruction based upon the aligned 5' or 3' untranslated region of multiple retrotransposons. Based on the Accession numbers provided herein, polypeptides or nucleic acid sequences can be aligned, e.g., by using routine sequence analysis tools as Basic Local Alignment Search Tool (BLAST) or CD-Search for conserved domain analysis. Molecular reconstructions can be created based upon sequence consensus, e.g. using approaches described in *Ivics et al., Cell 1997, 501 – 510 ; Wagstaff et al., Molecular Biology and Evolution 2013, 88-99.* In some embodiments, the retrotransposon from which the 5' or 3' untranslated region or polypeptide is derived is a young or a recently active

mobile element, as assessed via phylogenetic methods such as those described in Boissinot et al., Molecular Biology and Evolution 2000, 915-928.

Table Z1: Exemplary reverse transcriptase domains from different types of sources.

Sources include Group II intron, non-LTR retrotransposon, retrovirus, LTR retrotransposon, diversity-generating retroelement, retron, telomerase, retroplasmid, and evolved DNA polymerase. Also included are the associated RT signatures from the InterPro, pfam, and cd databases. Although the evolved polymerase RTX can perform RNA-dependent DNA polymerization, no RT signatures were identified by InterProScan, so polymerase signatures are included instead.

Protein	Type	Accession	UniProt	Sequence	RT signatures
Marathon RT	Group II intron	CBK92290.1	D4JMT6	MDTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKEHLAKNGETIKGQLRTRKYKPQPARRVEIPKPDGGVRNLGVPTVTDRFIQQAIAQVLTPIYEEQFHDHSYGFRPNRCAQQAILTALNIMNDGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDVISIVRKYLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDKEMEKRGLNFVRYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKVNMTKSKVDRPSGLKYLGFGFYFDPRAHQFKAKPHAKSVAKFKKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGSMKTLCKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARRVAYTGKRIAYVCNKGAVNVAISNKRLASFGLISMLDYYIEKCVTC	IPR000477, PF00078, cd01651
TGIRT, trt	Group II intron	AAT72329.1	Q6DKY2	MALLERILADRNLITALKRVEANQGAPGIGDVSTDQLRDIYRAHWSTIRAQLLAGTYRPAPVRRVGIPKGPGGTRQLGITPVVDRLIQQIALQELTPIFDPDFSPSSFGFRPGRNAHDAVRQAQGYIQEYGRYVVDMDLKEFFDRVNHDLIMSRVARKVDKKRVLKLIRYALQAGVMIEGVKVQTEEGTQPGGPLSPLLANILLDDLDKELEKRGLKFCYRADDCNI	IPR000477, PF00078, cd01651

				YVSKLRAGQRVKQSIQRFLEKTL KLKVNEEKSVADRPWKRAFGLF SFTPERKARIRLAPRSIQRLKQRI RQLTNPNWSISMPREIHRVNQYV GMWIGYFRLVTEPSVLQTIEGWI RRRLRLCWQLQWKRVRTRIREL RALGLKETAVMEIANRTKGAWR TTKPQTLHQALGKYTWTAQGLK TSLQRYFELRQG	
LtrA	Group II intron	AAB06503.1	P0A3U0	MKPTMAILERISKNSQENIDEVFT RLYRYLLRPDIYYVAYQNLYSN KGASTKGILDDTADGFSEEKIKKI IQSLKDGTYYPQPVRRMYIAKKN SKKMRPLGIPTFTDKLIQEAVRIIL ESIYEPVFEDVSHGFRPQRSCHTA LKTIKREFGGARWFVEGDIKGCF DNIDHVTLIGLINLKIKDMKMSQ LIYKFLKAGYLENWQYHKTYSG TPQGGILSPLLANIYLHELDKFVL QLKMKFDRESPERITPEYRELHN EIKRISHRLKKLEGEEKAKVLLE YQEKRKRLPTLPCTSQTNKVLKY VRYADDFIISVKGSKEDCQWIKE QLKLFIHNKLKMELSEEKTLITHS SQPARFLGYDIRVRRSGTIKRSG KVKKRTLNGSVELLIPLQDKIRQ FIFDKKIAIQKKDSSWFPVHRKYL IRSTDLEIITIYNSELRGICNYYGL ASNFNQLNYFAYLMEYSCLKTIA SKHKGTLSKTISMFKDGSGSWGI PYEIKQGKQRRYFANFSECKSPY QFTDEISQAPVLYGYARNTLENR LKAKCCELCGTSDENTSYEIHHV NKVKNLKGKEKWEMAMIAKQR KTLVVCFHCHRHVIHKHK	IPR000477, PF00078, cd01651
R2Bm	non-LTR retrotransposon	AAB59214.1	V9H052	MMASTALSLMGRCNPDGCTRGK HVTAAPMDGPRGPSSLAGTFGW GLAIPAGEPCGRVCSPATVGFFP VAKKSNKENRPEASGLPLESERT GDNPTVRGSAGADPVGQDAPG WTCQFCERTFSTNRGLGVHKRR AHPVETNTDAAPMMVKRRWHG EEIDLLARTEARLLAERGQCSGG DLFGALPGFGRTLEAIKGQRRRE PYRALVQAHLARFGSQPGPSSGG CSAEPDFRRASGAEEAGEERCAE	IPR000477, PF00078, cd01650

				DAAAYDPSAVGQMSPDAARVLS ELLEGAGRRRACRAMRPKTAGR RNDLHDDRTASAHKTSRQKRRA EYARVQELYKKCRSRAAAEVID GACGGVGHSLEEMETYWRPILE RVSDAPGPTPEALHALGRAEWH GGNRDYTQLWKPISVEEIKASRF DWRTSPGPDGIRSGQWRAVPVH LKAEMFNAWMARGEIPEILRQC RTVFVPKVERPGGPGEYRPISIAS IPLRHFHSILARRLLACCPPDARQ RGFICADGTLENSAVLDAVLGDS RKKLRECHVAVLDFAKAFDTVS HEALVELLRLRGMPEQFCGYIAH LYDTASTTLAVNNEMSSPVKVG RGVRQGDPLSPILFNVVMDLILA SLPERVGYRLEMELVSALAYAD DLVLLAGSKVGMQESISAVDCV GRQMGLRLNCRKSAVLSMIPDG HRKKHHYLTERTFNIGGKPLRQV SCVERWRYLGVDFEASGCVTLE HSISSALNNISRAPLKPQQRLEILR AHLIPRFQHGFVLGNISDDRLRM LDVQIRKAVGQWLRLPADVPKA YYHAAVQDGGLAIPSVRATIPDL IVRRFGGLDSSPWSVARAAAKSD KIRKKLRWAWKQLRRFSRVDST TQRPSVRLFWREHLHASVDGRE LRESTRTPTSTKWIRERCAQITGR DFVQFVHTHINALPSRIRGSRGR RGGGESSLTCRAGCKVRETTAHI LQQCHRTHGGRILRHNKIVSFVA KAMEENKWTVELEPRLRTSVGL RKPDIIASRDGVGVIVDVQVVSG QRSLDELHREKRNKYGNHGELV ELVAGRLGLPKAECVRATSCTIS WRGVWSLTSYKELRSIIGLREPT LQIVPILALRGSHMNWTRFNQMT SVMGGGVG	
LINE-1	non-LTR retrotransposon	AAC51271.1	O00370	MTGSNSHITILTLNVNGLNSPIKR HRLASWIKSQDPSVCCIQETHLT CRDTHRLKIKGWRKIYQANGKQ KKAGVAILVSDKTDFKPTKIKRD KEGHYIMVKGSIQQEELTILNIYA PNTGAPRFIKQVLSDLQRDLDSH TLIMGDFNTPLSILDRSTRQKVN	IPR000477, PF00078, cd01650

				KDTQELNSALHQTDLIDIYRTLH PKSTEYTFFSAPHHTYSKIDHIVG SKALLSKCKRTEIITNYLSDHSAI KLELRIKNLTQSRSTTWKLNNLL LNDYWVHNEMKAEIKMFFETNE NKDTTYQNLWDAFKAVCRGKFI ALNAYKRKQERSKIDTLTSQLKE LEKQEQTHSKASRRQEITKIRAEL KEIETQKTLQKINESRSWFFERIN KIDRPLARLIKKKREKNQIDTIKN DKGDITTDPTEIQTTIREYYKHLY ANKLENLEEMDTFLDTYTLPRLN QEEVESLNRPITGSEIVAIINSLPT KKSPGPDGFTAEFYQRYKEELVP FLLKLFQSIEKEGILPNSFYEASIIL IPKPGRDTTKKENFRPISLMNIDA KILNKILANRIQQHIKKLIHHDQV GFIPGMQGWFNIRKSINVIQHINR AKDKNHVIISIDAEKAFDKIQQPF MLKTLNKLGIDGMYLKIIRAIYD KPTANIILNGQKLEAFPLKTGTRQ GCPLSPLLFNIVLEVLARAIRQEK EIKGIQLGKEEVKLSLFADDMIV YLENPIVSAQNLLKLISNFSKVSG YKINVQKSQAFLYNNNRQTESQI MGELPFTIASKRIKYLGIQLTRDV KDLFKENYKPLLKEIKEDTNKW KNIPCSWVGRINIVKMAILPKVIY RFNAIPIKLPMTFFTELEKTTLKFI WNQKRARIAKSILSQKNKAGGIT LPDFKLYYKATVTKTAWYWYQ NRDIDQWNRTEPSEIMPHIYNYLI FDKPEKNKQWGKDSLLNKWCW ENWLAICRKLKLDPFLTPYTKINS RWIKDLNVKPKTIKTLEENLGITI QDIGVGKDFMSKTPKAMATKDK IDKWDLIKLKSFCTAKETTIRVNR QPTTWEKIFATYSSDKGLISRIYN ELKQIYKKKTNNPIKKWAKDMN RHFSKEDIYAAKKHMKKCSSSLA IREMQIKTTMRYHLTPVRMAIIK KSGNNRCWRGCGEIGTLVHCW WDCKLVQPLWKSVWRFLRDLEL EIPFDPAIPLLGIYPKDYKSCCYK DTCTRMFIAALFTIAKTWNQPNC PTMIDWIKKMWHIYTMEYYAAI	

				KNDEFISFVGTWMKLETIILSKLS QEQKTKHRIFSLIGGN	
Penelope	non-LTR retrotransposon	AAL14979.1	Q95VB5	MERSPEPSININGRHAVCTATNM SYAKIKTKYKDSKRTINKFQLTL VKLTKLKSSLKFLLKCRKSNLIPN FIKNLTQHLTILTTDNKTHPDITR TLTRHTHFYHTKILNLLIKHKHN LLQEQTKHMQKAKTNIEQLMTT DDAKAFFESERNIENKITTTLKKR QETKHDKLRDQRNLALADNNTQ REWFVNKTKIEFPPNVVALLAKG PKFALPISKRDFPLLKYIADGEEL VQTIKEKETQESARTKFSLLVKE HKTKNNQNSRDRAILDTVEQTR KLLKENINIKILSSDKGNKTVAM DEDEYKNKMTNILDDLCAYRTL RLDPTSRLQTKNNTFVAQLFKM GLISKDERNKMTTTTAVPPRIYG LPKIHKEGTPLRPICSSIGSPSYGL CKYIIQILKNLTMDSRYNIKNAV DFKDRVNNSQIREEETLVSFDVV SLFPSIPIELALDTIRQKWTKLEEH TNIPKQLFMDIVRFCIEENRYFKY EDKIYTQLKGMPMGSPASPVIAD ILMEELLDKITDKLKIKPRLLTKY VDDLFAITNKIDVENILKELNSFH KQIKFTMELEKDGKLPFLDSIVSR MDNTLKIKWYRKPIASGRILNFN SNHPKSMIINTALGCMNRMMKIS DTIYHKEIEHEIKELLTKNDFPPNI IKTLLKRRQIERKKPTEPAKIYKS LIYVPRLSERLTNSDCYNKQDIK VAHKPTNTLQKFFNKIKSKIPMIE KSNVVYQIPCGGDNNNKCNSVYI GTTKSKLKTRISQHKSDFKLRHQ NNIQKTALMTHCIRSNHTPNFDE TTILQQEQHYNKRHTLEMLHIIN TPTYKRLNYKTDTENCAHLYRH LLNSQTTSVTISTSKSADV	IPR000477, PF00078, cd00304
M-MLV RT	Retrovirus	ADS42990.1	P03355[660-1330]	TLNIEDEHRLHETSKEPDVSLGST WLSDFPQAWAETGGMGLAVRQ APLIIPLKATSTPVSIKQYPMSQE ARLGIKPHIQRLLDQGILVPCQSP WNTPLLPVKKPGTNDYRPVQDL REVNKRVEDIHPTVPNPYNLLSG LPPSHQWYTVLDLKDAFFCLRLH	IPR000477, PF00078, cd03715

				PTSQPLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFDEALHRDL ADFRIQHPDLILLQYVDDLLLAA TSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLK EGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAE MAAPLYPLTKTGTLFNWGPDQQ KAYQEIKQALLTAPALGLPDLTK PFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPP CLRMVAAIAVLTKDAGKLTMGQ PLVILAPHAVEALVKQPPDRWLS NARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHT WYTDGSSLLQEGQRKAGAAVTT ETEVIWAKALPAGTSAQRAELIA LTQALKMAEGKKLNVYTDSRYA FATAHIHGEIYRRRGLLTSEGKEI KNKDEILALLKALFLPKRLSIIHC PGHQKGHSAEARGNRMADQAA RKAAITETPDTSTLL	
RSV RT	Retro virus	AAC82561.1	P03354[709-1567]	TVALHLAIPLKWKPDHTPVWIDQ WPLPEGKLVALTQLVEKELQLG HIEPSLSCWNTPVFVIRKASGSYR LLHDLRAVNAKLVPFGAVQQGA PVLSALPRGWPLMVLDLKDCFFS IPLAEQDREAFAFTLPSVNNQAP ARRFQWKVLPQGMTCSPTICQL VVGQVLEPLRLKHPSLCMLHYM DDLLLAASSHDGLEAAGEEVIST LERAGFTISPDKVQREPGVQYLG YKLGSTYVAPVGLVAEPRIATLW DVQKLVGSLQWLRPALGIPPRL MGPFYEQLRGSDPNEAREWNLD MKMAWREIVRLSTTAALERWDP ALPLEGAVARCEQGAIGVLGQG LSTHPRPCLWLFSTQPTKAFTAW LEVLTLLITKLRASAVRTFGKEV DILLLPACFREDLPLPEGILLALK GFAGKIRSSDTPSIFDIARPLHVSL KVRVTDHPVPGPTVFTDASSSTH KGVVVWREGPRWEIKEIADLGA SVQQLEARAVAMALLLWPTTPT NVVTDSAFVAKMLLKMGQEGV	IPR000477, PF00078, cd01645

				PSTAAAFILEDALSQRSAMAAVL HVRSHSEVPGFFTEGNDVADSQA TFQAYPLREAKDLHTALHIGPRA LSKACNISMQQAREVVQTCPHC NSAPALEAGVNPRGLGPLQIWQT DFTLEPRMAPRSWLAVTVDTASS AIVVTQHGRVTSVAVQHHWATA IAVLGRPKAIKTDNGSCFTSKSTR EWLARWGIAHTTGIPGNSQGQA MVERANRLLKDRIRVLAEGDGF MKRIPTSKQGELLAKAMYALNH FERGENTKTPIQKHWRPTVLTEG PPVKIRIETGEWEKGWNVLVWG RGYAAVKNRDTDKVIWVPSRKV KPDITQKDEVTKKDEASPLFAG	
AMV RT	Retro virus	HW606680.1	-	TVALHLAIPLKWKPNHTPVWIDQ WPLPEGKLVALTQLVEKELQLG HIEPSLSCWNTPVFVIRKASGSYR LLHDLRAVNAKLVPFGAVQQGA PVLSALPRGWPLMVLDLKDCFFS IPLAEQDREAFAFTLPSVNNQAP ARRFQWKVLPQGMTCSPTICQLI VGQILEPLRLKHPSLRMLHYMD DLLLAASSHDGLEAAGEEVISTL ERAGFTISPDKVQREPGVQYLGY KLGSTYVAPVGLVAEPRIATLWD VQKLVGSLQWLRPALGIPPRLM GPFYEQLRGSDPNEAREWNLDM KMAWREIVQLSTTAALERWDPA LPLEGAVARCEQGAIGVLGQGLS THPRPCLWLFSTQPTKAFTAWLE VLTLLITKLRASAVRTFGKEVDIL LLPACFREDLPLPEGILLALRGFA GKIRSSDTPSIFDIARPLHVSLKV RVTDHPVPGPTVFTDASSSTHKG VVVWREGPRWEIKEIADLGASV QQLEARAVAMALLLWPTTPTNV VTDSAFVAKMLLKMGQEGVPST AAAFILEDALSQRSAMAAVLHV RSHSEVPGFFTEGNDVADSQATF QAY	IPR000477, PF00078, cd01645
HIV RT	Retro virus	AAB50259.1	P04585[588-1147]	PISPIETVPVKLKPGMDGPKVKQ WPLTEEKIKALVEICTEMEKEGKI SKIGPENPYNTPVFAIKKKDSTK WRKLVDFRELNKRTQDFWEVQL GIPHPAGLKKKKSVTVLDVGDA	IPR000477, PF00078, cd01645

				YFSVPLDEDFRKYTAFTIPSINNE TPGIRYQYNVLPQGWKGSPAIFQ SSMTKILEPFRKQNPDIVIYQYM DDLYVGSDLEIGQHRTKIEELRQ HLLRWGLTTPDKKHQKEPPFLW MGYELHPDKWTVQPIVLPEKDS WTVNDIQKLVGKLNWASQIYPGI KVRQLCKLLRGTKALTEVIPLTE EAELELAENREILKEPVHGVYYD PSKDLIAEIQKQGQGQWTYQIYQ EPFKNLKTGKYARMRGAHTNDV KQLTEAVQKITTESIVIWGKTPKF KLPIQKETWETWWTEYWQATWI PEWEFVNTPPLVKLWYQLEKEPI VGAETFYVDGAANRETKLGKAG YVTNRGRQKVVTLTDTTNQKTE LQAIYLALQDSGLEVNIVTDSQY ALGIIQAQPDQSESELVNQIIEQLI KKEKVYLAWVPAHKGIGGNEQV DKLVSAGIRKVL	
Ty1	LTR retrotransposon	AAA66938.1	Q07163-1[1218-1755]	AVKAVKSIKPIRTTLRYDEAITYN KDIKEKEKYIEAYHKEVNQLLK MKTWDTDEYYDRKEIDPKRVIN SMFIFNKKRDGTHKARFVARGDI QHPDTYDSGMQSNTVHHYALM TSLSLALDNNYYITQLDISSAYLY ADIKEELYIRPPPHLGMNDKLIRL KKSLYGLKQSGANWYETIKSYLI QQCGMEEVRGWSCVFKNSQVTI CLFVDDMVLFSKNLNSNKRIIEK LKMQYDTKIINLGESDEEIQYDIL GLEIKYQRGKYMKLGMENSLTE KIPKLNVPLNPKGRKLSAPGQPG LYIDQDELEIDEDEYKEKVHEMQ KLIGLASYVGYKFRFDLLYYINT LAQHILFPSRQVLDMTYELIQFM WDTRDKQLIWHKNKPTEPDNKL VAISDASYGNQPYYKSQIGNIYL LNGKVIGGKSTKASLTCTSTTEA EIHAISESVPLLNNLSYLIQELNK KPIIKGLLTDSRSTISIIKSTNEEKF RNRFFGTKAMRLRDEVSGNNLY VYYIETKKNIADVMTKPLPIKTF KLLTNKWIH	IPR013103, PF07727

Brt	Diversity-generating retroelement	NP_958675.1	Q775D8	MGKRHRNLIDQITTWENLLDAYRKTSHGKRRTWGYLEFKEYDLANLLALQAELKAGNYERGPYREFLVYEPKPRLISALEFKDRLVQHALCNIVAPIFEAGLLPYTYACRPDKGTHAGVCHVQAELRRTRATHFLKSDFSKFFPSIDRAALYAMIDKKIHCAATRRLLRVVLPDEGVGIPIGSLTSQLFANVYGGAVDRLLHDELKQRHWARYMDDIVVLGDDPEELRAVFYRLRDFASERLGLKISHWQVAPVSRGINFLGYRIWPTHKLLRKSSVKRAKRKVANFIKHGEDESLQRFLASWSGHAQWADTHNLFTWMEEQYGIACH	IPR000477, PF00078, cd01646
RT86	Retron	AAA61471.1	P23070	MKSAEYLNTFRLRNLGLPVMNNLHDMSKATRISVETLRLLIYTADFRYRIYTVEKKGPEKRMRTIYQPSRELKALQGWVLRNILDKLSSSPFSIGFEKHQSILNNATPHIGANFILNIDLEDFFPSLTANKVFGVFHSLGYNRLISSVLTKICCYKNLLPQGAPSSPKLANLICSKLDYRIQGYAGSRGLIYTRYADDLTLSAQSMKKVVKARDFLFSIIPSEGLVINSKKTCISGPRSQRKVTGLVISQEKVGIGREKYKEIRAKIHHIFCGKSSEIEHVRGWLSFILSVDSKSHRRLITYISKLEKKYGKNPLNKAKT	IPR000477, PF00078, cd03487
TERT	Telomerase	AAG23289.1	O14746	MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDPAAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFVLVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPAPGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPARPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQR	IPR000477, PF00078, cd01648

				YWQMRPLFLELLGNHAQCPYGV LLKTHCPLRAAVTPAAGVCARE KPQGSVAAPEEEDTDPRRLVQLL RQHSSPWQVYGFVRACLRRLVP PGLWGSRHNERRFLRNTKKFISL GKHAKLSLQELTWKMSVRDCA WLRRSPGVGCVPAAEHRLREEIL AKFLHWLMSVYVVELLRSFFYV TETTFQKNRLFFYRKSVWSKLQS IGIRQHLKRVQLRELSEAEVRQH REARPALLTSRLRFIPKPDGLRPI VNMDYVVGARTFRREKRAERLT SRVKALFSVLNYERARRPGLLGA SVLGLDDIHRAWRTFVLRVRAQ DPPPELYFVKVDVTGAYDTIPQD RLTEVIASIIKPQNTYCVRRYAVV QKAAHGHVRKAFKSHVSTLTDL QPYMRQFVAHLQETSPLRDAVVI EQSSSLNEASSGLFDVFLRFMCH HAVRIRGKSYVQCQGIPQGSILST LLCSLCYGDMENKLFAGIRRDGL LLRLVDDFLLVTPHLTHAKTFLR TLVRGVPEYGCVVNLRKTVVNF PVEDEALGGTAFVQMPAHGLFP WCGLLLDTRTLEVQSDYSSYAR TSIRASLTFNRGFKAGRNMRRKL FGVLRLKCHSLFLDLQVNSLQTV CTNIYKILLLQAYRFHACVLQLPF HQQVWKNPTFFLRVISDTASLCY SILKAKNAGMSLGAKGAAGPLPS EAVQWLCHQAFLLKLTRHRVTY VPLLGSLRTAQTQLSRKLPGTTL TALEAAANPALPSDFKTILD	
Mauric eville RT	Retro plasm id	NC_0015 70.1	Q36578	MPNHRLPNCVSYLGENHELSWL HGMFGLLKRSNPQTGGILGWLN TGPNGFVKYMMNLMGHARDKG DAKEYWRLGRSLMKNEAFQVQ AFNHVCKHWYLDYKPHKIAKLL KEVREMVEIQPVCIDYKRVYIPK ANGKQRPLGVPTVPWRVYLHM WNVLLVWYRIPEQDNQHAYFPK RGVFTAWRALWPKLDSQNIYEF DLKNFFPSVDLAYLKDKLMESGI PQDISEYLTVLNRSLVVLTSEDKI PEPHRDVIFNSDGTPNPNLPKDV QGRILKDPDFVEILRRRGFTDIAT	cd00304

				NGVPQGASTSCGLATYNVKELF KRYDELIMYADDGILCRQDPSTP DFSVEEAGVVQEPAKSGWIKQN GEFKKSVKFLGLEFIPANIPPLGE GEVKDYPRLRGATRNGSKMELS TELQFLCYLSYKLRIKVLRDLYIQ VLGYLPSVPLLRYRSLAEAINELS PKRITIGQFITSSFEEFTAWSPLKR MGFFFSSPAGPTILSSIFNNSTNLQ EPSDSRLLYRKGSWVNIRFAAYL YSKLSEEKHGLVPKFLEKLREINF ALDKVDVTEIDSKLSRLMKFSVS AAYDEVGTLALKSLFKFRNSERE SIKASFKQLRENGKIAEFSEARRL WFEILKLIRLDLFNASSLACDDLL SHLQDRRSIKKWGSSDVLYLKSQ RLMRTNKKQLQLDFEKKKNSLK KKLIKRRAKELRDTFKGKENKEA	
RTX	Engineered polymerase	QFN49000.1	-	MILDTDYITEDGKPVIRIFKKENG EFKIEYDRTFEPYLYALLKDDSAI EEVKKITAERHGTVVTVKRVEK VQKKFLGRPVEVWKLYFTHPQD VPAIMDKIREHPAVIDIYEYDIPF AIRYLIDKGLVPMEGDEELKLLA FDIETLYHEGEEFAEGPILMISYA DEEGARVITWKNVDLPYVDVVS TEREMIKRFLRVVKEKDPDVLIT YNGDNFDFAYLKKRCEKLGINF ALGRDGSEPKIQRMGDRFAVEV KGRIHFDLYPVIRRTINLPTYTLE AVYEAVFGQPKEKVYAEEITTA WETGENLERVARYSMEDAKVTY ELGKEFLPMEAQLSRLIGQSLWD VSRSSTGNLVEWFLLRKAYERNE LAPNKPDEKELARRHQSHEGGYI KEPERGLWENIVYLDFRSLYPSII THNVSPDTLNREGCKEYDVAPQ VGHRFCKDFPGFIPSLLGDLLEER QKIKKRMKATIDPIERKLLDYRQ RAIKILANSLYGYYGYARARWY CKECAESVIAWGREYLTMTIKEI EEKYGFKVIYSDTDGFFATIPGA DAETVKKKAMEFLKYINAKLPG ALELEYEGFYKRGLFVTKKKYA VIDEEGKITTRGLEIVRRDWSEIA KETQARVLEALLKDGDVEKAVR	IPR006134, PF00136, cd05536

				IVKEVTEKLSKYEVPPEKLVIHK QITRDLKDYKATGPHVAVAKRL AARGVKIRPGTVISYIVLKGSGRI VDRAIPFDEFDPTKHKYDAEYYI EKQVLPAVERILRAFGYRKEDLR YQKTRQVGLSARLKPKGTLEGSS HHHHHH	

Table Z2: InterPro descriptions of signatures present in reverse transcriptases in Table Z1, as of March 4, 2020.

Signature	Database	Short Name	Description
cd00304	CDD	*RT_like*	RT_like: Reverse transcriptase (RT, RNA-dependent DNA polymerase)_like family. An RT gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. RTs occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. These elements can be divided into two major groups. One group contains retroviruses and DNA viruses whose propagation involves an RNA intermediate. They are grouped together with transposable elements containing long terminal repeats (LTRs). The other group, also called poly(A)-type retrotransposons, contain fungal mitochondrial introns and transposable elements that lack LTRs. [PMID: 1698615, PMID: 8828137, PMID: 10669612, PMID: 9878607, PMID: 7540934, PMID: 7523679, PMID: 8648598]
cd01645	CDD	*RT_Rtv*	RT_Rtv: Reverse transcriptases (RTs) from retroviruses (Rtvs). RTs catalyze the conversion of single-stranded RNA into double-stranded viral DNA for integration into host chromosomes. Proteins in this subfamily contain long terminal repeats (LTRs) and are multifunctional enzymes with RNA-directed DNA polymerase, DNA directed DNA polymerase, and ribonuclease hybrid (RNase H) activities. The viral RNA genome enters the cytoplasm as part of a nucleoprotein complex, and the process of reverse transcription generates in the cytoplasm forming a linear DNA duplex via an intricate series of steps. This duplex DNA is colinear with its RNA template, but contains terminal duplications known as LTRs that are not present in viral RNA. It has been proposed that two specialized template switches, known as strand-transfer reactions or "jumps", are required to generate the LTRs.

			[PMID: 9831551, PMID: 15107837, PMID: 11080630, PMID: 10799511, PMID: 7523679, PMID: 7540934, PMID: 8648598, PMID: 1698615]
cd01646	CDD	*RT_Bac_retron_I*	RT_Bac_retron_I: Reverse transcriptases (RTs) in bacterial retrotransposons or retrons. The polymerase reaction of this enzyme leads to the production of a unique RNA-DNA complex called msDNA (multicopy single-stranded (ss)DNA) in which a small ssDNA branches out from a small ssRNA molecule via a 2'-5'phosphodiester linkage. Bacterial retron RTs produce cDNA corresponding to only a small portion of the retron genome. [PMID: 1698615, PMID: 16093702, PMID: 8828137]
cd01648	CDD	*TERT*	TERT: Telomerase reverse transcriptase (TERT). Telomerase is a ribonucleoprotein (RNP) that synthesizes telomeric DNA repeats. The telomerase RNA subunit provides the template for synthesis of these repeats. The catalytic subunit of RNP is known as telomerase reverse transcriptase (TERT). The reverse transcriptase (RT) domain is located in the C-terminal region of the TERT polypeptide. Single amino acid substitutions in this region lead to telomere shortening and senescence. Telomerase is an enzyme that, in certain cells, maintains the physical ends of chromosomes (telomeres) during replication. In somatic cells, replication of the lagging strand requires the continual presence of an RNA primer approximately 200 nucleotides upstream, which is complementary to the template strand. Since there is a region of DNA less than 200 base pairs from the end of the chromosome where this is not possible, the chromosome is continually shortened. However, a surplus of repetitive DNA at the chromosome ends protects against the erosion of gene-encoding DNA. Telomerase is not normally expressed in somatic cells. It has been suggested that exogenous TERT may extend the lifespan of, or even immortalize, the cell. However, recent studies have shown that telomerase activity can be induced by a number of oncogenes. Conversely, the oncogene c-myc can be activated in human TERT immortalized cells. Sequence comparisons place the telomerase proteins in the RT family but reveal hallmarks that distinguish them from retroviral and retrotransposon relatives. [PMID: 9110970, PMID: 9288757, PMID: 9389643, PMID: 9671703, PMID: 9671704, PMID: 10333526, PMID: 11250070, PMID: 15363846, PMID: 16416120, PMID: 16649103, PMID: 16793225, PMID:

			10860859, PMID: 9252327, PMID: 11602347, PMID: 1698615, PMID: 8828137, PMID: 10866187]
cd01650	CDD	*RT_nLTR_like*	RT_nLTR: Non-LTR (long terminal repeat) retrotransposon and non-LTR retrovirus reverse transcriptase (RT). This subfamily contains both non-LTR retrotransposons and non-LTR retrovirus RTs. RTs catalyze the conversion of single-stranded RNA into double-stranded DNA for integration into host chromosomes. RT is a multifunctional enzyme with RNA-directed DNA polymerase, DNA directed DNA polymerase and ribonuclease hybrid (RNase H) activities. [PMID: 1698615, PMID: 10605110, PMID: 10628860, PMID: 11734649, PMID: 12117499, PMID: 12777502, PMID: 14871946, PMID: 15939396, PMID: 16271150, PMID: 16356661, PMID: 2463954, PMID: 3040362, PMID: 3656436, PMID: 7512193, PMID: 7534829, PMID: 7659515, PMID: 8524653, PMID: 9190061, PMID: 9218812, PMID: 9332379, PMID: 9364772, PMID: 8828137]
cd01651	CDD	*RT_G2_intron*	RT_G2_intron: Reverse transcriptases (RTs) with group II intron origin. RT transcribes DNA using RNA as template. Proteins in this subfamily are found in bacterial and mitochondrial group II introns. Their most probable ancestor was a retrotransposable element with both gag-like and pol-like genes. This subfamily of proteins appears to have captured the RT sequences from transposable elements, which lack long terminal repeats (LTRs). [PMID: 1698615, PMID: 8828137, PMID: 12403467, PMID: 11058141, PMID: 11054545, PMID: 10760141, PMID: 10488235, PMID: 9680217, PMID: 9491607, PMID: 7994604, PMID: 7823908, PMID: 3129199, PMID: 2531370, PMID: 2476655]
cd03487	CDD	*RT_Bac_retron_II*	RT_Bac_retron_II: Reverse transcriptases (RTs) in bacterial retrotransposons or retrons. The polymerase reaction of this enzyme leads to the production of a unique RNA-DNA complex called msDNA (multicopy single-stranded (ss)DNA) in which a small ssDNA branches out from a small ssRNA molecule via a 2'-5'phosphodiester linkage. Bacterial retron RTs produce cDNA corresponding to only a small portion of the retron genome. [PMID: 1698615, PMID: 8828137, PMID: 11292805, PMID: 9281493, PMID: 2465092, PMID: 1722556, PMID: 1701261, PMID: 1689062]
cd03715	CDD	*RT_ZFREV_like*	RT_ZFREV_like: A subfamily of reverse transcriptases (RTs) found in sequences similar to the intact endogenous retrovirus ZFERV from zebrafish and to Moloney murine

			leukemia virus RT. An RT gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. RTs occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. These elements can be divided into two major groups. One group contains retroviruses and DNA viruses whose propagation involves an RNA intermediate. They are grouped together with transposable elements containing long terminal repeats (LTRs). The other group, also called poly(A)-type retrotransposons, contain fungal mitochondrial introns and transposable elements that lack LTRs. Phylogenetic analysis suggests that ZFERV belongs to a distinct group of retroviruses. [PMID: 14694121, PMID: 2410413, PMID: 9684890, PMID: 10669612, PMID: 1698615, PMID: 8828137]
cd05536	CDD	*POLBc_B3*	DNA polymerase type-B B3 subfamily catalytic domain. Archaeal proteins that are involved in DNA replication are similar to those from eukaryotes. Some members of the archaea also possess multiple family B DNA polymerases (B1, B2 and B3). So far there is no specific function(s) has been assigned for different members of the archaea type B DNA polymerases. Phylogenetic analyses of eubacterial, archaeal, and eukaryotic family B DNA polymerases are support independent gene duplications during the evolution of archaeal and eukaryotic family B DNA polymerases. Structural comparison of the thermostable DNA polymerase type B to its mesostable homolog suggests several adaptations to high temperature such as shorter loops, disulfide bridges, and increasing electrostatic interaction at subdomain interfaces. [PMID: 10997874, PMID: 11178906, PMID: 10860752, PMID: 10097083, PMID: 10545321]
cd05780	CDD	*DNA_polB_Kod1_like_exo*	The 3'-5' exonuclease domain of archaeal family-B DNA polymerases with similarity to Pyrococcus kodakaraensis Kod1, including polymerases from Desulfurococcus (D. Tok Pol) and Thermococcus gorgonarius (Tgo Pol). Kod1, D. Tok Pol, and Tgo Pol are thermostable enzymes that exhibit both polymerase and 3'-5' exonuclease activities. They are family-B DNA polymerases. Their amino termini harbor a DEDDy-type DnaQ-like 3'-5' exonuclease domain that contains three sequence motifs termed ExoI, ExoII and ExoIII, with a specific YX(3)D pattern at ExoIII. These motifs are clustered around the active site and are involved in metal binding and catalysis. The exonuclease domain of family B polymerases

			contains a beta hairpin structure that plays an important role in active site switching in the event of nucleotide misincorporation. Members of this subfamily show similarity to eukaryotic DNA polymerases involved in DNA replication. Some archaea possess multiple family-B DNA polymerases. Phylogenetic analyses of eubacterial, archaeal, and eukaryotic family-B DNA polymerases support independent gene duplications during the evolution of archaeal and eukaryotic family-B DNA polymerases. [PMID: 18355915, PMID: 16019029, PMID: 11178906, PMID: 10860752, PMID: 10097083, PMID: 10545321, PMID: 9098062, PMID: 12459442, PMID: 16230118, PMID: 11988770, PMID: 11222749, PMID: 17098747, PMID: 8594362, PMID: 9729885]
PF00078	Pfam	*RVT_1*	A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. [PMID: 1698615]
PF00136	Pfam	*DNA_pol_B*	This region of DNA polymerase B appears to consist of more than one structural domain, possibly including elongation, DNA-binding and dNTP binding activities. [PMID: 9757117, PMID: 8679562]
PF07727	Pfam	*RVT_2*	A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. This Pfam entry includes reverse transcriptases not recognised by the Pfam:PF00078 model. [PMID: 1698615]
IPR000477	InterPro	*RT_dom*	The use of an RNA template to produce DNA, for integration into the host genome and exploitation of a host cell, is a strategy employed in the replication of retroid elements, such as the retroviruses and bacterial retrons. The enzyme catalysing polymerisation is an RNA-directed DNA-polymerase, or reverse trancriptase (RT) (2.7.7.49). Reverse transcriptase occurs in a variety of mobile elements, including retrotransposons, retroviruses, group II introns [PMID: 12758069], bacterial msDNAs, hepadnaviruses, and caulimoviruses. Retroviral reverse transcriptase is synthesised as part of the POL polyprotein that contains; an aspartyl protease, a reverse transcriptase, RNase H and integrase. POL polyprotein undergoes specific enzymatic cleavage to yield the mature proteins.

			The discovery of retroelements in the prokaryotes raises intriguing questions concerning their roles in bacteria and the origin and evolution of reverse transcriptases and whether the bacterial reverse transcriptases are older than eukaryotic reverse transcriptases [PMID: 8828137]. Several crystal structures of the reverse transcriptase (RT) domain have been determined [PMID: 1377403].
IPR006134	InterPro	DNA-dir_DNA_pol_B_multi_dom	DNA is the biological information that instructs cells how to exist in an ordered fashion: accurate replication is thus one of the most important events in the life cycle of a cell. This function is performed by DNA- directed DNA-polymerases 2.7.7.7) by adding nucleotide triphosphate (dNTP) residues to the 5' end of the growing chain of DNA, using a complementary DNA chain as a template. Small RNA molecules are generally used as primers for chain elongation, although terminal proteins may also be used for the de novo synthesis of a DNA chain. Even though there are 2 different methods of priming, these are mediated by 2 very similar polymerases classes, A and B, with similar methods of chain elongation. A number of DNA polymerases have been grouped under the designation of DNA polymerase family B. Six regions of similarity (numbered from I to VI) are found in all or a subset of the B family polymerases. The most conserved region (I) includes a conserved tetrapeptide with two aspartate residues. It has been suggested that it may be involved in binding a magnesium ion. All sequences in the B family contain a characteristic DTDS motif, and possess many functional domains, including a 5'-3' elongation domain, a 3'-5' exonuclease domain [PMID: 8679562], a DNA binding domain, and binding domains for both dNTP's and pyrophosphate [PMID: 9757117]. This domain of DNA polymerase B appears to consist of more than one activities, possibly including elongation, DNA-binding and dNTP binding [PMID: 9757117].
IPR013103	InterPro	*RVT_2*	A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. This entry includes reverse transcriptases not recognised by IPR000477 [PMID: 1698615].

Tables 3A and 3B (below) show exemplary Gene Writer proteins and associated sequences from a variety of retrotransposases, identified using data mining. Column 1 indicates the family to which the retrotransposon belongs. Column 2 lists the element name. Column 3 indicates an accession number, if any. Column 4 lists an organism in which the retrotransposase is found. Column 5 lists the predicted 5' untranslated region, and column 6 lists the predicted 3' untranslated region; both are segments that are predicted to allow the template RNA to bind the retrotransposase of column 7. (It is understood that columns 5-6 show the DNA sequence, and that an RNA sequence according to any of columns 5-6 would typically include uracil rather than thymidine.) Column 7 lists the predicted retrotransposase amino acid sequence. For Table 3B, column 8 lists the predicted RT domain present based on sequence analysis, column 9 lists the start codon position, and column 10 lists the stop codon position.

Table 3A.

1. Family	2. Element	3. Accession	4. Organism	5. Predicted 5'UTR	6. Predicted 3'UTR	7. Predicted Amino Acid Sequence
R2	R2-1_TG	.	Taeniopygia guttata	GTCTAGTTACAACT GGGCATCGCTGCAG AGATCGCACCTCCT CGTGGTCCCGCTGG TAGCCCTTCGAAGG GTGACTAAGTCGAT CTCTGCCCCAGGTA CGGAGCCGTTGGG ACTCACCAGTCCAA CGTAACTCCTGCCT AAATTCGGTGAAAC AAATTCCTCGGTAA AAAGCCCC (SEQ ID NO: 1140)	TTCAGGTTATTTAGATGCTT AGTTTTTGTACCTTTCTTGTT TTGTTTAGGATTTTGATAGT GTTAGTATTTTTATATTTTTG TACGATTGCATAAATGTTCTT TTTTATACAGTTCTGTTTTAA TAAAATAGACGATAGCTAG AGACGTTAGGGCAGCCACA AGCCAGTTAGGTAGCGGAT AGTAGGTAGGAACAGACTT TTACTATTTCATAACGCGTC AATTACCACCTGATTTGGAC CAATTCACGGGATTTGTCCA AGGTGGACGGGCCACCTTT ACTTAACCCGGAAAAGGAA CATATATAATTTATGTGTGT TCGATAAAA (SEQ ID NO: 1263)	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVVECVHF AAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKESE DAPKTSPPTPGKHGLEQTGEEKVMVTVPDKNPPPCPCCGTRVNSVLNLIE HLKVSHGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETEKAPAGEWI CEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRC WTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEP REEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAAA GGRTMDEIRRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSAGALN TFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLS EIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSK GSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLI PKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAA GCSENLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYIITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLA MDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILET FCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESE KYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLI YIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTRDGGLGITK LAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDR ENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKW TKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFL ARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEA KKKDWVVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVRYEAAKSSLEE AAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTEL GLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1016)

R2	R2-1_Gfo	.	Geospiza fortis	AGACTTAAGTGAGT TTGGTTACAACTGG GCATAGCTGCAGAG ACCGCGCCTCCTCG CGGCCCCGCTGGTA AGCCCTTAACAGGG TGACTAA (SEQ ID NO: 1141)	GGTAGATAATCTTTGTATAG TGGGGGGGGATCTCATGTA CCGGGTTTCTTTTATTTGAT TTTCAATAAAACAGACGGT AGCTAGGTTCGCAAGGCAG CCACAAGCCAAAGATAGGT AGGGTGCTCATAGTGAGTA GGGACAGTGCCTTTTGATTC ACAACGCGTCAATACCATCT GACACGGATACCCTTACCG GACTTGTCATGATCTCCCAG ACTTGTCCAAGGTGGACGG GCCACCTTTACTTAACCCGG AAAAGGAACATATATTAATT ATATGTGTTCGGAAAA (SEQ ID NO: 1264)	VGLCPSPGVDGTHQPNDSFQNFGETNFSVQVARLVTRNLAPRSVRGNG FGSGMATHPVPADESGHESDPFLVGRSCGQPARLTRQSVGTQTSRDDIL PSKTTKLTENELDLLVNFSLELYRSDLQGFVQEGIHFSVNREVLEGFPEVYE QPAPQPAVGDDLNTSLPPDNNICVLEKGSSEAVEDGTPEVAHPVPETQG KESPNNIVMVTLPNKNPPCPCCRVRLHSVLALIEHLKGSHGKKRACFRCV KCGRENFNYHSTVCHIAKCKGPKVEKAPVGEWICEVCGRDFTTKIGLGQ HKRLAHPLVRNQERIDASQPKETSNRGAHKRCWTKEEEEMLIKLEVQFE GHRNINKLIAEHLTTKTSKQISDKRRLLPRKQLTDLSKGVAGQKVLDPGLS HQPQLGVVDNGLGGGHLPGGPAAEGRTIEPLGHHLDKDNGHREIADQ HKAGRLQAHYRKKIRKRLSEGMISNFPFEVFEQLLDCQEAQPLINQAAQD CFGCLDSASQIRKALRKQNTQKDQGDQPKRPAQKWMKKRAVKRGHFL RFQKLFHLDRGKLAKIILDDVECLSCDIPPSEIYSVFKARWETPGQFAGLG DFEINRKANNKAFRDLITAKEILKNVREMTKGSAPGPDGIALGDIRKMDP EYTRTAELFNLWLTSGEIPDMVRGCRTVLIPKSSKPERLKDINNWRPITIG SILRLFSRIITARLTKACPLNPRQRSFISAAGCSENLKLLQTIIRTAKNEHRP LGVVFVDIAKAFDTVSHQHIIHVLQRRRVDPHIIGLVKNNMYKDISTVITTKK NTYTDKIQIQVGVKQGDPLSPLLFNLAMDPLLCKLEEHGKGFHRGQSKIT AMAFADDLVLLSDSWEDMNANIKILETFCDLTGLKTQGQKCHGFYIKPT KDSYTVNNCAAWTINGTPLNMINPGESEKYLGLQFDPWVGIAKTSLPEK LDFWLERIDRAPLKPFQKLDILKTYTIPRLTYVADHSEMKAGALEALDRTIR SAVKDWLHLPSSTCDAILYTSMKDGGLGVTKLVGLIPSVQARRLHRIAQS PEETMKDFLEKAQMEKMYEKLWVQAGGKRKRMPSIWEALPEVVPSIDT ATTSEWEAPNPKSKYPRPCNWRRKEFKKWTKLIAQGWGIRCFKGDKISN NWIRHYRYIPHRKLLTAIQLRASVYPTREFLARGREDNCVKSCRHCEAAEE SCAHIIGMCPVVRDARIKRHNRICERLMEEAGKRDWTVFQEPHIRDVTK ELYKPDLIFVKEGLALVVDVTIRFESTKTTLEEAAAEKVNKYKHLETEVRNL TNAKDVIFMGFPLGARGQWYNKNFELLDTLGLPRSRQDIIAKTLSTDALIS SVDIIHMFASRGRRQHA (SEQ ID NO: 1386)
R2	R2-1_ZA	.	Zonotrichia albicollis	CGACTTGAGAAGGT CTGGTTACAACTGG GCATAGCTGCAGAG ATCGCGCCTCCTCG TGGCCCCGCTGGTA AGCCCTTAACAGGG TGACTAAGTCGATC TCTGCCCCAGTCCA GGAGCCGCTGGGT TTCACCAGCCCAGC	GTAGTCACATTGCACTTTCT GTAACTTGCACTGGGTGTG GGATGTGGGCCTGGGGTGT GGGTTATGGGGTATATATG TGGGATATTCTGGTGGGAA TGTCCATTCACTGTATGCCT ATCTTTTTAATAAAAAGACG GTAGCTAGGTTCGCGAAGC AGCCACAAGCCAATAGCCA GTTAGGTAGCTCATAGTGG	NKFLGKSRVAYCLKPGPPVSDRGKEFGSGLTTHPEPESESGHDPTVPNPG PSLGAGEGAQPLPLLRVSVGTQTCEEDFITSRPTKLPGIESELGPLVKFSLE VYRSDLKGDVQFEGIHFPDNWGVLEGFPEVYEQLAPQPNGGDELNHSL PGDREGDVLEKDSSEKEKEAAPEALPSVQRARSEQLPDNIVKVTVPDKNP PCPCCGVRLNSVLALIEHLKGSHGRRRVCFRCAKCGRENFNHHSTVCHY AKCKGPQIERPPVGEWICEVCGRDFTTKIGLGQHKRHMHAMVRNQERI DASQPKETSNRGAHKRCWTKEEEELLMKLEVQFENHKNINKLIAEQLTTK TAKQISDKRRMLLKKGRGTTGNLETEPGMSHQSQAKVKDNGLGGDHLP GGPVVDKGTIGKPGQHLDTDNSHQITAGKKKGGGLQARYRRRIMKRLA AGTINIFPKVFKELINDQEARPLINQTTEDCFGLLDSACQIRTALREKGKSQ

			GATTCCTTCCAAATT CGGTGA (SEQ ID NO: 1142)	GTAGGTGACAGGAACCTTT GACTCAGAACGCGTCCATT AACATCTAGAACGGACCAA ACTTCGGACATGCACCGATT AACCGGATTTGTCCAAGGT GGACGGGCCACCTTTACTT AACCCGGAAAGGGAACATA TATAGTTATATGTGTTCGTA ATA (SEQ ID NO: 1265)	EERPRKQYQKWMKKRAIKRGDYLRFQRLFHLDRGKLARIILDNTESLSCDI SPSEIYSVFKARWETPGHFNGLGDFEIKGKANNKAFRDFITAKEIEKNVRE MSKGSAPGPDGIALGDIKKMDPGYSRTAELFNLWLTAGDIPDMVRGCR TVLIPKSTTPERLKDINNWRPITIGSILLRLFSRIITARMTKACPLNPRQRGFI SAPGCSENLKLLQSIIRTAKNEHKPLGVIFVDIAKAFDTVSHQHIIHVLQQR RVDPHIVGLVNNMYKDISTYVTTKKNTHTDKIQIRVGVKQGDPLSPLLFN LAMDPLLCKLEESGKGFHRGQSSITAMAFADDLVLLSDSWENMKENIKIL ETFCNLTGLKTQGQKCHGFYIKPTKDSYTINNCPAWTINGTPLNMINPGE SEKYLGLQIDPWTGVAKYDLSTKLKIWLESIDRAPLKPLQKLDILKTYTIPRL TYLADHSEMKAGALEALDQQIRTAVKDWLHLPSCTCDAILYVSTRDGGL GVTKLAGLIPSVQARRLHRIAQSPDETMKDFLEKAQMEKMYEKLWVQA GGKKKGMPSIWEALPMTVPPTNTGNLSEWEAPNPKSKYPKPCDWRRK ELKKWTKLESQGRGVKNFRNDTISNDWIQYYRRIPHRKLLTAIQLRANVY PTREFLARGRGDNYYKFCRHCEADLETCGHIIGFCPVTKDARIKRHNRICD RLCEEAAKREWVVFKEPHLRDATTELFKPDVIFVKEDRALVVDVTVRYES AKTTLEAAAMEKVDKYKHLEAEVKELTNAKDVVFMGFPLGARGKFYKG NFNLLETLGLPKTRQLSVAKTLSTYALMSSVDIVHMFASRSRKPNV (SEQ ID NO: 1387)	MESTAKGKSYWMARRPVEGATEGSLGRVPFVTRDPKRKPEAKRTLTHG LGLRECSVVLTRLIEGRRGRDHTPSGWNAQRGMPNDESSVEEPNGPIPS NPIPTGTQALPEPMADGEQGEHPGVVVTLPLRDLNCPLCGGSASTAVKV QRHLAFRHGTVPVRFSCESCGKTSPGCHSVLCHIPKCRGPTGEPPEKVVK CEGCSRTFGTRRACSIHEMHVHSEIRNRKRIAQDRQEKGTSTDGEGRAG VERADAGEGPSGEGIPPKRPRRARTPREPSEPPANPPILSPQPDLPPGGL RDLLREVASGWVRAARDGGTVIDSVLAAWLDGNDRLPELVDAATQRTL QGLPAGRLARRPATFVAPNRRRGRWGRRLKLLAKRRAYHDCQIRFRKDP ARLAANILDGKSETSCPINEQAIHEHFRNKWANPSPFGGLGRFGTENRA NNAHLLGPISKSEVQTSLRNASNASTPGPDGVGKRDISNWDPECETLTQ LFNMWWFTGVIPSRLKKSRTVLLPKSSDPGAEMEIGNWRPITIGSMVLR LFTRVINTRLTEACPLHPRQRGFFRRSPGCSENLEVLECLLRHSKEKRSQLA VVFVDFAQAFDTVSHEHMLSVLEQMNVDPHMVNLIREIYTNSCTSVELG RKEGPDIPVRVGVKQGDPLSPLLFNLALDPLIQSLERTGKGCEAEGHKVT ALAFADDLALVAGSWEGMAHNLALVDEFCLTTGLTVQPKKCHSFMVRP CRGAFTVNDCPPWVLGGKALQLTNIENSIKYLGVKVNPWAGIEKPDLTV ALDRWCKRIGKSLLKPSQKVYILNQFAIPRLFYLADHGGAGDVMLQNLD GTIRKAVKKWLHLPPSTCNGLLYARNCNGGLGICKLTRHIPSMQARRMF RLANSSDPLMKAMMRGSRVEQKFKKAWMRAGGEESALPRVFGANQY QEGEEVANDLVPRCPMPSDWRLEEFQHWMGLPIQGVGIAGFFRNRVA


NGWLRKPAGFKERHYIAALQLRACVYPTLEFQQRGRSKAGAACRRCSSR
LESSSHILGKCPAVQGARIRRHNKICDLLKAEAETRGWEVRREWAFRTPA
GELRRLDLVLILGDEALVIDVTVRYEFAPDTLQNAGKDKVSYYGPHKEAIA
RELGVRRVDIHGFPLGARGLWLASNSKVLELMGLSRERVKVFSRLLSRRV
LLYSIDIMRTFYATLQ (SEQ ID NO: 1388)
CGCTGCTGCGAGCCTGAGG
TCGATGGTTAGAGGTGAAA
TACTTGGGAGGAGACACAG
CCTCCGGAGAGCCCCTCCC
GGGTGGTCATCATGGCAAC
CGGGTGAAACCTTACGGTT
TCACTTACGAAACAGCACCA
TAACAGCGCCGTAATAGCG
CACCGGTGTGACTACTGTCC
AGTGCTGATATTCTCATCTG
GAGAATACAACACGGGTAA
TGGCAGAGTATTCAAAACC
CAAATGTTTACGATCGACCA
ACGGAGTCGTTCCCTTGCAT
CTAGGCCGGACCCGAAACT
GCCGTAATTGCCCGTCCCCA
AGGTAGCCTCTTAGAAAAC
CGAAGCCCGGTCGGGGCG
GTGGTTGCGGCGGCGCTGC
GGGGGCCTGCTGCTCGGGC
GGCGTCGGTGTGCCGCGGT
GGTTGCGGTGGTGCGGCG
GGGATCTCGGTCCTTGCGG
TGCCGCTGTGCCGCCGCGG
TCGCGTCGGTGGCGCTGGG
GTGGTGGCCCGAGTGGCGT
CGGCGTGCCACTGCCCATA
GTCGCCCGCGGGGGCGACC
GATCTGGAGGGGCGAGGG
GGCTCGCGGGACTTTAACG
AGAAACGGAACGCAACTTC
TCGCATCGCTCCCGGGACTT
TCCCCCCTCGTTCAGCCGAG
GGATGCCAAAAGGCATGAA
AGGTAAGTACCATACCGGT
CCGCAAAACTCTCTTCTGAC
TCGGTTCTCTGTTGGTTTTC
TAGAGTAACAACGAGGTGG

					AGGAGAGGGACATGGCAG GGACTCCCATTCGTGCCAG CGGGTGGGGACAGATCGA AGGAACGGTTCGAGGGCGT AACAGACGAGAGGGAATCC GGTCACACATTGATGCCAT GCCTAAATAGGCGAGGTTT GTATTTCTACTTTGTGGGTT CAGTATAGTCGGAGCATAT GGTCGGTTGTCCCGTTGTTT TCACGGCGGGCAAGCGACT ATCATGATAAAGTAGAATG GGAGACGGGCTCCCTGACA AACCCGGAAAGGCGCCCCC CCGTGGTTCGTAGCAGCTG ACGGATCACGCTCGAAGAA AAATGAGTGAGAGGGGAC GCCGCAACCAC (SEQ ID NO: 1266)	MLRGGVGTPPAGGAGAVGPGMASPGGCSVRFSPGGRRLLGHRTGGLS PSVSWRLKRLSVSLRRWSGPGLLGADGAGGGAAVASPRGTQVLGSGAG RRWLGHGSRGSSPSAARGLRRLTVRLKRLSGGLLSPKACRDAEEGSSSSP GFRNPKGLGGRGLTPLGSRRFCRLTVSLNRWRGSLVKLNASSRASGRRTP VKPACDSRAGRGSEHAEGGGVSAAPMVLRSRRKLTFSVDGDSNSGDRA RSGSVSAARPGHLLVDGESASSRSGPAGDARLAGPSTRSRRKGCLPPVDF ENPKKRTRLMAKMTNGNPTSHVPCPAPCSNGHEGGGRVAVIEGRLPEL SGSRISGIQPALPVETSFVGQSTGRGADGDANANSSPPSPNLGGSVGMV PAVRDGTPPLGRPGEDHSRECAGGNTPLWMLEDSFRCDYCPREFGTRA GRSLHMRRAHLAEYDGAGFCWGERLSEFAATRLWSTEETKKLAVFCERG VPSPSECRAIAASLGAGKTHHQVRSKCRLVFEAIRRRELLEVAAATERLEK SARRKQPAVPPAPVHGVRGVLRGLLGKRVPREGGTTGSTSARIVRRDDC RQGAVASASLNLIRRLGRKATGRSGRRRVLGRPPRMDVRRSVRMRRMR RFLYRLARLGWAKLAMFVLDGQMGASCPVPLVEVSAVFRERWSIVRAFL GLGQFGGFGTADNAGFGKLIDPAEVRAHLQSIKNRSSPGPDGITKVALSK WDPEGIKLAHMYSTWLVSAGIPKVFKKCRTTLIPKTGDVSLHGDVGQW RPITIASLVLRLYSRILTERMTVACPSHPRQRGFIASPGCSENLMLLEGCMS LSKAGNGSLAVVFVDFAKAFDTVSHEHLLSVLVQKGLDQHMVELIKDSYE NSVTKVHCQEGCSTDIAMKVGVKQGDSMSPLLFNLALDPLIQQLEREGR
R2	R2-1_GA	.	Gasterosteus aculeatus	CATATTGGGGTCTC AGGAGGAGACACA GGGTCTGTTGCGGC TCCGGTAAACGGTA CCGGAGTCGGTTAA GCATCGTTTGGGCC CGCCTCCACGTGGT GGTCCGCGGTAACA CCAATAGGGTGGCT AAGAGGCCCAGTA ATTTCCCCGAATTGT CTTCCCCCCCGCGC GGGGGGGACCCCC CTTTAGTGTCGGAG CGGTCGCGCCTCCG CGTTTGGGGTGTCG CAGGCGTGAGCCTT CGTCCCCTTAAGTTC AGACGGTCCCGGCT	GGAGGGGAGTAGGTCTCTA CTCTGACCCGAAGGGCCCC CCCGTTTCAGACCTGATTCT AGGCTACCTGTGCCTAATTG GGGGGGTCCCAAAGAGAT GTTGTCTGTTGTAGAAGGG TTTGCGCCACTGACTGCACG GAAGGGTGGGCCTCGACA GGTAGGGGTTACATGACTC CGTGCTGCTCAGCAGACCC GCGCCTCTGAGACCGGGTA GGGCTACTTGAACAAGCGA CGCCCTGGTGTATGTCCGTA TCCTAACCTGGTTTGGGAA AGCCGATACCGGCAATGCC CGCCACAGGTGTCGCGCAC CCCACGGGATGACGTATGG GCCCCGGGGGACCTCATGG ATACTCCACTGGACTTGCAC

	TCTTGCCGGGCCAA	AATCCTGGTGTACTGGATG	GFPVNGKSITAMAFADDLAIVSDSWEGMRANLDILVDFCELTGMRTQP	
	CCCCCGGTGCAGCG	CAGCGACGTTGGTGACATA	SKCHGFLIEKSGSRSYKVNRCEPWLLNDTALHMVGPKESIKYLGVQVNP	
	TTCTCCCATGTTGG	AGCAATCGCTAAGTCGGGG	WTGIFAEDTVAKLRQWVVAISKTPLRPLDKVSLLCQFAVPRVIFVADHC	
	ATCGGCACCCAGCC	TAGGGGAGGTGGGGACCT	MLSAKALTEMDRSIRQAVKRWLHLARCTTNGLLYSRKSSGGLGIPKLSMI	
	CCGGGTGCCATGCG	CGGCACGGCTGTAGGAACG	VPAMQARRLLGLSRSKDETVRWMFLETTDHVAFERAWLRAGGSPDEV	
	AGTTCAGACATTTT	GGTGTATGGGCTCCGGCAG	PELGPDLVEGSPAEGNADPVSTVRPRKRIVPCDWRQVEFDRWAGQLVQ	
	GTTTATGTATCGTCT	CCGTCGTCACTCCCATACAA	GKGIRTFEADKISNCWLYDYPPNKLKPGDFTAAVQLRANVYPTRELAGR	
	GCGTGGTTGACTTG	CACAGGGGCTGCATCCTGG	GRTDTIDVCCRHCGEAPETCWHILALCPKVKRCRIQRHHKVCQVLVAEA	
	CTAAGCTCATTTCCT	TGGCCGGTGCTAGTTGGTT	ERHGWEVEREKRWMLPSGECVAPDLICWLDELALIVDVTVRYEFDEESL	
	CCTCTCACTGCGTC	CTGGAAGCCCGCCCGGGCT	ERARIEKECKYRPLIPVIRASRVQTKKVTVYGFPLGARGKWPAKNELLLAD	
	CCCCCAGGTGCTGA	GGTTCGCAGAAGCAGGGTG	LGLSKARTRSFAKLLSRRVLLHSLDVMRTFMR (SEQ ID NO: 1389)	
	TCGGTTGAAGAGG	CGCCCAGGGTAGGTTTGGT		
	ATTCGTCGTTGACC	ATATCTGGGTCCGGTGCGA		
	TCGGCGGTGAATTT	TACCTATCGATGGGCAGCG		
	GGGATTGTATTATA	AGGGCCGCCTCGTGACGCG		
	CAGGTAGGTATAGA	CTGTGTGGAGCTGGAGCCG		
	GGGCGTGCGG	GCCTGGGTATGAACAGTTC		
	(SEQ ID NO: 1144)	TTGCGGATGTGGCGTAGCT		
		AGATAGTACCCGTGGTTGT		
		GGGCGTGGTGTCGACCAAA		
		TGTTGTCCTGTGTGCACATA		
		GGCCAAGGGTTACGTGGGT		
		GGCAGTCAGAAGCACCCGC		
		ACCTGGAAGTGATTGCCCC		
		GGGATCCCGGCTCTCTGTG		
		AAGAGCTACCTTGAGGAAA		
		GGTGTTCCGCTGGAACTCA		
		AGACCCTACAGTAGGGGAT		
		ATCAACTGGCTTTGAGGTG		
		CTGTGATTCCGGAACCAGG		
		GCGAGGGCGAGTACTTAGA		
		GCATGTCCAAAAGCCCGGG		
		GAACGTTCCGGGGGGCCTGC		
		TTGGGTCGTTGGACCCACA		
		TCCGTAAAACGATGGATCT		
		CGCGTCGGCGCTCGGGAGA		
		ACTTCCCGCATGAACGCTG		
		ATTGCATGTGAGAACGCCC		

				CCACGGCGGCGGGGCAGG CGCTCCCCCTGGGTGTAAG GCTCGGGGGGGTCACGGCT CCGCTCTAAAAG (SEQ ID NO: 1267)	GCCTTGCACAGTAGTCCAG CGGTAAGGGTGTAGATCAG GCCCGTCTGTTTCTTCCCCG GAGCTCGCTCCCCTTGGCTTC CCTTATATTTAACATCAGAA ACAGACATTAAACATCTACT GATCCAATTTCGCCGGCGT ACGGCCACGATCGGGAGG GTGGGAATCTCGGGGATCT TCCGATCCTAATCCATGATG ATTACGACCTGAGTCACTAA AGACGATGGCATGATGATC CGGCGATGAAAA (SEQ ID NO: 1268)	MMASTALSLMGRCNPDGCTRGKHVTAAPMDGPRGPSSLAGTFGWGL AIPAGEPCGRVCSPATVGFFPVAKKSNKENRPEASGLPLESERTGDNPTV RGSAGADPVGQDAPGWTCQFCERTFSTNRGLGVHKRRAHPVETNTDA APMMVKRRWHGEEIDLLARTEARLLAERGQCSGGDLFGALPGFGRTLE AIKGQRRREPYRALVQAHLARFGSQPGPSSGGCSAEPDFRRASGAEEAV EERCAEDAAAYDPSAVGQMSPDAARVLSELLEGAGRRRACRAMRPKTA GRRNDLHDDRTASAHKTSRQKRRAEYARVQELYKKCRSRAAAEVIDGAC GGVGHSLEEMETYWRPILERVSDAPGPTPEALHALGRAEWHGGNRDYT QLWKPISVEEIKASRFDWRTSPGPDGIRSGQWRAVPVHLKAEMFNAW MARGEIPEILRQCRTVFVPKVERPGGPGEYRPIIIASIPLRHFHSILARRLLA CCPPDARQRGFICADGTLENSAVLDAVLGDSRKKLRECHVAVLDFAKAF DTVSHEALVELLRLRGMPEQFCGYIAHLYDTASTTLAVNNEMSSPVKVG RGVRQGDPLSPILFNVVMDLILASLPERVGYRLEMELVSALAYADDLVLL AGSKVGMQESISAVDCVGKQMGLRLNCRKSAVLSMIPDGHRKKHHYLT ERTFNIGGKPLRQVSCVERWRYLGVDFEASGCVTLEHSISSALNNISRAPL KPQQRLEILRAHLIPRFQHGFVLGNISDDRLRMLDVQIRKAVGQWLRLP ADVPKAYYHAAVQDGGLAIPSVRATIPDLIVRRFGGLDSSPWSVARAAA KSDKIRKKLRWAWKQLRRFSRVDSTTQRPSVRLFWREHLHASVDGRELR ESTRTPTSTKWIRERCAQITGRDFVQFVHTHINALPSRIRGSRGRRGGGE SSLTCRAGCKVRETTAHILQQCHRTHGGRILRHNKIVSFVAKAMEENKW TVELEPRLRTSVGLRKPDIIASRDGVGVIVDVQVVSGQRSLDELHREKRNK YGNHGELVELVAGRLGLPKAECVRATSCTISWRGVWSLTSYKELRSIIGLR EPTLQIVPILALRGSHMNWTRFNQMTSVMGGGVG (SEQ ID NO: 1390)
R2	R2_BM	AB076841	Bombyx mori	GGGCGATACGCATA ATTTTAATTTCCCGA TTGAAATCCAGTCG TCTTAATCTGGTGA CCAGTGGCGCGGTC ACCAGTATAGTGCA CAGGACGTGAATG GCTCCGAGGCTGGC GGAGTCACTCACTA TAAGTGTGAGAGAC GATGTCCTGTGCCA AGTATACGTCCAAC CCTAACGGGTTAAG TGAAATTAGTTGCT CATAACAGGGACG GTGTACCTGTTTGC TCGTGGCTGGCTAT CGAATGGACGGGA CCAATACACCCCCC TGTTAGTAATGGGG TAAGAGAGAGCGG TCTGAAACTATGGC CGAAATCACGACGC CCCACTCCTACCCAT AACCTGCACGTGGT ACCGCCGCACATTG ACCGATACGGGAG GAGGGGCAGCACT TGAATCACGTAGTC TTGGTGTAGCCATT GCGGGACTACAGCC CTCGTAAGTGCCGC CTTAGAACGCAACG	

				GGGCAATAGGTGG GCCGGGGCGCTAG CGGGGGGAGTAA TCTCCCCTGTTGGC GTGCACCGCACTGC TCCCACTGGGGGCA GTGTCATCCGGAAA CAGGTGGGCCGGG GCGCCACCAGGGG GGAGCAATCCCTCC TG (SEQ ID NO: 1145)	
R2	R8Hm-A	.	Hydra vulgaris	TTCAAGTGGATGAA GCTGGGAAGGTAA TCTGTAGTTGGTTG AGTTGGTTGCAGAT TACTGCTGTCGATT TTGCTTTCTATTGAA AGCCTGTCTCTACG GGTCCTGAAGCTTG AATTTTGGTAGCTA TAGTTTTGTGGGAG GAAAGTGGAATTTT GTACCATCTTTTGTC TCTCGTATCTACTAT AGTAAATCCGGTCA TGCAGCCTCTACGC GGCGCAACTAGAA ACTTGGATCAGTGA TCAAGGCTAATGCA TGCCGGGTCTCCTC AGATTAGGAGTATA ATACAAATCTGACT TCATCACTAAGAGG CTATGGGGCTAACG ATCCTATAGTCTCG (SEQ ID NO: 1146)	TAAATGCCAAAAGTTGCTT GGGCTAAATGATACGTACG CTAGAAAAAGCGACTTGCT GCACGGATGACGGTTCATC AGAGCCCGATATGTGCATG TCAAGGCGGCAGGGAGAAT CACTAGTGTAGCTGTTCTTT CCATTACGACTTACGCGGTT AACGTGGCACGATAGATTT ACACCAGGAAATAATACGT GAAGGGTTCCACCATATAC TGGAGTTTAGATCTATGAG GGAAACATTTGTAATAAGT CAGTCTGGTAACCTGGCGC CGCTGTTGAGTCAAATTAAC TATGTCAATACTCATTAAGT TATCGACTTTGATATGGCAT GGGGTGATTCCGCGTTATA TCAAAGTCAAACATGATGA TTGCAATGAGAAACTACCA CGCTTGGTCACGTTTGTGA GGAGAACATCTCATTCAAG CCTCCCGGATGTCGGCACC CGCTGACATCTTCTGGCTTA TGAAAATTTTCATTAATTTTT	MNLLIVTSSIKESDVPSSGKGGVAVNNITAGASGKDTCVIIHPGTDGIWC CTECVEIHNSGKDLKRHLAKRHPSVTISGYKCNLCPFVSERQLSVGTHLRY CRGVKEVVKREFACASCSFSSDTFSGLQVHMQRKHIAEWNDQLKEKTEF AWTDRELRELAEKELTTPSFRYNKIFYAALGTSRTYDAVRKIRYNDRYKSAI AEMRSQIADAAAAAQERDVERGLVSAHSDRGKEMLPVVETKSDIQVNN DIKKDIELTPNSRQKQTNLALARPAVIEVEEDLGRQDVKQYLASLRQDDY TSPAERSIFAYCREETNWSATKRQVLKISRTTRGLRQPKKVRPFEFPEGFK PNRNMRKWRKYRFLQECYREKRAETVSKILDGTFIDEPEEEIRPELEEVQR MYIDRLEKRTQLDTTKIVQTDEVFCLQSYGRITIGEVRDALGASKKDSASG PDGLLLQDVRRLGPLLLCNIFNMWYLHGIPVEENRCRTILLYKSGDRHLAS NYRPVTIGNMLNRLYAKIWDKRIRKNVRLHVRQKAFIPVDGCFENVKTIQ CVLQSYRKRKLEHNVVFIDLAKAFDTVLHDSIRKALWRKGVPSGVVKVVD SLYAGAVTSISVGKTKTRSICINSGVKQGCPLSPLLFNLILDELAERIEATGC GLDLDGHVLSSMAFADDYVLLAKDSVEMNELIRVCSTFFKEKGLSVNPGK CQSLRVLPVKEKKRSMKVLVRPHRWWRIKDQDVDIPSMTYDSLGKYLG VSIDPTGKIALPIEEWKNWMTKLKECKLKPEQKVKILKEVVCSRVNYVLR MSECGISELRSWTRFVRNWAKNIIHLPTWCSSDWIHSIKGLGIPDVSKGI VIQRMRASEKMSTSEDGIVRVVGARLVQKNRVLWEKAGFEGIELKAARR HCEVERLNNIGNITNGVALKTIAAVSSVNRYWMIEDNLKSGNKILVWKA MAGAIPTKINLSRGVADQTLKKCRRCGLTAETDGHILAGCHTSSDAYSKR HNMLCDKLAKELKLNGGPNRRVWWRERTCFTSTGRRYRPDIIVKDDSKITV IDMTCPYEKSEGHLIQCESAKVTKYEPLKLDKYWTRELEGANGIVAEKVEL MGLAIGAIGTIMRSTLRKLCELKSGRIVRRLQMIACNNSAQIIKGHLSRAT RRNLR (SEQ ID NO: 1391)

R2	R8Hm-B	.	Hydra vulgaris	CTTGGGGTCACTGA CACATTTTTCGGTA GCCATAGTTTTTTG AGAGGAAGAGTGG AAGTTTTTCCATGA GTCGTCTCTCGTAT AAACTGTGGTAAAT CCGGCCATCCAGCC TCTACGCGGCGCAA CTAGAAACTTGGAT CAGTGATCAAGGCT AATGGATGACGGG ACTCCATGGATAAG GAGATATAAAGATC TTATTTGAACGCAT CTTAAGGGGTTATG GGGCTAACACCCCC TTAATTCTGGTGCA CATTTATTGACCGTT (SEQ ID NO: 1147)	GTAAGTCATGGGCGGCTTG AAAGC (SEQ ID NO: 1269)	MSNRITIGDVPSVGKGGLTVNKQTAGADGAEACVVIHPGAKGIWSSPA
					ATGCCCGAGGTAGTTGGGA TAATGATGCACAAGCTCGT AAGGCGACTTGCTGCACGT ATGCCGCTAAACGCTTAGCT CGATGAGTGCATGTCAAGA CGGTCGGGAGTATGATCAG TGGAGCTGACTTTCCAGAC AACTCACGCGGATTCGCGT GCGGTGGATACAACACCTG GTATAACATATGAAGGGTT CCATCTAGTACAGGGATAA CGATCCATGGGAGCAAACT AATTAGTTGGAGGTAATCC AACGCCGCTGTTGAGTCAG TTTTTAACCGCCAGTCAACT CTTGTAGGTTATCGGTCTTC GGCAGACCTTGGACCGCCT AGCGCCGGCCAACAGTTTG TCGTCGACTAACATGATGAT TTGCGAGAGAAACCCACGC TTTGTCACTTATGTGAGGAT AAAATCTCTTGTCCATATGA TCCTTTGAAGGGAACAGCG CTTTGAGCTTGCTCGGCGTT GGCACCTTTAGTCTGTAATA TTTTCTTGATATTATGGACG AAAAAGGTAGTATGGTTGC A (SEQ ID NO: 1270)	CLRKFTIGKELRAHLAQIHKLAPSAVRYRCNKCPYEGDVQLSVGTHLRYCK
						GIAGVVEEKKQFACAICNFSSDTFSGLQVHKQRKHVVEWNEQLKEKTEF
						AWTDRELRELAVKEVTIPFSVVNTETFAVLDITTRTKDAVRKIRYTDRYKSI
						LAEVRAQVNAVAEEAPQASDESQITLLVNTGRGAELQPAVINITDSIELVT
						DVNEVEMVTSNSTNEEQPINAPVEPAVIEADLGRQDAKLYLASLRQSDC
						TNASDRWTLAYCRGEVDWCKTKSRLFKVSRHARGLRQPQRVENWEFPE
						GFRPNRNLRKWRKYSFLQSCYRTKKKETVSKILDGTFKDTPEEEIRPELEEV
						QRVYVDRLEVRTQLDTTRTVHIDERFDLVSYGRITIREVQDAISASKKDAS
						GGPDGLLLQDVKKASPRQLCIIFNMWYLHGIPVVENRCRTILLHKGGEKH
						LTSNYRPVTIGNMLNRVYAKIWDRRIRKNLQLHVRQKAFVPLDGCFENV
						KTIQCILQSYRRSRREHNVVFVDLAKAFDTILHDSIEKALLRKGIPRSVIKVV
						DSLYAGAVTSITVGKTKTRPICINSGVKQGCPLSPLLFNLVIDELAERLEAT
						GCGLDLEGHVISSMAFADDYVLLAKDSVEMNVLMNVCNTFFEEKGLAV
						NPAKCQSLRVLPVKGKRSMKVLTRTHRWWKINNQDVEIPSMTYESVGK
						YLGVMIDPAGKIALPIEEWKLWLTRLRECKLKPDQKVKVLKEVVCARANY
						VLRMSGCGICELRKWSRFVRGWVKSIIHFPAWCNSEWMHSSKGLGIPD
						VVSGIVIQRMRAAEKMAKSTDGVVRVVGARIVQTNRVLWKRAGLAGIE
						LDAARKFCEVKRVNKIGNQTNGGALKTIAESSVSRHWLLEKNIRPGNKIL
						VWKAMAGVIPTKINLSRGVADQTLKKCRCCGLTAETDCHILAGCPTSRD
						AYSKRHNLLCDKLAKELRLNGGPSRRVWRERMCLSGNGRRYKPDIVVKD
						DGVITVIDMACPYEKSERHLSQCEDAKVAKYEPLRLDRSWTQELEGNNG
						RSANEISVVGIAVGAIGTITRKTQRILSKLKLAKVGRPLQIIACNESAQIIRR
						HLSGSRLRNLR (SEQ ID NO: 1392)
R2	R9Av	GQ398057	Adineta vaga	GAAATAGTTTGCAA TGGTAGGTGTATGG CGCCTCTGTGTCTCT CTTTCGCTGGATAT AGTTTGACGATTTT GTACCAGGTATCTG TTTCTTGTGAGTTCA GCACCAGTTTGAAC	ACTAGTCTCCTTCTTCTATTA GTCAGTCTAATTAATTTTTC TTACATTCTACATCTAGTTC CATTATTAAATTGGTATGAT CAGTGCTATCTCTGCTACAC TCAATGCTTAATCGTATGTT ATTGACAGTCTGACACTTGA TTACTCTTACGACATATGCA	MNLPIREHAVSVHNINKFNYLCQLCSKSYDTINSVKAHYVACRRQKNASS
						TTAVPTNVINNNQLAINTNQVISRNPLQCVECLMKQVDFYAKDTKALVT
						HMRTKHAAAYEESKKVATRRVAWSPDEDQILAELEVKLKKIQKGQLLSRL
						VVEYNKCADKSKAPSRSKDAIRTRRQQHDYKLLLRSLQSQQPPVGSEDSD
						SDISSSNNNPLTTTHNVTPTPDSSNVVLLIQKIRESVDSIVKITNLKLNTNM
						LNAASAFINQNNNMDPLELSMRGIEEDVKAIRDKELQKPTRNVPSSTTSR
						KPTRNAKRLEKSKKYGYYQHLYYNNKKKLVAEILDGETSGAKPPPMNLVE
						DYYRNIWSRSTIDDSPVNNIKTVNSDSIFAPISRDEIKLALSNTKKDSAAGP

			AGGCTTAGCGATAG ACCTTCGAACTTGA AACACTGTTGTGAA GCTGGCTGGGCCCC TGCAGATTTTCTCG ATTAGAACGTGAGT GTTACGTCCAGAAT GACCCACCAGTGGT TAGTTCTACGTTGC CCTGGAAAGGAGA AAAGTTGAGCTAAA ATCGCACGGCCTAG TTGTTTATCAAATA GGCACGGTGAGGA ACTCTTCTATGTACC CTGACTAAAGTACT CACTTGTGCGCTGG GTTTGCTCCCCCTC GCATTGACTTATCT GATCGCACTACCCA CCAAACGAAACATA AACTTAGCTCGTGG TATCAGTCCACAGC GTGTGCAGTCGGAT TCAGGGGAGCGTG TTAGTGACAAGCAG GATAATATTAACAT AGTTAATGTTAAGG CGTTCAACATTCCTT ATCCAATTGGAAGA GTTGACTGTGAAGT TTGTCATGAAGACA TTGGACAA (SEQ ID NO: 1148)	CTGTTTGCTTCAGAGAAACC ACTGTTCATATAGTGAAGTT CCTCAGTTTTCTGTTGATAT ATTCTTCTTTCATTCTCGCTT CTCCTTTTCTACTGTGTTCTT TTTATCAGTTTTTTTGTGGAA AAATTGAGAATAAATAAAG T (SEQ ID NO: 1271)	DAVTIKEAKAIIDNLYVAYNIWLGVQGIPEQLKLNKTILIPKGNSDLSLLKN WRPITISSIILRVYNRLLAYRMNKIFKTNDKQVGFKPVNGCGINISWLHSLL KHARLNKNSIYACLVDVSKAFDSVSHQSIVRALTMNGAPSLLVKLIMDQY TNVNTVITCSGSISNKINISSGVKQGDPLSSLLFNLVIDELFDVIKDQYGYTI DNIGTTNARCFADDLTLISSSRMGMNKLLELTTKFFKERGLNVNPSKCMS IGMSKGYKGKKSKIESEPLFSITDAQIPMLGYIDKTTRYLGVNFTSIGAIDA KRIKKDLQDTLDKLEHLKLKAQCKMDLLRTYMIPRFMFQLIHTELYPKLLIK MDILIRKLAKRILHLPISTSSEFFYLPFKEGGLQLTSLKEAVGLAKIKLHKKIM SSNDPMLCYLIESQRSRIVEHFMKDLKLGDSLTLNEMNNIKECFMKEKRIS FAQKIHGVGFEVFSSSPLTNQWINGEIKTMTTKTYINSIKLRTNTLETRVTT SRGLNIIKTCRRCHVADESLMHVLQCCSSTKGLRYSRHHKICAKVANKLV MNGYGVFREKSYPDPNNSGSYLRPDIIAVKNGHVIVLDVTVVYEVTGATF INAYQTKINKYNAIMVQIEQMFNCVNGELHGLVIGSRGSIHHSQLHIWH QMGFSSIELKYVAIGCMEDSLRIMSTFSKAIT (SEQ ID NO: 1393)
R2	R2OI	LC349444	Oryzias latipes	CGCACAGGGGACA CAGAGCCTGCCCAA GTACCGCTCCCGAG GGAGCGGGAAACG	GGGGGACAGCTGGGAGTC TCGGCATGATTACAAATCTT GCGCTGCACTCGGATGTCG TCCCCGTGACGGACACATT	MGTDTVYVGQDYPSGLSKRVPARLVAGPMLRERSCHAHVFRAGHMW NWRTSLPSGRWDQPALEKSRVLTRSVATATDPEITSYPGKSVSTSTQVQE EDWCSRESGWISPGLAPEEPSVVSEITASMVATMRVATEEVVLEPQPEQ VVTILPEHGRNVPPGLAEQDTASPIEVSVLLPDLAENCPLCGVPSGGLRLL

			GGGGGGTGACTAT CCCCTGGGGTCCGG CGAGAGCGCTGGTC TACGGACCAGGGG TGGCTGTGGGCAG GCTGCTCCTCAGGC CAGTTGATTAGTTA CGCATGGGCTGTAC CTCCACGTGGTCCC GCTGGTAACGACTT GTCGGCTAAATCAG CCCGCCCACCATCT GGGATATGGTTGAC CGTCTAACCCCAGT ACTCAGGTCACAAA CAAA (SEQ ID NO: 1149)	AATCCGGAAAGCGAGTGGT GACTCGCCTCAAG (SEQ ID NO: 1272)	GKHFAVRHAGVPVTYECRKCAWRSPNSHSISCHVPKCRGRARMPSGDP GIACDLCEARFATEVGVAQHKRHVHPVEWNKVRLERRGARGGGIKATK LWSVAEVETLIRLIREHGDSGATYQLIADELGRGKTAEQVRSKKRLLRIDT ASNSPDDAEVEEERLESLAVRSSSRSPPSLVATRVVREAVARGESEGGEEIR AIAALIRDVDQNPCLIETSASDIISKLGRRVDGPKRPRPVVREQTQEKGW VRRLARRKREYREAQYLYSRDQARLAAQILDGAASQECALPVDQVYGAF REKWETVGQFHGLGEFRTGARADNWEFYSPILAAEVKENLMRMANGT APGPDRISKKALLDWDPRGEQLARLYTTWLIGGVIPRVFKECRTKLLPKSS DPVELQDIGGWRPVTIGSMVTRLFSRILTMRLLTRACPINPRQRGFLASSS GCAENLLIFDEIVRRSRRDGGPLAVVFVDFARAFDSISHEHILCVLEEGGLD RHVIGLIRNSYVDCVTRVGCVEGMTPPIQMKVGVKQGDPMSPLLFNLA MDPLIHKLETAGTGLKWGDLSIATLAFADDLVLVSDSEEGMGRSLGILEK FCQLTGLRVQPRKCHGFFMDKGVVNGCGTWEICGSPIHMIPPGESVRYL GVQVGPGRGVMEPDLIPTVHTWIERISEAPLKPSQRMRVLNSFALPRIIY QADLGKVTVTKLAQIDGIVRKAVKKWLHLSPSTCNGLLYSRNRDGGLGLL KLERLIPSVRTKRIYRMSRSPDIWTRRMTSHSVSKSDWEMLWVQAGGE RGSAPVMGAVEAAPTDVERSPDDYPDWRREENLAWSALRVQGVGADQ FRGDRTSSSWIAEPASVGFAQRHWLAALALRAGVYPTREFLARGKEKSG AACRRCPARLESCSHILGQCPFVQANRIARHNKVCVLLATEAERFGWTVI REFRLEDAAGGLKIPDLVCKKADTVLIVDVTVRYEMDGETLKRAASEKVK HYLPVGQQITDKVGGRCFKVMGFPVGARGKWPASNNTVLAELGVPAG RMRTFARLVSRRTLLYSLDILRDFMREPAGRGTRVALIPAATGAAN (SEQ ID NO: 1394)	GIDGYMFGYARASGSTSVSIQSSSMTEGETNERATPRASDSSSVSIQSSC VTEGECLPPTDNCNPSVENQLPCVTEGRFERVGSLVTVRLPFRKVACDLC SKEFLTYSKFAVHQANFHNSETQACCTYCGKSDGNHHSIACHVPKCPWR RTVTFAANLSNFLCDLCNDSFKTKSGLSQHKRHKHPCSRNAERILSLGVRT PSARPRQVVWSEEETRTLREVEVVYSGQKNINVLCAGHLPGKTSKQVSD KRRDLHRIRSSNVHGTPTTQSRGDPVEQVEEYEELDWEGMHPFPDPDS KFCSYLDQLRDQKGLTEPVWQEIEIVAQEWVENLAHVQSSWNNHERTTK QVPENNTPARRPFKRRLHRVERYKRFQRMYDLQRKRLAEEILDGREAVT GAAATATCGA (SEQ ID NO: 1273)	CNLKKEEIKDHYDQVYGVSNDRVSLDDCPRPPGANNTDLLKPFTPTEVM DSLQGMKNGAPGPDKITLPFLQKRLKNGIHVSLANVFNLWQFSGRIPEC MKSNRSVLIPKGKSNLRDVRNWRPITISSIVLRLYTRILARRLERAVQINPR QRGFVPQAGCRDNIFLLQSAMRRAKRKGTLALGLLDLSKAFDTVGHKHL LTSLERFAVHPHFVRIVEDMYSGCSTSFRVGSQSTRPIVLMRGVKQGDP MSPILFNIALDPLLRQLEEESRGFMFREGQAPVSSLAYADDMALLAKDHA SLQSMLGTVDKFCSGNGLGLNIAKSAGLLIRGANKTFTVNDCPSWLVNG
R2	R2_LP	AF015814	Limulus polyphemus	TGGGAGGAGACCC AAACTATCCTAGGA TGGGGCGGAACCG ACCATATGAGCCAT ATTAACATTGCCCA CACTATCCTCTGGA GGTACCTCCTCGTG GTACGGCTGGATAT AGGTAAATCCTGTA ACCAAATCCTCCAA CCCGTGAAGGAGA ACACTAAAACCCAT ATAGTGGCCTCGCC AACCACTATATGTC CAACGGCAGGAGA	ATTTTGTCTCTTTCCCCAATG ATGTCTACTAGCACGCTGCC GAAGCTAGATAGATTGAGG AATCTGCGTAATCTGTAATG ATTACGCCTCATGGGCATCT ATCGGTAGCGTCGACCCTG ACGTTAAATTGGGTAATAA GAAATATCGA (SEQ ID NO: 1273)	

			AGCTATCTCCCGGA		ETLPMIGPEQTYRYLGASICPWTGINSGPVKPTLEKWIANITESPLKPHQR
			TGGGAAGGAAAAAC		VDILCKYALPRLFYQLELGTLNFKELKELDSMVKQAVKRWCHLPACTADG
			CCTAAACCGTGATG		LLYSRHRDGGLAVVKLESLVPCLKIKTNLRLVHSTDPVISSLAESDGLVGAIE
			GGAACTTACCGGCC		GIAQKAGLPIPTPDQRSGTYHSNWRDMERRSWERLALHGQGVELFKGS
			CCATCAGCTATTGG		RSANHWLPRPVGMKPHHWVKCLAMRANVYPTKRGLSRGNLSKNKDS
			GTACCCGGTAGGG		AKCRGCTSMRETLCHLSGQCPKLKSMRIRRHNKICEHLIAEASFKGWKVL
			ACTTGCAACCCTAC		QEPTLVTDNGERRRPDLIFHRDDKAVVVDVTVRYEISKDTLREAYASKVR
			CCTGTATTTGCATTT		RYGCLTEQIKDLTGATSVVFHGFPMGARGAWFPESSDVMADLNIRSKYF
			TATAGGGAACCGGT		EEFLCRRTILYTLDLLWKSNNEQYLERLAP (SEQ ID NO: 1395)
			CGGCCCTATATCAG		
			AGTAGACCGTTTAT		
			TAAATATGGGTGAA		
			AATATTAACAGTAA		
			AAGCTATGGTTTGG		
			CGTCCGTGTGGTGC		
			CAGGGCGGCGGCC		
			AAACCCGAGCTACT		
			TGGCACCAACTGGG		
			GATGGTAGCTTCCG		
			AGCGATTCCCTGGC		
			GACGTGGGACCGA		
			TCGACGATGGAGTC		
			CAAACATCCGGAAT		
			AGAGGAATTGAGA		
			AATACCTATTCCACC		
			ACCGGCTCACATAC		
			CCAAGGTGAACCCG		
			GTGCAACTAGAGTA		
			CAACCTATCTGTGG		
			CGGTAGGTGCCGA		
			ACCACTCAGGTGAC		
			GGGCTTGTTTATTG		
			ATGTCTCCCTACGA		
			GACACGAATTGTGA		
			CAAATCCACTCCGG		
			TGGACAATTACCCG		
			ATCTATGAACCTGT		
			TACCGATATTAGAC		

NeSL	NeSL-1	Z8205 8	Caenorhabditis elegans	AAGAAAATAAAGA ACTGACAACGCCTA GAGCTTCAGGCAGC ATGTCTGTAAGTAT CCAGTCATCGAGCG TGACTGAGGGGCGA AATTGATAATAACT CTGAAACTGA (SEQ ID NO: 1150)	GCTCACTTTCTATCG TGTTAACCGTACGT TTACACTCCCAGTG AGTGTAATAAAGGT TATTCGATAGAGGG TGTCTCCCTCTTTCT TGGGTAATTCTTCG GCGGTCCGGGGTCT CTCCCTCGTCTTTTT TTTAAACTTTTCTTT CTCATCCACTCTTTT GCTCCTTTTTACTAA CTCTTGTACTCTATA GTCTTTTCTCATCCC CCATCCGCCGTTGG GCAAAGTTTATTTA CTTTGTTAAATCCAT ATTTTATCTCTCTCA CCCGTACAGAAAGC GTCTCCTTCTCAAAC GCTTTTCTGTACTTT TTCTTATATTTTCAT TAACATATTTTTCCT GTTTATACTAACCTA ACCTCCATTGTCAAT TACTAACTAACTTGT ACAACGGATTTCG (SEQ ID NO: 1151)	CCTCCAGGGCACGCCGCAC GCCAAAAGTCCTGGCATAA CTCTGCAAATAACATCAAAC GTCAATCAACTCCACAAACT CTCCACTCTCTTCAAGTCTTC TCGGTGCTTCCAACACCACA ATGGTGAAAGCTCCTTCACC TTTTCCCTCCAAAATTCTTCC CATGTGGGGAAGTCCTGTT CTTGTAAGCTCTCCGGAGG CTGCAAGAGCAGAAGAAAT TCTTCTTTCTGACAAGGTCA GAAGGAAGTCCTGTTCTTG AGGCGTCCATCCCGGGCGT CATAGGAGAGATCAGATGC ACCTTCTAGCAGGAGCTAG AAGGGCTGCCCTGTCTTGA GATCCCCACGGGGGTCAAT AGACGGGAGGGGCTGCTG GCTTTCTCTTTTTAAGAGGA AGCACCAATCCGGAGATCC TTAGGGGTCAAAGGATTAA AAGGCAGCAGGTCCAATTC TCCTCACTGACTTCGGTCAG AGAGGAGTCCCGCCTTGGA GACCTCCCCGGGGAGGTTG CTGAAGAGGCGGAAGCTCC TTCTAGCAAGAGCTAGAGG GAGTTCCCAGTCCTGAAAC	MLRRKGRHRMVMVNSVKWQPSAHAEAIGTGKSWAPQRSQASEHGW QSNAMFDPPNRILFARDSWSLNQSTHLQNQRSGSGLGIRPGQVRNNM VGGGPHRAGDPKRRVELVSIQGSEVTVRTIYPSDEIFSCYSKSCDIKTKAG YGPEDLKHLTRHIKNEHGLKARWAYQCGLCNEKSDPSVSEGHKWMEAH MVAVHQSSAEKRIKSYQKCTGARVAEQLQAAAPSLTVPGKHKSGSRDA AKDSMTPTKDDDPKTRIYQTRSVVKKSTQKTAEPTDEGSRGPKYASIFQK SVKARKSLALLCELSSPKPMNPLPTNELTLKEGNSRELAKEEAPSEGIDDIV IIDLDESEESPPRRKRFNTWCLDHESSREAWLDDTAIFWYISYLCRGSTKY SALDPCLWSMYKVKGSRYILDRLESSITYFFPICEEDHWTLLVLKDNSYYA NSLHQEPRGPVRDFINDSKRARKEFKVQVPLQRDSFNCGVHICLMTNSI MAGGKWHSEEDVRNFRKRLKKTLQEEGYELYSVNSLGIPFQAPTTEQM DYKETRCKRSYASVLTQISPPAKRPDCKPDNNIFVPTKDCAAEGNPQEKG RNESPEEINTEHIVVAGKPANNISPRCRSTSEMLFEMVKATTSSGRSSLGT MTQDEFIRTSTIAEAVPLMSIKLPPMELPRKILPPIPPRKPTQTNGGQKGK QQRVPTGKPDTLNAKVRNWFNNQLESYAMEGRSFQRLEWLTEVLTASI QKAAAGDEGIVDIICKRNPPLEVAKGEMCTQTENKRKTTNNAARIADPI QSSKGAGDVKASYWKERARTYNRIIGSKEELCKIPIDQLEDFFKKSTSRTN VQESIMKEKSSKIPALKIGNWMEKKFIGKEVAFALRKTKDTAQGADGLRY HHLQWFDPSGELLAKVYNECQRHRKIPKHWKEAETILLFKNGDQSKPEN WRPISLMPVIYKLYSSLWNRRIRAVPNVLSKCQRGFQEREGCNESLAILRT AIDVAKGKRRNLAVAWLDLTNAFGSIPHELIEYALTAYGFPQMVVDVVK DMYQGASMRVKNATEKSDRIPIMSGVKQGDPISPTLFNICLETVIRRHLE SANGHQCLKTRIKVLAFADDMAILTDSPDQLQRELSKLDNDCTPLNLIFKP AKCASLVIQKGVVRSASIKLKGNAIRCLDENTTYKYLGVQTGSAARISAMD LLEKVTKELECVVKSDLTPPQKLDCLKTFTLSKLTYMYGNSIPLITEIKMFAN IVIRGVKVMHRIPVRGSPLEYIHLPVKDGGLGVACPKTTCMITFLVSTLKKL WSDDEYIKTLFTSLAEEVVKKESKKSTVTMDDIADYLNVEERINRSEFGYN SITRLRDVMRNLAITGDSPLYRLKMVVKNGKIALLVQATSESMERIYTEED AKKLQRSLKDQVNKALKHRFNTTKVVKSKVVRVVQQHPASNRFVTKGG

					NLSLACHRFVHKARLNLLACNYNNYDKSKSKVCRRCGKDLETQWHILQN
					CPFGFSKKITERHDAVLHKVKTLIESGGKKNWTMKIDEELPGFSRLRPDIC
					LKSPDEKQIILADVACPYEHGVEAMERSWQAKIDKYETGFAHLRKSGTKL
					TVLPIIIGSLGSWWKPTGDSLKELGIKGSVINSAIPELCATVLEHSKNTYWN
					HIFGEAYIPNPMRNGHAKPAGNGWKKERLQKAPVRPTN (SEQ ID NO: 1396)
				CCTTGCGGTTGATGATGGA	
				ATGGAAGAGTACTTCGGTA	
				CTGCTCGTTGCTCTCTCTGC	
				GTTTTACTGCCGAGGGCCG	
				GATTTGCTCGAATCGCGAA	
				AGGTCTCAATCGACCATTCA	
				AGATGACGGCTTATCTAAG	
				GTCCGAAAGCAGTTGGGAG	
				AGTAACGTGTTCTCCTACCT	
				TTCAAGTTGAATGGTCGTTT	
				TACTGTTTGGGATAGCTGA	
				CTTGATGCTAGTACGCTTCA	
				TCTGTGGATGACGCTCCCCA	
				AGCAGTCAAGTAGACTTGA	
				AAGGTGCCCTCGCCCTAGTT	
				AGCTCTTAGACCTTATGGGT	
				CGCCATGGTTGTGGACGGG	
				TATGCTTGCCGGAGCCGAG	
				TCGTGTTTCTTAGAACCAAC	
				CTCGACGAGGCGAAAGCTT	
				GCACAAGTTAGCACAATTG	
				TGGTAGGGCCGACTAGAAA	
				ATGAGTCCCTTAGGGGGTT	
				ACGCCTTGGCGAAAGTGAG	
				GACAATTGGCATTGACGGG	
				TGCTTCGGCACTAGGCAAA	
				GGCGCCACCACACTGTCCA	
				ATCTCTAAAAAGTTCACATT	
				CATCGAAGAACTACCGGAA	
				CCAACCACACATGTGTTGAA	
				ACCTACACGGTGGAAGGGA	
				AAGGAAAGCTTCGCTGGAA	
				CGAAAAGAACGGATAGGTT	
				CCCCTTCTTGATGGCTGTGA	
				GGCTTAGGATGGACGGGA	
				AGGCCGTGAGGCCTCAGGC	
				GGGTAACTCGGCCAGACGC	
				TAGTTGATCTTCGGATCACG	

				ACAGCCCTGGCTAAGAGGA ACCCTGGATGGAGTGTGAA GGATGGGCGGGTAGGGGG TTAAGCCTGTTGACAGACC ACCGACTGCAGTCACAAAA TCAGTGATTATGCGGGTGG ACCAATCTGTTGGCGGGTG TTTCCCTCTACCTGACCCCG CAATATGGTATGTACGATCC TCGGATCTAAAATTCATAAT GGCCCACCACAACCATAAA CCTCCCTAGCAGCTGGTGG TCCCGATAATTCGGGTTCTT GCCACTACTGCGACCCAGG CTCGCC (SEQ ID NO: 1274)	MSLQRAKNARGDPGRCNLCSADYRDLKDHLNKQHSTHFFVPSDLRGSS LVACPRCGTPCSAGTGLSRHQSRYCGLTAPRIRRNRVGNSTNTSRCPPSN TAASPIVPSPSPERPSPPQPAEVVASLEPLSEAEEVLEVAQVDAETVDTLE GTRRAPESVPRSAEEGSTRVRELNMTAPEEEHRGEEESSHTNPTAPAGLE NAVSSTLGPSPGTLPSLLPSQECANERFLYLAHLPVRSKPLPNNLVTDFM DAAERCALAYIAQPSDSTLLAFLALPKVGLTQALAPEQPLRPSTFLKQFPHI PWPEQPPARRPPSNIRPDTTKQVIKLVENGRLGAAERVLEEDASVAELD QGVIDQLITKHPKGPSCPFGNAVGPTPGKAPDIDTIQKALDSFKPDTAPG VSGWSVPLLKTAAKREPVKQFLQLLCAAIANNTAPGRSMLRTSRLIPLKK DDGSIRPIAVGELIYRLCAKALIISHFQPDFLLPFQLGVKSIGGVEPIVRLTER VLEGSAGAEFSFLASLDASNAFNRVDRAEMAAAVKTHAPTLWRTCKWA YGDSSDLVCGDKILQSSQGVRQGDPFGPLFFSITLRPTLNAILSQSLGPSTQ ALAYLDDIYLFSNDSQVLSKTTQFLADKQHIIKLNEKKCKLISFDEIRQEGFK MLGTMVGGKEKRAEFLEGRIRKEMAKVGKLKDLPHQHALLLLRFCIQQN LRHLQRSLRSDDLVDLWERLDTMLWEEVKRMRMRQREDTAEEEALGR SLTKLPARLGGLGLLSFKDVAPLAYRSAAEASDTLLDNLGLLSSPEEPPTPIP QRTRCAELWESQQEAILHNLGDTERKRLTENASRLGRSWLSVIPYLQPLR LSNVEIASGLHDRTLVGSSIPVCRFCGSDSPLGHDELCRARNPWTQRRHN AINRVIYQHLKQIQGATVEIEPHTLSGQRRNDLRVRGSSALAFTDYDLKVY SLGDRDARSTVTPCAPNGKLADFCLDRCVNWLDKVGQVVSKNAPKVTG GVFKPIILSTGGLMSRSTADEWKDWRDAMPVGGFEKMEKRIGVELVKA RARTLVL (SEQ ID NO: 1397)
CRE	Cnl1	.	Cryptococcus neoformans	CCCTCTTAATACCCC ATAACACATAACAA CCCCCTAATCAACG TTCTCTGCACCTTAA ACACCACCAAC (SEQ ID NO: 1152)	TGAGGAAGAGGAGGTTGG ATTATTTTTTCTTTTCTTTAA TAAGTTGTTTATTTAAGTAG TTTCTTTCATTCGGGCAACC CACACGACAACCCAATAAA TTAAACAACGAAAAATGCA ACCTCTATAACCC (SEQ ID NO: 1275)	

CRE	CRE-12_CCri	.	Chondrus crispus	ACGCCCCCTATCCA TTTCTGCCAGCCTCC CATCGGCTCGCCGT CTCCGCAACCCCTCT TCCTCGGCTGTACC AGTTCCGCTCCCAC AACCTCCCTCGCCA CA (SEQ ID NO: 1153)	TAAGTCCTTGACGCCTGCCC CGTGATACAGCATCGGTAC CCCTAGCATTTGAATAAAAA A (SEQ ID NO: 1276)	MSQPNISSAETPLSQLPTPVPTPPSPSNPSLSLPTVRDLLLCPIRSSHVYSSI PSSCLHSFTMLLIKTVRAASATMTPTESHRAFIHLHILPIAVLRRSFRGETG WRSRTGQHHALRQRIRRASSGRHWAALWHEALAAHQVDLDYRTRHSR RYQASATSRHRIGRAMRLAADAQYGRAMSALKAKPLPDLHAAATRDTL TALHPPPASPVQPLSPTDLPPVPEITEGQVLRAARALNPTSAAGPDHLSP RILQLLARTTISPEAGVTGLSALTNLVRRLARGDIPDRTAPLLAAATLIPLQP RPHKIRPIAVGQALRRLVTKVLLPPAIQDTRDHLLPEQLANSVASGMDAI VHDTRMLMHRHGRNPDYIMVSVDARNAFNTFSRQSLLDRLPLQTPSLA RFLNLIYGRTVPDLVLPSSPRFLMKSQEGTQQGDPASMLLFSLAIQPLLRR LTRECRLDLNRWYADDGTLVGPISEVIKALRILRDDGPQSGFHVNINKCR AYWPTVMPEKLSELLRIFPLHVECGEGGVALLGAPLGTDAFVRRHLMNK VQSCHASLSLLDEIPDARTRFHLHRVTGSVCKVEHVFRLTPPHLSLPAATK FDEQQIAAYSRLNDVAVSTSMATQIGLPFRLGGHGFTPLSPFIHASYAASL IEAAPVRVKGPHNPSESFYRRMARRHIVHVLGALNPEVRTRGILGTHSPL GPFEPEALLSRPERVHHTLIQAMQGATSRLYWEHTAWDLDPLPRNHSA ASVRRRARYNSLRAPGAASFLCSHPSLTSRVPSAVWSCMLRRHLDTPVY CDSIRPLICSHCCKPMDARGDHAAICRHGFGVVHRHNTVRNLLARHAFR AAGLCCDLEVPSLLPNTANRPADILVQPAPPPSGALPDRPTAYDVTVRSP YCRSTMSLAAKGLAGAAEAADLDKLRVHSRTVRDAFHLQPDSPLPLLDW HFVPLAFDTLGATSSRTMAVLEYLAHRIANRTYSSYGTAKIRLLQRISFAV WSSLASATLSRMPYHGAALSSPAQV (SEQ ID NO: 1398)
CRE	CRE-13_CCri	.	Chondrus crispus	CNCCAGCCAMCGA TCCCGCCGCCACTC GCMGCCCGGCCGT CTCGACCGCCACCT CCCCGASGCCCCAG CCCATC (SEQ ID NO: 1154)	TAATTCACCTTCATATCTGC TAGTGTCTCTGTAAGCGCAC CCCTCATGCATTGATAAAAT TACCCCCCA (SEQ ID NO: 1277)	MAXXPXISPPGAPPAPLRYRMLQCPPPLPKXXXXPVPHPMSSPIRXRLPH RXMRGPPSXTPPRDMHRPHGTPGPHSHRXCGRPPXHCTHASXQPRXA XHXLQXPKLRSPPPHPHVSPLILCXGPLPTPMTQPRMKRALSXSAKAPPT KRPSASQGPAASSHDXPRTPPPXPPRPPPYRFPPPPTLDQHXFALSXAYPH PXPRRPPSPXRXLRHSFPPPRFGXQTFSSIPGPRLHSTVLLLIRLVRAATAAN TPETTTLXSCTFTCSRLPFFERPSXAXLAGGPRAVNFMLSACXYGERVXDE SGXSYXXKXHCITSHPPRPRRYSRQHSRNHHTXFLHLHLLPTAXLREAFRG EXGWRSSRGQLHALRLXIRRACTGREWGLLXXEALDAHSXRTEWQHTH ARRPSPPVSPSARAARAMRLASQAQYGRAMRTFTNPPLADLNDPATM ERLQALHPTPTVPVVPLPPSAQPRPPEVTXEAVXRAVRRLNPNSAAGPD RMSPKLLHLLAHTPISPEAGVTGLSALTNLVSRLARGSLPPCTIPLASAATL LPLQPRPGKIRPIAIGQALRRLVTKXLLPAAIDDCRDHLAPEQXANGIPNGI DAIVHDARMLVRRHGNDPHYXMVSIDASNAFNNFSRQQVLDQLPTRA PSLSRYLDMVYARAPSPLVLPSXPPTILHSRXGSQQGDPASMLLFSLALQP LTRLISRECXLXMNRWYADDGTIIGRIDEVXKALDIITKEGPRFQFFLNPSK TRVFWPSRQXDLLSPLMTVGPLRVIDEGGVXLLGAPIGSPSXMAQYIREX LNTCKTALAHLDHIPEARMRFHLHRVSASACRLQHLFRLVPPDFAXPFAQ

						QFDRDQLXAYXRFNSVTMSPRIVPKYGCXFXTXATASPHWHLPYTPXTLL
						ASSIPLQHGYKVPTFPPSLSISVLHEARCGSFFEIYLHSXNPHTSRCDYVAK
						NRAQIRLXFSHGGHGLTSLASTIHASYAASLIDTAPARLQGPHFPAVSQY
						QRFARGPLRVVLRNLPSFVQPAHFSMTEXDLGCLEPXALLARPERIHTFLL
						QAQYSAAASSYWQXPLWESFPNPGDHSAASLRKRVRYNSLLAPGATSFL
						TAHPAATSRVHNATWSTMLRRHLDAPVTNDSISPLRCXHCSKPMDARG
						DHAXIXSHGFGTLHRHNTVRNVLARQLFRVAGLAYSLEVPFLIPNTAARP
						ADILVQPPPPAPGLPPDXPTAYDVTICSPFRRGMLYHAARHRGGAADAA
						SVRKXKALERTIRXALLIEDDNXPPPLDWHFQPLSFDALGAPSQSTVHVIE
						DHAKLMALRNSCTIATAKSRIQQRLSFAIWSSAAAAILSRLPTHAADISYPI
						EV (SEQ ID NO: 1399)
CRE	CRE-1_ACas	.	Acanthamoebacastellanii	TAACCCTAACCCTCT	TAAGCCGCGCGACGAGGAC	MATTTISRSPSSSSSSSSSARSRASASTSASVASIPRLFRDGRFHCPLAHCQT
				CCCTCGGCCCCTCT	GGCCAGGACGACCAGGAC	RTSTWQDLSAHLTRMHDGDVPRDVAAACGIVQCLHEGCRKWFRGAA
				ACCCTAAAGCGCCC	GACGGCGACGGCGACCACC	GLASHRGKARHAPPPAPRAALAVAAVPRADSRGRTPAPTPSVAPPXAGP
				TAATCGACCGGCGA	TAGCACCGCACGCGCCACG	PPRAAPRAAPSPLPCPPALPHPPPSASPPTSSVTSPCSPPTTPPSQPSPDLF
				CGCCCTAATCGCTA	ACATATTGTCGCGCGCTGTA	SGFANAPTTPSPPSTPXSSPAGSPIPAARRFVLPVATPYPAPAPRANRPKL
				CCCTCTACGCCCTA	CAGGCGGCTAGGTCGAGCC	SPVARPFVPKARAGAIPEASSPVTPQDRAVSRRREDAAAAPSSAPGLGLAD
				ATCGACTTTGGCGC	CAGCCGACCGTTCTGAGCC	EHEDDDTYGGDTIALTAPHAPRETRAPFEFEACFLEEEAPATAGDLPPYA
				CAAAGCGACTTTCC	TCAGTCGGCTTGAGCCCCC	RAFLACPSARLQEIPRRLKSAWQAAAKTIAEAALDCHTAGDTQGYNAHL
				CCGGCCGATTTTCT	GGCTTCCCAAGGCCTACCG	RLFIELPARGLAVPTNCRGAARTKLQRERLLDIAAGRIPAIPDPPCDAPGA
				TCCTGCCTTTTTCTT	GGGCGGCTCTTTTTCGCCCT	DDALRGFPVSGTTAGDVSNDDDSGGVHDRPAATASARQAKRLVEQGLS
				TTCTCTCCAAGCGA	GGTTTTTGCCGGCCTGTTTT	SRALRALERGEPAVASADTLGRLEALHPPNPTDRGLWPGAPKAAIPRVT
				CGCGCCTTTTACTTT	TTCTCTCCCCCTTTTCCCCCC	AKHLAQVAKELPRGSAPGPSGWTFELVQAAIDRQPTGTVAAFLIDMAQ
				GCCGCCGTTCTGTT	TTTTCCATTTGTACTTAGTTT	RALRGTLHWRGLLTASRLVALKKPDGGVRPIAVGEALYRVIGRLVLKADR
				TTTTCTTTTCTCTTT	TTCCTTCGGCCGCGGCAGCT	VMSSADATQVVGRHQYGVAYPGGVEAPVHAVRELHDSGQLRAVVSLD
				GCACTTCGCTTCTAC	TGTTGCCCGGCATAGTGTTA	WRNAFNSLDRVHTALLIADRAPALARLYEWSYREDSVLVLPRAFEKAGLP
				TTCACACCTCCTCCT	ATATGTTTAAAAAACGTGTA	ASLLSQAGVRQGDVLGPLFFAIGAAPVLDEIDAIPYVTPRAYLDDIFVTIPH
				CCTCCTTCTCGACCC	AATAAATAACTGTTTAACCC	GVTDAATKAAVAATFATAEREGAAAGLRLNRCKSAVWAADAEALLPPH
				GCGCGGCCTCGAGC	TAACCCTAACCCTAA (SEQ	AAGAREDVESCAPVREGLKILGAPVGSPAFVAKSLDGIIKRAIGTLDLVAD
				GACTTGCTGCAGCG	ID NO: 1278)	AELPLQHKLVLLRQCVAQIPTFWARAVPDAGPALAVWDTALLRRTGALV
				GCTCCCGGCCTCCC		GLDVRDGSLQADIARLPVRLGGLGLRSMKDTAPRAFVASILFAAALANTR
				CCACGCGGCCTGCT		RSELTCSASTARRLRAALPELARTDACNDEAAWRRSIARGVFPDVDKLGT
				ACTCCCGCTTTCTAG		TQLQRVLQGMADSKSAHRTRRQVPFLFAAVFEDAATPGSGAWLAAIPS
				ACGCCCCCGGTCTT		DPTLVLPDAELAEAVRIKLLTTTANAAGVCPACHKTGIDPSHAYTCVSLSH
				GCTCTCAGTCTCCC		LRTARHDVVVRRVELACKTEKPVREHVLAIPPVAPTDNNNNGDEDGSPV
				GCATCGAAGCGGTA		TTADDNADGHAVATKRRPETRASARAAAAAATAAAAAAIINDNSLLSDD
				GTCGGGGTACGTG		DDDDDHDDNCHGEERGEGERNVTCPGHYTATPFAADDTLDNSDEDNE
				CTCAAGTGACTCAA		DNAHEDDDEDGKDDNDDDVYNNCNSSSSDGDEGGDDLDYEYSDQSVT

						RSVDAATGESPNPERPTTPTRALLRADLWLPATSTAVDVMVAAACRRSR AKAFDRAVSRKAAKYGPAVADGSIAKVVPFVVSPFGVLSRPAKAFLKRA MGDTTAAKQAKARLRLAVAAVRGTARLSYAWGACAALIVGGN (SEQ ID NO: 1400)
				GCCTCTTTTCAGCCT	TAGCACCACCATCTATTCAT	MAPLPWNAATSSPPSPVPLTNDKKKDSTLPTATSKNLSKNNNNKNNNT
				CGGCGCTCTCTCAA	ATCCACACACTGACCACCTC	NRINNIKNNDNTNDGSNKINLKLPPAAVKITNPYKNKKKNKKKNNAGKS
				TCCGCCTCAGTCTTA	CACCTTCACAACTCCACTCT	NPKTNQNPNSSPLSDNDDDDTDSSNITINRRLKFGTDDLAPPNPPSNTN
				GCCTTTCAAGTTGC	CAATTCCCCTGACTACTAAG	TIGTATAATAATATTTATAATATTATNTTTTTTTTTNNTTGDNLASNINNN
				TCGATTACGCTCTC	AAATATTTCATGGTGGTTAC	NNNNNSGSNNSNTNNINNTDGNGSNNRPPPRVVTVDPRSDLPGAEISA
				GAATCGCTCTCTCTC	ATTGGAGGTATCCCAACAC	ANKMLDEVYGDHVHDNPGSHLSGLISSSQDQLWQGYFRRLIPHNQSLY
				TCAGTCTCAGTCTC	CAAGCACAAAATGAACCCA	DCPKGKLGKDITNEYSNLFEAIMNGKCNMEKLLVFPVVVLQRRHGVTKN
				AGTCTCAATCTCGA	CTAACCCTCTCATCCTATCC	ADVKRRLLSRLTAWKEGKFKYLVEDTHRDLIAKQSKARGDTTPAHRAKVY
				TCTTGCCTTCGCCTT	ACGGGGACCACCTTTGAGC	SSKLMRGHLQSAVNYITDREGGGILYPYDVDEKSGHTVSRVLQDKHPSM
				CGTCTCGACGCCTT	AGAACACCATTCATATTACA	RDPGPTAMPAYESVPELPTLEITADTVEIVAGKLSGGAGLSGVDSIQLKHL
				GCTCTCGWAATCG	ACCTTTAGCTAGATTAAGAT	LLHHGQASQRLRNVCAKFGRWLANEHPPWASYRAMLANRLIALDKMP
				CTGCCACTACGTGC	AATTATTTAGTACATATTTT	GIRPVGIGDTWRRFFAKLVLAVSMSYATDCCGSDQLCAGLRAGVDGAIH
				CAGCTTTTTCGTGC		GLSAMWREMESEENTGFVLIDADNAFNEVSRINMLWTIRHEWPAGAR
				CTTGTCTTCGTGTC		
				GACCCGGACCGTTT		
				GCAAGCCCTCGCCT		
				TCGTACCCCGCTCT		
				CGTAGCCGTTCTCA		
				TCGCTGAAGCGTTC		
				TACGCGCTGGCAGC		
				AAGCCTCGGCCCTA		
				GCTTGTAGCGCCGC		
				CGGTGGCCGCTCGC CAAC (SEQ ID NO: 1155)		
CRE	Cre-1_FCy	.	Fragilariopsis cylindrus	ATCAATCTAATACT GAAGGCAATACCAA ACTCAACCCGAAAT CAAAATCGTTAGAA TCAATATACGACCC CCGCTGCTGTACAT GTCCAGCCGGATCT CGTTGTAAAGAAGA TTGCAGCTGTAAAA AAGTTGGACTTCTT TGTTCTTCCTGTGA AGAAGTTGATTGTG GCTGTTCAAATTCTT		

			TTCATAATAAAGAA TTA (SEQ ID NO: 1156)	ATACTATTAAAAAAAAAAA AAAAA (SEQ ID NO: 1279)		FAFNCYRHHSLLVVRNPGGKPFTFFSKEGVTQGDPFAMIAYGVALLPLIR KLKELNVLLVQSWYADDASAAGKFDEILRLFQDLLRMGPDFGYFPNASK SILITHPDNVVAAHHFFNETHGLGFKISTGSRFLGGFIGDTTSRDEYVSTKI ADWIHGTKELAAVARLKYPHAAYTGITKCLQHKWSFTQRVIPGIDDLFQP LEDELTNNLLPALFGDPPSTMDDKLRLLTALPVKHAGLALPNPVTSSATN YKNSTLMSSHLLLAVQGKINFSLQDHRDTCQSSLSASRELRQTENDSSLT NLLAALPPAAAGGQPSTTRAIKRAGETGLWLTTIPNHINGNILGCDEFIDAI RLRYQKVPHNLPAKCDGCGSAFDVGHALQCKSGGLIIRRHDELNLELASL AKMALRESAIRAEPEINPSASIMDSPTTITAIDTNGDRGDLLIKGFWDNG MDAIIDVRITDTDAKSYRTRDPKKVLQSQEKEKKKKYLDQCLLQRRAFTPF VVSVDGLIGYEASNVLKQLSKRLADKWNKPYSVTCGIVRSRISIACARASN QCLRGSRIPFKTMSRQIQWEDGAGAGLYRIVR (SEQ ID NO: 1401)
CRE	Cre-1_HM	.	Hydra vulgaris	TAACTTGTATTTTTAAATTG TTTTATTAGTTT (SEQ ID NO: 1280)		MNMVSICKRCDRSFTTLKGLNIHKGQCKIFVSNTNKQINNVVNNELTTP NKNKVEINTILNCDEISVEHYSTNTPYLPKINICESIIDPNDYLWGHMPFSF LLNHVNTIYDEIVFYHKNLFKVPSGKGGKMFIEELTFWLKQFNNRTKLNGI AMKCFMIVPSLMLQKPSIRSKAKEHAECLVRRITLWRNGNFSELMREIRY IQSKINTSKKKRTFEDISRIFAKLMMEGKVAAALKVLDRESSGILQCSESVL KELKSKHPDETPVQDNCLLYGPLQNTPECLFDSIDEISIFNSALQTKGSAGP SGMDADLYRRVLCSKCFGPSCKTLREEIATFTKNIATKSYQPDIVQPYIACR LIPLDKNPGIRPIGIGEVLRRIVGKTISHHCQKEIKEAAGPLQTCAGHGAGA EAAIHAMQKIFHQEDTDGVLLIDARNAFNCLNRSVALHNIQITCPILAMY LVNTYRKPAKLFIYGGETIFSKEGTTQGDPLAMPWYSLSTVTIINTLKLVIP DVKQVWLADDATAAGKLQSLKKWYKCLEDVGGLYGYYVNQSKCWLIVK SDNQAEEAKLIFGNSINITTQGKRHLGAALGSEAYKKVYCEDLVSKWSKEL NNLCEIATTQPQAAYSAFIKGYRSKFTYFLRTIEAFENFVTPVEKILSEKLLP VLFGTDCSIIKENRDLLALNPSEGGLGICNLITEAKEQHTASKKITNLHIKSIL DQSDVMKEKDDFGKTFSEIKTKTNMDKSKKKKEEVKKIHAGLPENLKLLV EQACDKGASSWLNTLPIKEQHLDLNKEEFKDALRLRYNVPLANLPSYCAC GEKFDELHAMSCKKGGFVCNRHDNIRDLLTVCLNKVCTDVQAEPHLIPLT NEKFNFKTANTNDEARLDIKAKGFWRKGETAFFDVRVTHVNSKSSKKQP TKHIFRRHEDAKKREYLERVLEVEHGTFTPLIFGTNGGFGDECKRFTALLA QKLSLKMGERYGAVINWLRTRLSMEITRASLLCLRGSRTPFRHYNTDDVG LENVQCGLI (SEQ ID NO: 1402)

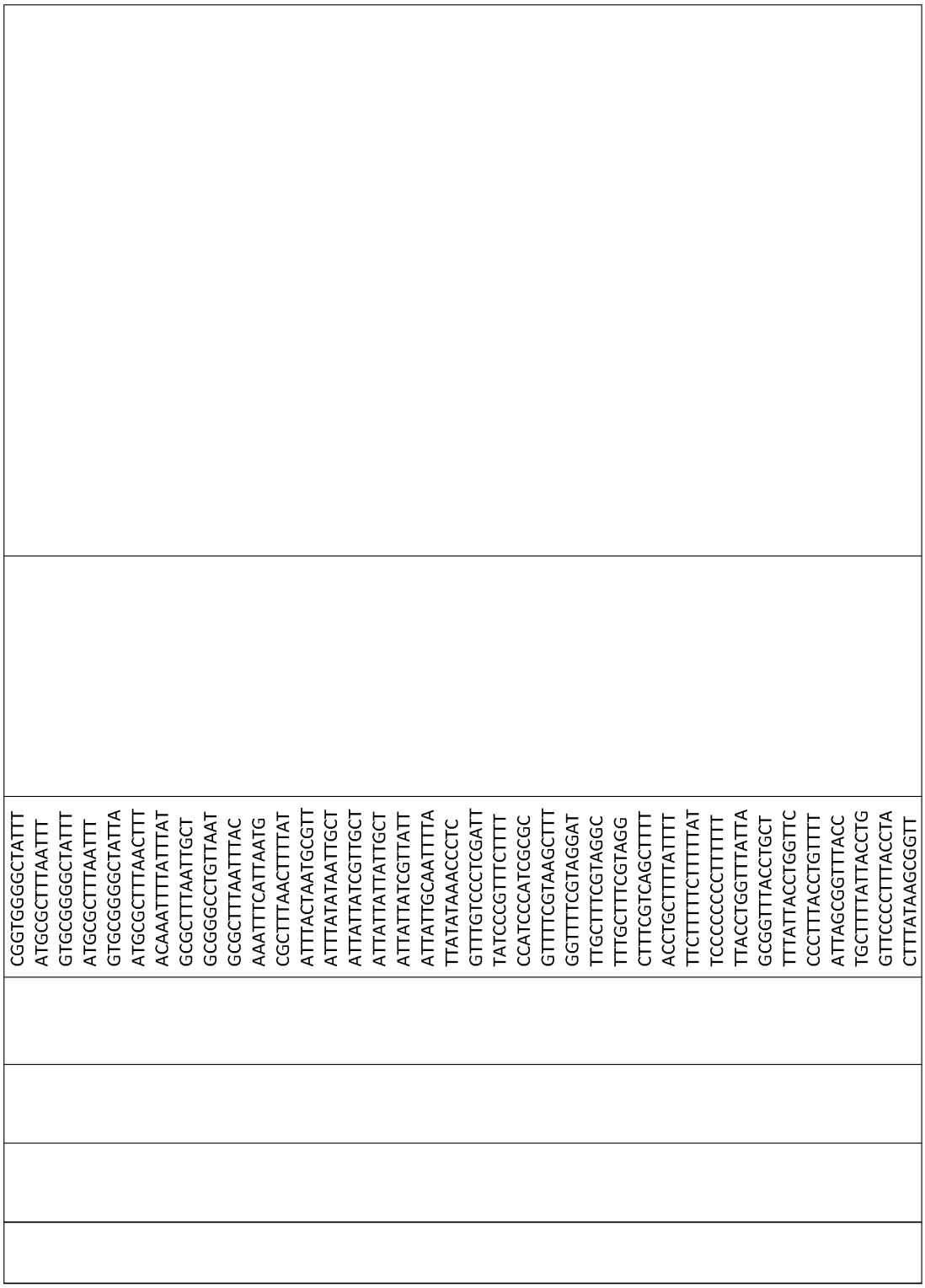
CRE	CRE-1_LSa	.	Lactuca sativa	ACATTAAATTAGAG AGGTTGATGTTTCA ATGGAAGAAGATG AAATTCCAAGAAGC	TGAACTATATTTTATATATT AAAAAAA (SEQ ID NO: 1281)	MASSSTSSSDICLCPFRSFHCCPNGEVGSKGIXRMISHIKRHHLLTEDRKC VLREALSSDVGLFMAVEETLKAFGQWMCGKCMTLHALSRYCHHPDGR VXFVTGADGSSRYIVGILKPSTKESVTNALGGLVFDVGLLDRVFKEPITTVK SIPHSCRLAFSQALKTALYKVIAQPGSVDAWICLLLLPRCTLQVFRPKNRQ

AAAGGAAGAAAAA
ATAAAAATAAAAAG
TAGTAAATCCTTTCA
TAACAATAATCATT
CTATTATTAAATTTA
AAGGAATATTTTGG
TTTTGTACTAAATCA
TGCGTTCATATTTCA
CCGAAGAAGGGGG
CTGCTATATTTTTGT
TTGAAGTTGTTTAT
CTTAAAACTTTAAA
CTTGTGTTCAACCA
ACCGTAAACATTAG
TTCGCTGTTCGCTC
AAATTATCTACAAT
ATAAAATTTATCAA
TCTTTTTTCGTTACG
GTAAACAATAAACA
ATAAAATAACTATA
GTTATTTTATTGTTT
ACCGCATATTGTTT
AACTATAGTTAAAC
AAAGTATTTGTTTAT
GGAACATTACCAGT
ATCTCTTGTTAAGG
TAAACAACAAAACA
TAGACGGCATCTCT
TTTTAAGGTAATTA
AGTATACGGCTAAT
AATAAAAATATACA
GCTAATAATAAAAT
CTTCA (SEQ ID NO: 1157)

				TATTTTTGTTGCCCA CCAAGTGTTTGATA AAATGTCCAAACTA ATTTTTCTCTTGTTG CAGCTTTATTGTTCA AGATAATGTAGTTT GCTTAGTTTGAGCG TTCCTTGTGCACACC AACAGTGTGTTGGT GTGCCATTTCCTTTC CTTCCTTTTTAACTA TTGCTTCATAGCTTA AGCTTCATCTCGAG GCTTGTTCTCTTGT (SEQ ID NO: 1158)	ECRSGNRKSLQQSSILKSLDTWGKEDGIRKLVQNMLDNPEVGAMGQG GGILQKESTSSNTNIRQCLRKVADGHFTAAVKVLCSSGVAPYNGDTIKAL EDKHPFRPPPSMPSPIISEPPLVADFDCVFGCIKSFPKGTSCGRDGLRAQH XLDALCGEGSAIATDLIRAITSVVNLWLAGRCPTILAEFVASAPLTPLIKPD NGIRPIAVGTIWRRLVSKVAMKGVGKEMAKYLNDFQFGVGVSGGAEVV LHSANRVLSEHHADGSLAMLTVDFSNAFNLVDRSALLHEVKRMCPSISL WVNFLYGQAARLYIGDQHIWSATGVQQGDPLGPLLFALVLHPLVHKIRD NCKLLLHAWYLDDGTVIGDSEEVARVLNIIRVNGPGLGLELNIKKTEIFWP SCDGRKLRADLFPTDIGRPSLGVKLLGGAVSRDAGFISGLAMKRAVNAV DLMGLLPQLCDPQSELLLLRSCMGIAKLFFGLRTCQPVHIEEAALFFDKGL RRSIEDMVVCGGPFFGDIQWRLASLPIRFGGLGLYSAYEVSSYAFVASRA QSWALQDHILRDSGICGMDSDYLCAMTRLRDTIPGFDCSGFTNKDTPPK SQKALACALFSKIVKDMEVDFDMTVRQKAVFECLRAPHAQDFLLTIPIDG LGQHMSPVEYRTILRYRLMIPLFPIDEICPVCRKACLDTFGEHAVHCRELP GFKYRHDVVRDVLFDACRRAGISAKKEAPVNFLTDPQDGRSTLRPADILV FGWVGGKHACVDLTGVSPLVGLRSGGFTAGHAALKAAACKVAKHENA CIENQHVFVPFAFDTFGFLAPEAVELLNRVQRVMHSNVISPRSTDVVFKR ISFAIQKGLAAQLVARLPSIDMY (SEQ ID NO: 1403)
CRE	Cre-1_MB	.	Monosiga brevicollis	CATCTTGGCGTGAA CCACGTTGTCAGAC AAAATCTGCAACCC CGCTCTTTGCGGCC CGCGTTTTGGCGGC GCCCTCGCTCCCAC CGTGTCCGCTCGCT TGCTCGCTTGCTTG CCCCGCGGGAC (SEQ ID NO: 1159)	TAGGTAGGCACCGTCTCGG GGGTCCCTCTGTGGGGATC CCTGTGTGCACCTGTCGCTC CCTAGGTGGTTCCTCGTTGT GTCTTTGATGGCTTGACTT GTATTTTTGTTTTAATTTTGC TTTAATTTTTGCTGTATTTGT GTGGTATTTTTGCTGAATTT TTGTAAGGTCCTTTGTATGA TGTCTTTGTCTTCTGTGGTC GGTTGTTTTCCTCAATCCGA CGTTGTGTCTCGTTGGATGT GAGCGTGCCGTGGTGTTCT TTGTGTTTGTGCTGTGATGG CTTGTAGTTGTGATGTGTGA CTGCCTTTTTGGGTGTCTTG TGTTTGAAATGGCCGTATCT CTGGTTATACTTGGTCGTTT TGTACGATTTTTGTTTCTAT GTGCGTGATTCTTCGCGCTT	MATESGGEDSWTQVRGAKRPSAESPPSNTTTSPSQTHRSAKHTKHGSA RHDRNHVFPDPMTTPLRPHARHSVPTARASSHVPSTSPAAGATESSARA VVPAAEPVTRTSNGGGEQHPIIGNTSNASPRTPRTPSSPRSFAQVAAAM PAAATATSSAPMTEDLSASVPSEPNGSGEQQPSPESTGQTHHSIPNTPS DFLTMSSDESDSPPRSTALRAPTPIAPPAHDGDGDTNGSATPEPLVQSPT PAQMVLPYPSGTQQTHSDPSPPSASPPATTILPAAISHPVEHSEHANSAP LGEVSESETHNTAGEHSESEQDVLLSDPAPPIAANVLDAQRKVLLKTSGH RQLLACPFGLCKCKGPRLDRKAWVNHVLREHPYDEQATDLVKQVMEAK LVAQCNKCHLFFEAAGISQHRSRCGANLKRATEALFHAAGHDLLEIMRG AWPQQCVGSRISVCELLKLARHPLMQRSRYPSNATETKLMAATLSQLY WSAVHSDYTAEEREMCWALILALPSMLLSAPSTALSTIDLRNMFHDRLR WLVTGQLGRVVDAMRKAVARKQSRRGQLNAGAGAHPNDAVDQSLRS LVRDPDLADEAWANHVTNRLNRGQIAKAFDADKARAVIGNSEVQAVR DLLVPPGLTPYIASTPASTSTLAPATAVSSPTVSFTKGELPKALAATKGVTD PYGWSGELLASIYRIKEHFSQVLGPRQGSTSDPTAPSDGDAPQGPTTATG GPQVALNKIFHHIANNTVPESIRHALCSINYTILEKANGKFRPVGTDSIFNK VVNRALLEQQQPHIAHLLQASPELAVGVKDGISAAVGMAFGELQACEST PGWTMLSLDFKSAFNYTDRARLHEIVADKVPGLLRAFERHYARPTTHCIV DKFFKVIDIDVGQGIVQGNELSPFFFALYSCEVLGLLDATTDYRCKVIKYLD DIVLMGPAEDVAADVEIVKARAESAGLHLQPSKSRFYMPRHHSASITAIK

				GTACTTCTTGGCATGATAGA AGCCAATGAATGTGTCTTGT TCTCTTGTGTTGTTTTGCGT GCCGTCGTGATTTTGATGTC GGGGTTGCACAGCTTTGCT TTCAGCTCTGAGGTTCAAAC ACCTAATTTA (SEQ ID NO: 1282)	SVLPDAVRETANTGMTVLGTPIGRREWMKKQLNDKAKHIAGKLNDML TTGVSLQALLTAMQVVPSLINHLYTLPPSLTSGLSELLNRACKDTFVKAFFA KVNLSAPAGAEGHDVTLEQLLEARLFTRANTGGFGLHDLVERGPVAYVC NMAKLATRYPRVYDRLLEDASRAADFEAHVQRAGFQMATVKDAATQR PAEIIALRSKAALDDLMAKCALDLQQAYLASREWGVSTVLTMRGRDKLR RLSDTTFAIAVVSMMGFGLHELINVKPTDKCPLCSSKTPQPRLTREHLLTC RPIKRHNALRDEMGRLLRYATLSHVWVEKSGYNANGQSCRIDLHCRNPF PGGALGPALPDLGIDVTVRTAQPPTTSQACIKVGAALRRAEKEKRDYYTG FNHGKTLIVPAAMTTTGGFASSFVDLLGQLARCAEARGVYQPGLDEAFV PRWKGRFAALVHQMNADHIQRHFGGVCLRSS (SEQ ID NO: 1404)
				TGAGCTCTTATAAATTTATA TTATAGCATTTTGTTTTA (SEQ ID NO: 1283)	MSSCKVTIPHVCPYCKVELKTICGINRHILKCKKNPLQIPSLQKTNTSLTLEP NTKVIPSITKQNDIIIASTSSNNLAFNQKKDYTLTPTYSRKTTPVSILSSMKM TPISITSHIVRRKLPELPSQTTNHLFNENFINVPFLPEIMNHLPVPNNNVM WGVYSYQQFKLFVDSTYDEIVNYRRNIFNIPSGKAGKEFIEELTFWLRKFN STSSLNSIALKVTMILPNLLLLQKPSAKSKSKEHTLCLTRRIDLWKKGDTSLLL KEVRNIQKKFVNSKXKRSMDDISRIFAKLIMEGKITAALKFLEKEASSGILPL SDNTLKDLKSKHPEPSRVEDYSLLFGPIDLIPKCFFDCIDEQLVMKAAFATK GSAGPSGMDADIYRRILCSKNFIKEGKELRKEIAKMTQNLLTETYEPTFLEA FTACRLIPLDKNPGIRPIGVGEVLRRIIGKVISWSFNSEIKEAAGPLQTCAG HGAGAEAAVHAMKEIFDNVQTDAILLIDAKNAFNCMNRQVALHNIQIC PLISIYLINTYRNPSRLFVAGGKEISSQEGTTQGDPLAMPWYSCNTTIIIEHL LVNYPQVKQVWLADDAAASGSIANLHSWYQHLIDEGCKHGYYVNQSKC WLIVKSPSLAENAGIVFGKSVNITTEGQRHLGSVIGSQNFKNKYCTEKVAK WLTELKQLCKVAETQPQAAFIAFTKGFRSKFTYFLRTIPKFEQYLAPVDEIL SHLLLPTLFGKDTPFEDHIRKLFTLTPRDGGLGIPILVEEAPHQFLSSVKLTK NLVQQIIDQDKILKTKNSSGNVLEDLEKILTTDRLKHRKEKIIAVDSMQPDS MLRNIQQTRSECASTWLNALPLENQGFVLNKEEFRDALCLRYNFDLKNIP RICECGEPFNVTHALSCKKGGFISSRHDNIRNLFTTLLKRVCINVQSEPHLI PLDNENFYFHTANKSNQARLDIKANGFWRNGQTAFFDVRVTHVNSMS NKNLDIAAIFRKHEKEKKREYGERVREVEHGSLTPLVFGTNGGMGKECH RFVRRLAEKLAEKQNEKYSVVMTWLRTKLSFEILRSTILCLRGSRTPWTKK NDFEIGVDFKMDALEARI (SEQ ID NO: 1405)
CRE	CRE-2_HMa	.	Hydra vulgaris	AATTTAAAAAAAAA AAATCGTTTATTTAT GGCATAATACTGTT TGTAATTTTTGAAA ATTCGTGCAACAAC TGCAGTTAAATTGA AGAGCTGAAATTTA AGATCTGAGCTTTT CAATCAGAGTTTTTT ACCCTAAAACATTA AATTTTATCATAACA AAAATCGTTCTAAT ATTATTAAACTTAA AGAAATTCGTTCTT ATATCAAATCTTATT TCAGTGTTTCACAG ACGAAGGGTTTTAC TAGATTTTTATTTTT TCAACTTTTGAATTT GTTTATTATAAAACT GTAAACTAGTGTGC AACCAACCGTAAAA AWTAGTTAGCTGTT CACCCAAAATATTA TTCAGTATGAAAAT ATTTAATCTTCTTTA TTCGCAGTAAACAA TAAAATATCTAGTT		

				AAACAAAATATTTC TTAATAATAAAAAA CAAAAACTTTTTCTT AACAAGTACA (SEQ ID NO: 1160)	TAATAGGTAACGTCCCTATT TTTGTCTTTGGTTTTGTTTTT ATCTTTGTTTTTGTTTTTGTT TTCGTTTTTGTTTTTGTTTTC GTTTTTGTTTTTTTTTTTGTT TTTGTTTTTGTTTTTGCCTTT GTTTTTGTTTTTATCTTTATT TTTGTTTTTGTTTTTACTTTG TTTTATTTGTTTTATATTTAC CTTTTGATTTTTTCTATTTTT CCCACCCTTATTATTATAAC CCCAACCTACTAATATTTTTT CTTTTTTCTTTTTTCTTTTTAC GGTTTTATTTTCCCGTTTGTT TTTTCTATTTTTATTTGTACGA CAAAACCCTTAGCAAATAA GCTTAGAATATATAATAAAGC GCGAATTAAAAA (SEQ ID NO: 1284)	MVCPTCNGVYADYNDHIRKKHPDERYTALQLQPLGLTPCPICKTACKND LGVKTHLSKIHKISGASKISTQPRIRTENTDNTNSVPTSSFNPVLPEIQTLTP GLNNSRWADNPRKRRAADTPSPTRGRNTRPRRFSYTDIDLTNDEPADNP RANNPRVNNPRVNNEPPSSPNSLPSISEFHTPGTLPLTNSNISLKDQHDKI TGPILQKPLIQKLIEYSKIPIPEHHLHARQAKIFADAANRIAKNFIQSPTEKTL FNLLILPRIFGIGLINGKVTKIMQNFPSQIPPIPKIDFPSEKTDSDPVLNAKKL LEKGYIGRAAKAIIDPTPVAPETPESLNILREKHPIGQNNPFNTKSQPISGR QITEKAILLAISSIGREKAPGLSGWTRSLLDAAIKIPTQNDVIPALRLLTDMI RQGTAPGRELLCASRLIGLSKPDGGVRPIAVGDLLYKIAFKAILNTLWSPN CLLPYQLGVNSIGGVEPAIFTLEEAIMGPNINGIKSITSLDLKNAFNSVSRA AIASSVAKYAPTFYRSTCWAYNQPSILTENGSVLASAQGIRQGDPLGPLL FSLAFRPTLETIQKSLPYTYIAAYLDDVYILSKTPVKDKIAKIIEKSPFTLNSAK TTETDIDTLKTNGLKTLGSFIGPTELRKEFLQNKIQNFESSINALKKLPKQYG LLILRKSTQLLLRHLLRTLNSQDLWELWEKTDKLIADFVINLTVTKRKKRPIT DFVTPLITLPIKDGGFGLLRHNGIAQDIYFAAKDLTTEIRHKIQRISNDFPQ NQSPTATEILHLLHNGVLADCKNGLTNAQLNALTENASYLGRKWLNILPI QKSNRLTDWEMAEAVRLRLLAPVKPLTHPCNHCGNRTNININHEDVCKGA VRKYTARHDQINRSFVNSLKSRPEIDVEIEPDLNNENNVVNNANTTTENPT PSPNGQNDTGCLFTTPIRSGTRNGQNGLRADFAVINGVSKYYYDVQIVAI NKDSGNTNPLNTLADAANNKRRKYQFLDPFFHPIIISAGGLMEKDTAQA YKQIQKLIGPVAAHWLDTSISLILLRSRTTAAISIAKNRPRA (SEQ ID NO: 1406)
CRE	MoTe R1	JQ747 487	Magnap orthe oryzae			


CGGTGGGGGCTATTT
ATGCGCTTTAATTT
GTGCGGGGGCTATTT
ATGCGCTTTAATTT
GTGCGGGGGCTATTA
ATGCGCTTTAACTTT
ACAAATTTTATTTAT
GCGCTTTAATTGCT
GCGGGCCTGTTAAT
GCGCTTTAATTTAC
AAATTTCATTAATG
CGCTTTAACTTTTAT
ATTTACTAATGCGTT
ATTTATATAATTGCT
ATTATTATCGTTGCT
ATTATTATTATTGCT
ATTATTATCGTTATT
ATTATTGCAATTTTA
TTATATAAACCCTC
GTTTGTCCCTCGATT
TATCCCGTTTCTTTT
CCATCCCATCGCGC
GTTTTCGTAAGCTTT
GGTTTTCGTAGGAT
TTGCTTTCGTAGGC
TTTGCTTTCGTAGG
CTTTCGTCAGCTTTT
ACCTGCTTTTATTTT
TTCTTTTTCTTTTTAT
TCCCCCCCCTTTTTT
TTACCTGGTTTATTA
GCGGTTTACCTGCT
TTTATTACCTGGTTC
CCCTTTACCTGTTTT
ATTAGCGGTTTACC
TGCTTTTATTACCTG
GTTCCCCTTTACCTA
CTTTATAAGCGGTT

HERO	HERO-2_BF	.	Branchiostoma floridae	TTTTCAGTCTGGCTC AGCCAGTGACCGCC GGGAAAGTCCGGC TGACTACCACGAAT AGGGTGGTGACAG CTGGATAGACAGAC GACAGCTCGGAAA	TGATTAAAGACCCGAAACA CCCAATGACCCCGGGTTCAT CACTGATGATGTGTCCCTGT TCGCACTACCAGAGTGTATT CTAGAG (SEQ ID NO: 1285)	MNAVCVCGKVCKNQRGLRIHQTKMACLRRVQAEHRSGAVATTVEPVL SASAPGQTEEDQGPEAPHSARNLRATPAPPQGRKSDHHRVKWPAANS KEWSQFDEDVDMILESVSRGSTDQKLQSMCTVIMSMGAERFGTIGQRK PTDTMKPNRREVKIRQLRQELKSLRRSFKASTSGEERAALAELTHHLREKL RTLRRAEWHKKKKGKERARKRSAFITNPFGFTKRLLGQKRSGNLTCPVEEI NLHLSNTFSDASRDVDLGPCPLLVTSPEPEVHFDISEPTLKEVRETVKAAR SSSAPGPSGVVYKVYKHCPRLVVRLWRILKVVWWRRGKVAADWRQAEGV

						WIPKEEESSKVDQFRLISLLSVEGKIFFKIVAQRLIKYLLDNQYIDTSVQKGG
						VPGVPGCLEHTGVVTQLIREAKENRGDLAVLWLDLANAYGSIPHKLVETA
						LTRHHVPESIQNLILDYYSNFWLRAGSSTATSAWQRLEKGIITGCTISVPLF
						ALAMNMIVKGAEAGCRGPVSRSGTRQPPIRAFMDDLTVMTATVPVCR
						WLLQGLERLITWARMSFKPAKSRSLVLKKGKVAERFRFTLGGTQIPTVSE
						KPVKSLGKVFNSSLKDTASVQQTRSDLTTWLEGIDKTGLPGSFKAWMFQ
						HGVLPRVLWPLLVYEVPMTMVEQLERTISRFLRKWLGLPRSLSNIALYGR
						STKLQLPLSGLTEEFKVTRAREVLMYRDSSDSKVSSAGIHVRTGRKWKAQ
						EAVDQAEARLRHSVLVGSVAVGRAGLGSCPKPRYDKVSGKEKRLLIQDEI
						RAGEEEDRRCRMVGMRKQGAWTRWEHADSRKVTWPELCRAEPSRIKF
						LISSVYDVLPSPANLHVWGLAETPSCQLCQRRGTLEHILSCCPKALGEGRY
						RWRHDQVLRVLADTVSNAIQSSRSQQPPKKSIVFVRAGEKTRQQPTSAG
						GLLSTARDWQLLVDLGRQLKFPEHIVATSLRPDMVLVSESTRQVVLLELT
						VPWEERISEANERKRAKYAELVVQSQSNGWRARCVPVEVGCRGFAGQS
						LAYVLKLLGVRGFRLRKSIRDILEAAEKASRWLWFRRGEPWKPHGHRSG
						NDQPRLGRPGEGVW (SEQ ID NO: 1407)
HERO	HERO-2_DR	.	Danio rerio	GACGGCATTGGGG CAGTATGGGTTGGC ACCCCTAACTGCAT CTCCCCTAGGAGAG CATCCCGCAACACG CTACAAAGAACCAC AAAGAGCAATACCC CCAGGGATGCCCGA GAGGGGGGGAGGA TGAGCATCCCATTC GGACGGTCCAATCG GTATTGACCCCAGC AAACGGAGAATCG ACA (SEQ ID NO: 1162)	TTCAAGCCTGGCGC AGCCAGTGACTCCT AGGAATAGACTAG GTGGCAACCAAGA ATAGTTTGGTCGAC TACTGGAGAGACA GTTGACGGCACGG AAAGACGGCACTTG GGACAGTATGGGTT AGCACCCCAGCCTG TGTCTTTCGTGAGA GAGAACCCAAACAA GCTACGGAAAGCCC CACAGAGATATACC CCCAGGAGATCCCG AGAGGGGGGGAGG ATGAGATCTCCAAT CGGACGGATCAAA GGTTA (SEQ ID NO: 1163)	TGATCAACCCCGGCTGGGT CACCTGGGTGAGAGTGTAT GATGTTGAGAGACCCGAAA CACTCAATGATCCCAGGATA CATCACTGATGATGTGTCCC AAATGCATCCATGAGATGT TTCTTGCATAA (SEQ ID NO: 1286)	MTHANEQTTNKIYVTCICGKLCKNHWGLKIHQARMKCLEQESKVQRTG PEPGETQEEPGPEATHRAKSLHVPEPQTPSEVVQQRIKWPPASKGSEWL QFDEDVSNIIQAIAKGDADSRLKTMTTIIFSYALERFGCIEKGKTKPTTPYT MNRRATQIHHLRQELRSLKKLYKKATDEEKQPLAELKNILRKKLMILRRAE WHRRRGRERARKRAAFITNPFGFTKQLLGDKRSGRLECSIEEVNRFIEETV SDPLREQELEPNKALISPTPPAREFSLRGPSLKEVKEIIKASRSASTPGPSGIP YLVYKRCPGLLLHLWKILKVIWQRGRVAEQWRCAEGVWIPKEENSKNIN QFRIISLLSVEGKVFFSIVSRRLTEFLLENNYIDPSVQKGGIPGAPGCLEHTG VVTQLIREAHENRGDLVVLWLDLANAYGSIPHKLVELALHRHHVPSKIKD LILDYYNNFKMRVTSGSETSSWHRIGKGIITGCTISVILFALAMNMVVKSA EVECRGPLTKSGVRQPPIRAYMDDLTITTTTVPGSRWILQGLERLIAWAR MSFKPSKSRSMVLKKGKVVDKFHFSISGSVIPTITEQPVKSLGKLFDSSLKD SAAIQKSKKELGAWLAKVDKSGLPGRFKAWIYQHSILPRVLWPLLIYAVP MSTVESLERKISGFLRKWLGLPRSLTSAALYGTSNTLQLPFSGLTEEFMVV RTREALQYRDSRDGKVSSACIEVRTGRKWNAGKAVEVAESRLQQKALVG TVATGRAGLGYFPKTLVSQVKGKERHHLLQGEVRASVEEERVSRVVGLR QQGAWTRWNTLQRRITWANILQADFQRVRFLVQAAVTDVDLPSPSNLHV WGKNETPSCLLCSGRGSLEHLLSSCPKALADGRYRWRHDQVLKAIAASL ASAINTSKNHRAPRKAVHFIKAGEKPRALPQLTTGLLHKASDWQLEVDLG KQLRFPHHIAATRLRPDIIAISEASRQLIILELTVPWEERIEEANERKRAKYQ ELVEECRERGWRTYYEPIEIGCRGFAGRSLCKVLSRLGITGVAKKRAIRSAS EAAEKATRWLWIKRADPWTAVGTQVGT (SEQ ID NO: 1408)

HERO	HERO-3_BF	.	Branchiostoma floridae	CTGACCAGCAGACG GGAAGCCCGCGAC CAACTAGTCTCCGC AAATATTGCACACA GGGCGACCCTATGG AGCTGATTCAGTCA AATTTCCTCTGAGA TATACCGATAACTA TCTACAGAAACTGC ACAGTTAGTTTGGA AAGAGCTTTTCTAC TGAAAGACAGCAA AATCCGCCACTTTA GACGAGCGTCAAG ACTGCCCTCCCCAT AACCAAT (SEQ ID NO: 1164)	TAGAAACCCACAAGGCTGA GAAATGTAGAGCATCTGTA TGGACAATATTGATGATTG AAATGTTGTGATTTTAGATC AAATTTAGAAATATGAAAA CCGAACTAAACTAAATATAA TGTTTTTTTTAAAGTAATGA TAAGCAATACCCACATTGTG CAATACTATCTATGTTATGT CCTTTGTCCCCCCTGCATGT TTGGTCAATAATGACCATCG TGTCCTGGGCTCCGTGTACC TTTCTTTACTATGAATAAAG AATGATTTTACTAC (SEQ ID NO: 1287)	MALPAVRSGPASTWTLLITLVIVAAKGTDGFMSFKLPLLSTDTWSGYNN DVKTLLGPLHHELATNEMSPKLAGEGFSDIMCDFMASKPEFSHTTEESHS EGYISHEPQSLAQVKRLKNKLRKKAFRADATPEDRKAFRDAIKTYSFMKR QQKRKETTKSAAHQEKEYHKNFWKFAGKCAKGQLDIPPVKPAFSVYYAN EYYKNKYSHPTRVDFNKLLWFPHLPVEEQLPANSFDMSPVRPKDIKAVLS KRCATSAPGPDGIMYGHLKHLPACHLFLSTLFSKLLESGDPPTSWSSGNV SLIHKDGSPEAAENFRMICLTSCVSKIFHQILSERWAKYMTCNDLIDPETQ KAFLTGINGCVEHVQVMREILAHAKKNRRTVHITWFDLADAFGSVEHELI YYQMERNGFPPIITYIKNLYSRLKGKVKGPGWESDPPFPGRGVFQGDNL SPIIFLTVFQPILQHLKGVEQQHGYNLNDKHYVTLPFADDFCLITTNKRQH QKLITQISSNTKSMNLKLKPRKCKSMSIVSGKPSDISFTIDGDPVKTTKDAP EKFLGGYITFLSKTKETYDILAKTIETTVENINKSAIRNEYKLRVVMEYAFPS WRYMLMVHDLTDTQLQKLDSIHTKAIKTWLRMQPSATNAILYNTRGLN FKSISDLYLEAHALAYSRSVLKADEKVKHALQAKLDRESQWTRKMQKWG IGKCHTIHQQAIHVAKDSEWTSVRKHVKQQVTDMRHDVWTKHQENLL QQGQMLQLLEEEKCDLTWRSAMYNLPRGILSFAVRASIDALPTLCNLTT WGKRNTDKCKLCGNRETLHHVLNHCGVALQQGRYTFRHNSVLKHITDTI IESIDTSRINATIYADIQGYTTNGGTIPVHTIPTTQKPDLIIYLPEQKTLHIHEL TVPFEKNIKTSHDRKVNKYSTLAADLETAGISATLTCFEVGSRGLVTPENKT RLRTLFKIVKAKPPKTLFTDISRIAMLSSYAIWNSRHEPYWESETLL (SEQ ID NO: 1409)
HERO	HERO Dr	.	Danio rerio	AAAGCAGTAGAG (SEQ ID NO: 1165)	TAGCATGCCACTTGGACAC AGGCCGGGGTCTGATCAGC CTCGGTCGGGTCGCCTGGA GGAGGGTGTCTGTTGCAAG ACCCGAAACACCCTGTGAG CCCAGGAAACAACACTGAT GATGTGTCCAAGGTTGTGC ATCAGGAGATGTTTCTGTA AC (SEQ ID NO: 1288)	MTTHRAEVTTSGKTQEEPGPEATHSAQSLLVSPTPAAGRSPATQSCPQV TAAHNSPQSPQSQQVAVTRSDCVPLAQPRIQWPQSSKKAEWLQFDKD VNQILEVTGKGGVDQRLSTMTTLIVNIAAERFGTVTPKPTPSTYTPSHRVK EIKRLRKELKLLKRQYKAAGEVERAGLEDLRGILRKQLVNLCRAEYHRKRR RERARKRAAFLANPFKLTKQLLGQKRTGKLTCSKEAINNHLKATYSDPNR EQPLGPCGALLTPPEPTSEFNMKEPCRSEVEEVVRRARSSSAPGPSGVPY KVYKNCPKLLHRLWKALKVIWRRGKIAQPWRYAEGVYIPKEEKSENIDQF RVISLLSVESKIFFSIVAKRLSNFLLSNKYIDTSMQKGGIPGVPGCLEHTGVV TQLIREAREGRGDLAVLWLDLTNAYGSIPHKLVEVALEKHHVPQKVKDLII DYYSKFSLRVSSGQLTSDWHQLEVGIITGCTISVTLFALAMNMMVKAAE TECRGPLSKSGVRQPPIRAFMDDLTVTTTSVPGARWILQGLERLVAWAR MSFKPAKSRSLVLRKGKVRDEFRFLGQHQIPSVTERPVKSLGKAFNCSL NDRDSIRETSTAMEAWLKAVDKSGLPGRFKAWVVQHGILPRLLWPLLIY EVPMTVVEGFEQKVSSYLRRWLGLPRSLSNIALYGNTNKLKLPFGSVREE FIVARTREHLQYSGSRDAKVSGAGIVIRTGRKWRAAEAVEQAETRLKHKA ILGAVAQGRAGLGSLAATRYDSASGRERQRLVQEEVRASVEEERTSRAVA MRQQGAWMKWEQAMERNVTWKDIWTWNPLRIRFLIQGVYDVLPSP

						SNLYIWGRVETPACPLCSKPGTLEHILSSCSKALGEGRYRWRHDQVLKSIA EAISKGIKDSRYRQATAKVIQFIKEGQRPERTAKNCSAGLLSTARDWVMT VDLERQLKIPPHITQSTLRPDIILVSEATKQLILLELTVPWEERMEEAQERK RGKYQELVEQCRANGWRTRCMPVEVGSRGFASYTLSKAYGTLGITGTN RRRALSNNVEAAEKASRWLWLKRGEQWGQ (SEQ ID NO: 1410)
HERO	HEROFor	.	Takifugu rubripes	AGACTAGGTGACAA CCAAGAACAGTTW GGTCGACTACTGGA AAGACAGTTGGCA GCTCGGAAAGACG GCACCCGGGACAGT ATGGGTTAGCACCC CAGCCTGTATCTTTC GCGAGAAGGAACC CAAACAAGCTACGG AAAGCCCTACAGAG AAACACCCCCAGGA GATCCCGAGAGGG GGGGAGGATGAGA TCTCCAATCGGACG GACCTAACGTTA (SEQ ID NO: 1166)	TGATCACCCCGGCTGGGTC GCCTGGGCGAGGGTGTATG ATGTCGTGAGACCCGAAAC ACCCTATGAACCCAGGATA CATCCTGACGATGTGTCCCA GTGCATCCAGGAGATGTAK CTTTAAGT (SEQ ID NO: 1289)	MTPAMEMTTTVTCICSKLCKNQRGLKIHQARMKCLEREVEVQRTGPGP GETQEEPGQEATHRSQSLHVPEPPNPNRVVQQQRIKWPPANRRSEWL QFDEDVSNIIQATAKGDVDSRLQAISTIIVSYGSERFGRIEKGNTETTSYTM NRRSFKIHQLRKELRTLKKQFKRAXDGDKQALKELYNILRKKLKTLRRAEW HRRRGRERARKRAAFIANPFRFSKQLLGDKRSGRLECSREEVNRFLQNT MSDPLRGQDLGPNRALISPAPPSAEFKLAEPSLKEVEEVIKAARSASSPGP SGVPYLVYKRCPEILRHLWKALKVIWRRGRVADQWRCAEGLWIPKEEDS KNINQFRTISLLSVEGKVFFSIVSRRLTEFLLKNNYIDTSVQKGGIPGVPGCL EHNGVVTQLIREAHESKGELAVLWLDLTNAYGSIPHKLVELALHLHHVPS KIKDLILDYYNNFRLRVTSGSVTSDWHRLEKGIITGCTISVVLFVLAMNMV VKAAEVECRGPLSRSGVRQPPIRAYMDDLTVTTTSVPGCRWILQGLERLI LWARMSFKPTKSRSMVLKKGKVVDKFRFSISGTVIPSITEQPVKSLGKLFD SSLKDTAAIQKSTEELGGWLTKVDKSGLPGRFKAWIYQYSILPRVLWPLLV YAVPVTTVESFERKISSFLRRWLGLPRSLNSAALYGTSNTLQLPFSGLTEEF KVARTREALQYRDSRDCKVSSAGIEVKTGRKWKAEKAVXVAESRLRQKA LVGAVATGRTGLGYFPKTQVSHARGKERNHLLQEEVRAGVEEERVGRA VGLRQQGAWTRWESALQRKVTWSNIMQADFHRVRFLVAAVYDALPSP ANLHAWGKSETPTCSLCSGRGSLEHLLSSCPKSLADGRYRWRHDQVLKA VAESIALAISTXKHHHAPKKAISFIKAGERPRAGPQITTGLLHTAXDWQLH VDLGKQLIFPQHIATTSLRPDMIIISEASKHLIMLELTVPWEERIEEANERK RAKYQELVEECRGRGWRTFYEPIEVGCRGFAGRSLCKAFGRLGVTGTAK KRAIKXASEAAERATRWXWLKRADPWVATGTQAGS (SEQ ID NO: 1411)
HERO	HEROTn	.	Tetraodon nigroviridis	AGATTGGTCTGGCT AAGCCAGTGACGTC CAGGAACAGACTG GCTGACGACCACGA ATAGAGTGGTGACA GCTTGGATAGACAG CTGACAGCAGGGA AAGACGGCAACCG GGGCAGGAAGGGC TAGCAACCCAGCCT	TGATCACTCCCAGTCGGGTC GCCTGGGTGAGGGGGTCTG ATGTTGAAAGACCCGAAAC CCCCGATGACCCCAGGTAC TATCACTGACGATGTGTCCA AGACATGCATCAATAGGTG TATTTAGAAATC (SEQ ID NO: 1290)	MATTQASVKPTAVATCVCGKICKNPRGLKIHQTKMGCLASVQPEQRAR FSLSESREVPARAEPYGPQQPHSPEALGETQEERGQESPHSAQNLRAQV AQAPDNPQHRRVKWPPASKVSEWQQLDEDLEGILESTAKGGVDRKL QTMTTLVISFATERYGTMEKRAAPEKYTKNRRAEKISQLRQELRVLKKQF KGASEDQKPGLAELRCTLRKKLLTLRRAEWHRRRRAKERAKKRAAFLANPF GFTKQLLGQKRSAHLECAKEEVDSYLHDTFSDAERENSLGECRVLISPPEP ACSFNTKAPTWKEIQTVVRAARNNSAPGPNGVPYLVYKRCPKLLARLWK ILRVIWRRGKVAHQWRWAEGVWVPKEEKSTLIEQFRTISLLNVEGKIFFSI LSHRLSDFLLKNQYIDSSVQKGGIPGVPGCLEHCGVVTQLIREAREGRGSL AVLWLDLANAYGSIPHKLVEMALARHHVPGPIKTLIMDYYDSFHLRVTSG

						SVTSEWHRLEKGIITGCTISVIIFALAMNMLAKSAEPECRGPITKSGIRQPPI
						RAFMDDLTVTTTSVPGCRWILQGLERLMTWARMRFKPGKSRSLVLKAG
					GCATCTTCCGTGAG	KVTDRFRFYLGGTQIPSVSEKPVKSLGKMFDGSLKDAASIRETNDQLGH
					GAAGAACCCAAAAC	WLTLVDKSGLPGKFKAWVYQHGILPRILWPLLVYEFPISTVEGLERRVSSC
					TTGCTACGAAGAGC	LRRWLGLPRSLSSNALYGNNNKLTLPFSSLAEEFMVTRAREVLQYRESKD
					CCGAAGCAAAGATA	PKVALAGIEVRTGRRWRAQEAVDQAESRLHHKELVGAVATGRAGLGTT
					CCCCCAGGGGAGCC	PTTHLSRLKGKERRDQVQLEVRASIEEQRASQWVGLRQQGAWTRWEE
					CGAGAGGGGGGGA	AMARKISWPELWRAEPLRIRFLIQSVYDVLPSPSNLFLWGKVESPSCPLC
					GAATGAGCTCCCCA	QGRGTLEHILSSCPKALGEGRYRWRHDQVLKAIAESISSAMEYSKRLPLPG
					AACGGACGGATAA	RGVRFVRAGEQPPPQPRAQPGLLATARDWQLRVDLGKQLKFPENIVET
					C (SEQ ID NO:	NLRPDIVLHSQSSKQVILLELTVPWEERMEEAYERKAGKYAELVEDCRRA
					1167)	GWRSRCLPIEVGGRGFAGKSLCKAFSLLGITGMRRRKAICAASEAAERAS
						RWLWIQRDKPWTSASWTQAGN (SEQ ID NO: 1412)
NeSL	LIN9_SM	.	Schmidtea mediterranea	AAACGACATCATGA	TAAAATGGCAAAAAGATAT	MMDSRQLNTPKIRKYQNPKMTNDIMKSYNYAVLSDVTPQETTQTTTHL
				ACGCTTGGCCGCAA	TTCAAGATGAATTGTGGAC	NVDIDNETTQPKQPLTKSGKPKSKPIAVSYKFKDATFIWDTTPQTNPPRD
				CAATCCAGTTATCC	TCATCTAAAAAATGACCACC	CTKLIDKTRPRKTIFKKSAFQSYLKKELSNETFVEVKTFLMATHKYREKDEN
				CTGCGGTAACATTG	TTGAGTCCAAATATGCCTAG	SRLLAYRIINRYVMETANEFKETEFDMARFAKFFTIPENWLKHLKPYSTAT
				TGGAACTCATAAGA	CTATCATGGTTGCTGATGG	ETSPADRIKVQKLVDLTCRYPFKTQEEQTSVANFLHFFTQRSIGISRDYKF
				CAAGTACTAAAAGA	AAACAGTAAGGCACCTGAT	QKFIPFMARKNTRPETTSTMVTTSPTEQNRLPMVIITPLEEPKSEHRRPEK
				AGAATTAGAAAAAT	AGCTAACTTTTCACTGTGAA	RGASNDTIVLSDEEFPLLKRRTLPTRKSKNPTGAGNVPTETECTDEVKFILN
				TAGAAGAAAAAATT	TATCTTCAGATATTCACAGT	NEYQIECKECGKVWENVRNGLNHLRQKHDFPNRTDVMVSCVRCEVPIK
				GAAAATAATTTATT	GACACGAAAGGACACCACT	GAECVNHIKNHKKDDKEESEAGSLVANTQDIPNESSLSQAAIEVYLRNILK
				TATAAAATTTAAAA	AGTAAAAACCACTAGTTTTT	MKENQERNIQYLEPSTANFLINRNLRAFYQNVKIEKLIGWEQVIWLIHW
				ATTTAAATAAATTTA	TCTGACACCTCTTGCTACAA	NKCHWIVYLANCDSKTSVILDSDNQMTLQQRCNIKAKFDKFLEGTFEEKT
				AAAATTTAAATTTA	ACTCTGTAAAAATCAAAAG	VLGTLERKVPQQPNNFDCGIYVIQYISDFLKDPQRIDYHTPDSKRIRKEIGE
				AATTTAAATGAAGA	GATCGATAGGCCGCGCTTT	LILEEMKNPASKIKNPNKEIQSLLQKFRLLQINVNDVFHWFAAEYQKSLPK
				TAAAAATTTATTTAA	CACGGTCTGTATTCGTACTG	IRTKRDGKLNKLSCSYQIQRLFGLAPKRAVKEIYFQETSTADLETRVLNEHF
				TCCAATAAATAATC	AAAATCAAGATCAAGGAAG	KKDESTMKECKIKNGNHYQDWITKAQIDNKEILEALKNSTDSAPGEDNIP
				AAGAAAATCAAGA	CTTTTCCCCTTTTAGTCAACA	LRQWIIWNNDGVLFDMFNYIKRTHDIPDMWKNYTTTLLIKPGKSQESNI
				AA (SEQ ID NO:	CCAGGTTTCTGTCCTAGTTG	PANWRPISILPTSYRIFMKVLNKRVLEWANRGELISKWQKAVDKANGCD
				1168)	AGCTTCCCTTGGGACATCTG	EHSYVIQALIEKANRSYYKNEQCHLAFLDLADAFGSIPFQVIWHTLKNMG
					CGTTACCATTTGACAGATGT	MDEETINLLKEIYKDCSTKYKCGKNESEKIKITKGVRQGCPLSMTLFSLCIQ
					ACCGCCCCAGTCAAACTCCC	YLIQGIAEKKKGATIAGQEVCILAYADDLVIVANTAKDMQMLLTTIENLAK
					CACCTGACACTGTCCTCAAA	QADLIFKPAKCGYYRDPRDKKSMMKIYGKEISIVDEKNVYTYLGVRIGDTK
					ACAGTTCAATTGCATCCGAA	KKDLNVRFEEVKKKTTAIFKSKLRSDQKLEAYNIFCQSKFVYILQGEDIAKT
					GATCGCAATTTTTTCACTAA	KIETYDEEIKKMIKEDILKLQDKSPFTDFVIYSPREKGGLGITKIIDEQTIQTIN
					AATAAATTAACAAAAGTTA	RTAKLLNSSHRAIRAIIYEELIQVANLRGEKEINTIEEALKWLEGTNKYKKNS
					ATTATACTGCTTCATTGAGT	NAKTTWITRVREAFQTLEKKHKIKVRFVPKENCIGYKIKCDTQEKIVELDN

					AAGTAGAAAAACAATC (SEQ ID NO: 1291)	SKELSKSLHWMIKEAYYKEWKALKCQGYIISLKTSEFMEWKMPRGLPDP DWRFLTKVKANMILDVNMKQANQGGRLGSTKCRKCEDKESASHVINHC ASGNWSRVEKHNQVQNELAKELTKRNISFEKDSIPKETKESLRPDLVIRLK DKIMIVDIKCPFDEESAIESARNKNIDKYRELAKEIQAKTGLQTTVSTFVVC SLGTWDKRNNELLRQMGIRYEESKEMRINMIQKAIHGSRKTYDHHRNF NNG (SEQ ID NO: 1413)
NeSL	NeSL-1_CBre	.	Caenorhabditis brenneri	AAGGACGCTGGTTT AAGGCCGAATTCGT TCGTTCTTTTTCTGG CGGTCTTGCTTTGA GCTTGGTTTCCGAT CCT (SEQ ID NO: 1169)	TAAACCCACACGAGAMCTA CGACGCCATAAGATCAGGC ATGTACGGATGTGAATGAG ACTGATGAACGGAATGAGC ACGTGCCCATAAGATCGGG TATKAAAGAWCAGAGACG ATCCCTAMCATCGGGAAAA CACGAGTTATACTGCTTCAC TGAMCTCGCTAAGCTCTCA TAATGACCGAACTTGTTCGC AACTGCCTCCTCAACCGGGC GGGTGTGAGAAGGGAGGT CGCCTTGAGGCGGACGCAA TGAGGGATGTGTGCAGGTT CCCCCTCTTGAGATCCGAAA GTCTAAAAGTACTAGACCG AAAGATCGAGGACGGACG GGATGGCCGCGAGGCACAC GGCGGGTAACACAGCCAGA TAACCTAGTAGATCTTCGGA TCTCGTCGGCCTGGAGATA TGTGGAACCCTGGGAAAGG AGAAAGTTGTTTGTTGGGC TGGCAAGAGTGAAGTTTGA ATGTGAACCACCGTCATGC AACCACTAAACCAGTGGCG ATGCGGGTGGAGTCATCAC AGGAAAATGTTTCTGTTGCT TGACTTATCAGTGTTTGATA TCGCCCTCAGGCACAAGTA TGAAGGCCCCCACCCACAT AAACTCCCTAGCAACTGGTA	MPLXISDCVHLVSAEGDTMNGRSTCGPLSRSSSVVSRSRSSPSPSVPPHP SPSIGPDTGLSAGIIGTSRGCSLWLPEVDNAILSQWLRKGLERDHEVLVCG FEAAKPLSLSKARLLRKTPRNTGVVRHILEFDGRLVHTNCNETECVLSTLXS XXAVEVVRISLKCEPREPCEPKCVLSILCSDKIVXISFECETREPFPFFXDRKF REPIPFVFERMYDPRDPIPSFICWMYDLRQRMTPGTLPXNPLSXENKDS WGRPAVIKNEIRSMRSYLEENVKENRLNLLRRLRGGGEGKKMIRKLVAEK KSDTEAVCRILYPLDDRYECFVDGCETTSTMGYGSSDLKYMTTHIKKEHG VKVQWTYECSLCNKQAPFMGGAASKWVTAHMATKHTETVKLKLKPSIS TTAKVAAKLDEIAVSLPKPRQVRVLRDPDEVKEKVAKPTLASTREEVKRN ALRNMAPLVELSSQNQLTGAERPEETSEAMRLEECRTPEKIAELEGKIQT RTVTKKLSALKESMEKRTREEKVGKPSLAPIHEEVKKTARRSLAPLVEPSTF THLTGASRLQAVRDAFSKANKDAAAKRRSSLAKPARLSEIMNTTFTKETV NETKEPVNDTDESIATIQPQVRVYRFNTWCLDHETTREAWLTGEVVDW FMGKVTEKKDQYRVFDSLVWSMYKFHGVGYVLDLMRDPLTYFLPICEH DHWVLLVIDEKGIWYGDSKGAEPCREIAKFIEETKRERRMFPVPVPLQRD GVNCGVHICLMVKSIVNGEPWYTEEEVKVFRRNVKRGLKEFGFELYSERI VYVGDDSIKVNDEHDDDVVFLSEETNNTFFTIEQAEDPAEEDAQHLESPV KPVKLMELKIPKIEIKKKEIRRKPKQQIEKKRKVPTGKPDELLVRVRLWLER EVQSYFDSGKRFQRLEWILDVLTAAIHKATAGDEQAIERIEKRSPPLEVEE GEMSTQTEPKKRERKEKESGCEMKASHKEMYFKNRSKAFNVIIGKDSKQ CEIPIETLQKFFEGTTAETNVPAEVLKEMGSRLPKLEALDWMEANFIESEV SDAMKKTKDTAPGVDGLRYHHLKWFDPEYKMLTLLYNECKNHRKIPSH WKEAETILLYKGGDETRPDNWRPISLMPTIYKLYSSLWNRRIRSVGGVMS KCQRGFQEREGCNESIGILRTAIDVAKGKRRNLSVAWLDLTNAFGSVPHE LIKSTLESYGFPEMVTEIVMDMYRGASIRIKSKNEKSEQIVIKSGVKQGGDPI SPTLFNMCLENVIRRHLDSASGHRCIKTKVKVLAFADDMAILAENRDQLQ TELNKLDKECESLNLIFKPVKCASLIIERGMVNKNAEVVLRGKPIRNLDEN GSYKYLGVHTGIATRVSTMQLLESVTKEMDLVNQSGMAPFQKLDCLKTF VLPKLTYMYANAIPKLTELKVFANLTMRMVKEIHEIPIKGSPLEYVQLPPS QGGLGVACPKITALITFLVNVMKKLWSSDSYIRKLYRDYLDEVAETETGM EEMTKEDIAKYLSGDVPIDKKAFGYNTFTRVRDVCNSLTXIXGAPLHKLKI VERDGDFAILVQATKEGMEKIFTCAQEKKLQQLLKAEVNTALAHRFFTEK

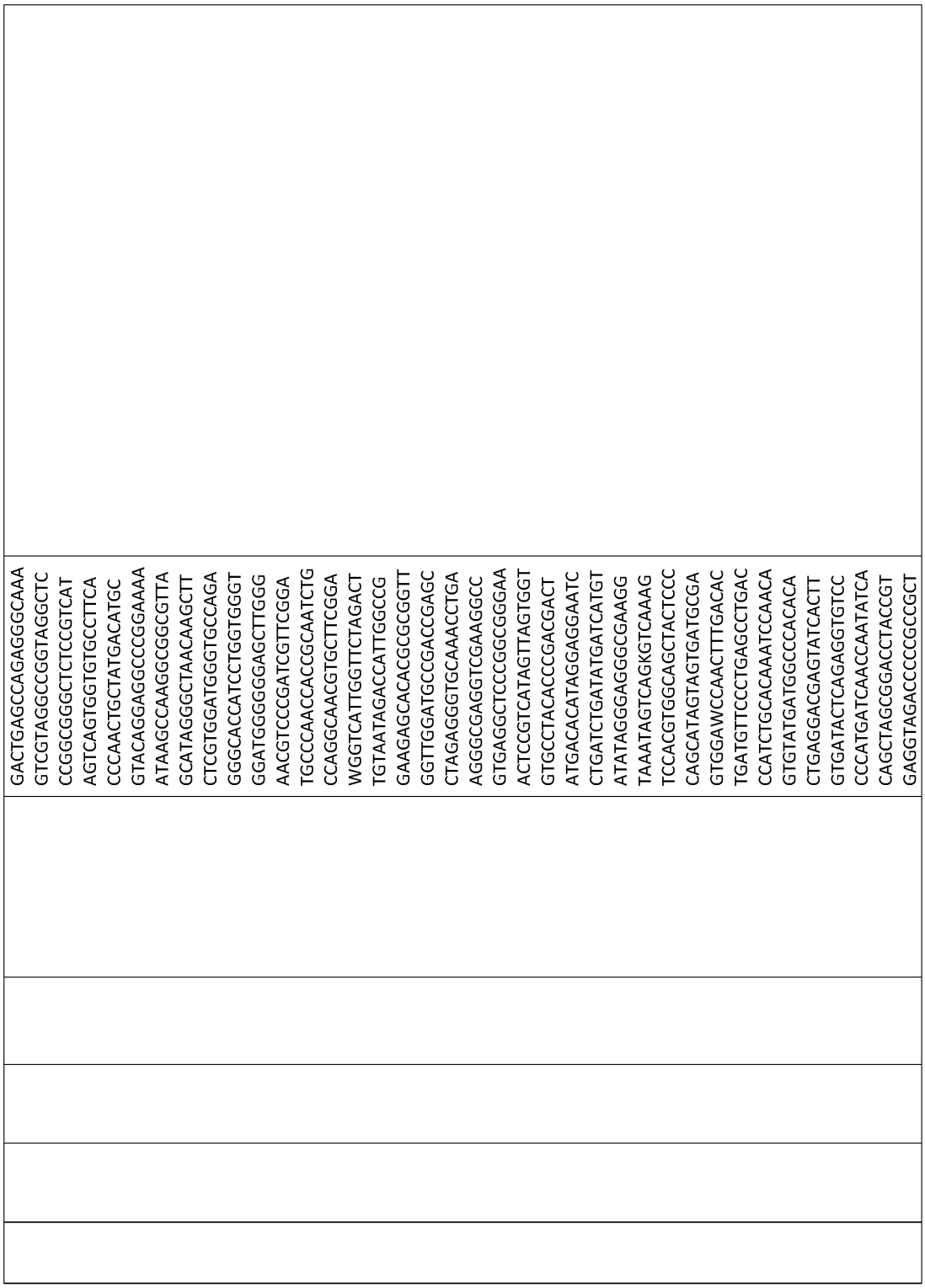
				GTCCAGCAAGCGCTGGTWC TTGCTACTATTGCGCCCCAG GCTCGCCC (SEQ ID NO: 1292)	PVKSAVMSVMRQYPQSNAFVKNGKNVSIAVHSWIHKARLNALHCNFN TYGENKSKVCRRCGKDVETQLHILQXCEYGLPKLINERHDAVLHVVRNLIR KGSKKDWKLKIDETVSSCNQLRPDIYMCSPDGKEVIMADVTCPYESGM QAMQESWNRKVTKYEGGFSHFXKMGKKFTVLPIVVGSLGTWWKPTTN SLVQLGIEKXTIRRVIPELCSMTMEYSKDVYWNHIFGDTFRKPPMRFGVE KPKGNSWKKEGSEPKGAASSD (SEQ ID NO: 1414)
NeSL	NeSL-1_Clap	.	Caenorhabditis japonica	GCGCCCCGGGTTAC ATTGTCGGGGCCAC CTTTCTCTTGGAGT AGAGTACAGTCTAC TAATTTTTTGATAAG CTAGTCGGGTCCGA ACCACTAGAGTTTG CTTGAAAATGCGTC AAACCAGCATTTTA GAACTCGCCCAAAA GTTCGGCCCCGACC CCCAAACAAATGGG ACCTTCTTGACGATT TTCCCTGAAAATCG GAGGATGGAATGG TCCCCTATTCTTGTA AATAGKACTGTGCA ATACCCCTTCGTCAT CTGTGGGGAACAG ATGACACGTGACGT CATCCGTGTAGACG TCACGTTTTCCCGT GCCTGCGGGGAGCCC CCAATCGAGCAATT TTTGCTCTTTTGAGT GTCTGGAACGCTTG AAACCCCAGACAAA TCAGGCCCAGTCGT CGGAAAATTTCTTTT TGAAATTTTTTGGC GCCTGCGAAAAAAA TTTTTTAACCGCCAC	TAAAAGCCAAAAGCCACGG AGCATCGGGGAAAGAAAAAT GGAAAAGGACTGAAAACG AGACTGAAAAATCCCAAAC AAAACAAATCCAAAACAAA CTGAAAAAAAAAAAAAAAC AAAACAAAAACTGGACAGA CACTGGAAACAGTGTCAGG CAAAGTCGCCGATTATACT GTTCCACGCCTTAAAAGTCC CGAAATGGCGCAAAACAAC CTGAATCTATCTGAAAGTGC TCCAAACCACGCACAACTCG GAGAAAATCAGGGACAAGT TGCTTCACGCAACGGGCTG GGACAGGTACCCCCTCCTG AAACCGCGAGGTTGAGGAT GGACGGGAAGGCCGCGAG GCTTATGGCGGGTAACTCG GTTGGTGTGCTAGTAGATG ATTTATATCCGACAGCCCCA ACTAAGAGGAATCCTGGGA AAGGAAAACTTGAAAAAGT TTTTACAGGGCTGGTAATA GTTCAGCACAATTGTAGTCT ACTGTCTTGCAACCACAACA AACCAGTGGTTCTGCGGGT AGATCAAACTATAATTTGTG TGTTTTCTTTTACTTGACCCG GGCAACACATTATACCACGT CCACAAGGACGAATTCATA ATGGCCCCTCCCTAAATAAA	WRRPAPKQTKNSSLHHLGHEVKRIARLKPGIFEFHAKPKNSSLHHLGHGV KRXARLKPGIFEFHAKXKNSSLHHLGHEVKRIARLKPGIFEFHAKPKNSSLH HLGHEVRRNSRLKPGIFGFYQKSKNSSLHHLGHEVRRIARLKPGILEFHAK NRIKSGLKVTFLSDLXAHAGALACSRFLASTLKTEHCRQKSFKPVGFLLHFL KNSSINEVASLRNVKKXFLEFFSGKPIGGMASFSRTKITFFKLCLKNFVLSAE NPPIIRQKTNQNKASXVQIARGGHLSDCLPSQKMAGVLGRLFLSVQSTLS HRPFDTLLRSDDDKRGRKTIKLQFFIKENLVTPXVARDVKILXKQTKNNSG NSDSNSETKNFSKNKVSRQNGPLIGGGNHKKIGENQITRTLEIESKSDDN KVLVLRILYPTNDWYKCYSQWCQHKSLVGYGAHDLKYLTDHIKSTHSKKV EWSYQCSICDAKAEGTGTKAARWITAHMPKVHGIEATHRIKQNSEKTTN VKTANSLQEMALSLQKPKNGPKKVVMATSTTPEKKISELESKIQTREVAK QLSALKESAQKNQQGNKTKNVKSSLKTIAENTNETKKISARKSLINYLKPE DVLNHIPKEPKPASAKXGLQELTGAQRLQETRRRFMAGNRRDSIARRESL SLGKISNSFKIELKNAPEKTTLKKPAVTQKQNTSQNVSSSTVVKENKTGND VITIDDTETVKRKINTWCLDHESTENAWMADDIIFWYIQKQIEISLDNKKF KVIDPLIWTTYRIYGVECVQDELVGFEKYFFPICENGHWVLLIIDDKRVWY SDSLADKPIEVIEDLINKLNRTQGKFNQTVPKQKDGFNCGVHVCLVAKSV ITENFWYTEKDVNDFRKTVKLWLFSEGFELYSEPYKQIQNKNISVNSEKN QISDNEKNWGDKTQTVNESTLKERDEDIFLLRPHISVGVALKTEDEKNQK AENLKAPQKLKAIRRLKILKTCLKKLTAVKGKPEETERAAIPNLMAIKLKTP PKVEPVRRNPEKGENYXKSQPNKKRQIPTGKPPDELVKKVREWFEIQFQA YFEDGKSFQRLEWXTGLLTAAIHKASAGDEQAVGKIIKRCPPLEIEEGEMA TQTETKQKPKNQKSTKGANSSSSIREAYAENRARTFNKIIGKDDKCEIPIEK IEKFFENTTSNTNVPTETLARITSDLPKLEIGSWIEEEFREKEVAEALKKTKD TAPGVDGLRYHHLSWFDPKXKLLTKLYNECREHKKIPGHWKEAETVLLYK GGDETQAENWRPISLMPTICKLYSSLWNKRIKSVTGVLSKCQRGFQERE GCNESIAILRTAIEAAKGTKKSLSIAWLDLTNAFGSVPHESIEATLIAYGFPG MVTEVIKDMYNGASIRVKTKNEKSKQILIKSGVKQGDPISPTLFNICLESVI XRHLKSADGHKCIXSNIKLLAFADDMAILSDSKTKLQQELQKMDDDCTPL NLIFKPAKCASLIEWGKVQKDQKIKLKGQFIRSLAEQDTYKYLGVQTGIET RVSAMQLMKKTVSELDKINCSALAXWQKLDAVKTFVLPKMTYMYANTV PKLSELKEFANITMRAIKVMQNIPVKGSPLEYVQLPIGKGGLGVACPKTTA

			AAACCCCCGGGAG GCGCGGWTAGGGA TATCGATGTCATCG ACTCGTCGGTGATC TTTGATTTTCTCTCT GCGTCTCCTATTTTG GAACAGTCTCGACC AAAAAACCGGGCCT GGCAACCCACCGAA TCCGGATGTCGGAG GGATTTGGCAAGAA ATGTTGGAAATAAC GAAATTTCGTTATTT TCAGCACAATTGTC AAACCGGCAAGAA AACTGGATGGACAA GACACACAATTTAC CGGAAATTGTGCTT GTTACGTCGAATTT CCCAATTTTGAAAA AATTCCTCGTTCCAC TGGTCGGGACGCG AGGTCAGACGATCT GCACGTCTGAAACC CAAAATCTTCGGAT TTTATGCAGTAG (SEQ ID NO: 1170)	CTCCCTAGCAACTGGTGGTC CGGCGAAGCCGGTTCTTGC CACTATTGCGCCCCAGGCTC GCCC (SEQ ID NO: 1293)	LITYLVSTMKKLWSTDDYIRKLHTDYLKMVAIKETKTKEVTLEDLASYLSDD KTVCKKAVGYNSFTRVREICKTLSKNKGALLSQLKIIAKDGKLAILVQAKKD GKTKIFTHDHVKTLQKXLKKEINEALLHRFTTEKRVKSEVVRVVQEYPQCN SFVRDGGKVSIGAHRFVHKARLNLLACNYNTWQDAATKQCRRCGYEKE TQWHILSSCPKSMGGKITERHDSVLKTVKEMIQTGSLKNWKLKLDHELP GSTRLRPDIYLRSPNGSEIILGDVTIPYEHGIEAMQTAWQKKIEKYEEGFKY LRSTGKKLTIVPIVVGALGSWWKPTTDSLVSLGIDKNTVKRAIPEICSTVLE YSKNIYWNHIFGDSYQKVPMFFGGEKPKGQSWKKVKPPEGKTASNHEP PG (SEQ ID NO: 1415)		
NeSL	NeSL-1_CRem	.	Caenorhabditis remanei	CGCGAACCAGTCAT (SEQ ID NO: 1171)	TAGCCGATCGTAAAAGAAA CCGAGCCGTAACAACAAGC AAAGTAAACAAAAGAAAAA TCAATAAAAAGGAAGGTTG ACCTCAGACCCCGAGGAGG GAAGAGAGACACCSAGAAA AAGAGAGACGCAGAGAAA AGGAGAGACACCTCTCATA AGGAGAGGTAGGTCAATCC AAATGTAAACAGAAAAAAC CAGTGGGGAGGAAAGAAA	MTVFIDRGIGERGQMAVCSLHRYFSFSPFSPIPPYVVNNGSFGENGCGTD KSLLPVIEVVVREVKINWSENILVVECLIMVKSGERVVVKRQNLEKVIQNL ARINSTLFSNLGNQIFCVVPRIKDSTNKEQGYRKEKQXKFHVSFRSIKSQV PPYLRGGGDVMEDTEIRGIRKLEPEAQLDSSKPLICRVLYPTQGYMYKCFY PKCKGHSNGSTDLRSLKKHMVDKHFTNIEFAYKCATCMFLTTGKSATALK SIKAHMASHHKVTMEPGKKSLVQKLNARLEEAAPSLPMPRNRSKVIQLT PEKSISELEKKKQTRSVAKQLSTLKESAQKKEEEVKIAEVKKREPRLSIIPESN VRRSLAAGLEQCINPEQSVAQRIREKREEYAKASREAAAKRRSSLAMKPA RLPDKENEITLQETKKIDDPIVIDLEKECILTTVLQVPRNQFNSWCLEHETTI DAWLTDEVIHMYMCTITENRKYFMAIDPVLWPVYVRNGAEDLLRRTSC PGTFFFPICESNHWVLLVIEHDVYWYLDPKGEEPKGNVEILLESMKRKRQ

					GACTGATTTCACCCACTAAAATGAATTTGGAAACAGAATTTGGAAGAGAAAAGAGAAAGGGAAACCTAAAGAAAATAGTTCTCTTGCCAAAATTCTGTAGAGGAATACTTTGTCAAAACATGATAGAAACCAGTAATCTGGTACGAAAGACAAGTAAGACCTGAACTGACAAGAAGGAAGTCAGAAAGAAATACCGCTCACAAAGCCTGTGATCGATTCTCTTACCTACTGAACTTGTTCTCTTGGCCTCGTAACCGGCTAAAGGGAGAAGGAATGTTAATTGGAGATAGACATAAAGATAGGTGGAGTGAAGGTCCTGTTCTTGAAACTAGGAGGAATGTGGAAAGAGCAGAAGGCCGCGAGGCTTTAGACGGGTAACTCAGTCAGTTGCTAGTGGTCTTCGGATCCAACGGCTTCGGACATAGTGAGGAACCCTGGGTACGGAGAAGAAATGGAAAAGAGATAGGGCGGGCAAAGGCTAAGTTCATACACTGTCATGCAACCACTAAACCAGTGGGATCTGCGGGTGAATCACTTTCGAAAAGAAGTGAATGGACGTGCTGATGTCTGACTTTAAAGAAGTCTGAAATTAAAAAAACAGATATAAAGGCCCCTCACTATAAACTCCACAGCAACAGGTGGTCCGGCGAGGCCGGTTCTTGCCACCATTGCACCCCAGGCTCGTC (SEQ ID NO: 1294)	YYEFPPPSQRDNVNCGVHVCLMAKSIVDECGYNWYSEEDVRSFRTNMKDILKSKGYELCPEPYNRQNLLKTEKQKEVILEEMIDSFVVEDDMTFTVHRDSDHGDDEVEHLKTIEQEPENEISEIENVEGSVDSVIPKLMEMRVQTPPVINEKRGKKRVSAKEKPRKQKEKEQKVPTGKPDELVKRVRVWFEKEFKSYVEDGKSFQRLEWXTDVLTAAIQKASAGDEKAVELIEKRCPPLEXEEGEMCTQTEKKKKPKSGKGNGGQESMKSLMASYSENRAKTYNRIIGKHSKQCEIPIAKVQKFFEGTTAETNVPKETLKEMCSRLPKVEVGTWIEGEFSESEVTEALKKTKDTAPGVDGLRYHHLKWFDPELKMLSQIYNECREHRKIPKHWKEAETILLYKGGDESKXDNWRPISLMPTIYKLYSSLWNRRIRAVKGVMSKCQRGFQEREGCNESIGILRTAIDVAKGKKRNIAVAWLDLTNAFGSVPHELIKETLESYGFPEIVVDVVEDMYRDASIRVTTRTEKSDQIMIKSGVKQGDPISPTLFNMCLESVIRRHLDRSVGHRCLKTKIKVLAFADDMAVLAESSEQLQKELTAMDADCSALNLLFKPAKCASLILEKGIVNRLNEVVLRGKPIRNLMENETYKYLGVQTGTETRVSIMDHITEVSREIDLVNMSQLAMHQKLDILKAFILPKMTYMYQNTTPKLSELKVFANLVMRSVKEFHNIPLKGSPLEYVQLPVGKGGLGVACPKNTALLTFLVTIMKKLWSSDSYIRKLYTDYLEEVAKVEIGKFEVNLNDLAEFLSDERAVDSKLFGFNAFTRVREVVRSLCKNKDSPLHSLKIIEREGKLAISVQATEESIEKIFTEDQEKKLMYLLKGELNTALQHRFFTQKVFKSEVMRVVQQHPQSNSFVRNGGKMSFSAQRFVHPGRLNQLPCNYNTWAKGRTKLCRRCAKNENETQSHILQVCDYSIGNIIKERHDAVLYKFRELIKRGSKGHWLERTDRTVPNTGSQLKPDLYLESPDGKHVILADVTVPYERGIEGMQKAWNEKINKYTDGYKEIFRRQGKSLVVLPLVVGSLGTWWKPTEESLIKLGVEKTTVRRIIPETCGMVAEYSKNCYWRHIYGEKYVQTPMINGGKKPEGNDWKKCEKGIEVPKVTN (SEQ ID NO: 1416)

NeSL	NeSL-1_TV	.	Trichomonas vaginalis	GGGTGAGTAGTCTAGTGGT (SEQ ID NO: 1172)	TAAGAAGAGATAAGACGAG TGAGAAGAACAGAAGCATA GTAGGATTGGCAGAGCTTA AGCGATGTCACTCGGTACG AAACGTGTACCAAACACCG GATTCCGTGCTAGGAATCA CAAGCCAAAATAAAAGAGA CACCACGAAAATTACTCACC CTCCCTCAAACAGATAATAA TATTAACCTCCCATCCATCA GTCCGTATGGTCTGATAAC AGACTAGCACCACATCCAT GATACACTCATTGGAGTGA AAACCACCAACAACAAATC CACCTAGACCAAATCCTGCC CCACCTCCACCCAAGTAGCT CGCTTCGCTCGCTCACCTAA AACTTTGCTCGCTCGCTTCG CTCGCTCGTCTTAACCCTTT CCGAATAAACACTTACAATT CCCGGCTCGCCCCATTTTTT (SEQ ID NO: 1295)	MIPVLGTGGPEKLPLQSYVYCGNTAITDSFTPTAKTILKPEEQNLDIVLKNI AALNPENYSDLIRSLSKMEFRLDYPKEIENYWISEKLFSQSIASLPISLLVAS MFSPEDRDLSTEPFHCNADGCNFHCDNCERMVEHIREHHNTDPMINTF ETTEDTFRRITAIKIDKTGIEELNPLKYRCSYCDELFTEAEDHAIHMISHLTE KLSPDISFFFNDILRLYKTIDKPTVQNLFPETQVAIFDTLEETNRFRLIVGRE AIETIEEAFPPSPPGTDRKPSIIITDTCQLRFVPCMDEPPKGDLGILTLLLRD FSAHNIPIKSLNNKELIADKDIDYSPDFVEGALANAEEHDTTNSQNNNGR YINSAEKLTEFLIQCEDYLTNIKTLEDLERFYTTIKDYRVVNKEVIAEDTPIFVY FLVEEGKLPKPGLRCPLESYEGHEDKAFESLRKLCDHFKGEIAKTSFDPKVH TIDIWVEFLAQAYGTGTFVVKDENGNIDLDTHVFKCPYADCSYTNNDRSK LMDHMKTKKHAKNVYIERYGFFWGIVIEGVNRPKGIVYPTLKDIKEHACR KCPEAGCNTYVTELSDIKEHLKKKHKSTTAGVDGEIAHTDATYCWITKEEL DALHAERARERAEQVDNTPVQQIINADNNEENNENQEDNGNNEEADA LDPPNNTTETEDEAVHAVIINPPATEEEEVAIIAEARRNIPELQQQAEERGC VTPKMTSLVRLKLLKGGGELFNKKLTPLATRYAAATGNTEADKIKVDYLTLK CNAALREMIYTNNHSESKFMTAENGEDTAPPPRISEDTRDRIQKAANEIK GTLIKVVKHISHARCLKDSTRDDEHNKFVEMIAKIKNDLRDNKFEQYNIEE IFQGPISDQSILNIVNTEDNNEFIKKMDYINRILGTPQDASPYARKKLQACF ADNPTKTLRNIILADKVPQQSLKPSEYLDYYGPQWANEAEGYENFLHHD YALPERYGQVFANDFLDFMTNESKIIEVIRNKNHLSAHGLDGIPNSVYML FPVSAAKFLSILFRSIISGHIPDCWKLSKTVMLFKKDDPSLAKNWRPIGITS CTYRIFMTLVNKALQMIPMFHAMQKGFVRGATLSEHIAVANEVLCQSTR TQSEMFQTAIDFTNAFGTVPHQLIFDSLEAKKVPDSIINLLKDLYKGARTAI YTRHAHSEIVPVRRGVIQGCPLSPILFNCCLDPLLYAVQRRHFEDGYRFQD KAGQYSIAIQAYADDVLVISPTHEGMQRILNTVDEFQKIAKLKVAPQKCV TLAKTSTAIQPFRIGPDEIPIKTSMDNITYLGIPISGTKTSRFAAATGILEKVK AQIRVVFASHLALSQKIIALRVFILPQLDFYMFHNVFRVNDLKATDQMIR GLIDKEAPTSNIPVSFFYMPKNKGGFGLVKLELRQPQLVLTKFARLWLSQ QAETKAFFHTMAQEEKSFRKVVEDQENGFLGIKMENGKIVQKNERSKRT NCFITQAAKAADKLEVRFKEWDKGGIQVRGVGENATDWVRSKHIGQISP LIGRVIQQRQYEEFKKDETHSHTFCEPAALAESHDIMKRPQAVPNNLYSA AIALRTNTAPTPANMHFHNPEVLANCPLCGCQSCTLFHTLNMCRNRFSL YKWRHNIICDDIYQFIHDHYPGVTIKCSARITSDGYQTTGPELDDTVKDLL PDLVVYDEANKMIKIEVTCPYGTDNNVGNSLDAAYDKKVNKYKSLAEQT ERLFNWTTTLSIIVVSSLGVIPLRTKLDALRISPADHIQLLKRLSMHAIAASA CIVFEKVPEFFGMRCRPLPGRVTAPNAAIPPNNNENNNNDTDHGQENCQQ ATSEEQPTNNGNAQEDNGQGEQINNSTEQTISVDQIIEEDAENNAIEQA LDQPDEDEFLN (SEQ ID NO: 1417)

NeSL	NeSL-2_CBre	.	Caenorhabditis brenneri	GACTCGCCTTGGGGAAGGTWTTTCAGGGGKSAATTGCCGMAGGCAAGGCAGCCCCCSMMTAGCTTACAAAGTAAGTACMCATTTTCATTTCTTGTGAATTCTTTAAACATATTTTTCTTGTTTTTTGATTTCTTTTTTCTCTACCTTCCCCCAATTCTTCCCCTCATCTTGTGTATACATCCCCCTCCTCCAACCAATCAATACATTGACCTCTCTCTTCTGTCAAAAAATCAATACTAGTATATTGTCCCTTGTATAGTATTATTTGACGTCGTCTTTGTATTAGGAGTAGGTAACAAWCTGTGTATGGCTTCAAAAAGCATGCACAAACWCCTGTCAAAAAGTAWTTCCCATCMTGTGAATAGCTCAACGACWKGAAGMCCAATGAT (SEQ ID NO: 1173)	TAAACCGGCTCCTCTGGGAGGAGGTATGTCAGAGGACATTCTCCGTGGGCGGATGGGAGGAGTAGGGTAACGACCCGTCATTCTGGATGCCTAAACCACCACAATCTGTCAAGGCAAAGTGCCCCAAAAGCACACGCGTGGATCGGTTTGGATGCCGACTGAGCCAGAGGGCAAAGTCGAAGGCCGGTAGGCTCCCGGCGGGTTGTCCGTCATAGTCAGTGGTGCGCCTACACCCAACTGCTATGACACACAAGGACAACCCAAAATAAATAAGCCAAGGCGGCGTTAGCTTCGAGCTAACAAGCTCCCCGAGAGGATGGTTGCCACAGGGCACCATCCTGGGGAACGACCCGATCTTTCGGATGCCCAACCACCGCCAATCTGTCAGGCAACGTGCCCCAAAAGCACACGTGCGGAGCGGTTGGATGCCGACTGAGCCAGAGGGCAAAGTCGTAGGCCGGTAGGCTCCCGGCGGGTTCTCCGTCGTAGTCAGTGGTGTGCCTACACCTAACTGCTATGACAAGCGTATAGGAGGCCCGGAAAAACAAGCCAAGGCGGCGTTAGCTGAGAGCTAACAAGCTTCTCGTGGATGGGTGCCAGAGGGCACCATCCTGGTGGGTGGATGGGGGGAGCTTGGGAACGTCCCGATCCTTCGGATGCCCAGACCACCGCAATCTGTCAAGGCACCGTGCTCCAAAAGCACACGCGCGGGTTGGTTTGGATGCC	MRYHXSNXPAXRTSDNXWRSIXKDVRRPDPSTIEEKSRYNRSIGIPDSLKXRSSAVRSXSSXPPSGPQDVRLXNSPSLDDRRRLVDCETTLGSYREWTDKPMMGKMTYAAVTKRAPPRPQTGGARLSTNLLADEMEIKYRDTNDIRLVIDLPNPHLIKCPLCKSCISARGRGANALKYMKRHIADAHHLNADFVYKCSRCQEHEPENVCGAKWIVNHLKRVHGYTLEDAVSTAKPSTRQQIANAFNDSAPFIDARKTSDVPEKKSREAGLEKFLAPTKSEDTREKTPPSTRKSSESSEASIQSTIQETLSESSDTLTVQEIINISSEDEMDEEPPKRRVNVWALIHENGKDAWIDSDLMVIFLESRARGYESCSIIDPLNFICTDMSYLTTIVRRRMEEGYKKIIFPLCANDHWTLVTITGSTATFYDPMGNEPTETVKKMIDELDLEMQLAPSNSPRQRDSWNCGVFVMKMAEAYIKDTQWDLTDVDTDVKTFRRSLLTELKAKFNIFAEDIQTYRPPSRKALTRNSQSPVVVCHKCSRPATPIQDVSRMEVEEAPVLVPTPEEPPQEWTFVGKNRKRGVTSRTPNTSPEAKRPAFPPVPLKPSANRWHFPEEETEKMEVSSADEVKNSTPPKPPKIPNLLAMKIASPVPLKRGNPSKKHGKGHMMNTARKGPTKKEMPKGEPANLIVKIRSWFDEQLKMYKDEGSNLQRLTWLSDSLTAAIGKAFNGNKYIVDQIIKRNPPPLVEKGAMSTQTSRKRDEFKPRERMAQEPNEPLRIQYAKNRQKTFFKIIGKQSEQCTINIETVEQHFRKTLKAPVVSENAIKTVCGSIKKVLMPKTIEDPISSVEVKSILTKVKDTSPGTDGVKYSNLRWFDPEGERLAKLFEECRKHREIPSHWKEAETILLPKDCSDEEKKKPENWRPIALMATIYKLYSAVWSRRISGVQGVISPCQRGFQSLDGCNESIGILRMCIDTASVLNRNLSCSWLDLTNAFGSVPHELIRRSLESFGYPQSVIQIVTDMYKGATMKVKTADQKTQSIKIEAGVKQGDPISPTLFNICLEGIIRMHQMREKGYDCVGHKVRCLAFADDLAILTNNKDEMQEVIDKLDADCRSVSLIFKPRKCASLTIVRGAVDKYAKIRINGDAIRTMADRDTYRYLGVKTGVGGRASETEALIQVVKELQKVHETDLAPHQKLDILKTFLLPRLQHLYRNATPKLSELREFENVVMKSVKRYHNIPIKGSPVEYVQIPVKKGGLGVLSPRLTCLITFLTSTLCKLWSDDPFISSIHKDALSRITVKAMGLTTQSATIKETCEYLNTRKAVTKGGYSLFCRMNESLRTLSVIQGAPLKSMEFIPVNNEIGIAVQATKDSEIKVFTKADSLKLMSKLKDLVRSAMLKRFLEEKSVKSRVTQVLQHHPQSNRFVRDGRNCSIAAQRFVHPARLNLLSCNANTYDVNHPKGCRRCQADFESQQHILQNCHYSLAGGITQRHDRVMNRILQEIGNGRKAHYKIMVDMETGATRERPDIIMEERDGPEVLLADVTVPYENGVQAVERAWDKKIEKYKHFLDYYRKIGKKATILPLVVGSLGTYWPDTSHSLKMLGLSDGQIRNVIPEICQIALESSKNIYWKHILGDSYKTVEGLFCQRNNKEVRFEGKGEKHHVSQRFQPLKCEKVRTMKSTKEEGRSRSNAKKGPNWRRSKSESDGRSVSKGRYWRDPSNKPPHSKMTQSALAKR (SEQ ID NO: 1418)


GACTGAGCCAGAGGGCAAA
GTCGTAGGCCGGTAGGCTC
CCGGCGGGCTCTCCGTCAT
AGTCAGTGGTGTGCCTTCA
CCCAACTGCTATGACATGC
GTACAGGAGGCCCGGAAAA
ATAAGCCAAGGCGGCGTTA
GCATAGGGCTAACAAGCTT
CTCGTGGATGGGTGCCAGA
GGGCACCATCCTGGTGGGT
GGATGGGGGGAGCTTGGG
AACGTCCCGATCGTTCGGA
TGCCCAACCACCGCAATCTG
CCAGGCAACGTGCTTCGGA
WGGTCATTGGTTCTAGACT
TGTAATAGACCATTGGCCG
GAAGAGCACACGCGCGGTT
GGTTGGATGCCGACCGAGC
CTAGAGGGTGCAAACCTGA
AGGGCGAGGTCGAAGGCC
GTGAGGCTCCCGGCGGGAA
ACTCCGTCATAGTTAGTGGT
GTGCCTACACCCGACGACT
ATGACACATAGGAGGAATC
CTGATCTGATATGATCATGT
ATATAGGGAGGGCGAAGG
TAAATAGTCAGKGTCAAAG
TCCACGTGGCAGCTACTCCC
CAGCATAGTAGTGATGCGA
GTGGAWCCAACTTTGACAC
TGATGTTCCCTGAGCCTGAC
CCATCTGCACAAATCCAACA
GTGTATGATGGCCCACACA
CTGAGGACGAGTATCACTT
GTGATACTCAGAGGTGTCC
CCCATGATCAACCAATATCA
CAGCTAGCGGACCTACCGT
GAGGTAGACCCCCGCCGCT

NeSL	NeSL-2_CRem	.	Caenorhabditis remanei	CCAACTCTCATCGT ATTAACCTACGGTA TTCACTCCTAGTGA GTGTAATAAAGGTT AATTACGTTTTCTCT TGCMAGAGAAAAA GAAAATTCGAATCC TTTTTGTGTAACTCA CAAACTGACAGAGA CCTATCGAATTTCCT TTGTTTCGTATATA GGAATAGTCACTCT GGACCACGAAGTG GACAGTTGTCGGCG GACTTCCAGAGTGG AGAGAAAAAGGTGT GAAGAGAGGAGGT CTAGAAACACTTCG GCTGTCTAGGACCA GTTCCTGAGTGGAA AGAGGAAGGTCTA GAAACACTTCGGCT GTCTAGGACCAGTT CGTGAGATCTCTCG TGGAGAGTTGAAA ACAGTCAGCTGAGG CTACTGTATTTCTTG ATAGCCCCCGCCCCC AATCCCCCTCCCCCC CCCCTCGACAGATT TTTCTGTTTGACCTC CTGGAATTTGCGAG GAGTGCGCGAGAA TTTTCGAATTCTTCG CGCGTTTTCTCGAA ATTTTCCAGAAGAT	GTAGCAGGCTCGCCTC (SEQ ID NO: 1296) TGAATACCGTCAGATAAGC CCCCAACATAAAAATAAAA GTCGGCGTTAGCTAACCAC TAAACCGGCTCCTCATTGG GGGAGAGTATCATTCCGGT GCTCTCCGTTTGGGCGGTA GGGAGGAGTTGGGTAGCG ACCCGGAAGTATGGATGCC CAACCACCGCAATCTGATCT GGCATTGTGTTTCGGATGG TCTCTGTCTCTAGATCTGAA ATAGAGCTCTGGCCTGAAG AACACACGCGCGGACCGGT TGGATGCCGACTCGATCTG GAGGGTGCAAACCTGAAAG GGAAAGTTGAAGGCCGTGA GGCTCCTGGCGGGGAAACTC CGTCATAGTCAGTGGTGTG CCAACACCCGACGACTATG ACATAGTTGGAGGAATCCT GATCTGATAATAATCATTGT TCATATAAGGGAGGGGGGAT GGTAAATACCCAGGGTCCG AAACCATCAAAGCAGCTAC TGACCAGCATAGTAGTGAT GAACACATAGACCCTGGGG TTCCCTGAACTCGACCCATC TGCACAAACCCACTTTGTAC AAATGAACCAAACTGATGA AGAGTTTAATGATTTCTTAC ATCACAGCTAGCGGACCTA CCGTGAGGTAGACTCCCGC CGCTGTAGCAGGCTCGCCA TTG (SEQ ID NO: 1297)	MTNVYLKPVNDNQTNKTGDNSRNTMSNSQCEMTWKPVARTYAQAAS TNPADDKTVTVLGCKYNLLKLGNTPQTSKRSPPKPSRGGARISSVVTLTDE LEITHREEGKITFAIDLPNKNNILCPLCRECTQTRGRGSSFTKHMKLHVKEK HQLDATFIYKCSMCNEYEPEKKCGTKWIQTHLQKVHNYKYDESAIVVPV PPNTRQQIANELNNAAPFVDIRKPKAAAVEEKKTENGALLKFLTKSNKDD QVKSPSXDIPDAESPEKETQALTIDPKGNNSPSKSSIRSSQSSASSVCQEIQ EIITLSEDEDPKGARPKPGINVWSLINETGKDAYIDTDIMMAFLKMRVEN CDSVNIIDPLNYQFPARVDLVPLIQRNLEDGKKRVVFPICADEHWTLLTIS NGIAAFYDPTGSRMSSYIEELVNELGLIIPKEQDEQPRQRDSYNCGVFVM KMAEAFIQDTEWEMEEVEEDVKNFRRNLLEELKPNYEIFAEKIKYYNSPG KSFAQSRPTSRSSQCAVCPTCSRSATPMMDVGNMEVDPVPQQQETPK SREPEQDEGWKVVGKARKRGVVTERSPNISPEAKRQFTGPEIKVVSPGKF HPLVGETEEMEVTCDSPPTKEPTTEPKVTPSLPAMKIASPEVTKKQTSKKK GKYGKKKQXTKKAQPPKGEPTKKAQPKGEPAKLIEQVRTWFDKQMKSY QEQGSNIQTLTWIADSLTAAIFKANSGNKYLVDKITARCPPPLLNEGEMA TQTSRRTEAVKPKDRFVKESNEPLRIQYAKNRAKTFNVIIGKHSARCEIDIN VVENHFRQTLKAQPVTEEALNTVCSGIKKAKVDPSIEGPISSGEVKAILAKI KDTSPGTDGVKYSDLKWFDPEGERLALLFDECRQHGKIPSHWKEAETVL LPKDCTEEERKKPENWRPISLMATVYKLYSSVWNRRISSVKGVISDCQRG FQAIDGCNESIGILRMCIDTATVLNRNLSCSWLDLTNAFGSVPHELIRRSL AAFGYPESVINIISDMYNGSSMRVKTAEQKTQNIMIEAGVKQGDPISPTL FNICLEGIIRRHQTRKTGYNCVGNDVRCLAFADDLAILTNNQDEMQDVL NQLDKDCRSVALIFKPKKCASLTIKKGSVDQYARIKIHGMPIRTMSDGDTY KYLGVQTGNGGRASESESLTQIAAELQMVHDTDLAPNQKLDVLKAFILP RLQHMYRNATPKLTELKEFENTVMKSVKMYHNIPIKGSPLEYVQIPVKN GGLGVMSPRFTCLITFLASTLFKLWSDDEYISSIHKKALSRITAKVMGLKTQ KATLQEQCEYLNTKKAITKGGYSLFSRMNEAIRTLSVNLGAPLKSMQFIPE NGEIALEVQASENSQIKVFSKADSMKLVTKLKDLVKSAMLKNFLENKKVK SKVVQVLQHHPQSNKFVNDGKNXSISSQKFVHPARLSQLVCNGNSYSKD LPKNCRWCGYECESQAHILQHCTYSLSSGITQRHDRVLNRILXEVIKGRKN NDYYDIMVDTEPGPTRERPDIIMIQKDGPEVLLADVTVPYENGVVAIEAA WDWKMEKYSHFIDYFARLGKRAVILPLVVGSLGTYWPDTSNSLRMLGLS DGQIRNLIPDISMIALESSKQIYWRHIFGDSYRIVSDLYCRKDQQEIRFGDE PMENVQVSDRFQPFKTREREKKSEEEKKRRSKSKKGKTWRGSKKQTDSR QSGKSNQNQGGFQRSVGQGVSR (SEQ ID NO: 1419)

NeSL	NeSL-4_CRem	chrUn	Caenorhabditis remanei	TCGAGCGGAGAATC TTCGAGAAAGTGAG CTGAATTTCGCGCG AATTTTCCGCGATTT TCAAATTATCGATTT TTGTCGGAAAATTT ATTTTCTGGCAAAA TTTGATTGAGTTCA CGCGGGAGAGAAG GAATTGTTGGAAAA GGGTATTGATTTTT GTGGCGGAGGAAA CTCCCACTGAATCA ATAACTCTCAAAGG AGAACTCATCGAAC AACCTCGGGTGACC TGAATCTTGGGCGA AATTTTCGCATTGA CACAAGATAAMAC AAATTACTGTKGAA AATAAATCAGAACA AACTGTCAAAAAGA GAGACAAAAAGTAT TGATTAACAACATC (SEQ ID NO: 1174)	CCCTTTTCTATCGTA TTAACTACGATAAC CGCTCATTTGAGTG TAAAAAAGGTTCCC CCCTCCTCGCCTGC CTTACCCACGCATCT CTGCCTCTGGGAAG GCGGAGGGTCAAC TTGCGGGTCTGTGG ATTTCCTTTCCTATC CACCGCCCATATTCT CTGTCGAAAGCCTA CCTAGATCAGCCGG	TAACATGCCTTGGAAGGCA CCACGCCAAAAGTCCTGGC AACTGATTTGAATAATGTAT AAAAGTAACTGGAACCAAA TGCCCGATAGGTAGGGCGG GAGAAAATGACCTAGAAAA CACAAAGTCCCAAGCCCCC GGATTCGAAAGACCTATAG GAAGTCAGTGAATAGAGAG AAATATCAAACAAATCTCAC CCATTCACAAGGACTTACTG GTCGAGTAGAAAACAAGCC AAAACATCAAGCACGACGC	MTKTEWSWRHRSRSVGIVVKIDTSDYANVRVHVAADLSNEDGHTSH NNGIILPIPMKPSVDRFCQIQYPPRGYYVPHPQSQKGHDAKPSRHWNEE AQPPYYHNNNHGRRGRSAKPSGRRPPRKPILQEESLAAHPQIPGDTASA VPLYSDVVNNENKSQGKPPQGSHRRSGRPGTKPSVPVGEAEQETNSRPI APEPIVKFKHDKHGWTTVQGSHSSGRPVPKPSPVVSEANRFQLLQEGD FPPLTTSESSQEEIKVPNYQRIVSPIPLPSEEDSKLPTKSNYRAPKGRKSRNY KKPQQQNPKKYQQRLPYQPKVNNAPTDRMAPEQLKGGGGKTAHNDIE EMEIEEDTDEKIIQVKRIKIVNKLTPHHFVCMMTYPTDNIYRCFVKGCTAT SQGGWGAEDLKYLTVHIRQEHKIKVEWTYECGICGDLSGGAGKHISKWI KPHMRKKHNRDAPTNFKMGSRSSGKPKITELLEESAPSCSNPRRKTLNQ KKTAIITQVTPEKLKTGYQTRSVTKALSVLKESRQKELEVLREEEKANAKQK SKLHPFFTKAPHIDGVKPTVRRELSKMITPGGEHKGTKIPMVHTKRGLIQ KINRKAKKAKPMHLDESTIIEASQLDVITIDDDEDDNMTPMRRRFNTW

			AAAAAGGGGTAACTTTGGG	CLDHETTQEAWLTDDVINWYLKDLCFGNEQYMLVDPLVWLIYKMGGM
		GAGTTTTTCCTATCC	CAACTAATTAACGGATACCT	AGVEQRFKSKKTCLFPICEADHWILLVFDETNLCYANSLGSQPNGQVKNF
		CATTCAGGCGATCG	CCGTGTATCAGGCAAAGCC	IQQLNRKLCSFEKEVPLQKDSVNCGVHVCLIAKSIVNGQFWYDDSDVRTF
		CTCAAGGCTGTTTT	GCCACCAACAGCAAATTACT	RTNAKAALKAQGYELFSEAPKQIENPDSSHREDIKENSMEMCSESLMIVA
		ATCGACACTCCTTCT	GCCCGATAGGTAGGGCGTG	TPQRSEAPMELVDTEPSDLESPKSDRVVYEDCITALSDVSEPRMTPEKSET
		TGACAAGTATTTAT	AGAAAATGACCTACAACCT	PEVPVVEERDLDWPKLESPKSDRVVYEDCITDLSDVSEQRMTPEKCETPE
		TTCTTGACAAATTCT	CCAAGACCCGAGCCCACGG	APLVVECVELERLPKDLPVTDRSTVVAIPEAVKLEEKSEVVIPRLMELSYTV
		ATTTTTCCTTTTATC	AATCGAAAGACCTATAGGA	PPEPSPVVEYTQPYTHTHTKPKVKATCQMGKKRKVPTGKPDELIQIVRQ
		GATTTTCTCTTATTT	AGTCAGTGAATTGATGGAA	WFEKEFNDYVTEGRNFQRLEWLTNLLTAAIQKASAGDEETIEKIRKRCPP
		ATCGATTCTTGTGA	ATACAAAACCAAATTTCTTC	PEVRENEMSTQTSQRQKPTTTNQKKRSRNTTQSDTQANTYWRNRAKT
		AAAT (SEQ ID NO: 1175)	CATTCACAAGGACTTACTG	YNQIIGQDFKQCDIPIAILEEFYKKTTSVTNVPQETLVKVTSRLPRLDIGKWI
			GTCGAGTAGAGCACAAGCC	EDPFTEQEVFGALKKTKDTAPGTDGLRYYHLQWFDPDCKMLSSIYNECQ
			AAAATATCAAGTATGACGC	HHLKIPAQWKEAETILLFKSGDESKPDNWRPISLMPTIYKLYSSLWNRRIR
			AAAAATGGGTAACCTTGGG	TVKGIMSKCQRGFQEREGCNESIGILRSAIDVAKGKRSHLSVAWLDLTNA
			CATCCAATCAACGGATACCT	FGSVPHELIESTLSAYGFPEMVVHIVKDMYKDASIRVKNRTEKSEQIMIKS
			CTGCGTATCAGGCAAAGTC	GVKQGDPISPTLFNMCLETVIRRHLKESSGHKCIDTRIKLLAFADDMAVLA
			GCCACCAAACTGTACTACTC	ESKEQLQKELTEMDEDCTPLNLIFKPAKCASLIIEFGKVRTHEQIMLKREPI
			CGAAAAAACCAAGAAACAT	RNLNDDGTYKYLGVHTGADARTSEEELIISVTKEVDLVNRSALTPPQKLDC
			GATTTTCCCACTCCGTTAAA	LKTFTLPKMTYMYANAIPKLTELSAFANMVMRGVKIIHYIPVRGSPLEYIQ
			GCATCTCAACCAAGCTAAA	IPTGKGGLGVPCPRITALITFLVSTMKKLWSDDEYIRKLYNSYLKKVVEAET
			GCGGTAAGGTTATCATGTC	GIVEVSTKDLAEYLSNKVPSRKHEFGYNCYSRIREVCNGLALNQAAPLYKL
			AAAAGGTGTAGCTACAGCA	EFIEQDNELAVVVQPTEESKERIFTKDHVKKLQSLLKASVNDALLHRFLTTK
			ACCTAAAGCCCGAAAGGTA	PVKSEVVQVLQQHPQSNSFVRMGGKVSISVHVWIHRSRLNQLTCNYNIF
			GGGCCGTATAAAAAGACCT	DPKQPKNCRRCGYKNETQWHILQDCTYGWAKLIRERHDAVHHKVVTM
			ACACCCTCCAAGACCTAAAC	ICAGAKKNWGRKIDQELPGFTSLRPDICLTSPDGKEVIFADVCVPYSRTRN
			CCACGAACTCGAACGACCT	IEFAWKEKIRKYTEGYSHLVAQGIKVTVLPIAIGSLGTWWTPTNESLYQLG
			ACAGGAAGTCCGTGAATGG	ISKSDIRSAIPLLCSTVMEYSKNAYWNHIYGNSYTSVPLRYGHQKPDGDD
			AGAGAAATATCTCACCAAA	WKKELSCEPVLALQQQ (SEQ ID NO: 1420)
			TCTCTTCCATTCACAAAGGC	
			TAACTGGTCAAGTAGAGCA	
			CAAGCTAAGCCTCCAAGCA	
			CGAAGTGATATGGGTAATT	
			TAGGCAACCAATCAACGGA	
			TACCTCCGTGTATCAGGCAA	
			AGTCGCCACAAACACTGTA	
			CTACTCCGTTACTCCCAAAC	
			ACATGGATCTCCTTCTCTCA	
			CCAAAAAGCTTTATAACCAA	

					GCTAACGGTGGAAAGGACA TCATGTCACGAGGAGTAGC TACAGTAACCTCTCTCTTGA GACTGCAAAGTCGAGGATG GATTGGGAAGGCCGCGAG GCAAAAGGCGGGTAACTCG GCCAGACGCTAGTGATCTT CGGATCCGACAGCCCTGGC CTTAGAGGAACCCTGGGAT AAGGAGCACGACGGGAAG GATGTTCCGCAAGGATTTCC CTTCCCATTAGTCAGGGCTG GCAGTTGGTAATATAGCCTT TCTACACACCACCGTCTTGC ACCCACTAAACCAGTGGGA TATGCGGGTGGACTCAATG TAGAAAGGTGTTCCCACTG CCTGACTCGCCAACTTTATA TGTCTTGTCAACATAATGGC CCCTCACTATAAACTCCCTA GCAACTGGTGGTCCGGCGA AGCCGGTTCTTGCCACTATT GCGCCCCAGGCTCGCC (SEQ ID NO: 1298)	MNVDLDATIKSIGMNTKETTYPNSQLRVETTPCTSTTIMHASCNTTSTISY SPLPSAVSLPESPASSITITTDDNCDIIETPYPLPQTNGDLSEILKDIEANKD TTMSNKVLDCDSDSGDDRDMIIENDRESDMDLFSQSLLNTNQSDERRE KNLTENAPTEITTEKSYFDIISKASDNTTSKKLLNVKNELTAGLPPMPPVTN TAKFIRNVRPEDIADPTLYRLDSRGKLGCRTQYKKPGCGDIAVYDYEAIVE HAAFIHTIPFNEQNNVDCQPCHPKKGKDVHTIVLIKYADIFNHIEAHSHV VQTAITDNMKTYLRLTKENXFYCSYRNNKKKNKCKKAFNLESNMMDITE HMKTHTGYSFDXNLNILCYCGIWKPFTELIAHIKTEHLQEYINSIPNKENIH NTTTIVSPLNFAGILASGETQNIPDEEIIKPRDLPENLAFNRNIENELSWSQ HLVKAYIFSYAVKTSTIFINPYTCNALIQCNYKTFFETFPFKDFAKWNEIVLP IHNNTSSWSFFFLNKKKRVAMIIDPTADDSHTLHFELATDILRTILNVQNIF EDLNFPLTEVEYPVCHEANLSAFXVCHFLKCLMSDLPIDIPDIDHMKETM RPIIRKYNCAKFPESDVRNYRVLIEDLIYQLNLDTITCEEILCEIERINGRLNP KRYFKESKPKTDIIHLQKKKSAELLCVKRLKFQISQKTEIGKIWENDDVDHR
NeSL	NeSL-4_SM	.	Schmidtea mediterranea	TTAAATCATTTTTAA ATGTGTTTGAATAT CTTAAATTATCAAAT CATATTAATATCAAT GCTAAAAAAAAATC GTGCKCATCAGGCG CACGAAAATAATGG ACACAACTCGTCGA CCTGCTGTCGACTC ACAGAGAACCTCAA TTTGGAAGAATGGG AAGCCTATAATGCT ACAATTCCGCCAAC CCCTATTTGAATGA	TGAGTGTGCTACGAGGCAG CGCTGGTAATTGCATCGGC GTTGCAGATTTGTGTACGAT AGATAAAAACCAATAGTAA TAAATGCTGAGCCTAGCTC GCATATCTAAGCCGAAAGG CAGCATATATATGAGACAA TTTAAAAAAAAAA (SEQ ID NO: 1299)		

			CAGATAGTCAAATA		PPMARFLKTFASQDCPVSNTSSINLPYYMDTDTDXCTDCENLSHIMKNL
			TCAAAAAATATACA		DSSAPGMDLITGGDWKKISPKHELITAICNCILRNKVCPEKWKLFRTVLILK
			AACTGCTGTCAAGC		PGKMSESFRANSWRPLAIMDTAYRIFTTLLNNRLLQWIRNGNLISPNQK
			GTGACTCACTTCCTT		AIGIPDGCAEHNATLHFAIDRAKRCKTELHIVWLDIADXFGSLPHDLIWYT
			CCAATCGAAAAATA		LANMGLKNETLTLIKELYKDVKTIFDCQGTLSEPVPITKGVKQGCPLSMTL
			GGAAKATGTAAGA		FCLSIDYILKSILTNYPFLLHDLNISILAYADDLVLLSDSYLEIKKSLESTVELAA
			AACATGAAAGTCAA		FANLKFKPSKSGYLSINNVNSDILKLHLYNEEIPTISENNKYRYLGVDFSYKR
			GCTGAAAAACCAAT		NQDVDGRLGSALALTRSLFKSYLHPAQKLNAYKTFIHSKLIFSLRNCVIGHR
			AATATGTCCTAAAA		ILDCDRNRVTQGREKQLGFDQEIKALLKTMIGDKFQAXNNYFPYTHCKLG
			TAAAACAATTTGAA		GLGITSAIDEYLIQSITGITRLFHSSNLSFRKMLITELAHSRGGKNFEAGLKW
			AATATGCAAAAAAT		LNCEVNKAFPNTSFFVKFQKSALALKRKFCICVNLKFVEDNFSLEMTYKKR
			ACCTATAAAATCAC		TSYVNHQNLSTLSKELHDFVGLYYAEQXCQMRVQGHIATAIGDSITAKYLI
			AGCCGAATAAATTC		ASDILNDAQYYFLVRARNNLLNLNYNAYRLKYNIGTKCRLCHLDEETQAH
			CCATCCGTTCTAAG		XFNHCRAKPNARRVKHENVLVSIVAFLEKIGFEIDVEKSPKYISIPTKLKPD
			CAGAAACCGCTACG		MVIRSKRNKDIHVLDLKVPYDSGEGFEKAREDNYVKYKYKDLAIEIGKAFNQ
			AACTACTGCAAGAA		KATISAVVIGCLGTWDKKNNAALSKIGLTKTEIISLARIACPNAVIACYHIYR
			TCGGATCAAGTATA		EHVSFTKSAMALPFSLA (SEQ ID NO: 1421)
			TTAATTTCCCCCCM		
			GGGGGAAATTAAT		
			ATACTTGTTAKAAA		
			ATTAATTTTTTTAATA		
			AAAATAAATAAATC		
			GAATAAATATAAAA		
			TAAAAATAAATCAA		
			ATTAAACTTTTATTA		
			ACAATAAAATCGCA		
			GTAAGTAAATTTCC		
			ACTGTTATTAAATTT		
			AAAACAAAATTCCT		
			TTAAAAATGCCTCT		
			CTTTTTCAGTAATAA		
			CACCTTTTCTTGCTT		
			TTATTACTATTTCTT		
			GTGTACTGTACAAA		
			TCGAGCACAGTTAT		
			TGCAAATAGGACAT		
			AGAAATTCCTTTTTA		
			AGTAAATTTAAATC		

			CATGAGAAATAAAA TAAAATCCTTTTGAT TCAAAGTTTCTATGT TGCTTTCTAATAGA ATGGTGTAAGCATT AATGGGTCTTGATT TTTATAAATTAAATA TATTTAATCTATTAA ATTAATATGTTTTTA TTAATTATTAATTTT TATAGTGGGGGGA AATTAATATACTTG ATCCCAAGAATCAA CTGATGATGAAGAA TATGTTATTTCAAAA TACATACAAGAAGC TGGAAAAAACAAAT CAATCGCTACA (SEQ ID NO: 1176)		TGATCCGTGTGTTTGTGTCG TATGATTGTTTCCGTGTGTG TCTATATTTTTCTTTTTTATA CTTTCAATTACCTCGTTGTA ATGTTATAACTTCATATGGA ATATATGTAATTTAGTTTAG TTTAGTTAGTTTAGTTTAGT TTAGTTTAGTTTAGTTAGTT TAGTTAGTTTAGTTAGT (SEQ ID NO: 1300)	TTGRNLGQWSCYSRSIQQSNYSFKLSSTEVGELVEQSPAPLQSPQFSNNY NNLNINNNLYYSLNTFNQSNNLCCLVNIEFFPTQHLLGDIVNSGCINYMN NYNNFDNINLYINSNVLSYNNYNHSFLASPYTTNITEHADINNMHVQEVN MQQDNNTQHAITQQVSLQATSLQHTLDEMIVQFNTAVRLKKKHKVAKI FRGHNHRKDLPTLPAREQYKTKPKLAIREVLHRKTTATSSPSENAIKAFFSS YSRPAELFTGQELLESSWFPVHPEDDFEFRIPGRDQIAKYIKFASKSAAGL DWITYEDIKLGDPSGEILQPIFEYIVQNNICPSEGKASRTIMIPKPGKSDYS DPSSWRPITITSAVYRLLMKYLTWELYNWILLNQMLSRSQKSLGKFEGCH DHNAMLNMLIQDVRRQTNPSNPINKNKRLYIVFLDFTNAFGSVPLDTLM YVPQRFGLGTSALTLIKNLYLDNYTNVTCGESKIENVKLNKGVKQGCPLS MLLFNIFINIIIRAIEAMPDVHGYPLGDMDIRILAYADDIALISDSHKDLQE MVYKAEYIGRILGLLFNPSKCALMDIPHDKKRTPPILVNGEMIKCVGKAD PYKYLGTFRSWFRKLDIKELLQMMMDETKLITESNLHPHQKIHAYETFIHS QLPFHLRHSRIPFSDFITNRKTNKTTNNSNDSEKSIQKAYDPESGQLFLNT FALPSGCAKDFFYITKDAGGPQLTSGLDEYLIQSIMYIFRLLGSEDPTLNSAI KHDLISHLNLKGFVNINFSQAISIFNSNFTDRTDHFSHLSRTEWARLQLAR KKLKSTLAIQTNVCLINGHLVLTLSLENNVLLIDSKEKGDVKKIHASLMGFL RLAHLIRLQKHGWSKLLFSATTHHEILNKRILNGHVPYKIWYFIHRARLGLL PTKLFSVSNLCRKCGGKKETMSHALVNCPMMQTLINERHDALEISLVQIL
NeSL	R5	AY216701	Girardia tigrina			

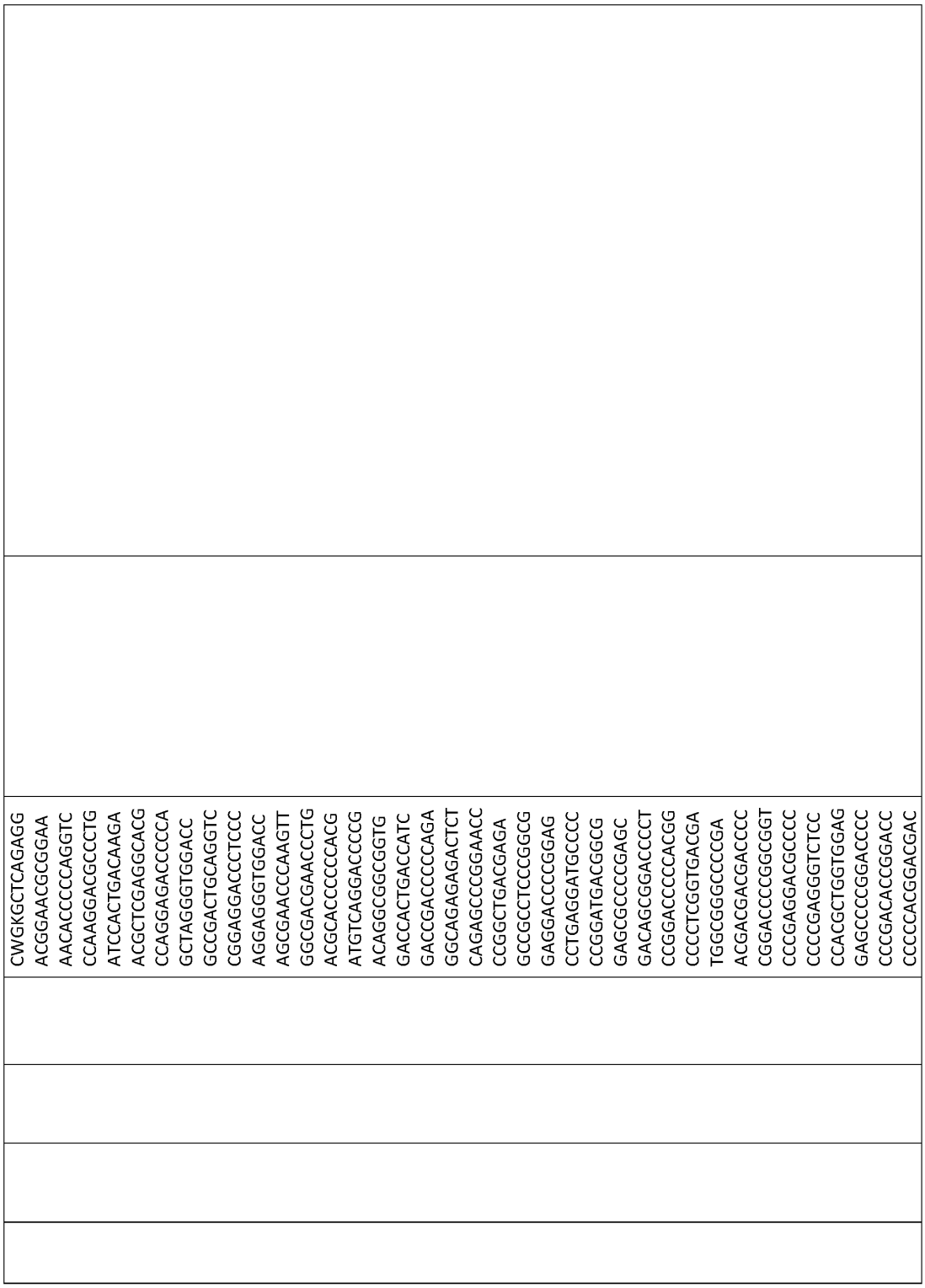
				SSKFQGTVIRQKTYVNELRPDITMESDTQYYLVEVKCPFDTKMSFELRTQ QTTDKYNIIIEILEDVHPGKEVRLVTFIVGTLGSWGPQNSDFLRDLGFSKD EIDQVKTRLMLQNINSSCEQWKRFVQYAPTITPGPIPDAESEDDQGTSD NGPTAATVQGPVIGDEEEELQIYDSGLDESSDDEPDPDDAELLFTIDIEQY LNSVITD (SEQ ID NO: 1422)
		CTAAGAAAACAACT		
		GAAAATAAAAATGA		
		AATTCCTATTCAAA		
		AGAGTAAAGATAA		
		GAAAAAGAAGTCTA		
		AAATTAATACCGAA		
		AAACTAACTGAAAA		
		TATTACTACTTCTGA		
		AATACCACTTGAAA		
		TTGCTCCTTCCATAC		
		CTTTACCTTCAGCAA		
		GTACCTCGGGTTCT		
		CAACAACCGGCCAA		
		TCCTCCAGAAGACG		
		CTACTCTAAGTGAT		
		ACGGATCTCTTCCTT		
		ACACAGGATGATCC		
		CGATAGTCTTATTCT		
		TTCTGGAAGTACTC		
		AACCAACCTTTGTT		
		GACCTCAACCCTTC		
		ACAGCAATCGGAAC		
		TTCCTTCAAATACTG		
		ACAGCCAAAGATTT		
		GAGGCGGGTGAAA		
		CACCCAAAATCATA		
		ACTTCTTACAGGGA		
		TGACCTTTTCTACTC		
		TACAGTCCTTCACTA		
		CAACTCAGATACAG		
		GTTACGGTATAAGT		
		GTTGACAATGGGG		
		AGCAGAGGTTTCGA		
		ATTCTTGCTAGGAA		
		TCTTGTCAGGAAAA		
		CCAAGGATAAGTTC		
		CCCTCTTTATATGCT		
		GGACAAGTAATTAG		

				ACACACAGTCTTCTT CAATCACTTCAACC AGGCATACTACGCC AATAATATAACTGA TAGTAAAGGTAATC TAATTGAGTTTTCT GATGATAAGCCTTT TCAAAGTATACCGA CTGACCCAAAAACT GAACTAGAGCAAAT TAGGAGAGAGAGA CAACATCTAGTTGA TAGAGCTCTTAGAC ATAATCAGTTACGG GAAACTTATATTTTA AATAAACTTAATAA TAATAATGGGGGGG GGTGGCGAACATTT GAAAAGGAAAAAG ATCAAAGTCAATAC GGATGATGTCTCCA GCAATGATGGAGA CAGAAAACATAG (SEQ ID NO: 1177)	TAACCGAGACCGCCGACCA GGGAAATAACCCACTTCCTT CCCTGACGAACCAAGGGAC GCACCCCACCCATGTACTTA TTCGCTACACCGATACTGAC TTGGACTCCTTATACATTCC ATGGGTGGCGTACCCGAGC CCACTTATCCACTGACACTT TAAAAACTCTTGCACCCCAA TCTGGGTCTATGCCGGTTAT GCGATATGTATGTATCTCTT CATCCTTGCAACCGATACCT GTAATCCCCTCATAACCCAAG CCTGACCCCAGATGTACAG	MESPAXIFEKIDAALXIYSAAAXLXXNSLSLSPXXAXXSXXAAPASSTPQKT QXKPIPXTTLGASRKXRTTXKDEXIXXWXKKAPVDTSXGRXSTRRTALRDL TSRSXNIXXALQEEDPRRTPPXSRDQDAERRPAAPEKAATRGAPPTIQDQ DADRCPAGRDATGGAPRRPRTRMLXAAPLGRMPPEEPPPTTRDQDAD RRPAAPERDAPEGTTSSTPDPETTYHPPVRRRAAPRGTHSXAXDLDAAR CPSGQRDIVASESSTPPGATSPPQASLPDXEESPAESAGTTEVRPTEGEA GEDDCIYLQYPXPTGLLLCPFCLPXHGVQTLGALSKHVRKAHNKRIAFRCS RCDAPFETQKKCKXHXATCKGPLTTAKVNPTDTLRVPTPTPTDGPASAP QPASPEPQXVRGDQPPTEGSATPASRTDDATKRTSPASRIPTLDPAVRGI TATSQVSDLTRCLSDLIKTIRHNTDTRRXSAPPQVTSCRPAVGATSTAPQA ARRDPANGGASRSPQIPRPDDPAPGRPNTSSKVTQRDSDRQKPHAPPRT PQPDTTRRRTRTIPSASKHDRAPTKPNTGVSRTPLPPGRSSAASETPRAA PPPPHQDPRLKTHLNTAPQSEGQQGHRLSPQHLNPRRQRSRRNDGRE QRVATPWQSAWMEELAKAEDFETFDTLMDRLTAELSAEITARRREPQE

				TACCTTCCCTCTTAACTCGT GTATATTTAATTTTAAACAT TAACTTTAATAAAATTTTTA AA (SEQ ID NO: 1301)	ASRATRRFPAPTRNNTAREGRRGDVGRRYDPAAASRIQKLYRTNRTKA MREILDGTSSYCAIQPERLYSYFKDVFDHEAQTNLQRPECLLPLPRINLTED LERDFSPQEVQARLMRTKNTAPGKDGIRYHLLKKRDPGCLVLAAIFTKCK QFHRVPRSWKKSMTVLIHKKGERDDPGNWRPISLCSTIYKLYASCLAARI TDWSVCGGAVSSVQKGFMSCEGCYEHNFLLQTAIQEARRSKRQCAVA WLDLTNAFGSIPHHHIFATLGEFGMPETFIQILRDLYKDCTTTIRATDGET DAIPIRRGVKQGCPLSPIIFNLAMEPLIRAISSGPTGFDLHGKKISILAYADD LALVADSSESLQQMLDVTSQAAEWMGLRFNPKKCASLHVDGGARALV RPSRFLIQGEPMASLEEGEVYQHLGTPTGVRVRQTPEDTIAEILRDAAQID SSLLAPWQKINALNTFLIPRISFVLRGSAVAKVPLNKADSTIRQLVKKWLYL PQRASTDIIYISHRQGGANVPRMGDLCDVAVMTHAFRLLTCPDPTVRSI AQEAVRDVVRKRIARAPSEQDIATYLSGSLEAEFGREGGDLSSLWSRARN ASRRLGKRIGCCWKWCEERRELGILVPRIKTPDHTIVTPTARAMLERTLKD AIRCHYAENLKRKPDQGKVFEVSSKWDASNHFLPGGSFTRFADWRFVH RARLNCVPLNGAIRHGNRDKRCRKCGYANETLPHVLCGCKQHSGAWRH RHNAIQNRLVKAIPPSLGKITLDSAIPGTDSRLRPDIVVTDAEKKKVLMVD VTVPFENRSPAFHEARARKALKYTPLAETLRAQGYEVQIHALIVGALGSW DPHNEPVLRACGVGRRYARLMRQLMVSDTIRWSRDIYTEHITGHRQYH TE (SEQ ID NO: 1423)	MAAKSVACPHDGCANKYASEASLRRHIKNKHATDEEGDETSHSCPHCH RPFSTARGLSVHIGKSHRQAPPEPTRPPPAPAPADPGLDPDPGPTVTPPS RDDEDREEPDDDPVEIADLSCPHCAQALPSAHGLANHLRACKDHRVPAP GAPRSGPPSSRYWTAVEHHRYVEAMARFADHPDLLARAAAHIGTRTYK QVDSHRTKVIAAEREGRPVRTLDPTMDWRMRPYCASTTARWLAEQGR SPVAPRSPCPEPHAPPPAAALLYIPATPPAPTPRAPVAPPKLAPPAESTVP ATPDGNPEAPAPPFSAPGPPTPKALPPPPPSRRNLRPHLVPKDAWQGV ADAVAPAASRLLRTPLAHLSTEQWATFEAALAGLEATLHHAARSAEAVP TRCASRAREDAERQLREARKTREIFGKAAALYAAGKDPTATIERIPPEVRL HLPTPGSAEWPARAAAAARRVIRRAVARADRLRKRMGILDSDRDLQRLFN ANQKKAVRQILAPSTKAPRCQLDPAAVEEAYIQTLAKPPPIDPSPPWKNS VQWPRPPTAADDGGSPFSVAEVRAQLRRLPNGSAPGIDGIPYEAYKRTK LDATLAHVFEVVRLNARLPARWDVARTVLLYKKGDPNDTGNWRPISLQ VTIYKIFTAALSKRLISWAGKHNTFSASQKGFLPAEGCHEHAFVLRSVLDD ARRHKQNVYLAWYDLRNAFGSVSHDLIAWCAAMLGLPRYLRDAIGAIYR HSALFVQVGDQETTGVIPMRCGVKQGCPLSPLLFNLCVEPALRCLRRTTG YKFYGTSITVEGQAYADDLLTAAPSAYHAARQVATIEEWANWAGVSFVV QALSLDAPAGKCAALAINFEGGLMHSIDPALKVQGAAIPAMSRNNVYRY LGVHVGLTDALGQANELLEKASRDARTICASGLEPWQKVVAIKTFILSRLP

						FFFHNGKIQRGRCQQFDRELRENLRAALRLPVCTTNAFFHSRVASGGLGI LPIAEEQQVYLAAHVFKLLTSPDLSIRAIARHQLAEVTHARHTTPVQDGEA SPFFGWLMRGQEVASTTPSGDVSSIWFAAAGAYSRMGWSVRDALHPT LTVGPGVQFEGRFQRANVIPALRASAFSRHAVEWSALRTQGRAAAYQH AVHPATHHWVHNSAGLTTKEYRFAIKCRLGLLPTRAAPHHRNGPTACRA CSYARETANHVLGHCPATKAEVIARHNRICRALAQAAEASWTSVLEDVPI PGVDSPLRPDIYCSRPGQCAIIEVAVSYEDAFNASMEGRAKQKTDKYAGL AATVEEQLRLQTRHAAFVVGFSGVVLPASVTATATSLDLPPKTWNVLLKR CVAASIKGSYTAWRRFRRSTP (SEQ ID NO: 1424)
NeSL	Utopia-1_AEc	.	Acromyrmex echinatior	GGTGCACAACGGATGCATCATACGTGTACCGGAGCATACGGGCTGTCACGGCGGCTGCATGCGCGATCTAGCTCGGAGATTTTATTTATTTATTTATTAATTTATTTATTTATTCATCGAGTGTGAGTGTTCGCGTTTTGCCGAGAAGCGATTTTCGTTAAGTGATACGCGCCGCGTTCATAGGTTAG (SEQ ID NO: 1180)	TAAATTATTTTGTCTTTGTCTTGGCCCCCCCTTTTTAAACCAAGCAGGAGAGAGTGGCCCAATGCCCAACTATTATATATTAACTATTTACTGTGATATTTATTATTTGACTGTTGGGCGGGCCCCTCTCTGCTGGTTTTATTTATATATATTTTTTACTCGCGTACTTTTTGTACTACTCTATTTTTCTTTTTATTTTAGCTATGCTATTTTTATCTCTTTCTTTGTCTCTATTTTCTTTCTTTTTTTCTTTCCTTTTCTTTTCTTTCTTTTATTCTTCTTTTATTTATCTTTTTTTCTTTTCTGTTGTGGGGCCCTGACCGTCCGAGTGTGAATGCCGCGAAAAACAATATTATGTTTTATACGAGTGTGCATGTGCGTGATATATTTATCTATTTTATTTTATTTATTTATTATAATTTATTGCCGCGCGCGCTCCTCCGGGACTTTTATTCGTTGACAATACTGTGATATTTTTCTGCKCAGGCTGGGGGGGGCTTGCCCCCCAGCCCCTTAGTTTTAATTGCCTATGCGGGGGGGGCTTTTGTCCCCCGCAAATGTATATATATATATATTTAGCGCGCG	VCSVRGCRREDSRRFYKFKFPLNFVKVPKTIVIGSAFQKSSVSARSQNHSR STRVPKTRQPRTSNTIGRYTAASANNYLTVIITGNYTVFAQWICYRECTW LLSKFVNFFLTIIGYFFQLRLVVIYEGPVILDTFSNCGSSLFMRGQXSKALLV RLNRSALAMADPQVHYIDYPLPPRVKCVKCFGAEGAGKVKGEYSDPPHL AKHLKKCHPGDTLNYKCSICDLRGTGKYPLRDVKAHYAECHVSPAVDAA GPSTRGSLGECSGAGQPTASRAAKATTRLAETVGGTDKRRAATSGSRQL TLPFAATPSPSTAAGEARAPRSXSTTPTSRSPSYAAVTAGPPSMRSTTTST TARSKTVAKGAAPNTTTTTTARRSGEAAATRKPPTTATVSKPRVLSVETV RLPVDDIQRAGVQNAAKPARAPSRPPQRTSPEAGGPRTTGAKEKCGEG AYKKLPANSGNPISTRTRRATSVPVEKSEGTARRERVSPHPPPKGIDIILSS TSEEEGTPYQPGGVGRLRLRRKKVTGPPPKMTPREGVVTRARRSTSAPV EKSALDARLTALDRTSSRATGNPTSQIAGGLYTSRGQPERTPPARLPSLSP TTRGSPSGSLGEIRTPISPATSLPATLTTCTVTTTTCGSPITSTGFTGGVGRLI TPPSLPQTNILPTIGEEGTSPCVAVVTTHPRPTGEDAPCEAPQPVSDHQR QSIGEPRRDTDTHLACDVATGNAHHLHGDDDHLWKPHNIHGLHRWRG EADNTXEPPPNEHPPDHRGGRNVTVRGGRHHPSLGGRSTSPLILPRPTT PEPERGQEERRLEGAAQPPTTPVVEGDNQWDGQWTVSVRRRARRQQ LNDTSPSNSESPPTAGPSRSPRIAPLSALIAASTSRHETSLNLNCTNGNICM DRTPPRNILPVXAERRRETSPQDRVEGDIGYGAGKVSAEHPSAPVNVRG VMSRGRATASSIVPPRANRGEGGRQHHSRRRPDAPVGQPSRDHPAPAT VARQRRRERVAARDALLDRAKDVATIADLEAFAASVAAFFGEDASATGA AARARDRSVRSREAGARRGVKGGERPEREGAGRPGSAPADPGASGEAR GDWVREAKRLQALYRANRRKAVREVLQGPADQCQVPKRQVQEYFERL YSGGEDLAGAGVEAERPDPSSPREVSAVLGPLAEREVDRRLRRMNNSAP GPDGVSYRDLRGADRGARLLTALYNICLRLEAVPASWKTSNTVLIHKKGD RGMLENWRPLALGDTVPKLFAALLADRLTDWAVTRGKLCSAQKGFLRD EGCYEHNFVLQEVLTHAKRSKRQAVVAWLDLSNAFGSIPHATIRRALIRS AVPRGLIAIWDSMYDGCTTRVRTAEGHTAPIPIRSGVRQGCPLSPIIFNLA IDSVVRVAAEXNDGYSLHGNTWSALAYADDIALLAQTPEGMERMLASV

				GCTTAGCCGCTTTTGTTTGT ATTACCCCAGAGGGGAATT GTCCCTCTGGGGAAAAAAA ATGATTGGAAAAATAAAGT GAGCTAA (SEQ ID NO: 1303)	EAEAASVGLRFNPAKCATLHVGAGNGGRVLPTSFQIQGETINPLAQGES YTHLGVPTGFSVDQTPYAAVGDIVSDLRAVDRSLLAPWQKIEMLGTFILS RLDFLLRGARVFKGPLTAVDLNIRRHVKSWLNLPQRASAEGVYMPPRW GGCGLLPLSDLADVLTVAHAYRMLTVRDGAVRELAWESLRGVVGRRIG HAPSCEDIASFLSGSLDGRMRGGGEASLWSSARNAALRQSERLSLRWR WVEATEEMTLECRGPRGAAIKIPPEARGQVVNRLRSAVAEHYASRLLSKP DQGKVFEVSSRSRVSNHFIRGGSFTRFADWRFIHKARLDVLPLNGARRW EANDKRCRRCGEVSETLPHVLCHCGIHSAAIQLRHDAVLHRLWKATRLP GVVRVNQRVEGVSDELGALRPDLVVRHEPSKSVVICDVTVPFENRWTAF EDARARKIAKYSPLAEELQRRGYRVVVTAFVVGALGSWDPRNEAVLRLLR VGNQYAAMMRRLIVSDTIRWSRDIYVEHVSGTRQYLAPSRPSGDLATPP RAVRRRWLAEERSAQDAARRGSDSVSVA (SEQ ID NO: 1425)
NeSL	Utopia_1_AMi	.	Alligator mississippiensis	TGCTGGAAAGACG GAGAACCGCTTCCT TTTTCCCTGCGCCTG GCCTGGTATTGCAG TACCTCCAGGATTA GCGCCAACTAGTCC GGCAGACTGTCGG AATACAGCAATAGA AAGWGAGCTGACT AGCAGCTTGCTTTC CTTCCTCCGGTGCA GCATGGGTTCTCGT CAGTCMTGACGGG CTAGGGAAGGCGG TGCTGCCAGTACGT CCGAAAGAGTGCC GGTTGCGCAAGCG ACCGCGCCACTCAG GTGAGTAGCCAAG GGTCTTACAGTTCA CCGGACCCGAWAA CGCGAAAACCCCAA CTCGGGCTAGTAGC CGAAGACCTGGGTC CCCCCCMGGTCAG AGTAGGCGAACGC	TGAACCCCCCCTCTGCACCA GATGGACCTTCACTTCGAG AGGATTCTTCAGCAATGGA CGACCCCGCTCCACCCGAA GAGGACCCCCGCGATGAGA CTCTATATGGACTGAGACA CTTTTTCTTCGAACCACTTCC TCCACCATTGCGGACCATTG TAACGGGTTTGTGTGTATCT ATCTTCTTTCTCTCTCAGCGT CGCGAACCCCCTCCCTCCCC TTCCCCTCCCCCTCCCCCCA CCCCCGGGCTTAGTTGGCT AACATTGTATCTCCTGTAAC CTAGTTGCGTTCCCCTCCTC ACCCCCATCCCTCTATTGTT AGTCCCTCGCTCGGGCGCT CTGTATTTCCCTACCGGCTT TGTCATCTTTTTTGGATTCA CAATCCTAAACATCTACTAA TAAAAGTCAATC (SEQ ID NO: 1304)	CHHAGLRPGTPNRTRRPDQTAPLPDPRGHPMPPNRRGSRSRPEEPSRR EPPXPRACQGLRVWSPPQQRMPTPWQTLWLEELSRATTFKAFEASVAR LTEELSAAARPGQPRGGNNRPATRRDHRLQPQRRPRRQRYDPAAASRI QKLYRANRPKAVREILEGPSAFCQVPRETLFNYFSRVFNPPAEAAAPRPA TVEALTPVPPAEGFEDAFTPQEVEARLKRTRDTAPGRDGIRYSLLKKRDP GCLVLSVLFNRCREFRRTPTTWKRAMTVLIHKKGDPTDPGNWRPIALCS TVAKLYASCLAARITDWAVTGGAVSRSQKGFMSTEGCYEHNFTLQMAL DNARRTRKQCAVAWLDISNAFGSVPHRHIFGTLRELGLPDGVIDLVRELY HGCTTTVRATDGETAEIPIRSGVRQGCPLSPIIFNLAMEPLLRAVAGGPG GLDLYGQKLSVLAYADDLVLLAPDATQLQQMLDVTSEAARWMGLRFNV AKCASLHIDGRQKSRVLDSTLTIQGQAMRHLRDGEAYCHLGTPTGHRAK QTPEETINGIVQDAHKLDSSLLAPWQKIDAVNTFLIPRVAFVLRGSAVPKT PLKKADAEIRRLLKKWLHLPLRASNEVLHIPYRQGGANVPRMGDLCDIAV VTHAFRLLTCPDXTVSIIAASALEETARKRIGRQPTRRDLATFLSGSLEGEFS RDGGDFASLWSRARNATRRLGKRIGCAWTWTEERRELGVSLQPAPHAD RVTVTPRTRTFLERFLKDAVRNKYAGDLRAKPDQGKVFDVTSKWDSSNH FMPSGSFTRFADWRFLHRARLNCLPLNGAVRFGHRDKRCRRCGVVAETL PHVLCSCKPHARAWQLCHNAVQDRLVRAIPAAAGEISVNRTVPGCESQ MRPDIVITNEEAKKVVIVDVTIPFENRRQAFTDARARKREKYAPLADILRG RGYDVTVDALIVGTLGAWDPSNESVLHACRVSRRYAKLMRCLMVSDTIR WSRDIYVEHITGHRQYTDPTRRTAAGPDPEGTA (SEQ ID NO: 1426)


CWGKGCTCAGAGG
ACGGAACGCGGAA
AACACCCCCAGGTC
CCAAGGACGCCCTG
ATCCACTGACAAGA
ACGCTCGAGGCACG
CCAGGAGACCCCCA
GCTAGGGTGGACC
GCCGACTGCAGGTC
CGGAGGACCCTCCC
AGGAGGGTGGACC
AGCGAACCCAAGTT
GGCGACGAACCCTG
ACGCACCCCCCACG
ATGTCAGGACCCCG
ACAGGCGGCGGTG
GACCACTGACCATC
GACCGACCCCCAGA
GGCAGAGAGACTCT
CAGAGCCCGGAACC
CCGGCTGACGAGA
GCCGCCTCCCGGCG
GAGGACCCCGGAG
CCTGAGGATGCCCC
CCGGATGACGGCG
GAGCGCCCCGAGC
GACAGCGGACCCCT
CCGGACCCCCACGG
CCCCTCGGTGACGA
TGGCGGGCCCCGA
ACGACGACGACCCC
CGGACCCCGGCGGT
CCCGAGGACGCCCC
CCCCGAGGGTCTCC
CCACGCTGGTGGAG
GAGCCCCGGACCCC
CCCGACACCGGACC
CCCCCACGGACGAC

				CCAGGCGAAGGCG		
				TAGACATGACAGCA		
				CTCACGTTCCTCCCC		
				TTCCCCCTCCCGGC		
				GAAGCTGTTCTGCC		
				CGACCTGCCACCCG		
				CCAAGACAGTACAG		
				GTCGCACGGCGACA		
				TGAACAAGCACCTA		
				CGGCGCTTCCACCA		
				GCTGCGCCTAGCCT		
				TCTACTGCGCCCTCT		
				GCGGCACCGAGTAC		
				GAGGCCCTGAAGCT		
				CCTGAAGAACCACC		
				AGAAGGGATGCGA		
				GGGCCACGGAGCC		
				GAGAGGAGACCCG		
				GCACGCTGGTGAG		
				GTCCGCTGCCCCGG		
				CCCGCCGGACCCAG		
				GCCGCGGTGCGAA		
				GGCCCGCCAGACTG		
				GCCACCCCGCCGAC		
				AACCCCACCGGACC		
				AGACCTCCAGGGAC		
				CACCCGACGGAGA		
				GACCTGCCCCAGTG		
NeSL	Utopia_1_CMy	.	Chelonia mydas	A (SEQ ID NO: 1181)	TGAGCCGGTACGACATCGT	MTTKKVLGASTTLQTSSTKGKNSGCSKDPLRDAVPGRSWILRPACRDITT
				CTTCTTCTTATGAATA	GCATCAACTATGAGAAAGG	RRNIPPAPQQQQPPMESPPTLQLQDALRRPSPTPAAAQVADAGGALAA
				CTTGCAACACCTGC	GACTGAGAGACTTTTTCCAT	LHTIKRGISVDWTSISPKXXQRXTSASPDACPASETTQRDXRXLLDARPAG
				ACTGAAGATGGATT	TGGACCATATGAACTGGAA	PLDPTRPHQDEPASDTADAAGTPLLQGNEDTIYLQYPLAADMLICPICSP
				CTCCGGCTGCTATT	CCATAAACTCACTGAACATT	PQSFHLLGVVTRHLKRCHSKRVAFSCALCSLPFETQKQCKMHQVACRKC
				TTTGAAAAACTGAT	AAATCTCACCAAATGAGGG	LKGTTQSPAPAPSPPAARRPAAPEPQRRKXTSQAAVKKPAPVARPAERD
				GCTGCTTTGAAGGT	TAAATCCATCCTCCATCATCG	AAIEKVPAASGNITQVLASRRPVSPSHVAKXISMLRRLSAASPPVQHVPV
				GTATTCTGCTGCTG	TATCCACTCATTATACTCCA	PRRISAPPRIAARDPVAGRASAAPQTALRTPAAGGASTTPQTALRTPTAG
				CTACCTTGGAAGGA		

			AATTCTCTCTCTGCT CCTGAGACATCCCC AGCTGCACCGTGTA CCACCACCACCACT GCTGCTGCTCCACA GAAGGTTTCTCGGA CA (SEQ ID NO: 1182)	CACCTGAACATAGCCATTAT ATGAACAACATACCCCCATA TCTCAATGTCTGTACTTTGA CCCGTTAACCTTTTACCCCC AATCGGGGATATTGCAGAT TATGTATTCCTTACGCCACC CGATCCTAAACCGAATTTCG CACCCCTTGATAATCTGTAC CTTATTCCCTGATAACCAGA AACTTCTATGCTTAAACTCT GTACCGTTTTTTTTTATTTCA ACATCATCTTAATAAAATTA TTAAA (SEQ ID NO: 1305)	GASAMPQTTLPXPRRPDWRNQPRSHSKAPGLHRQTDQHGPQVHSAG HCLREISRSSSNRLGSSHSAAATHRRTGGVPATPEPDRVSPTTSNAXIPPEI PPQHPTEGNPDPRDRRQADHTAGSEPAPDEVEDXEGQRPMVRAATP WQTAWTEELQAAASFDDFDLLVDRLTRELSAEIAPRRSSNQENAPPAHR TPAPNHNTTTRGARSRDASRRYDPAAASRIQKLYRANRSKAMREILDGP SPYCTIPSERLYSYFKDVFDRIARNDAQRPECLRPLPRVDEAGVLETDXTP KEVMARLSKTKNTAPGKDGIPYSLLKKRDPGCLVLATLFNQCKRFCRTPSS WKKAMTVLVYKKGERDDPSNWRPISLCSTMYKLYASCLASRITEWSVSG GAISSIQKGFMSCEGCYEHNFVLQTTIETARRARRQCAVAWLDLANAFG SMPHHHIFATLQEFGMPENFLRVIREVYEGCSTTIRSVEGETAEIPIRSGVK QGCPLSPIIFNLAMEPLLRAISNGTDGFNLHGERVSVLAYADDLVLTADD PESLQGMLDATSRAADWMGLRFNAKKCATLHIDGSKRDSVQTTGFQIQ GEPVIPLAEGQAYQHLGTPTGFRVRQTPEDTIQEILQDAAKIDASLLAPW QKINALNTFLIPRISFVLRGSAVAKVPLNKADKIVRQLVKKWLFLPQRASN ELVYIAHRHGGANVPRMGDLCDIAVITHAFRLLTCPDAMVRNIAANALH DATKKRIGRAPSNQDIATFLSGSLDGEFGRDGRDIASLWSRARNATRRLG KRIGCRWEWCEERQELGVLVPQIRSNDNTIVTPSARGMLERTLKAAIHSL YVETLKRKPDQGKAFELTSKWDASQPLPRRGRLHPFRRLAVHPPCPAQL RPAQRSRPPREPRQALQEVRLLQRDPAPRPVQLQAPLQSLAAAPQCHPE PPGESHRTAPGGGRRELRHPRYPASGTPANHFLAGGGFTRFADWRFIH RARLNCVPLNGAVRHGNRDKRCRKCGYSNETLPHVLCSCKPHSRAWQL RHNAIQNRLVKAIAPRLGEVAVNCAIPGTDSQLRPDVVVTDEAQKKIILV DVTVSFENRTPAFREARARKLEKYAPLADTLRAKGYEVQMDALIVGALG AWDPCNERVLRTCGIGRRYARLMRRLMVSDTIRWSRDIYIEHITGHRQY QEV (SEQ ID NO: 1427)
NeSL	Utopia-1_CPB	.	Chrysemys picta bellii	TTTTTTCTGATGCTT GACTGCAAACACCC ATCCAGAAGATGGA ATCTCCTGCAGCCA TTTTTGAAAAAATT GATGCTGCTTTAAA GATATACTCCATTCT CCTWKTTTGKAAGA AAACTCTTTTTCAGC TTCAGCTATTCTGTC ATCGGCTGCTGCTG TTCCTGCTTCCCAGA AAGCTCAGCMAAA	TGAGCCAGAGTGACATCGT TCTCCCACTACGAGAAAGG GACCAAGTGACCTTCTCCGT TGGATCATATGAACTGGAA CCATAAACTCCCTGAACATT AAATCTCACCAAATGAGGG TCAATCCATCCTCATCATCA TATCCACTCATTATAMTCCA CACCCGAACACAGCCACTCT ATGAACTTCATACCCTCATA TCTCAATGTCTGTACTTTGA CCCATCAACCTTTTACCCCC AATCGGGGATATTGCAGAT	MTQDQDADCCPAGKDATRGAPPMTQDQDADRCPAAPERDAPEGTTS STPDPKTTYHPAVRRRAARRGMHLRAQDLDAARCPSGQRDNVASESSA PPRATSPPQASLPDPEESPGESAGTTEIRPTEGEAGEEDRIYLQYPLPTGLL LCPFCLPVHGVQTLAALSKHVRKTYNKRIAFRCSRCDLPFETQKKCKFHQ ATCRGPPTTAKVNPTDILRVPTLTPTDDLASAPQPASPESQQIRGDQPPT EGSVTPASRTDDATKRTSPVSRIPTLDPAVRGTTATSQVNNLTRRLSDLIK TIRHNTDTRRCSAPPQVTSCRPAVGATSIVPQAARRDPANGGASRSPQIP QPDPAPGRPNTSSKVTQRASDRQKPHAPPRTHQPDAARRRTRTIPSASK HDRAPTKPSTGASRTPLPPGRSSAASETPRAALPTTPGPPPQDPPEHRST VRGTTRPQTVPAAPEPAETTQQEERRPRARVATPWQSAWMEELAKAE DFENFDTLMDRLTAELSAEITARRREPQEAARATRRFPAPSRNNTAREGR RGDVGRRYDPAAASRIQKLYRMNRTKAMREILDGTSSYCAIQPERLYSYF KDVFDHEAQTNLRRPECLSPLPRIDLTEDLERDFSPQEVQARLSRTKNTAP

			ACCTATCCTGAAGA CCWCCCTTGGTGCC TCACGGAAGACCCG GASCACCTGCAAGA ACCAAAACATTAGG AGCTGGCTGAAGA AACCCCCCGTGGAT ACCTCWGCAGGGA GACCTGGSTCCAG MAGGACAKCTCTTC GGGACCTCMCATCS AGGAGCAAGAATA TCTCAACAGCTCTTC AGGAGGGGGACCC CCGGAGAACCCTGC CCGCTTCCCAGAAC CAGGATGCTGATCG CCGCCCCACCGGGA AGGATGCCACCGCA GGAGCCCCCCCA (SEQ ID NO: 1183)	TATGTATTCCTCATGCCACC TGATCTTAAACCAAACTTTG CACCCTCGATAATCTGTATG TTATTCCCTGATAACCAGAA ACTTCTATGCTCAAACTCTG TTCACTATTTTTTTTAACATC ATCTTAATAAAATTTTTAAA TCTGTT (SEQ ID NO: 1306)	GKDGIRYPLLKKRDPGCLVLAAIFNKCKQFHRVPRSWKKSMTVLIHKKGX RDDPGNWRPISLCSTIYKLYASCLAARITDWSVCGGAVSSVQKGFMSCE GCYEHNFLLQTAIQEARRSKRQCAVAWLDLTNAFGSIPHHHIFATLGEFG MPETFIQILRDLYKDCTTTIRATDGETDAIPIRRGVKQGCPLSPIIFNLAME PLIRAISSGPTGFDLHGKKLSILAYADDLVLTADDPESLQGMLDATSRATD WMGLRFNAKKCATLHIDGSKRDSVQTTGFQIQGEPVIPLAEGQAYQHL GTPTGFRVRQTPEDTIQEILQDAAKIDASLLAPWQKINALNTFLIPRISFTL RGSAVAKVPLNKADKIIRKLVKKWLFLPQRASNELVYIAHRHGGANVPR MGDLCDVAVITHAFRLLTCPDATVRNIAANALRDATEKRIGRAPSNQDIA TFLSGSLDGEFGRDGRDIASLWSRTRNATRRLGKRIGCRWEWCEERQEL GIRVPQIRSDDNTIVTPTARGLLERTLKAAIRSLYVETLKRKPDQGKAFELT SKWDASNHFLDGGGFTRFADWRFIHRARLNCVPLNGAVRHGNRDKRC RKCGYPNETLPHVLCSCKPHSRAWQLRHNAIQNRLVKAIAPRLGEISVNC TIAGTDSQLRPDVVVTDEAQKKIILVDVTVSFENRTPAFREARARKLEKYA PLADTLRAKGYEVQMDALIVGALGAWDPCNERVLRTCGIGRRYARLMR RLMVSDAIRWSRDIYIEHITGHRQYQEA (SEQ ID NO: 1428)
NeSL	Utopia_1_DYak	.	Drosophila yakuba	TAAAAAATTAAAATGCCTTA AAAATAAATAAATATATCAA AATTTAAAAAAAAAAACGA GGAACAAATAAACACAAAT TCTGAAAGATTTATATAATT TAAAAWATAAATCGAAAAT AAATGTTGAAAACAAAAAA AAAATAATAATAATAATAA AAACACAATAACACTCACCC GGCCTGCCCCAGAGGCAGG TAAACATTTACTGGCCATAT GGCTTTTTTTTTAA (SEQ ID NO: 1307)	YAPGYEAAQSPCGREPPRDHHRRPRDACGSSHSPEPCLTTPRLLPETVSA EPCDDESQRTRYASPHKQARTLHDAEPRDASREHAPSCAEPRCHRCQW THWKDCCPHSTNTTDGPEGTDRCADTITSPATAACPQRSPCPLGSSNGC DETAPEKRQPAADLVHTAPFAVLVRAGPFADLVRAGPFADHHQDDDPL PHRSGSLGPLCSKQKDPRKTHQHRHSGQAGNQTHTDIPRAAPSRRAAIC LMANAAATREDLLRAATSLSEMAAANQPTRSPTGGGEPTSQGRRGPQA LADAAKRIQQIYRTNIPRAMRKVLRTLLTAVFSACLRTGHVPDLCKKSRTV LIHKKGDRTDLSNWRPLSMGDTIPKLFAAVMADRLTAFLTNGGRLSEEQ KGFLQHEGCHEHNFVLGQVLEESRRQGKDLVMGWLDLSNAFGSIPHAT IMDAVAGMGIPSRIRTIIHQLATGAATTAKTIDGMSEEIPIEAGVRQGCP ASPILFNIAIERVLRKIKTVNAGYLLYGSRISPLAYADDLVLIASSPEEMRSLL RAADDAAIEAGLHFNPKKCATLHLTGKKSSRRAVQTGFLVRGTPIPAMTE GDAYEYLGIPLGLKKNQTPRAAMEAIVGDIAKIDDSLLAPWQKIDAARTF VAPKLDFVLRSGATLRAPLRHLDTVIKKHIKKWLYLPQRASAEVVYTPLKK GGAGILPSSILADVLTIAQAHRMVSCPGEVVSRIASEGLREAVKRKINREP SGDEMAHFLSGSTLSGETASFGDAGFWSRVRMATKRQAVHLGVRWA WRGGELLVESRGQRNRPVATDSNSRSQLIQRLRCAAQDEFLTILINKPDQ

				CCGTATGCATGGAT GCGTCCGGGATGCA TCCCACTAGGTCGC TGGGCGAATACGG CACATACGCTGCGG CATATCAGCACATA ACCCGGCGCCACCC ACAAGTGGTTATTA CATACCGTTGCCGG GTCTGTGGCGCTGA (SEQ ID NO: 1184)	GKVAKLSTLTPVSNAFIRDGSFTRFADWRFIHRARLGVLPLNGAIRWGSG DKRCRVCGYQLESVPHVLCHCMHHSNAMQQRHNAVMDRLAKAGSRL GTPRVNCRVEGVAEDMAALRPDLVWRDERSRKIVIVDVTVPFENGAEA FDNARGEKEEKYRPLAEALRAMGYQVKLEAFIVGALGSWDPKNERVLKT LGVSRFYAGLMRRLMVADTIRWSRDIYVEHVSGIRQFTLPSGAPSN (SEQ ID NO: 1429)
NeSL	Utopia-1_Gav	.	Gavialis gangeticus	CGCTGGAAAGACG GAGAACCGCTTCTT TTTCCTGCGCCCGG CCTGGTATTGCACT TCCTCCAGGACCAG CGCCAACCTAGTCC GGCAGACTGCCGG AATAATAGCCTCAG AAAGAGAGCTGGC TAGCAGCCCTCTTTT CTTTCCTCCGGTGC AGCGTGGGTTCTTG TCAGTCCTGATGGG CTAGGGAAGGCGG TGCCGCCAGTACGT CCGAAAGAGCGCC GGTTGCGCGAGCG ACCGCGCCGCTCAG GCGAGTAGCCCAA GGGTCTTACGGTTC GCCGGACCCGATAA CGCGAAAGCCCCGA CTCGGGCCAGTAGC CGAAGACCNTGGG CCTCCCTCCCCAGG TCGGAGTAGGCGA ACGCCCGTGCTCGG	TGAACCGCCCCCCCTCCGCG CCAGACGGACCTTCACTTCA CTCCGAGAGGATTCTTCGA CCACGGACGACCCCGCTCC ACCCGAAGAGGACCCCCGC GATGAGACTCTATACGGAC TGAGGCACTTCCTTCGAACC ACTTCCTCCACCATTGCGGA CCATTGTAACGGGTTTGTGT GTATCTATCTCCTTTCTCTCT CAGCGTCGCGAACCCCCTC CCCCACCCCCCACCCCCGGG CTTAGTTGGCTAACATTGTA TCTCCTGTAACCTAGTCGCG TTCCCCTCCTCACCCCCATCC CTCTATTGTTAGTCCCTCGC TCGGGCGATCTGTATTTCCC TATCGGCTTTGTCATCTTTTT TCTGGATTCCCGATCCTAAA CATTTACTAATAAAAGTCAA TCTGTTCTTT (SEQ ID NO: 1308)	MSGPRQAAADPRPSTDPRRQRDSQSPEPRLTRAASRRRTPDPEDAPRT TAEHPERRRTPPDPRGPSATTAGPERRRPPPDPGGPEDDPPEGLPTLVEEP RTPPTPDPPDGRPRRGCRRGSAHVPPLPPPCEAAVPDLPPAKAVQVAQ RHEQTPTALPPAAPSVLLLPLRHRVRGPEAPEEPPQGMPGPRGREETRH AGEVRRPTTRAAARRPARPAAPPATPPDQTSGDRPTERPAPATPPRRSA PRDPRPDVTPRPDGPPPGPPGPPDAPDPPRIPEPPGEPEPPGALQLPSVP GSPGAETSAQQRMPTPRQALWLEELSRATAFEAFEASVARLTEELSAAA RPGQPRRGADNGPTTRRDHRPQPQRRPRRQRYDPAAASRIQKLYRANR PKAAREILEGPSAFCQVPRETLFNYFSRVFNPPAEAAAPRPATVEALTPVP PAEGFEEAFTPREVEARLKRTRDTAPGRDGIRYGLLKKRDPGCLVLSVLFN RCREFRRTPAAWKRAMTVLIHKKGDPTDPGNWRPIALCSTVAKLYASCL AARITDWAVTGGAVSRSQKGFMSTEGCYEHNFTLQMALDNARRTRKQ CAVAWLDISNAFGSVPHRRIFGTLRELGLPDGVIDLVRELYHGCTTVRAT DGETAEIPIRSGVRQGCPLSPIIFNLAMEPLLRAVAGGPGGLDLYGQKLSV LAYADDLVLLAPDATQLQQMLDVTSEAARWMGLRFNVAKCASLHIDGR QKSRVLDSTLTIQGQAMRHLRDGEAYCHLGTPTGHRAKQTPEETINGIV QDAHKLDSSLLAPWQKIDAANTFLIPRVAFVLRGSAVPKTPLKKADAEIR RLLKKWLHLPLRASNEVLHIPYRQGGANVPRMGDLCDIAVVTHAFRLLT CPDATVSIIAASALEETARKRIARQPTGRDLATFLSGSLEGEFGRDGGDFA SLWSRARNATRRLGKRIGCAWTWTEECRELGVSLQPAPHADRVTVTPR TRTFLERFLKDAVRNKYAGDLRAKPDQGKVFDVTSKWDASNHFMPSGS FTRFADWRFLHRARLNCLPLNGAVRFGHRDKRCRRCGYAAETLPHVLCS CKPHARAWQLRHNAVQDRLVRAIPAAAGEISVNRTVPGCESQMRPDIV ITNEEAKKVVIVDVTIPFENRRQAFTDARARKREKYAPLADTLRGRGYDVT VDALIVGTLGAWDPSNESVLRACRVSRRYAKLMRCLMVSDTIRWSRDIY VEHITGHRQYSDPTRRAAAGPDPEGTA (SEQ ID NO: 1430)

				AGGACGGAACGTG GACAAAACACCCCC AGGTCCCAATGACG CCCTGATCCACTGA CAAGAACGCTCGAG GCACNCCAGGAGA CCCCCAGCTAGGGC AGACCGCCGACCAC GGGTCGCGGAGGA CCCTCCCAGGAGGG TGGACCAGCGAACC CGAGTCGGCGACG AACCCCGACGCACC CCCCCCGCG (SEQ ID NO: 1185)	TGAATAGCATTTATATTGTG TTCCAAACAACATACTCATT ATTATATCTAAACATTTTTTT TTCTGTTCCTGACAATCTAC GTAAAGTCTGCTAACCAACT GGCATGATGAAATAAGATA AAATCCCCTTACACATTAAT TTCTTGTCACATCATAATGC TTTGTCAAAGCAATGTCCTA CATAATATCTCGATGTCACC CCAATTAATTTTACATCCTTC GGTAACCTTTATACCGTTGG ATCAACATATATGATTTGTA AAACTGTTATTTCTGAGTTT TTTCTATGCTAATAAA (SEQ ID NO: 1309)	MSCPREGSDHLGPDPETPALHQGSDIRVTSSRLRGSRGKSSRQPSSRHQ VPASEASATAQQTAANECQVCGSSFATSSGLRRHMARLHRAASADPEG AAPASITEIFDYPLPSRWKCSACSENFFNQQTLKRHQTRHHPATTFAYAF RCSSCRSEFDSARRAANHWQVHKKERSQLSGTEPQASSQARVSMAHSP PPLPNTSWAELASNPAEIPSFVWESPPKNRPSVEEFGSSLPTDVTMMSQ SPPPQVQSSPVPALTPLSPAATASSSPPGAARQLTPPTQTNTPVTQRARL QPEADVVPELPPSVTEHPVSDAQHWVDAVSSASDWSEFEAVCDQFVIH AVAVSRPNLARPQQQDRQRSGDHPPRQQRGQHRPTFDVREASRIQKLY RTSKKRAIRHILKEKSPSFSGSESDVLDFFREVYSAKEVDEEAVGKLASSLFD VPQGDDSATSLSLPTSAKEIGARLSRMTNSAPGKDRLEYRHIRRADGSFSI SEAIFNKCLAEGRIPAPWKTASTILLHKAGPTDDPANFRPIALQSCLYKLF MAVLADRLTKWACENQYLSPEQKSARPCEGCFEHSFLLSAALKDCRRNQ KTICIGWLDLRNAFGSIPHPVIKIVLSSLGVPDSLVTLLMDAYNGASTSFTL TGGQTDTVPIRSGVKQGCPMSPILFNLAIELIIRAVKKNASDNHLGVTVQ GKNLSILAYADDLVLLSRDTEGLQSLLQVAGSSASTLQMQFKPQKCATLTL DCKRGTNVRQSAHHIQGAAIPSLTEEERYRYLGVPIGLPRLTSLQESSRKLS SDIETISSSLLAPWQKLDAIKTFVIPILQYTLRATEYLKSDLKPLRAAIIKHVKK ICHLPVRSSNAFVFASRPSGGLAFVDPGVDADILVVTQAVRTLASDDDTV RAVALGQLTSVVHRTVHSAPSDDCIDKFLSGSSSEGPLANSGNSGQASSL WSRTRAASRRLKIRIVGASSGDIKVESGGGRAIPSKKVTAGLRSDHHNEMS EKLRSLPDQGKVARALSLDSFANATSWLTSGSFIRFCDWRFIHRARLNCL PTNAAVRRWKQNANTKCRRCDHQLETLPHIINNCRPNMVPIRRRHNSI QERLVKAIHYGDIYQDQHVPGDPNPRERPDITVVEGNKVTIVDITIPFDN

						GPDALSTAANAKVMKYDTLRQELASRGMDVEVHAFVIGSLGSWHGDN ERVLGRLGISRRYRTLMRRLCCIDAIKGSRDIYIEHVTGHRQY (SEQ ID NO: 1431)
NeSL	Utopia-1_NVit	.	Nasonia vitripennis	CCATTCCTTCGTACG GGTTTTCGTGCCGG CATAGCCGGGTGG GAGACTCGCGCGG GGGAGGTCATATCT CACCACCATCCTCG AGCTTTTTGCTGCA CCTGA (SEQ ID NO: 1187)	TAGTGGGGCCATAACACCT AGGCCCCACAGTGTGGCGA TGTCCATGTGTGTCGTCCTT ACTTATTTATTTATTTGTCCT GTCGAGCTGTTTAATCTATT CATATGTATGTTGTGTGT GTCGTCCCTCGTAGCAGTTT TATTCCGTCCAACCAGAGGT CGACCAATATTAAAATAAG CATGGCTTGAAGCAGGCCA AGCGCCGTGTTCTAACCCC GTTTTAGGGGAAGTTACTT AACCTAAAAATACAACTTTT CC (SEQ ID NO: 1310)	SGXTGREVKCITVNVLMEQQPHTKAIREGDFIVILLPQSDDETLCCPLCVG RGRYSGKTRVECLNRHVKEVHPDLTTTFRCWGCGFAAPGDKKYPRKIVT QHCATCVPEVSSAPSGRVDGERRVNTRRRLGIAAATEASPVRRTRRNGL ASPPVEQNISQSAAPPEPARVPQHPEIVALGESADDEVFRSPVNSPPRD WRAAAPQQAASSSPXAVPGITAATPSNTTRTGNGSAXSILAEHPIPAPPP TNTTEANGRADIPRSGRAPPPGXQAARRRAPTTEQRRIVGLLEAATGRE QLEEATTQAMLFLARLTGRRPEPRNAIRPGXRQRHPAQGDVQAQAPDR IXEAKKLQRLYRTSKKRAVQKILAGPXMNCQIDKNTITAHFVELAARRDG GEDWPDVFDREEPTAASGEALCTPITREEVFRRLKGRNNTSPGPDGITYR DLAKAXPGAHVLAALYNXIWRIEATPALWGVSNTTLIYKKGDAMDISNW RPISLGDTVPKLFAAILADRIKRWAVANGRYSASQKGFLEFEGCYEHNFVL QEAIREAKGGRKELVVAWLDLASAFTSVPHSSILQALEGHGLPSKARNIIS SLYTGMTTRFHTAEGPTDPILIQSGVRQGCPLSPDVFNLTLEVVLREIQRT GEGYTIEGRRISHLAYADDVAILADSPAGMRRLLFAAERGARAVGLTFNP AKCATLHIAGRGEEAVRPTEFSVQGTPVRALASGEAYEHLGIPTGYQVRQ TPINTLRDLLADIGSIDRSLLAXWQKLDAVGTFLLPRLDFTMQGAHIDKGF LTEADKIIKKAAKSWLSLPQRASAELVFLPPSQGGGGLLTVAHSYKMLYSS DVTVSTIAGSTLRRTVSERLKKRASNIDIARFLSGDLDLPRSTSPSTFWTKV RSAALRIKTKLGLRWSWCQGGEVLLMACGDPRAPGTRVSPQTKHLVTT SLRRCLNRHYAESLLAKKDQGKVFEVTRRSGQSNHFLRSGSFTRFCDWRF IHRARLDVLPLNAAKRWQRGMDKRCRRCGSDLETLPHVLSHCGPHSAA RQKRHNNIQDRLVKAASRCPGTISVNQTVVGVRGPDAALRPDIVVRDD VNRRVTIVDVAVPFENRLEAFDGVREAKIAKYTPLARQLTDSGYTVTVEA FVVGALGAWDPRNERVLSLLSISRYYAILMRRLMVSDTIRWSRDIYVEHV SGIRQYRE (SEQ ID NO: 1432)
NeSL	Utopia-1_PCa	.	Phytophthora capsici	GCCCGCCGGTGGA GTAGCCATGTTGGC CACCACCGCCCAAG TCTCCGCCGCAGCT GCGACTGCTGCTGC TCATCGAGTCGCAG TAGTCGCAGCTGCA CAAGCACCACCTCC TGCCGGACGCGCCG CCGTGGAGCACCAC	TAGACGGCACAGTTCTGGC CCACGTAGGCCGAAAGGGC CCCACCCATGTAGGGAACC GCCCTCGGGAAATCCATTCC GGTGTTCGACTGAAGAGAT GCTCCTTCGCCTTGACGGA GGTACATCTCGACAGTCGA ACTTCAACTCGCAACATATC CGATACAGTTACAAACCAC AGTTAGATATCAGATAGGA	MVVSRITARLEATPAPRWDPPLPRRVIASRIADRLVPATAPCRSALNAAF PSPSRDTVTESFTQEDRQLEPLTRHVDEETKDSELPGRAPTVLDEESKDN DATAGEWLLRFDGACRANPGPGGAGAVLFNPSGAASWTCSHFMPGAT ETNNTAEYTAFLLGARAAADHGATLLRVQGDSQLVLRQVKGIYGAKSTR LRRLRDAVRAELARVGQFSLHHIDRQDNAHADRLANRALDMKSTLVEC ATHPGRNACTTTLTTSAAAESPASPPPVGARDTPMADAGEERLADVDD GEVYAAMRLGPGEVPERRPRLRLRQLSDEELEAASEMVERLGAALSAKIT DAEDWASAEGYITALPYMLYDKLQSYSQAPRGPQQPVLTRSPRGDDRP ASSEPNASSTTGGVASEHQPRRRRRRRGRRKGRRQRRRNPRRSGREGATG GHQQHKKHKPRPPRETQHHREHRLDEALDELHALERTDPHNRPAIAKA

		GCGCGCGCTGAGCC	ACCTTCCTTTAGGAAGCTAA	RRRVGRIRSAINQQLLRHKFDTDEKACVDGILSTARAERAARAATPSPPA
		GTACCAAGACCAAG	CGGGTACACTGGATGGTAA	SGAPTTTVSAPGAIVTNDDGTCPIPSDKLWRHFDAVNTPRLDFDAEAPG
		GNTTCCAGGCTCGC	ATACACATACATTTC (SEQ	SAAFRAAMDHLPAATRLLDLLKEAPSTDEIETQLQHVKASSSPGLDGVGY
		GCGCGCGTGGMGC	ID NO: 1311)	DVYKRFTIQLLPVLRAAFRCCWLYKKVPQSWKLGVVRLLHKKGPREDPA
		TCCCAGCCAGCGAC		NWRPICLQQAIYKLYTGILARRLTRWMDANDRHAPGQKGFRAVNGCGE
		GCGTTCAGCAGCAG		HNFLAATLIDNARRKHRPLYEVWYDFRNAFGSVPFALLWDSLQRLGVPP
		GGTCGGCTGCGGC		DYVDMCKGLYNQASFVVGNAVDGSTAPVEQRVGVFQGCPLSPQLFNA
		AGCTCGACGCTGCA		AISPLLYALRRLPDTGVQLSSVDRPGASAYADDLKIFSGTKAGITQQHELV
		CGGGGACGGCAGC		ATFLRWTGMQANPAKCRSMGVRRNTNGAVEADNVHLELDDTPIPSMT
		GACGGCAGGCACC		HMQSYTYLGIGDGFDHVRRRVELAPKLKTLKHDVTALVESGLAPWQVVK
		AAGACCACGACGAC		AVKVYLYPRVEYALRHLRPDDQHLESFDLHLRRGLRHLLRLPKNATNEFFY
		CAAGCCGTCGCTGC		APVSRGGLGLLPLVELHAALQIAHGWQTLHSPDPAIRRVAREQLYQIADA
		CGTCCCCATGGACG		RHRLDKDHWPHRREELCELLLNGELGTSAHAPPKRRNGDIGSLWVDVR
		TCGACCAAGGTGCT		KNLKTFGLKVATAPANQETGVPAQPLQLRVPHHAEWLDHGNVLRHVKL
		CGGTGGCCGACAG		HIKNLHWQTWCALSDQGKTARVHGGVGSAFLTRPRGMWESDYRFAV
		CGGATCAMCACCC		AARLNVVDTVNTLSRRRLRAHDRCRYPACRWKETLAHVLNHCPGTMDA
		GCTGTCGCCGCCAC		VRGRHDDALKEIEHTLRASSGDRRELRVNQTVPGLPGPPLRPDIQVYNH
		CGGAGCTCCGAGTC		DKRTVAVVDLAVAFDEQPSEDPESSGLAKAVQIKKAKYAGIKEHLENQG
		GGCGGCAAGCGCC		WKVHLSAIVYGSLGSVAASNHKVYTEHLGLLKRDAKRLDRQLSSACIQSS
		GCCGCCTGAACGAC		RRIWNFHCAKHRARQHEHQAPPSQATRGRRVTETGGNPSRTDRR
		GGCGACGACGAAG		(SEQ ID NO: 1433)
		ACATCCGCGAGCTG		
		GCCGAGCTTCTGCT		
		GCCCGACGAGGAG		
		GAGGCCGACGACC		
		ACAAACCAGCGCCC		
		AGGTTACCCGCGAC		
		CAGCGCTCATCCGG		
		CCTCCGTCCTCGCT		
		GTGTACGCGCACAA		
		CGCGCAGCGCTTCA		
		ACTGCACGTTGTGC		
		GTGTACACGGCTGC		
		CAGCTTCGCTGCTC		
		TTACGCGACACAGG		
		GACTCTCGGCACCG		
		GCGCGTGACCTTCC		
		TGGACAGGTTCTCG		

NeSL	Utopia-1_PI	.	Phytophthora infestans	AGCTCGGCCTCGCG GCTGCCTTCCCAGG CGCCGCCGACTTCG CGCTCTGGCGCGGC CCACACGCCGCCGC	TAAGCTGGTCATCATGACC GACAGGGCACTACCCAGGT AGGGAACCGCCCTTAAAAA ACCCAGGAAGACACAAACA CCCTCCACTTAGTGACATAC	MLADPAALAAGLARAPPPPSAPQDPSPAFPAGPAGQNPRAAAPARVEV HTVVAPPGRAGGMLPDPGLVEEPIQATYAHDAAQFECALCPYVAESMA VLVQHRRSAHRGTRFKDIFTSGCQCSLVFYARIVAASHAVACARRNQRA VPPAPTPVAPTRPEATPQPTGYLAAAMTAAAAAASSDTVVAAATNMQS AVPAAAKTTGLQLVPPELEPALPQRASCHAGKRRRLNADEAVTPCTPTA

CGAGCCTCCAAGCG		ATATTTTAGCCTAGATTTCA	RVSPQTEVAMAPHDAPQDDTVLQREAAEPQPDPAATPGAQVQRVEDT	
CGCCCGTTGGCTTT		GTTACGGAGAGGTTACTAA	TAAQDDTVQQDHDADTAQVSPPRRTPTRWGPRPSSTQEPSPMTGEPA	
CGCAGACGCAGGG		CTGGTAAATACGAACACAT	ATLAARRPLTPAATGTRATRWGPCHRAIGAAAIARLVTGLSTEPAQPQR	
CTCGCGGCGACGGC		ATTCTGTTCTAATCAGTGTG	RQPPPPQEPPSQPEPLAAAATAAADIAATVAADIAAAAAANAAMDVDG	
CCTCGCCAGCCCCC		AAAACTGGTTTTCGCCTTTT	GPAADETWLLRFDGACRRNPGPGGAGAALFAPSGAVVWTCSHYMPSR	
AAGACCCCCCCTAC		GGCGGACTTTTTCACTCGCA	SETNNTAEYTALLLGVQSAVHHGASHLEVEGDSSLVIAQVKGTFACRNA	
GATGTGGCACCACC		TTTTTGGGCAATCGTCTGCG	RLRQLRNRVRHALRSVDTHKLRHIDRQANAHADRLANRALDQRRTSSEC	
CGGCAGGGCGGCC		GCTAGCTTGCTAGCGGCGG	GTHGSCMDSCLAVPTALAAQETPPAAPPSTSATPAEGNAMDDIAAEIAA	
GGCAGGCTGCCCG		ACGAGCGGTCTCCGGGGGC	RDEGETFPVLPIGPGSAPERQPRLRLRQLSDEERDAAADALQELADTMA	
ACTCGGTATCGCCG		GTTCACCTTTCCCCCGCGAG	SKIEDADSWTSGEGYISSIPERIREVLQPYATAPPQPGRSRRQQRRRPPRV	
GGTGCTACACTCTC		GCCAACTACACCGATCTTCT	TRNQREHRLDEALDDMAATQQATPRDQRAVRRAARRRVGRVRASMAQ	
AGCCGCCACAGCTC		CTACACTTTTCTAATTCGCCT	QELRHEFAKDESKCVAKILKTASTETAAEDEHPETCPIDAATLHAHFTGVN	
GGGCCTTGGCGGTC		CCGTCTTCGGTCTTCGGTTG	APRTDFDYDATSGREFRAAMSDLPPATVEIDAFDAELTIDEVEDQLTRAA	
CGCCATTGGCCCTT		TCGGGCTTTTTTCTTTTTGAC	KTSSPGHDGIDYGIYSRFAAQLVPLLHAVFQFCWRHRRVPRLWKVGIVRL	
GGAGCTCGACAGC		CAATCAGAGCGCGCCATGC	IHKKGDPRQPTNWRPICLQPTIYKLYSGLLAHRLSRWLEGNDRLPMAQK	
GACAGCAGCGACG		GCCTCTTCTGGCCAATCAGA	GFRAFNGCHEHNFMATTLLDQTRRQHRKLYQVWYDLRNAFGSLPQQL	
ACGAGGACGCTCAA		GACCGGGCCCTGTCCTCGG	MWRVLRHLGVDSGFIDRCRDIYRDSAFVVANAADGATDPVRQEVGVY	
GACCCCCACGCCGC		ACAGCGAGGCCTCCACGGC	QGCPLSPLLFVAALVPLVRRLEKLDGVGVPLADGVRPCTTAYADDLKVFS	
CGCCCCAGAACCCC		CAGCCAATCGAGTCTCGGC	DSAAGIRKCHDTVAGFLAWTGLRANPGKCASLAVTTNARGNPTRDSSM	
CAGAAGACGTCGC		AGCGACGCGTCTTTCTATAG	RLEVHDAAITTLSLHESYRYLGVGDGYDHVRHRLQLEPKLKQLKREAVALL	
GAGTGTGCTTGCCC		CGCAGCTGACGAGGCCGAT	TSGLAPWQVVRALKVVYPKVEYALRHLRPLQSQLQAFDRVVSKGLRHL	
CACCCGGCAGGGC		CTGGCGGCCCCCGATTGGT	LSLPRSATSEVLYAPTSSGGLGLQPLVELHRALQLAHAWQMLHSKDPAIQ	
AGGCAGC (SEQ ID		CCGACTTTCGGCCAATCAGC	AVARAQACQVVRKRYRLQEDHWRGRDDELVRSFLNSELAASPHAEVLR	
NO: 1189)		GACGACGAGGGGGCAGGG	RNGDIASLWSDVQRWLRIYHLRFEHCDETEAHGPLSFRVPHHNKWLTH	
		GTTTACACTTTTGCCCCCGT	KTVLRHVKLHLKIRHQTRWKGMVDQGKTVRVHGGVGAKFMTTGAGLS	
		TTCGACTTCAACTTCAGGCC	DDDYRFGVKGRLNQVDTNSVLKRKRLRAHTTCRDPTCSSAETLAHVLNH	
		AAAATGGCGATTTGGACCC	CESNMDAIRQRHDDALEQIGSKIRGALDRAKSPTELRLNQTVPEYTGAAL	
		TCCACGCGCCGTGCCACTG	RPDIVLRNVAAKTMVIADLAVTFEDQAARARHSSLQLSHDHKTLKYQPIV	
		CTCGGCACCGGCGGCGATT	AELQHKGWRVQTAAIVYGTLGSVQPSNFKAYTEKFKLHKREARQLDLQL	
		CAGCGGGTGCAACTTCGGG	SSHCIRASHRIWGWHCRQHRDRQRSGTASRASRGSGGAPRRTSQAPAR	
		CACGTGTGCAACACATGCA	R (SEQ ID NO: 1434)	
		GCGCCCATTGCACGCCAAG		
		CGGCATCGCGGGACGACGC		
		CTCGGCCGCCCAAGCGCAG		
		CCCCGCCCTTCCAGCACGAC		
		CTCGCGCCGTTTGGCGGAT		
		CGCCATCAAGACGTGCGAG		
		AGCCAGGCGGGGTCGGGC		

NeSL	Utopia_1_PMi	.	Patiria miniata	CTGATGTGGATACC TTGGAATTACTCAA CCGTGTCGGAGTCT TTTGTCTTTTGCGCC CAACACCTCATGGA TACCATGCTTGTCG CTGGAGCGACGTTA CAAGCGTGAGGGC GCCCTCCATGCCGG ACAGCTGGTCTGTG CC (SEQ ID NO: 1190)	AAAATATACTTACTCTAAGT ATGCCCGAATCCCTGCCCTC TCAGGCTGAACGCGGCCCC ATACTTGATCTAAGTATGG GAGGATCCCTGGCCTCTCA GGCTGTACGCGAGACCC (SEQ ID NO: 1312)	MCLKSFSSTSGLRRHMARLHRQPSPDASTPSTMTEVFPYPLPKVWPCVV
					GATTAGCGAACACTAATATC CTGCCMTAGACGTGATTGC TAATCCGCAAACCAACCGG ATCTACAGCCTGAACACTGA ACTTTAATCTTCACCCATGT CACATCTGGACACTAGGTTT TTGCTCTGTTTGTGTTTCCT GCCTTTWCMTTGGAWCTT TCGCMCTGGAATTTATTTG TCGCTTGGATTATTTTTTTTC TCACAATTTGGATCTATTTT CGTTTGTTCACTTCGAACTC TAGCTGCCCTTTCTTCGGAC ACTGAACTTTAATCTTCGCC ATGGCTGTCAGTCGCCGGT TCACTTGCTGCGGTGGGAT CCTGTTGTGATAATCCCCGT GCATTGCCCATGGATTTATT TCCGCCTTAGTTGTTCCTAA CCTTGGATTTATTGCTGTGG GTGATGCCCGGGTTTTGTTT ACATCGGGATCCCGCTGCG GCTCGGTGTTCCATGCGACT GGCAGCCCCTTTGTTTACTC TSGACTCTATTCATTGTTGT ATTTCTTCCACACTGGCCAG TGATCACACTGCTGTGTTTC CCGGGAAGATATCCTCTGC GGTTTTCACGCTCTGGGTGT CTCCCCGGGCAACGCACTG GTTGCTTGCTGCGCCATCAC	MRKLLRIDAIKGSRDVYIEHMCGHRQYT (SEQ ID NO: 1435)

NeSL	Utopia-1_PS	.	Phytophthora sojae	AGACGAGCAACGC GCTGGGGCCCAAG ACCTGGCACGAACG ACACTGCCCAGGCT	TGAAGCTGCACAAGCGCGA GGCTCGACAGCTGGACCTT CAGCTGTCGAGCCACTGCA TCCGCGCCAGCCACCGCAC	CCTTTTCGTTTATATTCATTT TCAGTCTGCCGTTATCTTGG CCAGCGCTCATTCTTTTGTG ATGGCCGTGGACTGACCCT CTGCGGTTTTCTGCGGTCAS CACTCTCGGTGAATGCTGT GCCACCATTTTTCATTTGTTT ACTTTTTCAGCTACAATTAT CCTGGCCAGCTTTCACTCTT TTGTGATGGCCGTGGACGG ACCCTCTGCTGGTTTTCACG CTCCGGGTTTGTCTGCGGTC AGCACCCTTGGTGAGTGCT GGTCGTTTGTTTGCTCATTT TGCTTAGTTCACCATTATCT TTMTTCTCTTTTGTWTGGTT TTCCTAGCGGTTGTCTGGG AGTTGAGCTGCAGTTGTCT GGTCTTGGTTTTACCCCCAT TTGTTTTCTTTTAACGCGGG GCGTATTGCCTTGACCGGC CGTCTCAGCTTTTCTCCTAG AGCAACCTTCCGTTCATCCA ACTTTTAGTTTTCTCAGTTCT TGGCCATTCCGGTTGGTTAA TTTTTATTTATACTTAATTTT ATGTTTACATTTTCTGGTTG GAGACCATTTTAGCTTGTTT TAATAGCTTTTCTTCTTTAAT TAATACCCTCTGCCATTGAG GGTTTTTATTACTATTAATTT TGTTTACTCTTTGTAACTTGT TTGATTGAATATTTTAATAA ACCAC (SEQ ID NO: 1313)	MDVDGGPAMPEPWVLRFDGACRRNPGPGGAGAALFKPCGTVAWTCS HYMPNSSETNNTAEYTALLLGVQSAVHHGASHLEIEGDSHLVVAQVKGT FACRNPRLRQLRNRVRHALRAVTSLTLKHIDRKANAHADRLANRALDLK RSLAECGEHQGAMESCLHMNPAAQRQREQPAPPARPACAPTRAESAS

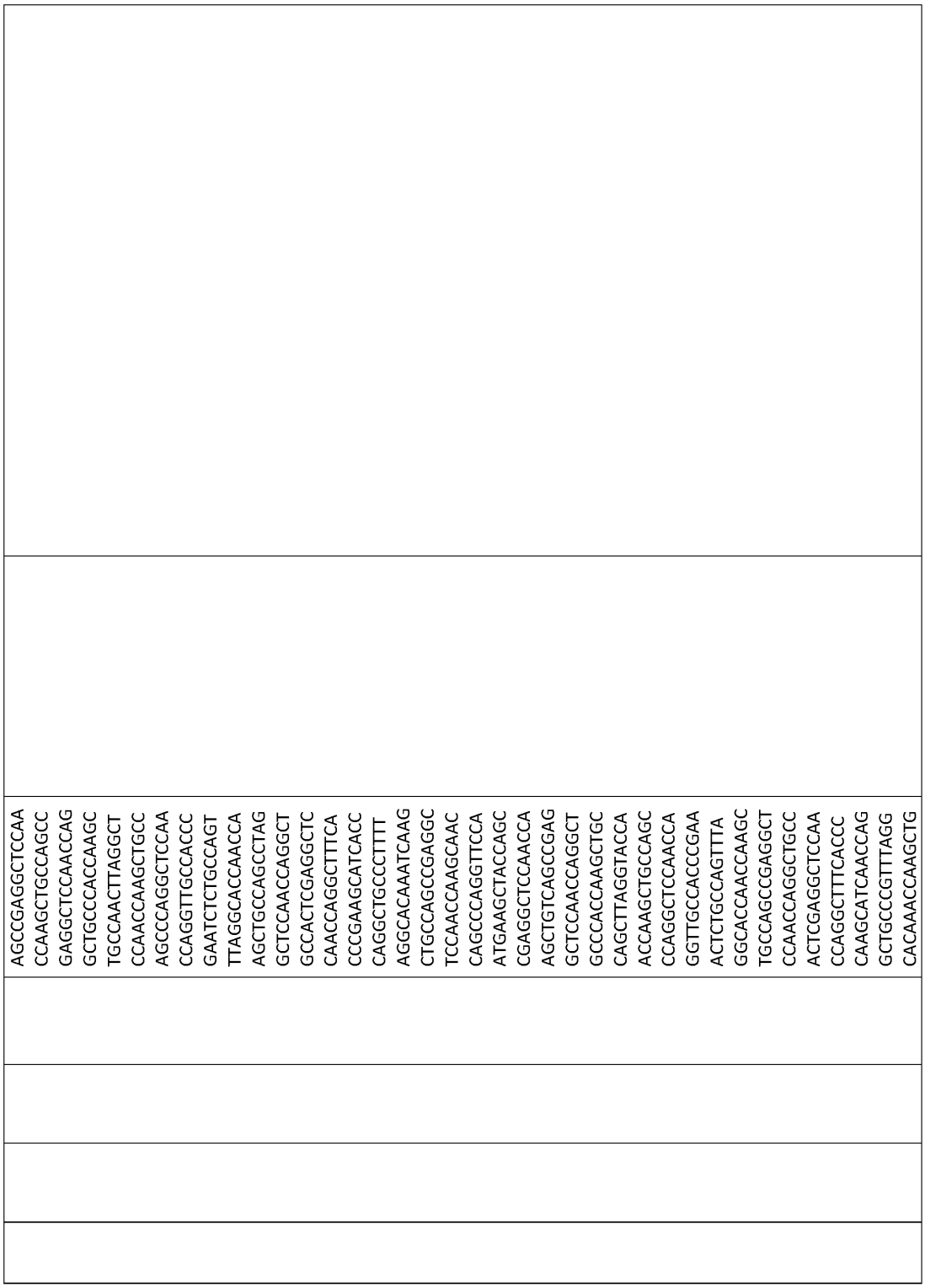
			GTGAACGACGAGC ACGCTGCTAACCCG GCAGCGCACCGGCC TCTGGGCTCCGCTG CACCGGTTACCGGT GCAACACGGTGGG GTCCACGTCACGGC GCGATCGGGGCAG CGGCTGTAGCTCGC CTGCTCACAGGCCT ACCCACGGCACCAG CACCAGCTACGCGC CGGCCTGCTTCGGC TCGGCGCTGCCCAG ACCCGCCCGCTCCC CCCGCGGCCACGAC GACAGCCCCCGATG CG (SEQ ID NO: 1191)	CTGGGGCTGGTACTGCCGG CGCCACCGCGAAGGACAAC GGAGCGGCAACGCCTCGCG AGCGCCGCGCGTGGGTCTGGG GGGGGCCCGCGGCGCACAT CGCAGGCTCGGGCACGGCG GTAAGCTGGTCATTTGACC AACAGGGCACTACCCAGGT AGGGAACCGCCCTTCAAAA ACCCAGGAAGACACAAACA CCCTCCCTTTAGTGACATGC ATATTTTAGCCTACATTTCA GTTACGGAGAGGTTACTAA CTGGTAAACACGAACACAC AT (SEQ ID NO: 1314)	DHDEDIDAEIAARDGGEAFPTLPIGPGTAPARQPRLRLRQLTEDEQEAAA SALQAMAEELACKIEDADSWTSGDGYISAIPSRIRQLLQPFTAAQPHPRP PLQQQRQRPPRVTRTQREHRLDEALDEMAAVQQERPTSRSAVRRARRR VGRIRASMRQQQLRHDFARNESKCVEDILRAASAETAAEEHPETCPIDSG TLHEHFTAVNSPRINFLPDEACGALFREAMADVGTPQERRSALTDELTM DEVEDQLMQAATNSSPGHDGVGYDIYKKFAAQLVPLLHAAFQSCWRH HRVPALWKVGFVRLIHKKGDPNDPANWRPICLQTAIYKLYSGLLARRLSA YLEANGLLLMAQKGFRAYNGCHEHNFVATTLLDQTRRMRRRLYQVWY DLRNAFGSVHQDMLWYVLRLLGVERAFVERCDDIYEDSYFVVGNAADG ATEPVRQEVGVYQGCPLSPLLFIAALVPLLRALEKLDGVGVALADGVRPC TTAYADDLKVFSDSAAGITRCHAVVEKFLEWTVLQANPGKCAFLAVTRN ARGNPAHDKDMKLSLHDEEVSSIKLHDSYRYLGVGDGFDHVRHRLQLEP KLQQIKREAVALMQSGLAPWQVVKALKTYVYPKVEYALRHLRPLQSQLQ GFDRVVAKGLRHLLRLPRSATNEVLYAPTSSGGLGLQPLVEMHRALQIAH AWQMLHSKDPAIREVARAQVWQVARKRHRLREEHWRERDDELVRCFL NSELAASPHAEALRRHGDIGSLWSDVQRWLRIYHLSLVVQDDRNGLDPL GLRVPHHAKWLDHKSVLRHVKLHLKIRHQTRWKGLADQGKTVRAHGG VGAKFMSTWAGLSDDDYRFGVKARLNQIDTNAVLKRKRLRSHKTCRDP TCSSAETLAHVLNHCESNMDAIRQRHDDALEQIGSKIRNALKRGKSTAEL RLNQTVPEYTGAALRPDIVLRIVAAKKMVIADLAVTFEEHAAGARHSSLQ LSHDHKTLKYQPIVAELQLKGWQVQTAAIVYGSLGSVQPSNSTPTRKS (SEQ ID NO: 1436)
NeSL	Utopia-1_PU	ADOS010013 21	Pythium ultimum	GCGGTGTACGCGCA CAACGCCGCGCTCT TCGAGTGCACGTTG TGCGCGCACACCGC GCGGGATCTCGCCG CGCTCCAGCAGCAT CGGCGCTCCGCGCA CCGCAGCGTCCGCT TTGTGGATCACTTC CACAGCGGATGCGC GTGCGGCGTGAGC TTCCACTCGCGTGC GGCGGCAACCAAG CACGCGCGCGAATG TCCAGAGAGCGCGT TCTCGGTCGCCGCC	TGATGCGGGTCATATTGAC CGAAAGGCACCATCCACG TAGGACACCGCCCTCAAAA ACCCAGTTCAGTTTATTGAC ACCCTCCACTTAGTGACATG CATATTCAAACCGATACATA TTCGTTAGGAGAGGTTACT AACTGGTAATATATCACCAT TTC (SEQ ID NO: 1315)	MGTQSARERGAPSAPHSHTLGPRTPPRPPACSKHGELESAAGGRDGQC SDGAERERDAERDIRANERDCNGDGDGDDADSDSDDRNDARRRSRRP RATATTTTSAPTTTTTTTTSATTSATTPATDSSPWVLRFDGACRRNPGPG GAGAALFEPGGAVVWTVSHYLPGSETNNTAEYSAMLLGVRSAIHHGAT RLRVEGDSHLALSQVRGTFACTNRRLRKLRNRVQAALRELGDYRLVHIDR QANAHADRLANRALDLRKTKVDCGPHATTTDACVQPAEILAPTARLSSS SSSSSSSSSDEPMPGLEEPAADDETDADAEADIAMRDGGEIFPTLQIGPG SAPAQQPRLRLRQLSDDESEAAARTLEHFANDMASKIADADDWRSGEG YISAIPVRLRELLAPYAVPIRSPPRNASSRPPRPRQSRPPPRPPRVTRHQREHR LDEALDDLAAAQRSTSTDQRSIRNARRRVGRIRTAQAQSDLRSQFATNE RACVESILRAAKPDGTEPQASAGTCPIDRATLHAHFAGVNTPRERFDFDD ALGADFRAALDVLPPPDQAADAFADELSLGEVEDQLDRVVASSSPGLDG VGYDVFKRFRLQLLPLLHAAYQCCWRHRRVPATWKVGLVRLLHKKGDR AEPNNWRPICLQQAIYKVYSGLLARRLSRWLEANERFTTAQKGFREFNG CHEHNFVASSLLDQTRRLHRKLYAVWYDLRNAFGSMPQPLMWRVLAR LGVDTAFLQRCEDIYADSFFVVGNAADGATDPVRQEVGVYQGCPLSPLL

						FISALIPLLRALQRLPGVGVPLADGVRPCTTAYADDLKVFSDSAAGIQQCH GTVARFLRWTGLRANASKCALLPVTTTARGNPAIDDTLQLELHGDAIARL TLQSSYAYLGVGDGFDHVQHRVQLAPKLAELKRDAVALLRSGLAPWQVL KAIKVYLYPRIEYALRHLRPLQSQLEGFDRAVAKGFRHLLRLPANATNELLY APVSSGGLGLLPLVELHKALQIAHGWQMLHSKDAAVQAIARAQVRQVV QKRYTLDADHWQGRDDELVQLFLNSELAASPHATIKRRNGDIGSLWSD VQRHLKTLQLRLETREPTADAPDSPNGLLHLRVPHHRKWLSHKTVLRHM KLHIRLCHKHKWQSMSDQGRTVRAHGQAGSHFVSRGVGLWDADYRF ALQARLNQLDTNSTLKRRRQRTNATCRAPNCSRTETLAHVLNHCETNM DVIRQRHDGALEQIGAAINAAIKGRRTDTEVRLNQTVPEFNGPAWRPDI QVRDARSKTMVIADLAITFEDQPNDQSASSSLQHSREHKIAKYQPIAAAL ERQGWRVHTSAIVYGSLRSVHPSNFTVYTELLGLLKRDARRLNTTLSCHCI RSSRRVWNWHCGQHRARQHQRCQEGRAHGSGGNQRAEGGTATT (SEQ ID NO: 1437)
					TGATGGGAGAGTGAGGAA TCTTCTCCACTGTGCAAAAC	MATVRLKYPYPPEGILCGPCAANTNAPQTRPYSDKSGLAKHLKLYHKATL VVECRHCGHEESDLRKMKKHISTNHPVAAAAAPTVPPRLGPTAPPPRV
NeSL	Utopia-1_SM	.	Strigamia	GCCGCGCGCACTGC AGCGGCCAACACCG CAGGTATGTCTCTC GGCGCCGACGAAC GCGACCACCTCGCG TCCGTCGGCGCCTT GCATGATGTTGCAT CCCCCGCTTTTGGC AACATTTTGCCGGT TGCGTTTGCGACCG CGGCAGACGCATCA AGCGCCACCGTGAT CGCAGACGCAGCCA TGCAGCACAGTGCT GTGCCCTCTGCTGC CGCCCAATCCCCTC GGCGTGCGCACGTC CCCCCCGTGCCGCG CGCCACCACGACAC CATCCGCGCTGCGG ATTGGTGGCAAACG CCGCCGCCTGAACG ACGACGGCGACAA CGAAAACAGCGAC GGCCGCGACGCCG ACATCGAGATGCGC GCCGACGACACCGA CGCACCAGCGCCGA CCAACCCCGCGACC AGTGCGGCTGCAAC GCCCGCGCGCACAG CACCAGCACCAACG GATGCCGCGACCGC GCGCCGCCGCCACA CGCG (SEQ ID NO: 1192)	GGAGTGTTCTTTTC GGAGACGCCGCCTA	

		maritima	CTTTAGAGGAGAGA ATCCCCACGGGCAT CCTCATTTGATCTGA TCCATCGAGTATCT GCGAATAGTCGGC GCACTCCTTTTGCCA TGATCCCGGGGGTC TCATGGTAAAAAGG TTTGTGGCACGGCT TAGTTGACGCCCCT CTTCCACGTCACTC GGCCTGCATCGATC GACTGCTCTCTCCTC TTCCTCCCTCCCTCC TCTTCACGCTCTCTT CACGTGACCCCTTC CCATCCCGCCCCTC GGCTTTTGGCAAAG ATCTGTGTGTCCTC CAAAGCACCCATCT ACCATTTGCTCGAG TTGCGATTGGTCGA AGCTGCCACGCCAC TCGTCTACTCTGCCT CTCTGACCCCTTCCC CCCTCTCTCCCTCTG CCGCCCTCGCCTCT CGGGTCCCTTCCCA TCTCCCTCCTCCCAT CCACGTGCTCCCTC CTCTCGTCACGTGA TCGTTGCGGCTCGA C (SEQ ID NO: 1193)	CATACAGTCAGAAGATGCT AACTACTAGTTTGATACCCT GTGCCCCCTGCAATGTCCCG CGTGTCGTACCCAAGCCCG GCTGGCATTGAGACACATT AGGCTCTCGCTCCCCCGTAT ACTCTCATAATTTCGTGTAC GCTAATCCTACCCTACCCCT CCCTTTGACCACTCACCCAA CCATGTGTATAGCTGTGCTG GTGATCCCGGGGCGGTTAT TCACTGGTTATCATATCATT CTAAAATGATCTTTGATCTC TCAATTAACAACTAACTTAT TTCTGTTTGTTTCATTGTTTC ACCTCGTAAGAGGAAGTTC ATTGTGCGATAAATCAA (SEQ ID NO: 1316)	ILRPRFIPRPRTPSPSSSSSSASSPASSRRSVSLPPASPPVSSASSPAARSGR NSPDSQGTAPVTPIGTVRNSPAGSPALSYSTASPIASTITTPRHLSPASPAL SAGPGSLGASPPVSPTAATVPPAPPATVPAVMAATVPFVAATTVPSVGS STVPQRPAGPRRPPPFPIDDWIGRIARVSSLPELDAVSRLLEDEVVKRRPP DPNARPASLHPTRRPPPPSTRPRPCHGTGGTSVSLAALTSSCIREDHRGSL PLLCTSGWDSPWPLSPSPSSHCPSSNPCPSSSTPPSSLLGPPRHSHLTWR GSGSTTPSHRHCSRARYHHLLWRLPYPSSPRLPYPPSALARYPSVPPALD DPLPSLSTIGSSEGSPECHLCRNWTPSQGCSRMKWSRDAPLTPTPDPPHSI RHAALPLHLLALVPAMGPAELQSLWRRSRPRAFAKITEGASPFCALPVGT VHGHFLQVHQATAHLPTPVPLPPLRPPPRSSDPLVTPISPGEVLDRLRRAT DTAPGPDTIIYSEWRAIDPTGRLLSSLFQKVQTFGAPTRWKESTTTLIHKG GDHTAMSSWRPIALLSTVAKIYGSILSHRLTTWAVQNGRLSLSQKGFLPF RGCLDQNYLVQSCLQDARRNKKTLSLAFLDLKNAFGSIPHLTIRHSLEWL GLAPSSIDILEASFLGSSTRVRTETGLTPPISLDTGVVQGAPLSPILFNLAIEP LLRTVPSAHSGFSLHGHVVSVVAYADDLAILAPSTPALQSQLDAISGMAD WAGLSFNPAKCATVTLTGKDNSRDTLSLQGSPVPSISDGDAYKHLGVPT GTTTFPSGTDAIKKMTTDLQAIDHSDLAPAQKLDALRTFIMPQLSFHLSH GSVPKAPLTQLDKKIKRAAKHWLFLPQRASNEILYMSHLHGGQSLLPLSV LADIGQVTHAVALLQSRDPAVADLALRTCREVASKRAKKTVNGPELAQYL SGSTDGIYCTPTSDIPSLWTTARAATRRLSSTLPLTWTSPLPSGVPFLSING SPLSPFRVQSTLTNAIRHNHLSTLIAKRDQGNSYRTSHDPDPSNYWVKGG DFLRFCDWRFIHRARLNLLPVNGARRWDANSIKTCRRCGAPNETLAHVL NVCPVGLPEMKKRHDAIHARIKKALRPSPHTVVHHDRTVPGCGPLRPDIL RISERDKSVAIVDIHVPFDNGTDAVERAHETKRAKYELIRRHYEHQGYRVT FDSLVVTALGRLWRGSEAALQALQISSQYSKLLRKLLVADAIHGSRNVYA HHMTGMVM (SEQ ID NO: 1438)
			AAGGCTCAAACCAG GCTGCCAACCAAGC TGCCAGCTTAGGCA CCAACCAAGCTGCC	TGACGAAATGTTCAATATAT GACATTCTAATGTTCATGTA TTTTTTGTTTGCTTGACAAA GCTAGATGAATATATTCCCT	MSHSITEVFDYPLPSRWKCTVCLENFFNQQTLKRHQARHHQTTSFLYVF RCSACQAEFDSARKASNHWQSHKRKPILSQPAVNEIPSSGLDPSPPRSRP PVEVIGSSFPDDVSMLSEPSTPSTSLQMDPEVVHPPSRSISFSPMHLSPTQ PASPIQIGVEVSFNSSSSLQMDPEVVQPPSPSISFSPMHLSPTQPASPIQI
NeSL	Utopia-1_SP	AAGJ02140537	Strongylocentrotus		

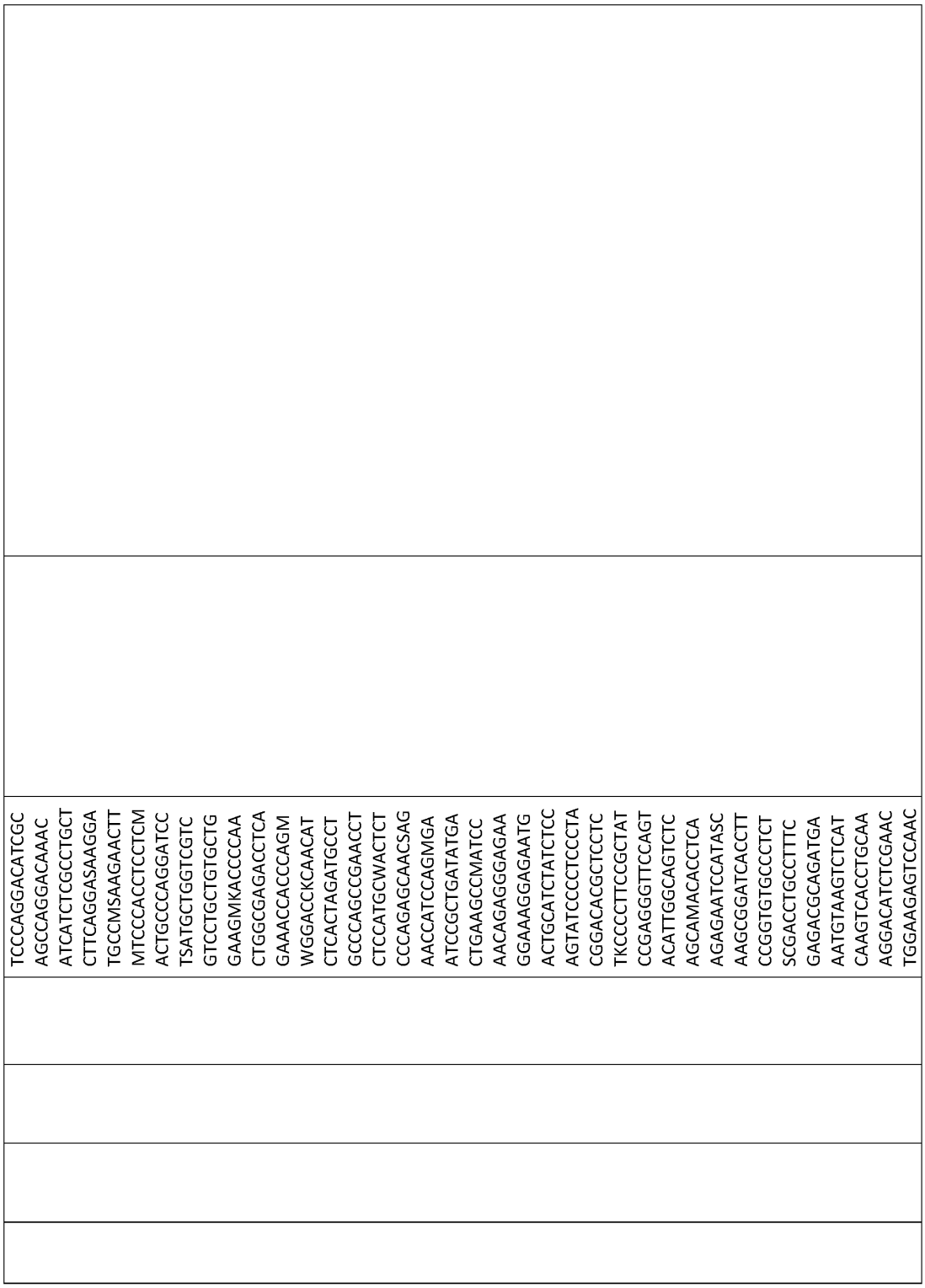
		purpura tus	AGTCCAGGCTCCAA	GTCCTACTACATACATCTCG
			CCAGGCTGCCCACC	ATGTGCCCTTGCTAGACGAT
			AATCTGCCAGCTTA	CCATTGGTAACCATGATATC
			GGCTCCAACCAAGC	AATGGATTAACACATGATCT
			TGCCCACCTAGCTG	GTAAACAATATATGATTTAC
			CCAGCTTAGGCACC	ACAATGTATTTATTGTTTCA
			AACCAAGCTGCCAG	ATAAATCTGTTCTTTTCACTT
			CCGAGGATCCAACC	TAATACATGAAACATGACT
			AGGCTGCCCACCTA	GTGCCTTCTCCAACTGGAG
			GCTGCCAGCTTAGG	ACCTACATAATTTGTTAAAA
			CTCCAACCAAGCCG	TGATATAAATATTTCGAAGA
			CCAGCCCAGGCTCC	TGAAATTATTATTAATAA
			AACCAGGTTGCCAC	(SEQ ID NO: 1317)
			CCGAAAGTCTGCCA	
			GTTTAGGCACCAAC	
			CAAGCTGCCAGCCG	
			AGGCTCCAACCAGG	
			CTGCCACTCGAGGC	
			TCCAACCAGGCTTT	
			CACCCGAAGCATTA	
			CCCAGGCTGCCCGT	
			TTAGGCACAAACCA	
			GGCTGCCAGCCGA	
			GGCTCCAACCAAGC	
			AACCAGCCCAGGTT	
			CCAATCAAGCTGCC	
			AGCCGAGGCTCCAA	
			GTAAGCTGCCAGCC	
			GAGGCTCCAACCAG	
			GCTGCCACTCGAGG	
			CTCCAACCAGGCTT	
			TCACCCGAAGCATT	
			ACCCAGGCTGCCCG	
			TTTAGGCACAAACC	
			AAGCTGCCAGCCGA	
			GGCTCCAACCAAGC	
			AACCAGCCCAGGTT	
			CCAATCAAGCTGCC	

GVEVSFNSSGSLQMDPEVVQPPSPSISFSPMHLSPTQPVSPIQIGIEVSFT
SSSPLQMDPEVVQPPSPSMSYSPMHLSLTQPDSPIPVDIDVIPAAEVPLP
DIEIPPSPDRHPAAEVPLPDIEIPPSPDRHPVAEVPLPDIEIPPSPDRHPQSP
PRPVMMEQPVHTPPPADTQQANGPQHWVTVLANATNWEDFGRVCV
EFANHAVEAARSRQDAPQVRPAAQRQPRRPTRPRQPTFDVREASRLQK
LYKRSKKRAVRHILRDDAPSFSGSNEQLLDYFKEIYAPPEIDENRAQQLAE
SLFTDLEEAKESAAALMSPISQQEISTRLSRMSNSAPGKDRLEYRHIRQAD
GACRVTHIMFNRCLQEHRIPSAWKEATTILIHKSGTTDDPANFRPIALQSC
LYKLFMGILSDRMTQWACNHNLLSPEQKSARPCEGCHEHTFLLSSVIKDT
KRNQKTANIAWLDLRNAFGSIPHQAIHAVLTTIGAPVSLVMLLKDTYTGA
STSFLSTSGETDPIQIQSGVKQGCPMSAILFNLTIELIIRAVKKKKATDDGLGL
VVHGQRLSIMAYADDLVLMSKTPEGLDAILSVASEQAETLRLAFKPTKCA
SLSLSCRHGTSVLPREYTVQGHLMPALDEEEQYRYLGVPFGLPRFTNLKD
LIGKLKGNIETIASSLLAPWQKLDAIKTFVQPGLSFVLRAADYLKSDLRSLKS
AITTNVKKICQLPLRAANAYIFAAKESGGLAFIDPNVDADIQVITQAVRVLS
SDDEVVQTIATSQLKSVVHRTIHAVPTEEDIDNYLSGSNEGLLANSGNSG
QASSLWSRTRSAARRLHLTLRATTSGTVVVNQQADIDHTRDILPASITRG
LRLIQRTTNAEKLKSLPDQGKVARSLSNDPFANGSSWHATGKFIRFCDW
RFIHRARLNCLPTNVATKRWKANANGKNGHQQETLPHVLNHCLPNMV
PIRRRHDNIQQRLVTAIRHGDVFVNQHVPGDPNPRERPDITVIEGNKVT
VIDISVPFDNGPNACTTAAQAKVEKYSALRQALRDMGRDVEVHGFIVGA
LGTWHQGNERALGRLGVSRWYRTLMRKLCCIDAIQASRDIWVEHVTGH
RQYE (SEQ ID NO: 1439)


AGCCGAGGCTCCAA
CCAAGCTGCCAGCC
GAGGCTCCAACCAG
GCTGCCCACCAAGC
TGCCAACTTAGGCT
CCAACCAAGCTGCC
AGCCCAGGCTCCAA
CCAGGTTGCCACCC
GAATCTCTGCCAGT
TTAGGCACCAACCA
AGCTGCCAGCCTAG
GCTCCAACCAGGCT
GCCACTCGAGGCTC
CAACCAGGCTTTCA
CCCGAAGCATCACC
CAGGCTGCCCTTTT
AGGCACAAATCAAG
CTGCCAGCCGAGGC
TCCAACCAAGCAAC
CAGCCCAGGTTCCA
ATGAAGCTACCAGC
CGAGGCTCCAACCA
AGCTGTCAGCCGAG
GCTCCAACCAGGCT
GCCCACCAAGCTGC
CAGCTTAGGTACCA
ACCAAGCTGCCAGC
CCAGGCTCCAACCA
GGTTGCCACCCGAA
ACTCTGCCAGTTTA
GGCACCAACCAAGC
TGCCAGCCGAGGCT
CCAACCAGGCTGCC
ACTCGAGGCTCCAA
CCAGGCTTTCACCC
CAAGCATCAACCAG
GCTGCCCGTTTAGG
CACAAACCAAGCTG

				CCAGCCGAGGCTCC AACCAGGCCTCCAA CCAAGCTACCAGCC GAAGCTCTGCCAGA TTAGGCACCAACCA AGCTGCCACCGTTA GAGGCCCAGAGCC CACCAATCCCATAA CGTGTGAGATATGT GAAGCTTCCTTCCA CACCTCCGCCGGTC TCCGTCGCCACACG GCCAGGCTTCATCG CACCACTGCTGAAC ACACTGACGACAGC (SEQ ID NO: 1194)	TGATCCGTCCCGAACCAAC GGAACCACATTGCCGCATG ACTTCGATTTCGCTTTTGCT CTTTTTGTATTTAATTTTGCT ATTAACAATTCAGTTTGTTA ACTGTTTTGTATTCATTTGA AGATCCAAATAAAAC (SEQ ID NO: 1318)	MCSAKTPALKSGRRRGKEVNYEGQIVRVERRRGSRSSTSATDLGTRMVT RGRKKLMEASVREAGHHGGESASTVDVDVVESKKITGKTARRNRRAPSG DGKRRESCGAECGQAVCGNAVADRSEASSPRTPNVSKSGRDKCGQPTIK ASTPSPPKRKPTTSSSPRTPCLSKRGARSKIPSTPDTPSTSGGSGKQRVLVS PLLRTEKLPDLEVLQRTEEQVTVRATFPIAQAVVCPLGCEKPYTAVRPDG QFAHQTLTRHFMRVHNCHSVQWHYRCRNCNTDFLPADHRYPLRVVNT HVRSCVSRWEITRKLGESEDLHGVRCDLCDVVGVSKRAVGLHRRRHANE NIMQNTGTAAQIEALSKQVGEIRVAGDYSQFKFGKRVRQYVAPTQRRD GLDEEEVHEAEEEEVPAESRTILGEPSTATAIGAEEISATGPVRADTAAQQ MICRIGQWCVWPQDYHSIPAPQCWTDTLMDLMIEQIVLQRYPDGAGV SVMSCSAVSAAIHHEISAEFAAQVMSSHDASLYCIIPVNVRNHWQMIVL DVAERVVHYYCSLREHNTVVLSSLLSLVELSGKHTGCTSWKIETHDGAPV QTNAFDCGPFSCLFLKHLLHGIDMNFGDRESAALRTDLKFMIDAVSTPVV PATDKLKKKPDGSATQLTQFQQKFLSASDNWQSPDVDLQAVYDEVVESI VSGHNEPSSRNTSRLQKKSPGKGGKGQVRRRSAVTRDPAWLKSASAVQ KAFNSAPARTVNAILRRPNACPSFTATQVADHYFNLRPAVTSLAPEVIDIL PPPATDHSMLVAELSESEVVWEKMQKAPNSAPGADRITIRMVRMADPG AMILTRFYRACLLRKWVPLQWKQSVCKLLYKDGDKERLANWRPIALEPV LQRVLSAVVASRVTNWARANGLISLEAQKGFQPADGTSEHNFVMEVAI QEARRTNAQLAISWLDISNAFGTVSHQLLFSLLERYGLDPTTFSFIQNLYK DATIVVKGANGTHVTARWSVGVRQGDPCSGILFCLFVEPLLRSVLPSLPC EAETTAVNVLGQPITALAYADDIALFAPSIGVMQQQLCKIQGMASAMGF
NeSL	Utopia-1_TSP	.	Trichinella spiralis			

						RFNPKKCASLYLNRAVVNAATFTISGEEIPALVHGDTFRYLGVAAGLGKP
						QTPFSLLRENLREAELIFRSKLAPWQKMDAYRTYVLPRLTFQLMIAKFNNI
						KQSAGQYDRAILRLVKRCFQLPVETSTDFIRAPRQCGGLGVPSLRELYATA
						KVSRALKMLWSPCRVVSSLAASQLQRVASAYFAKRLRDVEAADLSTFMN
						AARSTPLDRSGYPTCLWMDARKQMSYLTKVAGVDCYFLVGEAGTSFFIR
						NGLGQTVSVLSPLRKNKVMSVLGGAIQTRHLDAWLQCKRQGKTASCIVL
						DRSSSRFITTGRYTSFAAMRFALPARLDLLPCRARSSMRSYQNCRRCGYD
						RETLPHILQHCRQFSAPAYQARHDAVQGRLETVMRRRFPNLRVNRALPE
						IGSNKRPDLVVVDEEKRLVILLDIAIVFENTAAAFVDARTRKWAHYEKEILA
						YRLRGYSVTYDAIVVGALGTWDPKNDAILKRIGVVSQRYLRLMKVLVVSE
						MLEHSSRIYRKHLGLRDLLPDTGTKRRPVGTTETDPPGGDLRQKKRNTIS
						ARASGGKCLERRFTSPVGTPSQRGELQCQPCPGPRRPALAGIAPNPPSLQ
						PRKPPPRQHQKPVTKSTAH (SEQ ID NO: 1440)
NeSL	Utopia-2_CMy	.	Cheloniamydas	GGCAGAAACTGCAC	TGAGCTGGAGTGCCGATGA	MLQLRLPTPQTLRLLHPSQLPQSHSTKRWSNIYEERARAPKDTTIERVSA
				STTCTAGAAGACTC	GAAGCGCAGTCGGGAAAGT	ASKIPKLDPAKRRIGAPLQLMQGNSISRQLSASSQYVQHNAVVRRVSAPP
				ACTGCCTATCCTGA	AACTGAAATACTTTCCTCAT	HTGNFTGSCRRKLALPHRPPSETQPAESYLHCWTTQRDPTLADQHGLQ
				GGAAGACTACCGCT	GGATTGTATTTTCTAAATGG	DHPTGLPPEGSHCIKDPRQNSSTEGQRRSGNSQARRIPKRASTAASKTVL
				TTGGAGATGGATTC	ACAACCTACCTAATTCTCAA	PKRTSAALKSVREDTALVLEDPAKWSSQHREGXRQQANPTAVFQPEPAE
				TACTGCCGCTGCTT	TTACTGAGGGACAATCTCC	IEQQPXVRAATPWQAAWMEELARTASFXDFDLLVDRLTKDLSAEIVSGR
				CAAAAGAAAATCTT	ACTCATTGATATATTTTGCT	KGTQENTPTAHRQNQNNMREARRRNISRCYDPAAASRIQKLYRSNRPK
				CATGCTGCTTCGGA	TTCCACAACCAAATCTCTGT	AMREILDGPSSYCAIPSERLFLYFKGVFDRVAQNDMQRPECLXPXPRVDY
				GGCTCCAGGACAG	ACAACTTTCATGAGTGATG	AEDXEQDFTSWEVEARLTKTKNTAPGKDGIRYNFLKKRDPGCLVLTAIFN
				ASTGAGAAGATGAT	TACCCGAGTACTTGGATTCT	KCKQFRRTPSSWKKSMMVLVYKKGKQDNPNTRRPISLCSTMYKLYASCL
				CGCTCCCTCGCCGA	AATATCTAAACTGTATTGTT	AARITDWSVNGGAISSIQKGFMSCKGCYEHNFVLQTAIHMARRAWRQC
				TTCCACAGAAACCT	AAATCTATTCACCTAAATTT	AIAWLDLANAFGSMPHQHIFDMLREFGMPENFLQLVRELYEGCTTTICS
				TCAACTGCCGCTCG	GGGTTATTGCTGATTATGTA	MEGETPEIPIRSGVKQGCPLSPIVFNLAMEPLIRAISSGLGGFDLYDNRVNI
				GACCGCTGCTGCTC	CTCTATGTATCATATGACTT	LAYADDLVLIADNPESLQQMLDITSQAANWMGLRFNARKCASLHIDGSR
				CACGGAGCGCCCAT	TTAAAAACAAACTTTGTATT	RDSVQATSFQIQGEPMIFLEDGQAYQHLGTPTGFRVQQTPEDTIAEILRD
				CGGGGAACAGCCTC	TGTGGATAATCTAAGCACT	VARIDSSLLAPWQKINALNTFLIPRISFVLRGSAMVKVPLNKADNTIRQLV
				TAAGAGACCCTCAG	ATACCCAGATGTACAGACA	KKWMFLPQRASNELVYISHRQGGANVPRMGDLCDVAVITHAFRLLTCP
				CGCTTCCAGGACAT	CTCTTTCCCAACCTATGTAT	DAMVRNIAESALQDAVKKRIARTPSNQDVATYLSGSLEGEFGRDGGDFA
				CGCAGATGAGCAG	TATATTTTTTTAACATTAGCT	SLWTRARNATRRLEKRIGCHWTWCEERQELGVLVPQVKNTDHTIITPRA
				CATCGATTGGAAAA	TTAATAAAATTTTTTAAA	RTMLERTLKDAIRCQYVENLKRKPDQGKAFEVTCKWDASNHFLPGGSFT
				GCAGCGCCTCCCTG	(SEQ ID NO: 1319)	RFADWRFIHRARLNCVPLNGAVRHGNRDKRCRKCGYANETLPHVLCSC
				AAGAAAACCCGTG		KPHSRAWQLRHNAIQDRLARAIPPPVGKVAVNSAIPGTDSQLRPDIVITN
				GAGATGCTGCTGCA		EDRKKIIMVDVTVPFENRTPAFHDARARKVEKYAPLAETLRAKGYQVQT
				GGGAWATCTTGCA		HALIVGALGAWDPSNERVLRECGIGQRYARLMRQLMVSDAIRWSRDIYI
				CCTCGAGGACGGCC		EHITGHRQYQEG (SEQ ID NO: 1441)


TCCCAGGACATCGC
AGCCAGGACAAAC
ATCATCTCGCCTGCT
CTTCAGGASAAGGA
TGCCMSAAGAACTT
MTCCCACCTCCTCM
ACTGCCCAGGATCC
TSATGCTGGTCGTC
GTCCTGCTGTGCTG
GAAGMKACCCCAA
CTGGCGAGACCTCA
GAAACCACCCAGM
WGGACCKCAACAT
CTCACTAGATGCCT
GCCCAGCCGAACCT
CTCCATGCWACTCT
CCCAGAGCAACSAG
AACCATCCAGMGA
ATCCGCTGATATGA
CTGAAGCCMATCC
AACAGAGGGAGAA
GGAAAGGAGAATG
ACTGCATCTATCTCC
AGTATCCCCTCCCTA
CGGACACGCTCCTC
TKCCCCTTCCGCTAT
CCGAGGGTTCCAGT
ACATTGGCAGTCTC
AGCAMACACCTCA
AGAGAATCCATASC
AAGCGGATCACCTT
CCGGTGTGCCCTCT
SCGACCTGCCTTTC
GAGACGCAGATGA
AATGTAAGTCTCAT
CAAGTCACCTGCAA
AGGACATCTCGAAC
TGGAAGAGTCCAAC

NeSL	Utopia-2_PCa	.	Phytophthora capsici	TTTACCAGTCTATGT TGCCGCCACCCCAT CTCTGCTCCGAAAG CAGAAACACCAC (SEQ ID NO: 1196)	TAAGCTGGTCATTTGACCG ACAGGGCACTACCCAGGTA GGGAACCGCCCTTCAGAAA CCCAGGAAGACACAAACAC CCTCCCTTTAGTGACATACA TATTTTAGGCTACATTTCAG TTACGGAGAGGTTACTAAC TGGTAAATAAAAAGCACTTT (SEQ ID NO: 1320)	MQDMEEELLLDVEMETETTEPQTSTAXDATTTTDRPTRWGPHPRAVAA AAIAQLVTGEXAXPALPSRQDRRPAPRSHAPTRSRWGPRHQAVGAAAI ASLATGLPASAAPVSRATKHGEGRRRLQTRWGPRVSIPRAARRPGSRW GPPRAAGASGQLPASASGATGQLPEHVEAITTTPRVASDADEGPTPPDP WILRFDGACRRNPGPGGAGAALFKPSGAVVWTCSHYMPSSNETNNTA EYTALLLGVQSAVHHGATRLDIEGDSSLVIAQVKGTFACRNAKLRQLRNR VRHALRSVEKYTLRHIDRKANAHADRLANRALDRRSSSSECEPHGSCME RCCGTDTTPAVQGPTPQAAAAVPVQVWPQWQRQTMVAWTTSHGGR CRDCSTRCGRSLPSLAHRPRLSPRRQPRLRLRQLSDEERDXAADALQELS DVMASKIVDADSWDTGEGYISSIPERIREVLQPYTTRPPRPGHQQQQRR RPPRVTRNQREHRLDEALDDMQATQQAAPRDQRAIHRARRRVGRVRA SMAKQELRQAFAKDESKCVSKILAGASAETAAEEHVDECPIDAATLHAHF TGTNAPRTDFDYDAACGQEFRGALDSMQPPTVATDAFEEELTIDEVEDQ LTRAAKTSSPGHDGIGYDIYSRFAAQLVPLLHAAYQFCWLHRRVPALWKL GIVRLIHKKGDPMQPTNWRPICLQPAIYKIYSGLLARRLSRWMEQNQRL PMAQKGFRAFNGCHEHNFVATTLLDQTRRSHRRLYQVWYDLRNAFGSL PQQLMWSVLRHLGVDASFIARCKNIYQDSAFVVANAVDGATDPVRQEV GVYQGCPLSPLLFISALVPLIRRLEKLDGVGVPLAEGVRPCATAYADDIKVF SDSAAGIRKCHDAVTRFLEWTGLRANPGKCASLAVTTNARGNPVRDDG VHLELQGEVIAPLSLHDSYRYLGVGDGFDHVRHRLQLEPKLQQIKREAVA LMQSGLAGWQVVKALKTFVYPKVEYALRHLRPLQSQLQGFDRAVVRGL RHLLRLPQSATTEFFYTPTSGGGLGLQSLVEMHQALQVAHAWQMLHSK DAAVVAVAKEQVCQVARKRYRLQEEHWRGRGDELVRLFLNSELAASPF ADCLRRNGDIGSLWTDVQRTLRLHHLSLTAQDDRDGQDPLALRVPHHT KWLDHKTVLRHVKLHMKIRHQTRWKGLVDQGKTVRVHGGLGAKFVST GAGLSDDAYRFGVKARLNQVDTNAVLKRKRLRSSKTCRDPTCSSAETLA HALNHCASNMDAIRQRHDDALEQIGSKIRGALERAKSTTELRLNQTVPEY TGAALRPDIVLRNVAAKKMVIADLAVTFEDHAAGARHSSLQLSHDHKTL KYQPIVAELRVQGWQVQTAAIVYGSLGSVQPSNFKTYTEKLKLHKREAR QLDLQLSSHCIQASHRIWGWHCRRHREGQRSGNTSRASRGSGGTPRRT SQVRARR (SEQ ID NO: 1442)

NeSL	Utopia-2_PI	.	Phytophthora infestans	GCTCGGCCTCGCGG CTGCCTTCCCAGGC GCCGCCGACTTCGC GCTCTGGCGCGGCC CACACGCCGCCGCC GAGCCTCCAAGCGC GCCCGTTGGCTTCC GCAGACGCAGGGC TCGCGGCGACGGCC	CACTTCCACAGCGG GTGCACGTGTGGCA TCGGCTTCCAGTCG CGCGCGGCAGCTAC GCGACACGCTCAAG CCTGTGCAGACAGC ACACACGCCACCGT AGCTGCCTCGCGCG ACCCGGCMNCCAS TGCSSCCSGNGCCG GMAGCGASGAGGA GGACCNCGCACCCC CCGGTCCTCTSSTCG CGGCAGCATCTGCA GCTGCAGCAGCCGC ATCAAGCGCCACTC CAGCTGCAGATACT GCCACCACGCAGAG CGCCGTGCCCATCG CTACTCCTGGGCCC CAGTACGCCCCCCA CGTGCTGGAGCCAC CTCCAGAGCTCCGA GTTTCCGGCAAACG CCGMCGCCTCAAC ACGCCGATCGACCT GCAGCCGCTGGAC GTGGACGCGCTG (SEQ ID NO: 1197)	TAGGCGGAAACCAGGCCCA AGACGGCCGACAGGGCCCC ACCCAGGTAGGGAACCGCC CTAGAAACCCATTTCGGTG GTCGACTCGAAGGCCTTAC CTATTTTTTCCTTAGACATTC AATTAGGTAGCGACCAAAT TACAAATTTGGTAACGAGT AAGCCAAATGGTAATACAC	MLADPAALAAGLARAPPPPSAPQDPSPAFPAGPAGQNPRAAAPARVEV HTVVAPPGRAGGMLPDPGLVDSSPAAATAATPAPVAATATTARAAARV AVEHHAHAEPNQEHLPMARVLVEPMQVDECSSCDRSTLTADDGSGDD VAAPSSMLSNDVAAPMDVDSGTSCPPTLQQPLQRPRALHVGSKRRRLD ADDGEEAHQLQEEEEAGIHAPALRLSAASAQPASVLAVYTHNASRFDCT LCAYTAGSFASLLTHRNSRHRRTAFLDRFSAGCACGVPFASRLAAARHAQ ACASLSSAPSAEASSAAGTSSPTADGADSTVSAVAHAEPGLPHHNDTELT ASPPLVSSSDVEVQATKTEATDNRWGAPLPRVLVASRIAGRLAQVPPPR WGPPLPRTTIAARIATRLAATPAPRWDPPLPRSLVVSRIAARLLPALPDAP

CTCGCCAGCCCCCA	AAAACTTTTCTGTTCTAATC	ACEEEAKDSDTMDWAPTWTNEETKESEPHDEAPGQVDEETIDDADGE
AGACCCCCCCTACG	AGTGTGAAAACTGGTTTTC	WLLRFDGACRANPGPGGAGAALFKPSGPVVWTCSHYDPSTTATNNTAE
ATGTGGCACCACCC	GCCTTTTGGCGGACTTTTTC	YTALLLGARAAAADHGVTKLRIEGDSTLVIQQVRGIFATRSTRLRALRNKVK
GGCAGGGCGGCCG	ACTCGCATTTTTGGGCAATC	LELARVGSFSLHHIDRQANGHADRLANAGLDRRRTKLECSVHPDGRGCT
GCAGGCTGCCCGAC	GTCTGCGGCTAGCTTGCTA	NTSVATAAPTAPAAPLPSARPPASTAAPSPDDDHSDQGDIDDGEVYAA
TCGGTATCGCCGGG	GCGGCCGGACGAGCGGTCTC	MCISPDAVPHRRPRLRLRRLTDEESEEAGNVVERLAASLAAKIADAPDWE
TGCTACACTCTCAG	CGGGGGCGTTCACCTTTCCC	TAEGYITALPYALYDKLQPYSQSQHQPPRQQQQQQQRQRPRQQQQTRQ
CCGCCACAGCTCGG	CCGCGAGGCCAACTACACC	RRQRRCKRGGGSQHRQRKTRRRRPPRVTRHHREHRIDEALDDLHALESR
GCCTTGGCGGTCCG	GATCTTCTCTACACTTTTCTA	RPQDRTAISKARRRVGRIRSALDQHQLRHRFDTDEKACVDGILAAARDK
CCATTGGCCCTTGG	ATTCGCCTCCGTCTTCGGTC	DRAASVTTTAQTAAPPHSAPASAPSSAVDDGICPIPGDLLHAFFTDVNTP
AGCTCGACAGCGAC	TTCGGCTGTCGGATTTTTTT	RTEFDADSPIGARFREALAQLPAAIAATELLMEPPSPDEVEDQLQRVRGT
AGCAGCGACGACG	CTTTTTGACCAATCAGAGCG	SSPGLDGVGYDVYKTFTQQLLPALHAAFSRCWTDQRVPQSWKLGVVRL
AGGACGCTCAAGAC	CGCCATGCGACTCTTCTGGC	LFKKGDRQDPANWRPICLQQAVYKLYAGILAHRFTRWLDANTRHADAQ
CCCCACGCCGCCGC	CAATCAGAGACCGGGCCCT	KGFRAVNGCGEHNFLAATLTDNARRRRRELHVVWWYDIKNAFGSVPHELL
CCCAGGACCCCCCG	GTCCTCGGACAGCGAGGCC	WEVLRRMGVPAQFIACCQGIYDAAAFTVGNAADGTTAPIQLRLGVFQG
CAGACGTCGCGAGT	TCCACGGCCAGCCAATCAA	CPLSPHLFTAVISPLLHALKRLPGTGVQLSAVDRPGASAYADDLKVFSDTK
GTGCTTGCCCCACC	GTCTCGGCAGCGACGCGTC	DGITRQHQLVTDFLRWTGMVANPSKCSTMSVQRDNRGVLKTANLTLQ
CGGCAAGGCAGGC	TTTCTATAGCGCAGCTGACG	LDGAQIPALGMTEAYAYLGIGDGFDHVRRRVELAPKLRELKADTTALMQ
AGC (SEQ ID NO: 1198)	AGGCCGATCTGGCGGCCCCC	SGLAPWQVVKALKVYIYPRVEYALRHLRPFQQQLQGFDRHLARGLRHLL
	CGATTGGTCCGACTTTCGGC	RLPTSATTEFLYAPTSRGGLGLLPLTEVHGALQIAHAWQTLHSPDPAIRRI
	CAATCAGCGACGACGAGGG	ARVQLRQVADARHRLDAEHWKERGEELCERLLNSQLGTSAHAPPKRRN
	GGCAGGGGTTTACACTTTT	CDIGSLWVDVQRHLRSLGLQLQTAPADTHTGAPAQPLQLRVPHHDKWL
	GCCCCCGTTTCGACTTCAAC	THKDVLRHVKLHIKNNHWHRWTSMRDQGKTARAHGGEGSGFLTQPR
	TTCAGGCCAAAATGGCGAT	GMWEADYRFAVAGRLNQVDTYSVLKRRRLRSHDRCRQPGCHRAETLA
	TTCGACCCTCCACGCGCCGT	HVLNHCPGTMDAVRGRHDGALKRIERELHASATDRRDRVELRVNQTVP
	GCCACTGCTCGGCACCGGC	SLAGPALRPDLQLYNHTKKTVAVVDLAVAFEEQASDDASSSALSLIASHKR
	GGCGATTCAGCGGGTGCAA	AKYDRIKRHLERQGWKVHLSALVYGSLGAVASGNYQVYTTHLGLLKRDA
	CTTCGGGCACGTGTGCAAC	KRLDRQLSVECIQSSRRIWNLHCSQHRTRQHQARPSQGPRGSRATETGG
	ACATGCAGCGCCCATTGCA	TPSQTSRR (SEQ ID NO: 1443)
	CGCCAAGCGGCATCGCGGG	
	ACGACGCCTCGGCCGCTCA	
	AGCGCAGCCCCGCCCTTCC	
	AGCACGACCTCGCGCCGTT	
	TGGCGGATCGCCATCAAGA	
	CGTGCGAGAGCCAGGCGG	
	GGTCGGGCAAAATATACTT	
	ACTCTAAGTATGCCCGAATC	
	CCTGCCCTCTCAGGCTGAAC	

NeSL	Utopia-2_PR	.	Phytophthora ramorum	TCAAGCCCCGCCGC CAAGCCCAGCTGCG GCTGTTGCCGCCCC TCCAGCAGCAGCAG TCGCAGCTGAACCT ACAGCCCCTCCTGC TGATCGCGCCGCCG TGGAGCCCCGCGCG CGCGCCGAGCCGCC CCAAGAACAAGCAC CCCCAGCTAGCGCG CGCGTGGAGCCC (SEQ ID NO: 1199)	GCGGCCCCATACTTGATCTA AGTATGGGAGGGATCCCTG GCCTCTCAGGCTGTACGCG AGACCCGTACGGCCGAATC CCCTGGCCTCTCAGCCTGTA CGCGGGGC (SEQ ID NO: 1321)	MSGDVVSSDGSSRTTDASGDGDDGAGSSDAAGDVGVVAMDVDQGA
					TGAGCACCTTGGGTTGCTC AAGCGTGATGCGAAGCGGC TGGACCGGCAGCTCTCGGT GGCGTGCATCCAGTCCAGC CGCCGCATCTGGAACCTGC ACTGCAGCCAGCACCGCGC GCGCCAGCACCAAGCACCA GGGGGAAGTCGGGCGGCG GAGACCGGGGGGGACTCCG CCGCGCACCGGCCGCCGCT AGACGGCACACAGGCCCAC AGCGGCCGACAGGGCCACA CCCAGGTAGGGAACCGCCC TCAAACCCCGCCGGTACATT ATGGTCCGACACCTATGAG GTGCAACCTGTACACAAGT TACACACCACATAGCGACTA CCAGGTATTTACTACCTGGA AGCCAAGGATTAACCGGTC GGTAATACACATAACTTT (SEQ ID NO: 1322)	RRQQPPWQRVGGKRRRLNDVDDEDTRELAELLLEEEDEAGDHAPAPRL
					SAASARPASVLSVYAHNAQRFQCTLCTYTAASFASLKRHRDSRHRRTAFL
					DRFSAGCACGAPFASRLAAANHAHACASLNRTLSVAATPAAGELSPTAG
					AANATVKAATVTPDSPRQDPPELAASPPLASSPDVAVQAADMQAPTR
					WDPPLPRTLVATRVASRLTDLTPPRWGPPLPRATVVSRIAARLEAAPTPR
					WGPPLPRVVVASRIAERLAPPELAADDETKDGEEDQSFTEPVAAARSXG
					GEDANGEWLLRFDGACRANPGPGGAGAALFKPSGPVVWTCSHYMPSS
					SETNNTAEYTALLLGMRAAADHGATRVHVEGDSTLVIQQVRGIFATRST
					RLRGLRKSVKAEMARMEHVTLHHIDRQANGHADRLANAALDRRKTKLE
					CGLHPDGQGCSSTAATTAVPSVVPDRPPSSTAAAPTPSAEPDETEQGDI
					DDGEVYAAMCIGPDSIPERRPRLLRLRQLSETEEEEAGAIVERLAATLAGKI
					TDASDWATAEGYITALPYTLYDKLQPFAQHRHQPRPQHRQQPQRDPPL
					GTHDGDHGQPSTSRSRRRRRRRAKDRLRRRPPRITRHHHREHRLDEALDDL
					RAVEHASPHDRPAVARARRRVGRVNSAIAQQQLRHKFDKDEKACVDGI
					LAAARASRGLATPSASASRHPPPVPSTAADDGSCPIPSDELHAFFTAVNT
					PAGTFEPMAPVGAPFRSAVAHLPAATSQPELLSDAPTTDDIEDQLQRAR
					GSSSPGLDGVGYDIYKAFAAQLLPALHAAFACCWRHKQVPQSWKVGVV
					RLLFKKGERTEPANWRPICLQQAIYKLYAGVLARRLTRWLDANGRHADT
					QKGFRAMNGCGEHNFLAATLVDQARRKRRELHVVWYDFANAFGSVPH
					DLLWEALERQGVPSPFIACCRGLYADAVFTVGNAADGTTAPIALRVGVF
					QGCPLSPHLFTAAIAPLLHALKRLPDTGVQLSRVDCLGASAYADDLKIFSG
					TEGGTKRQHALVADFLRWTGMRANPAKCCTMSVQRDTRGVLKACNLG
					LQLDGAPIPALTMSASYAYLGIGDGFDHVRRRIELAPKLQELKHDATALLQ
					SGLAPWQVVKAVKVYLYPRVEYALRHLRPFHQQLEGFDRHLVRGLRHLL
					RLPANATTAFFYAPVSRGGLGLLPLTELHAALQVAHGWQMLNSKDPAIR
					RIARVQLRQIADARHRIDAQAWQDREEELAQLLLNSQLGASTGAPPKRR
					NGDIGSLWVDVQRHLRHLSLKLETAPACAETGTAAAMLQLRVPHHDK
					WLDHKTVLRHVKLHYKNKHWARWAAMXDQGKTARTHGGAGSGFLTR
					PRGMWEADYRFAVAARLNQLDTHSVLKRRRLRXHDRCRQPGCTQGGD
					AGARAQPLRRHHGRGPRPPRRRPQAHRARAARVVAGRPGPRRAPGQP

						DGAVARRPRATARPPAVQPHQEDGGGGRPGRGVRGAGERRPGELGA GTHRRTQAREVCRRQATPRAPRVEGPPLGARVRLARRGAGRQPQGAY (SEQ ID NO: 1444)
NeSL	Utopia-2_PU	.	Pythium ultimum	TGACTGGTGTTTGA	TAAGCGGGGGGGTCCAGACC	MDYDDSEFFDAICIPDEDADVLDDGDEGDEGGNDDESSEPLPLAITNAP
				TCACGATCAATGAG	CCACAAGAGAGAAGCAGG	SAPLHATMLCGTVTQPWLLRFDGACRRNPGPAGAGATLTRPNGIILWT
				GTGATTAACATGAG	AATCATGGTCCGCATGGAC	HYRYIPDKTATNNVAEYEALLDGLRCAAHHGVKHLRIEGDSNLVIEQVKG
				CCGGAGCAGGCCCC	CAGTAGGGCACGCTCCACA	IFACSTSLRPRRDQVREILRHFETYSFRHIDRALNRQADRLANQALDLLKT
				TTACACGCTGGTGC	AAGGTTATCGCCCTCAAACC	VSVCALSQTRVQDDTGAAHGCWHWTPPDASPTDDASTSILTQDVPVP
				TGTAATGGTTCAGG	CATCACACGAAGGATCTAA	MDIDDYDPDEPMNAADDPVSINAEREGGTVYPVLRLGPNVVPERQKRL
				AATGCTCTTATGAG	AAAGAAAGCAACAATCGAA	QIPWLPPREMQKLEKKIEVLGETFASRIRDAPDWFSAEGYITALTSELATLI
				TAACTCCACAGTAT	ATAGTAAATAGTAAATAGC	RQSTAATTGPNAARPCERTISKEKRRARRTTPLQRALAEAKHELQIIQPDA
				AATTTTTGTTAGCG	TTAGAAAGGTCAACATGCG	SRKSVRKAARRVKRISQAQQRHDLRRLFSTNERRCVEKILRDPPVGPSSTS
				GGGGTGAGCGCGT	AAATGCATGAGGAACGTAA	SSLPATDDDRCTIDPADLFAYFQTQATAPTNFDFDDEGGELFRSVLDELP
				GCGCCGCCCCCACC	GATGGTAATACATTTTTATA	RADQEVHLLEDEITRDEIEDQLSRISKSTAPGLDGITNAVYVRFKLQLLDAL
				TTTCTCTTTCGTTT	(SEQ ID NO: 1323)	QAAFNACWRYNRVPSMWKAAFVRLIYKKGNRAVPSNWRPICLQQTVY
				GTATTTGGTTCCTCT		KLYTAILASRLQRWMDANARFTMSQKGFRAFNGCHEHNFVATCLHDQ
				GGAGCGACTTCCGT		TRRLRKKLAIVWYDLRNAFGSLPHEYLWRVLARLGMPPQFVARVRQLYA
				GGCTTTTGTCGCGC		DASFTVESRDGTTDPVQLERGVYQGCPLSPYLFIAALIPLVRALHKLKDQH
				GCTATCGCGCCCGG		GIVLAPGVTDCVSAYADDIKIFARSGTGAKALHEIVVRFLSWTNMAANPA
				CATGAGGATCGTCG		KCALMVTDGARGGDDTDASMTLSIEGETIPRLTGKEGYVYLGVEDGLAH
				CTCCTGACGCCATG		ERRATCLRDSLKAASADVVRLLRSDLAPWQIVRAIKSHVLSRFDYVLRHLR
				TGCGCCAGTATATC		PFLSLFDGFDKMLVRGIKRLCQLPQTATSEFLFSPTSAGGLGFLPLKELFAA
				GGAGCCCCGTGCTC		LQIVHALQMLHSKDANVRAIARHQALQVVRKRYALQSDHWSDREEELL
				CGCTTGGGTGTGTG		EEFFNGTLERSPFALAKKVSGDIASLWTDVRVNLTKYGLKFGEAHGRRLQ
				TTCCGCCACGACGC		PLVSHTDKQLAPQQWASAIKTHMRLRHLKRWTTLVDQGKTARMHERI
				CACGCGATTTGCGT		GSAFLTRPSGVYDASYAFAVRARLNQVDTRSALKRKRIVNNSRCRVSGCS
				GCACGGCATGCGCT		ELETLAHVLNHCRFGSDSIRARHAETLLLIKTTMERELTRPGRQHQRLLVD
				GGCGACTTCCGCTC		ATVPEARDPVPSNDAAESNIGAMAPSISHLRPDIQLYDNKTMEAVIIDLA
				AGTGCGCGAACTCG		VAFEDQSTDDAASSSSFARVKGVKTKKYEVIKQFLEYKGYTVHVAALVYG
				CGGTGCATCGTCGA		SLGSVDTGNFAVVTERLGLRKGAVRRLECSLSARHINFAHRMWRRHAIA
				CGTGCGCACCGGTC		HTTGLRLIGTNSVQQQGVQRAPAEKQQHQRPVQRPSRAQAPRDQPSQ
				TCTGGCCTTCCAGG		QQQSQQSFQQQSQQQSQQQSQQQSQQSQRQRHAPTPTPVPVPVPVPV
				ACTGGTATGAGTCG		STPTLTPTPTPTRRPKPAPSSTQPQQGAPAQRRQQREQKKQPACRRRH
				TCGTGCGCGTGCAG		ATAVRETPPAAPTAARTATPTARPTATTRSTSTTRSTATATSTTRSSAPTS
				AGCGCTCTTCACCG		RPSAPRPRSSAPRPRSSAPTTRSAAPTTRSAAPTAIATTSVYKSRRTTNAIS
				CCCGTCTCGACGCG		GATTRRSASSKRAPMQPRTALTPTQQQQRQ (SEQ ID NO: 1445)
				TTCGTTCATTCGAG		
				CCGTTGCGACCACA		

NeSL	Utopia-3_PI	.	Phytophthora infestans	GTACGGCCGAATCCCTGGCCTCTCAGCCTGTACGCGGGGCTATACTTGGTCTAAGT (SEQ ID NO: 1201)	TAGACGGCAAAGTTCTGGCCCGCGTAGGCCGAAAGGGCCCCGCCCAGGTAGGGTAACGCCCTCGGGAAAACCATTCTGGTGTTTGGCTGTTTTTCAACAGTCGAACTTCAACTCGGAGCATATCAGATACACTTACTCACACATTTAGATATCAGATAGGGAACCTTTATTAGGGAGATAACGGGTACACCGG	MPRSLASEPVHSSASRLPSQAPPTSRSGAAHTPPPSLQARPLASADAGLAATALASPQDPPYDVAPPGRAAGRLPDSVSPGATLSAATARALAVRHWPLELDSDSSDDEDAQDPHAAAPGPPADVASVLAPPGRAGSMLADPAALAAGLAPPPSAPQDPSPALPAGPAGQNPRAAAPARVEVHTVVAPPGRAGGMLPDPGLVDSSPAAATAATPAPVAATATTARVAVEHHAHAEPNQEHLPMARVLVEPMQVDECSSCDRSTLTADDGSGDDVAAPSSMLSNDVAAPMDVDSGTSCPPTLQQPLQRPRALHVGSKRRRLNADDGEEAHQLQEEEEAGIHAPALRLSAASAQPASVLAVYTHNASRFDCTLCAYTAGSFASLKTHRNSRHRRTAFLDRFSAGCACGVPFASRLAAARHAQACASLSSAPLAEASSAAGASSHTVDGADSTVSAAGHAEPDLPRHNATELTASPPLVSSTDVEVQ

			ATGGTAAATATACAAAACCT	ATETEATENRWGTPLPRVLVASRIAGRLAQVPPPRWGPPLPRTTIAGRIA	
			TCTCTGTTCTAATCAGTGTG	TRLAATPAPRWSPPLPRSLVASRIAGRLLPALPDAPACEDEAKDSDEMD	
			AAAACTGGTTTTCGCCTTTT	WEASEPHVEAPGPVDEETIDDADGEWLLRFDGACRANPGPGGAGAAL	
			GGCGGACTTTTTCACTCGCA	FKPSGPVVWTCSHYDPSTTATNNTAEYTALLLGARAAADHGVTKLRVEG	
			TTTTTGGGCAATCGTCTGCG	DSTLVIQQVRGIFATRSTRLRALRNKVKLELARVGSFSLHHIDRQANGHA	
			GCTAGCTTGCTAGCGGCGG	DRLANAGLDRRRTQLECSVHPDGRGCTNTSVATAAPTASAAPSTPTRPP	
			ACGAGCGGTCTCCGGGGGC	ATTAAPFHSDQGHIDEDDERRADIDDGEIYAPMTLGPDEVPARRPRLRL	
			GTTCACCTTCCCCCGCGAG	RQLSDEELEAAGAIVERLSASLSAKITDAEDWGTAEGYITALPHLLYDKLLP	
			GCCAACTACACCGATCTTCT	YSRTAPRHQRPPPRPSRNQQDHPQPRRDQPQRNVDEQQHAESQQGED	
			CTACACTTTTCTAATTCGCCT	QRQQQPPTRRRRRRRGKRRGRRQRRHPRQPGQSTRASQQSRQRRPRPP	
			CCGTCTTCGGTCTTCGGTTG	RVTRHHREHRIDEALDELHTLERARPQDRSAIDKARRRVGRVRGAINQH	
			TCGGGCTTTTTTCTTTTTGAC	LLRHRFDTDEKACVADILEKAHAARAARTAQAAGAATSTGGAATAPTQ	
			CAATCAGAGCGCGCCATGC	QAATSALGDADDGTCPILADELWQYFTGTNTPRWEFNPATPVGEAFRT	
			GACTCTTCTGGCCAATCAGA	AMARLPPATRLRELLTEAPTADEIETQLQHVRGSSSPGLDGIGYDVYQRF	
			GACCGGCCCTGTCCTCGG	AQQLLPVLTASFKRCWTAKMVPQSWQVGVVRLLYKKGAHDDPANWR	
			ACAGCGAGGCCTCCACGGC	PICLQQAIYKLYTGVLARRLVRWLDVNDRHAPGQKGFRAVNGCGEHNF	
			CAGCCAATCAAGTCTCGGC	LAATLIDQARRKRRSLYEVWYDFRNAFGSVPFQLLWDSLQRLGAPADFI	
			AGCGACGCGTCTTTCTATAG	DMCKGLYHQAAFVIGNAADGPTAAIRQQVGVFQGCPLSPQLFNVAISPL	
			CGCAGCTGACGAGGCCGAT	LFALRRLPETGVQLSGDDRVGVSAYADDLKTFSSTKAGATKQHELVAAFL	
			CTGGCGGCCCCCGATTGGT	AWTGMKANAAKCSSMGVRRNSNGATEADNLDLALDGTPIPSMTHMQ	
			CCGACTTTCGGCCAATCAGC	SYTYLGIGDGFDHVHRRIELAPKLKTLKQDTTALLESGLAPWQVVKAVKV	
			GACGACGAGGGGGCAGGG	YLYPRVEYALRHLRPEDQLLESFDLHLRAGLRHLLRLPKNANNDFFYSPVS	
			GTTTACACTTTTGCCCCCGT	RGGLGLLPLVELHAALQIAHGWQMLNSTDPATRRIAREQLHQIADARHR	
			TTCGGCTTCAACTTCAGGCC	LDKAHWKERGDELCQLFLNLDLGTSAHAPPKRRNCDIGSLWVDVRKNL	
			AAAATGGCGATTTGGACCC	QAFGLKLETAPADAESGTPALPLQLRVPHHEKWLTHRDVLRHVKQHLKN	
			TCCACGCGCCGTGCCACTG	KHWRAWCAFQDQGRTARAHGGVGSSFITPRPGMWESDYRFAVAARL	
			CTCGGCACCGGCGGCGATT	NMVDTSATLARRRLRAHDRCRYPGCRWKESLEHVLNHCPGTMDAVRG	
			CAGCGGGTGCAACTTCGGG	RHDGVLREIEHALRAPSGARRELRVNQTVPGLPGPALRPDIQVYNHDQR	
			CACGTGTGCAACACATGCA	TVAVVDLAVAFDRQDRDDPETSGLAKAAAEKKAKYTGIQRHLERQGWK	
			GCGCCCATTGCACGCCAAG	VHLSALVYGSLGSVAPNNYKVYTEHLGLLKRDAKRLDRTLSVACIQSSRRI	
			CGGCATCGCGGGACGACGC	WNLHCAKHRARQHQTPSQSRGRRVTETGGAPSRTDRR (SEQ ID NO:	
			CTCGGCCGCTCAAGCGCAG	1446)	
			CCCCGCCCTTCCAGCACGAC		
			CTCGCGCCGTTTGGCGGAT		
			CGCCATCAAGACGTGCGAG		
			AGCCAGGCGCGGGTCGGGC		
			AAAATATACTTACTCTAAGT		
			ATGCCCGAATCCCTGCCCTC		

				TCAGGCTGTACGCGGCCCC ATACTTGACCTAAGTATGG GAGGATCCCTGGCCTCTCA GGCTGTACGCGRGGACCAA GTACAGCCGAATCCCTGGC CTCTCAGCCTGTACGCGGG GCTATACTTGGTCTAAGTAT GCCCCGGTCGCTGGCCTCT GAGCCCGTACGC (SEQ ID NO: 1324)	MQDQVDAEQQARNRWGPPLPRPLVASRVAARLGEVPPPRWGPPLPR GVVSRIAARLEAVPVPRGGPPLPRSFVATRIADRLAPPSPDLSLLDEEMK ESEPPDPTHHSADEDSTDAETADAVMEPAFVSDPPTATPREWRLQFDG ACRGGPNPGGAGALLYNPEGAVVWTGSHYMPGAKETNNSAEYTALLIG ARAAADHGARQLRIEGDSLLVIRQVKGLYATKSTRLRQLRNAVRHELARV GQHSLHHIDRQGNAFADRLANRALDLKSDKVECKEHPVAGACTTCMGS PSAGPPATPPPTTADIEMADAGSDDELRADIDDGEVYAPMRLEPGVIPT RRSRLRLRQLTDDEMEAAGEVVERLSAGLSAKIADADDWETAEGYITALP YMLYDKLQQYTQVRHGTARSPAPHPQRRDVQGQVETHREPRHETIGQ PDQPGEPSPTRRRRRGKRKGRRQRRHPRRTNCGGGGRQQRKQRHPRP PRGTRHHREHRIDEAIDELHALERARPQARPAIAKARRRVGRIRSAIDQQ LLRHRFDTAEKECVDGILAAARTARDARTTVRAAAATGTTATPETAVTSG TEQQDDNGTCPIPSEVLWRHFDSVNTPQRDFDPEAPEGAAFRSAMARL PAATRFMELLKEEPSTDGIEVQLQHASSTSSPGLDGVGYDVYKRFASQLL PVLKAAFKCCWTHKQVPQSWKLGVVRLLYKKGDREDPANWRPICLQQ AIYKIYTGVLARRLTRWQDANDRHAPGQKGFRPVNGCGEHNFLAAMLI DHARRKHRPLYEVWYDFRNAFGSVPLGLLWDALERTGVPAEYIAAVQG LYDHAAFMVGNAVDGSTAPILQRVGVFQGCPLSPPLFSAAISPLLHALQL LPSSGVQLSGDDRPGVSAYADDLKTFSGTKAGVTEQHELVAMFLRWTG MAANPAKCRSMGVRRNGNGAIEADHLELALDDTPIPTLTHLQSYTYLGI GDGFDHVRRRVALAPKLKLLKQDATALMESGLAPWQVVKAVKGYLYPR VEYALRHLRPDDQLLESFDLHLRRGLRHLLRLPKSANNDFVYAPVSRGGL GFLPLVELHAALQIAHGWQMINSPDPAIRRIAREQLHQVADARHRLDKD HWKQRGDELCELLLNGELGTSAHAPPKRRNGDIGSLWVDVRKNLKAFG LKLATAPADPESGAPAKPLQLCVPHHAEWLDHRNVLRHVKQHMKNKR WRAWCSHVDQGRTARAHGGVGSGFLTRPRGMWESDYRFAVAARLN MLDTVNVLARRRLRAHDRCRHPGCRWKETLAHVLNHCPGTMDSIRGR HDDALKEIERTLHASSGDRQGRTELRTNQTVPGLAGPALRPPDLQVYNHD
NeSL	Utopia-3_PR	.	Phytophthora ramorum	TGCGCGGCAGACCA AGACGCGCAGCCA ACAACAGACCGTGC AGCGTGCGGGTGG AAAGCGCCGCCGCC TGAACGCTGGTGAC GATGAAGACCAGC GAGAGCTGGCCGA GCTCCTGCTCGTGG ACGAGGACAAGGC TGGCGCCGAACACC CCGCGCTCAGGCTG CCCACGGCCAGCGC TCATCCGGCCTCCG TCCTCTCCGTGTAC GCGCACGCTGCAAC TCGCTTCGACTGCA CGCTGTGCACGTAC ACGGCTGCCAGCCT CGCTTCGCTCAAGC GCCACCGCTCGTCT CGGCACCGACGCAC GGCCTTCCTCGACA AGTTCTTGGCGGGC TGCGCGTGCGGCAC GCCCTTCGCATCGA GGTTGGCCGCAGCC AGACACGCGCAAG		

				CGTGCGCCAACCTC TGCACCACCTCGGC GACGACTTCGACGG CAGCAAAGGCATCA AGCCCCACTGCTGC CGGAGGCAGACCC ACCGTCCGTGCAGT GGTCACCGCCGCGC CCGACCTGCCCCGC CAGTATCCCTCGGA GCTCGTTGCGTCCC CCCCGCAGCCGAGC TCCACCAACGTTGC A (SEQ ID NO: 1202)	TAAACGGGTCACTTGACCG ACAGGGCACCACCCAGGTA GGGAACCGCCCTTTAAAAC CCAGGAAGACACAAACACC CTCCACATAGTGACATACAT ATTTTAGCCTAGATTTCAGT TACGGAGAGGTTACTAACT GGTACATAAAATTTACACATT CTGTTCTAATCAGTGTGAAA ACTGGTTTTCGCCTTTTGGC GGACTTTTTCACTCGCATTT TTGGGCAATCGTCTGCGGC TAGCTTGCTAGCGGCGGAC GAGCGGTCTCCGGGGGCGT ACACCTTTCCCCCGCGAGGC CAACTACACCGATCTTCTCT ACACTTTTCTAATTCGCCTTC GTCTTCGGTCTTCGGCTGTC GGATTTTTTTCTTTTTGACCA ATCAGAGTGCGCCATGCGA CTCTTCTGGCCAATCAGAGA CCGGGCCCTGTCCTCGGAC AGCGAGGCCTCCACGGCCA	QRTVAVVDLAIAFDEQPRDDPESSGLAKAAAEKKAKYAGIKRHLERQGW KVHLSALVYGSLGSVAPSNYKVYTEHLGLLKRDAKRLDRQLSVACIQSSRR IWNLHCAQHRARQHQDQPAPRGRRVTETGGTPSRTDRR (SEQ ID NO: 1447)
NeSL	Utopia-4_PI	AATU0100121 81.1	Phytophthora infestans	AGCTCGGCCTCGCG GCTGCCTTCCCAGG CGCCGCCGACTTCG CGCTCTGGCGCGGC CCACACGCCCGCCGC CGAGCCTCCAAGCG CGCCCGTTGGCTTT CGCAGACGCAGGG CTCGCGGCGACGGC CCTCGCCAGCCCCC AAGACCCCCCCTAC GATGTGGCACCACC CGGCAGGGCGGCC GGCAGGCTGCCCG ACTCGGTATCGCCG GGTGCTACACTCTC AGCCGCTACAGCTC GGGCCTTGGCGGTC CGCTATTGGCCCTT GGAGCTCGACGGC GACAGCAGCGACG ACGAGGACGCTCAA GACCCCCACGCCGC		

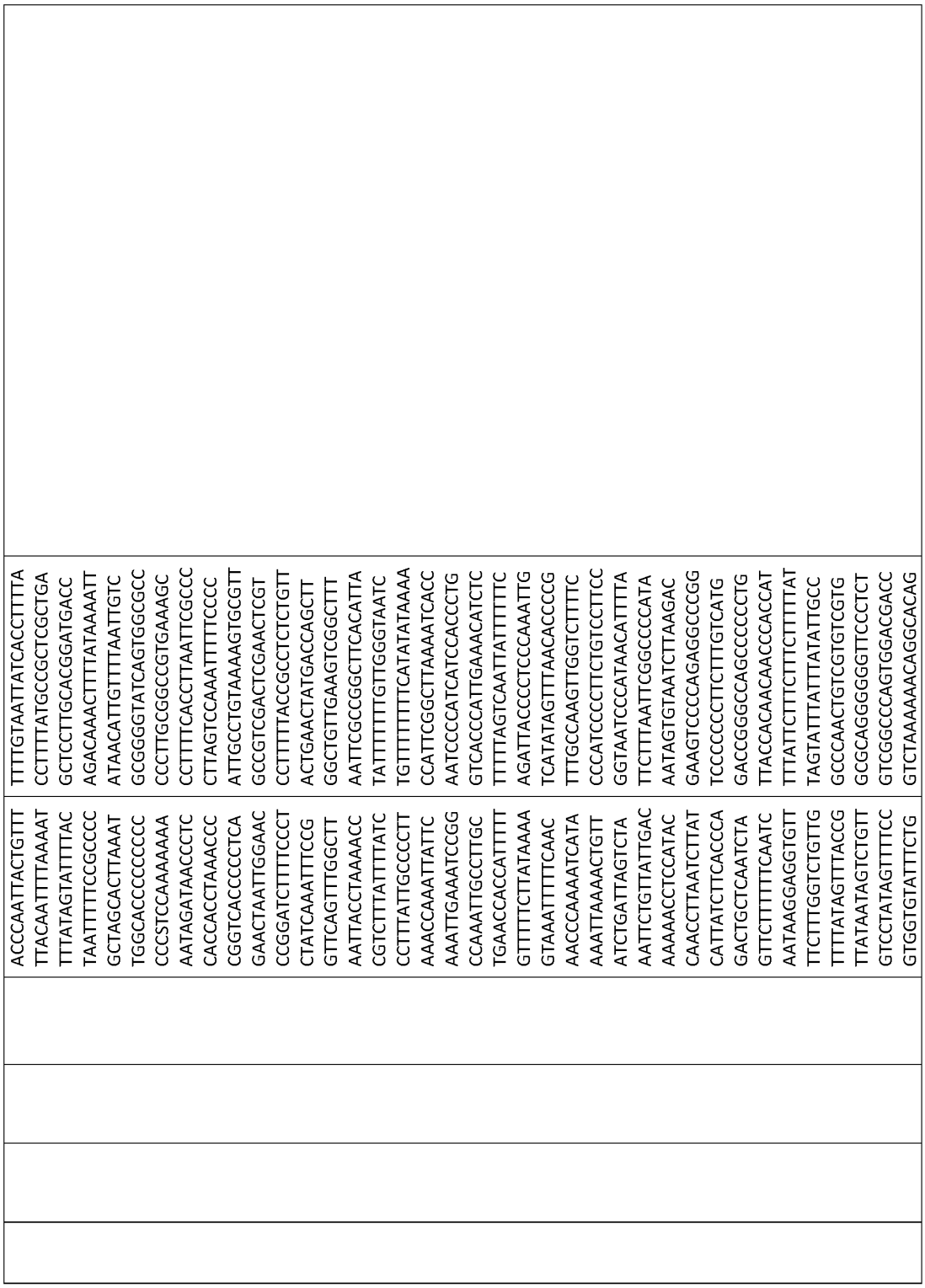
			CGCCCCAGAACCCC CAGAAGACGTCGC GAGTGTGCTTGCCC CACCCGGCAGGGC AGGCAGC (SEQ ID NO: 1203)	GCCAATCAAGTCTCGGCAG CGACGCGTCTTTCTATAGCG CAGCTGACGAGGCCGATCT GGCGGCCCCCGATTGGTCC GACTTTCGGCCAATCAGCG ACGACGAGGGGGCAGGAG TTTACACTTTGCCCCCGTTT CGACATCAACTTCAGGCCA AAATGGCGATTTCGACCCTC CACGCGCCGTGCCACTGCT CGGCACCGGCGGCGATTCA GCGGGTGCAACTTCGGGCA CGTGTGCAACACATGCAGC GCCCATTGCACGCCAAGCG GCATCGCGGGACGACGCCT CGGCCGCTCAAGCGCAGCC CCGCCCTTCCAGCACGACCT CGCGCCGTTTGGCGGATCG CCATCAAGACGTGCGAGAG CCAGGCGGGGTCGGGCAA AATATACTTACTCTAAGTAT GCCCGAATCCCTGCCCTCTC AGGCTGTACGCGGCCCCAT ACTTGATCTAAGTATGGGA GGATCCCTGGCCTCTCAGG CTGTACGCGAGACCCGTAC GGCCGAATCCCTGGCTTCTC AGCCTGTACGCGGGGCTAT ACTTGGTCTAAGTATGCCCC GGTCGCTGGCCTCTGAGCC CGTACACA (SEQ ID NO: 1326)	TAYADDIKVFCDSATGIQRCHALVTRFLEWTGLQANPAKCAFLPVTRSQ HSNPTRDRDIELRIHGEAIATLGLQESYRYLGVGDGFDHVRHRLQLEPKLK QIKREAVALLHSELVPWQILKALKVYIYPKVEYAYALRHLRPLKSQLQGFDSAI VRGLRHLLRLPENSHDGLFFSPTSAGGLGLLSLVELHEALQVAHAWQML HSKDPAIRAIARTQVGQVARKRFKLVEEHWRGREDDLAQRFLNTELAAS PHATETRRNGDIGSLWNDVRDTLQTLGLKFAAGDEEAAPGLLQLRVPHH TKWLSHSTVLRHVKLHMKLRRMDTWKSKVSQGTTVREHGGVGSRFITA GAGLSDAEYRFAIAPRAHLIDTNSTLKRRRLRANDTCRAPGCSYTEPPAHI LNKCSPNMDAIRKRHDDALERIADALRRKVEKSGGRLEVAINKTVPEYDG AALRPDIVLRNTETKRAIIADLAITHENQPTDATTSSALQQSRDNKITKYQT VAAAMMRAGWRVRVTGIVYGSLGSVLPSNFKVYTELLALLKRDARRLNR QLSSHCIRASARIWSAHCRRHRERQRSGNASRASRGSGGAPRRTSQASA RR (SEQ ID NO: 1448)	
NeSL	Utopia-4_PR	.	Phytophthora ramorum	CCAGGAATCACCCC CGCCGCCCCCAAGC GCTCCGCCGCCGAG CCCAGCTGCGGCTG CTGCCGCCCCACTG GTGACGGCAGTCGC	TAGACGGCACACAGGCCCA CAGCGGCCGACAGGGCCAC ACCCAGGTAGGGAACCGCC TTCAAACCCGGCCGGTACA TTATGGTCCGACACCTATGA GGTGCAACCGGTACACAAG	MSGDVVSSDGSSRTTDASGDGDDGAGSSDAAGDVGVVAMDVDQGA RRQQPPWQRVGGKRRRLNDVDDEDTRELAELLLEEEDEAGDHAPAPRL SAASARPASVLSVYAHNAQRFQCTLCTYTAASFASLKRHRDSRHRRTAFL DRFSAGCACGVPFASRLAAANHAHACDSLNRTFSVAAAPAAGELSPTAG AANATVKAATVTPDSPRQDPPKLAATPPLASSALVVDPDHAEQQARER WGPPLPRTLVAGRVAARLSEVPAPRWGPPLPRGVVAFRIGHRVLPPEM

			AGCTGCCCCTGCCA CCTCTCCTGTTGGC CGCGCCGCCGTGGA GCCCCGCGCGCGCG CCGAGCCGCCCCAA GAACAAGCACCCCC AGCTAGCGCGCGC GTGGAGCCC (SEQ ID NO: 1204)	TTACACACCACATAGCGACT ACCAGGTATTTACTACCTGG AAGCCAAGGATTAACCGGT CGGTAATACACATAACTTT (SEQ ID NO: 1327)	TSDEETKDDSSVQDGDRQDYPVAAMDVDSGMSGEWLLRFDGACRAN PGPGGAGAALSQPDGSVVWTCSHYMPSSSETNNTAEYTALLLGTRAAA DHGTTTLRVEGDSTLVIQQVRGIFATRSVTLRHLRDQVKLELARVGRFSLH HIDRQANAHADRLANRALDLRRTVSECGVHPDGNGCTPTAIDDRPLAPT QQPPDAPPPPPAADIEMEDPDDEDLADIDDGEVYAAMRVGPNATPQR RRRGRSGTAKKHRRQRPPRVTRHHREHRLDEALDDLHAVERSTPSDRTT VRRARRRVGRVNSAIEQQRLRHRFDTDFDEKACVTDILAKACATREAARTT ASGGDPPAGPATPAAGSADDGTCPILGEELWRFFDSVNTPRQEFAPDAP VGAAFRSALARLPAATSCKELLTAAPSAGEVEDQLQHVRGASSPGLDGV GYDVYQHFAAQLLPALTAAFKACWTAKRVPQSWKLGVVRLLHKKGARE DPANWRPICLQQAIYKLYTGLLARRLVRWLDANDRHAPGQKGFRAVNG CGEHNFLAATLVDQARRKRRTLFEVWYDFRNAFGSVPFALLWDALARL GVPDDYVTMCKGLYESAAFVVGNAIDGTTDPIALRVGVFQGCPLSPQLF NAAISPLLFALQRLPATGVQLSGDDCPGASAYADDLKIFSGTEDGIKRQH ALVADFLRWTGMAANPNKCCTMSVQRDGRGVLKTDDLQLDLAGTPIP ALSMSASYTYLGIGDGFDHVRRRVELAPALKQLKDDATTLLQSGLAPWQ VVKAVKTYLYPRVEYALRHLRPFQQQLEGFDRHLARGLRHLLRLPGNATA ECFYAPVSRGGLGLLPLTELHAALQVAHGWQLLNSKDPAIRRIARVQLRQ IADARHRIDSRAWEGRDEELCELLLNSQLGTSPDAPPKRRNGDIGSLWV DVQRHLRTLGLKFATAPACADAGSAATTLQLRVPHHDKWLDHRTVLRH VKLHVKHRHWSKWAAMRDQGKTARAHGGAGSGFLTRPRGMWEADY RFAVAARLNQLDTHSVLKRRRLRAHDHCRQPGCSRAETLAHVLNHCAG TMDAVRGRHDDALKHIERALHASSPGGQDRVELRVNQTVPSLAGPALR PDLQLYNHTKKMVAVVDLAVAFEEQASDDPESSALARIAAHKRAKYAGV KRHLERQGWKVHLSALVYGSLGAVPAGNHKVLTEHLGLLKRDAKRLDR QLSVACIQSSRRIWNLHCSQHRARQHQAPGGSRAAETGGTPPRTGRR (SEQ ID NO: 1449)	MARVLVEPMQVDECSSCDRSTLTADDGSGDDVAAPSSLNSNDVAAPM DVDSGTDCPPALQQPPQRPRALHVGSKRRRLDADDEEEARQLQEEEEA GIHAPALRLSAASAQPASVLAVYTHNASRFDCTLYAYTAGSFASLKTHRN SRHRRTAFLDRFSAGCACGVPFASRLAAARHAQACASLSSAPLAEASSAA GASSHTVDEADSTVSAAGHTEPDLPRHNATELTASPPLVSSPDVEVQAPE TEATENRWGTPLPRVLVASRIAGRLAQVPPPRWGPPLPRTTIAGRIATRL AATPAPRWDPPPLPRSLVVSRIAARLLPALPDAPACEEEEAKDSDTMDWAP TWTNEETKDSEPHDEAPGQVDEETIDDADGEWLLRFDGACRANPGPG GAGAALFKPSGPVVWTCSHYDPSTTATNNTAEYTALLLGARAAADHGV TKLRVEGDSTLVIQQVRGIFATRSTRLRALRNKVKLELARVGSFSLHHIDR QANGHADRLANAGLDRRRTKLECSVHPDGRGCTNTSVATAAPTAPAAP
NeSL	Utopia-5_PI	.	Phytophthora infestans	CTCAAGCCTAGCAG CGGCTACCGCAGCT ACTCCAGCTCCGGT AGCTGCTACTGCTA CAACTGCTCGCGCT GCTGCTCGCGTCGC CGTGGAGCACCACG CGCACGCTGAACCG AACCAAGAACATCT ACCG (SEQ ID NO: 1205)	TGACGCACCGTGACATAGT GCGGCACGTGAAGATGCAC ATGAAGCTCCGACACTGGG CCAAGTGGGCGGCCATGCG CGACCAAGGCAAGACAGCT CGTGCACATGGTGGGGTTG GTAGTGGCTTCCTCACACG GCCGCGAGGCCTGTGGGAA GCCGACTACCGGTTCGCGG TGGCCGGCCGCTTAAACCA GGTAGACACGCACAGTGTC	

			CTCAAGCGCCGGCGCCTCC	LPPARPPATTAAPSHDDDHSVQGDIDDGEVYAAMCIGPDAVPHRRPRL
			GAGCACATGACAGGTGCAG	RLRHLTDEESEEAGDVVERLAASLAAKIADAPDWETAEGYITALPYALYDK
			ACACCCAGGATGCACGCGC	LQPYSQAQPQPPSQQQQQQQQQRPRQQQQTRQRRQRRGKRGGGSQR
			TCCGAGACGCTGGCGCATG	RQRKTRRRRPPRVTRHHREHRIDEALDDLHAIESRRPQDRTAISKARRRV
			TGCTTAACCACTGCGACGG	GRIRSALDQHQLRHRFDTDEKVCVDAILAGARASQGATTAPPSATTDPP
			AACCATGGACGCAGTCCGT	APMDDSRCPIPGDDLWRFFDSVNTPRRSFDAEAPDGAAFREAMACLPA
			GGCCGGCCATGACGCCGCA	ATRAQELLTEAPTVDEVEDQIQHARASSSPGLDGVGYDIYKQFAAHLLPA
			CTCAAGATTATTGAGCGTG	LHTAFVCCWNHKRVPQSWKLGVVRLLHKKGDRQDPANWRPICLQQTI
			CGCTCCTCGCATCGTCGGCC	YKLYAGILSRRFVRWLDANARHAEAQKGFRAMNGCGEHNFLAATLVDH
			GACCAGCAGGACCGTGCTG	ARRKRKELHVVWYDLANAFGSVPHDLLWETLARQGVPPTFVDCCRGIYS
			AGCTCCGCGTGAACCAGAC	DAAFTIGNAADGTTAPIRLRVGVFQGCPLSPHLFTAAIAPLLHALKRLPVT
			CGTGCCGTCACTCGCCGGC	GVQLTGVDRPGAAAYADDLKTFSSSVDGIKRQHELVATFLRWTGMAAN
			CCCGCGCTACGGCCCGACC	LSKCSAMSVQRDSRGVLKTGDLCLKLNDAEIPALSMTASYAYLGIGDGFD
			TTCAGCTCTACAACCACACC	HVRRRLELAPMMKQLKHDATALMQSGLASWQVVKAVKVYLYPRIEYAL
			AAGAAGACGGTGGCGGTG	RHLRPFKQQLEAFDEHLRRGLRHLLRLPTNATSAFFSAPTSRGGLGLLPLT
			GTCGACCTGGCCGTGGCGT	ELHAALQIAHGWQILNSPDGATQRIAREQLREIPDARHRLDTAHWRNR
			TGAGGAGCAGGCGAGTGA	DAELCELLLNSQLGQSSHAPPKRRNCDIGSLWIDIRRQLGTLGLKFETAPG
			CGACGCGAGTAGCTCGGCA	RRSHQPARPAIAAFACRTTTSG (SEQ ID NO: 1450)
			CTGTCCCGGATCGCCAACC	
			ACAAGCGAGCCAAGTACGA	
			CCACATCAAGCTACACCTCG	
			AGCGCCAAGGATGGAAGGT	
			ACACCTCTCGGCACTCGTGT	
			ACGGGTCGCTTGGGGCGGT	
			CGCTAGTGGCAACTACCAG	
			GTGTACACCACACACCTGG	
			GGCTACTCAAGCGCGATGC	
			AAAGCGGCTGGACCGGCAG	
			CTGTCTGCCTAATGCATCCA	
			GTCCAGCCGCCGCATCTAG	
			AATCTACACTGCAGCCAGC	
			ACCGGACTCGCCAACACCA	
			GGCGAGGCCCAGCCAAGG	
			ACCAAGAGGCAGCCGGGC	
			GACGGAGACCGGGGGGAC	
			TCCGTCCCAGACAAGCCGC	
			CGCTAGGCGGAAACCAGGC	
			CCAAGACGGCCGACAGGGC	

					CCCACCCAGGTAGGGAACC GCCCTAGAAACCCATTTCG GTGGTCGACTCGCAAGCCT TACCTATATTTTAGACGTAG CGACCAAATTACAAATTTGG TAACGAGTAAGCCAAATGG TAATACACAAAACTTT (SEQ ID NO: 1328)	MSGDVVSSDGSSRTTDVSGDGDDGADGAGSSDAAGDVGVVAMDVD QGARRQRPPWQRVGGKRRRLNDVDDEDTRELAELLLEEEDEVGAQAP ALRLFAASAHPASVLSVVAHNAQRFVCTLCAYTAASFASLKRHRDSRHRR VSFVDKFSAGCACGTPFGSRLAAARHAQACASLSIPRTVTAPAAAGDLSP TATGANATASAAATSPDLPRPASPELAASPPQTSPFDVAIQADAAEQTA WTRWDPPLTRAAVAARVASRLAVVPAPRWGPPLSRTLVASRIAARLDA QTSRWGPPLPPAMVASRIASRLAAMPAPRWGPPLLRTVIASRIADRLLP PELAADEETKDDDVHMDNAASVDVDEESEVADVVMTDHDGEWLLRFD GACRANPGPGGAGAALFKPSGPVVWACSRYMPSSSATNNTAEYTALLL GARAAADHGATHLRVEGDSTLVIQQVRGIFAARSTRMRALRNQVQSEL ARVGSFSLHHIDRQDNAHADRLANRALDLRRTVIECGIHCDGVGCTATT TEVQSSSAPEIPTRPVADDHDEHEVVDVVDVCGVCGVCGDRGTCGVCD VSGDIDDGEVYAAMRTGPDAVPARRPRLRLRKLTDEEQEEAGTLAERLG ATLAAKIADARDWESAEGYITALPYLLYDKLLPYSQGPARSLPVRQHQRQ QQQPDGQFQRPTQSRSAARRQRRQRHRARRRPPRVTRHHREHRLDEA LDDLHAVERATPSDRRSIRRARRRVGRVNSAVEQQRLRHHFDTNEKGCV EILLAKARAQRSTTVARTAVGEPNSGAAEDDGTCPIPSERLHRHFTEVNT PGSSFDAMAPVGAPFRAALAHLPAATEASELLTEAPTPDEIEDQLQRAKG TTSPGLDGVGYDVVKAFSTQLLPVLHAAFQCCWQHHRVPQSRKQGIVR LLYKKGPREDPANWRPICLQQVIYKTYAGVLARRFTRWLAANGRHADA QKGFRTVNGCGEHNFLASTLIDHARRSRRELHMVWYDLKNAFGSVPQE LLWEVLQRMGVPPAFVEVCQGLYQDAAFTVGNAADGPTDLVRQLVGV FQGCPLSPHLFTAAISPLLHALDRLKDTGVRLSADDRPGASAYADDLKIFS GTADGVKRQHALVADFLRWTGMVANPNKCCTMSVQRDGRGVLKACD LELQLDGARIPSLIMNASYAYLGTGDGFDHVRRRIELVPALMQLKDDATA LLQSGLAPCQVVKAVKTYLFPRVEYALRHLRPFQQQLEGFNRHLVCGLRH LLRLPVSATTSFFFVPVSRGGLGLLPLTELHAAPADRRPAPPPPRPPRPLEGA GGENMRAADQLAARDVGPRPTQAPQRRHRLVVGRRPAPPPRTRPQAR DRAGVRGDRHRGGDAAASRAAPREVAGPPHGPAAGEAAHEEQALAAV GRDEGPRQDRPHPWWCRERLPHAASRPMGDRLPLCSGGSAQPAGHA
		Phytophthora ramorum	CCAGGCATCACCCC CGCCGCCCCCAAGC GCTCCGCCGCCGAG CGCAGCTGCGGCTG CTGCCGCCCCACTG GTGACGGCAGTCGC AGCTGCCCCTGCCA CCTCTCCTGTTGGC CGCGCCGCCGTGGA GCCCCGCGCGCGCG CCGAGCCGCCCCAA GAACAAGCACCCCC AGCTAGCGCGCGC GCGGAGCCC (SEQ ID NO: 1206)	TGATGCCCGCCAACTACAA GGTGCTTACTGAGCACCTT GGGCTGCTCAAGCGTGATG CGAAGCGGCTGGACCGGCA GCTGTCGGTGGCGTGCATC CAGTCCAGCCGCCGCATCT GGAACCTGCACTGCGCGCA GCATCGAGCACGGCAGCAC CAGGGCCAAGCGCCAAGG GGCAGTCGGGCGGCGGAG ACCGGGGGGACTCCGCCAC AGAGCGGCCGCCGCTAAGC GGACATCGGGCCCGTAGCG GCCGACAGGGCCACACCCA TGTAGGGAACCGCCCTCTA AACCCGCCCGGTACATTAT GGTCCGACACCTATGAGGT GCAACCGGTACACAAGTTA CACACCACATAGCGACTACC AGGTATTTACTACCTGGAA GCCAAGGATTAACCGGTCG GTAATACACATAACTTT (SEQ ID NO: 1329)	
		.				
	Utopia-5_PR					
NeSL						

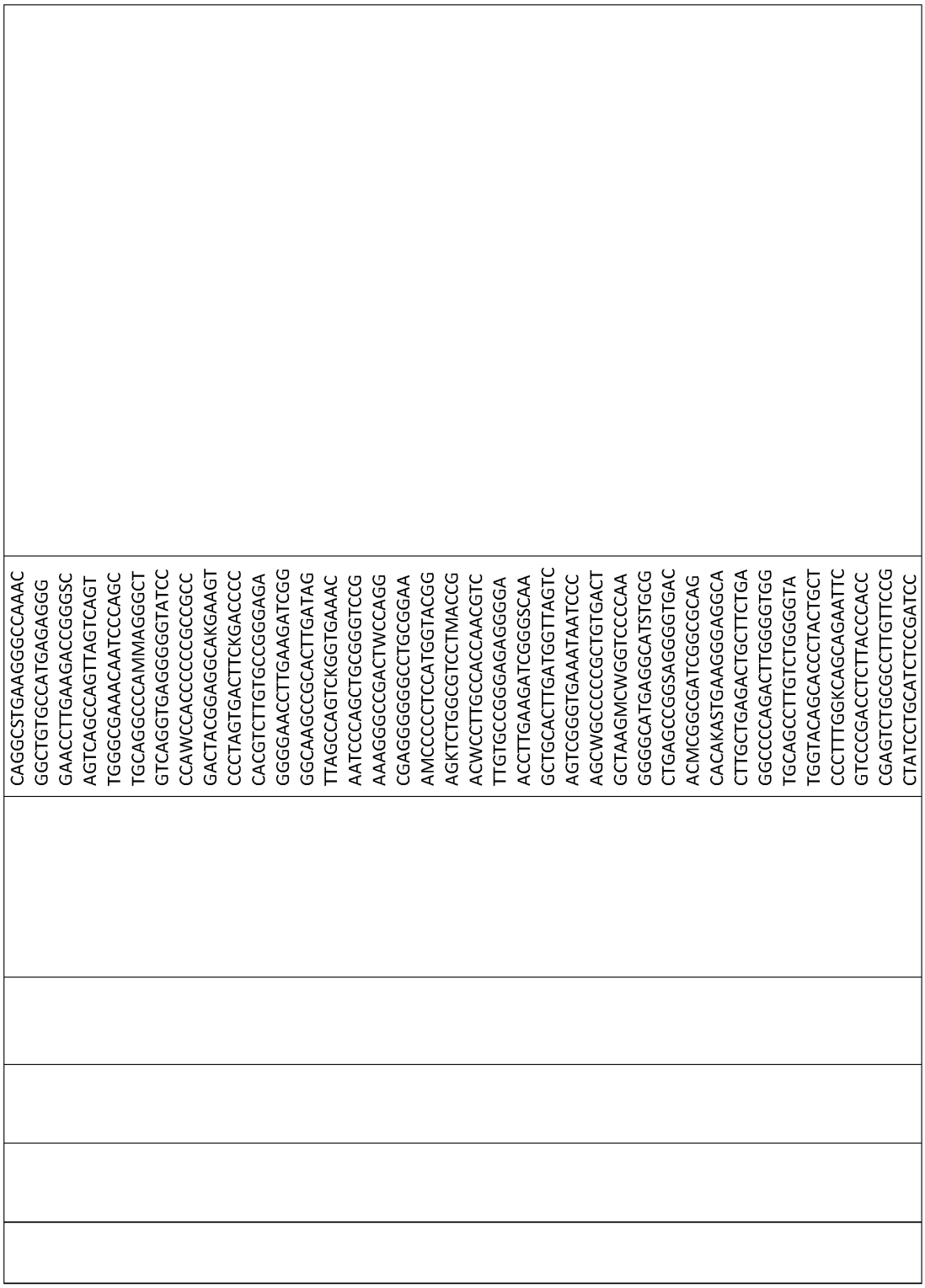
						QRAEAPAPPRARPVSTAGLLPCRDTGTRAESLRRHHGRGPRPPRRCAQD
						HRARAHRVVRAARTAPSSGSTRPCPRSPAPRCGPTSSSSTTPRRRWRWS
						TWPWRSRSRRATTPRALRWRASPRTSEPSTPASSGTSSAKGGRSTSRRS
						CTARWAR (SEQ ID NO: 1451)
NeSL	YURECi	.	Ciona intestinalis		TAGCCAAAAGATTTGGTGT	MATSSSSVSSGNVQTEVRCVYHGKGDLLLECPVAHCPSIHPTVATITKHL
					TCTTGGCGAGCTGCAGTCC	KKHHTPQFEQITTKNLTITYTCSQCSFSTTGLTQHHISKHYKTCKGVGAVQ
					CGGCAGAGTAGGCGAGATC	EGNKGRFCCPACGTRWALLCKARHHFNNVHFEYDTPPIAAFSGTPYKLK
					GTGAGCTCTACACCGCTTCA	KRKFTIINKALTYSCPLPLNQLLCPLWSCSLTILNKPLSSVQQETAHGDGSQ
					CGATTTGACGAGTTCTTCTT	GQSYVPTQLRQVLRARCHCGNPPIGKGHWASCQGKRPLSSPKGGRSSP
					GAGTCGATCCCCAACGGCA	TPPANLTLHFLNYLPFQLPSQSSSPQSSTLDPTACKARVPIPSFLRGDCEVT
					ACGCACTCCAAATTTAACGA	FFIIPSVNFYRPYLSYPLRMFWRNRTSSGHCSLHRVVRGFVRERLVPHPRS
					CCGCATGGCGACGCCACTG	KSPARTPLEFLCEFRLAGVFPDPGKVASLRPVPAPLTLCLSPPVAGPMISC
					ATCTCCCAGTCATCCGTGCT	EDHSAPPSVRSSSPIPNSPASVSSVEAHLSDLLDKVSSGELRPLSPTLPSSG
					CCGAGGGCACTTAAGGGGA	FFGPLLPPTPPPRPTPSAEKASPSGLSYLPCREVKIASIARPSPASQRVGCD
					ACCGATTCACTGATGCTGCA	ADRTGPSLNPNYQQTSPPSTPSFSPIVRPPKFPRSGAKVNSKSKPPGVRP
					AGTTGGCGGGGGGCGCACTC	RRAKPIEPGTESASPVDVDTISSSVQEPCTPENRTPEFFYERKWLVSILNIH
					ACCAATGGAAGATGAGGGA	EREGSNFFQFNRDLEYWTQLLSGSQKGGRAKRASYNRGAANQAMKNR
					GCCCAACGATGCCCCGAAC	DSGRKDFDPRPVAGHSSGGGTELGSRPRYPKGARLRADFWRDMKGTV
					TACCGCACCGTCCCCAGCAC	RKLLDGSNGERRCGIPLDIIERKFRQVSMPGWIDHRRYAAGASPSLVTQA
					CGTGTCCTTGTGAGTAATTA	ETDVAITSEEVEAVLSGLNVQSAPGSDGLSYRFWKGLDPSGRLLSCLFEIV
					TTTTCACTAATCCAGGGACG	RRHGRIPGAWPTCSVILLCKDAQGDVQDVGNWRPITICRTLYKLYAAVIA
					GGGCCATGAACTGAACTAT	RRIQTWAKQGGVLSRLQKGFMPVEGVFEHVFMLDTVLSDAKLRRKNLL
					GTCTGCACGCTGCCCCGCCT	AVFLDVRNAFGSVRHECLLKVLRHFDAPHYLVELVRDIYTGATCRVRSSV
					AGCCTCGGCCACAAATAAA	GETGDIPCDRGVRQGYPLSGILFNLVTEVLIPGLSAGNDGYRMACLGGKL
					TAGTCAAGCCGGGGCGTAC	TQVLAYADDLVVVTENRDQMLRQLGVCEEFGRWAGLAFNQRKCGLIG
					TACCAATTGATTCGGATTTG	WRTLRGGRRVALEDPLLLNGVEIPLLRPGEHYKYLGAMTGVMSVPRTGS
					GCGGGGCCCATCACGCGCC	QLIKDFRARLQRLFTSFLTPHQKLIALKRFLLPSLSFHLRVRPIARSELIALDR
					TCCGCCCCCACACACAAAAA	RVRECLRVAFRLTKPSCQAVFHTPTDMDGLGVPSVCSESSILTIAQGFKVL
					CACTTTTAAACTCTTCGGTT	TSPDGTVSATASARVKLYAAKFGGLTEAGPSDWARYLSGDDVNGNSTRK
					CCCCAACCACTACAACAAA	PGANLPSGLWTRVRCASRQLGAVWRVCPENGITVRVRNSVITSRDRRKL
					GCGAGCGGCGCCCTCTTAGA	IRSFHDCSNQQWKEQWMQHPNQEKTAAAHMAYADANRWVKQPSV
					TCCAATTTTAAAATTTTAAA	MEPHTYFFALRARLNLLPTRVSRAIYSRDQHPDILCRRCGASVESLPHVLN
					TCAGTGCACTCAACTTTTTT	HCPPNMSIILGRHNLVLQEVLNAVDKTQFKEISVDRTVPEHMSETGEALR
					ACGTGTTGTGTTTTTTGTAT	PDIVARRNDGSVVVVDVACPFDQKANFDEAAKRKLLKYDKLCCNIAAST
					TTTTTCCCACAGATTTTGTAT	GKPVECHSIVVGSLGSLAEGLSTSLRALGITDFARSKLVACHQG (SEQ ID
					TTTTATATTATATTTTATATA	NO: 1452)
					CACAACACTATTTTTATACA	
					CTACCTTGCACTGTCCCACT	



			GTCTACTATAGTTG GTGCCTCTAAAACA TTATTATACCGAAG AGTGCTAGCGATTT CTATTATTAACTCAC CCACCAGTACTGCT GTGCATCCACGCAG CACCACGCCTGCCA TCACGTACAGTGCG CTTGCTGCCTTGTG TTTTGTTTTGCAGAC ACCACTGGACGATA G (SEQ ID NO: 1207)	GCAGCAGGCACTTATCACC CGGTACCTCCGGGTACCGG AGGCTGGTTTTCAGCGCAC TGTACGTGATGGCCAATTTT ATTCATTGCATTTTATCCGC GTCGTGGTGTTTGCGTGGA TGCATTAATAAAAGATGAA ATTCC (SEQ ID NO: 1330)	
R2	PERER E-9	BN000 800	Schistos oma manson i	ATCTCACGTTTTAAT TTATTTTTGAACTAC TGCAGTCTGAGTGC TTCTAACGACCCGA AGGCTCAGAAACTA CCCACTTCTTGAACT GCTACTTTTTGCTGT TTATCCACAACAAC AGTTGTGATTCTAT TCTCCANATATTCCT TGTGCTTTTGTCAA CATTATTCTATACCA ACTGTACCACCTAC TTCTTCATCTCACGT TTTAATTCTGGTCTT ATTTTCTCATCATTA GTCACGGAGAGGG CCTATGAACGGTCC GTGACGCGAAATTC TATCCGCGATTTCG ACCTCTCCTGCTAGT GGTCCCCGAAGTAC GGTTCCTCTGGCCT GTCAGTTGTGTTAA	TAACGGCTGAACGAATAGC CCCCTTCACTCTTAGACATT CCCCCACTGTTGTTGCTTAT CTTCATGCTCTTGTGTTAAT TGACTGCTCTCTTCTGGGTT GACGTCTGATTGTCTCTCTC TCTTTCCATATTGCTTGCTCT GCCCGCTTACTTCCAATAGT TGTCATATTATGTCTTTGTTT ACTTGCCATGTCTAACGACA ATTACTTTATCTACCTTAGTT TGTCCTCTTGGTTTCGATTG CCTTCATATGTTCATGGCGG AATCTGATGTTTATAATGAC TATTCCTATTACCACCACTA CAACTACTATTATTATTTTCA TTACTATTAACATTATTATA AACATTATTACTATTATTATT ATTACTATTATTACTTCTACA ATTAATATTATGGCTACTCC TCTCAGCACACCAATAAAAT ATCAATCAAACATCTCAATT ATATCCACCTATTAAACTCT CTCTATTTCCCCTGAGTTAT	MPVSTGAETDITSSLPIPASSIVSPNYTLPDSSSTCLICFAIFPTHNILLSHAT AIHHISCPPTPVQDGSQQMSCVLCAAAFSSNRGLTQHIRHRHISEYNELIR QRIAVQPTSRIWSPFDDASLLSIANHEAHRFPTKNDLCQHISTILTRRTAE AVKRRLLHLQWSRSPTAITTSSNNHTITDIPNTEARYIFPVDLDEHPPLSD ATTPNASTHPLPELLVILTPLPSPTRLQNISESQTSHESNKNSMHTPPTYAC DPDETLGATPSSTIPSCFHSYQDPLAEQRGKLLRASASLLQSSCTRIRSSSLL AFLQNESTLMDEEHVSTFLNSHAEFVFPRTWTPSRPKHPSHAPANVSRK KRRKIEYAHIQRLFHHRPKDASNTVLDGRWRNPYVANHSMIPDFDCFW TTVFTKTNSPDSREITPIIPMTPSLIDPILPSDVTWALKEMHGTAGGIDRLT SYDLMRFGKNGLAGYLNMLLALAYLPTNLSTARVTFVPKSSSPVSPEDFR PISVAPVATRCLHKILAKRWMPLFPQERLQFAFLNRDGCFEAVNLLHSVI RHVHTRHTGASFALLDISRAFDTVSHDSIIRAAKRYGAPELLCRYLNNYYR RSTSCVNRTELHPTCGVKQGDPLSPLLFIMVLDEVLEGLDPMTHLTVDGE SLNYIAYADDLVVFAPNAELLQRKLDRISILLHEAGWSVNPEKSRTLDLISG GHSKITALSQTEFTIAGMRIPPLSAADTFDYLGIKFNFKGRCPVAHIDLLNN YLTEISCAPLKPQQRMKILKDNLLPRLLYPLTLGIVHLKTLKSMDRNIHTAIR KWLRLPSDTPLAYFHSPVAAGGLGILHLSSSVPFHRRKRLETLLSSPNRLLH KLPTSPTLASYSHLSQLPVRIGHETVTSREEASNSWVRRLHSSCDGKGLLL APLSTESHAWLRYPQSIFPSVYINAVKLRGGLLSTKVRRSRGGRVTNGLN CRGGCAHHETIHHILQHCALTHDIRCKRHNELCNLVAKKLRRQKIHFLQE PCIPLEKTYCKPDFIIIRDSIAYVLDVTVSDDGNTHASRLLKISKYGNERTVA SIKRFLTSSGYIITSVRQTPVVLTFRGILDRASSQSLRRLCFSSRDLGDLCLSA IQGSIKIYNTYMRGT (SEQ ID NO: 1453)

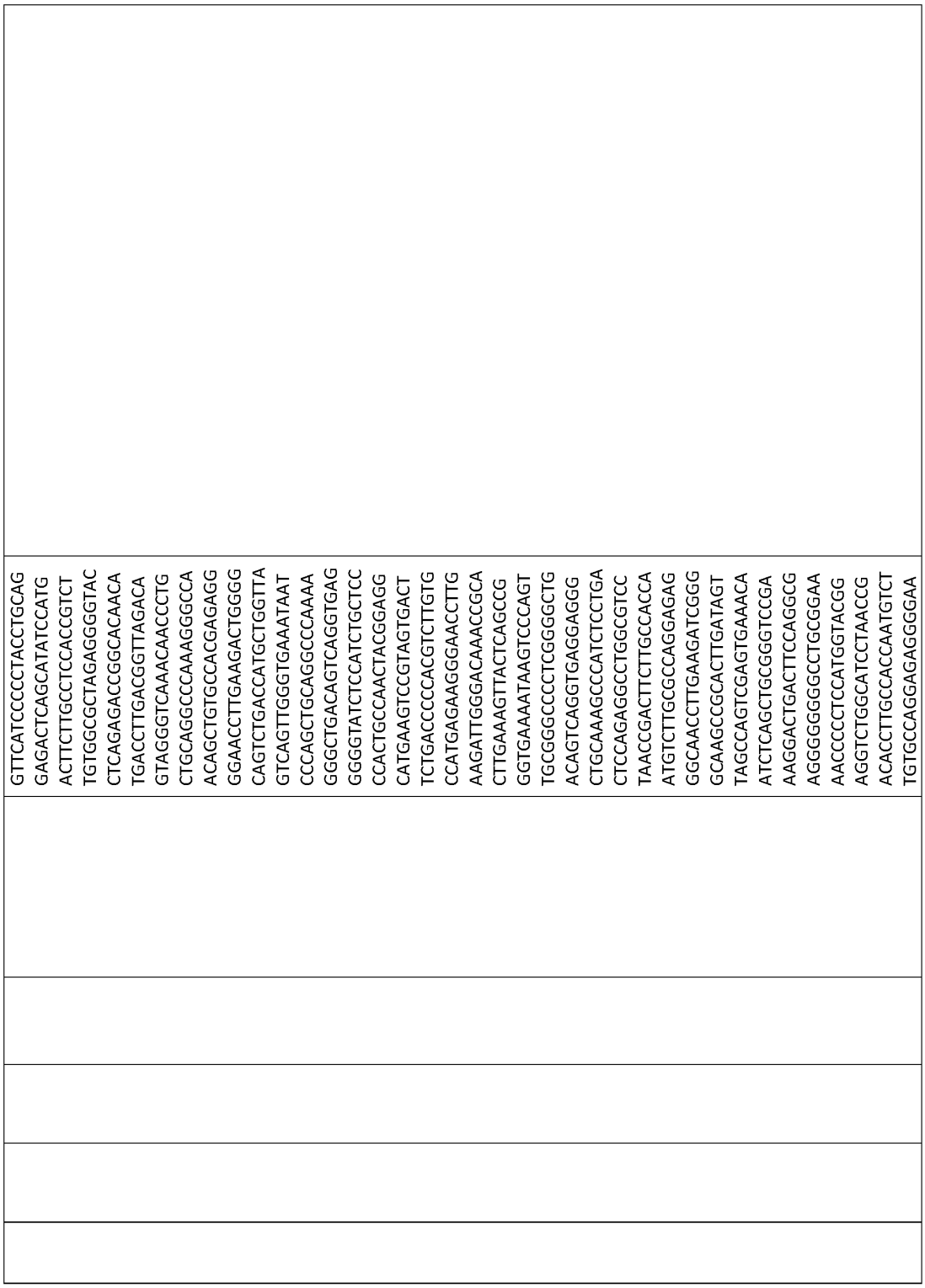
			AACTATATAATAACG (SEQ ID NO: 1208)	AAACTTACAATTCAGTCTAACCGAATATCTCTCTTTTACAAATCTTAAGTATGTAATTTTGTGCCAAACCCATTTGGGTCTGTACAATTTGATACTTAAAAATAAATGTTATTAGCC (SEQ ID NO: 1331)	IAKFDNNTNSASDAAPLSPGGAVADLSASEGTTDNDQAMSPAMSLXTV
R2	R2-1_BTe	.	Bombus terrestris	TCAATAGTTACTCG	TGAAAACACGATAACGATT	PLVGNRVACPXCEKREANLFFLNLSDLDRHLTQHHPDAPIXWSCIDCAKC
				GGGGCAGGCGGGA	ATGAATCAAAATAAGAAAG	FPKLHGARCHIPKCGGASSQARTGEFQCEACPMSFGSRRGLSTHERHAH
				TATTGGTCTTGCCTT	TAAACATCCCAGAAATTGTC	PAVRNIKRRGADPPEENTKSWKVEXVARLKGLWEIFKDHKYPNKEISKFL
				GCCCAAGTCACACT	TACGTCTTATTTGTTATCTAT	TTKTVDXXKYQRKKLNLIGXESPQEATSLATEGGCDLVSSGNASFGSPVG
				CCTACCTCCTCGTG	TTATTTGTTTA (SEQ ID NO: 1332)	RNENEEELIHEWKLSLKNEINKPTEVPPILKEVYNRLMLIWEEHQDDRDSL
				GTACCGCCGGTAAC		TESLDHFIRTALYELINKINKNQTDLKTKRAAKTKSPKNNRNSRKRFSYARC
				ACGCGCACGTCCAC		QELFHECPRRLADAVVNNDQAYLEPARQPPGSEEVRGLYEKLWGQVGS
				ATCAGCGAGGGGC		TYVPAPVTRVPKLSLSEIFPPIAAEDVGERIGKIRKKAAAGPDGLQRDHLTI
				GTACTCCCCCGGAT		PGLPIIMAKIYNILVYCSYFPSAWKENRTTLIPKINKPCSLVENWRPITISPIL
				GTGGCGGCGCGTG		GRIFSSIDGRIRRGTVLNMRQKGFTSENGCKINIELLNSALNYSKRNSGGI
				GCTAAACGGAGTGT		FTIVDISKVFDTVPHAALKPCLAKKGVPALIVDLIDEMYKNVKTTIKTKDGG
				GGCGACGAAGGAG		VEIMIRRGVKQGDPLSPLLFNLCLEPLLEEIEEQASGINVSEHRKVSVLAFA
				CGAAAGACTAACAA		DDIVLLGADAREAQHQINVLTDYLQSLMMNLSIEKCQTFEVVAKKDTWF
				CTATAACGGTCTTC		IKEPGLKIGNQIMPTVDPDEAFKYLGAKIGPWKGVHCGVIVPELLSVVKR
				CGTAACGGCTACTT		VRKLSLKPGQKVELLTKYIFPRYIYHLLVSPPSDTVLKLLDSEVRQEVKTILHL
				GGAGCCGTGAATA		VPSTATGFFYTPKACGGIGIPRFEHIIKLGTLKSAIKIANSIDPAVAGLIDDA
				ATGGAGCCTATATT		AIKKLKQTANSLRINWPASLEDIEKARKRLRKEHISQWADLKCQGQGVPD
				AAACCCTGGAACTC		FIKNKTGNLWLEDHSLLKPSRLIDALRLRTNTFGTRSVLARADKNIDVTCR
				GTTCCTTCGTTCTGT		RCRAQPETLGHILGLCQHTKGLRIKRHDEVKSLLEGRLKSKKNNEVFVEPT
				TGACGACTGGGAACG		IKAGGSLFKPDLVIKNGERVLVVDVTVRYENKNYLALAEKEKIEKYRPCLRA
				GCAAAGGACATGAT		LKEIFNAKGGEILPVVLGSRGTITPNTEKVLKRLGIANNDIKTILLNVLRSSIE
				TTGGATAACAATTG		LCNIFIDD (SEQ ID NO: 1454)
				GAAACTTAATATCG		
				AATTCACAATTAA		
				(SEQ ID NO: 1209)		
R2	R2-1_Crp	.	Crocodylus porosus	AGGCGTCTCCTTTA	TGAATCCCACTCTGGGGAC	VPPGAEARGRYHHPRXEXARQGEPPSXRVFLVXLPDSNPPCPICGDHVX
				AGGGCAACGGTCT	CCCCAAAAATTAGAAAAAC	XXSVLALHCVEGHXWAXVQYQCTHCGILCHIPRCQGRVXEXTGKDXXCP
				GGTTACGCGGTCGC	CCAAAACAGTTGTGTTTAA	ECPASFDEKAGLSQHKRHTVTXSXERVAGXLLRAXLRHGCWSVEEEETLT
				AGCAGKCTTGCKCC	GTGTGTTCTTGTTTGTCCCT	RLDAMFXGARNINQLIAAEXVSKMXKQISDKWRXLXLXPEQTTXGGXAE
				AGGTACCTCCWCG	TTGGCTTCACCTCCAAGTTG	SASVVXXESMTPEMEAQSPAXPPGKIRKIFTGQDGHAGGXAWENQEDF
				TGGTTCCCGCCGGG	CGATCCCCCCATCTCCCCTGC	HWTRWARRWLKRGQXLSDKVQEVLGXWVEGQPRIXAWVDXVSLDVL

		TGCCGMAGMCCCA	GCTGTCTTTCTGAATGACCA	TLFLGVPPGPQRAPSKKGPXEGGKPTSWMNKRAIKWGTFLRYQHLFGA
		GGSCTGTCGGTAGC	GTGGTGTTGAGGCTGGTGT	NRKLLAAXILDGAXRNQXTLLLEEVXQXYXGKWEAEPPFEGLGRFGXXRD
		TCGATCCTGGCACA	GACCTCGGTCACCTCCAAG	VDSFAFEALITXEEAVKHMMXMAXXSAPGPDKLTLRDLRRADPEGDALA
		GTAMGGCCAGGGG	GCCAAGTGCCCTGGCCCCG	ELFSLWXITGVVPDRLKEXQXVLIPKAVDFEKLRQLGNWRPITIXSIVLQLX
		AGTKCTTCCTTGCTG	AGTAGGACCAGGTGGCCCA	SRVLTARLTAACPIXPHQQGFISAPXCAENLKXPELIXRKVKXDRRPLGVA
		CWGGTGCCCCACA	GCTTGCTGGGCACCCGTCA	FVDXARAFDSVSHDXISWVLKAKGVDQHIVNLIEDSYQKVTMRVQVFSG
		AGCGTKTGGCAGS	CCGCCCAGGGCAAATGGAA	STPPISIKXGVKQGDPMSPLLFNIAMDPLIXKLKTVRQGVKVGSASLTTLA
		MCMCCTGCTTCWT	GGGATCTATCCTGACCACTA	FADGLXLLXDSWEGMQHNITTSXTPGRACNTTSRHPRGLLQPHGPTSAT
		CGCAAMAATMASA	CCAGGCTAAGTGTGGTGCT	XKMXGVLLESXMRLLYGEQLRGLEDXRPXXHDAXARRADTISGLEXRSLG
		GKGTSMTCAGTAGT	GCCTAGCCTGCCGTAAGGT	WDXQTRFGYATXLLAREDGDCXAQTNAEALCWXSXPFPGCAXRPXYAN
		CGGCCCCGCCGCTA	CAAGCGCCCTGCTGCCGCT	XGWVASEALDSMSRRXVKEWFHLPACTDXLLXSRHRDGGLGLLRLARXX
		GCCAAAACTGTTCG	CGGGTAKCAGTCCTCGTTCA	LAAXVRRPIRVATSSDEVTRKVSYACGISDEVERLXLAXGGDXSNVPRFED
		CCACSCAGTTACAG	CTCGTCCCTCCTAGTACCCT	PXAPKSXXVQGPHEAAQETPRVVRTQAIPWPSNWRAEEHSKWAQLSC
		ATGGTGCCTGTGCT	CCGCMTCTGCGCTTCTGCT	QGERVELFCNDPVSNGWINSRGQLAERLWIMALKLRSNIYPTREFLGRG
		GACCKGKKGCCCCG	ATCCACTGTGGCCGGWGAT	QAGTNIGCRHCTHPRETLGHILGICPAMQEARILRHNKLCKILAAEGKNC
		TGGGCTCGGWGGT	GCCGAGGWTGSWMCAWC	EWTVYYEPHLHNAAGELRKPDLIFVRDGTALVVDITVWYEGGPATLLSTT
		GCAGGGGWGGCC	CTCGACACCTCCAGGGCCA	AEKATKYLDLNTQIQELTGAEQVTYFGFPIGARGKWHADNWRVLSELGL
		GGCCCATGGCTGG	AGCGCCTTGGCCMCAGGTA	SNSRKERVTRLLSWRALLGSVDMVNIFVSKHRQENLLDEHCTPAEQVVS
		GCCAGACSTGGGCC	GGACCTGGCACCTGCCCAG	SYAS (SEQ ID NO: 1455)
		KTGGAGCCCGCTCC	GGGGCCAGACACTGCCTGC	
		CAMCCCAGAGTTCC	GGCAAGGGAAAGGGAGCC	
		CCCTCAAMGCCGCC	GTCCCTGACCGTTACCGGG	
		AGGKCAGMAMCAK	CTTGATGGTGCTGGCTAGC	
		CCGGGGAGKGMCA	CCGCCGTATGTCAAGCACTC	
		CGGCCCCGGCCGTG	CACAGCTGCTCGAGTTCGCT	
		A (SEQ ID NO: 1210)	GGCTTCACCTTCATCCCWCC	
			TAGTGCCTTCCGCCTCTGCG	
			CTATTTTCGTCCCGACTCGT	
			ACCTCCCCACCTCTGCGCTA	
			CMGCTATCCCTGMAAGGA	
			CCAAGTGGGCAGGGGSGTT	
			CGCCCCCCGCCCGCAGGAG	
			GCTCGGCGTATCCGTGGCK	
			TCWTGCCTCCACCGTCTTGT	
			GCCGCTAGAGGGGTACCTC	
			MGAGACCGGCGCAACACG	
			ACCTTGACGSTTAGACAGTA	
			GGGTGMAACAKCCCTGCTG	


CAGGCSTGAAGGGCCAAAC
GGCTGTGCCATGAGAGGG
GAACCTTGAAGACCGGGSC
AGTCAGCCAGTTAGTCAGT
TGGGCGAAACAATCCCAGC
TGCAGGCCCAMMAGGGCT
GTCAGGTGAGGGGGTATCC
CCAWCCACCCCCCGCCGCC
GACTACGGAGGCAKGAAGT
CCCTAGTGACTTCKGACCCC
CACGTCTTGTGCCGGGAGA
GGGGAACCTTGAAGATCGG
GGCAAGCCGCACTTGATAG
TTAGCCAGTCKGGTGAAAC
AATCCCAGCTGCGGGTCCG
AAAGGGCCGACTWCCAGG
CGAGGGGGGCCTGCGGAA
AMCCCCCTCCATGGTACGG
AGKTCTGGCGTCCTMACCG
ACWCCTTGCCACCAACGTC
TTGTGCCGGGAGAGGGGA
ACCTTGAAGATCGGGSCAA
GCTGCACTTGATGGTTAGTC
AGTCGGGTGAAATAATCCC
AGCWGCCCCCGCTGTGACT
GCTAAGMCWGGTCCCCAA
GGGGCATGAGGCATSTGCG
CTGAGCCGGSAGGGGTGAC
ACMCGGCGATCGGCGCAG
CACAKASTGAAGGGAGGCA
CTTGCTGAGACTGCTTCTGA
GGCCCCAGACTTGGGGTGG
TGCAGCCTTGTCTGGGGTA
TGGTACAGCACCCTACTGCT
CCCTTTGGKCAGCAGAATTC
GTCCCGACCTCTTACCCACC
CGAGTCTGCGCCTTGTTCCG
CTATCCTGCATCTCCGATCC

					ACCTCGCTGTCTCCCCGCTG CGCTGCTTTTCTCTCAAGTG GGTTAAATCTTGTCATGATT ACCTCCCACGTTTCCGCTCA AGGGCAATGCCCAAMATG ACGGGGATCGCTGGTGCAT GGCAGTCATGAGACCATCC GGACCCTCCGGTGGTCGCT ATAGTCATTTTKTGTTGCAT GGGGCATSCTGAGTCACTT AACCGAAAGACTCWAAATA ACTCAAAAGAGGKAMCCTC TGSGGTTCGGTAAA (SEQ ID NO: 1333)	FERRSNSWGYRPLEPRSVGTESNNNSPRSNITTSATSRPGDQPREAIAV VNLAGEIPCAVCGRLFNTRRGLGVHMSHQHKDELDTQRQREDVKLRWS EEEAWMMARKEVELEASGNLRFPNKKLAEVFTHRSSEAIKCFRKRGEYK AKLEQIRGQSTPTPEALDSITSQPRPSLLERNHQVSSSEAQPINPSEEQSN WEIMRILQGYRPVECSPRWRAQVLQTIVDRAQAVGKETTLQCLSNYLLE VFPLPNEPHTIGRSNLRRPRTRRQLRQQEYAQVQRRWDKNTGRCIKSLL DGTDESVMPNQEIMEPYWKQVMTNPSTCSCDNTRFRMEHSLETVWS AITPRDLRENKLKLSSAPGPDGITPRTARSVPLGIMILRIMNLILWCGKIPFS TRLARTIFIPKTVTANRPQDFRPITVPSVLVRQLNAVLASRLASKVNWDPR QRGFLPTDGCADNATLVDLILREHHKRWKSCYLATVDVSKAFDSVSHQA IIKTLQAYGAPTNFVSFIEEQYKGGGTSLNGAGWSSEVFIPARGVKQGDP LSPLLFNLIIDRLLRSYPREIGAKVGNTMTSAAAFADDLVLFAETPMGLQT LLDTTVGFLASVGLSLNADKCFTVSIKGQAKQKCTVVERRSFCVGERECPS LKRTEEWKYLGIRFTADGRARYSPADDLGPKLLRLTRAPLKPQQKLFALRT VLIPQLYHQLTLGSVMIGVLRKCDRLVRQFVRRWLDLPLDVPVAYFHAP HTCGGLGIPSIRWIAPMLRLKRLSNIKWPHLEQSEVASSFIDDELQRARDR LKAENVQLCSRPEIDSYFANRLYMSVDGCGLREAGHYGPQHGWVSQPT RLLTGKEYLHGVKLRINALPSKSRTTRGRHELERRCRAGCDAPETTNHILQ KCYRTHGRRVARHNSVVNAVKRGLERKGCVVHVEPSLQCDSGLNKPDL VGIRQNHIYVIDVQVVTDGHSLDQAHQRKVERYDRADIRSQMRRFFGA TGEIEFHSVTLNWRGIWSGQSVKRLIAKDLLIAEDTKLISVRAVNGGVTSF KYFMYCAGYTRS (SEQ ID NO: 1456)
R2	R2-1_DWi	.	Drosophila willistoni	GAAGCTGGGTCGG ATGAGCGCAGAAG GGGTGTTCTTTGGA ACACTGTAATTCAT AAGTCGTAAGTCTG ATCAAGTCGACTCG AAACCTCCTCGTGG TGTTTCCTGGGTGC TGTTGAGTTCCTAG TCTCTAGGTTCTTTT CAGTAGCTAA (SEQ ID NO: 1211)	TAGATGTACTAACCTCTAGC TTTTCTTATACTTTTGCCTGC TACCTTGGCATTACATCTAA AAAGGTACAAACATCGCAT TGTCATAAAGAGGTGGTTT TAGTACGTAGGCGCTGTGG GACTTCATTGTCCCGGTGAT GCAGTGAATCGTGCATACG AGATTGTCCAGTAGTTGGTT GCTCGTATCTTTAGAAGATT TCCTTCCTCGGCGATCAAAA AAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1334)	

R2	R2-1_Gav	.	Gavialis gangeticus	AGGCATCTCCTTKA AGGGTAATGGTCTG GTTACATGGTCATA GCAGGTTTGTGTCA GGTACCTCCCAGTG GTTCCCGCCGGGTG SCAMAGCCCCAGG GCTGTCGGTAGCTC GATCCTGGTACAGT ACGGCCAAGGGAG TTCTTCCTTGCTGTC GGGTGCCTCGCAAG CACKTGGCAGCCCC AATCGCTTCATTGC GAAAAACACAAAC GTCCTAAGGGGATG ATCAGCTAGTCAGT TCTGCCGCTAGCCA AAACTGTTTGCCAC CCAGTTACAGATAG CGTCTGTGCTGA (SEQ ID NO: 1212)	TGAGAAGTTGCGAGTTCTT ATGCAAGTTGAATACCACTC TKGKGACCCCAAAAAAWW AAACCCCAAAACAGTTGTG TTTAAGTGTGTTCTTGTTCG TCCCTTTGGCTTCACCTCSA AGTTGCGATCCCCCCATCTC CCCTGCGCTGCCTTTCAGAA CGGCCGGTGGTGTCGAGGC TGGCGCGACCTCGGTCACC TCCAAGGCCAAGTGCCCTG GCCCCGAGTAGGACTGAGT GGCCCAGCTCGCTGGGCAC CCGTCACCATCTGGGGCAA ATGGAAGGGATCTGTCCTG ACCACTACCAGGCTAAGTG TGGTGCGGCCTAGCCTGCC GTAAGGTCAAGCGCCCTGC TGCCACTCAGGTATCAGTCC TCGTTCACTTGTCCCTCCTA GTACCCTCTGCCTCTGCTCT TTTGCTATCCACTATGGCCA GTGATGTTGAGGTTGGTGC ATCCTTGGTCACCTCCAGGG CCAAGCGCCTTGGCCACAG GTAGGACCTGGCACCTGCC CAGGGGGCCAGACACTGCC TGTGGCAAGGGAAAGGGA GCCGTCCCTGACCGTTACCA GGCTTGAGATGGTGCTGGC TAGCCCACCATATGTCAAGC ACTCCACAGCTGCTTGAGTT TGTTGGCTTCACCTTCATCC CACCTAGTGTCTTCTGCCTC TGCACTATTTTCATCCCAAC TCGTACCTCCCCATCTCTGC GCTCCTGCTATCCCTGCAAG GACCAAGTAGGCAGGGGGG	PAAPRAWGAVEAGPWPGRTRAVEPAPSPESSPSEAARAAPAGEGHGP GHESPSVQRPEADTTAPGVSAPTREGEPPSTRVFLVRLPDSNPPCPICRD HVGKPSALALHCVESHAWADVQYQCTHCKKVSANKHSILCHIPCCQGR VPEWTGKDWACPECPASFNKKVGLSQHKRHVHPVTRNVERVAGSLSRA GLRPQTRRGCWSVEEEETLTCLDAMFRGARNINQLIAAEMVTKMPKQIS DKRRQLGLCPEQTTLGGDAESTSVVEEESMTPEMETQSPINPPGKIRKIL AQRARRWLKKGQGLSDKVREVLGAWVEGQPRIHAWVDSVSLDVLTLFL GVPSGPQRAPNKKRPKEGGKPTSWMNKCAVKWGTFLRYQHLFGANRK LLVAIVLDGADRNQCTLLLEEVFQAYREKWGLEEVLRAYRGKWEVESSFE GLGRFGVRRDADNFAFKALITPEEVVKHMMAMASKSAPGPDKLTLRDL RRADPEGDALAELFSLWLITGTVPDGLKECRSVLIPKTVDREKLGQLGNW RPITIGSIVLRLFSRVLTARLAAACPINPRQRGFIAAPGCAENLKVLELLLRK RKRDRQPLGVVFVDLARAFDSVSHDHISWVLKAKGVDEHIVNLIEDSYQ KVTTRVQVFNGVTPPISIKTGVKQGDPMSPLLFNIAMDPLIAKLETDGQG VKVGSASLTTLAFADDLVLLSDSWEGMLKNISILEDFCNLTGLRVQPKKC QGFFLNPTCDSFTVNNCEAWKIAGREITMLGPGESTRYLGLNVGPWVGI DKPDLGTQLSSWLERIGTAPLKPMQKLSLLVQYAIPRLNYQADYAGIGRV ALEALDSMNRRKVKEWFHLPACTSDGLLHSRHRDGGLGLPRLAKAIPEA QVRRLIRVATSSDEVTRKVSYACGISDEVERLWLARGGDMSSVPRFEDPE APRSPGVQGPCEAAQEIPSVVRKLAIPRPSNWRSKKHSKWAQLSCQGE GMELFCNDPVSNGWNNSRGQLAEHLQIVALKLRSNIYPTREFLGRSQAS TNVGCRHCTHPHETLGHILGICPAVQEARIIRHNKLCKILAAEGKKCEWTV YYELQLLNAAGELCKPDLIFVRDGTXLVVNVTVGYEGGPAXLLSTAAEKAT KYLDXNAQIQELTGAEQVTYFGFPIGARGKWHADNXRVLSELGLSNSRK ERVARLLXWRALLGSVDMVNIFASKHRQENLSDXALSPS (SEQ ID NO: 1457)

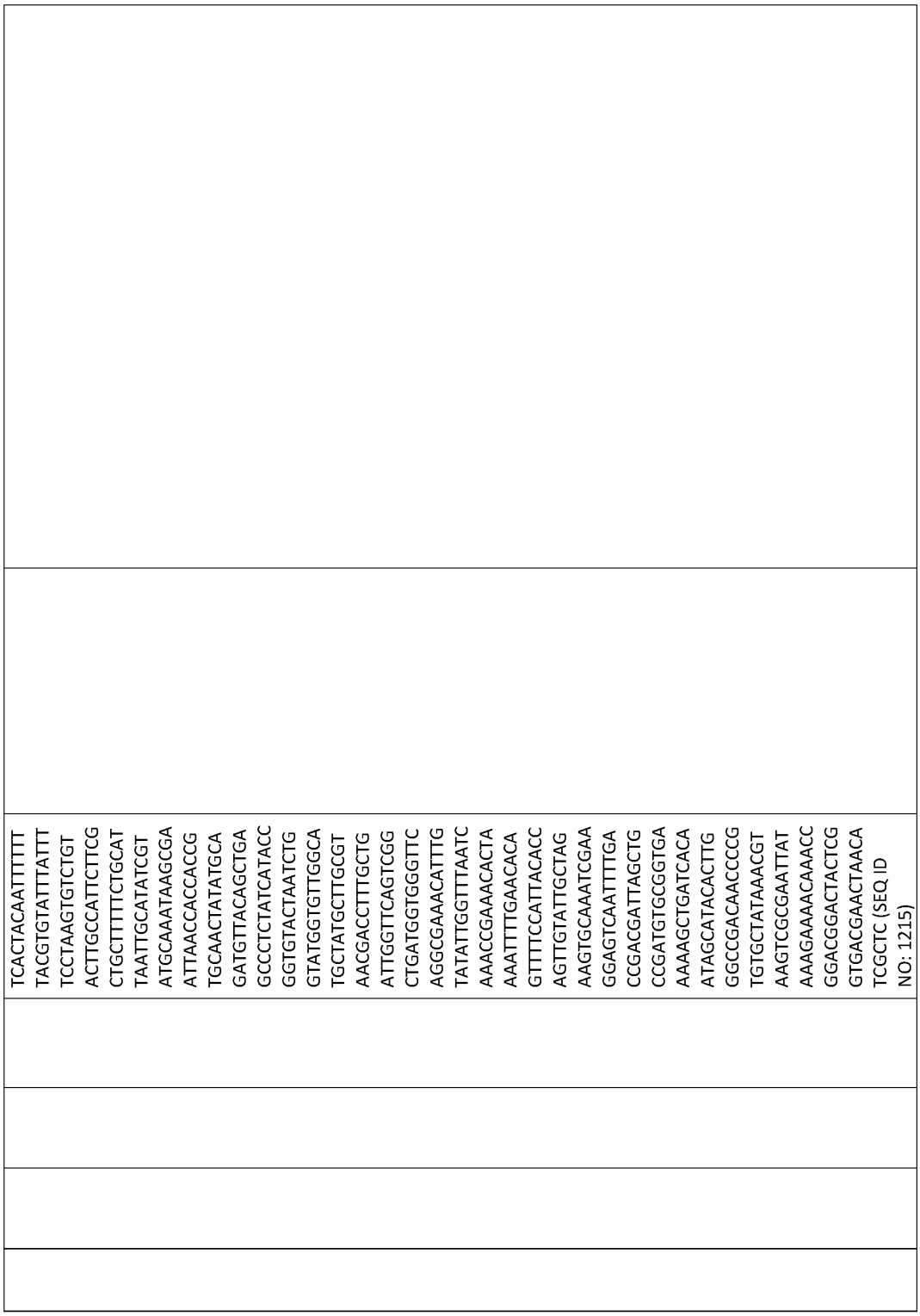

GTTCATCCCCCTACCTGCAG
GAGACTCAGCATATCCATG
ACTTCTTGCCTCCACCGTCT
TGTGGCGCTAGAGGGGTAC
CTCAGAGACCGGCACAACA
TGACCTTGACGGTTAGACA
GTAGGGTCAAACAACCCTG
CTGCAGGCCCAAAGGGCCA
ACAGCTGTGCCACGAGAGG
GGAACCTTGAAGACTGGGG
CAGTCTGACCATGCTGGTTA
GTCAGTTGGGTGAAATAAT
CCCAGCTGCAGGCCCAAAA
GGGCTGACAGTCAGGTGAG
GGGGTATCTCCATCTGCTCC
CCACTGCCAACTACGGAGG
CATGAAGTCCGTAGTGACT
TCTGACCCCCACGTCTTGTG
CCATGAGAAGGGGAACCTTG
AAGATTGGGACAAACCGCA
CTTGAAAGTTACTCAGCCG
GGTGAAAATAAGTCCCAGT
TGCGGGCCCCTCGGGGCTG
ACAGTCAGGTGAGGAGGG
CTGCAAAGCCCATCTCCTGA
CTCCAGAGGCCTGGCGTCC
TAACCGACTTCTTGCCACCA
ATGTCTTGCGCCAGGAGAG
GGCAACCTTGAAGATCGGG
GCAAGCCGCACTTGATAGT
TAGCCAGTCGAGTGAAACA
ATCTCAGCTGCGGGTCCGA
AAGGACTGACTTCCAGGCG
AGGGGGGGGCCTGCGGAA
AACCCCCTCCATGGTACGG
AGGTCTGGCATCCTAACCG
ACACCTTGCCACCAATGTCT
TGTGCCAGGGAGAGGGGAA

					CCTTGAAGACTGGGGCAAG	
					CCGCAGTTGATGGTTAGTC	
					AGTCGGGTGAAATAATCCC	
					GGCTGCACCCTGCTGTGAC	
					TGCTAAGCCCCGGTCCCCAA	
					GGGGCATGAGGCATGTGC	
					GCTGAGACGGGAGGGGTG	
					ACATCTGGCGATCAGCACA	
					GCACAGACTGAAGGGAGG	
					CACTTGCCGAGAATGCTTCT	
					GAGGCCCCAGACTTGGGGT	
					GGTGCAGCTTTGTCTCGTGT	
					ATAGTACAGCACCCTACTGC	
					TCCCTTTGGGCAGCAGAATT	
					TGTCCTGACCTCTTACCCAC	
					CCGAGTCTGCGCTTTTGTTC	
					CACCTCGCTGTCTCCCTGCT	
					GTGCTGTTTTTCTCTCAAGT	
					GGGTTAAATCTCAACATGA	
					TTATCTCCCACGTTTCCGCT	
					CAAGGGCAATGCCCAACAT	
					GACGGAGATCGTTGGTGCA	
					TGGTAGTCACGAGACCATC	
					CGGACCCTCCAGTGGTCGC	
					TATAGTCATTTTGTGTTGCA	
					TGGGGCATGCTGAGTCACT	
					TAACCGAAAGACTGTAAAT	
					AACTCAAAAGAGGTACCCT	
					CCGGGGTTCGGTAAA (SEQ ID NO: 1335)	
R2	R2-1_IS	.	Ixodes scapularis	GTTCCAAAGGAAG	TAGTGTGACGGAGTCCTCA	MQCTSRLADAPRFARVGVEGEGVGASGNGTDAQLWYGCTGCDEAFSS
				GCACTCCTTTGGTT	AGCCCCCACAAGTGCCTGC	LRGLRIHAAQKKHGNQDGLLRLPAGRPRKRRVGKSTTAGASDRVTTDPV
				CGTGATGAGATGTT	CAGGTGGCAGGAAAGGGC	PAPVPESPGLLPGLPGPSLPGCSDLPPGVLPGGWSASPGPLSWPPSLDA
				CATGGTGCTTGCCT	AACTACTGGTGAGCGACCC	GPLPGPSRVSPGPSRPSPGKPTGPPSLDAGPLPGPSRVSPGPSRPSPGKP
				AGCTGGAGAAATCC	AAGCAAGGCGGAGCCAAG	PGTPEPLPGSPGGRRGVSPGQPGSRTDPSSSAGAGHFVCPQCSRAFSSKI
				GACTCACACCTGCA	ACCAAGCTGGAGCCAAGAG	GMSQHQKHAHLEEYNAGINITRTKARWDPEETYLLARLEATLNPDHKNI
				CGTGGTCCCTGCCG	CAACTCCAGGAGGCAGGGG	NQTLHAALPRGSCRTLESIKAHRKQAAYRDLVTSLRSARESSEAQHVPDR
				CCTGCCAGTATGCC	TGGATATCAAGAGCAACCC	PLETPEPQTPANPQRDSKQAVIEALQSLIGRAPPGSFQGARLWDIARQA

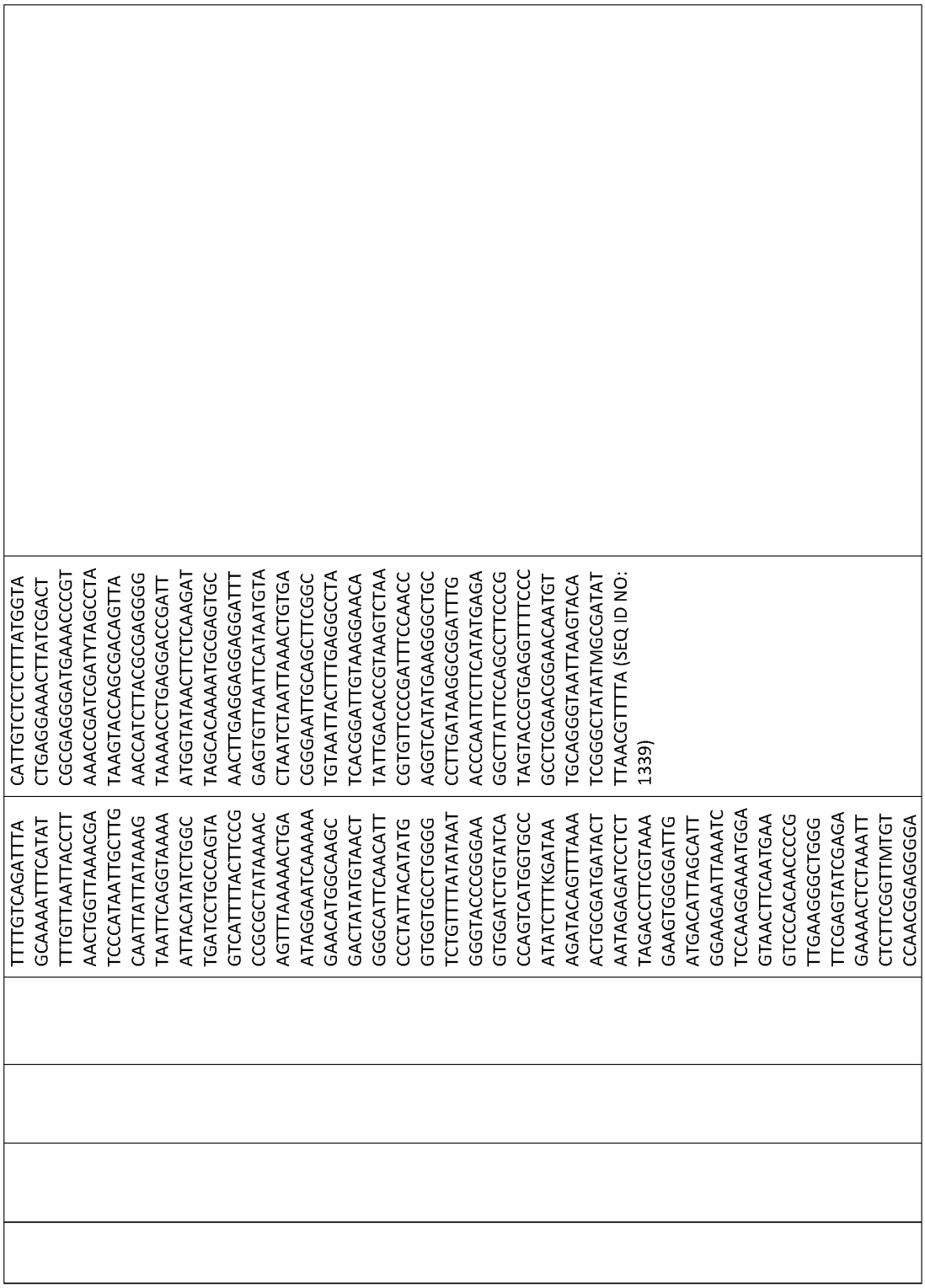
			GAGGAAACGGGTG CAACTTAATCCGTG GATACTGGTAGCAA CGTGAGCAACGGTA CGGTCCTTCGCGGA CCACCCTGGGCGTT CGGGTTGCCAGCCC GTTCGCCCGAAATA TCTTGGCCCTGAAA CTAAAAGAAAA (SEQ ID NO: 1213)	CAAGGGACACAGACCACGG GCAACTACTGGTGAGCGCC CAAGACAGGGGTGGATATT AAGAACAGCCCCACAAAGT GTTACCTATATTAACAATAA AGTTGAAGCCTCAACCACG CATTGCGGGTTAGATGGCG TGGCTTGGCCCGCCGCCAT GATGAGCTGGAACCCTCCA CCTGGTGGGCCGCACGAGA CCACCGGCTCTTTCTACTAA GGCCGGTCTCCGTGACTGC GGTTGGGATAAACTCCAAG CACTGAGCGGTAAAAAAAA AAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1336)	TRGTNILPLLNSYLRDVFTLPTKPTRKKPAVRPARSRRKQKKQEYARTQDL FRKKQSDCARAVLDGPTSSSVPGTGAFLQTWREIMTGPSPALEAPPLPTR GEVDLFFPATAQEIQSAEIAVNSAAGPDGFSARLLKSVPALLLRVMVNLLL LVRRVPAALRDARTTFIPKVPDAVDPSQFRPITVASVLQRLLHRILAKRALE AIPLNFRQRAFQPVDGCAENIWLLSTALNEAARTRRRPLHMASVDLTKAF DRVTTDAILRGARRAGLSGEFIGYLKELYTTSRTLLQFQGESLLVEPTTGVR QGDPLSPILFNLVLDEYLSSLDPDISFVSGDLRLDAMAFADDLIVFASTPAG LQDRLDALVEFFDPRGLRVNVKKSFTLSLQPGRDKKVKVVCDQIFTIGGT PLPASKVATPWRYLGMTFTPQGSINKGTSEQLDLLLTRTSKAPLKPQQRL VVLRNYLLPRLYHRLVLGPWSAALLLKMDTTIRGAIRRWMDLPHDTPLG FFHAPVTEGGLGINSLRASIPAMVLQRLDGLHFSTHPGAEVAIQLPFLTGL HRRAEAAAQYQGQRLLSKADVHRMWSARLHGSCDGRPLRESKRVPAA HRWAAEGTRLLSGRDFISITKLKINALPTLERTSRGQHKDIQCRAGCQAVE SLGHVLQACHRGHRGRIRRHDNIARYVCGRLTQIGWAVKWEPHYSVAG RTLKPDIVAHRGAETVVLDAQVVGTSMRLGFHHAQKKEKYSLPDLLHQV CEGRRDAARVSTITLNFRGVWAPESAQDLKSLGLTDNDLKLLTVRCLQG GAQCFRLHRRMTTVVKATGDEANALPAHSGLPPTQLGGRTLGPSAHNQ SARTT (SEQ ID NO: 1458)
R2	R2-1_MLe	.	Mnemiopsis leidyi	TGGGGGCCCCTTGG ACTTGCTCCCTGGG GCAGGACACCAGT GAAAGGAGATCCTC AAGACAGGACAGA GAGACAGGCACAC CAACCCTTCGAACC TGAGGAGACACCCA GATCTGGTTACCCC ATTCCTGTAACCAT GGTTGCTCTCCGTG CGCTGTTAAGGAAA CCCAGCCTAGTACC CTCGGGGAAAGGTT GCGTAGTACTTAGC AGTGTGCGGATCAA ACCTCTACCGGTCT CTCTAGCGATGAAA GTTTCTCCGACTGG AACTTGAGGGACTG	TAGTGACGAGAGAAAGTCGC TTTATCCTTACATAACAGTG TGATAGTCATCCTTATCACA ACTGTCCTGGCGAACAAAA TGGCAGAGGGATAGTACTC GGTCCAACCAGAAGGAAGC CCACACGATGCCAAACTTGC TGTAACCCACGTGAGCGGA AAACATCCTCCGAGTATAGT ATGATGGAAGGATACAGGA TGTGGACCGCTCTAGGCGG CGGACGGTGTAGACGGCG ACCTTACCATAGCTGCGACT GCTGTGCGATCCGGAACGG GCCTTTCTCACTTACCTCAA GCACCGTGTCGGATCGCGG GGTGGGCGCGGGATTCCTG CTGGGAAGTTGTCTGCACC CAAGTCGTCAGCTAACTGTC GGTACTTATACGCACCCAG	MSNTSHSKLNLKMDNKLKTSLETPSGVRADSIITRVRTSSNRGEHSNGVT YPRCEQGVAPLDTHGGICDAPPQVTVPATETDKQKKCEYCEFTYLKPRQI GTHMRKRHPQEWNDIKRTKFLSEKRQKRWLDEDFELLCIGQEEYLVLSSI GKQGKGINQYIQTKYFPTLSTDAIKSQRKSRRFSEYSEKRSRELQPCNTSS DPEELPNEAVTENSPLSFDPLDRDVVKKISSKDHGDQILLVQEHLINGRYQ EANTLAKAIFEKLSGKFPNLKTGDHRPGKQQTARKVGKKRVRGSGKKLSP SKQNRRELYAIVQKQWRTKKRSKVINQILTGNLNKEQSYTHTPDQLAQF WSTLFGRVSPRDDRPINHRRSVIPELDKPLSVEEVEAALKGAKDAATGID GVPISHLKHLGSAALTILYNGLYVTGSIPDPWVKRARTILIPKSNPPASPGDY RPISISSYFYRIYTSAISKRLASAVSLDDRQKGFIKEDGIRDNLSLIDTLINETK AGSKSLFMTFMDVKKAFDSVSHYAIARSLEWAGVPDGMRSVIADLYQD CTTDICGRSVKVTRGVKQGDPLSSTLFNLVIEMVMSNVPERLGIQFQGH RLFYLAFADDLVLLTRGPTANQKLVSLVHEQLARVGLELHPGKCKSIAIMA DPKRKTTFVDQGSSVLIGGEPVSSLGPQEWYKYLGIKLGSGGMPQGIYR DQLADLLAKTDSAPLKPQQRLYILRSHILPKFNHRLMFERVTCQTLEGLDK LIRTHVRKWLKLPKDTPGPAFYADKGSGGLGLITLRYRVPLLKLRRHKKM ADSPDPVIRLIPNAEPTISLLARWTKMCSLYGKQYQHKSELSKIIRDKYWT MCDGKGLRTEVPPDTAKKTLSLLFEDRTPLKPGQLIGAIGVRLNTLGTPAR NNRAKGYSPEANICDKCPGNRQATLGHISQTCPATHGRRVKRHDKIVNR IAKALKERGSVKNILTEPHLRHDKLPLRKPDLIVHTEKSVEIIDVQVVADQG

				GCTAGCCCAGCTAG CCTGTAGAGCAATG CGTATACGATGCCT TGGCGACAATGGC GACCGCTGCTTAGC AGACGGAGGTTAG TGAAAGGGCGACTT GCTGTTCATAGTCA CGTGAGTCGTCTAG AAACTGCACCGATG CTGCCCTCTGTTCCA GGAGAAGGACAGT GGAGGACTAAATC GTAGCGCGAGCGG TGTT (SEQ ID NO: 1214)	AGCCATGCTCTTACCAGGCC ACTTGACGGCGCTACACCG GTCTGTAGGAGGGTATCTC AGCGACTTGGACAAAATGG ATCTTGACTGCCTGAGGGC GCTTATCGAGTGACCCAGA GTAAGCTGGTAGGAAGAAT CTTGCAGTTGAGAGGGCTA GTAGGCCAGCCTCTGCTC GTCAACTGTCCTAGGCGTTA ACGTGTGCCTCCCTATGAG GTAACCGTGAAGTATTACTT GTTTCCCTGTAGAGACCATA TAACTAGCTGAGAAACGCC CTGCGGGTTAATATACAGTT ACCGCCGTCCTGCATCCCCG AGTTGTGACCAGCGTACAG CCATGTACCACGGCACTTCC AGCATTCTCTGGTCTAAGAA TGTATCAGCTGGCCCGCCG AAAGATAGAGCGACCCCGC CTCTATCACTAGAGTAAGCA GGCACGCGAAATATATGCA GGACGCCTTGGCTCAGTAG CTCTGGTACTGAGTATAGC ATGGCTGAACGCCCCTTTAA GTTCGGTGGCCCAGAGACT CTCCGAGACTTCCCAGTCCT GGAGAGGGTACACCTTTAC CAGAACTTTCTGTGGTCGGT TG (SEQ ID NO: 1337)	ISRHEDEDQQKKIVKYDVDGYKRAAYKMLGIDYGSIPCNVSAFTITWRGN LAPHSLKLASRLQFSPVLKYIVADSLVDTWGAFLIWGKTS (SEQ ID NO: 1459)
R2	R2-1_PM	.	Petromyzon marinus	CTATTAATGGGATG AAGAAGGGGGACA CGAGTTTGTGTGTG CATCCAGTTTCCAT GGTGCATGCAGGA GTGGTGGTTTAAAT GGCGAGACTCTACA	TAATTTAAGGTAAAATCGT GGGATTGTTTTGATGGCAA TCTGCCTAGTCGCGGCCTTC CATTTTGGGTAGGCAGCAG ACCCATCTATATAACAAACT ACTTTGCCTTTCATAGGGGT ACCCGACCCTACCAACTTTC	MNERLTDELTTEFILSDMFLWDYPCTDQNKCYPCNLVFLDHRTWSSHM ARVHPHANKTYKCRICNRTADSIHKIASHYGRTCKSLIGKTNAITTTIDETL FSCLHCSRGFTTKTGLGVHTRRTHPTEHEAILQQNTPGRKVRWGEEEVEI MAHKEAQQKDEDINMNQLIQNSVMPHRTLEAIKGKRRNIKYKELVRTLK ETTYKVENQCLVNLVLPTTSEITTTPSEGDQPAIRAEKEQSPTAAEDLQVII NDLKSQNFSHNQALLLLNSHVEKFLNRSKPIKRKDHVNQQEIDENRHRR QSKQTKYRRYQYLYHTNKKALLDEITSDRSGPSIYPTEESIRGTFVTLFESNS

		GGGCTTCCATGGCT	GGGGAAGTAAAAGAAA (SEQ ID NO: 1338)	PPDNIPSKLKNDQSCIDIVKAITLDELIKTLAIMKDKSPGQDNITLSDLRTLP
		ACACGGGATGCAA		IKYLLDILNIILYIQDIPQIWKQHRTRLIPKTKEELEKPSNWRPITISSIVIRLLH
		GGCATCAGACATTT		KILSYRLGQQLKLNYRQKAFLPVDGCFENSALLHFIIHNARQKHENTQIVSI
		TGGCACAGGCAATC		DLSKAFDSVSHESIIRALNRFNLSKESITYLTNIYKCNLTDIVFGSTIMRNINL
		CTTTTGGTCTCTACC		KRGVKQGDPLSPLLFNMIMDELLDNLPTYIGVNVGNQKVNSMMFADD
		GCAATCATGTCTTA		LILFAETECGMNKLLDITTKFLDDRHLKININKCNSLRFIKYGKQKTFSVATT
		GACCTCAGTAGCGA		SSYFINNEPINPVSYVKGFKYLGIEFDPRGKRSISCNLLAAMLNKLTRAPLK
		CCACTACAACCACA		PEKKVYLINNNLIPRIIHQLVLGKVTKGLLMSLDSEIRKTVKLLLRLPHDTPD
		GTGGTGACTGCTGT		SFFYTSVSNGGMGIRNLCDSVALSIINRHNKLITSDDLVIRALSQQSYTIAT
		TGAGTGAAGGACG		LKQAHIIAGSKFPSKSLNQNKWSNKLYQTTDGRGLVYCQSQTENNSWIT
		ACTGAGCGCTGGAT		GNHRTIKSYNYIDMVKLRINALPTKSRCNRGTLETKQCRFKCRSINNQISE
		AACAACTTTCTTGC		ETLAHILQKCDRSHYSRIARHDSLVQFLATAAQKLNWEVIKEPTLPSDTNK
		GTGGCCCAACATCG		AKPDLILVRDSHVLIVDVAVPWESRSLAHAYDFKVKKYATDKKMQAYLKT
		AAGCAACCACTTCG		IYPEKEIRTEALIISARGGWCALNNMVTKKVGLSSAWVKLALIKVMEGSV
		GAGCTGGCACAAG		KIWRSWSKG (SEQ ID NO: 1460)
		GCAAGAGGGCAGC		
		CCAAGGTGTGAATC		
		ATCTCAACTTCACTG		
		CAGGAAGAAATGCT		
		GTGCAAGGATGAG		
		TGTGAACGACACCA		
		ACGGGATTGTTGCT		
		GACCAGGAGGTGC		
		CAACCAAATTTGAA		
		TGGATTGACTTTGG		
		GCCTGGTTTCTCCT		
		GCGTGTATTGCACG		
		GAAAAACAAGTGG		
		CTACACGTGTGGCC		
		GTCGTGTCCTGGGG		
		TTTCGCAACACAAC		
		TCCACAAGATCGAC		
		AACTATGAGGATGA		
		CAATGTACTTAAAG		
		AACAAAGAGACTG		
		ACGCCAAAGGGGA		
		TTTAAACCGCCAAA		
		TCGTACATTGGGTC		


TCACTACAATTTTTT
TACGTGTATTTATTT
TCCTAAGTGTCTGT
ACTTGCCATTCTTCG
CTGCTTTTTCTGCAT
TAATTGCATATCGT
ATGCAAATAAGCGA
ATTAACCACCACCG
TGCAACTATATGCA
GATGTTACAGCTGA
GCCCTCTATCATACC
GGTGTACTAATCTG
GTATGGTGTTGGCA
TGCTATGCTTGCGT
AACGACCTTTGCTG
ATTGGTTCAGTCGG
CTGATGGTGGGTTC
AGGCGAAACATTTG
TATATTGGTTTAATC
AAACCGAAACACTA
AAATTTTGAACACA
GTTTTCCATTACACC
AGTTGTATTGCTAG
AAGTGCAAATCGAA
GGAGTCAATTTTGA
CCGACGATTAGCTG
CCGATGTGCGGTGA
AAAAGCTGATCACA
ATAGCATACACTTG
GGCCGACAACCCCG
TGTGCTATAAACGT
AAGTCGCGAATTAT
AAAGAAAACAAACC
GGACGGACTACTCG
GTGACGAACTAACA
TCGCTC (SEQ ID
NO: 1215)

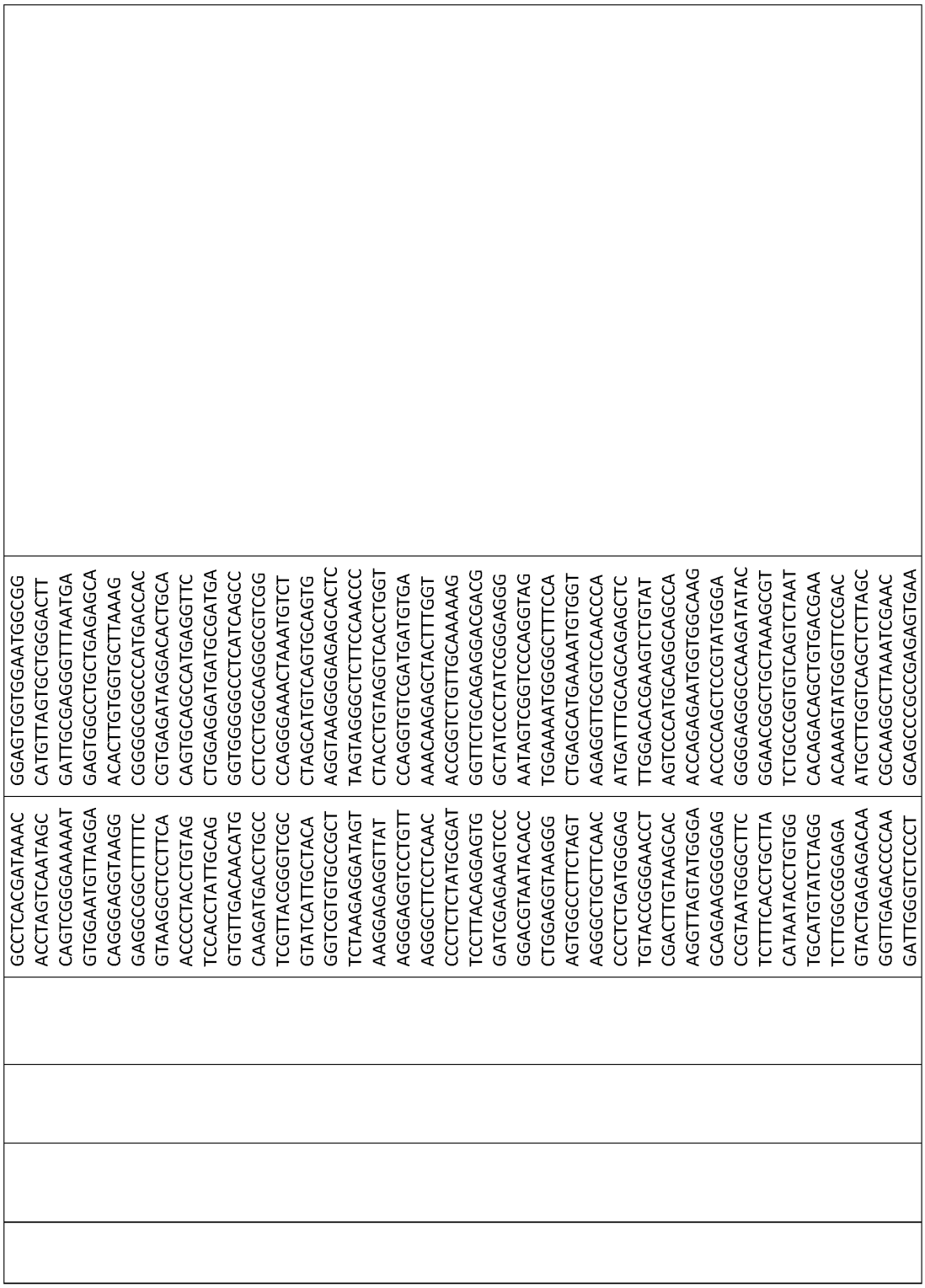
R2	R2-1_SM	.	Schmidtea mediterranea	CAGTGCTATTCGAA	GGCCACGCGCGTCGTCCTT	MKKVLNNETEKLPGSNLTFMCGFCDREFDTARGRGVHESRGHLVERDA
				TGTCAATGTGAAGA	GTTTGATCACTAGTGGATCA	AVQSRVKAVVSKKYYYSNEEDVALAKMQLXHADLAKSEXLEAMYLALGK
				AATTCAACTAAGCT	ACCTTCGACTCCCCGGAACT	GRTREAIEQHIRKSLRYKGVLEEQRKLLETARGNVRQNNVGVPASNATK
				CTGGTTAACGGCGG	GTGGGAGTGGCGGAAGAA	NLQRFLESLPLGTNRREERLDRIIRSNSIESQRLELIHYCNDMCQDFVQLD
				GAGTAACTATGACT	AGGCCAGAGGATGTCCTGA	CQXNPINAIRRRNPKRLSKKQLKRAKFSALQRLWIRDRKAAAQLVLKDKL
				CTCTTAAGGAATTA	AACCATATATTTATTTATAG	DSLLSNKEDSKDLGSYWQQVFERESELDRRPIPQVVENEELNSPVLEKEV
				AGAATTTACCTGCC	AAGTTTTACTTCATCCTATTT	EWAVKNIKKSTAAGPDGLTALALKKIPYSELVKLFNIILLVGFLPDVLKNSR
				KTAKTAAAARTGAA	ACGTATTTCAGTATGAAAAT	TILIPEVDNPQQGGGDYRPISINSVLTRTLNKILAKRVSEGDFGINGQKGFKS
				ATCMGTTGTTCAT	GAGTAAAGTTCTCGACTCG	VDGCLENLATVESILADARMKKNKLAVVFLDMSKAFDSVNHESIVRAGEI
				WGCAAGTGGTATT	ATGAGTTGGGGGCAACCAT	KGYPKLLMTYVKECYNDATTNVAGVTAKFNRGVKQGDPLSPALFNNVID
				GTACACCTTCCCGC	TGGGGGTCCTGAAGAGAG	LAIERVSGTGIGYNMGGKKYSVVAYADDLVLFGESREGLQIALTALLEELK
				GGTGCTAGTCGTTT	GCTCTCACTGTAAAAAATCT	LNGLTPNPAKSASLTFERSGPHWFASTDTVTALGDQIPAMGNIETYKYLG
				AAAACTAAGTTACA	CTTCGTGTCTGTTTATTCCTA	IKFNSCGVVKGSLPGIYTKKLELISKAPLKPQQRLAMLTDFLIPGVLHQAVF
				AACCACGAGGGGC	GGCACNTGCTGCATTATGA	GQTNAGDLRSLDKRTRRAVRSWCHLPSDTSTAFIHAKAKDGGMGIPSIR
				GTCCTGACGGACTG	AGCGGWGAAAGTAAAGTT	AEVQFGKLDRFGKLPNVKDERSKVLADNAHIKKKMLEKLGVGIPIKGVRC
				SAAAAGCATTGAGR	AGAGCTGAGAGATAGGTAC	KNKLEFYNKMREELIKSNDGIGLKEASLVPSANTWLKLSDLHMSGRTFVG
				GTCMTGAAGAGAG	TTGCTGCATTATGAAGCGG	CLKTRGNLMATVTRTSRGGQNPGIELNCKKGCQYQGSLNHIVQKCPVV
				GCTCTTATTGTACG	AGAAAGGCCTCGAATAAAT	KGLRIKRHDEVVKYVEEITKKAGWSATMEPIIPFEGSHRKPDLVLVRGDLG
				AATCTCTTCAACGA	AGGGTGTTAGAGTTATTGA	KVVDIQIVSDHCGLDEKNSCKIGKYDNDIIRNYVRGLGPSRVEVAAITLN
				TCGAAGTCTGGACC	TGGAGAGTATACTAGTAAG	WRGVWSRDSFNLIKRLGMTEMDAKIISMRVLASTAKMFKTCKKVLEPV
				GATATGAGAACTAA	CTTAAGCTGCGCCTCGCGC	CRTKTADCDGYGPEETSARPCHELNLKESSGT* (SEQ ID NO: 1461)
				TACATTAGTTGACA	GGTGCCCAAAAATATACTTA	
				GGTGAAAAATACTG	ATGAGAGCAATAACTCAAG	
				TTGATTACTTAGTTC	GNGAGTTTAATTCATATGC	
				TCAGTCATGTGGTA	GCATGCGGCACCAAGGTGC	
				TATTGCCAGTCAAT	TGAATGGCATCGATTAAAC	
				TACTACATWAATAT	CTCTCCTGTTGTAGAAGCAG	
				TAGTGTGGCTCTCA	GTCATAAATGGAGGRGGGC	
				AAGGAACACGATTG	AACCACTGAAACTTATGAG	
				RTCGGCAGTCCAAT	CCAGAAGAAGCTTAACTAC	
				GCGCGACTGGCGG	AWAAGTTTTAGGCAATTAC	
				GCTTGTTGTTTGCA	TGAACGGAGTTAACTGTTA	
				TTTGTTACCGGCTA	GTTAACACTACCATGTAGTT	
				CTTGAAAAGGTTAT	GTTTATAAAGCAAATATCAG	
				ATATAGCAGACGCT	GTTTCAGTCTATATACTAAA	
				TAAAGCGCGACTGT	AGTATTTTTTGATACCGTGG	
				AATTTACATYTCATT	TATATAGGCAACTAGTTAG	
				GCCCAGTATTTGTC	GAAATAGTAAGGGATGACG	

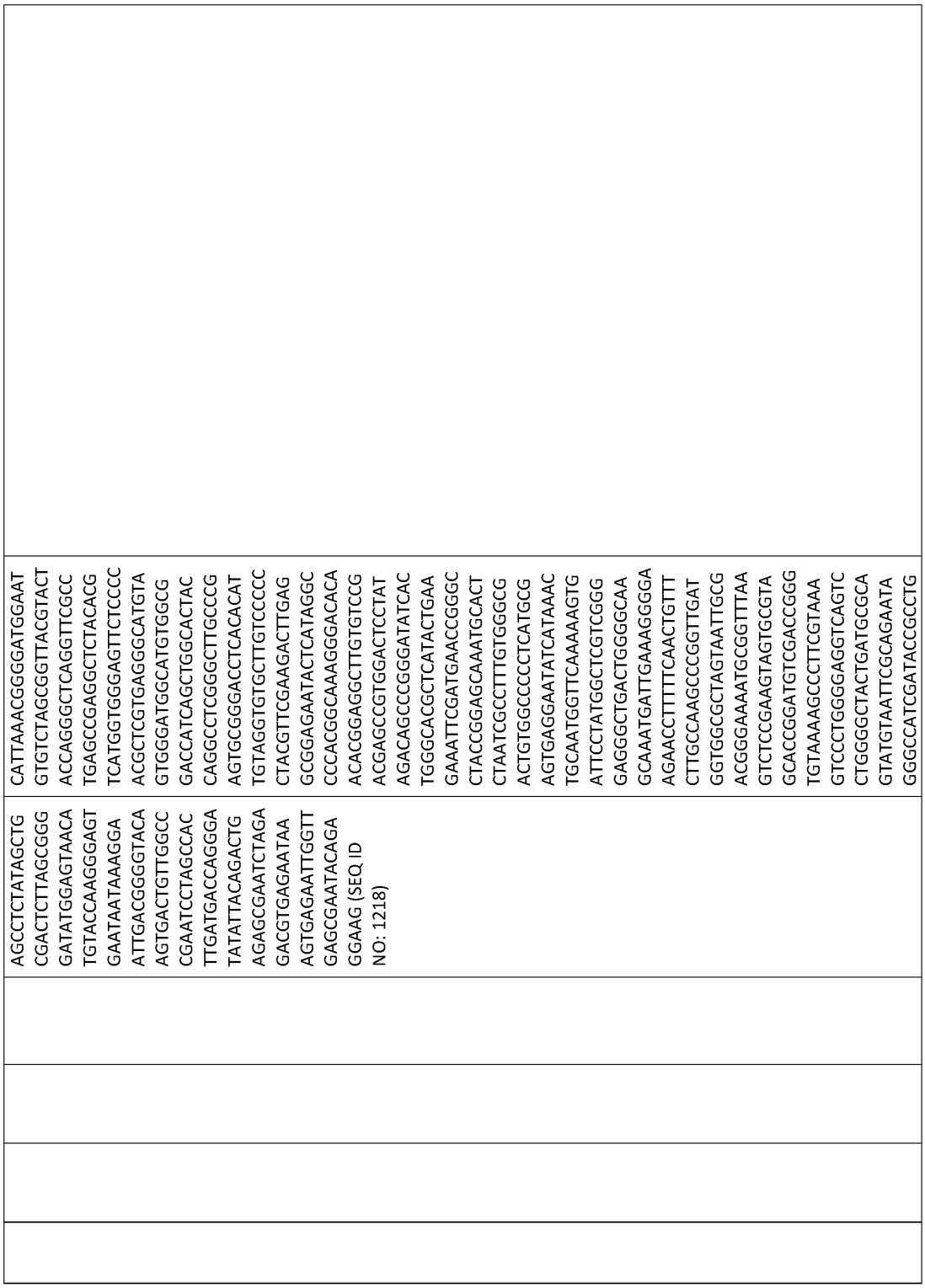

TTTTGTCAGATTTA
GCAAAATTTCATAT
TTTGTTAATTACCTT
AACTGGTTAAACGA
TCCCATAATTGCTTG
CAATTATTATAAAG
TAATTCAGGTAAAA
ATTACATATCTGGC
TGATCCTGCCAGTA
GTCATTTTACTTCCG
CCGCGCTATAAAAC
AGTTTAAAAACTGA
ATAGGAATCAAAAA
GAACATGGCAAGC
GACTATATGTAACT
GGGCATTCAACATT
CCCTATTACATATG
GTGGTGCCTGGGG
TCTGTTTTATATAAT
GGGTACCCGGGAA
GTGGATCTGTATCA
CCAGTCATGGTGCC
ATATCTTTKGATAA
AGATACAGTTTAAA
ACTGCGATGATACT
AATAGAGATCCTCT
TAGACCTTCGTAAA
GAAGTGGGGATTG
ATGACATTAGCATT
GGAAGAATTAAATC
TCCAAGGAAATGGA
GTAACTTCAATGAA
GTCCCACAACCCCG
TTGAAGGGCTGGG
TTCGAGTATCGAGA
GAAAACTCTAAATT
CTCTTCGGTTMTGT
CCAACGGAGGGGGA
CATTGTCTCTCTTTATGGTA
CTGAGGAAACTTATCGACT
CGCGAGGGATGAAACCCGT
AAACCGATCGATYTAGCCTA
TAAGTACCAGCGACAGTTA
AACCATCTTACGCGAGGGG
TAAAACCTGAGGACCGATT
ATGGTATAACTTCTCAAGAT
TAGCACAAAATGCGAGTGC
AACTTGAGGAGGAGGATTT
GAGTGTTAATTCATAATGTA
CTAATCTAATTAAACTGTGA
CGGGAATTGCAGCTTCGGC
TGTAATTACTTTGAGGCCTA
TCACGGATTGTAAGGAACA
TATTGACACCGTAAGTCTAA
CGTGTTCCCGATTTCCAACC
AGGTCATATGAAGGGCTGC
CCTTGATAAGGCGGATTTG
ACCCAATTCTTCATATGAGA
GGCTTATTCCAGCCTTCCCG
TAGTACCGTGAGGTTTTCCC
GCCTCGAACGGAACAATGT
TGCAGGGTAATTAAGTACA
TCGGGCTATATMGCGATAT
TTAACGTTTTA (SEQ ID NO: 1339)

R2	R2-1_SP	.	Strongylocentrotus purpuratus	CATTACTGTAAAATATCCTCTAAAAACAACT (SEQ ID NO: 1216)	TAAACCTTGCCTCCCCGGGC CCCCCTCAGTGACTAAGACA ACTTTCACCGTAATAATCAT ATATTTGTATACCATGTATT AATCTAGGTAACAATTGAA AGTAACATTGAACCGTATTC ATACTTTAATGAGTACAATA GTGAAGGGAACTGTATATT ACATACCTCGAGATAGAGG TTTTTGTACCTTAAGGGTTC GTGAGAATCCATACTGCAT AAAGGGGTTAACTTGTAAC TCAACCCCAGGGAGACAAG CCGACAAATTCGGCATTACT ATGTGAGGGTCATAAGTGT TAAAGGACCCATTGTATATA ACTTGTAAATGTACAGAGA ATTCGAGTTGAGTTCGAAG AATAAACAAACAAAAAAGA AAACAAAGAGAATTCGTTC GTCCATAGAAGCGGAAAAA GCGCAAAAAGGCGTCCTAA CCTCACGGTCTAACCCAGAA ACATTTAAGGGGGGAAGGA GCTCATTACTCCAATCCAAA ACCACGTTCCCTAGCCCAAA AGCGCGCTTCAATGAGGGA GCCGATATTGGAGGACAGA GTTCATCAAAAGCCACCTGT AAGTAGGCCCATTTCTGCCA AGGACATGCGCAAAAGGAA GCAGATTATCAAACCAGTC AAACAAGCACAAACATTGG GGGGATGGGATAACCCCG	MENSFAWEGTSSAEGRTTVEDSPSSSDDFVSNVGFKVAKADPTVWEEA NMSEDNTIIEDPPSSSDDFANNVGFKVTRADPTAWEEASTSTETEDLPSS SNFIDNVETQIDMAGPTAWEDADSNEDNIDEGTPNNINNNLAIVRGRA DAYACSCCERNFISLKAIGTHLKETHNKKVVFECAKCQHTFVKAHGLACH VPKCKEDADTPMLNRLLHGCGECGLAFNTRRGLSQHERHRHPSACLSTR RRSRLDGIARKKSLRNRRDIWTNDEIRLLKQLMIQYEHAKKINIKIAEHFN HKNAKQVMHKRRSLREKDMALGAPHDAPPPLAEEPIIEVVEGAREELEQ APVDVILPDLEALTVNDGRGGGSPVLTEGGESTRDMEENGGTDTRSPSP REERAGSTPWERGWQPRVDRGRGEYKGYGGERGDRHTVSLSQRGESG VDPLTPGGVVEDYDDSYLEDYFPGWDEDEHMHIIGRLDLSDESEQGEAA VSPRLGSFGDLLEEVNAMEGKDNLSEALAETLGLVLHEGHRVEYIKEKMN INVKQMATEILAHGANKGNPKRRKEAVAAKRNPGTRLDRAQRDNQAK AKAKEKKRIFSETQTQYKKNPHRLVEKLLDGKGDERCSVSLEVIQRTYMN RFSRESKEVDIGAYVDPETVEDNQGIVDPISKAEIERAISTTKKGSAPGPD GVTYDALKAYGNCQLYLLIMYNTWLAMGKVPSEAKTYRSILIPKGQGDP MDINNFRPLTLANVISRLYSKILTRRLDGAVSVCPRQRGFTHKASIEDNTLI LRELIMKSKRNKECLAVVLLDLAKAFDTVSHDLIIKALRRHRVHEHLISVIM DLYEGGTTSFTTDEGTTCPIAIRSEVKQGDSLSPVLFNLALDPLLATLEQRG KGVEIGGHTFVSLAYADDTALVSSSHLDMTANLDITVEYLNATGLSLNVR KCQGFLLTPINKSFLVNEAESWVVEREAIPWVEPGDTAKYLGVQVGPWS RPWPSIQPVIKRLTAYCESIDKAALKPRQRIHILTTYIAPRIAFEIAEGGYSTL VDCRGGIQYTRIREVDMTIRNYVRKWLFLPACLSNSFLYTRRGEGGLGLV SFYDYVPTERMRKLVRVCDSEDPVIAGAAASLGLRERAAKISAQTGLPVP VKPKGAHNAWRKVQKKKWKAQPTQGKGVSCYQHRLGNKWLGAPSFL TENDYIWAIKLRTNLVPTREAMGRGIIGRNQVECRHCHTTIETMGHISGY CQMVLDIRLIRHNRICKALIKAATATGLRVTEEPRIVGTDGKNYLPDLIFSA GAGEPCYVVDPTVVWDDDPKNLREAWRGKVRKYTPIIPAVEAMLHPSS VQIFGFVCGARGTWCPMNDDIAKIVGLKNSGISRTLQIVLCDTIRMVKAF MAR (SEQ ID NO: 1462)

			CTTATCTTCCATATA GATGCCAACCAGAC GGGTCGATGGGCA GTGAGAAACAAGA ATGGGGATAATTTG GAATTTGACATTCTT TCCTAAAGCAGACC TGAAGGGTTAAGA GTCACAAGCGGAGT CCGGACTGTCTTTA TAAACAGACAATAT TTTCTTATTTTACCA CCTATATGGGGATA TTTCCCAACTCGTAA TATAGGGCCCACTG TCAAGTGGCTGGTA TAAGTTACCCTGTG GGTAACAAATAAAT TCAAACAAAGA (SEQ ID NO: 1217)	GAAAGAGAGGGATCTTTAG ATAGTGGATGGAAGGGGT GGGACGTTGAATCCAACCA TGCCGTTTTTATGTTCCGA TAAAGAAGGATAAGGTCAC TCCAGCCTGACACACAAAG TGGGGTAAAAGAACTCCGC TCGTACGGACTCCAAATAG AA (SEQ ID NO: 1340)	
R2	R2-1_SSa	AGKD010724 55	Salmo salar	AATCTTTAACCCCG GACTCTTGGGGTTC TTACGACTCTGTAT GAGGAACAGTCGA AGAGAGGGCGCTA CCAATCCAAGTATA TGTCCCAAGAGGGC TGGGACAGGGTGG AAGAGTGCACCTCG CGATCTGGGGCAG GGAAGGAATGGAG AGAAGTCGAAGAA GGCTTGTAGAGAA GGGGCTCTCCTAGA TCCTAACCTGTATG ACGCCCGTAAAACG GTGACCCCAGTAGC GAATAAAGGAGGC	TAGATCCGTTTGTTATGATT GGAGGGAGCCTGCCGAGT GGTATGAGCGCTCCAACTA TTGAACCCATATGATTCCCG AGGCCTGGCCAGACGCCTA GATGCCTGCCACAATTGAA CGCAGCCCTAGCTTGCTAG GAGATCCATAGGAACTGGC CTATGGGGCGTCATGACGG TTGAAGTTCCTCCATAGCGT GCTTGGGAGGGGACGACA ATGACGAGTCATGACGTAC CGAGAGAACCCCAACCCAG GTTGGGGGAGAGAGCCAG CAAGAGCGGAGATGCTTGG TATACCAAGCTAGCAGAGA GAGGGTTGAAGAGGATGA CTACTGGGCTCAGAGTCAT	MSGKRIVEMSGCDEKICQNKHCLKRRWAWISGPKGETSPPRKRGTCEN VSFQDKSHASDPPDPLKAPEAREDAGSVAPQWVGEIKTPSLTSRDGVSEV VLPPQPVHAEGVSPASDSKDKATKITLLISLPVCDLRCGRCERPLETVGKA VRHFAVAHPTVSVVFKCQKCEKSSKNSHSISCHIPKCKGMTETRTDVEGD HGCDHCQEKFTTAMGLTQHKRHRHIVQYCKEKEGEMTARRKGEVEAV KWSEWEESEVARLSDGLAGLKMINRRIADSLGTGKTAEQVRQKRRRMR PEKVRCDKPKEAKDKSNLIKMLSIPSATPTPQTGLKGFLLGELNGVATKGE VQIGGVTLSLRGVEQDSALLNTSALELQRLLGGRAGSANPLSLQRERETTL PSERRKTKQGEYRRVQKMFRSNEKKIAKYILDGNGDGEAASPPLEIALAF KSRWEEVETFHGLGQFYSRGEADGVVFRSLISMSEVCENLGAIKNNTAA GPDGITKPALLEWDPTGAKLAAIFSIWLTSGTLPGPFKKCRTTLIPKTDDPI LLTQVAGWRPLTIGSVVLRLYSRILTHRLERACPINPRQRGFISSPGCSENL MILGGLIKRSWAKGERLAVVLVDFARAFDSVSHSHILEILRQRGLDEHIIGI VGDSYTDVTTTITVSGEQSPPIDMRVGVKQGDPMSPLLFNLALDPMIDT LERYGLGYRMGEQQITALAFADDLVLVSDSWEGMACNIRILEEFCRLTGL RIQPRKCHGFLIQKIQRARSVNLCKPWIVCGEELHMVGPEESVSYLGMK VSPWHGIMEPDPVERLCNWISSIGRSPLKPSQKVRMLNVYAAPRMTYQ ADHGGLGPIVLNVLDGMIRKAVKVWLHLPLCTCDGLLYSRCQDGGLGIV

		AGGTGACAATAGA	CTCACCCTAAAAGGCGGTG	KLACQIPSIQARRVYRLWHSKEAITRVVTRRTVEAEEYRGMWLRAGGSE
		GGGCAGGGCCGAC	GGGCATCGGTTGAACACCT	AGLPPLEDREEGAVQCTDTAGSVKPKNPVIPDWRRAEFLKWQNLTAQG
		TTCCCAGGTTTACAT	ACCCATACCGGGATGGGAG	VGVQVFGGDKNSNHWMANPETLGSKERHYIAGLQLRANVYPTREALSR
		TGTTGTACTTGTCA	GTGGTAGGCCGAAAAAGAA	GRPDLPKVCRQCLAGTESCAHILGQCPAVKDSRIRRHHKLCDLLASEAES
		ACATAAAGAGGTGT	CAGGAAGATGGTGGAGTAA	AGWTVIREMCCRTRAGALRRPDLVFVKTGFALVVDVTVRYEMAYDTLM
		CTCAATAGTTTGAA	GTTGAGAGCGGTTGCTCGG	GAAAEKVARYTPITPYVAMTLKARRVKVFGFPLGARGKWPGSNDRLLKA
		TCAACAAGGGAGA	GAAGTTATGTTGTGATAACT	MGVGGGRRKQLAKLFSRRALLYSLDVLRDFYRAEGETGDLDDESVDDHL
		GGAATACCGACCTG	CCATTAAGGCCGGTGGGCA	(SEQ ID NO: 1463)
		CTCCCTTGGGGCGG	TGGTGCGGATAATGGAAAC	
		GGGTACTGGTCTTA	TATAAAAACAATAAAAAGA	
		GCCCGGTTCCCCGC	AAGACCAAAAAAATGTTCT	
		AAGTTTCCTTTGCCT	GTTATGATGCCTTACACATG	
		GGGATGTGCCTGAC	TCTGGGAGACCCCATAAGG	
		TGGCTCCATCCCCTT	GTCTCCCCTTATACTTCACT	
		TCCCCATTAGGCAC	GGGAAACCCCATAAGGGTA	
		GGCTAGATGACGCA	TCCCCCTATATTTACTGGGA	
		CCGATGGGCGGGT	GACCCCATAAGGGTCTCCC	
		GTGTAGGTCGCTAC	CCTATAGATGTAGAGCGTA	
		CGAAGGGGACTGG	AGGGGTCTCCAAAGTACCG	
		GGGTGTCCGGTGA	GCCGATATGGCCTTATGGC	
		ACCAGGACTTCCCA	AAACTCTGGTGGTAGGGAC	
		AAATGGTCTCACAT	AAGGAGGTAAGGGCAGTG	
		TTTTAAGCGGCTTG	CCAACCCCTACTTGATCGGG	
		AGTATCGCCCAGTA	ACCATCCAGGGAATGCCAT	
		TCCTCGCGCGGCAC	CCTCCCGCGAAGGTGATGT	
		TGGGAACCCAGTCA	GGTGAGGTAAGGGGGGAG	
		ACCGCTCTGTGCCC	CCCGTCTTCGAGTTTCCCCA	
		CGGCGCAGGCGGG	ACCCCTACCCACAGGTGAG	
		GGTTTAATGTCTCC	AGGAGGAGAAGAGGAATC	
		CCGGCTTCACCGGC	TGTCCCCAACGGGAGGAGG	
		GCTTCGGCGACGAC	GTGAGGTGTAAGGGGGAG	
		GCAGAGGAGCACC	ACCTTCTAGTAGGGTCTTCT	
		CGGAGGCCCCCATG	CAGTCGCCTGACGTCCTGA	
		AACTTAAACCAACC	CTGTGGGGTGGATCAGTAC	
		TATCTTGAAATATG	CCTACAGGTGAGACCGGTG	
		GCCTCTCGTTCGGG	AGGTAAGGGTGTGGCCCTC	
		TGAAGGGCAGGTG	TTGAGGGCTGCGCCAACCC	
		GGAAGAGAGGGCT	CTACTCGAGGTAACCTGAG	


GCCTCACGATAAAC
ACCTAGTCAATAGC
CAGTCGGGAAAAAT
GTGGAATGTTAGGA
CAGGGAGGTAAGG
GAGGCGGCTTTTTC
GTAAGGCTCCTTCA
ACCCCTACCTGTAG
TCCACCTATTGCAG
GTGTTGACAACATG
CAAGATGACCTGCC
TCGTTACGGGTCGC
GTATCATTGCTACA
GGTCGTGTGCCGCT
TCTAAGAGGATAGT
AAGGAGAGGTTAT
AGGGAGGTCCTGTT
AGGGCTTCCTCAAC
CCCTCTCTATGCGAT
TCCTTACAGGAGTG
GATCGAGAAGTCCC
GGACGTAATACACC
CTGGAGGTAAGGG
AGTGGCCTTCTAGT
AGGGCTGCTTCAAC
CCCTCTGATGGGAG
TGTACCGGGAACCT
CGACTTGTAAGCAC
AGGTTAGTATGGGA
GCAGAAGGGGGAG
CCGTAATGGGCTTC
TCTTTCACCTGCTTA
CATAATACCTGTGG
TGCATGTATCTAGG
TCTTGGCGGGAGA
GTACTGAGAGACAA
GGTTGAGACCCCAA
GATTGGGTCTCCCT
GGAGTGGTGGAATGGCGG
CATGTTAGTGCTGGGACTT
GATTGCGAGGGTTTAATGA
GAGTGGCCTGCTGAGAGCA
ACACTTGTGGTGCTTAAAG
CGGGGCGGCCCATGACCAC
CGTGAGATAGGACACTGCA
CAGTGCAGCCATGAGGTTC
CTGGAGGATGATGCGATGA
GGTGGGGGCCTCATCAGCC
CCTCCTGGCAGGGCGTCGG
CCAGGGAAACTAAATGTCT
CTAGCATGTCAGTGCAGTG
AGGTAAGGGGAGAGCACTC
TAGTAGGGCTCTTCCAACCC
CTACCTGTAGGTCACCTGGT
CCAGGTGTCGATGATGTGA
AAACAAGAGCTACTTTGGT
ACCGGTCTGTTGCAAAAAG
GGTTCTGCAGAGGACGACG
GCTATCCCTATCGGGAGGG
AATAGTCGGTCCCAGGTAG
TGGAAAATGGGGCTTTCCA
CTGAGCATGAAAATGTGGT
AGAGGTTGCGTCCAACCCA
ATGATTTGCAGCAGAGCTC
TTGGACACGAAGTCTGTAT
AGTCCCATGCAGGCAGCCA
ACCAGAGAATGGTGGCAAG
ACCCCAGCTCCGTATGGGA
GGGGAGGGCCAAGATATAC
GGAACGGCTGCTAAAGCGT
TCTGCCGGTGTCAGTCTAAT
CACAGACAGCTGTGACGAA
ACAAAGTATGGGTTCCGAC
ATGCTTGGTCAGCTCTTAGC
CGCAAGGCTTAAATCGAAC
GCAGCCCGCCGAGAGTGAA


AGCCTCTATAGCTG
CGACTCTTAGCGGG
GATATGGAGTAACA
TGTACCAAGGGAGT
GAATAATAAAGGA
ATTGACGGGGTACA
AGTGACTGTTGGCC
CGAATCCTAGCCAC
TTGATGACCAGGGA
TATATTACAGACTG
AGAGCGAATCTAGA
GACGTGAGAATAA
AGTGAGAATTGGTT
GAGCGAATACAGA
GGAAG (SEQ ID
NO: 1218)
CATTAAACGGGGATGGAAT
GTGTCTAGCGGTTACGTACT
ACCAGGGCTCAGGTTCGCC
TGAGCCGAGGCTCTACACG
TCATGGTGGGAGTTCTCCCC
ACGCTCGTGAGGGCATGTA
GTGGGATGGCATGTGGCG
GACCATCAGCTGGCACTAC
CAGGCCTCGGGCTTGCCCG
AGTGCGGGACCTCACACAT
TGTAGGTGTGCTTGTCCCCC
CTACGTTCGAAGACTTGAG
GCGGAGAATACTCATAGGC
CCCACGGCAAAGGGACACA
ACACGGAGGCTTGTGTCCG
ACGAGCCGTGGACTCCTAT
AGACAGCCCGGGATATCAC
TGGGCACGCTCATACTGAA
GAAATTCGATGAACCGGGC
CTACCGGAGCAAATGCACT
CTAATCGCCTTTGTGGGCG
ACTGTGGCCCCCTCATGCG
AGTGAGGAATATCATAAAC
TGCAATGGTTCAAAAAGTG
ATTCCTATGGCTCGTCGGG
GAGGGCTGACTGGGGCAA
GCAAATGATTGAAAGGGGA
AGAACCTTTTTCAACTGTTT
CTTGCCAAGCCCGGTTGAT
GGTGGCGCTAGTAATTGCG
ACGGGAAAATGCGGTTTAA
GTCTCCGAAGTAGTGCGTA
GCACCGGATGTCGACCGGG
TGTAAAAGCCCTTCGTAAA
GTCCCTGGGGAGGTCAGTC
CTGGGGCTACTGATGCGCA
GTATGTAATTCGCAGAATA
GGGCCATCGATACCGCCTG

R2	R2-1_TCas	.	Triboliu m castane um	AGTCATAGAGCCAG AACCTCCTCGTGGT CCCGCTGGGCACAG GGATTAATTTTTCT GTGGCAAATTTGAC TGGCTTCAGAGAGC GTTTTTCGAAGTGG WCTGTGTGACTGC GTTCCCCCCTTAGTT GCTATWTCCGCTK MGATTAACATCTCA CCTCGACGTWTAA GATCATT (SEQ ID NO: 1219)	CGTGACTCGACTGGGTTTCC ACTTGAGGATATCCGACCG TAGCGTGCACCCTCTTGTAG TTGCGCCGGAAACGGCTGT GTTCCCTCACGTATGTGAG GAAACTCAACAATGTGAGT GGGTAAACGGCGGGACGA ACTATGGCTCTCGT (SEQ ID NO: 1341)	TGAAAAGAGTGGCAGTTGT GGAGACTTCCTTCGGACGT CGGGTCGGATTTTCGGACG CCAGGGTACCTCCACCGCT GGGTTCACAAACTAGGCGA ACATCTGCCGATACCCTCTT TAGGTCATAGGACCACATG TCTCTGCACGAGATTAACCC A (SEQ ID NO: 1342)	MSRRPGKSNEPPVRSRAMGLTTLSGTKTSNSGAQGPSTSAPMQNMAG GFVCDCGRSYALKTSLARHKKECGKNNAECRWCGTRFNTLAGTRQHER KAHFVQYQSDLAKALPQPESELMEKIAIVEARSXNGIFYKEMMASTGLTH QQVRSRREKPEYKGFLERARRSLAQTNIRAGSISPASTXAGSLESASPKAG CSSSASPGPTTRSRAPTKGVPXRSSNSARIVVEAQVHTRAPPNTGETEVA LRESRRTVPRLGXNPSRPCGISPLMAIAIDEDSVLGGLRVQAGPSPTAVHS VEAFPGTSSMTPMETDRVHNKSGIDPILEHNGTRQVRREESSTREDPVE QWSPNYPKTPVTMPNITTTADAXXTSYNRTPQTLPGNRRRRRSRSLPPVQ RKSASDXLESVDSLGPWAVFLQDQVDAGSLSGNDSLADLVRVALTKSDR GVLNDAVNRYLAQRAESLRIRKRGSKGKRKSKTGRHYGQTTSGSGQRAA LFKKHQDLFLKNRRGLAETILSGKEDFGPRPEPPVTSVEEFYGGIFESPSPP DNEPFEVRATGVEDPPPLTSPWTKRTLRSPCYWRRRPPPTYITMDEIKAA RAGWQISAPGSDQIPVAAVKTMSELELAILFNIILFRNVQPSAWGVLRTT LVPKDGDLRNPANWRPITISSALQRLLHRVLAARLSKLVSLSSSQRGFTEID GTLANALILHEYLQYRRQTGRTYXVVSLDVRKAFDTVSHCSVSRALGRFGI PSVIREYILATFGAQTTIKCGSVTTRPIRMLRGVRQGDPLSPVLFNLVMDE LLEKVNEKYEGGSLQSGERCAIMAFADDLILIADRDQDVPAMFDDVSTFL ERRGMSVNPAKCRALIAGAVSGRSVVRTGSSYKIHNTPIPNVDALDAFKY LGLEFGHKGVERPTIHNLSVWLNNLRRAPLKPDQKCLFIRQYVIPRLLYG MQNPQVTSRVLREADRLIRRHLKTYYHLNVHTPDSLIHASVSDGGLGIM ELRKAIPRIFLGRLVKLLNKNKDSVLSSVLQSNRVRTLMGKLSTMAGEVPE STFWRNQIASGPLSKGLEQAAEDSASRLWISEKPSGWSGRDHVRAVQLR TGNLPTKAIPSVPVGQRRCRHGCACDESISHVLQMCPLTHADRIRRHDE VVKKVARHCTSRGWTVEVEPHIRSRCGRLFKPDLAVHQPGGAIVIADVQ VSWDSESLTVPYERKRAKYDVPQFHQAAQHAWPGKALTFAPVIVGARG IWPRINNDRSAALQIPPVVRRACVNSVVKWGSSIHATFMRSVWANRLN PRPLRA (SEQ ID NO: 1464)

R2	R2-1_TGut	.	Tinamus guttatus	CTGGGGACCGTGG TTACAACCCGGGCT TAGCTGCAGAGACA GTACCTCCCCGTGG TTCCCGCCGGACCC CGTAACATCGGGTG ACTGAATCTGTCTCT GCCCCGGGAGTAGT TCCTCCTTGCCCTAT TGACCAGCGGTCGC CGGCTGCTCAATAG TATTCTAGGCGTGA AATATAGCGATAGT CCTAGTGGTTGTCT TACTGGGCCATAGC CCCTTGCTTCAGGG GTCATTCGCGAAGT CTCTCAGGAGAACT GGGGGTGGTGTTCT TCTGGGTATAGCTA AACCCCCTAGACTG TGTCCGATCC (SEQ ID NO: 1220)	TAGGGGGCTTGGCATTTCT CATTGCCTGCTCCTGAAAG GATATGGGTCCTGCGTCGC GTGGTAGGCAGACCCATTC GTCCGAGTAGGGGGCTTGG CAGTNTCCATTGCCTGTGCC CGAAAGGACGTGGGTCATC TGGTCTGTCTGCCTACACCT CTCTAGACTTGTAACATCTA GTCTGTCAACAAGATCAAA ATTCTTCACACAGACGACCG AGCTTGCTCAGTCTTCCTGT ACCCGCAGAATTTTGCTCTT GCTCTCCTTTGGCTGTGTCC TGGACGTGGGACTATTCCA TCTCGTCCCAAATGCCGCGT CCAATTATACCGGATTTGAC AAAGCGGACGGCCCGCTTT ATAAGCCGGAAAAGGTGCC TTGTAAAATTGCAAGGTTCA TTAAATAG (SEQ ID NO: 1343)	MGSWIVNFVSVATQTGEFPVDTARRAPVPVTSYPESECHXPLPLTFCNS DVTIWGGVRPEPVDCLGDLPEXYDALPGVAGPREXVGGSPPGEGVRSP GIASPSGTAVQHDFGSPILVPGAEAAEVSTPVVKVPQDHPACPCCGTRV VKVTALSEHLRRAHGRKRVLFQCSRCGRMNEKHHSIACHFPKCRGPPVE EGPLGAPEWCCEECGQKFNTKSGLSQHKRSVHPLTRNVERIEAARPKGK GKRGAHKGCWTEAEVAQLIELEGRFKNQRFINKLIAEHLPSKSAKQISDK RRQLAAATKTSSPEKRVTSSTSGESSPEVEKVEGIKREYRRRVGEWLCAGS LXDQTSFQKILEDVESGSEIVTGPLEELASFARGKLAAARVRHHRKHPAEA VPAREEQRWMKRRVGRRGLYLRFQRLFALDRRKLAGIILDDVESIKCPLP MEEVADVFRRRWEEVAPFTGSGSFRSLGKADNGAFKPMISAKEVMKNV XEMSRRSAXGPDGLSLRDLMKIDPQGSRMAELFNLWLLAGRVPDQVKA GRTVLIPKSADPGKIGNIDNWRPITIGSVXLRMFSRILSARLRRACPINRRQ RGFIAAPGCSENLKLLQALIKSAKRDHRTLGVVFVDLAKAFDSVNHQHIF QVLVQKGVDGHIIDILRDLYTNAGTYLESGSQRSGFIKILRGVKQGDPLSPI LFNLALDPLLCRLEDRGLGYKYGDQQIXSLAFADDLALLSDSWEGMQQSI RVVEEFCQRTGLRVQAPKCHGFLIRPTKESYTINDCDPWTIADMQLDMI DPGSSEKYLGLGIDPWIGLSRPELSEVLTRWVKNIGGAPLKPLQKVDILRS YALPRLLFIADHAGLSATCLHSLDLSIRSAVKGWLHLPPSTCDAIIYVSYKD GGLGLPRLASLIPNVQARRLVRIAQSEDDVIRSVVLQEGIQEEIRKVWISA GGRPEKVPSVTGEFPVMEAQAADEALSEWERRAPRTIYPIPCKWRKRE MENWTNLKSQGHGIRNFENDRISNDWLLHYGRIPHRKLITAIQLRANVY PTREFLARGLGEGAPRGCRHCPAEWESCSHIIGYCPAVQEARIKRHNDIC GVLAEEARKLGWVIFIEPHLRDNTNELFKPDLVLVKGSCAKVVDVTIRYES GLTTLSDAAAEKARKYQHLAGEVRALTSATTVDFLGFPIGARGKWYVGN NGLLSDLGFSTSRVVRIARALSKKALLSSVDIIHIFASRARQAQTSE (SEQ ID NO: 1465)
R2	R2-1_TSP	.	Trichinella spiralis	CTCCTGACTAACCT GATTTCGTCCGTGC GGCGGCGTTTTCTT TTCGCTCTCCGCTC GTCGAAATTTGCTG TAGTTGATTCGCTTT TCTTTGCGTTTTCTT CTACTTTCGCAGTTT TTTCTGCATTGCCAC G (SEQ ID NO: 1221)	TGAGGTTTTTGTTTTCTTTTT TCCTTTTACCATTCTTGTTCC ATTGTTGTTATTTGCTTTAAT CCTGTATTTTACCGCCGGCA ATTCCATTGTTATTATTACT GTTACTGTTATTATTGTTAC TATTGTTTTTACTTTTACTTA CTACTGTTATTATACTTTAAT TCGTTAACTTACGTTATTGT TACCACTACTTACTTTGCTCT CTCGCAAACGTTCGTTGTTG TTTCTTTTGGACCAGGTTTA	MSNRLANTAAAGGVPEKTSGTLDIPGQPSSSGEKRAISYPGPFGCNSCSF TSTTWLSLELHFKSVHNIRDFVFLCSKCKKSWPSINSVASHYPRCKGSVKA AVVPTSLANTCTTCGSSFGTFSGLQLHRKRAHPDVFAASCSKKTKARWS NDEFTLLARLEAGLDPACKNINQVLAERLMEYNITRGVEMIKGQRRKDQ YKALVRQLRSNSETQQCVGLAGSMDSNVPANDTSSSVASEVSITYPEYG AVMSCDLIKEATGMAIVDINELQSNLRKAFLSGRKLPMKFHGARETAQK KMANPRVAKFKRFQRLFRSNRRKLASHIFDKASLEQFGGSIDEASDHLEK FLSRPRLESDSYSVISGDKSIGVAHPILAEEVELELKASRPTAVGPDGIALED IKKLNTYDIASLFNLWLKAGDLPASVKASRTIFLPKSDGTTDISNCRPITIAS AMYRLFSRIITRRLAARLELNVRQKAFRPEMNGVFENSAILYALIKDAKVR SREICVTTLDLAKAFDTVPHSRILRALRKNNVDPESVDLISKMLTGTTYAEI KGLQGKLIPIRNGVRQGDPLSPLLFSLFIDEIIGRLQACGPAYDFHGEKICIL

				GAGAAATCGCACGCACAGC GGAACTGGACCGCTTAAGC CAGAAATAGTAAAGTAACA A (SEQ ID NO: 1344)	TGATGTATCCCTTCAATATA TTGTAATCCTCATTCGTCCC TATCCTTTCATTTGATAAAA GAAACTTTGTTGCTCCTTTT AATAGTTGGTCCCTCCTGTC CCTTTTCTGGAACCTGGTTG TATCGATTATTGAAAGTTGC AATAAACGGATTTTAA (SEQ ID NO: 1345)	AFADDLTLVADSAAGMKILLKAACDFLEESGMSLNAEKCRTLCITRSPRSR KTFVNPAAKFIISDWKTGISSEIPSLCATDTFRFLGHTFDGEGKIHIDTEEIR SMLKSVKSAPLKPEQKVALIRSHLLPRLQFLFSTAEADSRKAWLIDSIIRGC VKEILHSVKAGMCTDIFYIPSRDGGMGFTSLGEFSLFSRQKALAKMAGSS DPLSKRVAEFFIERWNIARDPKVIEAARRVYQKKRYQRFFQTYQSGGWN EFSGNTIGNAWLTNGRARGRNFIMAVKFRSNTAATRAENLRGRPGTKE CRFCKSATETLAHICQRCPANHGLVIQRHDAVVTFLGEVARKEGYQVMIE PKVSTPVGALKPDLLLIKADTAFIVDVGIAWEGGRPLKLVNKMKCDKYKT AIPAILETFHVGHAETYGVILGSRGCWLKSNDKALASIGLNITRKMKEHLS WLTFEIIFITQISRIYNSFMKK (SEQ ID NO: 1466)	MCILGGLTSHSREGGLSRGSSQLKTVKPQNEEDNGTTQLKAGSADSFPR PSGDLNPEEPLSIDICPVCFRQMKSYLGVRVHMQKMHLEEYNASIPDPV VSHTRWSDEEAAQLAFTEAKIEVDKLLPRGKGINKFLLELLPGRTLESIKSH RKRQSHKDLVRKYVKEFVDTLAADNDDDTIICQDNGDIFNDPIVGATDS QSETETVADPAEFKTFIELADDPTKPKVVAKLRNLIKDKPKSEILGSDILVRI LRRTLHGLPVEDELDQYLEVYFTGKIKQRRSKTQTALSKKQIKQRDYGRLQ ELYSRSRKRCANEILNPTSMSGGFGHQELSEFWTKTFGPDEQPTLGEVEII PKENCWWDIFSPISSDEIKASYPSIGKAAGPDNFSAYQLRKVPVWHLECL YNIFAFYKDIPSRLKDAKTILIPKKDNAESPGDFRPITLSSIITRHFHKILATRV NNFVRFHPMQRGFIQSDGCLENTALIQTVIREAKVRRKQVHITFCDVRKA FDSVRYDSIIAAIAKKGAPGSFIMYLSNLYRGNKTTLLTAGGETRITPTRGV RQGDPLSPILFNCVMDQILTALPSRTGFTLSAGDESVNVNCLAFADDIILIS KTKNGHQELLDVTQRILKENGLDLNPDKCCSLSLIPHSKTKKIKVVRADFV VNGVKVRSMSIGDSTCYLGVSINVTGQVAPVKMYQALCEKLDSAAIKPH QRLYILKHFVITKMFHPLILSTIAAHKIKNLDLISRRYVRKWLHLPHDCGSG MIHAIKVSDGGLGVPLLFRTIADLKVRRKEKLQVHENPIFRILAKLSTVSKEL ENCKKIASKTTDIQEKTFKEMLATYDGLSLKEARAVPEVHIKWVDSYDKRY KFAGRDFVQVIQARFNALPTRSRVWWRGRGADEKSLRCRAGCNARETLN HVSQSCFRTHRVRTARHDKILDFICERLDVVGVKYVREKPISFPGKKLIPDL IVENTDQALVLDLQIVGDNSELPLDERGKNKVIKYNCSEMQELYKRKKKT LAVKALTLHYKGLMAPETSNILRSFGFKSKDLEKMAYMALFGTVAAWGIF NRSTETMRSVANWPRPEEL (SEQ ID NO: 1467)
R2	R2-1_TUr	scaffold_6	Tetranychus urticae	CTCTCTTATTTTAAC ATATTCGATGTACT CGTACATTGAATAT GCTTTTATTTTTTTT CAAAGTTTTTTGGG TGCATACCCCTGGA AAATTCTGAGATGT ATAAATCTCCCATC AGCTTTGGCTGAAA CGTTGGCTAAGTTT TGTAGGTTGTTTGC CCCCTACTACTTAGT CGCAAATGGTATTT GCTAACAGTTGTTA AATTGTTACATTTAC AAGTCCTATCCAGT GCCTCCTCGTGGCG CTACCCGGTAACAC TTAGAGTAATCTGA GTGGCTAAACTGGA AGGGCGGAAAATG CAAACAGGCGGTTG GTAGATGCTTCGGC ATTTTGCCAAAAAT CCACGGCTTTTTAG CCCAACAACATCAG GGTGATGGACCCGC CAGCTTGTGGTCAG			

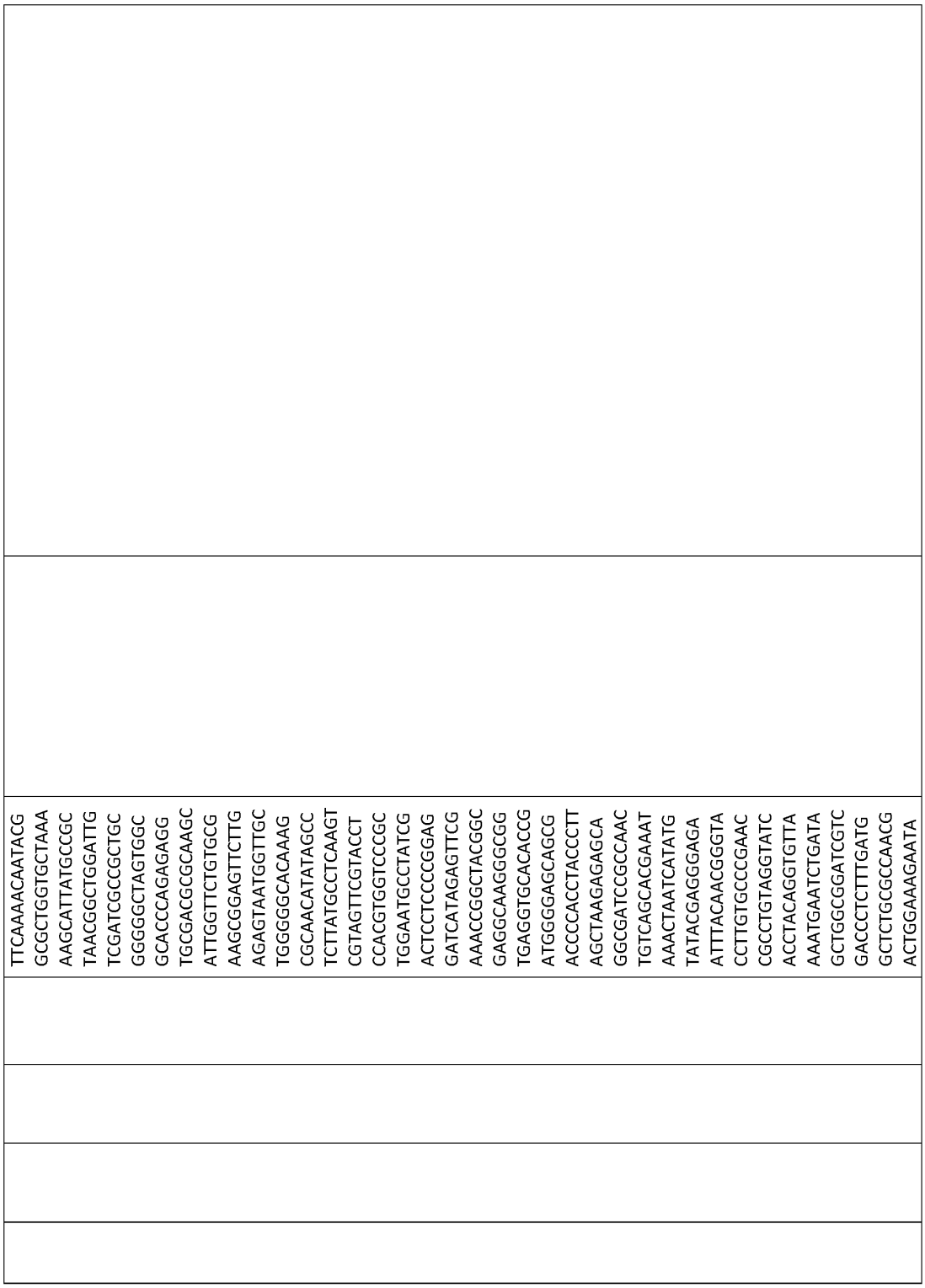
				GATCCCATCCATGA ATAAAGCATGGCTC TGCTTCTGGTGCAT CCTCAACGGGATCG GCTTCGGCTGGATG TAAGTCTTGCGGAG GC (SEQ ID NO: 1222)		
R2	R2-2_DWi	.	Drosophila willistoni	GAAGCTGGGAAGC TGGGTCGGATGAG CGCAGAAGGGGTG TTCTTCGGAGCACT GTAATTCATAAGTC GTAAGTCTGATCAA GTCGACTCGGAACC TCTTCGTGGTGTTTC CTGGGTGCTGTTGA GTTCCTAGTCTCTA GGTTCTCTCCAGTA GCTAA (SEQ ID NO: 1223)	TAGATGTACTAACCTCTAGT TTCTCTATACTTTTGCCTGCT ACCTTGGCATTACATCTAAA AAGGTACAAACATCGCATT GGCAAAAAGAGGTGGTTTT AGTACATAGGCGCTGTGGG ACTTCATTGTCCCGATGATG CAGCGAATCGTGCATACGA GATTGTCCAGTAGTTGGTT GCTCGTATCTTTAGAAGATT TCCTTCCTCGGCGATCAAAA NAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1346)	FERRSNSWGYQNLEPSNVGQDMNTVPRINNTTTTPATSRPGDQPREAI AVVNLAGEIPCAVCGRLFNTRRGFGVHMSHQHKDELDTQRQREDVKLR WSEEEAWMMARKEVELEASGNLRFPNKKLAEVFTHRSSEAIKCFRKRGE YKAKLEQIRGQSTPTPEALDSITSQPRPSLLERNHQVSSSEAQPINPSEEQS NWEIMRILQGYRPVECSPRWRAQVLQTIVDRAQAVGKETTLQCLSNYLL EVFPLPNEPHTIGRSNLRRPRTRRQLRQQEYAQVQRRWDKNTGRCIKSL LDGTDESVMPNQEIMEPYWKQVMTNPSTCSCENTRFRMEHSLETVWS AITPRDLRENKLKLSSAPGPDGITPRTARSVPLGIMLRIMNLILWCGKIPFS TRLARTIFIPKTVTANRPQDFRPITVPSVLVRQLNAVLASRLASKVNWDPR QRGFLPTDGCADNATLVDLILRLREHHKRWKSCYLATVDVSKAFDLVSHQAI IKTLQAYGAPTNFVSFIEEQYKGGGTSLNGAGWSSEVFIPARGVKQGDPL SPLLFNLIIDRLLRSYPREIGAKVGNTMTSAAAFADDLVLFAETPMGQTLL DTTLGFLASVGLSLNADKCFTVSIKGQAKQKCTVVERRSFCVGERECPSLK RTEEWKYLGIRFTADGRAQYSPADDLGPKLLRLTRAPLKPQQKLFAHRTV LIPQLYHQLTLGSVMIGVLGKCDRLVRQFVRRWLDLPLDVPVAYFHAPH TCGGLGIPSIRWIAPMLRLKRLSNIKWPHLEQSEVASSFIDDELQRARDRL KAENVQRCSRPEIDSYFANRLYMSVDGCGLREAGHYGPQHGWVSQPTR LLTGKEYLHGVKLRINALPSKSRTTRGRHELERRCRAGCDAPETTNHILQK CYRTHGRRVARHNSVVNAVKRGLERKGCVAHVEPSLQCDSGLNKPDLV GIRQNHIYVIDVQVVTDGHSLDQAHQRKVERYDRADIRSQMRRFFGVT GEIEFHSVTLNWRGIWSGQSVKRLIAKDLLIAEDTKLISVRAVNGGVTSFK YFMYCAGYTRS (SEQ ID NO: 1468)	
R2	R2-2_PM	.	Petromyzon marinus	CGGTGCGTTCCCTT GGGTAAGGAACAC GAGTCTTAGTGGCC TTGACCTCCACGTG GTCCCGCTGGTAAC ATCATCTCTTGATG ATGGCTAACAAGGC TAATGCACCCATTC	TGAGAATAATGAGGTGCTA ACCTCCCTGGGCCTGACCA AACCCAGAACACATCACTG GCCAAGATGATTTCCCGCA GCACGTTGCTTTTCTCTCTC GATACCCGAGAATGTTCTG CGGAACTACGCAGTCTATG AACAGTCACAGACAACCTCT	RPQTKLMTDKLKFSSQLARGLAKQRAMDGARVGDPPITVRPTETDLCNT EGSWGRRPMKLLFVSVSTQTQNEDALWASDVAKPMASRSALKMTSIPS MTFHNSSLEKEEEMNYDFYEQIKSLVESDDSSDDFTEDDEDVEESFLDISA EEPVLGKFPIDTKGTITVVLPSLEYICVICKQHMGKASELVAHFNIKHRDIPL VFKCAKCDKTNSNHRSIACHAPKCGGIKLTEESLPMVCECCQARFATLSG LSQHKRHAHPVTRNEERIKDGIKGTSQRGVHRSCWSLKEVEQLALLELQF QGKKNINKIIAEALGTKTNKQVSDKRRDLSKKTGAPMSDSLHFSSRPLETL SPPPNVTTGTSSILAQAAERLTNENSGTLEKPAMEAIKAWLNGEGQHDA

				CATCTCCTATCTCGC ATGGAGGCCGCTAT GCGTGATTACTAG (SEQ ID NO: 1224)	GATCCAAG (SEQ ID NO: 1347)	LVETATALMLCPMRLVKNKGKRSKPENDIIKPRILPTRSWMKKRAEKRGS FMKHQKLFFKNRSLLASLVLDGTERHECRIPNADVYRFYCEKWEKVLPFN GLGQFKSSGVANNEYFEPLISVEEVQTAIRAIKPTSAAGPDGLTRAAICAA DPEGRTLTALFNAWMITGIIPKELKKNRTILIPKVMDDEKLKELGNWRPIT IGSMILRLFSRIMTARLARACPLNPRQRGFIAASGCSENLKVLQDLMRHA KKLHRPLAVMFIDIAKAFDSVSHAHILWVLRHKKVDEHVVGIIQNAYDRC TTSFKSNGESTREISIRVGVKQGDPMSPLLFNLAMDPLICTLESHGVGYSI DTDHVTALAFADDLVLVSESWVGMAANLAILESFCGLSGLEVQARKCQG FMISPTKDSYTVNNCDPWTIKNKDVHMIQPDESTKYLGLKICPWTGIIRS DLHVQLKTRISKIDEAPLKPTQKVELLNAYALPRLLYPADHSDCKQSTLRVL DQEIIKAVKGWLHLPASTCDGLLYARARDGGLAILKLENAIPSVQVRRLLQ RIANSSDAIARNIASSQGVEEEYRSLWVRAGGDSEAIPTFFLRGSESKEPV YPRPCDWRKRESRRRCEKPVQGRGIVNFAQDRISNAWLGPRCGFKQCF FIAALQLRANIYPTRESINRGRDGASRSCRKCSARLESLSHILGQCPAVQKF KDCATQTEAEMVPHGPAGNALPGWSLSRTFLVNVPQYKNSRIARRNKIS DILADEAARLGWWVYKEPRFTSEAGELRKPALVFAKGEEALVIDVTVRFE LSRKTSSEAASHQVAYYTPPCDQVKVLTKASNVTFFGFQVGARGKVAP (SEQ ID NO: 1469)
R2	R2-2_SMed	.	Schmidtea mediterranea	AGTCATAGGGTGAA CTGCAATTCTGACA CGATGACCGAGCTG TGTCAGTTTGCAGC TAGTCGCTAAAGAC TCGATCAGTCCGCC AAGTGAGGTGGCC GGGTATCTGCAGCA CTAGAGCCACTGGT ATCAAGAGCAGAG ATACGCGAGTGGA AGTTGAGTACGACT ACCTTCACGGGGTC CTCCTGATAACCAC AGTGGACTGTGGG AACTAAATGTGTGC TCAGCGTTCCCTACT TTCTCGTAGGGTAA AGGGTATGATAACC CAGAGAATATCCCA	TAGAAGGGAAACAAAGGA AAAACGAAATGACTGGAAA CTATGAAGGATATAGCTGA AAGCCGCAAGGAAGGCTAA GTCCTGAAACCGATCTACAT CTTCGATCCCAAGAGGAAC TGTGGGTTAAGCTTGAGCC GACGGAAAAAGCGAATGCA TGTTAGACGACGAGGTACA GTCACCTCCTCGTGGTATTT GGCGGGCAATGCTCACTAA ATTAACTGTGAGTAGCTGA GAACTGTATGTGTATCATG AAAAAAAAA (SEQ ID NO: 1348)	RKELVTIKNLFEESGATAPAPVPLEVAVEVHQSSSVPEITDESTTTQEGSYS EPPIHRCENCGREFRTRAGVQQHRRKAHTNEFMEEKEKAAPTKKLRWT DEEKEILIESEIKIIKEGSLKEQHEINKILASRMPGRSQDGIAKIRQKQEHKA EIQRRLHGTVTTNETRGNRTSEITEPIRSLPINTKTWSEDEMKRMLAEEVK LRTKNEKDINKKLAEIFPNRTMGSIKSKRTKDKDYQQDLVKLTMQTISENPD NETDFNTSNTENNSTDAEKEVKNYLNMLLLTINEEEWLTSTLKEAATLAL QGKKTEASEKLNEYASKTLFPGLKITNQTRKREKKISKRETRRQEYAEIQKL YKKNISSAAEKAINGKWSIKPEEEYHNNKDLIKAWKPILEAPPFSDCRPIEN IKEMDYALMEISTAEIFLAIRAMGKTAPGPDGIKYSKLKKNIQSMAILFNT CLLTSFLPLPLKIARTILIPKQENPGILDYRPLTIASVVTRVFHSILAKKLDNN AQLSQRQKGFRKCDGVAENIVILETILTNSRSEKRPLCMAFVDLRKAFDSV GHESIIRGAKRVGVPPMLLEYISSSYQNASTNLFGEILNSRRGVRQGDPLS PILFNFVIDEALENLNRNIGYLLKEEKVSCLAFADDIVLIAETKGGLENHIEK LLEKLNGAGLELNASKCATLMVMKNGKEKSTYISTKAIKIKENDIPTMKAT ETYKYLGLQMGFKAREQNANEVITEGLENITRAPLKPQQRIHILRDFLIPRL IHKLVLGRVAKKSLKRIDQNIRKKVRNWLHLPKDTTAAFIHADAGDGGLG VPALEHTIPLLKRERITNLRKSNDPVTKECLRMEYTKQVLGKWSRPTKIGE TLATNKSQLKEAFRKQMLITLDGKGLKDHHETPTIHKWIRRGENMTGKQ FITAVKIRGNLVATKSRNSRGRPEQEKLCEAQCGRPDSLGHILQGCWRTH GMRVERHNNICRRIKAIMKGKESEVVEEPRLQTNEGLRKPDLLICHKGKIII

			TGGGAGATATCCAT GGAAAAAGCACCA CGTTAGACAATCCG ATGGTCTAACTCGG CTCCGAGGGGGCTAA CTATCCCAAAGGGC TTAA (SEQ ID NO: 1225)		CDAQVVADSSNCSLESENQRKIDYYKKDSVVSEARKLIGRVDEDIIIMAVT FNWRGAISKTSIRDLDMLLDIKSKEVIKMSRKIIRDNSIMVEMHRNRTEK RR (SEQ ID NO: 1470)
R2	R2-2_TCas	.	Triboliu m castane um	TGGAAGACCCCGCC CATGAGGCTTGGAG AGTGTGATCCTGAT CACACTTGAAAAGT TATGCTGAGTACGT CGTGAGAGTCGGTA ACTGTCCCAGGATG GTCTGGGATAGGCT AAACCTCAGCAGGG GAAAGTTGTAGGG GCCTGCCACCCCTA CACTTTTATAGATAT GGCATTCGATACCT CAAATAGAGCCTCG GACTTGGAGGAGC ATGGTTCCCCTCCTC CTCGTACTAGACCT GGAACCAACGGTCT TGACAACCCCATTG GACCTACGGGAGC GGACCATGCTCATG GACATGGATTCCGA AGACGAAGCGGGG GAACACGGACCCCC CGCCGATAATGCTC ACTTAACGTCAGGC GAACCCATCGAAAT CATCTTGATGTTACC CTTTCAAAGCAGGT CATGCGGCATATGT	TGATGCTCCTTTGGTTTTAC CATCTGTGGGGGCATCGGT CCTCACGGTTTCCTCGGGTT TCCTATTGTTTTTCCTAAACC CGACAAGGAGCCCTTTGGC CCTCCTCCTTAAACACCTCT CCTTCATCCTSTTAGTCCATT CCGGCTAAAATGATGAAGA CCGAGGAGTGTCACTCTCTT GGCGGGGTTAACCCGTCCA AGTGTAAATGTGACCTCGC CATTCGGGCTCTGATA (SEQ ID NO: 1349)	MKSRSFRRIGDCAAGSSRRGVRLTGKAGREGRFAASPHLSPRYLAGSVS GNVPSVPPGPGLGAGAPAFAAGRNADGGPAQNPCPYCARSFTTANGR GLHIRRAHPDDEANNAIDIERIHARWSHEETAMMARLEAGAIQRGGVRF MNQFLVPRMPGRTLEAVKSKRRDATYKALVQRFLQAPQINLPELRDGD APRQPPDPQQENPPEPPSFDGAIRGAVADLVGGVDWQRLGFQGDRLCN IARRACDGGDVSGQLLGWLRDVFPVKRVSTRGDQSDLDVDGALVSRRT ARTREYARVQELYRKEPKACLARILGDRREGANRAPNRDPAFIDFWRGV FSEASAEVEGWAEEVSDHGELARRVWDPISVEEVGRSRVRNGAAPGPD GIAVSVWNKLPPEAAALLFNVLLLGRCLPAELTRTRTVFIPKTDAPRTPAD YRPISIASVVARHFHRVLSARVQRIPDLFTKYQRGFLSGVDGIADNLSVLD TMLTMSRRCCKHLHLAALDVSKAFDTVSHFAIVRACRSIFGSAETVLEEG GRRHFVQVRXGVRQXDPLSPLLFNLVLDRALKRLSTDVGFRLTDATKVTA LAFADDVVLCATTARGLQTNLDVLEAELRLAGLLLNPNKCQALSLVASGR DHKVKLVTKPTFKVGQNTIHQVDASSIWKYLGIQFRGSGMCGCGSEGV AAGLKRITCAPLKPQQRMHLLRVFFLPKFYHAWTFGRLNAGVLRRLDVV VRTSVRTWLRLPHDIPVGYFHAPTKSGGLGIPQLSRFIPFLRLKRFDRLGRS AVDYVRECAFTDIADRKIRWCRERLSGIVDQVAGGRDALDAYWTAQLH QSVDGRALRESASVASSTQWLRCSTRAIPASDWLHYTAVHIGALPSRVRT SRGRRGGQDVSCRGGCLLDETPAHCIQVCHRTHGGRVLRHDAIAKRISA DLMELGWIVTREVSFRTTAGVFRPDMVAVKEGVTVILDVQIVSPAPTLD EAHRRKVAKYRDRADLARYLAEAAVARGRAPPANIRFASATISWRGVWS AESVGSLRELGLSARHFDRYTTMALCGSWRNWVRFNASTASRMGRGR GDASPRRHENQQ (SEQ ID NO: 1471)

				CTCAATGCCGGAAA GGGTGGCTCCCGG GCGTAACGGACATC TATCGCTGATGGAA SCTAGCTCGCTCTTG AAGAACGGCGGAC GGGGACTTGGAATC GTGGTGTGGTTCTG ATGTAAGTCCTGAA ATTATGGCGTGATG GCCCGCCCGCCCGA CCGGAGGGACTTA GAACCCCCTTCCGC GAGGGTCCTGTCTG TAGGTCCWCCATCT CCGTAAAACGAGTT GGAGGAAACCGCA GACGGGG (SEQ ID NO: 1226)	TAAATGTCGAACCGAATTTT GGGTAACGTGCACCCCACC ATCCTTAATCGGCAGCACG CAATAAAGCCGTGGGCAGT GGTTTTAGTGGGTAGTCATT AGGAGTCCCACAGTACCCA GCGAACATCTTAGTGGGTC TGCGTAAACGCATTTCCACT GCCTATCCTCCGGGAAAAA AAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAA AA (SEQ ID NO: 1350)	SGPATSTFGETKSRLCEPTSALGCRPGAVVIQWAQIHKEKRKRIVGWPLG HLGSPTSLKLRHPRLQAKRIVPVLAELMQCLCARHVSGRSPQKSAWAFTL EERTSACGVDAMFVCSTCQRSFATKIGLGVHVRRAHVEVANAAISVERV KDRWSEEERRIMAAVEVRGVLSGARFINEYIMSHLQTSRTLESVKGTRKN PKYKELVATLLEEARTSVREESPRSAVNDSATQPSGPSDTRSLRTEHLFTES TEPFEHRIRELIGDLEGVTDFRAELLVSIAEQQLQGDEVAESLTRWLGEVF KPENQQQQVQRKRRRQRKAPVSGQLPKWRERRRDYAAMQTLFHRNP SLAAGRVLDGKNESRPPDLPEMTAFWEPILTEQSAEHRAVGPASEKSELC SVWGPVEKEELLSSVPPLDTAVGPDGVTARQWRAVLPAVRALLYNIILKR GSFPASMLESRTVFLPKKQHSVVNPADFRPISIASVVVRQLHKILAMRLRRT NLVDERQRCMDDGCAENITVLASLLDDARHGLKELHLVSLDCAKAFDSV SHHAIDATLKECGLPAGFVQYISRTYSDSSTRLEVGRNRSEPIKTNRGVRQ GDPLSTLIFCLCFDRVARTLSPHIGYDLNNTRISTLLYADDAFLVSTTAPGM NILLRSVEESAGEVGLSFNTSKCSALSLIPSGKEKKMKVGTTPTFKTSQGFI TQITPSQEWRYLGVDFQYSGPKKASRSLKIELERISKAPLKPQQRLLILRVY LLPRYYHHLVLSRTTLGHLRGLDLQVRAAVRRWLSLPRDIPIAYFHTTAKE GGLGLPAFETSIPCLMLARLRSMETSTCKAARAAVQGFWVQKRIHWAT AALTKNGEALTCKADVDRWWASRLHKSVDGRELRECSGVGSSSTWVNS ALNITGRDYVQYHHVRINSLPTRIRTSRGVRREGMEVTCRAGCQVTETA
R2	R2-7_MR	.	Megachile rotunda ta	TCTAGTTAGCAAGC GGCCCCCTCTAA (SEQ ID NO: 1227)		

						AHVIQSCHRTHGGRILRHNAVCKVLASGLRDKGWEVREEPKLRTRQGLR KPDIVAIKDGVARVIDAQVVSGSGPLDEAHETKRKYYSDNGDVTAAIARE CNIAPSNVAYSSCTISWRGVWSPRSAADLLQVGLSKKLLGFITLRVLRGSH LNWTRWNKMTTMRVHHQRTGIG (SEQ ID NO: 1472)
R2	R2Amel	.	Apis mellifera	TGGTAATCAAATGC CTCGCTATTTTAGTA GCGGTAGCGCTCCG CCCGCGCAGGAACC ATTGACGCCGCCGT AGTGTGGGTGATTT TATATCCAACCAAT CACGTCAACTACGA TCATTTGTAATCACC GACGGTACTTGGTA GGGGTACCACATG GGCATTCTTGCTCA TTCCACAACGCCGC CTCCATCATGGCAA CAATTTAAAATATA TATAAATTCTTAAG GTTTGACCGTATTC ATATATATATATATA TATTTAATATTACAA CCATAAATCTTATAT CGAGCCTTCTATTT GGTCTCAAAAGCAA TACGTTGTCAGATC TTGTAGAACATCAG GAGTGAGCGGTGC GCTGTGGTATCCGT GCTTTGTGCCGCGG CGACAAACCAATAC GCTGCTGCCTGTCG CAAAGCAATACGCT GCTGATTCTCGGAT GCGGGTGTCGACG GTCACGCAAAGCGA TACGCTGGTGGGGT	TGATCGTTAAAAGTAAAAA TCTATTTATTTATTTTTATTC CTATATTATAACACATTATT TATTTATTTACTTATTGTTTT AAAGATGACGAAGCCGCAA GGCCAATCCAAATTTAACAA AAGAACGAGACTACTGGTC GACATTAAAAAGACGAAGC AGCTGCCAGCTGATAAACA ACAGAGCCCGTCTCGGCCT TTACACCGAGCGGTGCAAG TCCTGACGTACTATTGTACG TCTAGGGCGCGGGGCAGAT TCTACCGTGTAGAATCTGG GGCGACGCCTCCGCGAGGC ACTCCCTGGACAACGTACG CTAAAGCGTACGGCTAAGT GCGCCTCCCGAAAGGGTCC CCGTTCCTAATTTTTCCGAG CCCGCGGGCAGATCTCGTG GCAGTGACGCTAGAAAGTT AAGTCCGCGGACATATAAA ATTACAGCCTTAAATAATGA ACCCCACGAAGGAGGTATC CTCGAAATTCCGCCACGATC CTTCTGATCGTAGGCGCAA AACA (SEQ ID NO: 1351)	MSSNEEGASDTGAPGPGVPVADVSAADGRATYDDHGMSTDYEKQTIEL PLNGQIQCLWCHIEGRNQRFLQESQYLKHKDTQHPKGEIIWRCAACQKE FEKLHGCRCHLPKCKGRKEAKGVAKFKCDSCEESFLTQRGLSMHELHRH PAIRNLKRTQGTSRGNTRPINRASVWSKEETDLLIKLNERYKHLKQPNVAL KEYFPDKTLKQISDKRRLLPVQEPEDVATTDETGPPPSDSSEESIYESATED EGGGDMQQTAPNDSWKEPFIQSIRTNHLEEEDSLRKVEEAIERMAMNE GVTEQEVGTLLEQFVDSLTQSPTTERKGSRRKSQKTTKRKTTHNNRKKFL YAKHQELYKKSPRRLLELALSGESSSGREVVNLPEADSVGPLYKSLWGQIG PEKTHRNQPMCNNIDMSEIWTPIALESLVEKFKKIKSDTAAGADQIKKFH LRKKGALHVFAKLCNLLMLHRIYPAQWKTNRTTLIPKPGKSAEEVENWR PITIGSLLGRIYSAMIDRKLRSKIKQHIRQKGFTQEDGCKNNIAILSSALTKM KEDSGGIITIIDISKAFDTVPHGEISQSLMNKGVPSPICEYIQKMYIGCKTIIY CRDKKTLPVDILRGVKQGDPLSPLLFNLIIDPIIGTLDETTEGIKLENENISVL AFADDLVLLAKDKETADKQNRLINEYLDDLKMKVSAEKCTTFEIKRQNKT WFLGDPQLTLGQQRIPYADPEAAIKYLGTNFNPWRGLCKTSIKEIIDAART VKQLKLKPHQKINLIRTYLLPRYIHKLVANPPPLGTLDLIDKELKTIIKEILHLH PSTTDGLIYTDKSHGGLGIQRVANIVKLAKLKHSILMTRSEDNAVKIALNG QEGMVKRYATSIGLQWPCGIEEIEETRKKLKRADTNKWKTLISQGQGIKE FFGDKTGNAWLYNPEMLRPSRYLDALKLRTNTYGTKAALHRAKRDIDIN CRRCGVQVETLGHILGLCTHTKNKRIKRHDEICDLIAKNVSKEYVIFREPEV EVNGDRRKPDMVIKDHDKVYVVDVTVRYENNDSLNKAYKEKENKYKET AEIMRRDLKAKESRVLPVVIGSRGAVPRATIENLKVLGLQTKHALTASLIAL RSSIEMANEFLDYDHTT (SEQ ID NO: 1473)


TTCAAAACAATACG
GCGCTGGTGCTAAA
AAGCATTATGCCGC
TAACGGCTGGATTG
TCGATCGCCGCTGC
GGGGGCTAGTGGC
GCACCCAGAGAGG
TGCGACGCGCAAGC
ATTGGTTCTGTGCG
AAGCGGAGTTCTTG
AGAGTAATGGTTGC
TGGGGGCACAAAG
CGCAACATATAGCC
TCTTATGCCTCAAGT
CGTAGTTCGTACCT
CCACGTGGTCCCGC
TGGAATGCCTATCG
ACTCCTCCCCGGAG
GATCATAGAGTTCG
AAACCGGCTACGGC
GAGGCAAGGGCGG
TGAGGTGCACACCG
ATGGGGAGCAGCG
ACCCCACCTACCCTT
AGCTAAGAGAGCA
GGCGATCCGCCAAC
TGTCAGCACGAAAT
AAACTAATCATATG
TATACGAGGGGAGA
ATTTACAACGGGTA
CCTTGTGCCCGAAC
CGCCTGTAGGTATC
ACCTACAGGTGTTA
AAATGAATCTGATA
GCTGGCGGATCGTC
GACCCTCTTTGATG
GCTCTGCGCCAACG
ACTGGAAAGAATA

				GGAACGGAAGTCT AATGGAAGGAAAG TGTCGGGAGCACTA TAAATTCCCAAAGA AGAAAAGAAAAGA AAAAAAATAAAAAA CCCAAATTAA (SEQ ID NO: 1228)		
R2	R2B_NVi	.	Nasonia vitripennis	GACTAGACTATGGG TTCAGTCAGTCCCA AATAGCCGATCCTG GCGCGTCCGGCAGT AATGCCACGTATGA GTCGGTTACCCATC TCTAAACGCGTAGA GGTGGGGAGCTAA AGGCCAGGCGGTTT ACCCGACGTCGAAT TTCTCCAGGTCTGT GTCAGTCGACGGAA TAAAGGTACTACAA CATCTACTATCTATC GGGATCGGAAGAC GCCTTACAGCGTTT TCCGATTTTTGCTCT TTGAGCATTTTTCTT CAAATTGCGATAAC CGACCCGATCACGC GGGGCTTTGACAAA GCAATGCGTGGTCG GTAAGATGGTTGCA ATCTTTTCCACCTCG TTTCTTTTACGGAAC GAAAGCAATGCGT GTGGGGAACGTTA AAAACTCCCTTCAT GCATCCCAGGATTT ATCCTGCTTACTGC	TGACCTGAACAAAACGTGT TGTCTTGTCTTGTCTAAAAC TATTTATTCGAAATAAGGG GAGGCTAACTGCCTGCAAG TTGAACGCGAAAGTTAGAC CTTCCCACCTAAAGCCCAAA AGTGATCGGGGAATGAATC CGCGGGTGACCCCAGAGTT GGGTAAACCCTTGAAACGT TGGAGAAGCGGAAGAGAG TCCCGCCACCGAGCATCGA GTGCTGCGGCGCCCGAATG AAACCGATCGCGGATGGTG CAAGTCGTAGGACGGGGCA CGACCTAAGCCTCTGTCACG GCGGCGAAGCCAGGAATCA CCATGCAAAGGTGTGAACT GGGGCGGATACCTCCACGG GGTTTCCCTGGGCATCGCG CGAGCGATGGCCAAAGTCC GCTTTCTCAGCTACAAAACA AAAATGGTATGAGACTTCG TTAACACTAATTTTTCCGAG CCTAGCAGGCTCCCTTGACA ACGCTTATGAATCTGGAAA AGGACACAAAGTGGAAAAAA GCGCTGATGGTGGACAAAA GTCAGTTGAGACTTGATATC AGTTGTTTTGACTAAGAATT TTATTATCGTTGACTTTTAA	TFAPTHPMVRSGPCRKTKRPGSDYRESLIMDSGNNVASEPRGAVDVTSA APIGAELNAEPCEGRNQRREAALSAQTRRRNXARRARNAQQADEPGDD EEIETHGPLTIRTXEPMEIVAIAKNPQACPKCLQGGTQLLCMGSWELSRH INKEHPSVDVTWVCGACQRRCTTLRSWSCHVLHCKGRQEPKDLPFKCE HCSLSFDSQIGLSQHERHVHPEVRNDKRAAEANKPKGKSGRRPSIWSDE DLLLIRELESEYHGARNINEKIAEHFPDRTGRQVSDARRRKDYAALRGRG GPQGPAEGVEAIEEVDEGEIPEGEELVATDGAALESGPPENGGSAPAEQ VNAPALESSSQQDRECSPAVGSDEQIEDSSDDDEFSDALGEISLPEPLSVE RTTISPPPRDDWKGPMRWEICNASEEAGSYANWVTGLQELVRNNALSE IGLDSLYDQLIQIMRHPSDDNEQDRLQLNARGPPRRGHRKNRRRRRLTA ADRKRFAFARCQDLWNNNPKKLAELVIANDLSILQRRQAPGRTETQTLY NELWGRVGPNIEAPRRTEDPIPVSRIFTPITPQEIMGRIRRIKNDSAAGPD GVTKDDLRGRGVSIALSKLFNSILLAGYYPKAWRENRTTLLPKPEKDPADV KNWRPITISSMVSRVYSGLLDQRVRAVIKQCDRQKGFTEENGCFSNIQLL DDAVSNAKKAGGVITILDVSKAFDTVPHAVIQGCLEKKGIPETVAAYISSM YRDCSTAIRTRSGDVKIGMKRGVKQGDPLSPLIFNLVLEPLLERLQETSGV EIEGMNLSCAAFADDIVCFANTAPEAGRQLRMVADYLGRLDMSLSVSKC IAVEYVPHRKTWYTKNPGLEVNGNAVPSISPSETFKYLGAKVSPWKGLLE GFESDAFREVISRVQRLPLKPMQKVDLLQMYIFPRYTYGLITSPPAKAVLK TIDRIIRTRIKEILHLPESVSSSFLYTPRKQGGLGLLEVEKMVLIAALRNGLRA RQSHDPVTRAAMNSNAADDRLKSYADALRLHWPLTTKELDTYKYQLRLS YAQKWAEQKWQGQGVEEFAQDPVGNSWLQRYDLLPASRYIDAIKLRT NTYPTRALMKIIDGRVDSSCRKCQGSSETLGHILGRCRYTKDKRISRHNEI KDLLKARLAKNHQVMDEPQITVRGQRFKPDLVVKTNEGRVHVIDVTVRY EHRTYLDEGRTEKIGKYRQILSTLRRDLHSNAEEVIPIVIGSRGAIPRETRKA LSKLGIGKSDWLTISLIALRSSLEIVNAFMDD (SEQ ID NO: 1474)

			AAAGCAATGCGTGT GGAGCGACTTTACC ACGAGTCGCTCCAC CGCAAAGCAATGCG TATCGCGCAAAAGC AATGCGTGTGGGGG GACTTGTCAAAGAT CCCCCGCCGCAAAG CAATGCGTGTCGGC ACCACGTAGAGCAA AGCGTGTAGGCAG ACTTTGTCAAAAGT AGTTCTGCCGCAAA GCAATGCGTGTGGA GATCTTCGCCGGTG AAAGCAATACGTGT GGGCGAACTTAG (SEQ ID NO: 1229)	ATATTTTATTATTGACTGTT AATATACTGACTTGGGACC AAGTCATCTCTGTTACCCGG TACCGGTTCCTGTCATCAAA CCGGAAAGTCCGTCCCACG TAATGTGGTAGACGCAGGA G (SEQ ID NO: 1352)	
R2	R2Ci-B	AB097122	Ciona intestinalis	CGACGGTGAACCAC CTTGTCGCGGTGTA AGAGCTTTAGTGTC TCGAACAAGAAATA GCTTGTGTGCTGTC CTTCTGGGCGGTGC ACATACTTCTTAACC TCCCGAGGCCATGC CGGCGGGGGCTTT AGCCCCCGGCAGGT TTTACCATGCCGGA CGGGTTCGAGAGG TAGAGGCCAAACTA AGAGTTCACCAGCA GACTTCGCACGCGG CTGGCCACTGGCCG AAGTTTAAACAACA GGGCCGCATCTTCC CAAACTCAATATAT GGTGTTAAGTGAAC	TGACAGTAATATGAAAACA TCACATCTGACCGGCACAG AATCACCATGCCGTAATGCA CCCAACTAAGGATTCCAATG GGTAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAA AAAAAA (SEQ ID NO: 1353)	MGEWPWVSWSLTVLVEKWRPFTILQPYPMPGQLRVDVYLPRKTSYLM DKNIYENTTSPGGGPLCGEKTHRSDVIIPPPGFAPSTDTASNTLGENVDAS ATTSSANPLSQEPGWCESCSKLFKSQRGLRVHQRSKHPELYHSQNQPLP RSKARWSDEEMVIFAREEIANRKIRFINQHLHKVFPHRTLESIKGLRGKNV RYARIMADLEAEMTSQPEAATSLCTETSENLASSNVLPQTRGWAENLVE NIDTAHLANLGPLSQFEPGKPSSSTKEAINTEYNDWISKWLPSGAAHRER RANPPSTKLNARATRRLQYSRIQNLYKLNRSACAQEVLSGAWKVQSGEL NLKEVQPFWEKMFRKESAKDRRKPKPTGEVLWGLMEPLTIAEVGSTLKS TTPSAPGPDKLTLDGVKRIPIAELVSHYNLWLYAGYQPEGLREGITTLIPKI KGTRDPAKLRPITVSSFICRIFHRCLAQRMETSLPLGERQKAFRKVDGICH NIWSLRSLIHNSKDNLKELNITFLDVRKAFDSISHKSLGIAAARLGLPPPLIT YISNLYPNCSTKLKVNGKISKPIEVRRGVRQGDPLSPLLFNAVMDWALSEL DPRVGVQIGEQRINHLAFADDIILVSSTKIGMVSSINTLSRHLAKSGLEISA GKEGKSASMAIVVDGKKKMWTVDPLPRFKVNSQKIPALSITQQYKYLGI NIDAQGARNDAARILTEGLAELSRAPLKPQQRLYLLRVHLLPKLQHGLVLS SCAKRALTYLDKSVRSAIRRWLTLPKDTPTAFYHAKACDGGLGITRLEHTI PILKRNRMMKLTLSEDPVIMELVKLTYFTNLLHKYSNVKLLNSWPVTDKD SLARAEASMLHTSVDGRGLSNCSDVPRQSDWVTNGASLLSGRDFIGAIK VRGNLLPTKVSAARGRQREITCDCCRRPESLGHILQTCPRTWGPRISRHD SLLKRVRNQACLKNWTPIIEPSIPTNIGLRRPDLVLAKGNIAFLVDATVVA

				CGTGCCG (SEQ ID NO: 1230)		DNANMQLQHEAKVEKYNNSDIKEWIKVHCPGVDEVRVTSLTANWRGC LYGGSASFLTEDLGLPKAELSLLSAKINEKGYYLWCAHYRGTARLWNRPLR S (SEQ ID NO: 1475)
R2	R2C_NGi	.	Nasonia girault	CGGGTTCCCCCGAC TTCGGCTTGCCGTG GTCTGGGGCTCACT GCTTTTTGTGGAGT CATGGTTACATGGT GACCCTGGTTCCTC GCACCCCCGCTGGA AACTATCTGGGGAG GCCATGATTGGGTA ACGATAAAGGTCCT GGTCGTGTCCTCCT GAGATAGGCTGAAT GGGTCACTAAGTGG CACCTAA (SEQ ID NO: 1231)	TAGCGGACTGGACTGTCTG GAGGAGTGTTTAACTCGGG TTCTCATGGGAACCCGACA ACGTTGTTATCTTGTATGAC AATTCATAAAAAAAAAAAA AAAAAAAAAAAAAAAA (SEQ ID NO: 1354)	WVTSPRRPRYVGPQKKASDGNDGRAAARAEPTNPGGPDRADDDEG DVKFWCEFPGCDRFFMTRSGRGLHHKKGHPDWNDQRNLAGKQHRKEI WSEEERLLLAKKEAELAISGARFINVELRDFTARSLDAIKGQRKRPDYKILV EKFVRELRVRGIRQGVASRSQQAARAMAVAGAPAATSSGAPPVATQPPP SGRVLRSQVVEAPAMEIPVAESEGDSSGDELFEDVEPVRLSDLPPDRFTIY FAGLEIPGTEDIYAHRLHTICLMTTWRTKEEVRLELGLFLKDLFPSKGSQER PERTNLPDPRNRIERRRGEYKKCQDLWRRNKSTCVQRILKEDLSQGECLP RELMEPFWNATFTQNPGTAPVLPPPTEVYSSVWEPIRPENIKGNYPPQN TAAGIDGLTVGDLKGVSREMLARIFNLFMWCGKLPEHLCASRTILLPKKP GAKVPGEFRPITVTSVLIRTFHKVLAERLKVVPLDPRQRGFRESDGCAENV MLLDMTIRYHHERRRKMFLALLDMAKAFDSVSFESMREVLTTKGIPTPFI EYFMTHLEDSFTVLQHGNWQSGKIHPTCGVKQGDPLSPPIFNFIMDEM LKRLPKEIGVNLDGLFVNAMAFADDLSLVANTEQGLQILIDEATSFLGLCG LRANPNKCVTLAIKTIPKEKKTAIDPSSHFRIGNAVIPSLKRTDEWVYLGIKF NSNGRLISDAKPKLIKDLELLTKAPLKPQQRLWALKVIVIPGILYRGTLGSST AGYLRSLDCVIRAYVRRWLRLPGDCPNGYFHAAVADGGLGVHPIRYKA MVDRLARLRKLEKSAYITGPEAARYLQRQVSIAENRLRDGANRIMSDAS MLREFLRELLYKSFDGRPLENSSKVPGQHRWVEEPTRFLSGADYMNCIR ARIAALPTAARCARGRLKDKHCRAGCGNVETLNHVLQFCHRTHGTRIGR HDAVVKYVVGGLKKRGYAVKEEPKIVLQDVVYKPDMVATKEGKTLILDA QVLGDQRDMRLAHEDKLRKYGAPEFKRKIRSETGSATIKSLSVTLSWRGL WGPDSVKGLLEEGVILKKDLKILSTRVLIGALAGWRRFNERTSMATSGRR EEVTTRMVRRWKRRERVGVG (SEQ ID NO: 1476)	MSGKSSKPRTVSSGSSSQETPPSGSNACDICGKCFMKPVGLSRVHPSQY HARLEKNQPKAKKFRWTDEDLYFLAKKEAELLHLGSIKFVNKELAEFFPEK SVDQIRGQRRSETYKQQVLSIHSELLKLQTVADSPPPSRIPAKEVSAWLDF FLALPKTKNKFSEDKLDQLIRTAQDGTLILDDLDLYLREVLVQPTSQGEKQ AKLLPPPKSSREKRDREYARAQNLYRKNKTACVNAILDGNKKCENKIPDID DFWKTIFESHSPPDAEPVCVVVDEEPTNIWSWISFFEMNHNYPDSSTSP GPDGVTARMLRSIPARVLNKLLNLLLFIEDLPAVFKCHRTVLIPKIDNPTSP GEFRPITISSIVVRQLNKIIAARVSEGVPINPRQKAFRQIDGCAENVFLLDFI LRDAKTKIKSLSLATVDIKKAFDSVSHHSIFRAIRGARCPENLVNYIQNSYS GCTTQISVGGSISASKIPMNRGVKQGDPLSPVLFNLVINEIIRKLPASIGYPI NSELSINCIAYADDLILVTNTREGLKLLLGLLNEELPKRGLELNASKCFGLSL TALGKLKKTHLCTSDQLDLHGTLIKNLTAEESWVYLGVPFSHIGRSKSFSP

			ATCTTCCCAG (SEQ ID NO: 1232)	CTCTGGTTTTTGGAACGGCT TCCTTAGCCGGAATTTTGTC TGATGTATCTTGCTTGTGTC CTTGAAATATACGACCCAG GCTTGCGTCATTTAGGCTCT GGGAAA (SEQ ID NO: 1355)	DLEALLNKLQKSPLKLQQKLFALRVYLIPRLLHGLVLSRVAIGELKIMDKLIL KHLRVWLRLPKDTPLGFFYSPVKLGGLGIKNLRTNVLKCRKQRIERMLVSP DDVVRLVAESEIFLKETDKLKDLLTINGMCLDXRNVPRTGKNNKFWSERL YTSFDGKTLAYSEYFTQGGGWIREDKILQPAHVFAECIKLRINALPTKSRV AHGRPTKDRSCRAGCLDVQKVPTIETINHIAQVCPRTHGARIKRHDRLVQ FLSLNLRKNPKRNVLVEYNFRTVAGIRKPDIIVIEDTRAVILDVQVVGDSSN LEMEYLEKSRKYSNDANFINALQKLYPTVTNLTFHAVTFNNRGLIAKSTVA ALRMLGVPPRCIMILCVISLEKTLEVWRMFNQSTASARK (SEQ ID NO: 1477)
R2	R2LcA	.	Lepidurus couesii	TAGTGCTTGAGTGATGCCT ATCCTTTCTTTGATTAACTCT TACCATATACTTACCAGTTC TTACCCGTACTTACCCTGTA TACTTACCTGTGTGCGTACC TGTGTACTTGTCCTTTAGCC GCCTTGTGTTTTTACCATTG GTACTTACCTTGTGTGGTTG CCCGATACTTACCTTGTATT TGCCTTGTAATTCTGCATGA TATTTATTGTGTAGGTTCCT GATGCTTACCTGATTTGTCC CCCTCATCATCTTTAGTTTC GTTCTATTTCACTCCATTAT GGAGTTCCGTTTGTTTTTTG GTGGAGGTACAGCACCCTT TAAGCTGGAATTGAGTGAG TTTATGTACTTTGGATGGTT GTAATAAACTACCCGGAGG CAT (SEQ ID NO: 1356)	MSEESRPKQTASKRGAAVEKTMMSGTYVCTLCGRSFEKSVGLSLHTNR MHPEAYNKLKEAKKPVLKKARWSEEEVFLLAQKEAELSFIGGIKFMNIEL HKIFPERELEGIKGQRKNPTYKAQVVSLLAEIRESKANDSSSSSSSSSCDS ASLGISNWLEFLLALPKTSNQFQEGRLDRLISDALRGVDVLENLDAYLLEV FAKPMAQNPCPKPPPPAKNSRERRDREYSRVQNFYKKNRSACINSILDG NTRSQNVIPGLTKFWTETFEKNSPPDDEAPDQFVADEPRDMYKWITFYE MSQDYLDSSTAPGVDGFSAKQLRSMSPRVLNKILNLLLLSENLPNSFKM HKTVLIPKIDDPKSPGDFRPITISPVLARLLNKILAARLSKLVPISQRQKAFLP LRGLDLNAAKCHGLSIIASGKAKTTYVSAADSLDLDGQPIKNLGVLDTWT YLGIPFSHLGRAEKVSPDLTNLLNKLQKAPLKLQQKLYAVRNFVIPRALHG LILSKTNLKELNTLDRAIRVFLRTLLYLPKDTPLGFFHSPIKSGGLGITCFRTS VLKCRLQRIARMRSSCDGVIQAVAESDIFADEYAKLRDLIRINGNVLDTTE SIKRYWAQRLHSSVDGKTLAYMDYFPQGNLWMSEDKVSQRSYVFADC VKLRINAIPTRVRVSRGRPNKEMCCRAKCFDSQRMPAFESLNHITQVCP RTHGSRIQRHDKIAKFLFKNLNNCPSRSVLYEPHFVTVDGLRKPDIIIYDDS HMVVLDVQVVSDSANLEKEFECKAKKYANDVALRSAMLIKYPFIKSFSFV AATYNNRGLIAKSSVQVLRQLGLSPRSIMVSILICLEGTLETWRIFNQSTM NAH (SEQ ID NO: 1478)
R2	R2LcB	JN937619	Lepidurus couesii	TTTTGGGGTAGCAA TTGATCGATTCCCG CCTCCTCGTGGCGC TACCCTGGGATAAC CTCAAAGAAATTTG ACGGTAAAGCTAAG AGGAATTGATCACA CCGTGACGAATATC	TGATAATCGCTCCATCCTGC AACTAATTATGAATGCAAAT CTGTTAAGTGACATTAGTG ATACTTACCTGATACTTACC CTGGTATTTATTTGACCTAT ACTTACCCTGGTATCTACCT GACATATATTTATCTAACCA CCTACCTATGATGACTCCCG	MSGKSSKPRTVSSGSSSQETPPSGSNACDICGKCFMKPVGLSLHMSKVH PTQYHARLEKNQPKAKKFRWTDEDLYFLAKKEAELLLLGGIKFMNKELAE FFPEKSVDQIKGQRRSETYKQQVVSIHSELLKLQAVADSPPPSRIPAKEVS AWLDFLLALPKTKNKFSEDKLDQLIRTAQEGTPVLNDLDLYLREVLVQPTR QGERQAKPLPPPKSSREKRDREYARVQNFYRKNKTACVNAILDGNKKCE NKIPDIDEFWKAIFESQSPPDAEPVSYVVDEEPKNIWSWISFFEMNRNNYP DTSTSPGPDGVTARMLRSIPARVLNKLLNLLLFIEDLPAVFKCHRTVLIPKV DNPALPGEFRPITISSIIVRQLNKIIAARVSEGVPINPRQKAFRQIDGCAEN

			ATCAGTCACAGCTG CACGAATCCAGATA GATATATAACAGGC GAGTAATTCTTTTC GAAGCTCGTCAGTA ATCTTCCCAG (SEQ ID NO: 1234)	CGGAAACTCTCACTTACCTT ATTACCCACTTGGTCTTTTA TTTTCTCGTTCCTTATTACTT TGTTCCTTTGGTGTAGGGTT CTCTGGTTTTTGGAACGGCT TCCTTAGCCGGAATTTTGTC TGATGTATCTTGCTTGTGTC CTTGAAATATACGACCCAG GCTTGCGTCATTTAGGCTCT GGGAAA (SEQ ID NO: 1357)	VFLLDFILRDAKTKIKSLSLATVDIKKAFDSVSHHSIFRAIRGARCPENLVNYI QNSYSGCTTQISVGGSISTTKILMNRGVKQGDPLSPVLFNLVINEIIRKLPA SIGYPINSELSINCIAYADDLILVANTREGLKLLLNLLNEELPKRGLELNASKC FGLSLTALGKLKKTHLCTSDQLDLHGTLIKNLTAEESWVYLGVPFSHIGRSK SFSPDLEALLNKLQKSPLKLQQKLFALRVYLIPRLLHGLVLSRVAIGELKIMD KLILKHLRVWLRLPKDTPLGFFYSPVKLGGLGIKNLRTNVLKCRKQRIERM LVSPDDVVRLVAESEIFLKETDKLKDLLTINGMCLDXRNVPRTGKNNKFW SERLYTSFDGKTLAYSEYFTQGGGWIREDKILQPAHVFAECIKLRINALPTK SRVAHGRPTKDRSCRAGCLDVQKVPAIETINHIAQVCPRTHGARIKRHDR LVQFLSLNLRKNPKRNVLVEYNFRTVAGIRKPDIIVIEDTRAAILDVQVVG DSSNLEMEYLEKSRKYSNDATLSMRINALQKLYPTVTSLTFHAVTFNNRG LIAKSTVAALRMLGVPPRCIMILCVISLEKTLEVWRMFNQSTASARK (SEQ ID NO: 1479)
R2	R2Nve c-A	.	Nemato stella vectens is	GGTTGGGGCCTTCT CGTGGCGGAGTCG TGAGTAAGGGGTAT AGGGGTAAGGGAC ACCACGGACCGAG AACGGTTACCGCTC AAGGCGAGTGGTG GAAGGCATAAAATC GTAACGCCGCCCTC CGACCTGCTCCTGA AACTAATGCCAACC AACTGACTGTGGGG CTAACCTCCCCAGA GTCAGG (SEQ ID NO: 1235)	TGATGGTGGGTTACTCGCC TCTGTGTAACAGGCAAATG AAAGCTGCGCAAGCAGTCG ATGAGCCAAAGCCGCACAG CCCCCGACTGGGGTACAGG GCAGCCCTGGGCTATGCCC GAAGTCTTTTGACAGGTCC AAACTTCACCCTTGCCGCCA GTAGGGCACCAGGGCCAG GTGCAGGTGGGCGCTTTGT TCTATTTGGTTTCGCTTAATT TTTGTTAAATTTTTCCCGTG CGCCCACCCTTTTTAACCCT TTTAACACCAATTTTTTTTAC AACCCTCTACTCAATAATCC AAATGAATAAAAGCGGCAT AACAGGTGAACAC (SEQ ID NO: 1358)	MLRGTGNMNDKRDGSATADPTSALLGAVGDGSLVCNLCGLACKSRGG LSIHRRSKHATVYHAERQPAPRAKARWTNDEMILVARKQIASEKSRCSA VVEGMREAVPHRTFDAVKSLKTKNRNYTRILEQIRAECSEEEVIESGVLKD RTENVCVQTTSNVPGSAGRAASVELEGNIQVGHQLAQKTMAGNNSRK QPANHTNWAEFNIEEGNITLRKSKRKANGMPDATHRPGPPTVDSLKHP VCLLQGAADKRDEPHTVEQLYYNIEEGMPLAEEQQWSEKLFDAIDSSLLS VEVELGRIVPGCPDEETRQLIDREFLDFIHSYSREKPPQRGLAKSKPPPKGP KSLRRQQYRQLQRLWDKNRSAAAEQALTGKWQEVRTAAGVPLSLMEV PWREIFETPSTTDVREPAPAGPVLWQLLRPVTIAEVEDAISSKKSASGPD GVPCAALQTMGAASLAAHFNLWLLAGTQPKRLTECRTIFVPKEVNTHLP LHHRPITIGSVVVRLFHQILGKPMEAVLPLGSGQRGFRKGDGICQNIWLL HTLIRRSTDLLRPLKLVFLDVKKAFDSVSHESLLIAAKRLGVPGPLLTYINELY SRSETVFEVGGESSGSVKVSQGVKQGDPLSSTLFNCVIDWAVSDLDPHIG VLLGESRVSFLAYADDLVLLSETEAALTSQLNSIEKSLAHCGLKLSTGDSGK SASLNIVIDGKAKRWVVNPTPFLRASGGEIRSLVANETYKYLGINIGAQGV KAAEYNAFKEALDNLSRAPLKPQQRLFLLKTYLLPQLHHSLVLSRTTGKLL NSLDALVRKAVRGWLKLPHDTHRAFFYAHQADGGLNVPSLYHLIPLLRRS RYERLTRVEDPEIREVSRTDYFKRVLGAAAAATTVAGHRIDSKVTLRLAW REALYASADGRGLSQCPLVPEVHSWVTDVSGLQTGSQYISAVRLRGALLP TAVRKSRGRGGVNSNCDCCGRGQPEFLGHVLQTCPRTWGSRISRHNHV LSLIAKACRSRRWQVLEEPIIQTPAQLLKPDLVIWNHQAAYVVDVSVPGD NTPLSTCHNRKVAYYSGESVREWVRSKTGHNPTVSSVVINWRGAMAKE SYRLLTKDLRLAKTLPRLLVLRVLEGGHGIWLNFHRSTFAVGVT (SEQ ID NO: 1480)

R2	R2Sm-A	.	Schistosoma mansoni	ATGTTTTAATTTATT TTTGAACTACTACT GTCTGAGTGCTTCT TACAACCTGAAGGC TCAGAAACTACCCA CTTTTTGCTGTTTAT CCACAACAACAGTT GTGAATCTATTCTCC AAATATTCCTTGTG CTTTTGTCAACATTA TTCTATACCAACTGT ACCACCTACTTCTTC ATCTCACGTTTTAAT TCTGGTCTAATTTTC TCATCATTAGTCAC GGAGAGGGCCTAT GAACGGTCCGTGAC GCGAAATTCAATCC ACGAATTCGTCCTC TTCTGCTAGTGGTC CCCGAAATACGGTT CCTCTGGCCTGTCA GTTGTGTTAAAACT ATATAATAACG (SEQ ID NO: 1236)	TAGCCCCCTTCACTCTTAGA CATTCCCCCACTGTTGTTGC TTATCTTCATGTTTTTGTGTT AATTGACTGCTCTCTTCTGG GTTGATGTCTGATTGTCTCT CTCTCTTTCCATATTGCTTGC TCTCCCCCGCTTACTTCCAAT AGTTGTCATATTATGTCTTT GTTTACTTGCCATGTCTAAC GACAATTACTTTATCTACCT TAGTTGGTCCTCTTGGTTTG GTTGCCTTCATGTGTTCATG GCGGAATCTGATGTTTATA ATGACTATTCCTACTACCAC CATTACAACTATTATTATTA TCACTATTATTAACATTATT ATTACTTCTACAATTAGTAT TATGGCTACTCCTTTCAGCA CACCAATAAAATCTCAATCA AACATCTCACTTATTAAACT CTCTATTTCCCCTTCGTTATA AACTTACAATTCAGTTTAAC CGAATATCTCTCTTTTACAA ATCTTAAGTATGTAATTTTG TGCCAAGCCCATTTGGGTCT GTACAATTTGATACTTAAAA ATAAATGTTAT (SEQ ID NO: 1359)	MPVSTGAETDITSSLPIPASSIVSPNYTLPDSSSTCLICFAIFPTHNILLSHAT AIHHISCPPTPVQDGSQQMSCVLCAAAFSSNRGLTQHIRHRHISEYNELIR QRIAVQPTSRIWSPFDDASLLSIANHEAHRFPTKNDLYQHISTVLTRRTAE AVKRRLLHLQWSRSPTAITTSSNNHTTTDIPNTEARYIFPVDLDEHPPLSD ATTPDASTHPLPELLVILTPLPSPTRLQNISESQTSHESNRNSMHTPPTYAC DSDESLGVTPSSTIPSCFHSYRDPLAEQRSKLLRASASLLQSSCTRIRSSSLL AFLQNASTLMDEEHVSTFLNSHGEFVFPRTWTPSRPKHPSHAPANVSRK KRRKIEYAHIQTLFHHRPKDAANTVLDGRWRNPYVANHSMIPDFDCFW TTVFTKTNSPDSREITPIIPMTPSLIDPILPSDVTWALKEMHGTAGGIDRLT SYDLMRFGKNGLAGYLNMLLALAYLPTNLSTARVTFVPKSSSPVSPEDFR PISVAPVATRCLHKILAKRWMPLFPQERLQFAFLNRDGCFEAVNLLHSVI RHVHTRHAGASFALLDISRAFDTVSHDSIIRAAKRYGAPELLCRYLNNYYR RSTSCVNRTELHPTCGVKQGDPLSPLLFIMVLDELLEGLDPMTHLTVDGE SLNYIAYADDLVVFAPNAELLQRKLDRISLLLHEAGWSINPEKSRTLDLISG GHSKITALSQTEFTIAGMRIPPLSAADTFDYLGIKSNFKGRCPVAHIDLLNN YLTEISCAPLKPQQRMKILKDNLLPRLLYPLTLGIVHLKTLKSMDRNIHTAIR KWLRLPSDTPLAYFHSPVAAGGLGILHLSSSVPFHRRKRLETLLSSPNRLLH KLPTSPTLASYSHLSQLPVRIGHETVTSREEASNSWVRRLHSSCDGKGLLL APLSTESHAWLRYPQSIFPSVYINAVKLRGGLLSTKVRRSRGGRVTNGLN CRGGCAHHETIHHILQHCALTHDIRCKRHNELCNLVAKKLRRQKIHFLQE PCIPLEKTYCKPDFIIIRDSIAYVLDVTVSDDGNTHASRLLKISKYGNERTVA SIKRFLTSSGYIITSVRQTPVLTFRGILERASSQSLRRLCFSSRDLGDLCLSAI QGSIKIYNTYMRGTQRLNE (SEQ ID NO: 1481)
R2	R2Tc	EU854578	Triops cancriformis	TTTTTGGTCTGGCA TTTGATCGTTTCCGC CTCCTCGTGGCGCC AGACTGGGTAAGCT GATTTATCAGTGAG CTAAGAGAAACGAT CACCGCAGGAGTCC ATCTACCTGCGCTG CACGTGATTTCATT GTGCACTTGGCGAG	TAGATGACTGCCCTACCCCT TTGCTGCCGAAGAACTACT GAAGACTATTGAACTAACC AACTGTGTTAAGTAAGAAC TAATGCCTCTTTTTCCCTGC ATGTATCCCCTGATCAGTGA CTTATTTTCTTTTCCTGTTGC GCCCTTTGTTTAGTTATTTCC TTTAATTACTAGAATTATCT TTTCGTTCTCCGTCTAATTG	MSQKRRPEKAVPDEGATAHDVAQPDKSKCSVCGETFKGPASVTMHMV KKHPVEFNELKMAKKPVPKKVRWSEEEIFQLARTEAELTLQGVRFINVEL QKIFPAREIEGIKGQRKLAKYKELVKDQLDEIGRAPNPPEQEIGEDVPSPFK AWLELLLALPKTPNDFLEHKLDNIIVQALKEDVNSDQVFNDLNSYLKLILEP SGRAKSVPGEIIHGDPSGSAKTSVTKAPKPATVSSSRKKRRDAEFARIQRL YRKNRTSCINTILDGNTREHEAPKNMEGFWREIFERESPDDPDDPDDIFLE EEASDIWKYISFYEMCNLYPPPSTAPGPDGFSSKDLRRMTPRVLNKILNLL LHLRDLPQILKSHRTVLIPKTDLPTKPGDFRPITISNILVRHLNKILANRVSHL IPINERQKAFLPIDGCAENIFTLDFILHHARTKIKSLSMAILDISKAFDSVSHH SIFRALREARCPIGFIKFIENCYGGCFTKLFCGGVKYPSEVSMNRGVKQGD

			TTATCCCTTGTGGA GCTCGCTCGTTAGC TTTTGAG (SEQ ID NO: 1237)	CTTTTTCGGTGTAGGAACG GCTACCTAAAGCTGGAAGT GGGAAGTGTTTTCAATGTA CTTTGTGATTATAGAAATAT ATGACCCGAGGTGCATTGT TTGGCATTTCCTCGAGAAA (SEQ ID NO: 1360)	PLSPVLFNLVIDGLIRQIPSALGFNVSDQVKVSCIAYADDLILIATTRAGLKT LLDLTNSYLAKRGLSLNPDKCSALSIVASGKQKLVYIASSEHFDLAGQKMR NLNVGDSWRYLGIQFSHLGRAEKVTPDLTCLINRLQKAPLKLQQKLYALRI YLIPRLIHGLTLSKTNLGELKTLDKLIRKYIRAWLHLPDDTPMGYFYTPLKA GGLGLPSLRLVILNNRLERILRMKASQDIIVRTIAESETLGVEIRKLHDLLSID GTILDTSVKIHSFWAERLYSSYDGKCLCNSANFPPGNKWIGEDSLNQRSH IFADCLKLRINALPTRSRTARGRPLKDKPCRAGCRNSDGVKVIETLNHITQ VCERTHGARVKRHDRLVDFAVKGLQRPHRVVLKEPHYKTVNGVRKPDIV IKIPDHTYICDFQVVSDTSCLELEFRKKALKYAEDKGLCDQLTRDHPGELSF TAITFNTRGLIAKSSVTALRKLGMPPRSIMTLQKICMEGSLEIWRIFNQTT AMARN (SEQ ID NO: 1482)
R2	R2_DAn	.	Drosophila ananassae	AGAATATGGATTTG ATTGTGCAGAGGG GGTGCTATACCGTA ACTCGTAAGCCATG CAATCAGATCAAGT CGACTCAAAACCTC CTCGTGGTATTCTCT GGGTGCCAGTATTT ACTGGTAGCTGA (SEQ ID NO: 1238)	TAGCCAATGCACGGGTTCC AGATTAAGCTTGCTGCCGA AGCATACCATCAAAATCGG CATAAAATTCGCTTAATAAA GGAGGTGGTTTTAGTACGT AGGCGTCCCGGGACTTGTC TCGGGATGAATCGTGTGCATG CGTATAATTGGGATCGATA ACAAATACCAACTAAGTTAT TACTAATATATCGAAATACA TAAATATCCCGTCCTTACGT ATCTTTGAAGATTTCCATCC TCAGCGAACAAAAAAAAAAA AAAA (SEQ ID NO: 1361)	FERRKDPWGYRPPGTLKQIGATENNEPRNLNRFVRGESTASSLESTQFGT SAEVNLAGRVPCTICEMTFSSKRGLGVHMSHRHKDDLDAQRLRVDKKA RWSEEETLMMARKEVELAASGVRFLNKKLAEIFTHRSADAISSYRKRSEY KAKLEQIRGQSVPTPEAEEINTTQRRPSNSEQNRRVPRSEGGPIAPTEQT NNEILRVLQGLAPVVCLPRWRAEVLQNIVDNAQVSGQETTLQSLSSYLM EIFPPRNEPHILTRPRTEPRNMRQRRRQQYARVQRNWDKHPGRCIKSLL EEDDESVMPNQEVMEPYWRRVVMTQPSSSSIKRDMFNMEHSLERVWS AVNQRDLRATKVKLSSSPGPDGITPKTARSVPEGIMLRIMNLILWCGNLP YSIRLARTIFIPKKATANQPQDYRPISVPSVIVRQLNAILASRLSAAINWDT RQRGFLPTDGCADNTTIVDLVLREHHKRFKSCYIGTLDVSKAFDAVAHEA VYNTLASYGAPKGFINYLRKAYEGGGTMLAGNGWVSEAFIPARGVKQG DPLSPILFNLVIDRLLRSLPSEIGAKVGNAMTNAAAFADDIVLFAETPMGL QKLLDTTVCFLSSVGLTLNTDKCFTVSIKGQAKQKCTVVERRSFLIGGREC PSLKRTDEWKYLGIKFTAEGRARYDPAEDLGPKLLRLTRAPLKPQQKLFAL RTVLIPQLYHKLTLGSVTIGVLKKFDKLVRYTARKWLGLPVDVPVSFFHAP HKSGGLGLPSLRWTAPMLRLKRLSNIKWPHLERSEVASSFVEEEMRRAR DRLQAGSEELLTRSQVDSYLANRLHMSVDGCGLREAERFAPQHGWVSQ PTRLLTGKEYTDGIKLRINALPSRSRTTRGRHELERRCRAGCDAPETTNHIL QQCYRTHGRRIARHNGVVNFLKRGLERRGCVVHVEPSLQGETGLNKPDL VAIRQNRIYVIDTQIVTDGHSLDQAHQRKVGKYDTPDIRTNLRRSFGAFDI EFHSATVNWRGIWSGQSVKRLIASDLLSSGDSNIISVRVISGGLWSWRQF MYLSGYTRDWT (SEQ ID NO: 1483)
R2	R2_DM	X51967	Drosophila melanogaster	TTGGGGATCATGG GGTATTTGAGAGCA GAGGGGGAGTATT CTTCTGTAATTCGTA AGTCATATCATATG	TAGCTAAATCGTTTGGTTCA AAACATTTGCTTGCTGTCTT GGCATAACATCAATAAAGG CATAAACATCGCAAAATAAT GGTTATAATTAAATGGCTAT	MTTRPSVDIFPEDQYEPNAAATLSRVPCTVCGRSFNSKRGLGVHMRSRH PDELDEERRRVDIKARWSDEEKWMMARKEVELTANGCKHINKQLAVYF ANRSVEAIKKLRQRGDYKEKIEQIRGQSALAPEVANLTIRRRPSRSEQDHQ VTTSETTPITPFEQSNREILRTLRGYSPVECHSKWRAQELQTIIDRAHLEGK ETTLQCLSLYLLGIFPAQGVRHTLTRPPRRPRNRRESRRQQYAVVQRNW

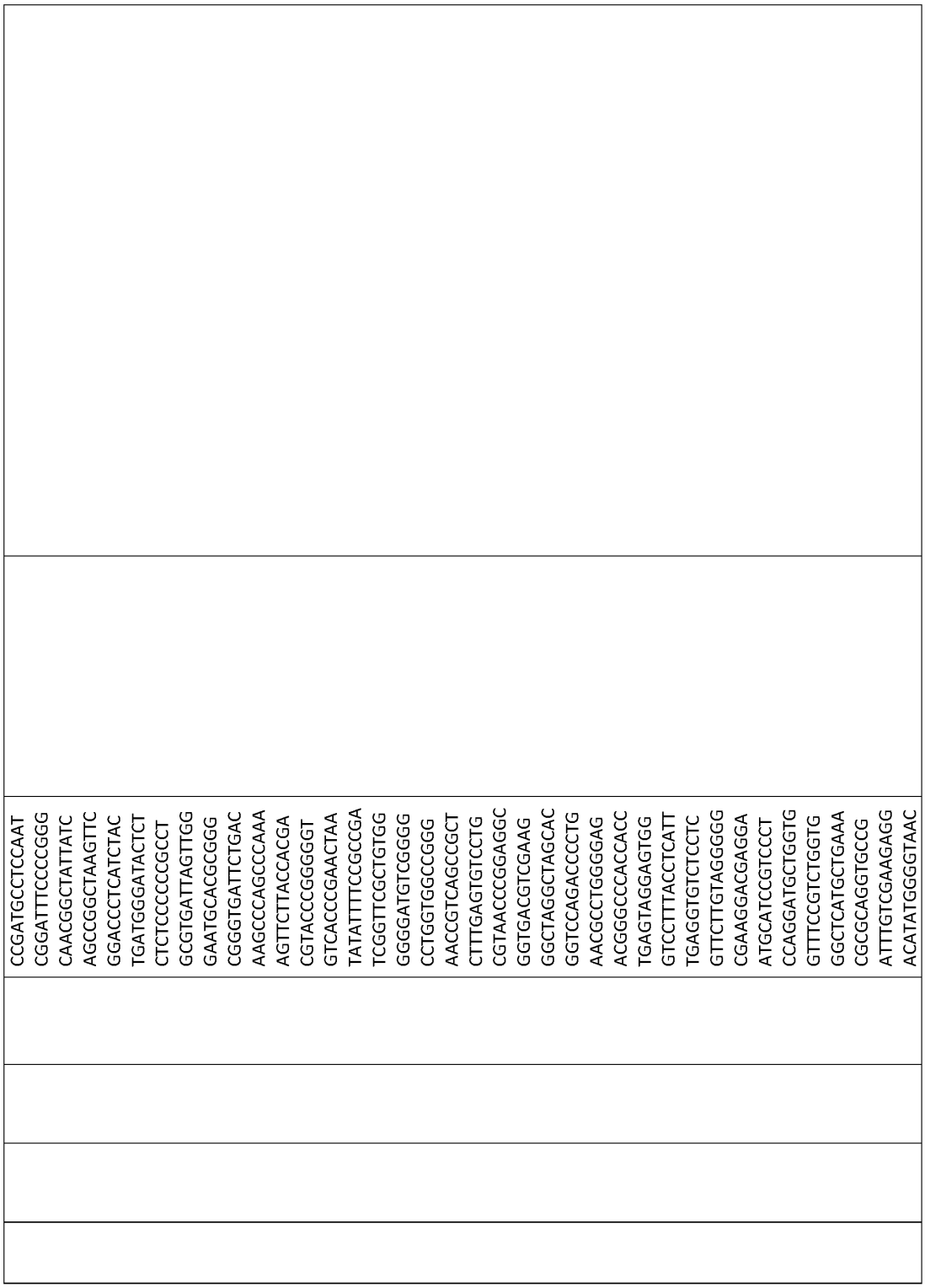
			ATGTGCGGAAGGG GAATTTTACTCTGTA ACTCACAAGTCTCT CCTTTACTCAAGTC GACTCAAAACCTCC TCGTGGTGGTCCCG GTAATGCTAAACTC GTTTAGCAGCTAAT TTGAGCGGAAAAAC TTTTCCGATGGGCT GGTTCCCCAGAGGA AATTTATTCATATTG GAACTACAAGCACA AATAACGAGCCTCG GATACCTTTACACA ATCTG (SEQ ID NO: 1239)	GAGGATGGTTTTAGTACGT AGGCGTTGCGGAACTTCGG TTCATATAGAGCAATGAATC GTGCATGCTAGGAAAAACTG ACCACACACAGTGTTGGCA GACCTAGTATCTTTCGAAGA TTTCCATACCTCCGCGATCA AAAAAAAAAAAAAAAAAAA AA (SEQ ID NO: 1362)	DKHKGRCIKSLLNGTDESVMPSQEIMVPYWREVMTQPSPSCSGEVIQ MDHSLERVWSAITEQDLRASRVSLSSSPGPDGITPKSAREVPSGIMLRIM NLILWCGNLPHSIRLARTVFIPKTVTAKRPQDFRPISVPSVLVRQLNAILAT RLNSSINWDPRQRGFLPTDGCADNATIVDLVLRHSHKHFRSCYIANLDVS KAFDSLSHASIYDTLRAYGAPKGFVDYVQNTYEGGGTSLNGDGWSSEEF VPARGVKQGDPLSPILFNLVMDRLLRTLPSEIGAKVGNAITNAAAFADDL VLFAETRMGLQVLLDKTLDFLSIVGLKLNADKCFTVGIKGQPKQKCTVLEA QSFYVGSSEIPSLKRTDEWKYLGINFTATGRVRCNPAEDIGPKLQRLTKAP LKPQQRLFALRTVLIPQLYHKLALGSVAIGVLRKTDKLIRYYVRRWLNLPLD VPIAFVHAPPKSGGLGIPSLRWVAPMLRLRRLSNIKWPHLTQNEVASSFL EAEKQRARDRLLAEQNELLSRPAIEKYWANKLYLSVDGSGLREGGHYGP QHGWVSQPTRLLTGKEYMDGIRLRINALPTKSRTTRGRHELERQCRAGC DAPETTNHIMQKCYRSHGRRVARHNCVVNRIKRGLEERGCVVIVEPSLQ CESGLNKPDLVALRQNHIDVIDTQIVTDGHSMDDAHQRKINRYDRPDIR TELRRRFEAAGDIEFHSATLNWRGIWSGQSVKRLIAKGLLSKYDSHIISVQ VMRGSLGCFKQFMYLSGFSRDWT (SEQ ID NO: 1484)
R2	R2_DPe	Drosophila persimilis	AAGATATGGATCTG AATAATAGCGTAGA AGGGGAGTCATTCC GTAATTCGTAAATC GTAAAAATCAGATC AAGTTGATTCAAGA CCTCCTCGTGGTAT CTTCTGGATGCTAT TAGACTGA (SEQ ID NO: 1240)	TAGCCTATACACTATGTTGG AGAGAAGACGCTTGCTACC TAGGCAAAATGTGAAATTA GGTATAAACATCGTGGTTG TAAAACTTGAGGTGGGTTT TTAGTACGTATGCGTGATTA CTTCGTAATCATGAATCGTG CATGCTAGTGGGGGTTTGGC CTCCACTAGTATCTTTGAAG ATTTTCCTTCCTCAGCGATC AAAAAAAA (SEQ ID NO: 1363)	SSFGLIVTNLNSETVLWGCQPLGQFSLIGTNMQNTTPRIINTNSLTNQIPT VSSLGAQSEHSAQVNPNSGYQCTICESSFRSKSGLGVHMSRRHKDEFDQ LRLRTDRKAQWSEEELSMMARKEIELAANGERYLNKKLAEVFTNRSVDAI KKCRQRERYKTKIEQLKGQAVPLPEALESETIQRRPSIRERDLLVTPPNTLG TTPTELSNSEILAVLQGYPPVVCNDQWRVEVLQSIVDGAQASGKEITLQR LSTYLMEVFPSQNDRPIQTRPPRRPRNRRQGRRQQQYALTQRNWDKHK GRCIKAILDGTEGTATMPSQGIMGSYWRQVMTQTSPTYSGTNTTFRTE HPLEGVWSPITLGDLRVHRVSLTKSPGPDGITPRTVRSIPSGVMLRIMNLI LWCGKLPVSIRQARTIFIPKVGNASRPQDFRPITVQSVMVRILNAILASRL TSSVDWDPRQRGFLPTDGCADNTTIVDLILRDHHKRCKSLYIATLDISKAF DSVSHAAVSATLTAYGAPKEFVDYVQNSYEVCGTTLNGDGWRSEEFIPA RGVRQGDPLSPIIFNLIIDQLLRSYPNEIGATIGDHTTNAAAFADDIVLFAE TRLGLQTMLDTTVDFLSSVGLTLNSDKCFTVGIKGQPKQKCTVVIPETFRI GSRSCPALKRTDEWKYLGITFTAQGRTRYSPADDLGPKLLRLTRSPLKPQ QKLFALRTVLIPQLYHKLTLGSVMIGVLRKCDILVRSTVRKWLGLPLDVST AFFHAPHTYGGLGIPSVRWVAPMLRMKRLSNIKWAHLAQSEAASSFLT DELNKARGRTLAGLNELTSRTEIETYWANRLYMSVDGRGLREAGLFRPQ HGWVCQPTRLLTGQDYRNSIKLRINALPSRSRTTRGRNELERQCRAGCD APETTNHILQNCYRTHGRRVARHNCVVNNLKRILEEKGHTVHVEPSLQLE TSVSKPDLVCIRDNHACVIDAQIITDGLFLDDVHHRKVEKYKRPEVISALRR

						EFGVSGNVEVLSATLNWRGIWSNQSVRRLIAKGLISSGDSNVISARVVTG GLYCFRQFMYLAGYTRDWT (SEQ ID NO: 1485)
R2	R2_DPs	.	Drosophila pseudoobscura	CAATTGGAAAGATATGGGTCTGAATAATAGCGTAGAAGGGGAGTCATTCCGTAATTCGTAAATCGTAAAAATCAGATCAAGTTGATTCAAGACCTCCTCGTGGTATCTTCTGGATGCTATTAGACTGA (SEQ ID NO: 1241)	TAGCCTATACACTATGTTGGAGAGAAGACGCTTGCTACCTAGGCATAATGTGAAATTAGGTATAAACATCGTGGTTGTAAAACTTGAGGTGGGTTTTTAGTACGTATGCGTGATTACTTCGTAATCATGAATCGTGCATAGTGGGGTTTGGCCTCCACTAGTATCTTTGAAGATTTTCCTTCCTCAGCGATCAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1364)	SSFGLIVTNLNSETVLWGCQPLGQFSLIGTNMQNTTPRIINTNSLTNQIPTVSSLGAQSEHSAQVNPNSGYQCTICESSFRSKSGLGVHMSRRHKDEFDQLRLRTDRKAQWSEEELSMMARKEIELAANGERYLNKKLAEVFTNRSVDAIKKCRQRERYKTKIEQLKGQAVPLPEALESETIQRRPSIRERDLLVTPPNTLGTTPTELSNRELLAVLQGYPPVVCNDQWRVEVLQSIVDGAQASGKEITLQRLSTYLMEVFPSQNDRPIQTRPPRPRNRRQGRRQQYALTQRNWDKHKLSTYLMEVFPSQNDRPIQTRPPRPRNRRQGRRQQYALTQRNWDKHKGRCIKAILDGTEGTATMPSQGIMGSYWRQVMTQTSPTYSGTNTTFRTEHPLEGVWSPITLGDLRVHRVSLTKSPGPDGITPRTVRSIPSGVMLRIMNLILWCGKLPVSIRQARTIFIPKVGNASRPQDFRPITVQSVMVRILNAILASRLTSSVDWDPRQRGFLPTDGCADNTTIVDLILRDHHKRCKSLYIATLDISKAFATSSVDWDPRQRGFLPTDGCADNTTIVDLILRDHHKRCKSLYIATLDISKAFDSVSHAAVSATLTAYGAPKEFVDYVQNSYEVCGTTLNGDGWRSEEFIPATSVSGNVEVLSATLNWRGIWSNQSVRRLIAKGLISSGDSNVISARVVTGGLYCFRQFMYLAGYTRDWTDEWKYLGINFTATGRVRCNPAEDIGPKLQRLTKAPLKPQQRMFALRTVLIEFGVSGNVEVLSATLNWRGIWSNQSVRRLIAKGLISSGDSNVISARVVTGGLYCFRQFMYLAGYTRDWTDEWKYLGINFTATGRVRCNPAEDIGPKLQRLTKAPLKPQQRMFALRTVLI
R2	R2_DSe	.	Drosophila sechellia	GGGATCAGGGGTAATTGCGAGCAGAGGGGGAGTATTTTTCTGTAATTCGTAAGTCATATCATATGGTGTGCGGAAGGGGAATTTTACTCTGTAACTCACAAGTCTCTCCTTTACTCAAGTCGACTCAAAACCTCCTCGTGGTGGTCCCCGGTAATGCTAAACTTGTTTAGCAGCTAA (SEQ ID NO: 1242)	TAGCTAAAACGTTTGGTTCAAAACATTTGCTTGCTGTCTTGGCATAACATCAATAAAGGCATAAACATCGCAAATAATGGTAATATATAAATTGGCTATGAGGATGGTTTTAGTACGTAGGCGTTGCGGAACTTCGGTTCAGATAGAGCAATGAATCGTGCATGCTAGGAAACTGAAGTGTTGACAGACCTAGTATCTTTCGATAGATTTCCATACCTCCGCGATCAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1365)	FERQNFSDGLVPQRKFIHIGTTNRNNEPRSNLRNLMTTRPSVDIFPEDQYEPNAAATLSRVPCTVCGRSFNSKRGLGVHMRSRHPDELDEERRRVDIKARWSEEEKWMMARKEVELTANGHKHINKQLAVYFANRSVEAIKKLRQRGDYKEKIEQIRRQSALVPEVANLTIRRRPSRSEQNHQVTTSETTPITPFEQSNREILRTLRGYSPVECHSKWRAQELQTIIDRAELEGKETTLQCLSLYLLGIFPAQGVRHTLTRPPRPRPNRRESRRQQYAVVQRNWDKHKGRCIKSLLNGTDESVMPSQEVMVPYWREVMTQPSPSSCSREVIQMDHSLERVWSAITEHDLRASRISLSSSPGPDGITPKTAREVPSGIMLRIMNLILWCGNLPHSIRLARTVFIPKTVTAKRPQDFRPISVPSVLVRQLNAILATRLNSSINWDPRQRGFLTVQSRIELSSSPGPDGITPKTAREVPSGIMLRIMNLILWCGNLPHSIRLARTVFIPKTVTAKRPQDFRPISVPSVLVRQLNAILATRLNSSINWDPRQRGFLPTDGCADNATIVDLILRDHHKRCKSLYIATLDISKAFDSVSHAAVSATLTAYGAPKEFVDYVQNTYEGGGTSLNGDGWSSEEFVPARGVKQGDPLSPILFATFTMGLQVLLDRTVLIEFGVSGNVEVLSATLNWRGIWSNQSVRRLIAKGLISSGDSNVISARVVTGGLYCFRQFMYLAGYTRDWTDEWKYLGLNADKCFTVGIKGQPKQKCTVLEAQSFYVGSREIPSLKRTTLIDEWKYLGINFTATGRVRCNPAEDIGPKLQRLTKAPLKPQQRMFALRTVLI

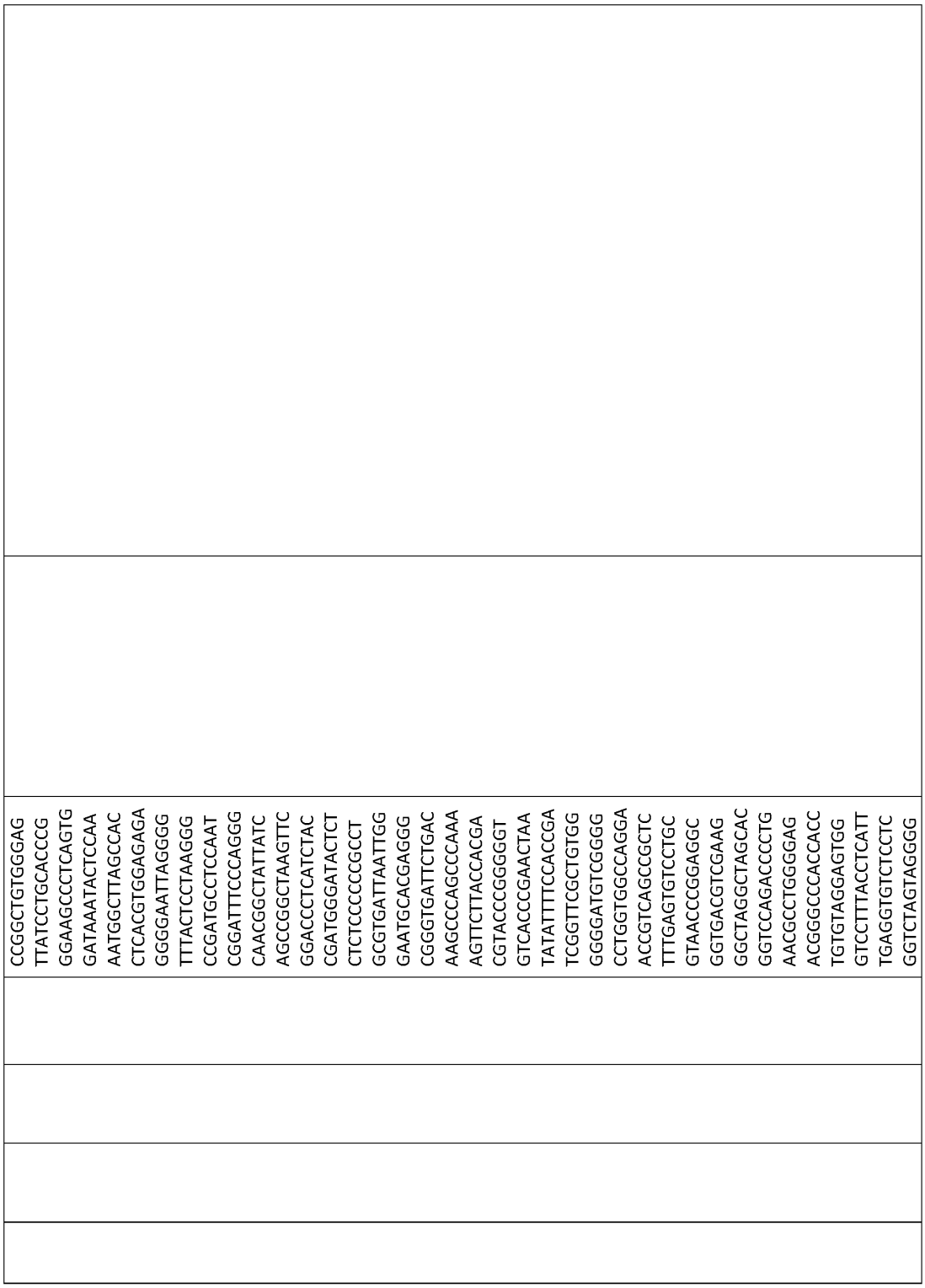
						PQLYHKLALGSVAIGILRKTDKLIRYYVRRWLNLPLDIPIAFIHAPPKSGGLG IPSLRWVAPMLRLRRLSNIKWPHLTQNEVASSFLEAEKQRARDRLLAEQ NELLSRPAIEKYWANKLYLSVDGSGLREAGHWGPQHGWVNQPTRLLTG KEYIDGIRLRINALPTKSRTTRGRHELERQCRAGCDAPETTNHIMQKCYRS HGRRIARHNCVVNRIKRGLEERGCVVIVEPSLQCESGLNKPDLVAILRQNH IDVIDIQIVTDGHSMDDAHQRKINRYDRPDIRTELRRRFEAAGDIEFHSAT LNWRGIWSGQSVKRLIAKGLLSKYDSHIISVQVMRGSLGCFKQFMYLSGF SRDWT (SEQ ID NO: 1487)
R2	R2_DSi	.	Drosophila simulans	GGGATCTGGGGTA ATTGCGAGCAGAG GGGGAGTATTTTTC TGTAATTCGTAAGT CATATCATATGGTG TGCGGAAGGGGAA TTTTACTCTGTAACT CACAAGTCTCTCCTT TACTCAAGTCGACT CAAAACCTCCTCGT GGTGGTCCCCGGTA ATGCTAA (SEQ ID NO: 1243)	TAGCTAAAACGTTTGGTTCA AAACATTTGCTTGCTGTCTT GGCATAACATCAATAAAGG CATAAACATCGCAAAATAAT GGTTATATATAAATGGCTAT GAGGATGGTTTTAGTACGT AGGCGTTGCGGAACTTCGG TTCAGATAGAGCAATGAAT CGTGCATGCTAGGAAAACT GACCACACGCAGTGTTGGC AGCCCTAGTATCTTTCGATA GATTTCCATACCTCCGCGAT CAAAAAAAAAAAAAAAAAA AAAAA (SEQ ID NO: 1366)	TCLAANLSGKNFSDGLVTQRKFTHIGTTNTNNEPRISLHNLMTTRPSVDIF PEDQYEPNAAATLSRVPCTVCGRSFNSKRGLGVHMRSRHPDELDEERRR VDIKARWSEEEKWMMARKEVELTANGHKHMNKQLAVYFANRSVEAIK KLRQRGDYKEKIEQIRGQSALVPEVANLTIRRRPSRSEQNHQVTTSETTPI TPFEQSNREILRTLRGYSPVECHSKWRAQELQTIIDRAELEGKETTLQCLSL YLLGIFPAQGVRHTLTRPPRRPRNRRESRRQQYAVVQRNWDKHKGRCIK SLLNGTDESVMPSQEVMVPYWREVMTQPSPSSCSGEVIQMDHSLERV WSAITEHDLRASRISLSSSPGPDGITPKSARE VPSGIMLRIMNLILWCGNL PHSIRLARTVFIPKTVTAKRPQDFRPISVPSVLVRQLNAILATRLNSSINWD PRQRGFLPTDGCADNATIVDLVLRHSHKHFRSCYIANLDVSKAFDSLSHA SIYDTLRAYGAPKGFVDYVQNTYEGGGTSLNGDGWSSEEFVPARGVKQ GDPLSPILFNLVMDRLLRNLPSEIGAKVGNAITNAAAFADDLVLFAETRM GLQVLLDKTLDFLSLVGLKLNADKCFTVGIKGQPKQKCTVLEAQSFYVGS REIPSLKRTDEWKYLGINFTATGRVRCNPAEDIGPKLQRLTKAPLKPQQR MFALRTVLIPQLYHKLALGSVAIGVLRKTDKLIRYYVRRWLNLPLDVPIAFI HAPPKSGGLGIPSLRWVAPMLRLRRLSNIKWPHLTQNEVASSFLEAEKQ RARDRLLAEQNELLSRPAIEKYWANKLYLSVDGSGLREAGHWGPQHGW VNQPTRLLTGKEYIDGIRLRINALPTKSRTTRGRHELERQCRAGCDAPETT NHIMQKCYRSHGRRVARHNCVVNRIKRGLEERGCVVIVEPSLQCESGLN KPDLVALRQDHIDVIDIQIVTDGHSMDDAHQRKINRYDRPDIRTELRRRF EAAGDIEFHSATLNWRGIWSGQSVKRLIAKGLLSKYDSHIISVQVMRGSL GCFKQFMYLSGFSRDWT (SEQ ID NO: 1488)
R2	R2_DYa	.	Drosophila yakuba	GGGAACATGGGGT AAAGGTGAGTAGA GGGGGAGTATTTTT TATACTCTGCAACTC ATAAGTCTTGCCTTT ACTCAAGTCGACTC AAAACCTCCTCGTG GTGTTTCCCGGTAA	TAGCTTAAAACGTTTGGTTC ACATACATCTGCCTGCTGCC TTGGCACAATATCAAAAAG GCATAAACATCGCACATATT GGTTATTTACGGCTATGAG GATGGTTTTAGTACGTAGG CGTTGCGGAACTTCGGTTC GGATAGAGCAATGAATCGT	FERRIFPKGLVPLTKDNHIGTTNLQNEPRIFTNDLLTTRPSVDHVPEDQYE PNAAATLSRVPCTVCDRSFNSKRGLGVHMRSRHPDELDEERRVDIKAR WSEEEKWMMARKEVELMANGFKHINKQLAVYFANRSVEAIKKLRQRG DYKEKIEQIRGQSALAPEVANLTIRRRPSRSEQDHQVPTSEASPITPLEQS NREILRTLRGYSPVVCPSKWRAQELQTIIDRAEFEGKETTLQCLSLYLQGIF PVQGVRHTLTRPPRRPRNRRESRRQQYAVIQRNWDKHKGRCIKSLLNG TDESVMPSREFMEPYWREVMTQPSPSSCNGEVIRTDHSLETVWSAITE QDLRASRVSLSSSPGPDGVTPKTAREVPSGIMLRIMNLILWCGNLPHSIRL

				TGTTAAACTTGTTTA GCAGCTAA (SEQ ID NO: 1244)	GCATGCTAGGAACTGACCA AATAACAGCAGCCCTAGTA TCTTTCGAAGATTTCCATAC CTTTGCGATCAAAAAAAAA AAAAAAAAAA (SEQ ID NO: 1367)	ARTIFIPKTVTAKRPQDFRPISVPSVLVRQLNAILATRLTSSIDWDPRQRGF SPTDGCADNATIVDLVLRHSHKYFKSCYIANLDVSKAFDSLSHAAIYGTLR AYGAPKGFVDYVQKTYEGGGISLNGEGWCSEEFVPARGVKQGDPLSPIL FNLVIDRLLRALPSEIGTKVGNAMINAAAFADDLVLFAETRMGLQTLLDK TVDFLSTVGLKLNADKCFTVGIKGQPKQKCTVLEAQSFCVGSREIPTLKRT DEWKYLGIHFTASGRVRCNPAEDIGPKLQRLSEAPLKPQQRLFALRTVLIP QLYHKLSLGSVTIGVLRKTDKLIRFYVRRWLNLPSDVPIAFVHAPPKCGGL GIPSLRWVAPMLRLRRLSNIKWPHLVQSEEASSFIEAEKQRARGRLIAEQ NELLSRPAIEKYWANRLYLSVDGGGLREAGHYGPQHGWVSQPTRLLTGK EYLDGIRLRINALPTKSRTTRGRHELERQCRAGCDAPETTNHIMQKCYRS HGRRVARHNCVVNRIKRGLEERGCVVIAEPSLQCESGLNKPPDLVVLRQN HIDVIDVQVVTDGHSMDEAHQRKINRYDRPDIRTELRRRFEAAGDIEFHS ATLNWRGIWSGQSVKRLIAKGLLSKYDSHIISVQVMRGSLGCFRQFMYLS GFSRDWT (SEQ ID NO: 1489)
R2	R2_KF	GU949558	Kalotermes flavicollis	GAGAGGTCACTGTT GCTGATCAGCTGGA CACCTAGCTGACTG CGTGCATGCGTGCA CGCCGCGCGCCGCC TCCTCGTGGTACCG CTGGTAACGCAGG GTCACTGAGACTAC CTGCGGCTAAAGGC GCGCGCGAGGGAT TGTAAGTCCACCAC CTCCACGTGGTTCC CCGCGGGCAACGG CATAGTTCATCTGC CGGCCAAGCGTGCC GTTCCCTCGATCCTC CTCTGATTAGGGAT AGGAGGGGGGCGG TGGCTCCCGCCGAC GACCTGGCAAACAC CTTGGACACGCTAA GATAATAGGCCGTC CTCCGGGCCGGCCG AATCATAAGCACAC	TAATGTCCCTTTGGCTTGC CCCCACCTGCTTAAAGGAAC TGGCAGGAAAGAGAGTGAT CCGTGCCATAGAAATATGG TTATCCGGGGCAAGTCACT AGCAATATGGGACTTCTCC GGGTCCGTGCGGTCCTTCC AACATGAGCTGGACGTAGT CCACTCTATGACTTGAACGA TACGGGGGCGTATCTCCCC CGGAAGAGGTCGCCTAGGC GACTTAGAA (SEQ ID NO: 1368)	MEVPLTSSLGSGATQAPGTPELLGEHTVERPGLDQGHSYGLLMDDVELP VRLPFFGPLICPGCRTLLTSEETISSHHRRVHPDARTRWVCYGCDSPFMTY RAIKCHLPKCSGRKVVTGDHICNGCTKRFESQRGLSLHKRRAHPGLRNEE MLEPPVRAERRPNAHKSSIWSIDEIRILEQYEAAYVGDLHINMKIAAHLPF KTNKQVSNYRNDRRKKSRTATDASQQGLGPNDGNRGIVPSGQSSPLFLE GSDAEGDEDVFNVLVPPTLGGLEPAGQVHSLSEGETSPLVGEADPCFMG GTPSAGEASGSTLLGPDPTPADGYSLVRKDLQLSVQTSPLLAVGSVGTES VQFERGVLSCGTPPEFLHPEQFAHCANNDPVLNASEEQVHAPLGEEAN DLPDNNHPSELGVDPEDPTCSPATEQVQPSSEEEADDPFAQFKAWRRR VASYALKIETGVLPAQVDDLLRRLRDGDTQSKVTCEEVEEVVLSLTRTILG GTAPKKRVEGRTKWTYKSRTNHEARKRIMYARCQDLYRRRPQRPVERA VGYQAEESLLDNQDERPSHGAFETFYTGLWGKSGQCNITMPPGVPRHT GHVLREVTPKDIYSRLRKLKKDYAPGPDGVTKLKVQSMGAYPSLLAKVYN LVMLTGYFSSCWKEHKTSLIPKDRGSPMDVSNWRPITIGSLLSRIYTGLIE RRLRTVSDIHQRQVGFMPVNGCAANLFIFDECIRQAKKEGTIVGSLIDVA KAFDTVPHEAILRALSSQGVDEHTMAHIRDMYSGIRTRINGKGSDIPLVR GVKQGDPLSPMLFNMVMDPLIRDLQRKGFRIGGHEIGALAFADDIVLLA DSIDGAQDHVDQVGRYMNKLGMTLNPRKSSSFLITAMRKTWICRDPGL SIGETKVPGARPSSALKYLGVNYTLSEGLESGALIDKLMQAVNRARGLALK PLQKVNLILERIIPKFLYGIILGGPSLTRLHAADKCVRMAVKEILHLHPSTTD HVLYARKKDGGMGIPRLAHLVRLASLRSGLALLASGDVAVQAAGMAGD LEGRCKKVANDLRLNWPVTLRDVVRASNKFKSQESKDWERLASQQHGV KDFRNDRLGNCWLYDPTVLSSSRYTDALRLRTNTFGVNVALRRADKDLE VNCRRCHGKPETLGHVLGECVAGKGMRIQRHDKMAAAFVATKCEEKGY

			A (SEQ ID NO: 1245)			QTTREQLFSIEQGKLKPDLVVIDGERALIVDVTVRFESGNALSRGASEKIEK YQPLADYFVSQGAVREANVLPIVVGSRGAITQATLKSLATLGLDVERVGK YLAICAVASSVEIACMHLDYT (SEQ ID NO: 1490)
R2	R2_RL	GU949555	Reticulitermes lucifugus	CTCACTGCTGTCAT	TGAGTTAGATATATATGAC	MMADYNNSVDHALEDNTRLIFARDAVLARVCGPFDNLECGLCGVLLTSL
				GCCATTGTTAATGC	GACTGATTCTTCCCTAACTA	QGVREHCHRAHHNLDLTFQCTKCDKGFSSYRGICCHFSKCIGARISVSEG
				GTTGGGTGATGGC	AAATATAATTTGGAATGTG	PLSCSECEREFDSKRALSTHERHMHPGIRNAKRLKDFNPRGGGGKTIHG
				GGGTGATGGGAGA	CATTACTTATACTTTATTGTA	NTKWTEEEVQLLVSLSKRFEGYKSINKEISLILTSKTCKQISDKRRYLNLHNG
				CGAGTTACAGCAGA	TTTATTGACATCCGTAGTAG	NGGLAAAEAVLEFCEDSHPEVTESDGAVLSEIMDEECHQSSVTMRSSIV
				GCTGGCTTCACGGG	TCTGTTTAGTGGATTTAGAT	HGDIGREVQGKELVRIPPDNSVMGNCVVLLRKLATDKRSDLDLSKDKELR
				GGGGGCTGTAGTT	CGTTAACCATTCTGTGACGT	IDIEKATKANRESADGRVIQSVAGNEIDPDTFQWKELLLGQVRGFPRVDE
				GCCCGAACCAGTCT	CGCTGATGTCTATATGTACC	NSELFDLDDKLTKELSSDSPVWNDNCELIVSDLCQVLCKKKYELGRQHHV
				GCTTCTGAGTGCCT	AGTGGCAGAATGCTGACTA	RKGKRHRGIHHKREKFRECQKIFRKSPRKLAEYLYRDKDLSHISKDASTPQ
				CCACGTGGCCTCGC	AGAACAAAATAATTTTAAAC	GIEQYYSQLWGEPELLESNTIEEKLPSSSLFDCLPPITPEEVEGRIHKIRPSSA
				TGGAAACGTCGGA	AATAGACAGCACTAAGAAA	PGLDGVRKIHLVGKGITLVLVKLYNLLFLTGGYPECWKRNRTVFIPKIGKDL
				GCTGCTTACGGCTA	TTCTGCAAACTGGAACGTG	SEVGGWRPLTIGSLLARMYSAFLERRIRRVTSLSLSQRGFTNIQGCHVNLT
				ACCCGTAGACAACT	GCCCACGGTAGGCCAATAT	ILKEGIRQAKVKNGGVIVSVDIEKAFDTIPHSVIFSRLASQGVPPLLRKIISN
				CCGGCGGCCAAACT	ACCGGAAAGGGAAAATGAC	MYKDVYTVIEGQCIPIKRGVKQGDPLSPLLFNIAIDPVLRSLEEFQGGLPLG
				CAGAAAACGGCCTT	ACATCCCCCCCTTAAA (SEQ	NSAIKILAFADDIILGASSAGQAQQMVDMLGIGLTSCGLGVSHRKCFGFQ
				ACTACCAGGGTCAT	ID NO: 1369)	IVNKNKTWTIVDPMITLNGSSLPFSGPEDRLPYLGVDINPWDRKSRYDAG
				CCCTCGCGGCGGCG		QRLISAAKRGSQLSLKPQQKINLITAFLLPKFLYILIEDPPSPAYLKSIDHDLR
				CTAAGACCTTACTG		QIYKNILHLPNCVSTAFMYSPKRDGGLGLPRLSCLVPLAHLKAGIKLGSLQ
				TATGTACTACGGAC		DSLVREITTSDRFVRTMGSVAHSLWASWSLTLQDIYKLKSALKRREAKAW
				TACAGTTGAGTAGC		ESCVSQGQGAAQFRGDSIGNNWLHNPGTYRPGQYIEALKLRANLTGVR
				GCGGGAGAGCACT		VNLKRSGYNVPITCRFCKDIPETQAHVLGLCPKTKGMRIQRHDSIVNRVR
				TGATGTGAGGGTA		DKLKTKSPVALMHEQNFTVEEGQVFKPDIVTILGEVGYVIDVTVRYDDRD
				GCAACTTGTTGCTA		YIKDASVEKIRKYEALKGYLKDLYPQLNKVEVLPLVFGSRGAVPGSTVHN
				CTTCGACTGTCTCCC		MGLLGFTKREMVHISRKVIADSLIISNFLEVY (SEQ ID NO: 1491)
				TGAAGATTTCCGAA		
				GGGGTGCCGCAAG		
				TCCATGGGGGCGG		
				TCCTGCTGTGGGAG		
				TTATCTTGCACCCG		
				GGAAGCCCTCAGTG		
				GATAAATACTCCAA		
				AATGGCTTAGCCAC		
				CTCACGTGGAGAGA		
				GGGGAATTAGGGGG		
				TTTACTCCTAAGGG		

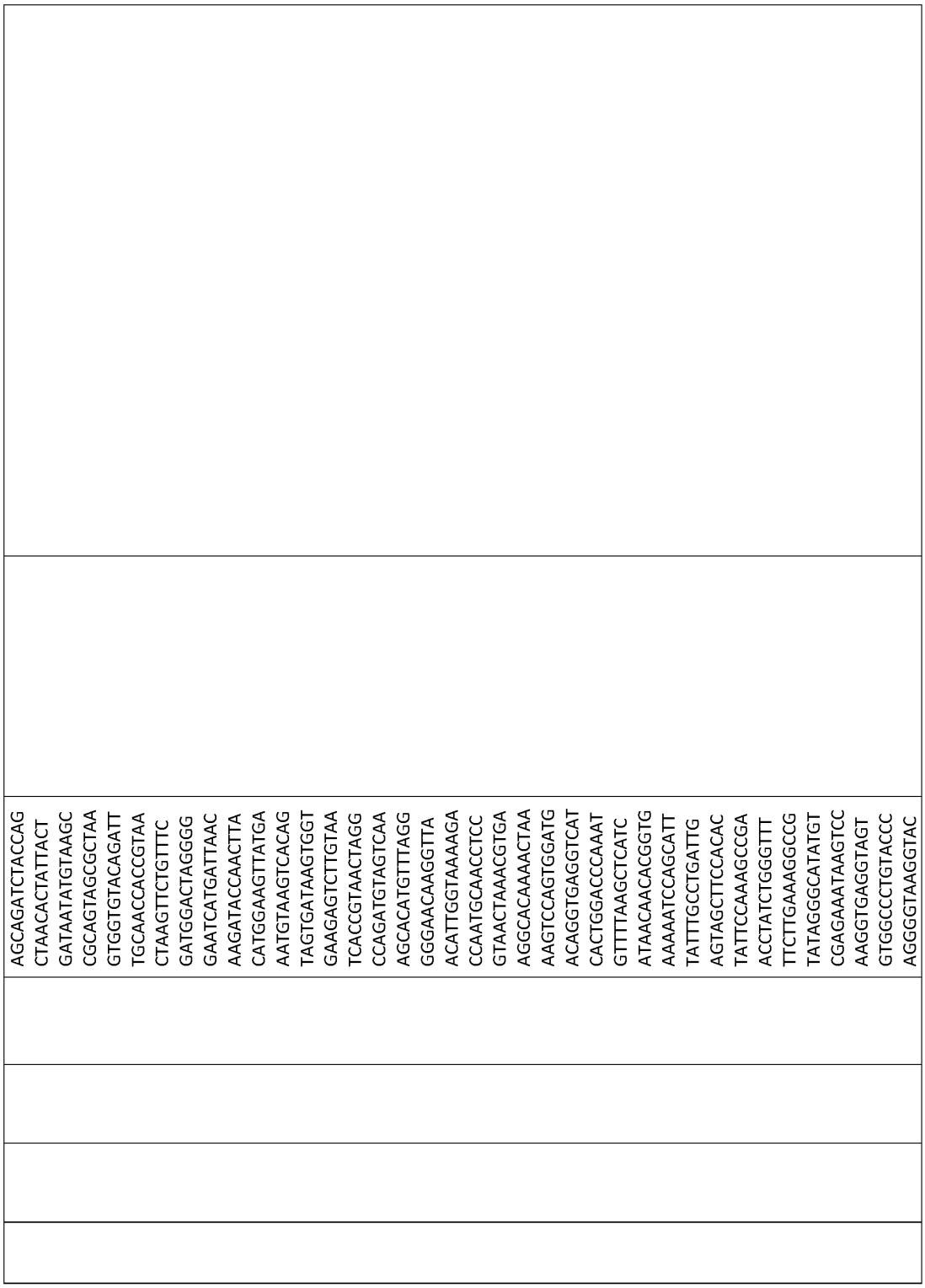

CCGATGCCTCCAAT
CGGATTTCCCCGGG
CAACGGCTATTATC
AGCCGGCTAAGTTC
GGACCCTCATCTAC
TGATGGGATACTCT
CTCTCCCCCCGCCT
GCGTGATTAGTTGG
GAATGCACGCGGG
CGGGTGATTCTGAC
AAGCCCAGCCCAAA
AGTTCTTACCACGA
CGTACCCGGGGGT
GTCACCCGAACTAA
TATATTTTCCGCCGA
TCGGTTCGCTGTGG
GGGGATGTCGGGGG
CCTGGTGGCCGGG
AACCGTCAGCCGCT
CTTTGAGTGTCCTG
CGTAACCCGGAGGC
GGTGACGTCGAAG
GGCTAGGCTAGCAC
GGTCCAGACCCCTG
AACGCCTGGGGAG
ACGGGCCCACCACC
TGAGTAGGAGTGG
GTCCTTTACCTCATT
TGAGGTGTCTCCTC
GTTCTTGTAGGGGG
CGAAGGACGAGGA
ATGCATCCGTCCCT
CCAGGATGCTGGTG
GTTTCCGTCTGGTG
GGCTCATGCTGAAA
CGCGCAGGTGCCG
ATTTGTCGAAGAGG
ACATATGGGGTAAC

			CCATAGAACCTAGG CAGGAGTAATCCCT TAGCTGGGGGGGT CAGCTGGTGGGTCC TGTCAATTTATCCTC CCCTCCATGCCAGA GCCGGTCCGAGGTT AGAGACGGACTCAT TTTCTCCTTTTTATA TGTC (SEQ ID NO: 1246)		
R2	R2_RU	GU949554	Reticulitermes urbis	CACACTGCTGTCAT GCCATTGTTAATTC GTTGGGTGATGGC GGGTGATGGGAGA CGAGTTACAGCAGA GCTGGCTTCACGGG GGGGCTGTAGTGG CCCGAACCAGTCTG CTTCTGAGTGCCCC CAAGTGGCCTCGCT GGAAACGTCGGAG CTGCTTACGGTTAA CCCGTAGACAACTC CGGCGGCCAAACTC AGAAAACGGCCTTA CTACCAGGGCCATC CCTTGCCGGTTGCGC TAAGACCTTACTGT ATGTACTACGGACT ACAGTTGAGTAGCG CGGGAGAGCACTC GATGTGAGGGTAG CAACTTGTTGCTTCT TCGACTGTCTCCCT GAAGAGTTCCGAA GAGGTGCCGCAAG TCCATGGGGGCGAC	TGAGTTAGATATATATAAC GACTGATTCTTCCCTAACTA AAATATAAATTTGGAATGTG CATTACTTATACTTTATTGTA TTTATTGACATCCGTAGTAG TCTGTTTAGTGGATTTAGAT CGTTAACCATTCTGTGACGT CGCTGATGTCTACATGTACC AGTGGCAGAATGCTGACTA AGAACAAAATAATTTTAAAC AATAGACAGCACTAAGAAA TTCTGCAAACTGGAACGTG GCCCACGGTAGGCCAATAT ACCGGAAAGGGAAAATGAC AAATCCCCCCTTAAA (SEQ ID NO: 1370)	MMADYNNSVDHALEDDTRFIFARDSVLARVCGHFDNLKCELCGVLLTSL QGVREHCHRSHHNLDLTFQCTKCDKGFSSYRGICCHFSKCRGARISVSEG PLSCSECERKFDSKRALSTHERHMHPGIRNAKRLKDFNPRGGGKTIHGNT KWTEEEVQLLVSLSKRFEGYKSINKEISLILTSKTCKQISDKRRYLNLHNGN GGLAAAEAVLVFCDDSHLEVTDSDGAVLSEIMDEEYYQSSLTMRSSIVHG DIGREVQGKDLVRIPPDNSVMGNCVVLLRKLATEKRSDLDLSKDKELRIDI EKATKANRESADGRVIQSVADNEIDPDTFQWKELLLGQVRGFPRVDENS ELFDLDDKLTSELSSDSPVWNDNCELIVSDLCHVLCKNKYELGRQHHVRK GKRHRGIHHKREKFRECQKIFRKSPRKLAEYLYRDKDLSHISKDVSTPQGIE QYYSQLWGKPELLESNTTEEMLPSSSLFDCLPPITPEEVEGRIHKIRPSSAP GLDGVGKIHLVGKGITLVLAKLYNLLFLTGGYPECWKRNRTVFIPKIGKDLS EVGGWRPLTIGSLLARMYSAFLERRIRRVTSLSSSQRGFTNIQGCHVNLTI LKEGIRQAKVKNGGVIVSVDIEKAFDTIPHSVIFSRLASQGVPPLLRKIISN MYKDVYTVIEGQCIPIKRGVKQGDPLFPLLFNIAIDPVLRSLEEFQGGLPLG NSAIKILAFDDDIILGASSAGQAQQMVDMLGIGLTSCGLGVSHRKCFGFQ IVNKNKTWAIVDPMITLNGSSLPFSGPEDRLPYLGVDTNPWDRKSRYDA GQRLISAAKRGSQLSLKPQQKINLITTFLLPKFLYILIEDPPSPAYLKSIDHDL RQIYKNILHLPNCVSTAFMYSPKRDGGLGLPRLSCLVPLAHIKAGIKLGSLQ DSLVREITTSDRFVRTMGSVSHSIGASWPLTLQDIYKLKSALKRREAKAWE SCVSQGQGAAQFRGDSIGNNWLHNPGTFRPGQYIEALKLRANSTGVRV NLKRSGYNVPITCRFCKDIPETQAHVLGLCPKTKGMRILRHDSIVNRVRD KLKTKSPVALMHEQNFTVEEGQVFKPDIVTILGEVGYVIDVTVRYEDRDYI KDASVEKIRKYEALKGYLKDLYPQLNKVEVLPLVFGSRGAVPGSTVHNMG LLGFTKREMVHISRKVITDSLIIISNFLEVY (SEQ ID NO: 1492)

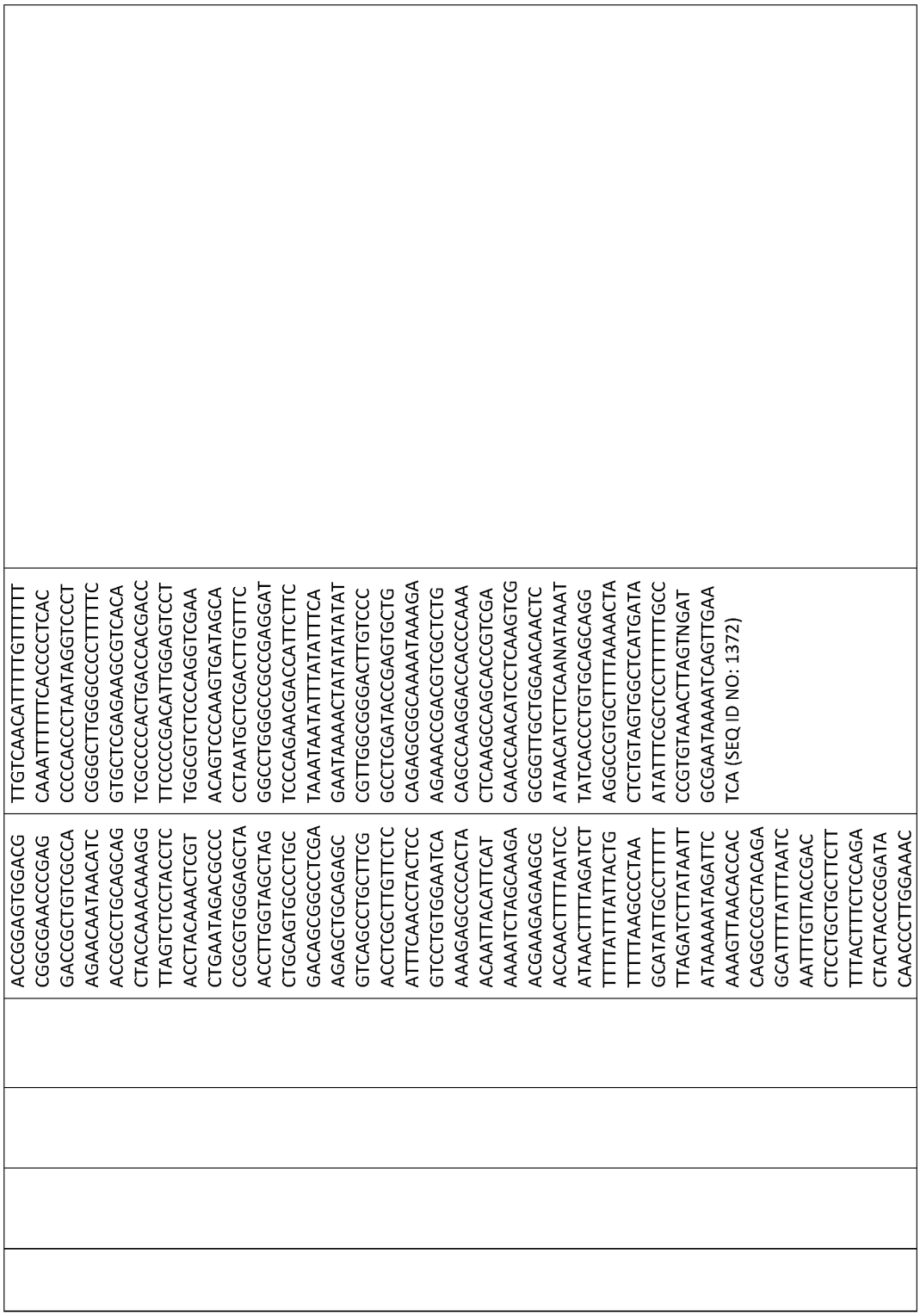

CCGGCTGTGGGAG
TTATCCTGCACCCG
GGAAGCCCTCAGTG
GATAAATACTCCAA
AATGGCTTAGCCAC
CTCACGTGGAGAGA
GGGGAATTAGGGG
TTTACTCCTAAGGG
CCGATGCCTCCAAT
CGGATTTCCCAGGG
CAACGGCTATTATC
AGCCGGCTAAGTTC
GGACCCTCATCTAC
CGATGGGATACTCT
CTCTCCCCCCGCCT
GCGTGATTAATTGG
GAATGCACGAGGG
CGGGTGATTCTGAC
AAGCCCAGCCCAAA
AGTTCTTACCACGA
CGTACCCGGGGGT
GTCACCCGAACTAA
TATATTTTCCACCGA
TCGGTTCGCTGTGG
GGGGATGTCGGGG
CCTGGTGGCCAGGA
ACCGTCAGCCGCTC
TTTGAGTGTCCTGC
GTAACCCGGAGGC
GGTGACGTCGAAG
GGCTAGGCTAGCAC
GGTCCAGACCCCTG
AACGCCTGGGGAG
ACGGGCCCACCACC
TGTGTAGGAGTGG
GTCCTTTACCTCATT
TGAGGTGTCTCCTC
GGTCTAGTAGGGG

			GCGAAGGACGAGG GATGCATCCGTCCC TCCAGGATGCTGGT GGTTTCCGTCTGGT GGGCTCATGCTGAA ACCCGCAGGTGCCG ATTTGTCGAAGAGG ACATATGGGGTAAC CCATAGAACCTAGG CAGGAGTAATCCCT TAGCTGGGGGGGT CAGCTGGTGGGTCC TGTCAATTTATCCTC CCCTCCATGCCAGA GCCGGTCCGAGGTT AGAGACGGACTCAT TTTCTCCTTTTTATA TGTC (SEQ ID NO: 1247)			
R2	RaR2	FJ461304	Rhynchosciara americana	CAGAACGTGGAGA AACGGAATAACTAC CCAGATCCGTTGGT TAACCGGTGGCAAA GTTAATCAAGGTTG CCATAGGCTTAATA ACCCTATGGAAATG TTTCCACACACCTCC ACGTGGTGCCTGCC GGAAATTGTTCTAG GGTGAACAAGCTAA GTTGTGAGAAACG GGCTCCACCACAAT ATGGAGCCTGCCAG GGCGCGAGACTCA GGACTCTCCATGTA CAAAGTGGTTAGTT GCAAAAAGAGTGC GCCTAGCATGACTG	TAGAAATGTGTGCGATAAG GTGTGAATAGAAGGGTTCA CCAAGAGGGAGACCTAGTT TGGACCTCAGAAATGGGGT CATAGGAGTGATAGGTTGT AAAGCCGTTGGGGAATCCG GCTACACATGGTATCTCAG GAGCCATTCATGCGCTGAT CTCATTAAGGCGTAATAAAC TGTGAAACAGATCCTGATA ATGCCGTGCTACCAAATGA TGTAACGAGGCGGAAATAA AATTAATCTGGGGCGTTCT GCGGAATGACTACTAAATA TAGCGATGCTATATATACAA ACGACTGATGGTAACACCG GCCTTA (SEQ ID NO: 1371)	MSNYNETNTSGGDNPRMATQTTGSLSSGPINQHTCELCCRTFGTRAGL GQHVRKTHPIESNQSINVERKKRRWSPEEIRRMANMEAQATINNIKHLT QYLATYLPQRTLNAIKGRRRDAEYKELVTGIIANLRSNSSTQQTNQVCNES EMSQRSKILQSIRESVRDLRSRRNKYAKALQELGEAALCGKMLNEEQLIH CIKSMFNTAKCPKGPRFRKTATHSGTNKQQRQQRYARVQKLYKMNRKV AAKMVLEETDKIQIKLPDHDPMFKFWESEFKEGEGMPERMPKDLKESP DLKAIWDPVTEEEVRKAKVANNTAAGPDGIQPKSWNRISLKYKTLIYNLL LYYEKVPHKLKVSRTVFIPKKKDGSSDPGEFRPLTICSVVLRGFNKILVQRL VSLYKYDERQTAYLPIDGVGTNIHVLAAILNDSNTKLSELHVALLDITKAFN RLHHTSIIKSLVGKGFPYGFITFIRRMYTGLQTMMQFEGHCKMTQVNRG VYQGDPLSGPIFLLAIEKGLQALDKEVGYDIGDVRVNAGAYADDTDLVAG TRLGLQDNINRFSSTIKQVGLEVNPRKSMTLSLVPSGKEKKMKVETGKPF RANDVPLKELSINDFWRYLGISYTNEGPERLSLTIEQDLERLTKAPLKPQQ RIHMLNAYVIPKYQDKLVLSKTTAKGLKRTDRQIRQYVRRWLKLPHDVPI AYLHAPVKSGGLNIPCLQYWIPLLRVNRVNKITESQRSVLAAVGKTALLTS TVYKCNQSLATLGGNPTMLAYRTYWEKELYAKVDGKDLQNARDDKAST RWNGMLHSDISGEDYLNYHKLRTNSVPTKVRTARGRPQKETSCRGGCK STETLQHVVQQCHRTHGGRTLRHDRIVGLLQHELRRDYNVLAKQELKTG IGLRKPDLVLIKDDTAHIVDVQVARCSKLNESHVRKRSKYDKKEIEVEVKSR

		ATAATTTTTCACTGA AATAACGTTGAACT TTATCTGTGTCATGT GCACAACACTATGG TGTCTGATCAAGCA CCATCAGTGGTGGA CCTGCTAATGTATT AGTAGAACGTGTCC AGGCGATAATGCAC ACACGGCCTCCGGG CCCATCGCTTTTTTT GAGATTCCCTAGAA ACTTCAGTGTGTGC GACAACTGTATAAC CCATAAGGGATGG ACAAAGGTTATACT AGGGGGTAAAAAC CCTAATCGGCTAAT GGCAAATGGGATG TAGAAATGCCAAAG ATTACTCGCACCGA ATAATGGTGGCCGA AAAGCGGGTAATC GAATGAAATGGTAA TGCTGTAGCGGAAA CATGATCACATTCT GTGACAGTAAACCA TTAGACCTAGGGGG AACTATGATTAACA AGATACCAGCTTAC ATGGAAGCAATGA AATGTAAGTCACAG TAGTGATAAGTGGT GAAGAGTCTTGTAA TCACCGTAACTAGG CCAGGTTCTGGGGA TGCCATGAACTTAG GGGGAGTATGGTT		YRVSKVMYEACTISYKGIWDKQSVMSMRRLGVSEYCLFKIVTSTLRGTWL CWKRFNMITSVRS (SEQ ID NO: 1493)


AGCAGATCTACCAG
CTAACACTATTACT
GATAATATGTAAGC
CGCAGTAGCGCTAA
GTGGTGTACAGATT
TGCAACCACCGTAA
CTAAGTTCTGTTTC
GATGGACTAGGGG
GAATCATGATTAAC
AAGATACCAACTTA
CATGGAAGTTATGA
AATGTAAGTCACAG
TAGTGATAAGTGGT
GAAGAGTCTTGTAA
TCACCGTAACTAGG
CCAGATGTAGTCAA
AGCACATGTTTAGG
GGGAACAAGGTTA
ACATTGGTAAAAGA
CCAATGCAACCTCC
GTAACTAAACGTGA
AGGCACAAAACTAA
AAGTCCAGTGGATG
ACAGGTGAGGTCAT
CACTGGACCCAAAT
GTTTTAAGCTCATC
ATAACAACACGGTG
AAAAATCCAGCATT
TATTTGCCTGATTG
AGTAGCTTCCACAC
TATTCCAAAGCCGA
ACCTATCTGGGTTT
TTCTTGAAAGGCCG
TATAGGGCATATGT
CGAGAAATAAGTCC
AAGGTGAGGTAGT
GTGGCCCTGTACCC
AGGGGTAAGGTAC

				TATACGGTCGAGTG GCTCAGTAGGCCTA ACTAGCCACTGAGT CACTATAATGACTA GTT (SEQ ID NO: 1248)		
R2	YURE-2_Cis	.	Ciona savignyi	CCAAAATTACTTCC	TAAGATCCGCGGCTGTGGC	MAGHKITMSEGKLLEVAVRYGGVRNVSYECPVPDCTKTFSQANNLIRHL
				AGCACCTCCACAGC	GCCGAACGAGCACCTGCCC	NNFGNTKHRAHNFTYFFTCEKCKIQIHSNTKHNISNHYKQCCATGGGPS
				AGACGAACGAAGA	ATTCTTCTTGTAGGGACTTT	CETGQYFCPACEQAGLGNLESALRHFQSSHPEFNLPPRSQFSKSHPNSYT
				AAGAAGACTTACGA	TTCACCCTCACTCCCCCCAA	LSLKPKDHLMKILYSGPLTPGQLVCPIKICLRSSAARLFHDVSKLRKHMLVD
				AATAAGAATAAGCA	TAGTTTTTTTTCGTTTTTTCG	HNRTLVYETTCGKCLRPVDTSKNMRKTTSHFEKCSGESFISSPSPIPQKTY
				GTTTAAAGACGAAG	TTTTTTCACCCCCACCCCACT	KLDLPSTSTPPPRKSPKLQPYKPIRTFKNPLTKSSQSKSDNPPKPTPFFSPRT
				CAGACGAACACCAC	CGCCTCTGGGCTGCACATCC	LERSASWPALSEVVDPLPKLKEKHPSLPCALDKCPPSPRIKPSTLVPPCHTA
				TCCACCAACGACGC	CACACGTAGGGACCTGTTT	NNSPKPTSPESPSTLKPLPRPIIRPSKPLEDWLTVRSVGPDREIVLNIGPRPR
				TCTGCAGCTACACC	ATATTATTTGCCTTTTATATG	PGPAAGSRTTSPPSTAPAKRVAANPIAAPLSGEPGATLDCGQTGRKVQP
				ACCACCATCTCGCC	TACCACTTTTTAAATATATTT	PKKRPTESAGSLPPPAEPATDLLTGREGLARLVEEYHLSGDFGAFCRDLER
				GCAACAAGAAGAA	TTGTACCCCACAAGATGCTT	WTALSSTNRRPKPRRGRYNRGAAARATRNRGRDDRQDPQDRDDQGG
				TTTTCCGCTGCTCTC	TTCGCCAAAAAAAAAAAAT	PGPVTCGRPQRYKRAAALRSAFGRDMKATVRRIIDGERGDARCEIDPKTI
				CACTCTGCTCCAAC	TTTTGTATCACATTTTTATAT	EGRFRDELSPPVREGPECSLPPWMAEAQAGEHAPSNDSQPGDAYDGPI
				ACCACCTCTCCTGCT	TTTGTAAAACACAGATTTTT	TALEVEMVLSTLNVGSAPGSDGLSYGFWRALDPKGLVLSELFEVCRIERR
				CTGTGGACTGCTGC	ATAAACTTTGCACTATTTTT	VPGPWKSSRVTLICKDAEGDLDDLGNWRPISICQTVYKIYAAVLARRLQS
				CTTGCTGCTGGACC	ATATAAACTTCGCACTTATT	WALDGGVISRSQKGFMPFEGVYEHVFLLDSVVADARATRRSLAVCWLD
				AACCTCTACCCGAA	TAAAATGAATCGCATCTTTT	LRNAFGSVDHTTIVEALSRFGAPAGLVEMISDIYTGGSCRIRTRAGFTPDIP
				GGAACCCTTCGAAC	TTATATACACCAACACAAAC	VGRGVRQGCPLSGIIFNLVMEVLLRGVEANNACGYRLSCAGGASVRVLA
				CCAGCAGCAAGGTA	AGGATGTGCAGCTCAGGGG	YADDVALVGSSRAEXKIQLGVCERFAAWAGFSFNNKKCAAMVLKHQRG
				CGTGTCACCACCTC	AACCAATCCTGCGTCCCTCC	GRRLLDSAPLRLCGEEVAILGPDSFYKYLGAHTGYGRQTGGQLVDRVERQ
				TCCACAGAGCCAAG	TAGCGGCGGGAGGGGCGC	VVRLFTSFLTPTQKLSALKRIVLPAMSFHLRVRPCAEGHLRRLDNTVRRCV
				GCCAGAGTGGATA	CCACCTACCCCCACGCTCCT	KTALRLPKGSCRAFFHTSPDAGGLGITSVVAECDILTVTQAFKMLSSPDHL
				GAGCAGCGGCCTCT	CTTGAGCACCAACAGGGAC	VSLVAKGRLGMHAARMGRSETASACAMADYLSGDSVMGHXSWKTGY
				ACAACCAAGTTCTC	TCCCTTCCGGAGCCCCTGCA	RMPADLWTATRAASRRLSLRFSPQPQGEFGLESGTFKIAPRERRSLTRRL
				CACTCCGACGACAA	CCCTCAACTTTTCTTATTTTT	HHRQNLWWRNQWAALPNQGKTVAAHSAYAASNNWVKGPSSLAPQA
				AACACCTGCCTCGT	AAAAAAAAAAAATCATATA	LFFGLKARLNQMPTRSVKACYSRAPNYDKSCRRCGAEVETLPHVLNHCP
				GGCCAGAGTCCTGC	TATTGATCTTGACGACGGG	KSMKSILERHDSVLAEVLAAIPRGTFASVDVDRTSREHFRRVGEALRPDIV
				CGAAGAAGAATCTT	GGCTACATTCAGCCCCCAAA	ARRHDGSVVVADVTCPFESCASALDTAAARKIEKYDQLCANLRQLYRKPV
				CGACCCCAACACCG	AACCCACCCACCATCCCCAA	ESHALVVGSLGSWGRTNNTALAALGIRGAVRSRLAKQLVNLSVEGSHNI
				TCTCGGCAGGGCTC	CGAGTGCCGGGGCATTGAA	WLRWSGGIPKDLVR (SEQ ID NO: 1494)
				CAAGCAACGACATC	GAGCTCCGGCACAATTAGC	
				AGCCACGAGTGCCC	ACTTAGCTTATTTATTATTTT	


ACCGGAGTGGACG
CGGCGAACCCGAG
GACCGCTGTCGCCA
AGAACAATAACATC
ACCGCCTGCAGCAG
CTACCAAACAAAGG
TTAGTCTCCTACCTC
ACCTACAAACTCGT
CTGAATAGACGCCC
CCGCGTGGGAGCTA
ACCTTGGTAGCTAG
CTGCAGTGCCCTGC
GACAGCGGCCTCGA
AGAGCTGCAGAGC
GTCAGCCTGCTTCG
ACCTCGCTTGTTCTC
ATTTCAACCTACTCC
GTCCTGTGGAATCA
AAAGAGCCCCACTA
ACAATTACATTCAT
AAAATCTAGCAAGA
ACGAAGAGAAGCG
ACCAACTTTTAATCC
ATAACTTTTAGATCT
TTTTATTTATTACTG
TTTTTAAGCCCTAA
GCATATTGCCTTTTT
TTAGATCTTATAATT
ATAAAAATAGATTC
AAAGTTAACACCAC
CAGGCCGCTACAGA
GCATTTTATTTAATC
AATTTGTTACCGAC
CTCCTGCTGCTTCTT
TTTACTTTCTCCAGA
CTACTACCCGGATA
CAACCCTTGGAAAC
TTGTCAACATTTTTGTTTTTT
CAAATTTTTTCACCCCTCAC
CCCCACCCTAATAGGTCCCT
CGGGCTTGGGCCCCTTTTTC
GTGCTCGAGAAGCGTCACA
TCGCCCCACTGACCACGACC
TTCCCCGACATTGGAGTCCT
TGGCGTCTCCCAGGTCGAA
ACAGTCCCAAGTGATAGCA
CCTAATGCTCGACTTGTTTC
GGCCTGGGCCGCCGAGGAT
TCCCAGAACGACCATTCTTC
TAAATAATATTTATATTTCA
GAATAAAACTATATATATAT
CGTTGGCGGGACTTGTCCC
GCCTCGATACCGAGTGCTG
CAGAGCGGCAAAATAAAGA
AGAAACCGACGTCGCTCTG
CAGCCAAGGACCACCCAAA
CTCAAGCCAGCACCGTCGA
CAACCAACATCCTCAAGTCG
GCGGTTGCTGGAACAACTC
ATAACATCTTCAANATAAAT
TATCACCCTGTGCAGCAGG
AGGCCGTGCTTTTAAAACTA
CTCTGTAGTGGCTCATGATA
ATATTTCGCTCCTTTTTTGCC
CCGTGTAAACTTAGTNGAT
GCGAATAAAATCAGTTGAA
TCA (SEQ ID NO: 1372)

				GAGAGGA (SEQ ID NO: 1249)		
R4	Dong	.	Bombyx mori	GCTAGCTCCCTAAA ATCCTACCTTACGTC CGAGGCGAACATCT GTCCACGTGGGGA GCGGAAACGCGTA CTATCGAAACTTAC GCGGCTAACAAGGT AAAGGTAACCCATT AATATGGAGACAA GACTAAAAAGAAA ATTGAGAGGGCCG CTTCCCGGGGGCGA TCGCCGGGGCACAC CTGGAGCTGGCGCT GGGTGTTCCAGCAT AGGATCCGTCAGCG GCGGAGAGTTGAG CAGGCGGGCTCTCG CCGGGGATTTACAA CCCGAAAATGCTAC AGCCGCCAACAGAA ATCACGATGTAGAA AGTAGGAGCAATA GCCCATGTGAACCT TACAGCCCGAGTAC CGGTTCATACAACC CCTCGGTACAATCA TCACCATCATCCTCG GGTCATAGAGGCTC ACCAACGTCGACTA TGG (SEQ ID NO: 1250)	TAAAAACTAGCATAATTATT AACTCATAACTAATGTATAT TACTTGGCCAAAAGCCCGT ATATACAGTTCCACCGGCTC TGTCGACAGACTGAACTGA GAAAGGGGAAACATATGG AAATAATAATAA (SEQ ID NO: 1373)	MLRRGRIFLPASTKAGKTRGRMKWSREVNLFIMRTYYVVTKLETDLTIYR KKLHEHFSLKYPNVIISQQRISDQKRAIERNKLLSQETLDRLKEEVRKQLED EQTNNVENEKLNSETYSHEYTTLTPQTILTKKTQQHTNIISSTQTSHSSTQT ESITLLLENEVDILNTNPTEGATQTQEVKDKFETNLTMYSGMDPKARPPL PKLKYSSKLNELIRLFNNDILVDYISPDTQLSDVHTLTYCTAVTISEQLKYKII AIEGNARHKKNFKPPWQQRLEKDIAKLRADIGKLTQYINNNRSKKVVQS VEQIFKNTKIHTSHENGNKKSQEFLDTLKQKLALKAHRLKRYNNSQKRKN ENTIFLTNEKLFYRNLIKPKTDRDNSNIDIPTAEQLEMYWARLWENSAKH NDKANWITEEKERWDTIEEMQFDDVTEEEITTITARLHNWKSPGIDKIH NFWFKKLICLHKTIAKNLTDIISGNQSIPEFIATGITYMIPKGDFSIEASQYR PITCLPTIYKILTTVITKKINSHIEHNNILAEEQKGCRRGHMGCKEQLIIDSTI MKHATTKNRNLHCTYIDYKKAFDSIPHSWLIQVLEIYKINPIIISFLRNIMTH WQTTLKLKNPPNFVTTRQIAIKKGIYQGDSLSPLWFCLALNPLSHQLHND RAGYRIKQQDNTETIISHLIYMDDIKLYAKNDKEMKKLIDTTTIFSNDISM QFGLDKCKTVHIIKGKVQPGDYTIDDTNTITAMEPSDLYKYLGFQQLKGL DHITIKQSLTSEYKKRINAICKTKLSGKHLIKALNTYAIPILTYSFGIIKWSKTD IEQIERITRTTLTKHNNLHPKSAIERLTIKRQDGGRGMIDIWHLWRKQIHS LKTFFYIKSDLSEIHRAIAQNDNNYTPLNLKQKELIDNTENLRNRNPQKD MEENWKKKALHGRHPHDLSQSHIDSKASNMWLKTGSLFPETEGFLIAIQ DQVINTKNYRKYIIKDPTIRDDKCRKCNTQPETIQHITGACSTLTQTDYTH RHNQLANIIHQQLALKHKLIQNTNTPYYNYKPQTVLENDSCKLYYDRAILT DRTIHYNRPDITLQDKNNKVTYIIDIAVPNTHNIQKTFTEKMTKYTELKEEI VRIWKQKKAYIVPIISTTGVVPNHIHNSLKLLDLKDNIFISLQKAAILNTCRI VRKFMQLEENQTYYTQ (SEQ ID NO: 1495)
R4	DongAG	AB097127	Anopheles gambiae	GAAGGCTAACCACA ATA (SEQ ID NO: 1251)	TAACATCCGGTGCAAACTCA TTAACATTAAGAAAAGAGA GAGGAGAAATGAGAATGA GATTCATTCACCTTTGGCAT	METRSMRKRTTRLPEEGAPTGAGPGTGDRASIQRLEDEMVQERSFSQR ALPVPRTQNRNGSPINHQGNAASANVAVADRQQSLILAGGRRQRIMW TREMNHYVIRCYVYYTRMETDMPGRVKMLGMFNDREPRFAHQLDLSK LYIRQRAIILPEELEFIKLEVRREFGEEEAGWRESSRISARLNTIDQNTSRAS

					TTGAATAGCCCGGGGTAGG TGAAAAGTTCCCAGCATATT GCTGAGAAGTGACAAAATT CGGATAATAATAATAATAAT AATAATAATAATAATAATAA TAATAATAATAATAATATGC ATAATAATA (SEQ ID NO: 1374)	EDRDLDEPTAPGLSVDIQHQMATAVTQFHGTDPLSRHRLPKLHYSYRLK TAVSIINQDVLPQYLDSVGSIEDLQLIVYSAAVAVVRTLWLRTYPQGDSEG RPCSKAEKPAWMRRLENRINATRTKIGRMQEYQRGNSSMKVVRQIAE MVKPKELRDLTDANITEVLDIHLQRLSALAKRLRRYAECSKRKEQNRMFN INEREFYNWIRNDKPNFREGLPDIGDFTQFWANLCEKPVQHNSEGMRL AEDERFSDGIEDMPVLVVNAQDIREATQYTRNGAAPGPDFVYNFWYKK LITIHEQIAACFNTVLEDSRKLPKFITGGVTYFLPKDQNTKNPAKYRPLTCLS NLNKVLSSVITQKVKDHCDTNNVMTEEQTGRRKNTQGCKDQVIIDAVIV GQAAKKQRNLDMAYIDYKKAYDSVPHSYLLKVLQLYKVDGNVIKLMQH AMGMWSTSLHVTDGKVVLRSRSLNIRRGIFQGDTFSTLWFCLAMNPLS RTLNQQCNFGYLLKSEEISTRITHTFFMDDLKLFAETVQKMHHLLKNVQG FSNDIKMEFGIGKCRSIHLHRGQVLDADSFRANEQEEIRHMVQGETYKFL GFLQLRGIHYAVIKKELQDKFLHRVSCILKSFLSVGNKVKAINTFAVALLTYS FGVMKWSNTDLEALERTIRVVSTKHQMRHPKASVERVILPRKIGGVGIID IQALCISQIHQLRSYFVESQNRHELYRTVYKADHGLSALHLAQQDYQLNC NIKTVDGKGATWKQKELHGTHTHQLNLEHIDKVSSSTWLVRCDLFCETE GFMVAIQDRVIATWNYRRCILREDVEDRCRKCNSGGESIEHVIAGCPVLA GSAYLDRHNDVAKIVHQQLALRHKLVERFLPCYRYLPDPVQENDCIKLY WDREIITDLILRANRPDILVYEKRKKRATIDIDIAVTLDHNVQTTFSTKVMK YHDLAEELKQTWYLEDIRIVPVIISATGIVPMALLRSLDELELLQRELPRIQKA VILRTCSTLRRFLNPYN (SEQ ID NO: 1496)
R4	R4-1_BX	CADV010081075	Bursaphelenchus xylophilus	GGGATCCTGGGTTC CTACTACCTCGCTCC ACCTCCTCGCGATG GATCCTGGGGAAGT CTCCGGACTGAGCT AAGAGAGCGTTAA AGTAGAGGGTGAC GGCGTAAGTACCTC CAAGTTGCGGTGGA GCGGAACATCTACT CTTCGGAGAGAGG GGAAGCTCTATGGC GGCGTTAGAAAGG TTGGACTACGGCAA CGCCAGGGAGATG GGGAAGGTTCATCA GGTGATACTAGTTC	TAAGAAAAGCATGAAATAA TAAGAAATCAGATAAGAAT AACAAGAATACTAATAAGT ATATCATGTAACTATGACAA AAAGAACGCACCAATAAGA ACATGCTTGAGTGGCCAGC TCTGCAGGCAAAAGTCGAA TTTGGAACAGCCGGTAATG GAAGACCTGCAACAAACGT GGGGTAGCAGGCAATATGT AACTATGACAGACCAAAAC TCCGAAACTCTGGTAATGA GCCCGTGCCCCCCAAGCAT GTGGTCTCGTTCGATGTAGT TAGGAACAGTTCTCTTTAAC CCGTGATGATTACGCCCTGT CTTAAATGGCAGGTGCCAC	MTCNNAVVFPPADGNPAGTADRNFAIRFPSSEPPGPSGIRPSEPLDGRT GIGDVEHAQAGNGGFLVDVLEYKEAHRYGSKCEFCYVQTKGTVCSKPRT DAWLKCEILFLLHHAYTANQNKSIELAESAFRRAGITRRSKATIAKRWSLI QRGKGTDYKEYWDEYFEKFRYECNPTPIVRRKRNRLAAGLQSPSSVPNG YEFERKRTCETPLDTKASSLPLICNLLTGIVGVENVEENMSVECTEPKELSG TANSSVPGLAEGVYERRHNNVNEPAAGCPQDVPVANNLIDSPTTNDRL EAEFKAQLDRAERSYMRRRLPRLKNLSPDERMWIGTTVERLRLETVSEPV CEQWRLANAGLYAAIRSIAVMRPLDAAREAHKTWLLNMKMTERKLRQ QIGWVETTRRTKNEARTERQEIVYRKVAKLRRERFPEMDLDSVSVHLKRK LELLKGRIQVRTAERLRRDTREAAGPYGKTALRGQGFAPNVKDATQYWS GLAQPSGQKCSENSAILSDWKELVECNLSSLPDQMEPLVVQGISRASPW KSPGPDGIFNYYWRQDFIVDWLKQLMLDSLRTGHYPWKLSSGRTVLLYK DGDPTKAENYRPITCLNGCFKMINSVVSEVILKRVENTIALPIEQMALRRK VWACVESQIWDQIKQRKLSDRTQKCKVAWVDFSKAYDSLNHDAIKFVI GVLKLPTGINNYLLDSMQNWSTHLELKSSGKVVRGPSYPIKRGVLQGDSL SPTLFVVVTSIIVRHIKTIESSDIQMYMDDIKLYGKDQETLTRLIKELQTVSN KLGLCMNLKKCAILGDDLPEEINGIEHLKESYKYLGVPQREITQVRATMAA

			GCTACTGTCATTCG ATGTATCCGGAACC ATACTCGCCAAGTT GTGAACTATGTGAA AGTCTGGATCCAAT CCAAGACCACGGG GCGCAATTAAAAGG TGTGAAGCAGCTTG CTGGTGATCACCAC GGTGGTACCTACAC CGGCGGGGGAACA TCTTTAATGCCGAG ATGACCGCGCGAAT AATAGGGAGCCGG CAAAAGCCGAGCGT AAGTGGAAAGGAT ACAGAAATTTGCTC AAAAGTATACCAAC CCGCCACTTATAAA AGACCGAGAAAAG GGTACCACTAGAGC TCATTATAAAACAT CT (SEQ ID NO: 1252)	CAAATACCGAACACTCGTTG AGGTATGGTGGTCCGAATG TGAAGCTGGGAGTACAATT TGGTACGAGAGCACCAGCG CCCCCGATCTAAGTGATGA CGCATGCGTCGGAACAATG AAGACGGCTGGCAAACATT CAGGAGTCGCAAA (SEQ ID NO: 1375)	LEKKILTEVDTSLGAAELSYRQRISRVNSKIAPLVRFVVQSMLVTPRDVLKV YNRLGGIDVEIRRRLVKYEIRYKKSNVARLYLDRKVGGIGFVNLCRIMVEA VAARAVYCRLAPSFNEFQDFLAEQNTSPITAAQTILDKCGINIELSTSTLGD VKKIVRNHYHELWLTAWKNTGLYKRWENDHVDIKRSSLWINRGNLSAN NARIGIGIQDNSIFCRGFVGNKCDTKYCRLCGDGIESVSHIVTGCPTHRTN LYIERHDCVARNVYAYLAIRYGIPVPHYTQRVKTIEKNGDQSVELYWNYK FPCTRALEACRPDIVLIDKVSKRTHIIEVAVSWRGRLQEMVDRKVYKYTV NGEYEADGSSRGWNIVRELNDQYGFPVEVYTLVIGAGGEILPCTVKDVER LTGGAATDNLIERMERSAVLGSCRIIKRHLAL (SEQ ID NO: 1497)
R4	R4-1_HG	.	Heterodera glycines	TGGCGATACTCGGA ACCTCCGGGGAGCC TGGTAGGAGTTGG CCTACAGGTCGCGA AAGTCCCTAGGTGC TGCACGGGTTGCGC TAATCCGAGGGCGC TGGGTTACCTTCCC ATCGGCCAAAAACG TCTGGGCCTTCTTA GCCGCGGGTCCAGT ATTTCTGTTGAGCC TGACAGTTCTTCCC GGATATGGCGAAA	TGAGGACTCANAATTGACA ATACACCTCAGA (SEQ ID NO: 1376)	MISCDLERETLTQMALFRARSDKTPTHAGIPAPDEVREGGCGQNRTNPA APRGKAAAIQRQNGITIPIXACAQSGLVRTQRVQQWSAVEESALKDVVV RNTDDRGLINWAKGVLPEWQRLCQLNPTMYMARSSPSLSNKWASLRR THVGPGCPSKEGSGPSQDLSDVKIQPARLAHDTVAELPQRTVPCGTDGH GVIDSDETETALAEVSRSSPFGEREPLDLGATERITRKRLRNAVRDVVPPR KRRVPSTPSRKEQDLVPEVDGPAPTDVLTHPPTESEPEPMLDPLSLVQLV RPQLGRAMGWAAEEMELGNVVMDVELKREFNREVRRVGRTPPDQMIY KRGAGPPLPQKREPERVALLEQLIAARVERGINRGLDWFLELNVAVFAAA RVLSRRERVETLADRLHINDSATLSEVSRRRAKAERKLRCAREQPWMSRR IRXLGVRVERLKQLADLVRQRIAGRGNRSSYEGPRRRFRLRPSLRSVTEAP VNPPLNGNEVYTFWHSLWAQSLRANTDDCQLREFKNQLSAARHTDLTS VGTSSLVQMFSAALRKMKKGKAPGPDGIRAAWWGVFRRIAPYVATWV VRVIRGAEPVANWICNGLTVLLPKSSDNADPSNYRPITCLNTCYKLFTAVI AQITASYVDVLGGLPRQQVALRKGVWGTSVSLMIDALTVADARRAKRPL

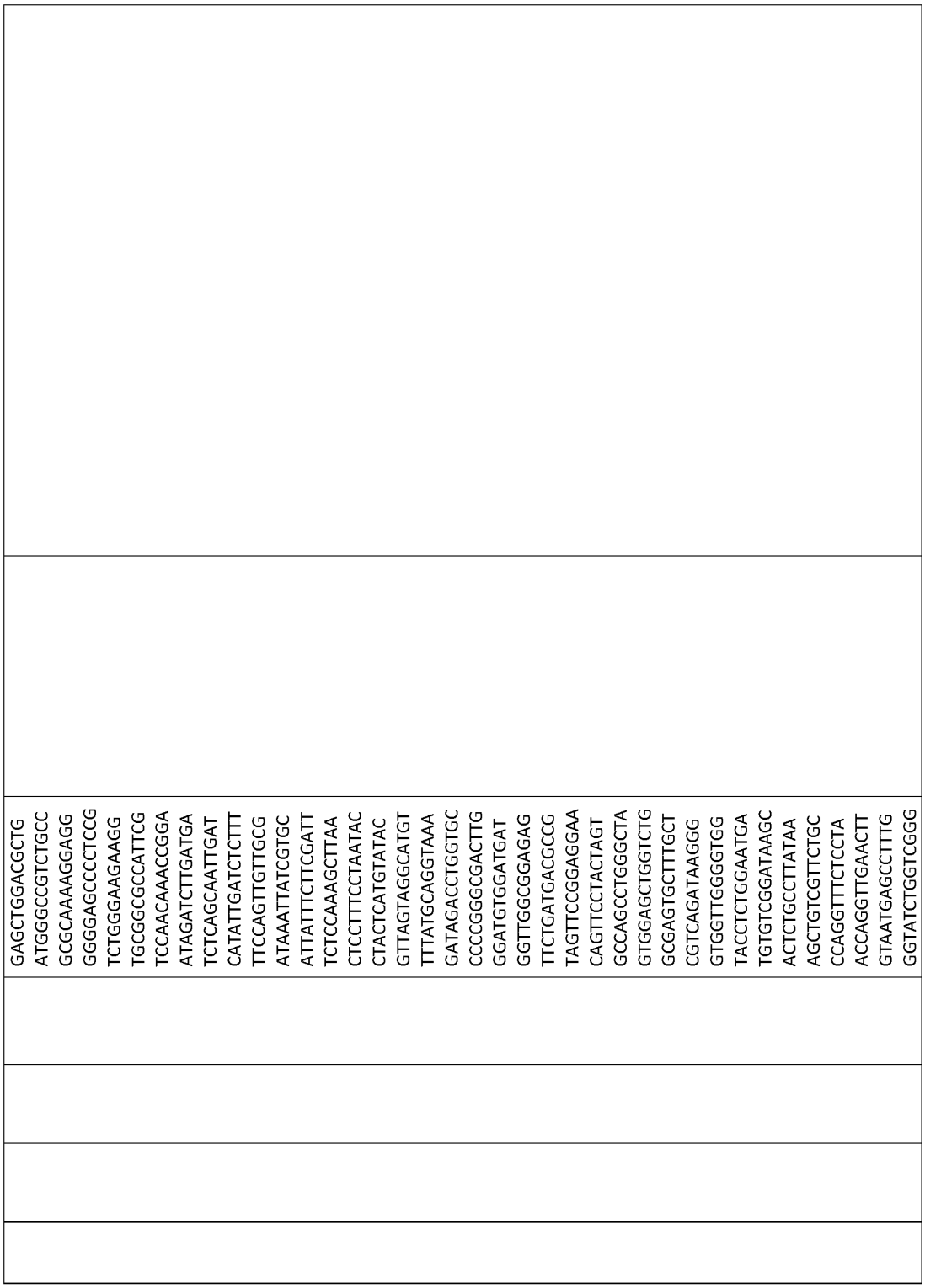
			GATTACAGGGCGGT		GVCWFDFKKAFDSVPHNLIRWILRVIGLPPVILSVIVSVMDQWATRLKIG
			ATTTCGTGAAACCT		GKVMPKTIPVRTGVFQGDTLSPLLFCLSVWPISFALDQFPQYQFRCANHL
			AAAAAWGGTCGGG		QQGFSVGHVFYMDDLKCYCPDREVLTAVIQQVQKSASALGLTIHYKKSA
			CCGAATGGCACGG		WLDQDGGKSGKAVLGVPXLVGTYKYLGMHERFMIVSKDSLESVRGKFM
			ACAGACTCACTTCG		GRLKTLWTSKLTFGQAMLGTKSXCMPVVRYVLQNLFLPKSEFNQTRLVL
			GAGTGAGCTCGGG		REWDRQIRDLLDECNIRQVFRSKTELYVSREEGGWGLPSMEDALEEEVV
			GGCATCCGTGTGTT		TKLAILVARQETEPLFRVCEALERKRCPTPLSLGLQILKDWGVGVELQGRT
			ACCCCGCTGCACCA		LLLNGNTVGPSQATRKLTGELVLRREAERLSRWRSKVKPGCGMTGGAW
			CGCCGAAGCTGTCA		RDVPGIDVHLSNRWLVKGALSPTVVSNSLAIRANTVILRGSGGGYTKGTL
			TAGCGAGCCCGAA		LRCRGCGNTGETRRHIVSACSLGRQKGAASRRHDNVCRILVRAICHKLNI
			GGGGAATGGCCAT		EPPNSANFPHVVVLEGSGAKMWIDFPFVVPHKIRHTRPDIVVLFEWNGV
			GGAGACTCCAGCCT		RRLSVIEVAVSDVANMQTQHIRKSHRYGTNSTEPFVAGVTPTYRNDCLA
			CACCCTGTAACTCG		AQLRAKFKAQQVDVIPIIVGTTGETLDGEFGRIRKGLPMLTKLQMPRLWS
			AACCTAAGTCCAGG		EIQRAVILGSYRILVEHLALPKGGA (SEQ ID NO: 1498)
			CCCCTTTCTGGTGTT		
			GGCTCGCACTGGTT		
			AGGAACACGACAG		
			TCTCGTGTAATCCC		
			ACACGCACCGAAAG		
			CCCTAGCCCTCGTA		
			GGCGAAGGCTGCG		
			TTGCTGGTTTCAGA		
			ACTGTGAGCWAGT		
			GGGGTTCGGATGG		
			CCGAGTGTACCAAC		
			CTGCTTGCTAACAG		
			GTAGCATAGAGTAA		
			TATGCTAGTAAGCA		
			GGGCACGAGAAAG		
			GGCCGTAAAGGCTC		
			CGGCTATGAAGGAC		
			CTTGCGACCACGCG		
			TGTGTCTCCCACGT		
			GCGGATTCTTGAAG		
			CCAGAGTCTTGCAC		
			TGCGCGCAGGATG		
			GAGCCTGTGCAACT		
			CCTCCCTCGCTGAT		

				CGCAGGAGTGGAG GATCACCACTCTTTT TACCTTGCTAGCTT GGGGTACCACCTTG AGCTGGGGCCGGC CTTGCTAGCTTGGG GTACGACCCTTTGA GCTGGGGCTGGCCT TACTAGCTTGGCGC GCCACCCTTGGAGC TGGGGTGGCGCAA GATCACTTGTATAC GGTCTAACCAATAC ATTTGAAAAGCGAT CAAAGCGAA (SEQ ID NO: 1253)	TGAGCCTTAGGTCGCGGGA TGTGACCCGGCGCCAGAGT GTAGAGCTGAACATCGCTC AACCGATCCAATTTGGGTC GTGAAATCCCCTCGATAATA ATAATAATAA (SEQ ID NO: 1377)	MKMSHNRDTPSNGVKGTSVRLGTSLVRSPVGEAGAVRERGTHPSESVS QDSDASVNATGEGSVREQAPLSPPGAEEATVPTQRRTRHKWSREDRVV LWECFVASKREGPGYLKRLKQLWDERGIPGNFPQASLSGQIRQICSKNLL SEEERLQIAARMEAQVASPSADEPARQVPTRPVTPPRSPPVEPARRPSIP SEETPDLGAVPSEIDSADPNRSPSRGPRHLPAHNMSQSESEDDVTDPDV SDQQRSDSLEPRDLLRNSSVESTPGHPNQELSDTLLSNYVPSEIDSDDPN QSPRRGPRHLPAHDMSLSDSMDEETEPDLSDQQRSDLLELRDLLRNSSV ETTPKGHPSLRHLPEPKIRAAAFRVNSVIGKIHTNNITETNALIKAGADLAV RILEVQPRPQRTQRKKDPPWKHRLEKNIAEIRKHLSWISEWRRGNLHDE EKKTLLESRYRCLEVGLTNLEDTLKQRLSAKRSKVRRFEARVAGFHQNQLF NTNQKRLYQTLRGEETSSDSPNAEESIRFWSDIWSKEVRHNNTAEWLHD VKEKNVAADPDLTITSQQLKKQLSKTKNWKAPGPDMVQGYWIKTFTSL HSRIAAQLNHCLQRGTVPTWMITTGKTVLIQKDKAKGTEVSNYRPITCLPL MWKVLTGIIYERVYQHLDSKKLLPDEQKGCRRNTRGTKDHLLVDKLLTKD ARSKKKNLSMAWVDYKKAFDMVPHSWILECLDIYGIAGNIRNLIATTMP NWKTQLTSANKHLGEVSIKRGIFQGDSLSPLLFVLTMIPLSETLNKAGQGY NYSRTMKLNHLLYMDDLKLYAKSKDQVEQLLNIVHQYSQDIKMQFGVS KCGVLNIERGEVTASEGITIEEGTIKDIEEAGYKYLGVMEYNTILHRTMKDS IRKEYLTRLRLILKSHLNGGNTIKAINTWAVPVVRYSAGIINWTKKDCTDM DIKTRKLMTIYRALHPRSCVDRLYINRREGGRGLISVEDCVEAEKRALSQH FRESDDPWARCLVEAKLLKETETADQFKERRRLDRTNKWKSMKMSGQY LEAVQDKIVPDSWNWLLRGELKRETEGTILAAQEQALRTRYIQNKIDKRN
R4	R4-1_PH	.	Parhyale hawaiensis	CCGACCGCCAGCGG GATAACTGGCAAAC CCTGTCTCGACCAC CGGCCCGTGAATCC ATCGGGGCGTATGA GTCTGACACAGGG GGGTGTTTAAGGTG ACCCTGTTGCGAGG AAATGCGCAGCAAA AGCCGGATGAGCCT TAGAACATCGAAGG CCAACGACAACCTG CCGAAAGACTGGC GACTAATCCAGTCA ACCTCCTGTAGGTC ACCGGCTGGTCATG CTGAATCTCAGCTT AACCAGGCGTACGA ACTTACAGTTGGAG GGTCGAGCACCCCT GATGGCTGAAAAG GACCATCAAAGTCG			

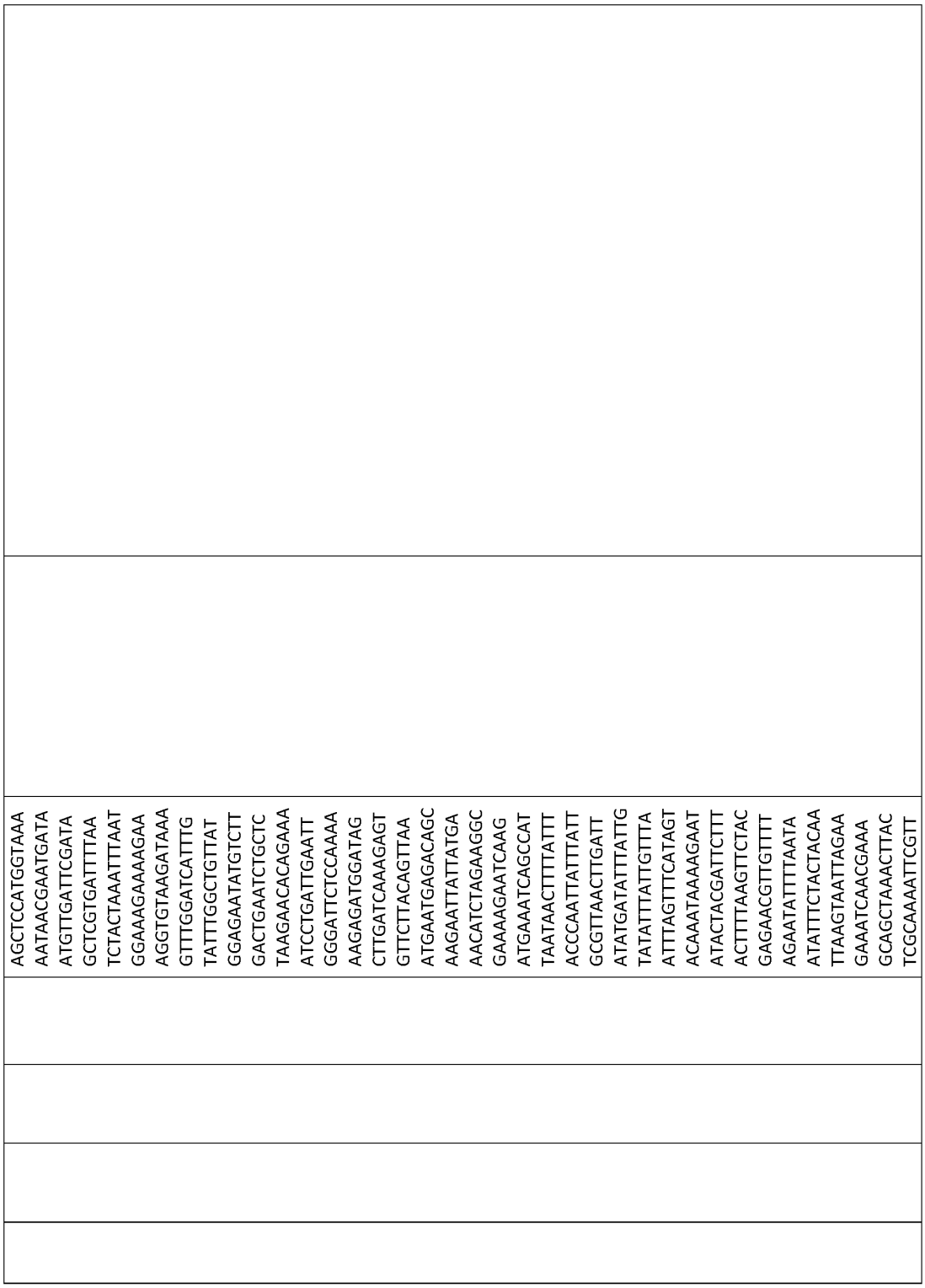
				AAGGTAGCCAACG AGAAGACCAACGG CTGATTCAGGCGGA AGAGTCAACTCGTT GAATGCGTTCGACA GCTTGGGGTAGAT GGAACTCCTAAGCC CTGAAAGGCAGTCC ATCTTCGCAGACGC TAAGGTGCCCCGCC GCCTGAGGGTTATC AGGCCCCGCCGCCT GAGGGTTACCAGG CAACC (SEQ ID NO: 1254)		VPSTCRICRSSDETINHVISECGVLAQKEYKRRHDKVARHLHWTLLRIHNF PVSERWYEHEPAPVVENEAVQIYWDKRMETDRVLHANRPDIVVKDKQE KSAKLIDISIPFDSRIVDKEAEKKEKYRDLAIELQRLWQMKVDVVPVVIGAL GAMSKNLKTALRELKCGHLHPGTLQKSALLGTAHIIRKVL (SEQ ID NO: 1499)
R4	R4-2_BX	CADV010090 48	Bursaphelenchus xylophilus	GAGGATCCTGGGTT CCTACTACCCTGCTC CATCTCCTCGCGAT GGATCCTTGGGGAA GTCTCCGGACTGAG CTAAGAGAGCGTTA AAGTAGAGGGGTGG CGGCGTAGTGACTT CCAAGTTGCGATGG GGCGGAACATCTAC TCTTCTGAGAGAGG GAAAGCCCTATGGC GGCGGTAGAAAGG TTGGGCTACGGCAA CACTTGCCATGATC AGATTCGATCAAAA TTAGCCTCTGGGGC TGGCAACCCTACAA CGGATTGTAAACTG AACTATGCTATGCA AAATGAAAATAAAA AATGGGGGCTTTAC AATCTAAGGTGTTG	TAACAAGTGTAATAAAAAC CACCCATGCGTAGTAAACC GATCAATTATCTAGCAAAAT CGCAGGTCAGAAGACCAAA GAACCGACCCAGAGGAATA GGACCAGAGCTGAAACTCT CAGATACGCCAACGGTCCT AATAAAACGTCGTTAAGTA AAGCATCGTTAAGTACAAA ACAAAAGCACTGTAAACGC GAGGCCCCCTCTTTGCCAAA CTCCGGTAATCCTCGTAGTA CGGTGCTTTTTCCCGCTTCA CGAATCAGAACGCTGCCAG ATCTTGTCCGACATGGGCC GTAGGGTTTGGGAGTACCA GCGTGGGGCGGAGAGCGT ACCTGGGTACACTGCATAA TCGGGTCTCAGAAAACCAC CTATGGTTTATTATTCTGTCT CCCATCTGCAGGGTAGCTTT TCGTTGGGCCATAGGAGCC TAGGGGCAAGAGTGCATGT	MVFNNCKPKHLCPAIRPTGQQETNGGSEGTAEPTAGPSRPAVSEDAAQ PVPLFEEGEYIRAHRDKTCPYCEVLWIGARSSKARSDSWPLCQILYLMKR NDDLRGQRTRYPLLESSLRAAGIARTKFAIIKCIRNVLRDRYVPNGPYSEH WKIYRANSGEVPQGATITKGKRSARVAGLPSPSQSGHHTKRIQAGTGIET ETTVTETNTTPEVSHEHRDPCGEPETSAANVDKVTELTEDGSETRGTANV ANGGVSVSDPGRKRQSSSQNRGNIETTNPELVGMWEDMFGVQLDGA MRTTERPRLPKLKHLSEPERLWIRAKLEQAWLQCVSYDVEQQQWLNANA VLYAAIRSVAASRPCKEAREAQKTWLDNKKKDEAKLRRLIGRISSVHSMP KGDRTPREKKLVKNITKLKNTHYPDMDWGGLLNHFKVKLSQLKEKISVRV AEHKRKVNRNAAGQYGKSVAGSAGLAPDVVSATAYWSGLAQPGPKKF KASSPIFQTWKDDVAKNLNTEPVLLYPIIKECIRKPSPWKAPGPDGIYNYY WQQEFVAQWIQTLVKRTLDIGRFPTALMCGRTVLLFKSGDKSMPQNYR PITCLNGCFKITNAVLTKVILQRVQDTCALPREQMALKPKVWSCMEAQL RDQALQSEIGDDCKTAWIDFSKAYDSLDHDALRFVIQTIALPAGMEEYLL KSLDSWRTQLVLSDAGKVVSGKPYPIKRGVLQGDSLSPALFVLTTSPIVAH LQRTCPTGRIQLYMDDIKLYGKTESDLCMLIKETQRVANKLGLNINLKKCA LFGKSIKQSIAGFDPLGDRTYKYLGIPQRDVADIKQAYDELKAKTVQTIGET MACDYLTTRQVINRLNSKIPPVVRFVTQSALCSAPMTRGLYNKITELDNV SRAELRKVLIYKATNVSRFYLATKEGGFGYASLQQVFVEAVVSRAIYCLRA PSLCDIREFILSKFDPVKVARIALARSKIDMDIERMDMASATRTIRQHYQA KWKTLFQQSKLYQKWVQHKIDIPNSSRWLQRGEISPRNCRIAVAVQDN TLLCRGFVGSKDPNKQCRLCNAGIETASHIVTECSTHRVHMYIERHDSVA RNIYAVLAKNCGFWIPHYSQKIPTVKITKSYELYWNYKFPCTQALEACRPD

			GCAGATCACGAAAA CTGCCCGTCGATGA GGGTGAGATAACAT CCTACGGAACAGCC CCTGCTGGACCAAA CCAAATCATCCACA AATTGGAGGGTTTT CTTGGTAGTTTCCCT TGGCACGTCTTGTT TCATAAGCCAGAAT AAAAACGATACCAT ACAGACATGAGGCT GGGTACCTGCCAGC CGCGACACGGAAA ACCGGTAGGTGCAA CCGGAGACAGCTA GGGGAAAGAAAAT AGTAAAGTGTCGAA AACAAAACAGGTA GCCCCTGGCTAGAG GGAATGGGACATT GTCCGATTAGGTTG CCTTGACCCAATGA AAGCCAACCGGGTT TACATTATCGGTCT ACCAAGGGCAATG ACCAGAGAAACTGC GACTATGCCGTAAC CCTTCGCAGATTGC CGATGAGAACCATC GATCGTAAGTCGAA GCCAAGCGGATTGA CCAGTGGAGGGTTA TCCCGACAACAGC (SEQ ID NO: 1255)	AGTCTTCAACGGCCATGCC AGGGAAACTTGTGAGAGGT GAGGGATAACTAGCATCAG ATAATATCAGTCATGAAATT AGTAACAACCAACGTTCACC GTCGTTGGCAAAACACCGA CTAACGATGCTAGTTAGAA AGAGTCGGGTCTTCCCAAA GTTAGGTGCTTGCACCGAA GCCGATCCGCTCTACCCACA GCTCTGCCCAGCGTT (SEQ ID NO: 1378)	IVLIDRAKKRILVVEVAVSYVTRLEQMTQRKLYKYGVNGEYQADGETRG WNICRELVQKYNMRIDLCIVVIGACGEILPCMVKEIEKISKVSGRQLLERC QRSAVLGTVRTVRRHLAN (SEQ ID NO: 1500)
R4	R4-2_HG	ABLA010003389	Heterodera glycines	TGTGGCGATACTCG GAACCTCCGGGGA GCCTGGTAGGAGTT	TAGTCGTAGCCCAGATGTC GACAGCCCCTACACCGGTA GATGGAAGGAGGTATGTAT	MRKSFLQHIPELSSHIAMSVPARNYPKMCSLAQGSSGTLSHNGKGVAM HRCPSDDCAGKDPPQRGSQKGNLRSVRWTPSEEKAVFEYWSRLEQHA MLNGSEARGTCAITRSQFLIHWDGERESRSLSDGVPEYPMRTERAYYER

GGCCTACAGGTCGC	CTAAGCCCACGGCAAGCCA	VRLLRQRGWQWDCANECLVIGQCAEPCRKPNVVAIKADKGMKRSLVK	
GAAAGTCCCTAGGT	CCAGTGGAAACGGTGACCA	GKLLSLPHVMGEINQVSVQVEVPLPSVPASVPQVEGVESKGFTETEPSNK	
GCTGCACGGGTTGC	CCTCTGTTCGCGGAAATGCC	PSLEGNPAEEGLRKPERVNVPVHGIISDSERKDLKDRFWSAYKTAKRSVG	
GCTAATCCGAGCCC	CCGTAGAGTATATGAAGGT	FRPALKIEPNRVNRAQWEVLDSCVVEVLKKRETSNGYRGCVLRHLNVAV	
CTTCCGGTGTTGGC	CGGAGCATTGAAGCACAAG	YAAGYVLAEGNKERRQVIRRQSAEWLLRQKSEINNIRRHIGWITDELTRR	
TAGCACTGGAGGG	TGAGAACCCTGGAAGTATG	RTGKNPTSRQLSNFAWLQRRYQVLGKPVRETRDLEVQRERLVSRLRLAQ	
GGGGGGGTCCAAA	GTGGTGAGCTGGTAACCTC	DRINSSMDREERVRKRMILPLRRKLEEPLGDSKLDTKQARTFWASLIGERK	
GAATACGTGAAGC	GAATTCCTCCTATGGGTGCT	EFGKIPELENWAAEEVRSKVTDGQGFASDHVDQTVWKKILGKARPLKAPG	
GGACCCTGTGTTTT	TGCGCCCGTAGGTCATTTGT	PDGIPNLLWKRLPSANQALFKWLMGIKRKQLSVPSWLTKGRVVLLPKGG	
CTCTGACCTCATAC	GTATGTAGGGATGGAAGG	DPVDPANYRPIACLNTQYKLVTGMVTAWVSEHLTTYSILPIEQRAMVSG	
GGGTATATCCCGAT	AATCTCGATAGCGCAAATC	TWGCTHAMVIDRAITSYAEATGLPLYVGFVDFAKAFDSVSQPWIRYALK	
GATATCTATACAGT	GGGATACCGCACTGGCGAT	VAGVHKRIRCLIGILMKCWSVRYEVFKSGRVLRSAPLAVKNGVLQGDTLS	
GTTCATTTTCATTTC	CATGCTGCCAGTGGGCCCT	PLLFCLSVAVVSSAVGSLFDFEVTIPGRGVMQQQNHLFYMDDFKGFAPS	
TTAACCTGCTTTTTC	GCTGAGTTGGGCGGGTTAG	EASLTRMLVTLERTASALGLKINKRKCALVHPRERENEETGSDIPVLGLRD	
CTCAAGAGAAATAT	CAACCCGTCCAGGGACGCA	TYKYLGIEERFGIVFEDAWDRVRTKMFERMRTLLCTEHTFGELRAAFASTI	
GACCACATTCGTTT	GATGCCAAAAGCCTGGTGA	APVARYLFLNVIVGGPSWSETLTKAKDMDLRIRRLLWERRDNEPGWRFK	
GCCATTTCGAAGGT	CACTCTAAAGGGAGTCTAC	HCSADRLYLRVQYGGLGFVSVEDTLSESIIYCWAYVQCRPELELARELFGT	
GCGACTTTTGAGGG	GGCGACGAGAACATCCCCT	LNRSARSGIKQSIAKGARKVFRSYALLSKNSAQRVSDLDGDASPGFRVGE	
AGTTCTCAGGGACA	AAGTC (SEQ ID NO: 1379)	MIFMEPTRGARAIVKILRKENDSRRLAAWKGRPMGGRVVSLPELDQVH	
TCAAGCTGTTCATG		SYHWLIRARIGRRSFRDCIAAQEGQLKARELMCPHINAKAKWCRRCGDG	
GACGGCTTGACGGC		RVETEQHILSGCAWSRTGTMLDRHNGVVRQVHTALCRKYGLPVSSHVV	
TCGTGGCGGCACGC		PLHAVIENEHAKILYDVALHTSPAGVLPREDGSTSYTGLRSTRPDMVIFDK	
AGCCGCGCGGGCC		KARTILVEISVPWRENLVKQELIKWRKYAINSMIEPLELAEAEIPGPNLKH	
AGTTAGGGCTGGT		ALGLAYGTSFPTVKVVPIVVGSCGEVLPNITKRLSELGIPKRGIPSLLESIQR	
GCCAATGAGGCAG		AAIIGSGHVIRAHLSVPRSESET (SEQ ID NO: 1501)	
CGAAGAAGTCTCTG			
AAACGCCAAAAGCA			
GCGGGAGAAGAGA			
GAGAAAGCGGGGA			
CGGGTTAGAGATAA			
TAGCATGATTGACG			
GCGAAGCAATACTT			
TGACTTCCAGGTAC			
GGAGGACGGCGGC			
AAAAAGCAAAAAG			
GGGACGAAACCGA			
TGGCATTGCCAAAG			
GGGCCGAAACCATT			


GAGCTGGACGCTG
ATGGGCCGTCTGCC
GCGCAAAAGGAGG
GGGGAGCCCCTCCG
TCTGGGAAGAAGG
TGCGGCGCCATTCG
TCCAACAAACCGGA
ATAGATCTTGATGA
TCTCAGCAATTGAT
CATATTGATCTCTTT
TTCCAGTTGTTGCG
ATAAATTATCGTGC
ATTATTTCTTCGATT
TCTCCAAAGCTTAA
CTCCTTTCCTAATAC
CTACTCATGTATAC
GTTAGTAGGCATGT
TTTATGCAGGTAAA
GATAGACCTGGTGC
CCCCGGGCGACTTG
GGATGTGGATGAT
GGTTGGCGGAGAG
TTCTGATGACGCCG
TAGTTCCGGAGGAA
CAGTTCCTACTAGT
GCCAGCCTGGGCTA
GTGGAGCTGGTCTG
GCGAGTGCTTTGCT
CGTCAGATAAGGG
GTGGTTGGGGTGG
TACCTCTGGAATGA
TGTGTCGGATAAGC
ACTCTGCCTTATAA
AGCTGTCGTTCTGC
CCAGGTTTCTCCTA
ACCAGGTTGAACTT
GTAATGAGCCTTTG
GGTATCTGGTCGGG

				GTCTGCGCGGAGAT AGCTCGTGAGGAG CTGTTATTACATTAT TGGTTGAACGCTTT TGGGTCGTACTGCA CAGATAAAT (SEQ ID NO: 1256)		
R4	R4-2_SRa	CACX010020 06	Strongyloides ratti	CGTCAGAAGAGCA GGTGTTTTTCAAAG CAAAGACTTATTCT ACGAAGGGGAAAG ATGATCAAACATGC AGATTTGGCTGCAA TGAAATAGAATCAA ACTACCATGTGGTG ACATCTTGCAAATA CCATTCATACACAA GAAGACACGATATG GTTGTGTACAATAT ATTAAAAAGACTAA AAGAAAAGTATAA GTTAGATGGTGATT TACAATATGGTAGA TCAGTATTTAATGG GAAGAAGGAAAAA ATTAGGATCAGATC TGGCCATAAGTTTT TGACTGAACAACCG TTAGTAAACAATAA ACCGGATATTGTTA TAACGATGAAAGAA GGGAAAAAAAAGA TAACGTACATACTA GAGATGTCGATACC ACATATCTAGAACT TGAAGATGCAAGA AAGAATAAAGTATA AAAAATATTGTGTG	TAATTAGATTTATATGTGCC ACTTTACCTCAAAGATATAT TAATACTAACTTAATTATAA TTTATTATGAATTATAATAT AATAAGTTTTAAAATAAAA (SEQ ID NO: 1380)	MQKFSVPKDSSQIFLVDSILNKHICSTKNKVKDVIKRRSIIKTLICAAGLTLR KLVCGKLGNNKYNSKINQLWKKERKIINCIEDLKHLIETNKRRHNFGKRLR NAKVSPSEMLKDYYNKLRYIKNEITACLDEHKKAILRTKFKLTPSIKIISNIQN HNDEEAELPKEEEFVKYYKELFTNKDGDDKETPHLDNWLKKFSKTLIVDW TINDKEILEALKYCGNFKAPGSDMVMKVCYKWFKSAQNYLIRWIKSTWY GEYTINKKDTNAVTFMIWKRDGKPKNDVKSYRPISCLNCDFKLLNKLIAN KIYESIEKILPINQMAVIKNKHGTCEALLYKSLVQSMKFRRTKDVKEIWCS WIDFSKCYDSISHKCLKKMIQSIKAPPIIHKLILDGIDSWNISICNGKNISKTK IPVKSGILQGKVASSLYFVLLTGEISYALNKEEQVPIETITPSNTLKINHISFID DYQLYATSQKKVEKLTIKLREIAEEMNLKLNPQKCGIYGTDDLGKRLMLKE SSLNFPYTSEYKYLGLVENSLDLKDINIQLFKDKILSKYSTIFESRLTTHQKRK VFNSTISPCAAYYLGNLITNKCSIQELLNECKKFDQMVRNQLVNQNIKKL QVSNSRIYLPKEYNSLGLNEIEIEVAANIIRKACYIKKRETLRGVDKLYIAMS KNGHRNTLSDALYITKKYSNFQINWNIMGMVKDQNNILLDAKKIIENIKE KRRNLWLEHWKKGNMTYANEAIKKEFHLPDLNIDSKYLMLCYAGSEEQII YNGHVSLVNQSSPSSRLCRKCNKLEETSYHVASVCEFHKKNLHLMRHNS AVYHIITELCRIMKVKCTLRYPEASGIIKSGNMKIAAGVKYTFGTAKIYHNK PDLVWYTPEVIYVIEVSISSLKNAKSQMKMKTARYAVNSTKKLENFAALN NLKKGENFVEILSHKANFKRVHFMPLVFCTFGEIPKETMKYLEKLNFSNEK IKTIASPIARYTGRTLKAHFTN (SEQ ID NO: 1502)

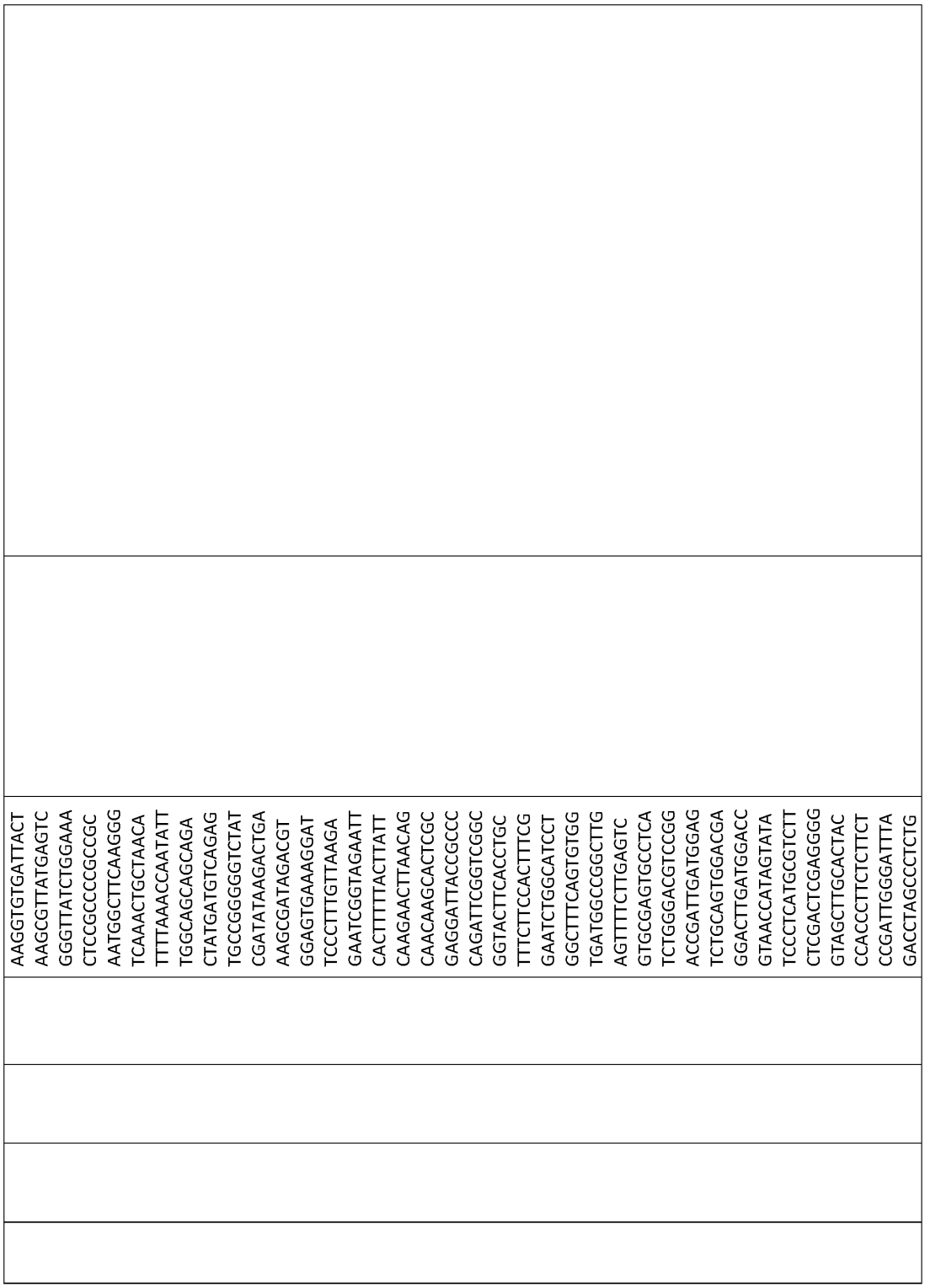

AGCTCCATGGTAAA
AATAACGAATGATA
ATGTTGATTCGATA
GCTCGTGATTTTAA
TCTACTAAATTTAAT
GGAAAGAAAAGAA
AGGTGTAAGATAAA
GTTTGGATCATTTG
TATTTGGCTGTTAT
GGAGAATATGTCTT
GACTGAATCTGCTC
TAAGAACACAGAAA
ATCCTGATTGAATT
GGGATTCTCCAAAA
AAGAGATGGATAG
CTTGATCAAAGAGT
GTTCTTACAGTTAA
ATGAATGAGACAGC
AAGAATTATTATGA
AACATCTAGAAGGC
GAAAAGAATCAAG
ATGAAATCAGCCAT
TAATAACTTTTATTT
ACCCAATTATTTATT
GCGTTAACTTGATT
ATATGATATTTATTG
TATATTTATTGTTTA
ATTTAGTTTCATAGT
ACAAATAAAAGAAT
ATACTACGATTCTTT
ACTTTTAAGTTCTAC
GAGAACGTTGTTTT
AGAATATTTTAATA
ATATTTCTACTACAA
TTAAGTAATTAGAA
GAAATCAACGAAA
GCAGCTAAACTTAC
TCGCAAAATTCGTT

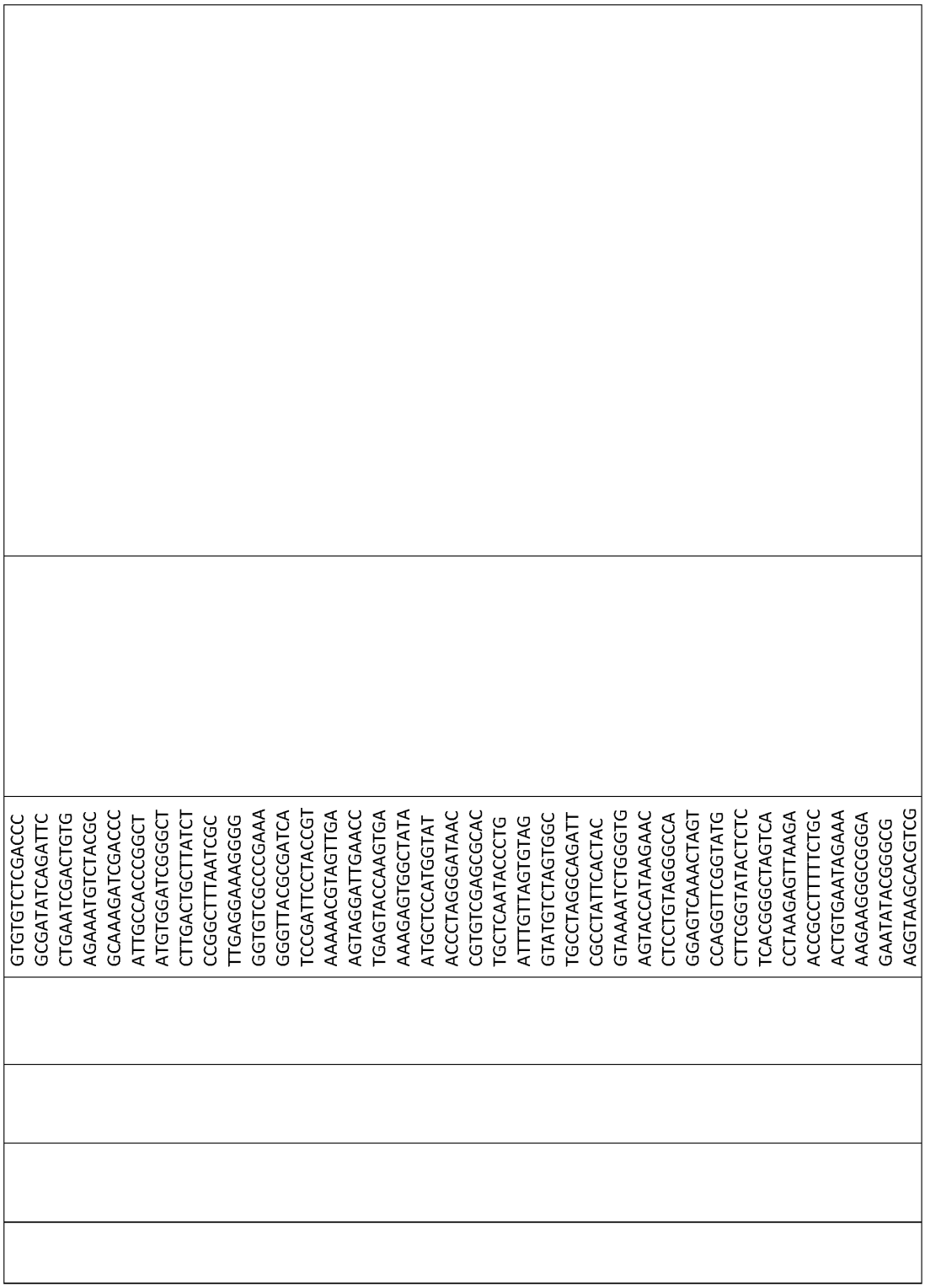
R4	R4-3_BX	CADV010088 32	Bursaphelenchus	GGGTCCTCGGTTCT TACTACCGTGCTCC ACCTCCTCGCGATG GACCCTGGGGTAG	TGAGACCACCCATGCGCAG AGTATCCGAATCAGTGAAA GTCCAAGTTTCAGGACAGA AACGTCAGATAAGTCCAAG	MISRSQADRPVEGQPVTAMSFHNLEPNNLYPENLRPTGSQDANRGVA DIAEEVTGPSGLVTNEEAARAPPLFVEGEYKRAHCGGGKCHYCRVLWIG ARSSKARTDSWNLCEILFLINKCMELGNVRRIYSPLESSLKEAGINRTRHAI VKCRLAVMRDRFVDNAPYSEHWRLYNACAENRAVVVPMDSATTVKKR	GATCGAGGCTGGA ATGGCACCAACTAA TATACTAACCAACA ATAGAAAAAAAAA GAAGAACTTCAGTG GACTTCAAATAAAA TAAAATTATTGGTA GAATTGTATGATAA AACTGAGAAAATTT GACTAAAACAAAAA GATTAGAGCAAATA TGTCATCATTTTCCT AATCATACAATTAA AGCGATGATGACG AAATTGAGAGAATT TGAAAGAGAAAAA AGGGAAAATTGTG AAATTAAAATGAGG GAAAATGAGGAAA AACCTGAGGAAAA ATTAAATTTTGACA ACTATGAAGAAGCA AAATTGAAAAGAG GAATTAAGTGTAAA AAAGAAGTAAAACC AATTGTTATTGAAA ACAAGGATTAGGA ATTTTTGAAAACGG AAAAATTATTTCTCA AGTTTGTCAACACA ATCAATTACCAAAG AAAGAAGGGTAAA TCAACCAGA (SEQ ID NO: 1257)

		xylophilus	GCCTCCGGGCTGAG	AGAAACGAGAGAAAACAAGTT
			CTAAGCAGAGCATT	CAAGTATGCAAGAGTTAAT
			AAAGTAAAGAGTG	CAATAAGAGAGTACCGTAA
			ACGGCGCAGTTGCT	ATGTATGACCCCCCCCTTTG
			TCCAAGTTGCGGTG	CCAAGTCGACAACTGTCAT
			GGGCGGAACATCTA	GCAGGTGTCTCTCTTTTCAC
			CTCTTCTGAGAGAG	CCGCCATATGGACCAAACG
			GGGAAGCCCTATG	CTATCCAGCCTCGCTCAGAA
			GCGGCGTTAGAAA	GAGCCTTAGGGCTGGGGA
			GGTTGGACTGCGG	GTACCACATGTGGCGGAAA
			CAACACTAGCCATG	CTGAATCTGGATGCGATGC
			ATCAGATTCGATCA	ATACCGGGAGCGCAGCGAA
			AAATAGCCTCTGGG	ATCACTTAACGCTGGTGCAC
			GCTGGCGACCCTAT	CTCCTGCTATCGTAGTACTT
			AACGGATTGTAAAC	CCTAGATAGATGAGTAGGG
			TGAACTATGCTAAC	TGGGCTAAAGGTAGTCGTC
			CTGTCAGTAAAGAC	TTCAATGGCAATACCGAGG
			AGAACGGGGGCTTT	GGTCTCGTGACAGGTGAGG
			GCAATCTAAGGTGT	GATAACTAGTATCAGATAAT
			TGGCAGACCACTAA	ATCAGTCATGAAATTAGCA
			AACTGCCCTTTGAT	ACAACCAACGTCACCGTCGT
			GAGGGTGAGATAA	TGGCAAAACACCGACTAAC
			CATCCTACGGAACA	GATGCTAGTTAGAAAGAGT
			GCCCCTGCTGATCC	CGGGTCTTCCCAAAGTTAG
			AAACAAAATCATCC	GTGCTTGCACCGAAGCCGA
			ACAAATCGGAGGG	TCCGCTCTACCCACAGCTCT
			TTTTCTTGGTAGTTT	GCCCAGCGTT (SEQ ID NO:
			CCCTTGGCACGTCT	1381)
			TGTTTCAAAAGCCA	
			GAATATAAAACAAT	
			ACCGAAGTACAGAA	
			TGGCTGGAACCTAA	
			CAGCCATGGCACAG	
			AAAACCGGTGGGT	
			GCACACCGGAGAC	
			AGCAGGGGAAAGA	
			TTGCAAAGTGCCGT	
			AAATAAAGATGGTA	

TARQAGLESPSQIGVAGKRVHEAETGTDRINAVIETNTTPLEDIDLSPETP
EGLAELPSTVEIMELTEDGSRSRGTANDADGGVSISDPLRNRPSSSQESR
NVPEQVDPDGELVWESLYGAQLRGAMRTTDRPRLPKLTKFSAAEQLWI
KSKVEKARLECVSYGIEQQWLRASAVLYATIKTVAACRPYNKAREAHKV
WLENKRAEEKRVRRIIGRIETVRTMPKGKRTDKQIRLARKINRLKRVSFPE
MDWHGFLNHFKAKLDLLKKLISVRVAEHERKISRKIAGTYGKSVSGQSGF
TPDVVAATTFWSGLAQPGPKKFKKNSLIFQTWKDSVVENMNTEPVLLH
PLIIECMNKPSPFKATGPDGIFNSYWRQGFIANWVKSLIQRTIQTGEFPAS
LMCGRTVLLYKNGDTAKPENYRPITCLNGCFKMTNAVITKVIVQRVQDT
CALPGEQMALKPKVWACMEAQLRDQALQSEIGNDCKTAWIDFSKAYD
SLDHDAIRFVIETLALPDGMEKYLLKSLESWKTKLVLSNRGKVATGRPYKI
KRGVLQGDSLSPALFVIATSPIVSHLKRVCPSGRIQLYMDDIKLYGKSETEL
RMLIKEVQKVANKLGLQMNLKKCSTYGAGLTESIAGFDPLGDRAYKYLG
VPQRSVADTNLAFGELEGKVIRSIEETMACEYLTMRQVVTRLNSVIGPLV
RFVAQSVLTSQAKVSWIYNKISDLDSKIRAKLAQTGLRYKKSNVARLYLSK
SKNGIGLVNVQQVLVEALVSRAIYCLRAPSLVEIREHILTAEFDPVGAART
VLRRSRIQLEIERVEMASAISAIKTNYQARWMTKFTQSKLYQKWVHHDI
DLANSNLWLERGEISPQNARIAVAAQDNTLLCRGFVGNRESEKQCRMC
NMGIETCSHILTECSYHRAHMYIERHDSVARNIYAVLAKDHGLWIPHYSQ
PVSSVTKTPTCELYWNYKFPCTRALEACRPDIVLIDRAKRTILIVEVAVSYV
TRLKQMVSRKVYKYGVNGEKGADGESRGWNMIRELSEVYNMKVNLCA
VVIGASGEVLPCTVKAIQSISSKTSSRQLLERCQRSAVLGSTRVVKRHLAEF
H (SEQ ID NO: 1503)

					GCTCTCTGACCTGA GGGAATGGGACAC TGTCCGATTAATGC CTTATCCCGAAGAC GGTCAAGGTTCTAT CTTATCGGTCTCCA AAGGGCTCTGGCCA TAGAAGCTGCGAG GAAGCCGTAACCCT ACGCAGATTGCCGC CGTGC (SEQ ID NO: 1258)	
R4	R4-4_BX	.	Bursaphelenchus xylophilus	TGCCAGCGGTGTTG ATTAGGTCCAAGTT CTTTGGCCAAAGAT CCGCCCTCGGTTTA GCAGTACCGAACGA GTATACCTTCAAGT GGTGGCACTGAATT AGACTGAATACTCT GAACTGTAGACTTT TGTGCAACTGTGTA TGGTGTGGAAGACT TGTTTGTACCACTAT CAGCTTTATTGGGG CTCGTTACTGTTTCA TACAGGTAGATGTC CCCTTTAGAGATTTC CCTGCAGTTTGCGC TCCAAGTCGCTAGC CTCTTGCGTGTAGT CAAAGGAATACATT CGCCGTCGGTGACA GGGCTATACCCGGC GACTACGGACTTGT TTATTACGTAGTGC AGCCTCGTTTAAGA CGAATGTGAAAAG	TAATTAGGCAGTGCTCCTA GCGGAGTGCCGTGAAGTG GTGTCAGTACCCGTCGTGT GGAAAGCCCAGGAGGGTTA GTACCGACAGTGGGAAACC CGCTGCACGCAACCTAAAG ACATTTGCCCTTCGGGGGA GAGGTATGAGACCACACTA GTCATGGTTGCTTGCGCAA GCATGACTCATTCATTGTAA GTTCGTATTATGAGGCCTCG CGGAATGCGAAAGTCCAAG AGCGCCAACTTGTCTCTTTG GGAGGTTCCTCCGGGTTCC TTAGGGTTGGCGTACCATG CTTCGTGTGAGACCTAGGC CGCTTGGACCTAGAGTTAA GCTGTTGCATGTTTGAGGA TGCCTTAGGCGGACGCCAT G (SEQ ID NO: 1382)	MEILWEDLRLKIEDRYGVTLPQRSASSLKNQYPKVILRGLPDSGLPWAGV QVNDTGQVVVVDHAEAATLRGSSPAAVDGEAEEPVVPPLPAAEVVESA ADAAVPDPQSEIVADQGVETRPVENPPANSRETETEPVEVEPYLEGQYK FFVSKILGKSMWRKPIKYPRRVPETLWRQANELIERSIRQGEVSIQSLNC MVYAAGCAVKSSLDKKDQEAKRRESEWYACRKAEIKALERYLNFIDLELK RRSASRPLTSRQRQNLGVLITKYGRARVRSGVRLSELQAMLRDALVGIRK CMAKRSADKKRKQGKFVPIQRYLEPSSAEPRLSPDTVRAYWNDIVGSSQ QSTSDSTIQDWSSNLSVPSQELNASKIMGWWRAAVSKSKPNKAAGPD GIPGVLWKRFRSASEWVCTWLYRLLQKRRIITPRWLSVGRVVLLPKKGPL EDPANYRPIACLNTVYKLITSVVEMAVREQIQACPGLVPYEQIANRKGVW GCTHASIVDRMITGASREGKGGGFPDLRVLFYDCKKAFDSVNRDHMFA VLRVANVNVKVVHLLHTLSQQWCVRYELRRNNRVERSSPLRVKRGLLQ GDTLSPTWFCLCMAPISASIKTLNPGPTLRPNMGNGRNRGQVAIQVSH VFYMDDLKVYCPRVADQRRMEQNIPQLFGEIGLSINASKSAAAAAVGRY VESELPVLGTKDEYKYLGIESGFVVNEVAALDRMQAVLLNRVEAILSVKEH TVGQRRDAIRAKAIPGGAYILGHIILSDLDPRGAAERMRRLDIEIRRLVKSA GILHDKCSTARIHLSCEQGGLAWPSMERAYYVAVAYSASYLLTSQDETISR ARDYFVSGRLSNKFTVYKHLTSIVDSLGLSVELPDPNGLPTGQPSVLARTIA RAIDAKLEAQWKETLLTYQRAGRVERADPTVVDHANSYHWLRKAWINE KAYQHAVSVMEGTLLEGVNPHGVLTMCRACKAPSASIAHIITGCAELRKS HMKVRHDGVTRWLYNALTEVDGSLPKFHYTQQIPAEMRGERLTVRYDS DIVTPNKPRHNRPDLVVFDSTRKVIYIVEVSVTWLSVLQKQYDNKLNRYA VNSNHEFSESIPYPPGVNLANEIRVLYPQFTGGVKVFPMIISPTGEVHMQ FVPHLAELLENPNIPRILEKIQRSVVLGTDYIIRSYFAM (SEQ ID NO: 1504)


AAGGTGTGATTACT
AAGCGTTATGAGTC
GGGTTATCTGGAAA
CTCCGCCCCGCCGC
AATGGCTTCAAGGG
TCAAACTGCTAACA
TTTTAAACCAATATT
TGGCAGCAGCAGA
CTATGATGTCAGAG
TGCCGGGGGTCTAT
CGATATAAGACTGA
AAGCGATAGACGT
GGAGTGAAAGGAT
TCCCTTTGTTAAGA
GAATCGGTAGAATT
CACTTTTTACTTATT
CAAGAACTTAACAG
CAACAAGCACTCGC
GAGGATTACCGCCC
CAGATTCGGTCGGC
GGTACTTCACCTGC
TTTCTTCCACTTTCG
GAATCTGGCATCCT
GGCTTTCAGTGTGG
TGATGGCCGGCTTG
AGTTTTCTTGAGTC
GTGCGAGTGCCTCA
TCTGGGACGTCCGG
ACCGATTGATGGAG
TCTGCAGTGGACGA
GGACTTGATGGACC
GTAACCATAGTATA
TCCCTCATGCGTCTT
CTCGACTCGAGGGG
GTAGCTTGCACTAC
CCACCCTTCTCTTCT
CCGATTGGGATTTA
GACCTAGCCCTCTG


GTGTGTCTCGACCC
GCGATATCAGATTC
CTGAATCGACTGTG
AGAAATGTCTACGC
GCAAAGATCGACCC
ATTGCCACCCGGCT
ATGTGGATCGGGCT
CTTGACTGCTTATCT
CCGGCTTTAATCGC
TTGAGGAAAGGGG
GGTGTCGCCCGAAA
GGGTTACGCGATCA
TCCGATTCCTACCGT
AAAAACGTAGTTGA
AGTAGGATTGAACC
TGAGTACCAAGTGA
AAAGAGTGGCTATA
ATGCTCCATGGTAT
ACCCTAGGGATAAC
CGTGTCGAGCGCAC
TGCTCAATACCCTG
ATTTGTTAGTGTAG
GTATGTCTAGTGGC
TGCCTAGGCAGATT
CGCCTATTCACTAC
GTAAAATCTGGGTG
AGTACCATAAGAAC
CTCCTGTAGGGCCA
GGAGTCAAACTAGT
CCAGGTTCGGTATG
CTTCGGTATACTCTC
TCACGGGCTAGTCA
CCTAAGAGTTAAGA
ACCGCCTTTTTCTGC
ACTGTGAATAGAAA
AAGAAGGGCGGGGA
GAATATACGGGGCG
AGGTAAGCACGTCG

				GAACGGGGGTGTGC CCAACTCGTACCGT ACGCCCGGACAAG GACCATCTTTCGTA CCCCGTTCACCCGG CACAAGTCCGATTG TCTCTCCCGAGAGG CGTCGGCGGAGGT TGGCTAACGCCGTC TCTCCAACCATCGG TTTGGGTTTAAGTA ACGCCCCAGTGGCC GGTAGCCAAACTGA TGGAGGCAGAAAT GCCGGTCCGTTGCT AAATGCGGGAACG GACCAAAACGCCG GTGTGGTGGATACA GGTGGAAGAGTCTT TTGGTCTACGCAAG AAAAGACCAGGCTC AAGTACGAATACGA CCGTCTTCGTGAAC GACGCCGTCGAGAC AAAGTTCGAAATCC TACGGTCGATAAGG CC (SEQ ID NO: 1259)	TAGTCGCTAAGGGGTCCGG AAATGGTCCGGTCCTGCGC TACCCGGTTCTGGTAGCAC GTTCAAGCGCTCAATCGCCT GCCTTGTAGGCAGTCCATCT GTGGAAGTCGCGCTCTTGA TACAGATGTGGACGGATGG AAGCAGATGATAGAGCCGG TGACGGCCCTACTAGCCAA ACGC (SEQ ID NO: 1383)	MPCSTNSFFERGTPEPHREPISGTDSSESLGMGTHRSPRLNDDEVINGPK GHESDPVHVVRAPRTLHPRRLELPIGVNNLGEASQLRQDSAIAEEAQLES TENHDGRRPPLRGGRKLWSEKEIATLRRLCEAYGNRQVCWKEVQRKFA DFHEERTVAALATKWGALKRPRAPMVGAPPTPDHDPERGPAGEGDGG TTSQENVPTDDPIPANGPTEGKESDVRPAVACRCTEPEEQLMESDVRPP AVVRLADPEQHTMKSGVKPVALDGSADLEERPKEKDIEQMGVDFEGEP RFRAFRKAFYGYFRWAVNSFDREPVKRVRRDCPKVFYAYADYLIATGSSK ALGPNQSRIGRLNGLVYAAARTIHQFWREEVGHRQQGEKGWYTKTKAT REDLQMLISMMESELARRKEKRKPGAKELENIHKLVARLGTRSTSGIVRRL EMTRQRLKLLEDRISLHEQEKRRKRLRKQFAETPSLKLLTKGAKDRGDTM
R4	R4_AL	U29445	Ascaris lumbricoides	GGGGCCGGTGGGT TTACTCACTTCTGAC CCACCACCAACGGA ACGAGGGAAAGCA GAGCTGGGGCCCTC TTCCGATTGGCATG GAACCGACCTCCAC GTGGTGGCCCTGG GCAACGGAATTCAA GAGAGGATTTAATC		

			CTCTCTATCATTTGC AAGATGGATGAGA TCGAGGTATCCGGC AAACAGGTTCCAAG TGAGCACCTTTCCC ATAGCTGGGAATAT GGGTTAGGCGTCCT CTGACATATAAGAG GAATCAGACTCGTT CGCGCCCGGTCATT AACATCGATCAGCG GGAGGGCCGGACT GAAGTAAATTTCCT GTTGGCCCGAGTGC AGGTGGAGCTCGG ACCCGAAAAACGAT CCCTAAGAGGACCA CAACCCGAAGGGAT GGACGCAGTCGCCC CGGCACGTTTGGTG TTGGTTATCCTGGA GTGTTGTGGGACG AATAGCT (SEQ ID NO: 1260)	TAACTTATTGTCAGAATTCC TTACTAGTAATAATAATTAT CGCTGAAAATCTCCACCCAA ATATTGCTTGGCTATATGCT CGCAATTTTTGGTTAACGTA CCCCAATGATTTGGGAGAA CAAAAATGGTAAAACTATA ATAATAATAATTATATTAAT AT (SEQ ID NO: 1384)	VTMKSVMDFWRPIIGRRVTSNPDQLQVLRDWRDEQKKAYPADLDLEK ADLEEKYEGAIRRIQPWKAPGPDGLHAHWWKALPSAKRLLGELVVDWL TTGKVTTGWMCRGRTILIPKKGDRGDPSNYRPITCLNTCYKVLTSVMNS VILSHLSRGEALPMNQRAMRKREWGCTHAMVLDRAMVMDAMAQKK HSLSVAWLDYRKAYDSVSHEYIRWAINSVNIPRSVQLTLKRLMSDWETR FESTQCRPKLRSDKMKVLNGIFQGDSLSPTLFVLCIAPISYALNKGVGQCQ SSSGWSAGYGFEIGHQFYMDDLKLYARTPAMILDSQIQVVSEVSEAMGL HLNLSKCAKAHYAPHGAGGAQEAVEGAEGSRKGEIPILGLRSTYKYLGVE QRLLPMEVALKEFEDKFMDRAETIFASELTWGQMATAYNTIAIAGLRYV YSNTNGASPKLLEALKRAATLDTRIRDLLRRHKCRFRNSFVERLYIPRECGG YGLKSVEDTLRESILATWSYIATNPHLAGQQYFFERLAARGKRTPMADGV KILLDLGVEPQVDLKRRTVTVDGIVFEDPTKLHRYLVGKLLKARTEARIRR WKEASLAGRLVNDTSIDMRLSCLWMKKGFVSARNLRDALAVQEGSLLT RACPALKGKGGQEVCRCCHAAPETAEHITSACRYWLPSLYVERHDSVAR NLYYVICCRYGITPVHYSNRVSPLSENSQCRVLWNMDMQTRTPMKHRK PDIVVFDLKREKILMFEVSIAHASGLLKQREIKINRYTVNSEELPDETITPYP PGPNLAADLAATYGWQVEFAPVVVGTCGEHVPAVKEDLQRTLDLKPHQ VEALLERISRSAVIGTARVVRAHLACS (SEQ ID NO: 1505)	ITYTANMALVTLFMENMENKRYNLRPLPGGRRGASGANAGCHSMRTV GDGGLSRRVPLEKNVTAEQSSSPLTSSSSHSPVSSIPSPSSTRTLLNSPNSS PTSSHSSLVIRSADVVQEALANYPAPTAGSIRARKKWTDIMNRYIWRTYLI ITKCETTLLNNYLEPLHQEFSSKFPEMQVTRQRIGDQRRAIIRNKLLSDDTL AQILIEVKELLQIGDQPLTQNNIHSTQLSHSNTRIKWSNELNEEIVKCYFEV TLLEVNKTSYRKNLYSLFISRNPHLSHLTEQRIADQRRLIFMNKSVHNDRII ELKREVEIKLANSNSLTKNITESNSPSSQTNEINDSAYVQSNLQPVEPLDQ HCINRHNLIEKHYVEQEFNNALIQFNNTNPETRPYIPRQKSSRKFSQIVSFL NSEVLPKHLNNELDFNALHNIIYTAXYTASLCNGTKFSFIDNYRPRNSKPS WQRRLESRIDKYRLQIGRLTQYISGNRNRKILKTVEEIKTQYKIHSHHEEPN TELPHFLDTLKQKLNATSNRLRRYLTCTKRKQQNNTFVNNEKHFYRTLSS TNQNTTTQLXEHPTENNLQQYWANIWETSIEHNADAEWLNKIPDXEIN XMKFKDISIETFNQIIQRTHNWKAPGTDNIHNYWYKKLTCTHSLLLKHIN QFIQSPCTLPLFITNGITYMLPKGLDPTNPANYRPITCLQTIYKIITACITDIIY

					KHIDQNNILAEQQKGCRKNSQGCKEQLTIDAIVMKQAHNKNXNTMYID YRKAFDSVPHSWLLYILKKYKIHPILITFLSSVMLSWKTRLKLINNNETLITD WIKIQRGIFQGDALSPLWFCLALNPLSELLNNTNTGFKLKHNNTYHIISHL MYMDDIKLYASNNKELKILADLTQSFSTDIRMEFGIEKCKVHSIKRGKSQQ NTYILNTGEQIESMDENSTYKYLGFQQAKQIQQKQTKIELTNKFKFRLNQI LRSQLNSRNIIKAINTYAIPILTYSFAIINWSQTDLSNLQRIINTHMTTHRKH HPKSCIQRLTISRLDGGRGLIDIRNLHNNLVTKFRNYFYAKAEISELHKFIVN IDNKYTPLNLNDRNIQLNQTLITKQQKIEAWSLKSLHGRHLADLSQTHVD KVASNEWLRRGDLFPETEAFMMAIQDQVIDTRNYQKHIIKRPNMVNDL CRRCYSSPETIQHITGACKTIVQTDYKHRHDQVAAIIHQHLAFKHSLITQA QKTPYYKYSPQAILESTNFKLYWDRTIITDKTVHYNRPDILLHDKVKXSVYL IDIAIPNTHNLASTFSNKIDKYTDLTIELKSQWKVQSVTTVPIVLSTTGVVP HTLHTSLETLGIHRLSYILLQKAAILNTCRIVRKFLSSNN (SEQ ID NO: 1506)
				TAGAGGACCCGAGTCTGAA GGAAGGAGGCACCGCCCA GGAGGGCGAGGAAGAGAT TTTTTTTTATATATATATATA TATATATA (SEQ ID NO: 1385)	MSGTXTDRVIPARTSPGSTRSASGVGEPGPPDVKLATGTRHSWSRAENV VLMECYYGSNPSERGYMQRMWEKWVLRNPTSSLTKKQLLAQCSNIRN KKLLSQLEIDEARRCASPTVQICYGKGEPGRQVSXGVISSSPPNIEIGYKAP MTDGLGTRAADLRERIMKSWGNSTTSLPRLTHKVPDQSLLEDMNTALS TIPTTTITETNQLMYAAATVILQMLGYKMKSMNSQKEQMAPWRRRLEA KIMATRREVSLLTELSRGVNLRTEXPKKYNKLSTTEALETAKQRLTALATRL KRYTREVEARRINKVFSTNPAKVYSQWQGNKMTTDPPRAETEQYWKSI WEKEATHNTXAQWLQDLQTEHSQLPEQDPVVITLADIQTRVSKMKSW TAPGPDKIHAYWLKKLTALHERLAAQMNQLLTSGNHPEWLTQGRTVLI MKDPQKGTIPSNYRPITCLSTTWKLLSGIIAAKISRHMDQYMSRAQKGIG NNTRGAKHQLLVDRAIAQDCRTRHTNLCTAWIDYKKAYDSMPHTWILE CLKLYNINRTLREFIQNSMKLWNTTLEANSKPIARVSIRCGIYQGDALSPLL FCIGLNPLSQIITKSGYGYQFRSGTTVSHLLYMDDIKLYAKNERDIDSLIHLT RIYSKDIGMSFGLDKCGRMISRRGKVIATDGVELPEGNITDVQDSYKYLGI PQANGNHEEAARRSATAKYLQRLRQVLKSQLNGKNKIQAINTYALPVIRY PAGIIPWPLEEIQATDIKTRKLNGKHKIQAINTYALPVIRYPAGIIPWPLEEI QATDIKTRKLLTMHGGFHPKSSVLRLYTKRKEGGRGLVSVRTTVQEETTS LREYIKKMAPTDRLLSECLRQQKPTKEEEPEGLSWKDKPLHGMYHRQIEE VADIEKTYQWLEKAGLKDSTEALLMAAQEQALSTRAIEARVVHTRQDPR CRLCGDAPETVQHITAGCKMLAGKAYMERHNQVAGIVVRNICTEYGLEV PGSRWETPPKVLENKQAKILWDFQIQTDKMVVANQPDIVVVDKHQKT VVVIDVAIPSDSNIRKKEHEKLEKYQGLKEEMERMWGMKATVVPVVIGT LGAVTPKLSRWLQQIPGTTSEISVQKSAVLGTAKILRRTLRLPGLW (SEQ ID NO: 1507)
R4	Rex6	.	Takifugu rubripes	TTCTATGCGCCTTAT GCGACTGGATAGG CCAGTGGTTTACGC CGCTGACTTTGGTG CGGAAGGTTGTCG GTTCGAATCCAGGC GAGCCCTTAGGCAA GGCTCCTTACGCAT ATATGCCTACACCT CGGT (SEQ ID NO: 1262)	

Table 3B.

1. Family	2. Element	3. Accession	4. Organism	5. Predicted 5'UTR	6. Predicted 3'UTR	7. Predicted Amino acid sequence	8. RT	9. Start	10. Stop
	Crack-3_SP	GCF_0000022 35.5	Strongylocentrotus purpuratus	AAGGTGGTTCAGTT CCTCCGTTTAGAAA CAGGGGCTTAAAA TTTGCCCATTTGAA TATTCATAGCTTAG TTGCTAAGATTGAC GAACTTAGATTACT TCTGAGAAATGCAT CATTTGATGTTATT GGTTTAAATGAAAC ATTTTGCGACTCC	TGAATGGCTCAGCCGTAC ATACACTACGTGCATGTTC TGTAATATTTACATGTCTT ATTCACATAATTGCGACTT AATTAATATTTTCCTATTG ACATTACTGCTTACTAGGT TAATGATATATGTTAATTT TGCAAACGAGTATTATTTT TGTAACTGTTCTGTATAAA TGTTATAAACGATGTAAAT ATGTTCATTCAGGGCCCCA TGGAAGACCACTACTTGCT GGTGAATGGGCTACCCTG TCAAAAGATCAGAAATAA ATAAATAAA	MIDDSELYIDDYILLRRDRTRHGGGVAVYISKR FNYKRRDDLISSTLECLWFELHYPNKSPILVGE FYRPPNSHVDFFEELDGNLDSALSEDMSCIIL GDFNCDCMPHQLDLNAEKLMFYTNMYGLS QLIDTATRVTNTSSTLIDLIFVSDKELYNEWGIF ETSISDHYLVYTVRDFSAKIHNKSEYAEFRSFK NLDEEKFYSELRSTPWQVIKDIDNVDVAWD MWLTLFNNIINKHIPVCKKRIRKKACPWISDEI VKLMTERDKVLKAAKANKSDDLWDTYKYLK NTVIHLMRKRKREYNADQISQNHNNSKQM WRCLRNLVPKSLPISPNSINVDGANVSDNANI AEAFNDFFISNATKITKDLRRNKSQLRVDEDN NIGPTHSQFNLNLITEDFVLKQIDNLSIDKATG EDHISCKLLKMTKNVIAESLCDIINKSLSTGVVP REWKKARVVPIFKSGDISSLNNYRPISILPIVSKI IERAVHQQLSEFLDVNKLLHPNQSGFRPKHST ATALAKLVNQWSLNIDNKEISGVAFIDLRKAF DTVDHELLLHKLACIGCSNRTIYWFKSYLGDR QQVTQFKGAKSHPSTIKLGVPQGSILGPLLFSI YVNSLPECTPEGIIDMYADDTTLTVSANDATV VEEKLTVALNRVMAWIKENRLVLNTEKTCVM LIGSPANLKKIESFTISLNGNLIKRVQFTKCLGLL IDEELKWSKQIENVSKLTQKNISMIKRAKSFLP QKSLKLLYNSLVLPRLDYCSVVWSNRFHSHTT KLQKIQKRAARIILNKTYDTPSADLFTSLKWM TLDKRFELSRVMMIYKCVHNLAPSYLQVGLIN	cd01650	452	710

	CR1-2_IS	GCF_0028928 25.2	Ixodes scapularis	TT	TGATCAGTTGAAGCGCTG ATGTAGTATTTGCTTTTTG CTCTCACACCTGGTGATGT CTCTGTTTTGTGCTTTTGAT TTCCCTTGTTAATCTGAGT GTTTTTCATACTTGTTAATA TTTTTATTCTCAGACTGGA GAACTACGTTCTCCTTCTG CACTCGTGTGTATGCGTGT TTGTATGTGTGTGTTTTTG CTCATTTTTTTTTAGTGTGC CAGCACTAAGACCGATTG GTTGTTCCTGGACACTTTA ATAAACATTGATTGATTGA TTGATTGA	PNDVHEHHTRQIDSGCLRVPRFHTDCYKCSPI VSSIFEWNKLDNSLRTASSVNSFKRLFKQSCN L MRQHNNILNDQGNALDLCFSNTDGIHVARS ELSMVPEDSYHPALDVVLSIFDPPVLSPSNGP SPHTSYKYSAGDYTGMYHHLSGVDWASILES SDINGQVYLLTTVVQEAMDMYIPLKTSKKSRF PIWFSRELRTLLKRKDFLHRRFKKSGLPKWEV EFKICRKRCKKILNRDRQAHSDSVESDLVKNP KKFWRYAKSRLGKNIESEIVLNQHSDSAKGM SELFADHFASVYINNNNCSSSILPHSDVFGDSL GSVVVDEECVRLCIKKMKPTLSAGHDGIPAAI VKAYHELLTPILCTIFNNSLTSGIFPDDWKLAV VVPVLKSGNSGDVLNYRPISLLSSFSKLFEMVI HQFLSFKFRSIVIPNQHGFMSGRSTSTNLVSF MSAASEVVCNRGQLDVFYFDLSKAFDVVNH NILLYKLSSYGVCGSVVNLLKSYLSDRHAYVRV NNTTSSLYKASSGVPQGSVLGPLLFNIFVNDV SRVILNSSFLQYADDIKLYKKISTLEDCIALQND AYSFGRWCLDNDLRLNHSKTKIMTYSRKTHD IFFPYSYHGELLSRVSELRDLGVVFDSTLRFDT HVIKVVQSALRTLGAVSRITKEFKSPSAFFTLYC SLVRSQLEYASVVWNGISKTNSSSVERVQNK FISIFKYRYLENGDPCLNGVCNSQLVNLLSLQQ RRVKADLLFLFKTLHGSINSPSLLSEVSLRVPRV PTRLQSSFYIARHFSNLNPIQRSAECYNRYSDT IDIFNSCGSKFGNAVQIFLLQEKG	cd01 650	316	567
	Jockey- 1_DSi	GCA_004382 185.1	Drosophila simulans	TCGAGCTTTATTGC AAGACCAACCAAG CAGATATTCTTCTC ATATCCGAAACCCA TTTTTCAGACAGGT CCTACTTCACCTTA AGGGGGTATGA	TAAAATTAAATTATATACA TACATTATTATATACATCAT ACA	MPRVSHPPEWFSILIRSSIKFDTQPPIQEDWV QCARISVSTTHGDITLAAAYCPPRFRATDDCF SSLLDSLGPKFVIGGDFNAKHTWWGSRVCN PKGRALCRQVLSRNLNCHGTGTPTYWPTDP AKTPDILDFAISGGLDPARIQTREVEDLLSDHS AIAVSVYASAMLRSAGKKLIFKTSDISKYHHW LNNSTVTNPPLDTPELIDAAVVNLTSQIQEAA TYAAPRFSNQTRDRSRDVHFWNPRVNELKS EKRRLRRVWFLSRNPRDKTALNRATKELKETL VRLRKEALDSFVGNLAPGDPHHNLWNVTKRI KRPLKRIPPIRRQDNSWCRTETERANAFAEHL	cd01 650	438	693

								126
							2	
						SSF5 6672		
				EEVFTPANRCSPNEAAETARLLSEPLFDSESIP QATEEEISNLIAAMGNHKAPGDDAINAIALKL LPPISKSLITTIINKCLQLGHFPSSWKRAEVVMI LKPGKHEANPASYRPISLLPVLSKILERVFLTRV LPVIEEAGLIPDHQFGFRRRHGTPEQGHRVA QYILDAFENKKYCLAVMLDIKQAFDRVWHPG LLLKLKNGLPQPYFKFLKSFLEGRRFAVRCGEA RSDVKEARAGVPQGSVLGPLLYNAYTADLPVI PSQYIMAATYADDTAFLTTADTPAEASQTM QEQLDLLNTWHTRWNISVNSDKSTATTFATR RGTCPPVSFNGFPIPEVANPNYLGFTLDRRLT WRPHLLKKRKQAENRVREFYWLMGRQSKLP TSTKLLIYKAIIRPIWTYGIQLWGTASKSNINIIQ TFENKTLRTVTNAHPFHDNATIHEVLSFPWV KDEIKKSSRQYMQRLQDHPNNLAQDLLDTS QRTRRLHRSHPLDLSGIRHAN	MAFDTVSHSLIIDKLRKCGLDDQVVRWIKNW LKGRRQRVVVNGAESSWRSVTSGVPQGSVL GPVLFNIFINDLDEGTECTLSKFADDTKLGGV ADTPEGCVAIQRDLDRLESWAGRNLMKFNK GKCRVLHLGKNNPMYQYRLGVDLLERVKGK GPGGPGGAEPQR			
					TGATGAAGGGAGTGGAAC ACCTCCCTTATGAGGAGA GGCTGAGGGAGCTGGGTC TCTTTAGCTTGGAGAAGA GGAGACTGAGGGGTGACC TCATCAGTGTTTACAAATA TGTAAAGGGTGGGTGTCA GGATGATGGAGCTAGGCT TTTTTCAGTGATATCCAGT GATAGGACAAGGGGCAAT GGGTGTAAACTGGAGCAT AGGAGGTTCCACGTTAAC ATCAGGAAGAACTTCTTTA CTGTAAGAGTGACAGAGC ACTGGAACAGGTTGCCCA GGGGGGTTGTGGAGTCTC CTACGTTGGAGATATTCAA GGCCCGCCTGGACAAGTT CCTGTGTGATGTACTCTAG GTTACCCTGCTCTTGCAGG GGGGTTGGACTAGATGAT CTTTTGAGGTCCCTTCCAA				
			Melopsittacus undulatus					
		GCF_012275295.1						
	CR1-4_MUn							

	Tad1-65B_BG	GCA_900519115.1	Blumeria graminis	AT	CCCTTGGGATTCTGTGATT CTGTGA	MKIAECGVPQGSPLSPLLFLYIAILLQDENKN	SSF56672	6 148
					TAAGTCCTGGCGAAATAG	TKFGYADDIAILSIGNTASEAVAEAQKEVDKL		
					AGGTTTTTCAGGACAAAG	VQLASENRIDFDSTKSELLVIGGGPKKKLDTSG		
					AGTTGTGAGGAAGGAGG	LSVQIQGHSVPTSPYVRWLGVWLDSQLNFK		
					GAATGCGCAGGGCGCTTA	QHVQEWCGKAQRITQFMRQTNSVQRSAAP		
					ATTTTCTTGGAACTTATTTC	GPMIKVVQACVLSTALYGAEAWWPGLTRITT		
					TTTATTTATGTTTTACGTAT	QRGKKVCTGVGWHTDLLDSTILKAVRAALPV		
					AGATTTGACAAGCAAACT	WRTTPNMALHRESGITPSAVILQQRNILAAA		
					ATGTACTTAGGTTGTAGTG	RIQRLDDWHPLVFRSREDPKDTISRLGLCICQ		
					TGTCCCGAGGGAGCATAG	KNRSFKFPEWYLTRLEISAKNVPMAEGPGKLY		
					ACCCGTATCACGCCTATTG	FPEAESIRKNTGSDFQDEADYAKQWIASLSAE		
					GGTGTAGGGCGTAAGCCT	TVCAYSDSSSCGPARSAWGYTLYKGGQVIGS		
					GATACGAAATGAAATCAA	DSGPLLGAEVYDAEIFGAQKALERALGKAGN		
					ATCAAATCAAA	CPIKVLLDNTTAVRALRTGQTKSSQEAVEVFT		
						RLARQARSVEIRWIPAHAGVMGNEDTDRLA		
						KEALRNLPDETDTSPSEMRDRPKYTFAALTRF		
						VRQRCDEVVESWWRDHRPSRYEELDLQMK		
						RKRPPELALPRWAHHRLITARTGHGDFIKYHE		
						RFGHEFENCECVCGREKRPWHFAECRIALQQ		
						WRNLKHEPPPRGREMLAEKGWERFLQFLTV		
						TKCYERQTLHRED		
	RTE-1_EL	GCF_011004845.1	Esox lucius	CTCCTTGAACTGCC	TAAGCGGAAGATGATGCA	MQPDPRTRKLALGTWNVTSLAGKEPEMVRE	cd01650	529 784
				ACCTTAACGTGGTG	ATGCAATGC	VERFRLEVVGITSTHGLGSGTTLLERGWTLHH		
				GAGGGGTTTGAGT		SGVAHGERRRAGVGLLIAPQLCRHVLEFTPV		
				ACCCGAGTGACCCT		NERVVSLRLRVGDRSLTVVCAYGPNGSAEYP		
				AGGAGCTATGTTGT		TFLESLGGVLESAPTGDSIVLLGDFNAHVGND		
				CTGGGGCTATATGC		SDTWRGVIGRNGPPDLNPSGVQLLDFCASHS		
				CCCTGGTAGGGTCT		LSITNTMFKHKGVHQCTWHQDTLGRRSMID		
				CCCAAGGCAAACA		FVVVSSDLRPYVLDTRVKRGAELSTDHHLVVS		
				GGTCCTAGGCGAT		WIRWRGRKLDRLGRPKRTVRVCWERLAESP		
				GGGTCAGACTAAG		VREIFNSHLRQSFDWIPREVGDIESEWTMFST		
				AGCGGTTCAAAACC		AIVEAAARSCGRKVSGACRGGNPRTRWWTP		
				CCTTAATGAGAAAT		EVRDAVKLKKESYQAWLACGTPEAADGYRQ		
				TGTGCCGTGACGTC		AKRTAARVVVEAKTRAWEEFGEAMEKDYRL		
				GCCCGGCATGGCG		ASKRFWQTIRRLRRGKQCPTNAVYSRGGQLL		
				CAGCCGGGGCCCC		TSTGDVVGRWKEYFEDLLNPAVTSSIEEAEDE		

	ORTE-3_AAe	GCF_002204515.2	Aedes aegypti	CATACGTTGAGTGA CTACCGTTGAGAGC GGACGTGTCTGTG CTTGCGTGCTGTTA CAGTGGTTCGAGT GAAAAGATCGTTTT TGCGATTGTCGTAG GAGGAATTGTGCT GGGGTTTTTGGGTT TCCGGAAGTGCCG TGTTGGAGCAGTT GGGAGGTTACCGT GCTGGGTTTCCGAC GGATTGGACTGGA TCCGTGATTTGGCC AGCAACTCGGCTG GGTGAAATAGGCT TTTCGGAGAGTAGT GAGAAAACGAAAA AGCAGCTGGTAGT GGAGTGCGGAGGA	ACCCTGGAGCCAG GCCCGAGGAGCGA CATGGGGCCGCCC TCCCGTGGGCTCAC CACCCACAGGAGG GACCATAAGGGGC CGGTGCGAGAGGA TCGGGCCGGCAGTC GAAGGCGGGGGCC TGA	TGATGCGATTCAGAACGA CTGGAGTTGAAGAAGAGA GGAGAGAAGATCACCATA TTGTGCATTGCCAGAATAA AGGTATTGTAAGTACAAAT CAATGTTTTAAATATGTAA AATGGAAGTCAAACAAGT ATACGTTATACTGGGAAG GTTTAATTTAATGGTAAGT GAAATTATTACAAAACATA ATCAATAGCACCAATCGTA ATTTTCATATATTATATATC AAACAGGACACCCAATTG AAGCAATATGGATGAGGA GAGTGGAATTGCCAACTC CTTATCGCTAACATACTGT TGATAGAGGGATGGTCCT TAAACCCCACTCATTGAGA CATCCTTTTGTATGATGAA ATGACCGGTAGATCGGTG ACCAGTATCTAGGTTATAA	GSEVDSSITRAEVTEVVKKLLGGKAPGVDEIR PEYLKSLDVVGILSWLTRLCNIAWQSGTVPLG WQTGVVVPLYKKGDRRVCSNYRGITLLSLPG KVYARVLERRIRPIVEPRIQEEQCGFRPGRGTL DQLYTLYRVMEGSWEFAQPVHMCFVDLEKA FDCVPRGILWRVLREYGVLGPLLRAVRSLYDR SRSLVRIAGSKSDLFPVHVGLRQGCPLSPVLFV IFMDRISRRSQGPEGVRFGDHTISSLLFADDV VVLAPSSQDLQHALGRFAAECEAVGMKISTS KSEAMVLSRKRVACPLQVGGECLPQVEEFKY LGVLFTSEGRMEREIDRRIGAASAVMRSMYR SVVVKKELSRKAKLSIYRSIYVPTLTYGHELWV MTERTRSRIQAAEMSFLRRVAGRSLRDRVRS SVTREELRVEPLLLHIERGQLRWLGHLFRMPP ERLPGKVFRSRPTGKRPRGRPRTRWRDYVSR LAWERLGVPPEELEEVSGEREVWASLLRLLPP RPGPG	MMTMTFNGIRRLGWFLSVLSLVKQRLLNSS HESNSLSKLILKQICVKLNGVNFTQEDVDELN SIILSKIEGWEQDLNKFIEFFDNIQFEVFDIAEM ANHCLLLAVSQIFECEVRVSNDKDDCKIIGEG YKNSWFGVLEIGMESSERVEKVTRALKGSSKI KLGKCGKDRQVNGMAGKGNEPLIVKEECLW YTTIGNWIIEIESKGEDDGKLYSGLMMNVIQK MNEGEIGEQEMMMEMLKQIDLFFEMYGEILN RLEEEVMINGNALYEIWKIVSKLAVALNREIIV LMRGGEEIVLGVTERKRKFASFWLVAEFQEV GVRYYSIKSSKLIQEWKENMSIKPAIDVFEKRA EEVSYNLRFKESNQQAVIKRIDGDGSCLIRAM LDQMGRIDQNYYMSDYNVTKIRNNIASYINN SRERFNVYLYEHIEEVGTIERVIENIREKSYWLG EEAIITVAELFEKQVEVYTNQGLSVRIGVRKDN ESALRIFYTGNHYDSVISIAEEEEAGEIEAEMINI NEAEGNKMQAGGESRAVKSNKVREVNESSS HSKKITQYGKQSTNTTGDSNNHEQSKKKVEK QLRKDSEIKDKENKALKIVSWNVRGCRRKVK RDEIDDILVANDVMIAALQEVNTIGDDLSTN NYSWQVKDNHTNKARGLAIVVRKNQRIRILQ	cd01650	1078	1338

		GGTAGAAGCCTAG	TTTTTGGAATGCCTTAGGT	AKECRKGIMWIKLTINGVKLIVINVHAPNSNQ	
		TGGAGGAGTGCGG	AAACCTTAAATAATAGCAA	KNFLSYLNEIVTRESDRKSLIVLGDFNAQIGSE	
		AGGGGGTGAAAGT	AGAACTGAAACAATCAAC	DLQPEDRVWIGTKLGHRYCNDNGELFKYFLH	
		GCGGTGAGGGCTG	TGGATAACATCATATAAAA	LSKLKNVSTKIGKSIEITWRTKDKSSQIDHVLV	
		TAGTCGGTAAATTC	CGGACAGTCTACTGTCTAA	PQYTSFRIRYIKGCWTDLKTDHVMLTIGLVFT	
		CGCATTTCAGATTG	CGGCTGTGCCATGCAACA	NNPVLKKSNKDTARINVDLLKIPEKQKCFREKL	
		CTGATTGTTAGTGT	ATGAGTGGTTACCCTTACA	SSIPTKTDIQTDTLDKCYEEMASRLRIAAYNTL	
		CCTGAAGAGAAGA	TACATACATACATACAT	KVSTVPNTPRRKYALDRLRDALKVSKKHPDMI	
		CTAAATTCCTTTAA		NYKYRVANRRHEYQIAVKEHREKEITDFFKNL	
		TTGGGGCAGCTGG		NDYDAGVRVRKTYNYLKDFVKRKKQKKAMIS	
		CGTGGTTAAGGGA		SRVWEDILKTCEGPEVEFCAEHDYTPLPNPPT	
		TTGTTCCGGGGAA		NDEIKSILFSACNGKSPGSDRVVAEYIKAADEE	
		GCGGAAAGGGAGT		TLERIEIIKAVYIRNDMPKEWKKSTQIPIPKKS	
		GTGAAGTGTTTGG		GASEADHFRRISLCNVVYKIYAKWIAKQLRAY	
		CCAGAGGAAGTTG		AGEPGLHQAAFTANRSTDDHIFVARRMMEE	
		GCGGAAAAAAAAA		YWNSGQDLFILALDISKAFDNVKLQSIGKVLID	
		TTGGAACGCCGAA		LGVPSRLVDRVLSCVKDEVTRIMWENQYSAE	
		GAAAGAAAAAAGT		VKRGKGIKQGCPLSPLLFNYLIQDVLKKVVAEI	
		TTTGAGGCGAGAG		PELKLMGINQLILPLILVFADDILIIARTKAELEKI	
		TGGCGGGGGGGGTG		LHELELQLAKVGLEINRDKSQVLIRYPNDKAQ	
		TTGCGGTCGAAAG		HPKEILLNNKIFKVSDSLAYLGICLTATLNRKLT	
		CTTCAACGGAAGA		CKQRCLNAVKSSRIIVEFCKTFKPSWELGKLM	
		AAACACGAGTGGT		YKTVIAPSLLYGTKVSVMTKKSRLKIGNYEKLIL	
		GTGAGTGAGAAGA		RNIYQYCKKPKDLKFNARKLLDGKTINRRIRV	
		CCTTGATCGGTTTT		GRISYFGHVRRREKNHPLKLAYKLKFDKKKEG	
		GGCGGTAAAGCTC		RPSLTWLKSLDQDFERYQNMNRSSWEELSK	
		GTTGCAGGAGTTC		DREKIKKKLDEIYKNEESEISDGESSDGDQNSR	
		GGAAAAAAAAAAA		YKHWKWKRRNEK	
		AAGAAAATTGGAA			
		GCGAGAGTGAGTG			
		CCACGTAGAAGCG			
		AAGTGGAGAAAAG			
		TGGCATCGCACACT			
		GTGAGAGGAGGAA			
		GTGGAGCATCGAA			
		ATAGGCAAGCGGC			
		TGAGTAGCGACAG			
		AGAGAGAGTAAGT			

			GAGAAAGCGCATC TGTAGCGGGGAGG CAGAAATAAGAAG CAGGGCTGCTGGT CGATCCGTGGCGA ATGCGGTGGAGAG CACGAAGGAAGAT TCCAACGCAGCGA GACTGAGTGCACC GGTTTCATTGGTGG CAGTATTGGTGAGT CATTCACTTTCTCGT TCCAATAATTTTTCC AAGTATCTCGAAGT TTGGTCCTC					208
	Proto2-1_OL	GCA_004348135.1	Oryzias latipes	CA	TAATTTATATCTTTTTACTT TTTATTATGTATACTGTATT TGTATCTTTTTTCTATGGAC CACATCTGAGTCTGAAATA AAAGTGTGTT	MVVKWGNVLSEPFYVSNGVQQGGILSPFLF NVYIHDLSTNLNLCKTGCVVGCQTFNHLMYA DDLVIFSPYSAGLQTLLKVCSDYGRDFDIKYNP SKSKVMIVRTKDDRNVNFPVFKLCGTTLEQTS SIKYLGHFITDDWRDDRDIQRQYCKLYAQAN MLMRKFSACSLQVKCSLFRAYCTPLYTAPLW WNYGRGTFHKFLVAYNDAFRLLLQVPRWYR ASQLFVDAHIPTCEALLRKLTYSLMERLDKSK NSIIMSLTDPKRSSCRYTSALRKHWIQRLYSF	SSF56672	13	208
	Penelope-2_Sln	GCA_010367695.1	Solenopsis invicta	AAAGATCAATGACC AGCTAATTAAAGAC TGGTTACGAACCAC GGGAAGATTCGTT AGGGAGCATCCTA ATATTTTGTTTACC AGAGCGGACAAGG GTAATGTAACAGTT GCCCTTGACAAGAA TGATTATATCTCGA AG	TAGGTTCATCACCCCTTCT CCTTCTATTCCTTTCGTTCG GGCCACCGAGGCTATCTC GTTTCGGTCTATATTATTT GTATCTATTCTCACCGAGA TGTAACATCCCATCTCAAG GCTGTGACACGAGTTTTG GTGCTTAGAGGTAATTTGT ATTTGATTTTCGCTTTGACT TATCTCCTCTTTTTCCCTTT TTTGTTGAATTTACATTACT TTATAGAAGTGTATTTTTT ACAATTTTATATCTAATGA CTTGTACAGCTTCCATTTTTT	MEGMLSDESTYVKLSKDPIQKLTKEARSLLSI WKDKGFIEQHIYRSLLTTDGVLPRAYGLPKIH KPNNPLRIIISSINSPLHSLAYFLHNIIYNSIPEAD SFVKNSFHLVNKLNGLHVESDHVLVSLDVVSL FTNVPKNLAVESVTTRWDLISKNTTVPVNEFL KALNLIIDSTYFKFNNTCYKQIFGLPMGSPLSPI LADLVMQDLECKALKLLSFHLPFYCRYVDDIIL AAPANSLSGILDTFNSMHDRLQFTMEREIEG KISFLDVMLINVNNRIIFDQYRKPTFSGRYLNF FSHHPLCHKKGVIFGLVDKLLLLSHPQFHCKN FINLVNILLENNYPLNFIFSMINSRLKHLFYKDD TYVEDKRVRRFFTIPYIESVSEKFSSFSCKFDCS VAFTIPSKLNKFITTGKDHTERLSCNDVVYKIN CKDCDSSYVGQTKRQLKTRVSEHKADIRKASS	cd00304	189	269

	PENELOPE	GCA_007989325.2	virilis group	GGTCGCCAGAGCC ATCAATAAATATCA ACGGAAGGCACGC CGTATGCACAGCAA CCAAC	ATTCTTATTTATTTATTTAT TGAATTTTTTACTTATATTC CGCTTATTCTTCTTTTTTAA CTATTTAATTTTCATTGTTG TATGCCACAAGTTTTATAG TCATCATTTTAACATCTTTC TCTTATTTATTTATTTATTT GTTAGCAACCAATCGATGT TGTTTCACAAAACTCTTGT CTCGGCACACGCTCTGTGA	PSVISLHRIGENHEFDWENVKILDKEPSVVKRS ISEMLHIKNQMHGLNKQSDTELLSDAYLPILE RIPPS		470	558
					TAAAATAATGTATGTAAAA TGTTCGAAATAATGTTTAA TTTATTGTATTATAATTGTT AATTGTTTTTTGTATCTTG GTGTTAGTGCCCTGAAGA CGGTTTGCCGATGTGCAA CCGAAATATATCGGAAGA GAATTGAATAAAAATTGTTT TTCATTGTTTGTTTTAACAA ACACGGACCTCGAGCCAG CCAACAAATAAATATTGAA ATATGGAAA	MSYAKIKTKYKDSKRTINKFQLTLVKLTKLKSSL KFLLKCRKSNLIPNFIKNLTQHLTILTTDNKTHP DITRTLTRHTHFYHTKILNLLIKHKHNLLQEQT KHMEKAKTNIEQLMTTDDAKAFFESERNIEN KITTLKKRQETKHDKLRDQRNLALADNNTQ REWFVNKTKIEFPPNVVALLAKGPKFALPISK RDFPLLKYIADGEELVQTIKEKETQESARTKFSL LVKEHKTKNNQNSRDRAILDTVEQTRKLLKEN INIKILSSDKGNKTVAMDEDEYKNKMTNILDD LCAYRTLRLDPTSRLQTKNNTFVAQLFKMGLI SKDERNKMTTTAVPPRIYGLPKIHKEGTPLR PICSSIGSPSYGLCKYIIQILKNLTMDSRYNIKN AVDFKDRVNNSQIREEETLVSFDVVSLFPSIPI ELALDTIRQKWTKLEEHTNIPKQLFMDIVRFCI QENRYFKYEDKIYTQLKGMPMGSPASPVIADI LMEELLDKITDKLKIKPRLLTKYVDDLFAITNKI DVENILKELNSFHKQIKFTMELEKDGKLPFLDS IVSRMDNTLKIKWYRKPIASGRILNFNSNHPK SMIINTALGCMNRMMKISDTIYHKEIEHEIKE LLTKNDFPPNIIKTLLKRRQIERKKPTEPAKIYKS LIYVPRLSERLTNSDCYNKQDIKVAHKPTNTL QKFFNKIKSKIPMIEKSNVVYQIPCGGDNNNK CNSVYIGTTKSKLKTRISQHKSDFKLRHQNNI QKTALMTHCIRSNHTPNFDETTILQQEQHYN KRHTLEMLHIINTPTYKRLNYKTDTENCAHLY RHLLNSQTTSVTISTSKSADV	cd00304		

	l-3_SPur	GCF_0000022 35.5	Strongylocent rotus purpuratus	GGTAAAAAGTCGC CTCATATAACTAAT AAGTCCATATCAGA CCTATTGATTGTAC AATGGAACGCACA AGGCTTGACGGCG CATGGCGCAAACTT TTTATCTTCCCTTAT TGATAACAATTTGT TAGGTATTCACCCA GATATA	TAGCTTTTAAAGCTATCAG GGATCCCCTTAACTTATGT CGATCCTGTAATAATATAT AAAATGAAGTATTTGCGG TTTTTCCATGAACAATTGC AACTAACTCCCATCTCGCC TTTTATGTGCATTTTATGC AATTTTAATATTGCAGCTT TTTTTTTTTTTTTTTTTTAT AAAACCATATCACTATTAA TTTCATTTCACCAAGGAAA GTCGCTCGTGGGCTTGATT TCCGAAGCCCACGCCAAA GCTATACACGAAATTGACT GGCGCAAATAATCTTTGA GATTGAAGAGCGCCATCT AAACGAAGAACCAACCAA CC	MCIQETWFNDFNVLSIPNYVCVCKNRPDGR RGGLVMYLKNDITFQAMDTPRNAEYQKVDI FCKNGVISIFNYYNPCKKIELIDLNNMIRNVPG KYLILGDFNSHNPLWGSNKLDTNGRTVESFIE NNDLVLLNDGSPTRLDPYYGTYSNLDLSLVSK ELARNTEWKVLSNNFGSDHNLVCLNFKHVY NRPGGENMGQRWSFRKVNWEEYREEIEEAL RKEGYENLGAELNIQGRYNVIVKVIKETSNKCI PRVQIAKRRKVNPWWNNACKKAVQERNRL KNRLRRYYDLNDVKKYLKKKAVAQKVIRRAQ YDYWLEYCRRMNRFTDIGKVWEQVKRMRD STMNKHIFHNLFVDGNILKESEEKANAMATY FKSVDESLSVTKNENENILYNNSSTNDDDCIN DDFNMSEMGNVLKEMKNSSSGKDDIYLIMI KNLPNKALVYLLNLYNDIWKISLFPHQWYEVI QIPILKSGKDPTQVTSYRPISLISTFSKIMEKMI KNRLTWFLEKNRLLNPIQSGFRKYRNVKDHL VRLETDIHISLKSKKYTVAVFLDLEKAYDSLWK TGLLRKLSILGIGGKMLLFIKAFLLSRISNVRINN YMSRHYTLENGIPQGSSLSPLLFTLMLNDLKI NVSDVKLSCYADDIAIWCSNKDLELARRKLQE ALHELEIWCWTWKLNVSPTKSAVLVFTNKQK KDILLQIKGHNLPICEQYKFLGVRFDSNLTWS VHIDSIIDKCKKRINLLRCVSGTPWGSKYEILM MIYKGLILSILDYGSEMYDSACISLTKKLDSIQY RALKIITGTIHNVSLEALQVLTGELPLRYRRRLL NDNFRVYLLSSTLNHPTKEVLGDDYDLNNLE WKIGEGPFKLRSFRDEVVVEGQTNLLPKIPE WRLARPQVSFEHHGKVQKKEMPPFEIKSIVL DSIGNSWNGFLHIYTDGSKLEDGKTGAAFYV SHYNVVKFFRLSHVCIMRAELVAILMALEWIN TFIESSSVVIFSDALSALQMIDSYFIRCAIVNEIL SIINILQFKDVCVAFEWIPSHCDIRGNEIVDRA AKKGAMKRCIDIVIPNTRSELAASRLIHYNKL WQAEWDNAEHGRFLYALHSDVSKPIHRQGL SRREECLIQQLRVGNCSLNYYLFQINQHDTGL CEVCQEDETIKHFIFYCTKYDKDRRLLKERIKLN NPNLKQLLMSNPIFTRELLGYIYDTNRFR	cd01 650	462	717

	L2-3_XT	GCF_0000041 95.4	Xenopus tropicalis	TATCCCTCTGCTGT	TAGCAATACCCTGTGCCCC	MFYPALLLFALSCFLKHLQPHPPSSQKGITITA	cd01 650	702	964
				ATACTGCCATATAA	ACCTCTCACAACTCTGGTC	SLLSSPSLSAHEVFKLLCSITPVNSFSSHKTKTY			
				CCCACTATATTGCA	ATGCCCATTCCCACACCTT	KSRSHLAFLSFLLLAAGDISPNPGPYPIPVLLGT			
				GCCTCTTTATTAAT	GTGTCTCAACCCTTTCTCCT	RSPNPPCTPLKYRTLTPIPKASIQISCALWNSR			
				GGTAATTCTCATCT	TGTAGATTGTAAGCTCTTT	SVCNKLTSIHDLFIAKSLNLFAITETWLSPTDTA			
				GTGTATCCCTCTGC	TGGGCAGGGCCCTCTTCA	SPAALSFGGLHLTHTPRPGNRAGGGVGLLLS			
				TGTATACTGCCATA	CCTCTTGTATCGGTTATTG	PRYSFKVLPPPPSLSFSSFEVHSIRLFSPVSLRIA			
				TAACCCACTATATT	ATTGCTTTATATGTTACTCT	VIYRPPGPTSQFFDNFSAWLPYFLHIPTSWTD			
				GCAGCCTCTTTATT	GTATGTCCAATGTATGTAA	LTILDSFSPPAIHLKSFHLLLLFHFPHLRFTLLGS			
				AATGGTAATTCTGT	CCCACTTATTGTACAGCGC	SPLFRFALRSYTDLLDRPHNSLTTFLPGFLTSFI			
				CATCTGTGTATCCC	TGCGGAATATGTTGGCGC	YRPPGPTSQFFDNFSAWLPYFLSSDSPSIVLG			
				TCTGCTGTATACTG	TTTATAAATAAATGTTAAT	DFNIPVNNPNAPATSKLLSLTSSFGLSLCSDSP			
				CCATATAACCCACT	GTAATGTAATGTAAT	SHSNGNSLDLIFTKCCSTSNFTNSPFPLSDHNL			
				ATATTGCAGCCTCT		LTFQLSLTPSTPSAVPQTYTYRDLQSLNTSHLS			
				TTATTAATGGTAAT		SSFDSLHSNILTVSCPNQATSVYYSTLTTALNE			
				TCTCATCTGTGTAT		LAPAKTKRMRPKPLQPWHTAHTKSLQKQSR			
				CCCTCTGCTGTATA		TLERRWRKSRSESDFCSYKSALLSYNTTLCQA			
				CTGCCATATAACCC		KQTYFTALINSLSKKPAQLFSTFNSLLSPPPPPP			
				ACTATATTGCAGCC		CASVTAQAIAEHFKNKIDTIRSDIAQLNPHNH			
				TCTTTATTAATGGT		SPPSLHAPQSLLGSFSPVTEEEVSKLLSSSHLTT			
				AATTCTCTCATCTT		CPLDPIPTKLLRNANPCLIKALTHLFNLSLSTGI			
				GTGTATCTGCTATG		FPSQLKHALVTPILKKPSLDPSNPTNLRPISLLP			
				CCAATAACTCATTC		FISKLLERLVYNRLTLFLSDNNLLDPLQSGFRQ			
				TATTAGGCCTACCT		QHSTETALTRLTNDLLSAKAKKHYSLLILLDLSA			
				TGATTTAGTACTTG		AFDTVDHPLLLQTLRSLGLRDTALSWFSSYLS			
				CCATAGCTGTGTCA		DRSFSVSYNGESSSPLPLSVGVPQGSVLGPLLF			
				AACTCACATCTGTG		SLYTSSLGKLISSFGFQYHLYADDTQIYLSSPDL			
				TCCACTCCTCAGTG		NPELLTRVSSCLSAISTWMSQRYLKLNLSKTEL			
				CCAATATTTCACAA		VLFPPSNAHIVPEVSITVNNSTITPSSQARCLG			
				GTGCTAGGAAAAT		VILDSALSFTPHIQSLIKSCHFHLRNISKIRSFIT			
				TGCAGTTAAACAAG		QDAAKILIHSLIISRLDYCNSLLIGLPLQRLSPLQ			
				GCAGTTCATGGAA		SIMNTAARLIHLSNRSSSAMPLCQSLHWLPLP			
				CTAACTCATATCTC		SRIKFKLMTLTFKAVHNSAPPYISELISRYHPTR			
				TTCTCACTTCTAAT		LLRSSTDLLLNSSLIPSSHARIQDFARAAPLLW			
				GGTAATTCTCATCT		NSLPQTVRLSPNLSAFKRSLKTHLFREAYPNLV			
				GTGTATCCCTCTGC					
				TGTATACTGCCATA					
				TAACCCACTATATT					

				GCAGCCTCTTTATT AATGGTAATTGTAA CTCATCTTGTATCT GCTATGCCAATAAC TTTTTCTGTTAGGC CTACCTTGATTTAG TACTTGTGATAGCT GTGTCAAACTCACA TCTGTGTCCACTCC TCAGTGAGCAAGA GATATTTTTACAGC CCCAAAACTTGGTT CTGGTGCTTGTAAA AGTCTCCCCCTCCC CACACAGCCTGCTT ACTTTGATCCCTGA TTGAACTGATTAAT TAGTCAATTAGTTT CTGGCTTTCCCCAA TCAGCTGCATATAA ATTCAGATGCAGCC TTGGGGATGGCAC TAGTGTCAGGGGC AGGCACTGGTGCT ACCATTTCCTAAGT GCCAACATTTCACA AGTGCTAGGAAGA AGGGAATGGCAGT TAAACAAGGCAGTT GGACAAGGCATTG ACAAGGCTACAAA CTATCTCCTATCAA CACTGAACAACGTG GAATCGCAAGAAA ACAACTTTGCTTAT TCATGGTAGTTTGT TATTTCCTATCCCA AA					

	L1-1_VVi	GCA_011037095.1	Vitis vinifera	GGAGTGTTCAGTG AATTATTGTGGGAC TAGAAAGAGTGCG GGGTTGCTTCAGAT TGCTAACTGAAGTT TGGGTTTGTTGGTC CAGTGAGGGGAGA GGGGCTTATTAAG ATGGATGTT	TAAGGGTGTGTGGATAGGGT GGCTCTTCGTTCCTTGGCA GGCTGCTAGGGAGCGTGG GGTGTTTTTCTTTCTTTTTC TTTTCTCCTTCTTGTACTTT TAGCCACTGCTTCGGTGGT TTTTAATACAATCTCTTTAT TTATCAAAAAAAAAAA	MKLRVLSWNVRGANDSEKRKVIKALIKSQKV DVVCLQETKMQEMSRMIVRSLGVGRCLDW KVLNSRGSSGGVLVFWDKRVLHLLEAEVGNF SVSCRFKNCEDDFCWLFSGVYGPSLMKERED FWAELGAVRGLWSDPWCVAGDFNVVRFPV ESSRGGRLSASMRRFSEIMEDLELRDLPLQGG SFTWKGGLNNQSHSRLDRFLISNEWEDHFSG SVQYVLPKPTSDHFPILLDGGGVRSGPMPFRF ENMWLKEEGFKEKMQGWWVGLNFSGSAS YVLVSKLKALKSLLRDWNKLEFGKVEVNKALA LSQVDFWDKMELSRTLSVQEVDARRGAKED FKKWALMEEISWRQKSREVWLREGDRNTKF FHKMANAHRRGNMLSRVKINGVWLTKENE VKEGVVNEFKAMLSSAGGWRPSLRGLSFERL EAVDAASLEEPFSEQEVMEALKGFCGDKAPG PDGFSMAFWQSSWEFVKEEVLGFFREFHNH GRFVKSLNATFIVLIPKKGGAEDLRDFRPISLV GGLYKWLAKVLANRLKRVVGKVVSKAQNAF VQGRQILDAVLVANEVLDSVLKNKEEVVMCK LDIEKAYDHVEWSFLFSVMRKMGFGEKWIR WMKWCISTVSFSVLVNGSSSGFFQSSRGLRQ GDPLSPYLFVLVMEAFSSLLRKAVAGGFVSAC KARSRGGEGVNVSHLLFADDTLVFCGASKVQ LLHLNWILMWFEAMSGLRINLDKSELIPVGC VNNVEELAAAIGCKVGSLPTSYLGLPLGAQYR SRAVWDGMEERMRKKLARWKSQYISKGGRI TLIRSTLANMPIYFMSMLSMPRKVRLRLERIQ REFLWGGGALERKIHLVKWELVCLEKDNGGL GVKSLSILNKALLCKWSWRFAMEREAFWNQ VIRGKYGEEQGGWSSKEARGETHGVGLWKT LRKEWEVVKSRLVFVVGNGKRIKFWKDIWC GDEPLCVSFPSLFALAVSKDAWVKDVWRCN EGGGSWSPLFSRPFNDWELEEVCSFFVALNR KQIQQGVDDRVIWRETNCGKFSVKSLYKSLV SGHPISFPSSAIWKVTVQPRVSFFGWEATWG KALTLDQLQRRGWALANRCYLCQRHEESIDH ILLHCDKVRTLWVLLYSMFGVQWVLPATVKE TLSGWNGSFVGKKRKGVWKASPLCLFWTV	cd01650	496	763

	R2-2_SMed	GCA_002600895.1	Schmidtea mediterranea	GTCATAGGGTGAA	TAGAAGGGAAACAAAGGA	WKTRNKVAFEEEELSIQRLKASFVYFLWSETK	cd01650	474	730
				CTGCAATTCTGACA	AAAACGAAATGACTGGTA	RSIKDGPSTLVDFVGWVASR			
				CGATGACCGAGCT	ACTATGAAGGATATAGCT	MRKELVTIKNLIEESGATAPAPVPLEVAVEVH			
				GTGTCAGTTTGCAG	GAAAGCCGAAAGGAAGG	QSSSVPEITDQSTTTQEGSYSEPPIHRCENCG			
				CTAGTCGCTAAAGA	CTAAGTCCTGAAACCGATC	REFRTRAGVQQHRRKAHTNEFMEEKEKAAP			
				CTCGATCAGTCCGC	TACATCTTCGATCCCAAGA	TKKLRWTDEEKEILIESEIKIIKEGSLKEQHEINK			
				CAAGTGAGGTGGC	GGAACTGTGGGTTAAGCT	ILASRMPGRSQDGIAKIRQKQEHKAEIQRRLH			
				CGGGTATCTGCAG	TGAGCCGACGGAAAAAGC	GTVTTDETRGSRTSEITEPIRSLPINTKNWSED			
				CACTAGAGCCACTG	GAATGCATGTTAGACGAC	ELKRMLAEEVKLRNKNEKDINKKLAEIFPNRT			
				GTATCAAGAGCAG	GAGGTACAGTCACCTCCTC	MGSIKSKRTKDKEYQDLVKLTMQTISENPDN			
				AGATACGCGAGTG	GTGGTATTTGGCGGGCAA	ETDFNTSNTENNSTDAEKEVKNYLNMLLLSN			
				GAAGTTGAGTACG	TGCTCACTAAATTAACTGT	NEEEWLTSTLKEAATLALQGKKTEASEKLNEY			
				ACTACCTTCACGGG	GAGTAGCTGAGAACTGTA	ASKTLFPGLKITNQTRKREKKISKRETRRQEYA			
				GTCCTCCTGATAAC	TGTGTATCATGAAAAAAA	EIQKLYKKNISSAAEKAINGKWSIKPEEEYHNN			
				CACAGTGGACTGT	AA	KDLIKAWKPILEAPPFSDCRPIENIKEMDYAL			
				GGGAACTAAATGT		MEISTAEIFLAIRAMGKTAPGPDGIKYSKLKKN			
				GTGCTCAGCGTTCC		IQSMAILFNTCLLTSFLPLPLKIARTILIPKQENP			
				CTACTTTCTCGTAG		GILDYRPLTIASVVTRVFHSILAKKLDNNAQLS			
				GGTAAAGGGTAAG		QRQKGFRKCDGVAENIVILETILTNSRSEKRPL			
				ATAACCCAGAGAAT		CMAFVDLRKAFDSVGHESIIRGAKRVGVPPM			
				ATCCCATGGGAGA		LLEYISSSYQNASTNLFGEILNSRRGVRQGDPL			
				TATCCATGGAAAAA		SPILFNFVIDEALENLNRNIGYLLKEEKVSCLAF			
				GCACCACGTTAGAC		ADDIVLIAETKGGLENHIEKLLEKLNGAGLELN			
				AATCCGATGGTCTA		ASKCATLMVMKNGKEKSTYISTKVIKIKENDIP			
				ACTCGGCTCCGAG		TMKATETYKYLGLQMGFKAREQNANEVITE			
				GGGCTAACTATCCC		GLENITRAPLKPQQRIHILRDFLIPRLIHKLVLG			
				AAAGGGCTTAA		RVAKKSLKRIDQNIRKKVRNWLHLPKDTTAAF			
						IHADAGDGGLGVPALEHTIPLLKRERITNLRKS			
						NDPVTKECLRMEYTKQVLGKWSRPTKIGETL			
						ATNKSQLKEAFRKQMLITLDGKGLKDHHETP			
						TIHKWIRRGENMTGKQFITAVKIRGNLVATKS			
						RNSRGRPEQEKLCEAQCGRPDSLGHILQGC			
						WRTHGMRVERHNNICRRIKAIMKGKESEVV			
						EEPRLQTNEGLRKPDLLICHKGKIIICDAQVVA			
						DSSNCSLESENQRKIDYYKKDSVVAEARKLIGR			
						DDEDIIIMAVTFNWRGAISKTSIRDLDMLLDIK			
						SKEVIKMSRKIIRDNSIMVEMHRNRTEKRR			

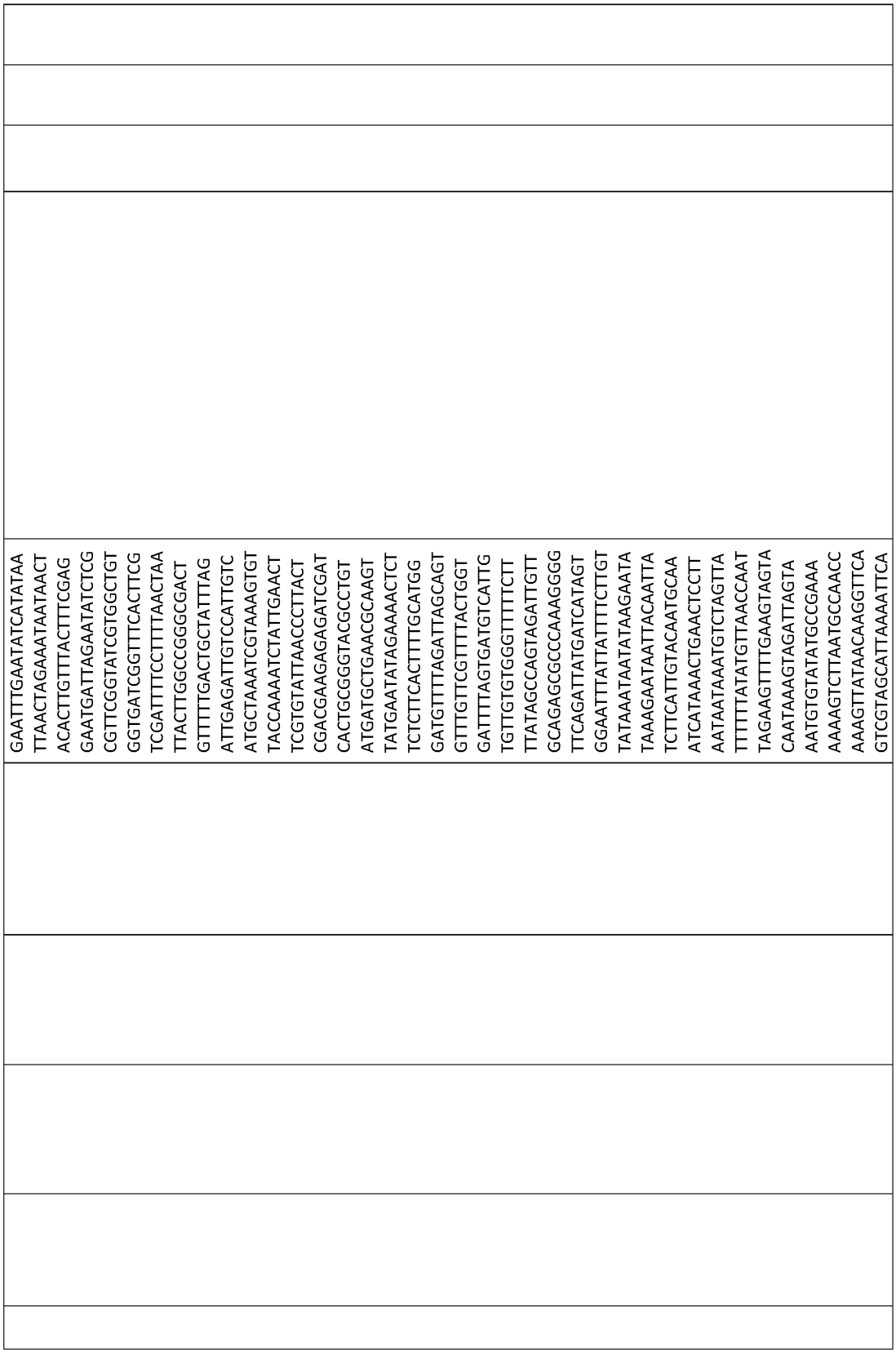
	Utopia-1_NVit	GCF_009193385.2	Nasonia vitripennis	CCATTCCTTCGTAC GGGTTTTCGTGCCG GCATAGCCGGGTG GGAGACTCGCGCG GGGGAGGTCATAT CTCACCACCATCTT CGAGCTTTTTGCTG CACCTGA	TAGTGGGGCCATAACACC TAGGCCCCACAGTGTGGC GATGTCCATGTGTGTCGTC CTTATTTATTTATTTATTTG TCCTGTCGAGCTGTTTAAT TTATTCACATGTATGTGTT GTGTGTGTCGTCCCTCGTA GCAGTTTTATTCCGTCCAA CCAGAGGTCGACCAAGAT TAAAATAAGCATGGCTTG AAGCAGGCCAAGCGCCAT GTTCTAACCCCGTTTTAGG GGAAGTTACTTAACCTAAA AATACAACTTTTCC	MSGLTGREVKCITVNVLMEQQPHTKAIREGD FIVILLPQSDDETLCCPLCVGRGRYSGKTRVEC LNRHVKEVHPDLTTTFRCWGCGFAAPGDKK YPRKIVTQHCATCVPEVSSAPSGRVDGERRV NTRRRLGIAAATEASPVRRTRRNGLASPPVEQ DISQSAAPPEPARVPQHPEIVALGESADDEVF RSPVNSPPRDWRAAAPQQAASSSPPAVPGIT AATPSNTTRTGNGSANSILAEHPIPTPPPTNT TEANGRADIPRSGRTPPPGQQAARRRAPTTE QRRIVGLLEAATGREQLEEATTQAMLFLARLT GRRPEPRNAIRPGPRQRHPAQEDVQAQAPD RIQEAKKLQRLYRTSKKRAVQKILAGPNMNC QIDKNTITAHFVELAARRDGGEDWPDVFDRE EPTAASGEALCTPITREEVFRRLKGRNNTSPG PDGITYRDLAKADPGAHVLAALYNSIWRIEAT PALWGVSNTTLIYKKGDAMDISNWRPISLGD TVPKLFAAILADRIKRWAVANGRYSASQKGFL EFEGCYEHNFVLQEAIREAKGGRKELVVAWL DLASAFTSVPHSSILQALEGHGLPSKARNIISSL YTGMTTRFHTAEGPTDPILIQSGVRQGCPLSP DVFNLTLEVVLREIQRTGEGYTIEGRRTSHLAY ADDVAILADSPAGMRRLLFAAERGARAVGLT FNPAKCATLHIAGRGEEAVRPTEFSVQGTPVR ALASEEAYEHLGIPTGYQVRQTPINTLRDLLA DIGSIDRSLLAPWQKLDAVGTFLLPRLDFIMQ GAHIDKGFLTEADKIIKKAAKSWLSLPQRASA ELVFLPPSQGGGGGLLTVAHSYKMLYSSDVTVS TIAGSTLRRTVSERLKKRASNIDIARFLSGDLDL PRSTSPSTFWTKVRSAALRIKTKLGLRWSWC QEGEVLLMACGDPRAPGTRVSPQTKHLVTTS LRRCLNRHYAESLLAKKDQGKVFEVTRRSGQ SNHFLRSGSFTRFCDWRFIHRARLDVLPLNAA KRWQRGMDKRCRRCGSDLETLPHVLSHCGP HSAARQKRHNNIQDRLVKAASRCPGTISVNQ TVVGVRGPDAALRPDIVVRDDVNRRVTIVDV AVPFENRLEAFDGVREAKIAKYTPLARQLTDS GYTVTVEAFVVGALGAWDPRNERVLSLLSISR	cd01650	477	740

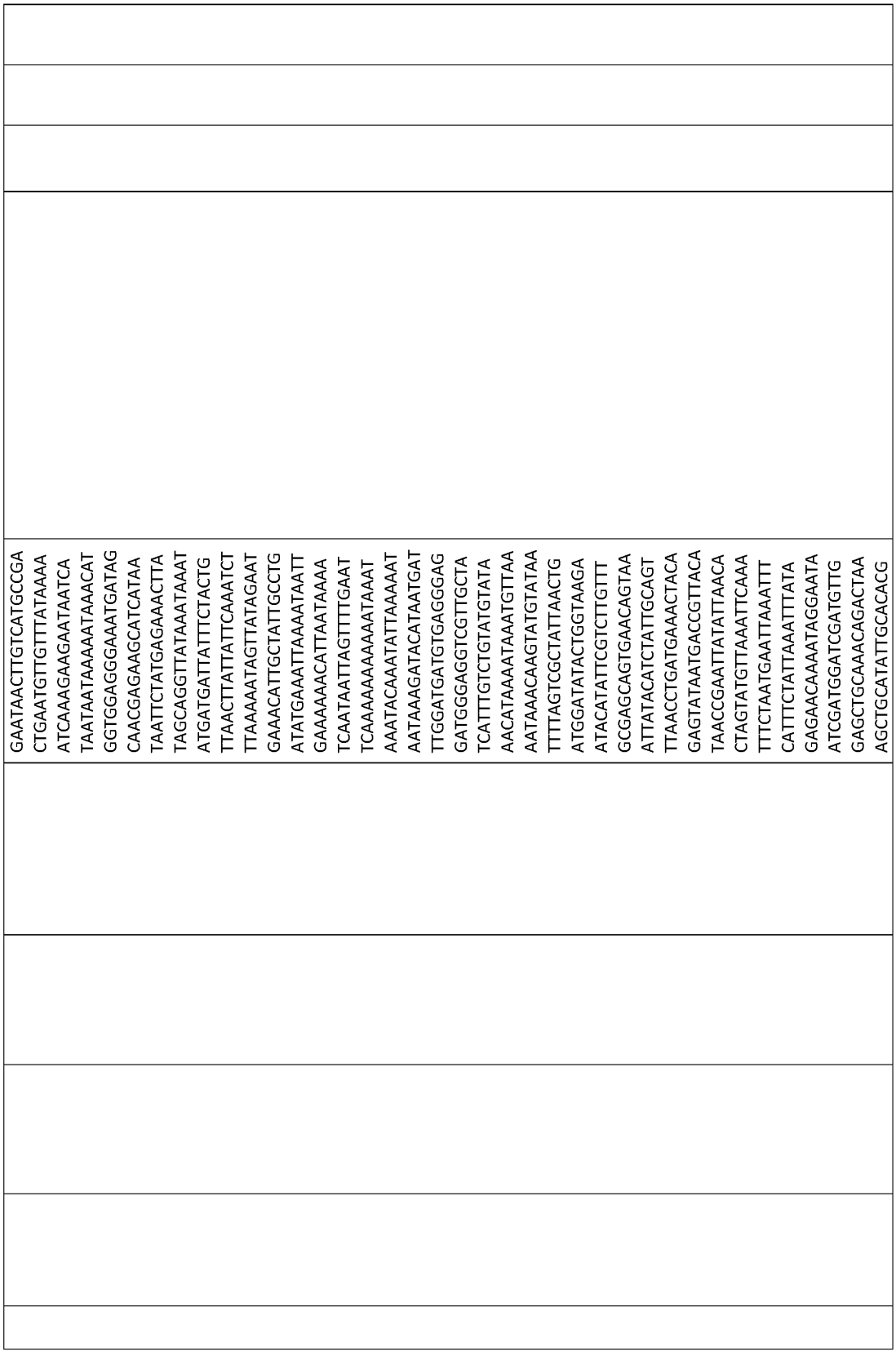
	LIN9_SM	GCA_002600895.1	Schmidtea mediterranea	AAACGACATCATGA ACGCTTGGCCGCA ACAATCCAGTTATC CCTGCGGTAACATT GTGGAACTCATAA GACAAGTACTAAA AGAAGAATTAGAA AAATTAGAAGAAA AAATTGAAAATAAT TTATTTATAAAATTT AAAAATTTAAATTT AAATTTAAATGAAG ATAAAAATTTATTT AATCCAATAAATAA TCAAGAAAATCAA GAAA	TAAAAGGGCAAAAAGATA TTTCAAGATGAATTGTGGA CTCATCTAAAAAATGACCA CCTCGAGCCCAAATAAGCC TAGCTATCATGGTTGCTGA TGGAAACAGTAAGGCACC TGATAGCTAACTCTTCACC GTGGGTGTCTTCAGACACT CACGGTGACAAGAAAGGA CACCACTAGTAAAAACCAC TAGTTTTCTGACACCTCTT GCTACAAACTCTGTAAAAA TCAAAAGGATCGATAGGC CGCGCTTTCACGGTCTGTA TTCGTACTGAAAATCAAGA TCAAGGAAGCTTTTCCCAT TTTAGTCAACACCAGGTTT CTGTCCTAGTTGAGCTTCC CTTGGGACATCTGCGTTAC CATTTGACAGATGTACCGC CCCAGTCAAACTCCCCACC TGTCCTCAAAACAGTTCAA TTGCATCCAAAGATCGCAA TTTTTTCACTAAAAAAAAT TAACAAAAGTTAATTATAC TGCTTCATTGAGTAAGTAG AAAAACAATC	YYAILMRRLMVSDTIRWSRDIYVEHVSGIRQY RE	MMDSRQLNTPKIRKYQNPKMTNDIMKSYNY AVLSDVTPQETTQQTTHLNVDIDKETTQPKQ PLTKSGKPKSKPIAVSYKFKDATFIWDTTPQT NPPKDCTKLIDKTRPRKTIFKKSAFQTYLKKELS NETFVEVKTFLMATHKYRFKDENSRLLAYRIIN RYVMETANEFKETEFDMARFAKFFTIPENWL KHLKPYSTATETSPADRIKVQKLVDLTCRYPFK TQEEQTSVANFLHFFTQRSIIGISRDYKFQKCI PFMARKTNGQETTSTINTTPPTEENRLPMVII TPLEEPKIEHRRPEKRGASNDTIVLSDEEFPLLK RRTLPTRKSKNPTGAGNVPTETECTDEVKFIL NNEYQIECKECGKVWWENVRNGMNHLRQKH DFPNRTDVMVSCVRCEVPIKGAECVNHIKNH KKDDKEESQAGSLVANTQDIPNESSLSQAAIE VYLRNILKMKENQERNIQYLEPSTANFLINRN LRAFYQNVKIEKLIGWEQVIWLIHWNKCHWI VYLANCDSKTSVILDSDNQMTLQQRCNIKAK FDKFLEGTFEEKAVLGTLERNVPQQPNNFDC GVYVIQYISDFLKDPQRMDYHTPDSKRIRKEI GELILEEMKNPASKIKNPNKEIQSLLQKFRLLQI NVNDVFHWFAEEYQKSLPKIRTKRDGKLNKL SCSYQIQRLFGLAPKRAVKEIYFQETSTADLET RVLNEHFKKDESMMKECKIKNGNHYQDWIT KAQIDNKEILEALKNSTDSAPGEDNIPLRQWII WNNDGVLFDMFNYIKRTHDIPEMWKNYTT TLLIKPGKSQESNIPANWRPISILPTSYRIFMKV LNKRVLDWANRGELISKWQKAVDKANGCDE HSYVIQALIEKANRSYYKNEQCNLAFLDLADA FGSIPFQVIWHTLKNMGMDEKTINLLKEIYKD CSTKYKCGKNESEKIKITKGVRQGCPLSMTLFS LCIQYLIQGIAEKKKGATIAGQEVCILAYADDL VIVANTAKDMQMLLTTIENLAKQADLIFKPAK CGYYRDPRDKKSMMKIYGKEISIVDEKNVYTY LGVRIGDTKKKDLNVRFEEVKKKTTAIFKSKLR SDQKLEAYNIFCQSKFVYILQGEDIAKTKIETY DEEIKKMIKEDILKLQDKSPFTDFVIYSPREKG	cd01650	793	1057

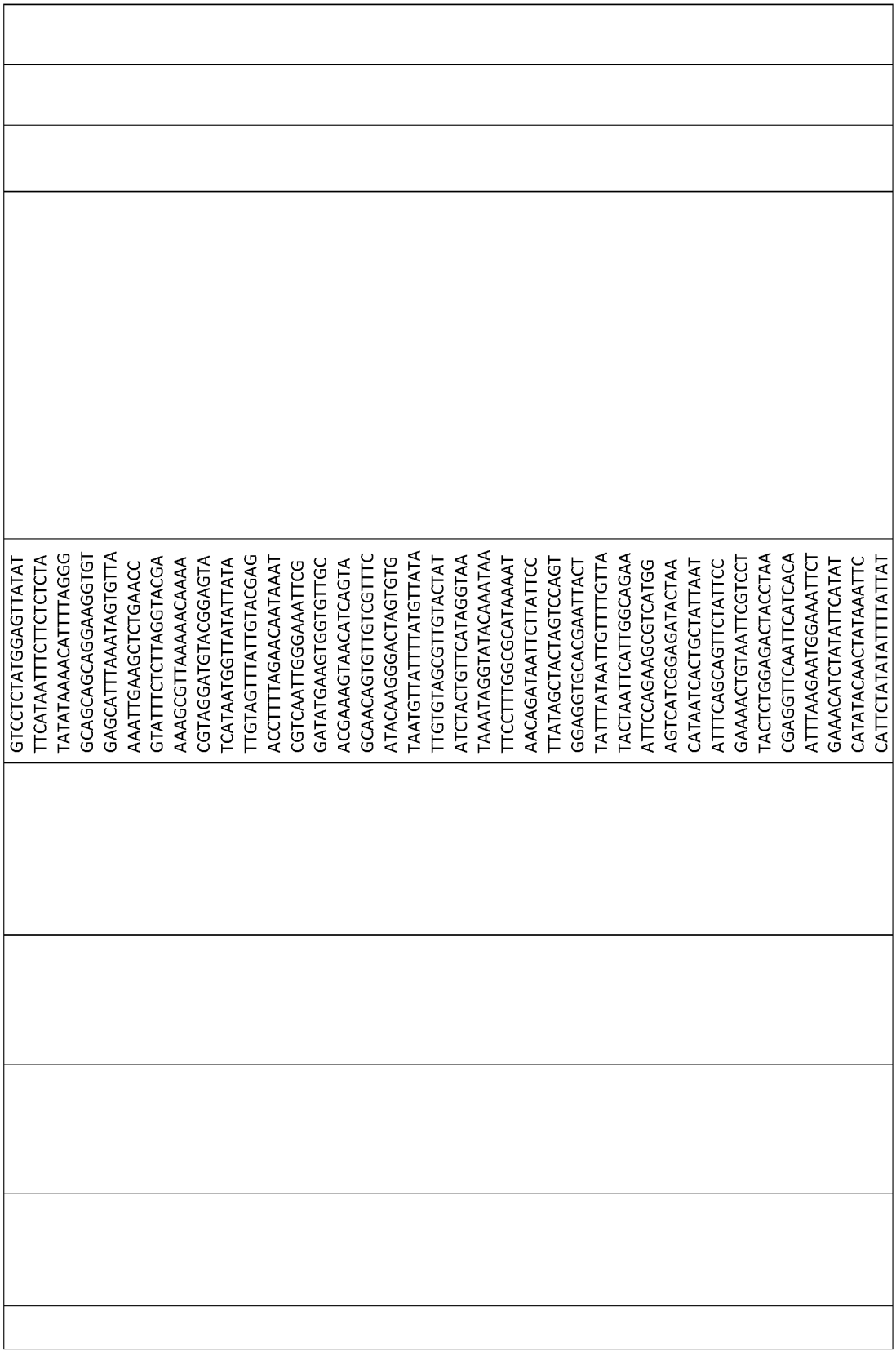
	R2-1_TG	GCA_012069535.1	Taeniopygia guttata	GTCTAGTTACAACT GGGCATCGCTGCA GAGATCGCACCTCC TCGTGGTCCCGCTG GTAGCCCTTCGAAG GGTGACTAAGTCG ATCTCTGCCCCAGG TACGGAGCCGTTG GGACTCACCAGTCC AACGTAACTCCTGC CTAAACTCGGTGAA ACAAATTCCTCGGT AAAAAGCCCC	TAATTCAGGTTATTTAGAT GCTTAGTTTATGTACCTTT CTTGATTTGTCTAGGATTT TGATAGTGTTATTATTTTT ATATTTTTGTACGATTGCA TAATGTTCTTTTTTATACAG CTCTGTTTTAATAAAATAG ACGATACCTAGAGACGTT AGGGCAGCCACAAGCCAG TTAGGTAGCGGATAGTAG GTAGGAACAGACTTTTACT ATTTCATAACGCGTCAATT ACCACCTGATTTGGACCAA TTCACGGGATTTGTCCAAG GTGGACGGGCCACCTTTA CTTAACCAGGAAAAGGAA CATATATAATTTATGTGTG TTCGATAAA	GLGITKIIDEQIIQTINRTAKLLNSSHRAIRAIIYE ELIQVANLRGEKEINSIEEALKWLEGTNKYKK NSNAKTTWITRVREAFQTLEKKHKIKVRFVPK ENGIGYKIKCDTQEKVEELDNSKELSKSLHW MIKEAYFKEWKTLKCQGYILSLKTSEFMEWK MPRGLPDPDWRFLTKVRANMLDINMKQAN QGGRLGNTKCRKCEDKESASHVINHCASGN WSRVEKHNQVQNELAKELTKRNISFEKDCIPK ETKGTLRPDLVIRLKDKIMIVDIKCPFDEESAIE SARNKNIDKYRELAKEIQAKTVLPTTISTFVVC SLGTWDKRNNELLRQMGIRYEESKEMRINMI HKAIHGSRKTYDQHRNFNNG	MASCPKPGPPVSAGAMSSESGLTTHSVLAIE RGPNSLANSGSDFGGGGLGLPVCLLRVSVGT QTSRSDWVDLVSWSHPGPASKSQRVDLVSL FPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAE NDLALLVNLEFYRSDLHVYECVHFAADWEGL SGLPEVYEQLAPQPCVGETLHSSLPRDSEPFV PEEGSSEKESEDAPKTSPPTPGKHGLEQTGEE KVMVTVPDKNPPCPCCGTRVNSVLNLIEHLK VSHGKRGVCFRCAKCGKENSNYHSVVCHFPK CRGPETEKAPAGEWICEVCNRDFTTKIGLGQ HKRLAHPAVRNQERIVASQPKETSNRGAHKR CWTKEEEELLIRLEAQFEGNKNINKLIAEHITT KTAKQISDKTRLLSRKPAEETREEPGVCHRTR RAAASLRTEPEMSHHAQAEARDNGPGRRPL PGRAAAGGRTMDEIRRHPDKGNGQQRPTK QKSEEQLQAYYKKTLEERLSAGALNTFPRAFK QVMEGRDIQLVINQTAQDCFGCLESISQIRTA TRDKKDTVTREKRPKKPFQKWMKDRAIKKG NYLRFQRLFYLDRGKLAKIILDDIECLSCDIPPS EIYSVFKTRWETTGNFKSLGDFKTYGKADNTA FRELITAKEIEKNVQEMSKGSAPGPDGITLGD VVKMDPEFSRTMEIFNLWLTTGKIPDMVRG CRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFS RIVTARLSKACPLNPRQRGFIRAAGCSENLKLL QTIIWSAKREHRPLGVVFVDIAKAFDTVSHQ HIIHALQQREVDPHIVGLVSNMYENISTYITTK	cd01650	690	956

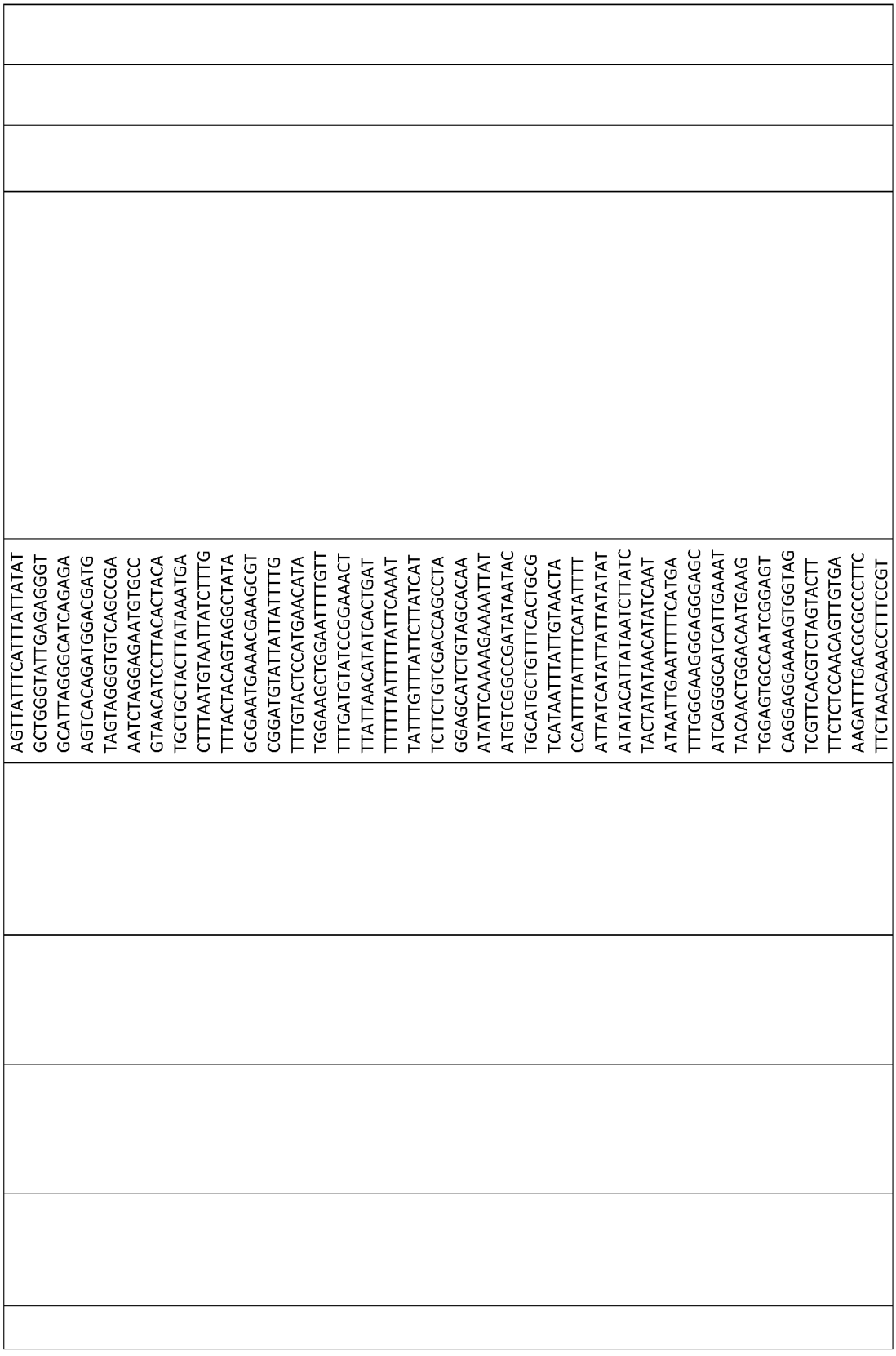
	ORTE-1_AAe	GCF_002204515.2	Aedes aegypti	ACGGATTAATGGT GTGAATGTGAACG GGAGGTTTCGAAA GACAAATGCATGA ACGGAGCTAGCGA GTGATAAATCGGG AGCGGGGAGACGA AGTGAGAGGTTGG TGCGAGAGAGACG AAGTGAAAGGTGG GTGATAGTAAAGT AGGTGAGATGTGT GAGAGGTAAATAT ATAGACACTGGAG TGAAGTGACCATTG GATTATTAGTAGTA GTGTAGGTCAATCA AGAAGTGGTGAAG GTGGTAGTGACGA	TAGTAAATAATTGAAGAA GTATCAAAAAAAAAAAAA AACATACACAAAAACCTAC AAAATTTACACACACGCAT AAACATATTGGAAAATAA CACATATGAGTATGCACAA ATACGAATTATACATATAA AAAGATACGAAAACAAAT ATACATACACAAAAACATA TGATAACTCACACGAAACA CTAATACACTCAATCACTC ATAAACACATACAAAATAA TCAATAATGAAAATTACTC TGAAAATACACAAAACAT AGATACTAAAATAAACAC ACACATAAATTACACATAA ACACATATACAGTATTGAA CAGAACATTTGCAACTTTT TCGGTTTTCCATACGGAAT	RNTHTDKIQIRVGVKQGDPMSPLLFNLAMD PLLCKLEESGKGYHRGQSSITAMAFADDLVLL SDSWENMNTNISILETFCNLTGLQTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPG ESEKYLGLQFDPWIGIARSGLSTKLDFWLQRI DQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTA LLETLDQKIRTAVKEWLHLPPCTCDAILYSSTR DGGLGITKLAGLIPSVQARRLHRIAQSSDDTM KCFMEKEKMEQLHKKLWIQAGGDRENIPSI WEAPPSSEPPNNVSTNSEWEAPTQKDKFPK PCNWRKNEFKKWTKLASQGRGIANFERDKIS NHWIQYYRSIPHRKLLTALQLRANVYPTREFL ARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHI RDSNKELYKPDLIFVKDARALVVDVTVRYEAA KSSLEEAAAEKVRKYKHLETEVRQLTNAKDVT FVGFPLGARGKWHQDNFKLLTELGLSKSRQV KMAETFSTVALFSSVDIVHMFASRARQSMV M	MDLFNFDINDFQTAVVQKIPAGSFITAIMLQI NKINSEFKNHQLNEMEFCCLIADFIRCNDHLF SKMWEEDIKEVVHKIFSMESGVEVHMAQAI VELFQTKICFYQKKREVEVIQPNSCSNTNMLT EINILFTFEDGTEKFYSIQSITDKKSSGKKESGY NEQPRNELYKVKEYESGREMMVKKIKGDGN CFIRAIIDQMNKSEDYYRCNLSDIIELRNKIAN HMNKNRERYEMYLIESLEENGFLKFIEDIRKP GVYLGHESMVAVTEIFDKCILLYQTDSETLRIG EEKDREEDVLRLLYTGENGVKNHFDSIRIVKG QTEENTGTEKTSVEEVKEKESYTSMHGKHND DSEEISTIRYRKTQDTPITGKQIKETGKLVQEE GKQWQWNKERNNQKKSDDKKAVEKNMDE RGNETRKITREKHKPINKDRAITKDFITDNRKE EILDLEEKRRRQRSKEREDRLEQLATQIKTNEE RSMLINGSQRQGWEENDSQERLDREKQAKT DKNTRGRIQMDVQGKTEEPLSIVNDEQVRKE NNTKDVKLGDKKENPRISTKQRNDRKRAKNE RVDEEGVQENSDTMDGGTKKHRNHVAEER	cd01650	1122	1384

		TGACGGTTAGGAT	GGCCGGTTTTGGTAGGTTT	GETRKGQVIESEESRATGVKATGEIGEENEEV	
		GCCATAGATAATAT	GATCTCGGTTGTTTGTGGG	KESIMLIATLNIRGCSKLEKRQEIDNMLESFNI	
		ATAAAGCGACTTG	CCGAATTGAGTGAGATTTT	DVAALQEVNVKAELITTRNYEWRIGGQGQN	
		GTGGTGTTGAAGA	CACGGAGCATTGGATTTC	KSRGLAILIRKDKGIDVMEWKEIGKYGLSISLLS	
		AATGAGAAGGGTG	GGCACGTGTTTTTGAATG	KGRKIILINVHGPNKNGYIFYSSLGKLISQDHLR	
		TAGAGAAGAGCGG	GTTTTTCCGATTATGGGTT	SNLIMVGDWNAQIGKDSVMPEDAECIGNRL	
		AAAAGAAAGGTGA	TGGATGTAGTGGCGGTTT	GFSNNNDNGEEFKMFLTIHKIKNMSTVIGKN	
		ATTGAAACACAGG	GACTGAAGTGAAATTTGA	TDITWKSGKKESQIDHVLKPREGKIEIRFIKGY	
		GGTGTTCAACTGAA	AATTTAATACCCCTACACA	WTGINTDHKMIVTGIRTREKENKEKGKKKIPI	
		TTGGGCTACGTGAC	AAAACTGTCTTCAGCAAAC	DASVLKYDTIKEKYQEALKKYEIRQEKGNTIEQ	
		AGGATTGGAGCGA	TTGTTCATAGCGTCAAAAT	DFKAVAGKMKKAANEVLRSARAPLTPIRKAA	
		TAAGAGAAAAATA	AGACAACTTTGCTGAAAAT	LNKLKTAINLSNKHPDIFPYRWKLKDRRMEF	
		AGTGAAAGGTATT	ATCATCAAGCTATTCTGTT	QAAIRSHNERKIKKFYKELNDFDVAVRIKKSY	
		AATTATTTAAAACT	AACGTGTTGAGTTATGTCA	QFLKGFLKRKKHKNVYIPMSQWNEALKESEG	
		AAGGTAAGACGAC	ATAAAAAAAAAATCGTAA	PEIQLMDEHDTCPVTVPPKEEEMMQIVQQQ	
		TGATGACTGGAAT	TAAGGGTCGCTTTGGTCA	CNGKSAGADGIRMELLKYADKQTQQELADI	
		GCATTCTAATACAC	AACCATTAATGCTTTTGAA	WKRVWIENHLPEDMEKTIQVPIPKNAKAKG	
		TGATGAATTAAGGT	CTCGTGGTTGGGGGGATC	VQDFRRISLCNVAYKPYAKWIKQRLREFTGEP	
		AAGACTAGATATAA	AGTCGGAGGCATGTGTTG	DINQAAFTEGRSTDDHMFVTRRILEEYWNA	
		GCAAATATATATTA	AGGTTCGGGTTGTGCCTG	GNPLYVAALDISKAFDNVSLLQLRSILAKLNVP	
		TGAGTTTTCTTTTTC	GTGTTCTGTGGGGGTTTCA	SHLIDRVLQCIRKEKIRVRWQNQYTTECNRGK	
		TTTATCAATTATAG	TTCGGATCGGTCCGTGGA	GVKQGCPLSPFLFNIVMQDVVRKVKEKVPEL	
		CATAACATTTATAA	TAGCTGGGGTCTGGCTTG	NLMNTGSLKLPLILIFADDILIITKDEEGLKRILIA	
		CTACACAAAGAGTA	CCGGGGTTGGTCGTTTCGT	LEESLREVGLEINNDKSQILIRAPNATKRPPNTI	
		CAAA	GTGGGGACCGGGGAGGT	NLNGRNYDVKTSMRYLGAWLTDTLNRPATT	
			TGCGAGTGTTTTGTCAAAT	KQRCVNAAKVSKLVVEFCKKFKPTWEIGKLIY	
			ACTGTAAATACATACAAGA	NTVIAPSILYGTKTATLTKRSRVQLGKYEKQILK	
			CAAAATAACACAAGCACA	DIWNQCRKTDERKFNVRKELKGKTINRRVRV	
			TGTATATATATGAGAAC	NRISYYGHIQRREETHPIKMAMKLKFNKKKH	
			ACATGCAGAAATAGACAT	GRPSFTWKKSLEQDFNRYNGITEEEWSQIAQ	
			TGACTTATAACACAAATTA	DRDKIKKKAEEIYRNNESEISDGEDSTEQDM	
			AAGCTGTTTTAAAAAGATA		
			ATTACTCAATTTTAATAAT		
			AAACACATAGCATTAGTAC		
			AAGAATGACATGCAGTCA		
			TCAGTCGTTTGAGTTGGTT		
			AAAAGAAAATAAAATTGT		
			TCGATTTAAAAAATGCAAG		


GAATTTGAATATCATATAA
TTAACTAGAAATAATAACT
ACACTTGTTTACTTTCGAG
GAATGATTAGAATATCTCG
CGTTCGGTATCGTGGCTGT
GGTGATCGGTTTCACTTCG
TCGATTTTCCTTTTAACTAA
TTACTTGGCCGGGCGACT
GTTTTTGACTGCTATTTAG
ATTGAGATTGTCCATTGTC
ATGCTAAATCGTAAAGTGT
TACCAAAATCTATTGAACT
TCGTGTATTAACCCTTACT
CGACGAAGAGAGATCGAT
CACTGCGGGTACGCCTGT
ATGATGCTGAACGCAAGT
TATGAATATAGAAAACTCT
TCTCTTCACTTTTGCATGG
GATGTTTTAGATTAGCAGT
GTTTGTTCGTTTTTACTGGT
GATTTTAGTGATGTCATTG
TGTTGTGTGGGTTTTTCTT
TTATAGCCAGTAGATTGTT
GCAGAGCGCCCAAAGGGG
TTCAGATTATGATCATAGT
GGAATTTATTATTTTCTTGT
TATAAATAATATAAGAATA
TAAAGAATAATTACAATTA
TCTTCATTGTACAATGCAA
ATCATAAACTGAACTCCTT
AATAATAAATGTCTAGTTA
TTTTTTATATGTTAACCAAT
TAGAAGTTTTGAAGTAGTA
CAATAAAGTAGATTAGTA
AATGTGTATATGCCGAAA
AAAAGTCTTAATGCCAACC
AAAGTTATAACAAGGTTCA
GTCGTAGCATTAAAATTCA


GAATAACTTGTCATGCCGA
CTGAATGTTGTTTATAAAA
ATCAAAGAAGAATAATCA
TAATAATAAAAATAAACAT
GGTGGAGGGAAATGATAG
CAACGAGAAGCATCATAA
TAATTCTATGAGAAACTTA
TAGCAGGTTATAAATAAAT
ATGATGATTATTTCTACTG
TTAACTTATTATTCAAATCT
TTAAAAATAGTTATAGAAT
GAAACATTGCTATTGCCTG
ATATGAAATTAAAATAATT
GAAAAAACATTAATAAAA
TCAATAATTAGTTTTGAAT
TCAAAAAAAAAAATAAAT
AAATACAAATATTAAAAAT
AATAAAGATACATAATGAT
TTGGATGATGTGAGGGAG
GATGGGAGGTCGTTGCTA
TCATTTGTCTGTATGTATA
AACATAAAATAAATGTTAA
AATAAACAAGTATGTATAA
TTTTAGTCGCTATTAACTG
ATGGATATACTGGTAAGA
ATACATATTCGTCTTGTTT
GCGAGCAGTGAACAGTAA
ATTATACATCTATTGCAGT
TTAACCTGATGAAACTACA
GAGTATAATGACCGTTACA
TAACCGAATTATATTAACA
CTAGTATGTTAAATTCAAA
TTTCTAATGAATTAAATTT
CATTTCTATTAAATTTATA
GAGAACAAAATAGGAATA
ATCGATGGATCGATGTTG
GAGCTGCAAACAGACTAA
AGCTGCATATTGCACACG


GTCCTCTATGGAGTTATAT
TTCATAATTTCTTCTCTCTA
TATATAAAACATTTTAGGG
GCAGCAGCAGGAAGGTGT
GAGCATTTAAATAGTGTTA
AAATTGAAGCTCTGAACC
GTATTTCTCTTAGGTACGA
AAAGCGTTAAAAACAAAA
CGTAGGATGTACGGAGTA
TCATAATGGTTATATTATA
TTGTAGTTTATTGTACGAG
ACCTTTTAGAACAATAAAT
CGTCAATTGGGAAATTCG
GATATGAAGTGGTGTTGC
ACGAAAGTAACATCAGTA
GCAACAGTGTTGTCGTTTC
ATACAAGGGACTAGTGTG
TAATGTTATTTTATGTTATA
TTGTGTAGCGTTGTACTAT
ATCTACTGTTCATAGGTAA
TAAATAGGTATACAAATAA
TTCCTTTGGCGCATAAAAT
AACAGATAATTCTTATTCC
TTATAGCTACTAGTCCAGT
GGAGGTGCACGAATTACT
TATTTATAATTGTTTTGTTA
TACTAATTCATTGGCAGAA
ATTCCAGAAGCGTCATGG
AGTCATCGGAGATACTAA
CATAATCACTGCTATTAAT
ATTTCAGCAGTTCTATTCC
GAAAACTGTAATTCGTCCT
TACTCTGGAGACTACCTAA
CGAGGTTCAATTCATCACA
ATTTAAGAATGGAAATTCT
GAAACATCTATATTCATAT
CATATACAACTATAAATTC
CATTCTATATATTTTATTAT


AGTTATTTCATTTATTATTATAT
GCTGGGTATTGAGAGGGT
GCATTAGGGCATCAGAGA
AGTCACAGATGGACGATG
TAGTAGGGTGTCAGCCGA
AATCTAGGAGAATGTGCC
GTAACATCCTTACACTACA
TGCTGCTACTTATAAATGA
CTTTAATGTAATTATCTTTG
TTTACTACAGTAGGCTATA
GCGAATGAAACGAAGCGT
CGGATGTATTATTATTTTG
TTTGTACTCCATGAACATA
TGGAAGCTGGAATTTTGTT
TTTGATGTATCCGGAAACT
TTATTAACATATCACTGAT
TTTTTTATTTTTATTCAAAT
TATTTGTTTATTCTTATCAT
TCTTCTGTCGACCAGCCTA
GGAGCATCTGTAGCACAA
ATATTCAAAAGAAAAATTAT
ATGTCGGCCGATATAATAC
TGCATGCTGTTTCACTGCG
TCATAATTTATTGTAACTA
CCATTTTATTTTCATATTTT
ATTATCATATTATTATATAT
ATATACATTATAATCTTATC
TACTATATAACATATCAAT
ATAATTGAATTTTTTCATGA
TTTGGGAAGGGAGGGAGC
ATCAGGGCATCATTGAAAT
TACAACTGGACAATGAAG
TGGAGTGCCAATCGGAGT
CAGGAGGAAAAGTGGTAG
TCGTTCACGTCTAGTACTT
TTCTCTCCAACAGTTGTGA
AAGATTTGACGCGCCCCTTC
TTCTAACAAACCTTTCCGT

							871
						674	
					SSF5 6672		
				GGAAAGCTCGACAAGTAA TGCAAATTCTAATACTCAA TCTGTTGGATCATATGGCT GGAAGGAGATTCTCTATTT CCCGCGGGTTTCGCGTAA GTGTTTCTACTTACGGGTC CGCCACCCCCCTTTGGCC CTTCATACAGCGATATGTT GAATGCAGATTGGTAATA AAATTTGACCTTACTGTTG AGTGGAGCCGCATGCAGA GCAAGACTCAAGCAAGGA TTGGTGGCTACCTACATAC ATACATACA	MPLFNRPPQCRQLCPNPLLSANNATIQVLDS PVSTHNETSRSTQLDAHRSNFVGGSQIAKVE SSDLDRSSPFNGVFYRIETGKIIAQPSSSPLPPS SSPLLPSSSPLLPLSSSSSSTPLPLSSSSSSPPPSS SSSPLLPLSSSSSPPPPSSSSSPLLPLSSSSSSTPL PLSSSSSPPPPSSSSPRPHPGPRKKRNKQKKRT KPRKLPPPTSSQPLPHTTEGLNPVFEDLFCPVS GCPGPSSIGTDEWILCNACLIWYHTSCLFLQAI PEGDWFCANCPQIDTQIGPQINSNTSSPANQ SYLGTHIQLTQSIPFTNPRASQPISTLSTPLSNFI SSKETCTSLTLEWLNNHIVPGGVIEKAILHDAY RRDSLGIFKPINDAEMGKLIKNSFSNIRTKRIRT ELHGTSKYHYVGIQFKSEHSADSVFDFQTIISS LRQNCPTLKQCPKACRIPVAQVLKSAVDKCIS ENTPEAWHLLLTFPYITLQCESKANIKNGQSL NSKIRNNLSIFESLDKNTYQDYLITLSSKNSAH DKPGSTIKDIGPKVISKVSEGNIRGAIRLLTSDD IIAPNSSETLETLVEKHPPQPLDNFTVLDEPNIP PFTTTVEQVISAIFSFPSGSAGGLDALRPQILK DLVSQHTGDVGTQLLESMTKFCNFMLAGKV PLDICPFIYGASLTALKKKSGGIRPIAVGCTYRR IVSKICVKVMNHVFQDYFSPVQVGFNVKNG AEAGAHAARRFFSFPHKSPTAFLKIDFKNAFN MVDRGKLLSSVHELVPTLYPYINQCYRFSSSLV			
				TAAGTGTATTTCATACTTT ATCAATGTTTATGTTAATA AATAATT			
				TGCATCATTGCCAA A			
		Folsomia candida					
		GCF_0022171 75.1					
		CRE- 1B_FoCa					

						WNSSLIESCRGVQQGDPLGPPLFCLILHPVVSNLGSEFNMWYLDDGTLGGDPTTVLDDFKHIISCANEVGLELNFDKCEVSVLGLQTNTQCILESFQSVAPGISCMTQENAELLGAPLTYESIEHVFRKKINAFSNMAGRLNSIPNHIAYYLLKHSLAIPRLVYLLRAAPIFKSPHLLEEFDSLLHDTLEQITNCQINPLSLEEIALPVKLGGLGIKQSKTLATPCFLASSFATTDLVSLILPSYVTALSDSLVTEAQDLWKSISCDELPRSTISSIQSSWEIPIFNQVCSRILDSSTTPINTARILASRSKESGAWLQALPSRNLGTLLDDNSFRIAIGLRLGLKLCFPHKCVCGTQIDELGLHGLSCKKSAGRHSRHTQCNDIIKRALISADIPSIREPLGCCRTDGKRPDGLTLIPYKQGRSLIWDFTCTDTFAPSYLSHTLRHAGTAAKLAETKKHKLYSELEKQYIFVPVAVETSGVWGMEGLKFIQDVGRRMKDVTGPMSTTFLIQRISLAVQRGNAASIMGAVPSQRGLEELFYILNDRSFYLQKSCKGPNGDLGRLSGSSKGFPNPPTSGFSSTPQAN			
	L2-22_NV	GCF_000209225.1	Nematostella vectensis	AAACGGGAGATCTAATCTCATCCTTGGCCTCAGTGAGACGTTTTTGGACGAATCATGGACGGATACCGAGTTGGCTATAGATGGTTACTCGCTGGTTAGAGCTGATCGCACATACAACTCTGGTGACGGGCTGCTAGTTTTCATACCCAGCCATATTTCCTTCAAAAGGCGCGTGGATCTTGAGTTTGAAGGTATTGAAGCCATTTGGCTTGAGCTATTATTCCACCGATCGAGCCCGTGCTTTGCTTGTTTTATTTACCGT	TAAAACCGGATATCCATTCTCATGTTAATTGTTATTGCATGTTAATTAATGTTTGTCTGTTTGCACTTGTTTCTATGTTGACGTCTGCCTGTTTTTGTAGTTAGTTTTGTCGTTAGTTTTTAGTTGCTTTATCAAGATGTAGTATTAGTGATGTCTTGTATAGTTTAATTTCCAAGTTTCCCCTTTGTATTTCCGTAAATGTCTAACATAGTTTTTTTTATTATTATCATAGGTACTTAGATATAGGGAGGGCCATATATTATAAAACTGTAAAAGGTTAATATGAAACCCTCGTTAAATAAAGTTTATTGTATTGTATTG	MFSHASSAISQQVMVLGDLNFDLLERPLPAPTKQYLSIFDSLGYQQLITKSTRPISNSLLDHIFVSSSDFVIEADTISNTASDHLPIFVAWKLKAGYTRPTGHKVLRFRSRKNFCIESFMHDLSLVPWSTLDIYDDVNDALDQWYKLFNDILTVHAPIKEHRIKRPLKLDWFNAEISAAIKARNCLHRIATSLKTDTSWSEYRRARNHVVHLIRKAKKTFYNDAIKNNLENPKNLWKIIRNISPAKHVNLPNRLTVNDTLYEDTFDIANFFNEHFSNISSQVCNDAVPEFPDRSVLTNFINFKHQPGSYFSIPPISEDFVLNSPSRLDFGKAAGLDELKVFFLKSAAPEIFRSLTKILNLSISSCVFPDPWKSAKVPPLFKDGSLFDRTNFRPISVLCILSKVLERHIHTSFYYFLTSQDLLTDFQFGFRWFRSCELATINLSDSILKNMDSGLPSGMLLIDLKKAFALVDHNTLLLKLGLYGCDELTLKWFKSYLSGRTQKTCFKGVLSNALPINVGVPQGSILGPLFFLLFINDLPLFLSSDLVTDTTMFADDTTFLISGASVPDIELKLNKLATDVHKWATLNRMVLNTKKTKSMLIGSHQKVSNLAESSLSVTLN	cd01650	374	636

				CGTCCTTCTTCTACC TCTGCATACATAGC TCTTTTGGACGAA		TGAGATGTCCATATTGAAT TACTGAAATGTAATGAGTT AATGTTAGTATTAAGTAGA AAATAAGAAACATTGTTA GATTATATGCTTGCTCCCT CGAAATATTTTACTTTTCG AGTGGCTGGACACAGTTA ATTATGTAACTTTTAAAGC TGCGTTGTTAAATAAAGCA AAAAAAATAAAA	SCPIEQVNIARVLGVVFDNHFTCEPHIELFCKY INSRISLLRSIAPFLTPAGAIHYYNACIHSQFPY CASVWGNCSQTLQLLLLRSQKRAARTILSADF SVPSVSLFSRLNWLYICDLIKQRKLWTLFSILIN PDCPLCLRSKFSFVSHERRTRSSTPNNLVLPQP RTDAGKRTFSFSAISLFNLLLPSLKQLAFSRFLN LNQLPNVTSLAKKKVRDIFHSHLNSVSHLEQL FCHACRFYIDCRCYFS			
	Crack-5_AAe	GCF_002204515.2	Aedes aegypti	GGTATTCCAACGG GAATGTCTTCTTAC GTAAGGATGAAAG TTCTAAGGCAATCC ATGTGAGTGATGTT AGTGTTCTATCATC GATTGTGTAATATT TCCTGATTAAGTTA A			MITLIKINYKAELINELNINFCLEEIKLDKENLNI LYLNTRSCRNKEEELYQLINEFDGSVHVVVFA ETWLYKNEIFNLLGYDAYHGCRDGRGGGVSI FVLSDYRSQLMLVINSEDNNFLMVELIDQNV KIMGVYNPGRSFNLFFEQLEQILHNYKRIIVCF DFNINLLDSSGELIRQYTYKIQSLGYIILNKLDA RYATRISNTIATVIAHIITDIFEYDLQFGIKDTER LISDHRTLTLSMRIENHKKNKPDHKKICVNYE NIIRSTSQNTFEICQQFSDIVVVLNEIMTKNRT EILSRKKAMVIRKPYINNALIAEIQNKNELYRQ YRTAPGGVLKDELYILYARKRNMLKNKIKFAK EQYYGRKMHKNINNYKKTWSTLREIIFQNGS ENITPQALTLKENGKIIDSAAQIAKCLNEYFINI GWNIQQNGTGNDNFLIYLNNMPETRFSFSV VETEKAKNVINSLKSGVAVGLNKISTKCIQKCV TLIAPKLTEIVNNVILSSQFEDCLKIAKIVPLYKS GDKFSKSNYRPVSIIPALSKIPECIMNEQIKHHL LQNELLNKNQFGFLPKKSTETASLELLNFIVKG LDDGEYVGCIFIDLRKAFDCISHELLIAKLSFYG FTPSAVSLLKSYLANRKQQCQVNGVLSEPLLIK SGVPQGSIIGPILFNIFINDLFNIPLKGKLQCYA DDAVLKYRAKDLETLQSMMQHDLQLLPSWF SANKMQINNKKTNFIIFSLKTSNVSLDLFINNE RLSQVTEVDYLGLAIDKNLNWKSHITLVKNKI TPFIFALRRVRGCLSTKSCWQLYNAFILPHLSY MNTVWGSAAASHLNVVLKVLQNRVIKTIRKLP FLFPSASLYSPEVLPINLLHKYNCLILVYKIKYKL VESNIRLTEFHDLHSHSTRARQNFYVSRPRTE LSLKHFIYSGITQFNALPNHLKLEIHLSSYKKKL KKYLFENN	cd01650	515	769

	CR1-54_AAe	GCF_002204515.2	Aedes aegypti	A	TAAGTAAATCAGTCTGTAGGAAATTTATCCAAGACGTTGCAAATAAATAAATAAATAAA	MWLRPDVLNSELCSNYTIFRCDRSEQTSSFQRGGGVLIAVNASFRCTSVSLNDCDSLEQVVVSVNLPMSTIYLCGIYIRPTSPPEAYSMHATAIQSICNFSSETDAIIVVGDYNLPGLSWISDEELNSYLPCNASTEAETTFVEAVMATGLYHVNSIRNVNNRILDLAFVSDVCDIEVLEAPSYLLRMDAHHKSFVLRVQIWETRNNSSSDSNDETDFDFSLCDFESLNASLSMINWTNELANKGTDEATAHFYDKLYDVLHSNVPRKRLRPKQKGKLPWWSPELRRLRNGVRKARKRYYRSKSHVVRQNLHVIESRYNKCKVFTFRNYINSIESNLKNNPNSFWSFVKSRKSKNRIPEQIRFQNRTSANVDESANLFADFFSSVNSTISPTLSTATRHSIPTHDITLSFLNFTEQSSVNSTMPYCTTQVLHVEKVQRSFLRYALRSLPWDVKTTLTSLNVKKGAGADRLPPSYWKECAESGSNSVDNYRGISILCCIAKVFESMMIHMVLYNATRHLISDSQHGFVKKLSTVSNLMCYTNFLSSEIERRQQVDAVYFDFSKAFDKVPHDLAIKKLEQMGLPDWITEWLRSYLSGRKAFVRIGDTQSRGFDITSGVPQGSVLGSLIFVLFINDLSHRLKSCKLFYADDLKIYRAITSVIDCYALPSDINELLLWCSENGMQLNNSKCKCITFTRSLSSIYFNYEINGSIDRVTTINDLGVTIDSKLKFSKHVRVTTAKAFAVLGFIRRNSQAFQDIYTLKTLYCSLVRTILEYAACVWSPYCTTQVLHVEKVQRSFLRYALRSLPWNDPVNLPDYESRCMLLNMETLSFRRTKLKQMFLFDLITHNIDCSDLLAEVGFLAPSRYLRSRQNLAPRTHRTAYGQNSPLTSCIRSFSIVNDVFDFNLSRVEFKRRISSIRNLSRVEFKRRISSIR	cd01650	460	710
	CR1-1_CBre	GCA_000143925.2	Caenorhabditis brenneri	TTTCGAAGACAGACGAGATGCGATCAAGACCCTCACTACATTCAACTCAATCAAAAGCCGCGTGACCGGTTGTCACAACGCCACCATTCGTCAGGACATGTCCAAACCGGAGCTGTCAAAATTT	TAACATGTAACTATTTGTTTGATGTATAAGAGGCCTATTGGTCTTTCACTCAATAAACTCATCATCATCA	MNCFSIKNKLPQLDLVSSTNNFDVICLSETKLDDSFPDSLLSLNNKFSVVRKDRNKHGGGVAILITKKVKFCPIEIPKHFSSPEISGVDILLRGTTVRVIVAYHPSHTSSLERLVQSLEFLLCSQKHCVVLGDFNMPHINWETLTACDSPCKTFLAFVTNNGLTQHVLSPTRLNPDNILDLCLSNTKIVRDVQVGELFSDHKIISVTLKLSNSAKKVVKRVQLFRKANFQLINYNLSMINWDVEMSCMSLEEKYEFFVSVLQHQALHVPCKNINVLAKKHGPKILSLQKE	cd01650	460	710

	LINE2H_CE	GCA_004526295.1	Caenorhabditis elegans	AGCTGGTTAATTGC ATGTAGGCTAATGC TCGTTCCATTGCAT CGCTTGAAAAACTT GAATTTGTCCATAA GTTCGTAGCCTCTA ATAGTTTAGACATC TTGTTTGTTACCGA AACATTTCTAAGTA ATGCTTTTACTGAT TTCTCATGCTCTTCA CCTCACTATACCTG CATAAGATCTGATC GGCTTGAGTGCCAT CCAAAAAATGCGG CGGAGGTACTGCT GTGTTTTTTCATAA AAAATGGGCTCTTA CGCCTGTCTCAATT	CGGGACTCGTGGA AGCAGGCCATCAA GAACAACAACGAA GCTGGACAAGGGA TCTGGACTGTGAG AAACCTCGAGCTG GTCAGAATTCCCTA CAAGGACAATGAA ACTCCTCACCCTTG GACAGTGCGAAAC AGAACCAAGTAAA GTTTTTCTCCCTTGC CCAT	TGATCTCATTTCCTCACTCA TTTGTACATGATGTCATAT TTTTTTATTTATTTTTTTATT ATTTTCTCATCAACTTTTCT TGGAAATCATTTGTATCAT CTGACCTCGGACATGTTCC AATCGTGTCACCTTATCAT TCACATTTCTTACCTCATCT CATTTTTATTCCTCATTTTG CTCTTAATTTTCAAGGCAA GAAATAGTGAATAAATAA ATTAA	KLRIWRAAGNSDNYKKISLILKEELIQEDQMN TEKSLTYGTSKKFFKFINSRYKSNDEIGVIKDSQ NIPVNSNILKAELLSDAFSKVFTVDNSSSTLFPS RTNNVLESMTFEPSVIEKVLLKLVPKCNTTPD EIPAIFLKKVGTSIAIPLSIIFSESFRTSQVPRLW KTAIVKPLHKKGPRTDPNNYRPISLTSSVCKV MEKLVRKNLTDYLNSQQLFSRAQFGFRSNM STESQLTSYFSTLISNHSLRKPTYSVVVDFKKAF DTVSFEKLTCRLMSYGIKGQLLDWIVCFLTNR TQRVNVNGVLSQEKFVLSGVPQGSVLGPLLF LLFINDIGDKFQSHFLLYADDLKLFSTNPSTIQT DLGLLSEWCDKWQMSVAPNKCENIIFNLTKS PRTLNFDFTIKGSKIPRTKKIRDLGLILSSDLPFS EHIDTILRRSHQRINIFFNILRRASFETFIKCYTIY VRPILEYGSIVFNPVLKEQIKRIESVQKSFVYRV FKKFHIPYLNYFDALSVCNLTSLELRRLHADLV FLYKVIVSKETLLYPPFFTLPSLPNIRKHKFYFRS NSKNSNKFFSQFISNRSLNCWNDLSPSVFPPT PHSINFISNISYCNFQKYLTLNPHNF	cd01650	309	557
					MKFHDLAQMVNFPTRFSKLNSGTFLDLIFTN NSGVLQNLKPGTPFINSDHISIFFSLSLPQPTP AHSAPPKSFMNFRKCDFFNLNRYMVSYNWP KQLSFFAPPDSKLNHFLNIFNELISDFTPQYPTL RVQSPKVTGRKLYRKLRKHHPDTESSVLKKRV KSRLKFLKKKNMAFENKLIHSKNSKSLFSYIRN RIKPSSSLTHFSVDNNNVDKPIEMAEVLVKEF SKSVVTHDPPFPTFPSCESTPSPVDTSPLTIAST ILKLKPKIGYSLDNINFFILKKCCDSISIPLSLIFKE FITTNSFPSCWKTSIVIPVHKKGNSSDPKNYRP IALTHPLSRLFERIVLNPTRKTFSHKISKFQFGFL NNRSCITAHLNASSNYHKILSVPNQFVDIILFD FAKAFDSVPHELLLCKLRNFGFDDNICLWFKA FISDRVSKVKIGDSLSNSSFKNSSGVLQGTVTG PFLFLVYINDLLDLFPPDVHVIAFADDLKLFSN NCQSLRSSIKIIENWCKIWQLKLAENKTKVLH VGKKNPKYKYYVNGQNIECCSKACDLGIWVD DQLAYEKHILVKVNAAMFKCRQILKNFRSLN MIMFYFKLFNTYIRPILEYGCEIYHPKSASLTKKL			

			GCACCTTCCTTTTAT CTCTCGCATTTTTGT CAAATTCTGGCATT AGACCATAAGCAAT CCAAAACTAGATTT ATACTGATATACCG CCCCCATCCACTAC CGCAAGTAGAACC GCTGAGCTCTTTAA AAATCTAAACCAGC TTCTCGACTGCCCC TCCTATACTTATGT CCTCCTCGGTGATT TCAATTTTGCAAAC ACCAAGTGGAATTC AAATTTTCTTCCCA CGGGCTCTGAGTAT TTCTATAATTGG			EQPLRYFSRHVFKKCKTQYDSYEHRLNQSNQ QSLKHRRVFLILKTFHNIITGKYHFPNLALYLKN SKSPRFPYQMTAIGKTNNNSFLHTHLTIWNL VTKHFPVPVSGYTFSKRISCLPLESYV			
	RTE-1_CRem	GCA_010183535.1	Caenorhabditis remanei	CTTCTTCTGTACTG AGCCTCTTCGCGGC GAAGAAGATCAAC AGATGTCACACTTC TACAAACCCGGCG AGTTCGTCACGCCG GGAGTGAAACATC TGGCCAAGAAGTA CAACGTCTCACGCA GGTCAGAGGTAGC GTCGAATACTCTCT GTGACCGGGCGGC TTTAGCATTGGTTT AGCAGACCATAGT AAAGTAATCAGAG ACACCCAAACCGCT CAAAACAGCACCA AAACCTGCCTCAGT CCTCCGCACTCGGC CTCTGCAACTGTTA	TGAGCGAAGAACGGATCG ACTAAGTAACTAAGTAAGT AAGTAA	MDVYHSKRTKLTRPKTRKTDLRICTFNCRSVA TDNRLAELLEETNRIKYDIIGLSETKRSAEVHM TNRDGTGVILGKRNDSSVSGGVGFIINKSLMP KIKEIKIVNHRIGYLTLQVNKNQYTIIQGYAPTA DYEEEDHSNFYENLEDVYKSCKSRYKMVIGDF NARIGERKGNEVFIGPHSMENRNDSGERLAT FCEVNRVFHMNSQFIKPTHRRWTYISPDKQH RHELDHILANGKYITDVSVVPSFTNGSDHRM LRANVHINAKQAKFEQVKRRKPPRRVLDPTA ALLATENLDSCEDQDIDKEYDTLVHVLKKAQD AAVTIPTNHSRNRLKDSTRLLLSKRRFTDRSDP NFKTLSKECRQAVKRDHENFAKDRLLNAANQ RKSLRKVARDINEYQSYIPCLLSGETGEKLTSRE KMEIEVRRFYSNLFATKQSKATTIIPQQTKALP PFLPEEIQHSLNSFQNGKAAGEDKLSADFLKS CHSSVHKLLANKFSRYLQEGKVPTKWKSSKTT LIFKKGDKENLENYRPICLLPVLYKVFTKCILNRI RKSLEEAQPVEQAGFRRSFSTIDHIHSVQRLLE VGREYQIPITMVFIDFKKAFDTIEPTALWESLK TQGIDSGYIKLLKECYNDCSTTITPFYNPVNIPI	cd01650	511	779

			CGGTAGATTCAGTA GCACGTCGAGAGA GGCGATCTAATCCC GGAGTT		TRGVRQGDPISPNLFSACLESAFRRMSWPHL KEDKDDYDNSPGIRINGRNLTHLRFADDIILIS KTPQIAEKMLQELVNSCETVGLEINASKTKVL RNKFASSHQIHIRKNNTMSAIEDVEEYVYLGR LLNTKNDLEPEIHRRRRSAWAALNNNIKNTTD ALTCPKIRAQLFDSIVLPALTYGSETWTFTKAL SERVRITHAALERKLVGITLTEQREKNLHREDI RKMSQVKDPLVFITKRKLSWAGHVMRRNDN RWTTLLQEWIPRNEKRPVGRPPMRWADSLK KEISVRQGTQLIEPWSTIAKDRKKWIAVIRAH TN		
	NeSL-2_CRem	GCA_010183535.1	Caenorhabditis remanei	CCAACTCTCATCGT ATTAACCTACGGTA TTCCTCCTAGTGAG TGTAATAAAGGTTA ATTACGTTTTCTCTT ACCAAAGAAAAAG AAATTTCGAATCCT TTTTGTGTAACTCA CAAACTGACAGAG ACCTATCGAATTTC TTTTGTTTCGTATAT AGGAATAGTCACTC TGGACCACGAAGT GGACAGTTGTTGG CGGACTTCCAAAGT TGAGAGAAAAGGT GTGAAGAGAGGAG GTCTAGAAACACTT CGGCTGTCTAGGA CCAGTTCCTGAGTG GAAAGAGGAAGGT CTAGAAACACTTCG GCTGTCTAGGACCA GTTCGTGAGATCTC TTGTGAAGAGTTG AAAACAGTTAGCTG AGGCTACTGTATTT	TGAATACCGTCAGATAAG CCCCCAACATAAAAATAAA AGTCGGCGTTAGCTAATCA CTAAACCGGCTCCTCATTG GGGGAGAGTATCATTCCG GTGCTCTCCGTTTGGGCG GTAGGGAGGAGTTGGGTA GCGACCCGGAAGTATGGA TGCCCAACCACCGCAATCT GATCTGGCAATGTGTTTCG GATGGTCTCTGTCTCTAGA TCTGAAATAGAGCTCTGGC CTGAAGAACACACGCGCG GACCGGTTGGATGCCGAC TTGATCTGGAGGGTGCAA ACCTGAAAGGGAAAGTTG AAGGCCGTGAGGCTCCTG GCGGGAAACTCCGTCATA GTCAGTGGTGTGCCAACA CCCGACGACTATGACATA GTTGGAGGAATCCTGATC TGATAATAATCATTGTTCA TATAAGGGAGGGGGATG GTAAATACCCAGGGTCCG AAACCATCAAAGCAGCTA CTGACCAGCATAGTAGTG ATGAACACATAGACCCTG	MTNVYLKPVNDNQTNKTGDNSRNTMSNSQ CEMTWKPVARTYAQAASTNPADDKTVTVLG CKYNLLKLGNTPQTTKRSPPKPSRGGARISSV YTLTDELEITHREEGKITFAIDLPNKNNILCPLC RECTQTRGRGSSFTKHMKLHVKEKHQLKATF IYKCSMCNEYEPEKKCGTKWIQTHLQKVHNY KYDESAIVVPVPPNTRQQIANELNNAAPFVDI RKPKANVVEEKKTENGALLKFLTKSNKDEQE HSQSNDSPNVESPEKANQALTIDPKGNNSPS KSSMRSSQSSASSICQEIQEIITLSEDEEPKAAR PKPGINVWSLINEKGKDAYIDTDIMMAFLKM RVENYDSVNIIDPLNYQFPARVDLVPLIQRNL EDGKKRVVFPICADEHWTLLTISNGIAAFYDP TGSRMSSYIEELVNELGLIIPKEQDEQPRQRDS YNCGVFVMKMAEAFIQDTEWEMEEVEEDV KNFRRNLLEELKPNYEIFAEKIKYYNSPGKSFA QSRPTSRSSQCAVCPTCSRSATPMMDVGN MEVDPVPQQQETPKSREPEQDEGWKVVGN TRKRGVVTERSPNISPEAKRQFTGPEIKVVSP GKFHPLVGETEEMEVTCDSPPTKEPTTEPKVT PSLPAMRIVSPEVTKKQTSKKKGKYGKKKQQ TKKAQPPKGEPTKKAQPKGEPAKLIEQVRTW FDKQMKSYQEQGSNIQTLTWIADSLTAAIFK ANNGNKYLVDKITARCPPPLLNEGEMATQTS RRTEAVKPKDRFVKESNEPLRIQYAKNRAKTF NVIIGKHSARCEIDINVVENHFRQTLKAQPVT EKALNTVCSGIKKAEVDPSIEGPISSGEVKAILA	cd01650	895	1156

	CR1-2_CRem	GCA_010183535.1	Caenorhabditis remanei	CTTGATAGCCCCGC CCCCAATCCCCCTC CCCCCCCCCTCGAC AGAATTTTCTGTTT GACCTCCTGGAATT TGCGAGGAGTGCG CGAGAATTTTCGAA TTCTTCGCGCGTTT TCTCGAAATTTTCC AGAAGATTCGAGC GGAGAATCTTCGA GAAAGTGAGCCGA ATTTCGCGCGAATT TTTCGCGATTTTCA AATTATCAATTTTT GTCGGAAAATTTAT TTTCTGGCAAAATT TGATTTAGTTCACG CGGGAGAGAAGGA TTTGGTGGAAAAG GGTATTGATTTTTG TGGCGGAGGAAAC TCCCACTGAATCAA TAACTCTCAAGGA GAACTCATCGAACA ACCTCGGGTGACCT GAATCTTGGGCGA AATTTTCGCATTGA CACAAGATAACACA AATTACTGTTGAAA ACAAATCAGAACA AACTGTCAAAAAG AAAGACAAAAAGT ATTGATTAACAACA TC		GGGTTCCCTGAACTCGACC CATCTGCACAAACCCACTT TGTACAAATGAACCAAACT GATGAAGAGTTTAATGATT TCTTACATCACAGCTAGCG GACCTACCGTGAGGTAGA CTCCCGCCGCTGTAGCAG GCTCGCCATTG	TGATTTTTTTTTTTCTTTCTT TATTCTCCCTCCCCCTTGCC CAACACCTATTTTATAATTC	MSAQSLALIDHLDEYLPLKSCILAETLITHLLLES SIFNHPFSLSSSLKSLPERPFPTRILPHLQEPPLL NSDHFSVDFSLITTVSNHSSASLPTSPTFRNEH	cd01650	350	599

	Penelope-6_CoFl	GCF_000648655.2	Copidosoma floridanum	ACTCCTGTAGCCAACAACAATGATTCAAAGTGGCTGTTTAATCTTTCTAATGTTACTATACCAGAAGAAGTCTATAATATTTTGAAA	TAGAGCAACAGATTTACTGTAATTTAAATTTGCGAGTTCGTTGACTTGTACTGGGGTGGGGTGGCTGTAGCGTTGTGTGACTTAAATATACTCATGTACAGTGTCACACTTACACTGACCGATAGTATATAAGCACGATCGTAACTTAAGAAGTGCGTATTTTTCGATCTACCATCTATCGTGATGCCTATAGTCGCTATTGTCGTGGCTATGTGAGTGTTCATTATTGCGATTTTCTTGTAAAGTATTGATGACTTGAAATATAATTGTTTCATATTTTTTATTGTTATTTAGAAAGTTGTACCTGATG	RLNYKKCDFKSLNADIANVDWDLVLSSISSPTDKFALFSKMLSDFIISHTPLSTPPKPHRTSFSIRQLKRARQRYSKLLKLPNPPADQIFRLRTLIASTSKRIKSNMVSEERKILSAPNSRSARLLIKKRTRTSLKIPPLYVNNELISSNAAKSSIFSSVFFSNYNCSPSSSLIPVSNNSSNNISPSELFLPWIIENTLRKVPPRCGFTTHHANFYIIRNCATTLAIPLSIIYSDSFLKSHVPASWKHAIIIPIPKKGSLTSPENYRPISLTDPFSVKIDDFVGTPSASIPSGVPQGSVSGPLLFLAHSKTALLRLGALQMGIDLVESWASDNSLPLAHSKTALLRLGPRNSSHPYTIGGSLIDSVDSVRDLGLLIEPNLKFTRHINRAVALALLRSKQLLKSFKSNSPQFYIFLFKTYVLPLVEYCSVVYSPPPSSKLAHKLETPLRFFSRVKILQRCNTPYCSYSDRLSQLDLFSMRHRRLSRKSRCSSSLVHSISEYRLLLSSHKSLDVVYFDFRKAFDQVDHQFLISKLDSFGIPSNIISWLTDFLSDRTFSVKIDDFVGTPSASIPSGVPQGSVSGPLLFLAHSKTALLRLGALQMGIDLVESWASDNSLPLAHSKTALLRLGPRNSSHPYTIGGSLIDSVDSVRDLGLLIEPNLKFTRHINRAVALALLRSKQLLKSFKSNSPQFYIFLFKTYVLPLVEYCSVVYSPPPSSKLAHKLETPLRFFSRVKILQRCNTPYCSYSDRLSQLDLFSMRHRRLKAQLLLYNLIIGASYFPSLETHVRLSSSSRRPMSLICINPNCSNFFSTVIPIWNALTINVPHFLCPSEFNSLLVNNIARF	MIRETSYFTTNMNTHSKFEEFFKLLETSLHFVPTVTDKIRTNLLKMFDKFLKSTIAQKFDNQKIINDINKTALFLKHNKELLVTKVDKGNPTVVLDRSEHINSVEEQLNNDKYYTELLKNLLNELIKSVTTFVKRWYDLGIFKDCEFDIRSSEISKNDLSKCYGLVKIHMAGNPIRIIVSAVNSPTYNLSKFFTKFLNDNIPKPNSCIKNSFHIKEKLNSFIIPQHHTIISLDVVSLFTNTSESIILKSIKNNWKYLRNKINLSSQEFCEGVSLILNSTVFQFNNKFYKQIFGSAMRSPVSPVLCDLALRCLEEDCLKKLSFFVPFYVRYVDDILICVRNNSIEETITVFNNHDRHLQLTVETPVDNEISFLDLRLIFESDGSIITNWYRKPTHSGRYLNYNSHHPLSQKIAIVYTLVDYSIKLSDSKFHKENLDISKKILFNNQYPLNFINKYIKKRLTILSINTPLNRVVSNNRGLMILPYFQGFDSIIKRYLYKNNIVCNDCDSCYIGQSSRRLKTRINEHMINDCKHR	SSF5 6672	203	380

	RTEX-15_CGi	GCA_902806645.1	Crassostrea gigas	T	AGGGCCTTGTAACGTGGT TTACTGTACGGGCCGGAA CGTTGTTTACAGCAATAAA CGATTTTGATGGCATCACG AAGAGAGTCATTATTTCTT GATAAAAATTATCTGCAAT CACGAGAGCCACTGACGA ACTA	REKSALVDHIAKTNHSFNFDEVKILDTESNRFK RLFSEMLCIDSYRNNINRMHDTQNLKLCYKN VITEHT		324	600
					TGAGTTTCACTTTGTATAA CTTTTCCTAACTCTATAACC TTTATATTAATATATGTTTA CCTCTGATCAATTTATATG TATATTTTAATGTTATTTTT GTCCTCATGTACCATATGT TTGAAATGGTTTTGAGCG AATAAATGAACTTGAACT	MAVTCRNCSTVDYVLASANVFEILEDLDIHDF CNLFSDVHNPVSFSLSTCYLSRNISTGKNHSG VKLWCPDKADQFRSNIDVHAVKSIENYLNEIS NSPGDLCSDNINKVVKDICTVFTDAARDTFGP SHFSPCIEKHNKPWFNRECKSARKQYHLAKS MHNKSKTDDNKNKLKEKSKIYKKAMDNAVK TYNTGITRKLMQLRSTNTKDFWNILNSSKRDS KSPVNIKDMFDFFKNINEGSLSPDSDTLHEPQ VDGSSNELLNCEINESEITDAVKKLKNGKAPG YDNIVNEHIASTLNTFLPIYKLFFNAVFDSGLV PEEWLVGIIKPIYKNKGDPTSPENYRPITLLSCL GKLFTSILSKRLELYSEEIELIHNNQAGFRKGFS TLDHILTLQFLSNTLMKCKKKLFCAFIDFKQAF DTVWRNGLWCKLLNYGINGKCFKYITNMYK GIKSLIKMNGTVSDFFSCNVGVRQGENLSPFL FSLYINDLENYLVEKNIAGLQSFTTSIENELCVY IIMCIYIKLLVLLYADDTVILAESGDDLQNALNE FECYCKEWKLNVNIDKTKVLIFSKGRMCKRVF YFNDRVLEIVKEFKYLGIIFSRAGSFLKAKKHLC DQAQKAMYGVIRKIRQFNLPVTCQLELFDKM VVPVLLYGCEVWGFEKLDSIERVHLKFLKHILC LKSSTPNYMVYGETGRFPLSVIVFTRMVSYW CKLISSTNTSITNILYRYLLNQTENTTSTNPWID CIRHVFNSCGLSNVWHSQFLNIVNQKWVST AVKQRLQDQFIQKWHADINNSSKGVIYKIFKT NFGYEKYLDILPLKFRKILGKFRTSNHHLPVEL GRWEGIPLNERFCPLCNTKRIADEFHYILECD AISQVRNKFIDKKFSARPNVLKFSYLMTTCNM KLLRKLCIFITKIYEVVCPP	cd01650		

	CR1-24_CQ	GCF_000209185.1	Culex quinquefasciatus		TAAATCATGTTTTGTTTAA ATTTATGTTTTGATAACTT GACTATATGTATGTTTGTA AAGTTCTTATTAGTTTTAA GCAACCAAACTTCATTGAG ACCACCGGGTCAGATGAT TTCAAATAAACAAATC	MEDPRLFVTVVPSHSRNYAFCTSHPSRPGPV FGPRRGVFQNRSVGKYPYAENDLVSLPDARL TSSNVSTNESIRQDSAYRSQDAASQQSNDNA DPGRPPTCYMEDLRLSVTVVPPHSRNYAFCT SHPSRPGPVFGPRRGVFQNRSVGKYPYAEND LVSLPDARLTSSNVSTNESIRQDSAYRSQDAA SQQSNDNADPGRPPTCYMEDLRLSVTVVPP HSRKYAFCSSHPSRPGPVFGPRRGVFQNHLV GKYPYAENDLVSLPDAHLTSSNVSTNESLRPD SAYRSQDAASQQSTDNADPGRPPIRIMEDPD PPDTVVPPHCCDSSRSGHQSRPGPVFGSREG VFQTPNAGKYQAAFQVGYALDPDPRLLSSDV QTTPVHSDDPTTPVHHDSRRSSAGQPVIFEPI RFYYQNVRGLNTKVDDFFLSCCDCDHDVILLT ETWLDGRVTSAQLFGNNFSVYHTYRNPANS KRKVGGGVLIAVRNSLDSVECVQAVAENLEQ MFVTIRTPSKRIFVGCWYLPPEKRFETELMEK HLSCIENIRRIAGPHDVIVVAGDYNQGSLVWK HLSGNRSYVDPQLTRTRTEKERLTHGVMLDG MANCNMHQCNLVKNDRNRILDLIFVSDDDC TVVAAADEPLVPLDAPHPALVFDIDASRTVFV QVDNNDRVLNYKRTNFDELQRTLADIDWTP VTSAMDVDTAVREFSNLLNFHLAQHTPQFRT PPKPEWSNARLKRLKKERAAAALKTFTLRRCR MTKLAFNAASSRYRRYNRYRYNVYIRKTQSQ LRRNPKKFWTFVKSKYKERGLPATMALGDST ATTNDEKCELFARHFSSVFREPEIISSTHHLDA VPRDLVDINTFVISRDMIEKATKKLKSSFSPGP DGIPSAVLKRCANQLIEPLLVIFNLSLRQAIFPT AWKKSFMFPIFKKGSKRSIENYRGITSLCSGSK LLEIIVSDVFLFNCRSYVAPEQYGFMPKRSVNT NLLEFTSVCIDKIAEGRQVYAIYTDLKAAFDRI DHDIALQKFSKLGFSPMLCRWLESYLKGRIIQ VKIGASVSLELTNGSGVAQGSNLGSIVFSGFF NDANFLFSGNCKLSYADDFKLFIPIDTLADCYV LQSRLDKFSEWCGHNRLELSSAKCYVISFNRK QSRELFTFDYHLGSHVLERLEVVKDLGVLLDL KLDFHLQLSSVIDKANRKLGFMFKAAQEFDD	cd01650	918	1167

					PLCLRTLYCALVRSHLETSAVVWAPYHQNWI DRIERVQRKFVWFALRRLPWNNPAQLPSYET RCNLLGIETLQNRRTISRAVFAAKVITSEIDSPN LLFQLNARVQPRNLRPTPGFLSRPLVRTVYAE NSPIRALATAFNEFYHLFDFHEPVSRFLERLRT ELRDRSRSSRTAENPQQRASRTRRSSR			182
	Crack-10_CQ	GCF_000209185.1	Culex quinquefasciatus	C	TAATTATTTTGTTAGTTAAT TCTCCGATTCATTCCCTCGT TTGTATTTGTATCCATCTTG AACACATTTATCTTGGTAG TTTTCTTTTATTTATTTTCA GCCGCTAATCGCCAGAAA CCGAACTCCTTAAAAGGTC TACGCCGATGGAGTTCTG GAGCCAGCCGCATCAACTT TGTCGTTACATTTTTATTTA ATTTTCAATGACATGTAAT CTACTTAATTTGTATTGCA CATTGAAAAGAAACAGTTT TTTCAGTCAGCACTTTGCT GGCTTTTTCTGTCAAATAA AAACAATCAATCAATCAAT CAATCAA	MLDNRDDVLGLFLDLSKAFDTVDRKILIDKLR YAGVRGIALDLFTSYLSNRAQFVCIDGIHSLLT LVNVGVPQGSVLGPLFFIIYLNDFSLLPLKGDL RLFADDSSLFYNNKSTNMNDLNLRDDLTIVIE YFRLNKLTLNIINKTNIINIKNSSRSIPNSLSLTTS KFPDLEVVSDTKYLGIILDNRLNWSAHINALIL KLHKITGIIFKIKHKLPQKVLFLIYHSLFHSILSYV TAVWGNACGLLINKLQVAQNRILKIILNLPIRS HTVDLYVKNNILPVRGIYVFQVCSFIYLCLNNK THSNTKFVRSNHQYGTRYHDLLNRPNVTTVP GERSINFNGAQLYNYFNNRFGKSSSLNIFKNQ LKQCLSRPDIIEKLLKSVNIFA	IPR000477	1	182
	Crack-10_CQ	GCF_000209185.1	Culex quinquefasciatus	C	TAATTATTTTGTTAGTTAAT TCTCCGATTCATTCCCTCGT TTGTATTTGTATCCATCTTG AACACATTTATCTTGGTAG TTTTCTTTTATTTATTTTCA GCCGCTAATCGCCAGAAA CCGAACTCCTTAAAAGGTC TACGCCGATGGAGTTCTG GAGCCAGCCGCATCAACTT TGTCGTTACATTTTTATTTA ATTTTCAATGACATGTAAT CTACTTAATTTGTATTGCA CATTGAAAAGAAACAGTTT TTTCAGTCAGCACTTTGCT GGCTTTTTCTGTCAAATAA	MLDNRDDVLGLFLDLSKAFDTVDRKILIDKLR YAGVRGIALDLFTSYLSNRAQFVCIDGIHSLLT LVNVGVPQGSVLGPLFFIIYLNDFSLLPLKGDL RLFADDSSLFYNNKSTNMNDLNLRDDLTIVIE YFRLNKLTLNIINKTNIINIKNSSRSIPNSLSLTTS KFPDLEVVSDTKYLGIILDNRLNWSAHINALIL KLHKITGIIFKIKHKLPQKVLFLIYHSLFHSILSYV TAVWGNACGLLINKLQVAQNRILKIILNLPIRS HTVDLYVKNNILPVRGIYVFQVCSFIYLCLNNK THSNTKFVRSNHQYGTRYHDLLNRPNVTTVP GERSINFNGAQLYNYFNNRFGKSSSLNIFKNQ LKQCLSRPDIIEKLLKSVNIFA	IPR000477	7	181

	L2B-6_CQ	GCF_0002091 85.1	Culex quinquefascia tus	AATTGG	AAACAATCAATCAATCAAT CAATCAA	TGAAGTGGCATCCCATAG GTGTGGCTGTGAAGGAGA GCATGTCATGACGGTCGG CTTCTCTACA	MLAIEIHHQQFNGIFGGVVHSPSSSDRDFLDS FENWLQQVMADDKTNVISGDFNICWNNEG YSRDLKNIADAVGLEQKVLEHTRISVRSRTTID LVFANVEDCTVKVLADWRITDHETLGINVEG VCKINMPEKKTEYTCWRKYSKERLQNTLLCDE TLWRNNLPVDEKAQQFSRALATAVGQLTET RMRRSADTKRWFTPHLQAVKLKRDEAYKIFK TTNCAEHWERYKQFRNEYVRELQKAKNSSV QQEIQSCSGDSKKLWRCLKDLIKPGGTPEKEII FDDKTEPCSDEETANRLNHYFIKSVKEIHDSIP TVNQPQQDPLPTVRTLDISEFSQFQPISMEKL RKIVAAMKNCAGVENISKRVLEDAMDVVGD KLLDIINSSLNLGVFPQEWKKSVVVPIPKVPNS TRAEDRRPINMLPIYEKVLETVVREQLVEYVD RAGLLIEEQSGFRKQHSCESALNLLMLKWKQ SIENNKFILTVFVDLKRAFETIDRRKLLEVLRRN RIKGDVWRWFKSYLEYRTQITRYNNSVSEAE AVELGVPQGSVLGPLLFILYINDLTKALRRACV NLFADDTVIYVAGDKLDECYEIMNNELAVFA DWLKWKKLKLNVSKTKYMVVTTRRSESNCTI SVDGELVERVNAIKYLGVMLDEKLSFAEHVDY TIRKAARKLGVLCRINRYLSFDNKIMVYKTLIA PHFDYCASILFLATNQQRKRMQIVQSKAMR MILRCDRLTPRILMLDSLQWMSIRQRVEYSTL VFIFKVVNGLAPQYLTKTVQCGRDVHRHFTR QAGDIRLLNFKKSCTQNSLFYRGYNLFNMLPE STRATTNLREFKKLCKPFVRQRPLE	cd01 650	397	652
	RTE-3_CQ	GCF_0002091 85.1	Culex quinquefascia tus	TGTGTGTGCGCCGT GCACATCGGTATTG TTTTTCGTACCGTTC CCGCGAAAGTGCA AAATTTACACGAAA ATACGCGTTAGTG GGGTGTGAAATTC ATTGCATCCCGTGTT CTGAACCCGTCTCC	TAGCGCCAATAAAAGTAA AGTAAGTAA	MVLRRDRGLVQATRTRRKTPVQEAHDASRT NRHGTGHHMRSHDWKLGTWNCRSLKFDGS IRILSDILRVRKFSIVALQEVGWIGAEEVQAYP RIGCTIYQSRGENKRLGTAFIVLGEMRDRVIG WTPLTDRMCVLRIKGRFFNISIINVHSPHSGS EDDDKDAFYEQLNWTYNSCPKHDVKIVIGDF NAQVGQEEEFRPVIGKFSAHVRTNENGLRLI DFATSKNMAVRSTCFQHNLRDKYTWRSPQG TESQIDHVVIDGRHFSDIIDVRTYRGANVDSD	cd01 650	569	829

				TGGCCCGGGAAAG TGATAGTCCGTGTT TTACGACCGCGAGT GAGACAAATCAATC ACCAGATTTGACCG AATATCGCGTTCCG CCACGGGCAATTG AAAAAAAGTACAA AGAACTGCTAAGG CATCGTCGTCGTCG TCGCGATCGATGGT CGTTGTATTGTGTC GCGCGGCGTGTGG AAGTGTACCGCGA AGACCATACTGACG CGTGTCGTCGTCGT CCTCGTCACTGTGT CTGTTGCTACGAGA GTCGTGTATAGAA GAGAAGCCGTTGG TGCACGGTGGGGT GAATTGCTAGGAA AGCTACAGAAAAA AAAAGTGCTGCAA AGTTTTTTGGGTGA AATTAAAAAGAAG GAAAAATCAAGAA GCGCAACCCTGCTG CATACACAGCCCCC CCCCCCCCCCCTCT AGTTCCGTTTCTGG GCGTCTGCACCCCA TGTTAGGGGCGGC CCAAACGGACTAG GTGGTCGATTCCTT CTATCCGTGAGCG GCTGTTCCCCACGT TAGGGGCGGCTCG		HYLVMVKMRQRLSLAKSVRYRRPPRLDLERL KLPEVASRYAHSLEAALPGEGELLEAPLEDCW RSVKAAITNAAESTIGFVERGRRNDWFDEEC RAILEEKNAARRAMLQYNLRDYEEAYGQKRR QQHQLFRAKVRHQEELEFEDMEQLHRSNET RKFYKKLNGSRNGFTPRVEMCRDKNGAILTN EREVIDRWKQHFDEHLNSAEAEAGVQGGRR EDFIGTAGEGEEPVPTMREVKDAIKKLKNNK AAGKDGIGAELIKMGPEKLASCLHRLIVRVWE SEQLPEEWKEGVICPIYKKGDKLDCENYRAITI LNAAYKVFSQILFSRLSPIAEGFVGSYQAGFV MGRSTTDQIFTVRQILQKCREYQVPTHHLFID FKAAYDSVDREELWKIMDENGFPGKLIRLIK MTMDGARCCVKISGAESDSFTSLGGLRQGD GISCLCFNVVLEGVMRRAGFNMRGTIFSKSN QFICFADDMDIVGRTFKAVADAYTDLKREAE KVGLRVNVAKTKYLLAGGTESLRARIGPSVTID GDEFEVVEEFVYLGSLVTSDNSCSREIRRRIIA GSRAYFGLHKSLRSRKFSLHTKCSIYKSLIRPVV LYGHETWTMLEEDLRALSVFERRVLRTIFGGV YENDGWRRRMNHELAQLYNEPSIRKVAKAG RLQWAGHVARMPERADHLSQRNQKINPAK LVFVSEPVGTRRRGVQRARWVDQVESDLES VGAPRNWRNAAMDRACWRRIVQQAKLMV			

	Kiri-7_CQ	GCF_0002091 85.1	Culex quinquefascia tus	AAAGAAGTGGCCG ACACCCCCCCCCC CCCCCCCACCCTT CTTTGAGCGTCTGT TCCCCAGGTTAGG GGCGGCTCAAAGA AACCGGTGTCCTGC TCTATCGTCGAGGT AAGCGTCTGTTCTC CAGGTTAGGGGCG GCTTACAGCAGATA GAGTTCGGACCCCC CACCCCTCCCCCC ATCGAGCGTCTGTT CCCCAGGTTAGGG GCGGCTCGAAACA GCGTCTGTACCCCA GGTTAGGGGCGGC TGAGTAAAAGTCCT TGTGTCGGCGTGG GACTGTAAACAGTA CCGGCACG		TAGATCCAGATTAGATTTT AGTTTTTTTTTCTCACCCG TTCAGAGAATTGACAATTT CTTTGGCTCCAAATGTACA ATCTCCTTATCCCTTCCTTG CCCTGATACAAAGTGTAAA CTAACAGGATATCGTTACC AAATAAATTACAATTACAA TTACAATTACA	MKNDPAGDVDADELNAFFSSGHRQLQAVD RSDNSSEPCHRTTADHGDNGFAFRHTDVNEI SRKILEVQTNATGTDDIPISFVKLLCPFVLPML AHLFNHIIDTSSFPAAWKKAIVTPIPKSSSPIQP KDFRPISVLPAVSKVLEKVLLGQISEHLNAAEP PLLAQNQSGYRKGYSTTTALTKVVHDVYSNL DENRCTVMVLVDFSLAFNCVNHQILRKKLNT EFKFTRSACDLITSFLSGRSQIVRFGNSMSAAL DVPDGTPQGSCLSALLFSLYINSLPRNLKCEW QLYADDLQVYLSGPIAEVDRIMRDVNEDLAAI ADWARVNQLFPNPKKTQAIVFNKTGTVTPQ ENIVFCDSIIPLSSQVINLGLHMDCNLTWKAQ VNDVVRKVYHTLRTFRRFGSVLSLQTRRKLVQ AVIVPFFTYCDTVYFPGLSAALREQLHRGFKS ALRFVHNLRRRDTTVALRNSIMGHDLPDNYQ LRVCCFMKKAFEGTLPDYIMQHLQRGRQERT	cd01 650	111	371

					AGFIIPRHTTSSGKSVLVHGASCWNSLPMAIK REARFNPFKRAVITHMQN			828
	Rex1-20_DRe	GCA_008692375.1	Danio rerio	ACGTGGC	ATTGTGCTGAACTTGAATT AGATATTGGATATGTACAA GTGTTTACTCGTTGTTTTT GGAACCAAGGTGTGCAGT ATGAAATGTAATCTTATTA TTGTTTTTATATGTATGTA ACTTATTTATCTATTTTTAG TATGGAGAGCTCTGCTGT GACATGTGAATTTCCCTCC GGGGATTAATAAAGTCAA ACTAAACTAAACTAAACTA A	MMAPYLWQPVAAALDVFNFVYLLYFLNIFLQ WLRTLYLHTFQFESIRKHWIIFIFILAFLIGKSDA VNSPAIADAVIYSREQLLALRHTSTFSVPTME VPTELKRRRRGCRAGVKRRLKKKRYKPSLPVII MGNVRSLGNKTEELTALVRTQREYRECSFMC FTETWLNDKTPDSVVALDRFHLVRADRCAKQ TGKNKGGGLALFVNERWCRPAHITTKEVVCN KDIELLAVGMRPYYLPREFSYVIAITVYIPPSAN ASNACEHINTVTASIQLKYPHSLIVISGDFNHV SLQSSLHTFTQYVDCGTRDCKTLDMLYANVK DAYASTPLPPLGRSDHNLIYIRSVYVPAVRKQ PPTIRTVQVWSDSVYEALMDCFDSTDWEVL CKSHDEDIDSLTDCITDYITFVLRTQYPQKRLN ATQTISHGLYNFCVENTVPTKKVKCYSNNKP WVTPELKALLNEKKRAFLAGDREEIKRVQKEL KIKIKEGKVIINLKWNINYNKIMLEMYGVIVCT KLEVTIEQVRNGLRRIKKKKSAGPDGISAGVLK ACADQICWVVHYIFNLSLRLERVPILWKTSCV VPVPKKPNPTEFSHYRPIALTAQLMKVFERLV LNYIKKLLIASEDRLQFAYKTGVGVDDAIIYLLH KSLSHLETSGNTVRITFFDFSSAFNTIQPALLKV KLENAGLQKGMVSWLTDYITNRPQYVRMH NCVSQVIMSSTGVPQGTVLAPYLFTFYTSDFQ YSSKNCHLQKFSDDSAIVGYIRNHDESEYRLLL KSFADWCDMNCLNLNTSKTKEIIVDFSRHPS QAVPVIIQESEIEIVSSYRYLGVYLNNKLDWSE NTIQVYKKGQSRIHLLRRLRSFGVCQTLLKTFY DSVVSSVILYAVTCWGGGLLEKEKNKLNKLIK RAGSVVGGALPTIEEIAQDRMLDKFVSIMNH DSHPLYDTVQACASSFSTRLIQPRCYRERYRKS FLPTVIRLYNQSHTR	cd01650	578	828
	L2-6_DRe	GCA_008692375.1	Danio rerio		GAAAAGAGATCATATCTCC CCTGTCCTTGCCTCATTGC ACTGGCTACCTATAAAATA CAGGGTTGATTTTAAAATT CTTCTTTTCACTTATAAAGC	MLTQPVSFSRVFHENSALEFSNSYTNFLSNAN ASLPSCSDTEQLAETFNSACDFALNEVAPLKS KRSRVSSIPQPWLNSSTRTLRQECRRAERRW KKDGLQVSYQILKDCLNSYQDSVSEAKSKYFS NLIAKSASKPKVLFKTINSVLNPTMCSVPVVNL	cd01650	260	521

								545
							292	
						cd01650		
				TTTGCACAATTCTGCCCCG GCCTATGTCAGCGAGCTTA TTAGGCCATATACAGCTTG CAGATTCTTAAGATCACAT GATCAGTTTCTTTTGTCTG TCCCTCGGTCTCGCTGCAA GTCAAAGGGTGATCGAGC TTTCAGTGTTGCGGCCCCA AGGCTCTGGAATAGTCTCC CTCTGAATATCAGAATCTC TCCTTCTTTGGATGCTTTTA AATCTAACTTAAAGACACA TTTTTACTCTCTTGCCTTTG ACTGATGCATGCATGTTAT TTTTGTATTATTTATATTTT AGTATTACTTTTAAATATA TTTTATTGTATTTTTAATAC CAGTTATTTTTAGTTTTATT GTCTAGTTTATTTTATTGTA AAGCACTTTGGATCAACTA CGGTTGTGTTAAATTGTGC TCTATAAATAAAGTTTGCC TTGCCTTGCCTTGCCT	ETCIKFSDFFAQKIQDIKSHLFPTDIDHSDSVPP SNSFTNFQPVSSSQLLNTISQMKLTFCPSDVL PCRLFKEVLDTICPSLLALINSSLINGVFPDVFK HAVVQPLLKRPNLDPAVPDHYRPISKLSFISKV LEKIVLSQLTSYLDIFYILDPFQSGFRAKHSTES ALLKVTNDILLSMDSGSSAILILLDLSAAFDTVD HNILLRRLECVVGVQGLALQWFSSYLKGRSFS VNIGSFTSPPKPIHCGVPQGSILGPLLFSLYML PLSSILLKHKISYHCYADDLQLYFPIKLGKDLSLE NLFLCLTDIKTWMTNNSLQLNENKTEVLILGP AATNNSIKTQLGALNDNLHNHAKNLGVIFDP LLQFDKHVNAVVKSSFFQLRLVAKVKRFLSKK RFRNRNPCFNLLSSRLLQFFVYWPTSVNTLSA TVGPKCSRQINNRN	MLDLCYGSIKGAMKAVAGPPVGFSDHNVVH LLPVYKSVLRREKVRKREIKVWSEDNSLALQG CFDCTDWSVFKETSESIDELTEVVCDYISFCTD MIIPHKTVIDFPNNKPWLSNDTKSLLQSKHKY FKQGDMSSVKQLKKAVRVQINKDKIRYKDKIE NQLSSNNLRSAWQGVKAMAGCKEDRGGSEI NLNGFSSAREFAEELNTFYLRFDEDDFKSEND NLKQHLSCSNSVNISRIQVKGLLQRTNTKKSP GPDNIGGRVLHTCAEQLSEIFHYIFSQSLMLQ RVPEIWKKSVIVPIAKTKFPRELNDFRPVALTS LVMKCFERFIKGEILLQSKDKFDPLQFAFRQSR GVEDANLTILNYLFNHLENPKTHARLLFIDFSS AFNTMKPHLLVEKLISLFNLDLNICGWILDFLV GRQQCVRVNGALSDFQFCSTGSPQGCCLSPL LFIMYTNDCRSIHRDCHIIKYADDSVIVSLLHD				
	Rex1-23_DRe	GCA_008692375.1	Danio rerio					

					QDLGHGPAVQEFTSWCDRFSLQLNASKTKD MVIDYRKASPTPNITNIKGLDIEIVDTHKYLGV VIDNMLTFQPNTQAVYKKVQQRLFFLRKLRS FQVSTSMMTLFYKQFIESVLSFCIVAWYGNLS VSSKNRLSSLVKVAGKIIGFNQTGLMVIYQNH VIRRAQGILCTPDHPLYSVFQLLPSGRRFRVPI CRTNRTRNSFIMVAIRSLNESRGEGDVSIY	cd01650	491	742
	REX1-7_DR	GCA_008692375.1	Danio rerio	TGCGCAGGCGCGA GCTTGCTGCAGCAT GTTTTGACTGCTCT GTCTTGTCTTAGTG TTTTAGAGTTTAAT TTGTTGTCTTTATTT TTATTTTTTGTTTT AGTTTATTAGTGTT AGTAAGCGTGCCCT CCTTAAACATGAGG GTAGAGAGAGCCA TCTCTGTGTTTTTTT GGTTACTGCTCTTT GTTTGAGCTTGCCG CACGCTGCTAATGC TAACACAGCTAGCG GTTGGTCCGGTTTT TGTTTATACCCCGG ACCAGCTGATCTCG TTGAGGCAGGATG CTTCGAGCACAGTG CGGTCTCTCGGGTA TCCCCGCGGGTTAC GGAGGGGGGAAGAC TAAGAGAGGCTGC CGTGGAGGACAAG GTGCGAGGGAAAA GAACGAGGAATGG GAAATTGTCTGATG AGCAGCTGAGAAG ACGGTATAAACCTA	TAATTTATGCAATATCTTG CTGCTCGTGCCATATATTA ATTAATTATTACGGGTTTT AATGCGTGGAAGTCTGTG CATTATGTTCTATGATGTG CAATAGTCTATTAGGTGCA ATATTTGTACGTTTTTTTAC TGGGTCTTAAATATTTTTA GTGCAATATCTTTGTTTTTT AAATGAGTGCTGAATGAC TTTACTACTTGAATTATTGT ATGTGTAATTGTTTTATGT GTAATATTATTGGTTTTGG TTAATATTATTTTCATTACT TGATATATGACTTATTAGT ATTAGTTACCACTGATTCT TTTAAGTGTAATATCTTAT ATAATAGTTTGTATACTTA TGTATTGTGGCTGTTTTCT TTTATAGATTGGTTGGATT TATGTGACTTATTTCACTT ACTTAATTTATTTATTTTGT CTTGTGGTTTTGTGTTTCT GTGTTCTGTGTGTTATGGC AGTCATTACATAAGCAATT TCCTGCGGGGATAAATAAA GTTTTCTAAGTCTAAGTCT AAGTCTAA	MTWLHECISDSVVDVAGFSSVRADRTRRESG KSKGGGLAVFVNERWCHPNHIVIKERICSPDI ELLAVGLRPYYLPREFSHVIAIVVYIPPSANATT ACDVIHSATARLQTMHQDAFVLISGDFNHVS LNRTLSTFSQYVECPTRENKTLDLLYANVEEAY SCIAIPPLGRSDHNLVYLQSTYLPLVKGSQLLL GQGSGQKRPVQPCKTAWRLRTGMLCACH MVWTLITSQTVSLSISTSVWIVLYHEIHTLLSY CTMKSIRCFPNNKPWVTKDIKALLNEKKVAF RSGDKEAAKEVQRRLSVRLKEGKERYKKKLEQ ELKQNNSRYVWRGMKNITGFKTSSPRTDGD VDRANQFNMFFNRFDSRPLSSSDTALTAPSA SAPPNSVTSNCSILPYFTPSPSSSFTIVPTEASIA SPPFIGVSDSYKPPPMIITEDQVRRELCRLHPAK APGPDGLTTQVLKLCASQLSGVLLRIFNLSLSI KKVPVMWKTSCLVPVPKKARPSVPSDYRPVA LTSHSMKVFERLVLGTLRPLVRSAQDSLQFAY QAKIGVEDAIIYLLHRAYSHLDKPSASLRITFFD FTSAFNTIQPARLGSKLSAMQVHAPLVAWIM DYLTDRPQYVRLQGNKSDTVLCSTGAPQGTV LSPFLFTLYTSDFQYNTEGCHLQKFSDDSAIVG CIKGGDDLEYRMAVDSFVDWCELNQLQLNI QKTKELVVDLRRSRRSPVTPLSIRGVDVEIVQ DYKYLGVYIDNKLDWSKNSLVTYKKGQSRLYF LRKLRSFRVCNTMLRMFYESVVASAIFYAVVC WGGSVKVADMKRFNKLIRKAGSVVGEELDN METVAEKRTLCRLRSIMHNVDHPLHNVVDEL RSTFSHRLTSLKCSTERHRKSFLPTAIRLHNVSL SHLPLQ			

			ACCTGCCGTCTGTT GTTATGGGCAACG TGAGGTCACTGGA AAATAAGATGGAC GAGCTTACGGCAC GGGTCAGATGCAC AGAGAATTTCGTG AGTGTAGTTTGATG TGTTTTACTGA					
	Tx1-49_DR	GCA_008692375.1	Danio rerio	C	TAAAAGGCTTTTGTAAGAC CTATAGCACTCAACTAATT AACAAAATTTTAATTATTT ATTGAGACATATTCGCATA TTAATATAAATGAGTAAAT AAATAAAATGAATAGTAT AAATGAATAAAATGGGTA TTAAACAAGATATAAT	MYYMQEKEKGGVLIGLDLEKAFDRVEHDFLF AVLKEFGFGENFRKWIQIFLYRYFNLYLVLNVF FFLTECLQPTRSIRQGCPLTALLYTLEAEPLGLA IKSDRSIKGIQIENNDNREAVYQYADDTTSLGI LERYCKASGAKVNVGKSLCMRVGRVKEITQH IPFKEEKEHTKMLGIWVGMDSNKADILNWE RVLNGIEKRLRFWKARNLTLKGKYSSKFDVKR KVFILNALMLSKVWYVLSVTALPEVYYEKIND VVLKFLWGNGRAKIAYYTLIGEVKEGGLGLLD SFLKMKTLTIKIIQRFLEGKMGVIWKSVLDYYL KKCGHFNMGCNVLLMKMKKRMIAGLPKFY VEVLESGLRIMRGIFRKCVDYTKSKGADSRTT TFTELNNFKTILSEGNVLYFKEWWDGGLRQV KDVLYEVIKGFVPLKVIVEDGYKRVVKAKLQR QCHELKKAIPKDCIVIIEKKKSWAMKMNYQR PILRRKMERSMSLCTVKSFYAIFR	SSF56672	8	159
	L2-42_DRe	GCA_008692375.1	Danio rerio	GACGTGTGCAAAAG	TAATCTTTATTGTGGTTGT TGCTGTAATACTTTTATTCC TATTATTACTTTTATTGTTA TCTTTTACTCTTTTAGAGTG TGTGCTATGGTTTTGTTCA GCACTTTGGTCAGTGCCAA ACTGAAGTAAATGTGCTTT ATAAATAAACTTTACTTAC TTACTTACTT	MAPVCLSACRLPRQFLNVCSATILSVLFLLLFM SCVDALISYDRQALLNIRGLLENQPRDCSITHD YCYSSFPFPTAAFVRRLPCCLIYQRKRRRRRRGN RGGVRVRMRKEISISLALHTACSFVPTYGHRG LHPARRSWDVRYTCHLPIFPDKRQLAEFGAP PRVRIRRRGVNLQHIKSLDYVHQSDLQVISHI NMALLNARSVSNKTFILNDFFTSHNLDFLFLT ETWIGAGESSVFGELCPLNCSFISTPRCIGKGG GVAMVFRNWFRCRTISVGDFTTFESQCALIE SDTTPFLCVLVYRPPKPDRGFIKEFSEFVSHIIT SYDRLLILGDFNIHVCCPGKPLVTEFMHVLDS FGFTQHIVGATHALGHTLDLIMSYGFSIENIVI EDASFSDHKPIVFNTTLSYFSGTTKSEGFHSRC	cd01650	673	932

								803
							526	
						cd01650		
					INSLTANQFSTMFNSNKVSTAILAAAEHSSSP DDLISLFNSSCSDILDCVAPLKYKYPKMKSQP WLDDLSRSLRQICRKAERRWRKDHLVVSLEIF RDAQVNFQKAAKKARATYFSDIINKHSHTPKI LFDTINAIISPRTSCKIEQSSEICERFKRFFTEKV ADIRSSFAPLCPSSCVDSTVSSATFCNFEPVSL SELRGLVFQTKTSTSPCDPISSNLLKETFNTIGS SIQAIINSCLVSGTVPAFLKHAVVQPLLKKPNA DINCLNNFRPISKLPFMSKVLERAVLSQLQPFL KSSHKLEPLQSGFTMHHSTETALLKVWNDLL LSVDSGSNAVLVLLDLSSAFDTVDHTILLSRLE HWVGIKGSALQWFQSYLKSRSFSVSVGQFM STSAPLLCGVPQGSILGPMLFNLYMLPLGNIIR KHNISFHFYADDSQLYVPLKSQNSLQSLINCLE DIKCWMAKNFLQLNNDKTEVILFGPSKTRMS YIGGLGSLAPYVKSQVKNLGVIFDSELCLEKQI SSVVKNSYYQLRIISRLKSILSFQDLEKVIHAFIT SRLDYCNSLYVGLPQSSLARLQLVQNAAARLL TGTKKHAHISPVLASLHWLPVNFRIQFKILLIV FKALNGQAPSYISELINLQSTPRSLRSSNKILLQ VPRSRLKLKGDRAFAVAAPRLWNQLPPDITS ASSLSIFKSRLKTHFYSIAFPGF	MLTGLSEDLRGIDDSRKTAVINNELLRLNVDI AALQETRLADSGTLKEKDYTFYWQGRAPDEP RQHGVAFAVKNNLLSMVEPGRNGTERLLTLR LNTTTGPLTLVSVYAPTLNATLETKDKFYGNLT SVINNIPDKEQLVLLGDFNARVGANHESWPS CLGKFGIGKMNENGGQRLLELCAFHNLCIANSY FQTKPQHKVSWRHPRSKHWHQLDLILVRRS AINCVLHVRSYHSADCDTDHSLVCCKIRLNPK RFHRLRKQGNPRINVSRMSQPDLTQKFAETL VKELDTAQTGDSALEMWESLRNTMQRTALA TFGKRTTKTHDWFEAKASTMIPCIEAKRAALT EYKRSPTQKNLQILRSTRSKAQHIARYCANEY WQELSNDIQKAAIAGNIRGMYDGIKKALGPT QCKTAPLKSSTGEIISDKQQQMERWVEHYSD LYSRQNTVSSAALDVIKCLPTMEELDEEPTAE ELRKAIDKLASGKAPGSDGIPPDLLKQCKDSLL	650		
				TAGGATACGTCATCCATGG TCAACACTGACCGACGGA GGCCTACTA				
			GGTTCCAAGGTTGT CATAGCCGGGGGC GGTAATGGGGATA AGCTCCCACTATCT ATAAAGTACCCCTA TACGGCGTGCGTCT CAAATAGCCTCTGA CAACCAAGTCCAGC TCCTGGCCTTCAAG TGTGGTTTAGCTAC CAAACCCGGCGGA ACTGTTTTCACTGA CAGGAGAAGGGGGC GCAGGCGGGTCAC TGGCGCCTTACAAC CAGTTGCTTCGGG					
		Danio rerio						
	GCA_008692375.1							
	RTE-1_DR							

					HPLHKALCQCWKEGAVPQDMRDAKIITLYK NKGERSDCNNYRGISLLSIVGKVFAKVILARLQ KLAERVYPESQCGFRAERSTVDMIFSLRQLQE KCREQQMPLYISFIDLTKAFDLVSRDGLFKILP KIGCPPKLQSLIESFHTDMKGTIQFNGSCSEPF SISSGVKQGCVLAPTLFGIFFALLLRHAFGSAT EGIYLRTRSDGKLFNLSRLKAKTKVRETLIRDM LFADDAAVTTHTQEELQSLMTRFSMACKDF GLTISLKKTNVLSQDTATPPTITVDDYQLDVV HQFTYLGSTITDNLSLDAELDRRIGKAASTLAR LTTRVWTNHRLTTATKMAVYNACVISTLLYG SETWTTYARQERRLNTFHLRGLRRILGITWQ DKVSNVKVLTRAGLPSMYTMLRQRRLRWLG HVCRMEDGRIPKDILYGELSSGKRTTGRPFLR FKDVVKRDMKALDINTKSWEDLAADRLKWR CTLTKQLKSGEKNMMRASEDKRVHRKVQSV SQGATYQCDCCGKKCLSRIGLFSHQRRCLRQ SVMPQYKS		
				GAGATGATATTCGT TAGCCTGGGAAGG CAGATCATCTAGG GGAAGGCAACCCT GTTTTCAAACCTCC GCTGCCTTGCGGCT ATATCCATTCATGG AAAAGGCTTCAGG AGTAAACCTCGAG GAAAAATCCGGAG TCGGAGTCCCTAAG GCAGTTTAACGCTG TTTGCAGCCTCACT CTGGCAACTCCTGC GACGGCGCCGATA CCAAACTGTAGCAG CCCTGCTGTTCCTTT GGATTTGTCGACAA CGTGGAGAGGGGGG GACCCGCTACATG GGCAACAGCCTGT CCTCCATAATACAT TGCCCTGGCTAGTA TCCGATCTCGCACA CCCTGGAGAGGAC ACTCCAGCCTCGCT AGCACTCTGGCGT GGATACAACACGG GGAGCAGTAGTTT ACCGGTTATAAGCC ACAGCTCGGTTGG CGTAGAGCAAGGC GCCAGGGACTGCT CCCGACGGTGGGA GGGATCTTCGGGT CCCACTGGACAGTC ACCGCCCGCCTCAA GCTGGGCAGCCCC			

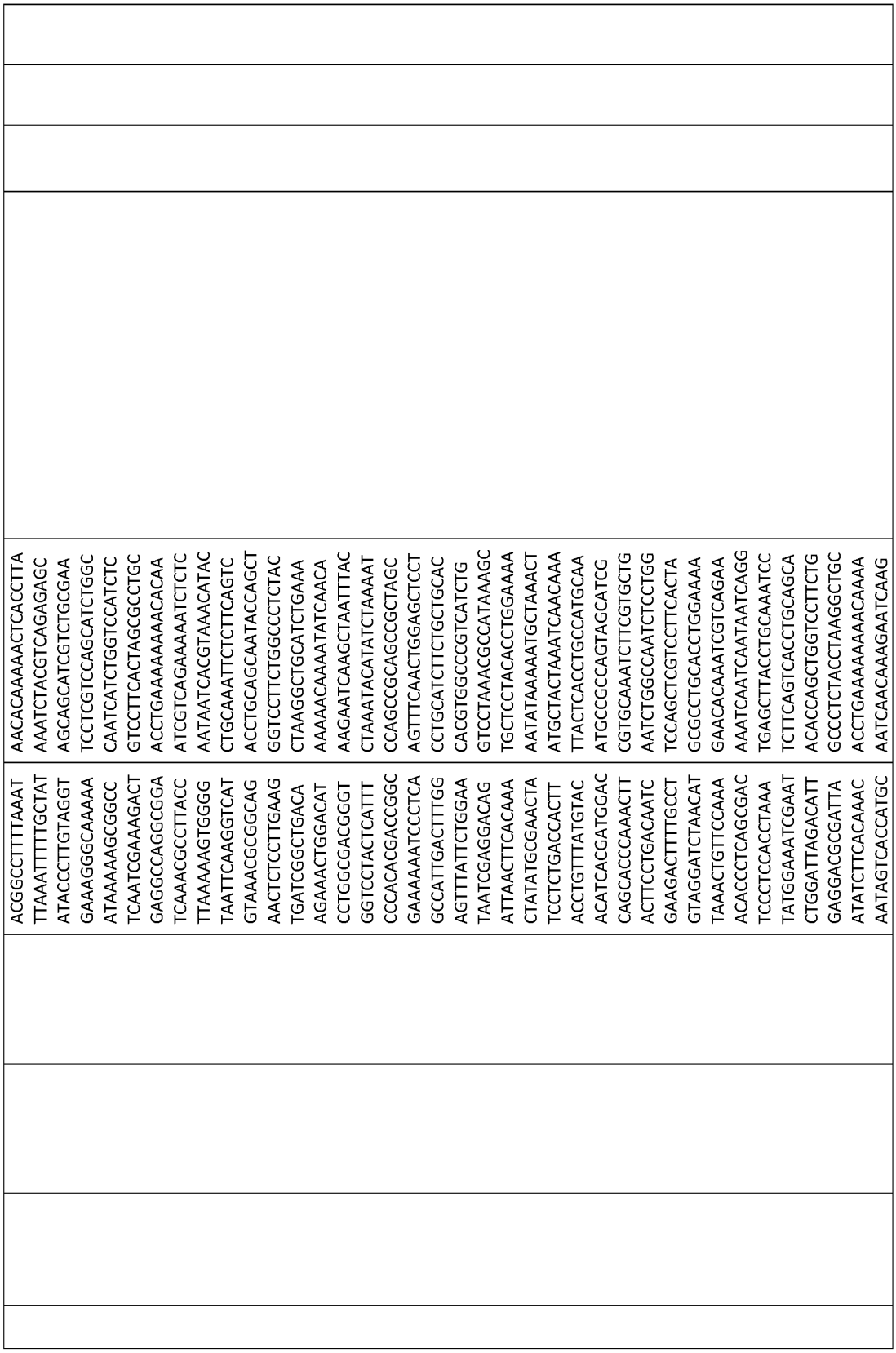
							834
							581
						cd01650	
	Rex1-26_DRe	GCA_008692375.1	Danio rerio	CAGTCAATTAGGTA CTGCCCCGCCACAG TCTGCTTTCCTCATT GGGTGCATGGGGA ATAGGAGCATTTCG AACAGCAGACTGC AACCATCGCACCAG ACAATAAAATAACA CAGAGAAAGAGAC CAGCTCTAAAACTG GGATGCTGGAATG TCCGTACA	GATGCAGGACTGCCAGAC TCAAGAACTCTTTCTTCCC ATCAGCCATCAGACTTCTA AACAGATAACCAACGCAC ATAACAGTCTATTTTTCTG CTCAGTACAACACTACATA CTCCATCCCATTTTGCACT ATTTTATTTTTTATTCTTTTT TTTTGCACATAATCCAATT GCACTAGGCACTTTTTTGC ATAATAAGCACACAATAAAA AACAACAAAAACAAAAAA CACTGTACACTGTTTACAT CTGTTTCATTACTCAAATC AGGTTTACATATTGTTTAC TTTCATAGATATTTATATA CTTATATTTTACATTCAATC TTCGGTCACTTCTGTGTAT ATATGTGTATTTATGTGTA TGTTGTGTATGATGTGTAT GTTGTGTGTTTGACTTCAC TGTGGACGGCAAAGTAAG AATCTCATTGTACAGGGA GACGTGTTTCCTTACTGTG CACATGACAATAAACAGTT	MSWRPSWQTCMLAPHINVVFLLFFVCFLSFP RCIHAMHYQHIADKLAGSCNLRQTFRIYQFYS TGLTTHCPLPQPLQRRRRKRHKRGKRGGLHA RLKSRANRPPLPSILLANVRSLDNKLDELRARI TSQREVRECCALIFTETWLSEKVPGTAVQLQT HSLHRGDRTTASGKAKGGGVCVFINNSWCG DVQTVHKHCSPDVEFLLLKCRPYYLPREFTAV FIAAVYIPPRANATAALGKLFNVFNAQEMAH PDAVIIAAGDFNQCNLRTVLPKYHQHVSIPTR ENNTLDNVYSNIRGAYRAAPRPHFGQSDHIS LFLYPAYRQRLKQTNPITKLVKIWNPQTESTL QDCFALTDWDVFKTAATKEDLSVNVQDYAE YVTGYISTCVENIIPTIQVKKFPNQKPWINSKV RHMLNARSLAFTSGNENEYKSAKYGLRSAIKE AKRQYKEKLDNTYSTASAGQMWRGLQHITD YKSTTNVTISASESLPDDLNTFYARFESSSINTE QRHTQTQTSQPALPPPVVTPAAVHKALRKIN PRKAAGPDNIPGQALKACASELVDVFTSIFNL SLSQSSVPTCFKTTTIVPLPKKSPLTCLNDYRPI ALTPIIAKCFERVVLPLIQSSIPDTLDPLQYAYR TNRSTSDAIAAALHTSLSHLEDKDTYIRMLFID YSSAFNTVIPHNLTHKLSELGLHPTLCDWLLD FLTGRLQSVRIGNKSSSTILTNIDTPQRCVLSPI LYTLFTHDCVASHKNNTIIKFADDTAVIGCITG GDEAAYRKEGASLVTWCDNNNLTLNTDKTK EMIVDMRKERKTHQPLFIRKLEVERVSSFKYL	cd01650	581	834

				GAATTGAATTGAATTGAAT T	GVHISEDLTWTLNTTQLVKKAQQRLYFLRKLR KLGLSSKILSNFYSCVVESILTNCITVWYGNAT VKDRKRLPRVVRTAEKIIRSPLPSLQTIYHHRV HKRTASILKDPTHPQHGYSHSYPQGGGIGV			743
	RTE-2_DAn	GCF_0032859 75.2	Drosophila ananassae	GGGGAGAATGGG GGAGGTTATATAA GGTAAAGTGTCTG ATTGCACTCAACTT GGTGAATGCCAGC ATGATCCTTATCAC TGCTCTCCCCAAGC CGACGCACTACCTC CACGTGGTTCTCCC TGGATTGTCACTCA ACACCCGTTGTTTG ACAAACTAAAGCG CGCGACGAGAGTC CTAGCTCTTCAAAC TAGGTTGTGGCGT AGGACTTGAAATCC ACCCACGGAAAAC AAATCTTCAGTGAA TGTAGCAGATGAA GCCTCGGAACGGA ACCAATCTAAAGTT GACGACCATGGCA CACGTAAAAAGGA CAATGATTTGAGA GTATGCACTTGGAA TGTACGCACATTGT ACAGACCAGGTGC TGCTCAACAGCTAG CAGATACCCTCAAT AAATGTAGGGCTG ACATCACTGCTATC CAGGAA	TAGGGTTGTGATGGCCAT TTATATATTATAT	MRWIGQGCLRKKNCDIYYSCHPERHEFGCGF VVGARLRRRVSHFLPINERLATIRITAKFFNISLI CAHAPTEEKDDVTKDAFYAELDRAYGRCPSH DIKILLGDFNAKVGREDIFGATVGQFSLHETTS SNGLRLIGFAAAHNMVVRSTGFRHLDIHKAS WLSPDRLTRNQIDHVVIDARHASNVLDVRSC RGPNIDSDHYLVAAKIRTRLCVAKVARRGLLR KLDIGKLQSQRTRDAFSTQVTGLLSRATPTPE DISNMWALISHSLRTSAEEVIGFQRTPKRNP WYDQECRDATAAKDAAYGRTLQAGATRAV VEVYRSRRRDEKRLFRRKKREQERRECESIESS RDRNEARNFYQRVKRLTQGFKTGGVACKDD DGNLITDAEIVLRLWRDHFSSLLAGSGHDDY GEYDPQTPIYGTDVDVPTPSHAEVKDAIQRLK NNKTAGADGLPAELFKAAGDMLVGSMHKLI SKIWLTESMPDDWNHSMICPILKKGDATLRT NYRGISLLPVAYKVLTSVLCENLKPHAEALIGP YQCGFRPGKSTIDQIFTLRQILEKSYENQIDTY HLFVDYKAAFDSPRRDRLYAAMSELGIPAKLI RLCMMTLSNTISSVRVGNDQSETFYTKCGLR QGDSLSCMLFNILMESIIRKAGVHRTGTILTN RCVQLLGYADDIDIIGRTKRDVTAAFGAIEKES AKVGLAVNMDKTKFMVCSSRESRRLDSQLSA GSHSFESVKEFIYLGSAISSTNDVSLEIKRRITLA NRCYFGLSRHFNSRALSRPTKITLYRTLILPVLL YGAECWTVVQSDAAALGVFERKILRKIFGPIC VGDAYRTRWNHELYELYGDVDIVSRVKIQRL RWLGHIARMEEDAPARKVFDAVIVGKRRRG RPRIRWQDQVMEALSTSGVTNWRARAQDR KAWRQIVQQAVTR	cd01 650	490	
	Jockey-14_DAn	GCF_0032859 75.2	Drosophila ananassae	AG	TGATACCACACCTGACCTA GTCTCCATTACTACCTCAC	MQDIDDAIEIFNRNIYHAAKIATRHIKTSNRN PTGQFLLDRKLLNLQAIKKYTKQIKMARQTPE	cd01 650	268	540

					AAKNYKAANERFLRAIKEAKTKRTNNILNKID	
					PLDRYRMQKLWKITNHFKRQPEPNWPLKLY	
					TGDGTSAKDWYSWTKSSEEKTEAFTRHLEER	
					FSPKLTNSESFRAHVISKKEEICTQKLLMTTPC	
					WMHRENYGHNSDKTNLPFLPITPNEVLAEVK	
					LLQNKKAPGPDQIDNKLIKLLPHKATLHLTHIY	
					NCMLRYSYFPKQWKRAAITMIKKPGKTPHEV	
					SSYRPISLLPGFSKIFETLLMHRMFKIQDFERAI	
					PSHQFGFRKQHSTEQQLLRVTQHILSAYEQK	
					MYCSAVFIDVSEAFDRVWHDGLLLKLAKLLP	
					HSLYKILQSYLTDRSFTMNSTNGIPSRTGKISA	
					GVPQGSVLGPILYTVFTSDMPLPEVISSQHTY	
					NPSYFENMMLSTYADDTIILCTKREPQLAIQT	
					TESFLKYITDWATSWCITLNPNKTIHVLFTKRS	
					VANLTGNKSPKMKGILIKNEPKHRYLGLILDR	
					KLLLHHHINQLQNRIRNVAAGLNWIIGTKSKL	
					SPQSKILLYKQIIAPIWQYALPLWGALTSNTQF	
					QRIEAAQSKLIRKILKAGRYTRNQVLRDAYNIK	
					SINEIYSEASNRLATSLASHPNTEARKILMDPV	
					TPNRLEMPRYLDQLNRFFLPLRQQLQQQNQ	
					LNTHQPALPTLISYEQQLLSQQASIKARQQDE	
					EREQNKRLKIEQDPHYRPSIGWVNAMSYKLR	
					HGITTRWEMADKIKYEPEYIQRLIMSPLLPSEI	
					PPPTPDQELDARLSTLKTQIINDLDLNLNYETN	
					LAEGDSRTRFLPNSSSPILQAPQIPCINLVEDH	
					PLGNPAIADIWKSTPEI	
			AGATGGCCATAACTACAG			
			AACCCGACGCCATGACAC			
			AGCAACGCCCCGCCAATA			
			GCCGTGCAGCAACAGCCC			
			ACTAATACGACGCACAGA			
			CGCCCGATCCAGCCGCCAC			
			TGCTCCGATTACAAAAACA			
			AGCGACTGCGTCCAGGCA			
			GAAACCCCGCTACGCAGC			
			AACGGCACATCAACACGA			
			CGCTCAGCCACCCGGACC			
			AGCCGTCACCACTCTGGCC			
			ACCAGACCAAACCCCCGA			
			CGACGCCCAGCAGAAACT			
			CCGCCGCGCAGCATCGCCT			
			CATCAGTATGACGCACAAC			
			CACCTGCATCAGCCGTCAC			
			TACTCTGGCCTCCAGACCA			
			AGCTACAGCAACCAGGCA			
			GCAAACCCGCCACGCAGC			
			AACGGCTTGCCAACACGC			
			ACACATGATCCTGGAACA			
			GCCGCTGCCGACTTGGCCT			
			CCAAAACCAACAACGGCG			
			TCCAGGCAGCTACATCGCA			
			GCGCTGCAGCAACGGCCA			
			GCCAACACGTCGCACAAG			
			AACCTAGAAAAGCAGCCA			
			CCATTCTGACCACCAGAAC			
			CAACAATGGCATCCAGGC			
			AACCACCCCGCACAAAGC			
			ACAGGCCCACCAACCACT			
			GAAGAAAACACCTGAAAG			
			CGCCGCCAATGGATTGAC			
			CACCAAACTCCCGACGAC			
			GCCCAGCAGTAACCCCGC			
			CGCGCAGCAACGGCTCAC			
			CAATACGACACACTGATAC			

	R2_DPe	GCF_0032860 85.1	Drosophila persimilis	AAGATATGGGTCT GAATAATAGCGTA GAAGGGGAGTCAT TCCGTAATTCGTAA ATCGTAAAAATCAG ATCAAGTTGATTCA AGACCTCCTCGTGG TATCTTCTGGATGC TATTAGACTGA	CTGGACCAACCGCCACTGC CCTGGCCAACAGACCAAA CGACGGCATCCAGGCAGC AACTCCGCCACGCAGCAA AGGCTTGTCAA	MSSFGLIVTNLNSETVLWGCQPLGQFSLIGTN MQNTTPRIINTNSLTNQIPTVSSLGAQSEHSA QVNPNSGYQCTICESSFRSKSGLGVHMSRRH KDEFDQLRLRTDRKAQWSEEELSMMARKEIE LAANGERYLNKKLAEVFTNRSVDAIKKCRQRE RYKTKIEQLKGQAVPLPEALESETIQRRPSIRER DLLVTPPNTLGTTPTELSNSEILAVLQGYPPVV CNDQWRVEVLQSIVDGAQASGKEITLQRLST YLMEVFPSQNDRPIQTRPPRRPRNRRQGRR QQYALTQRNWDKHKGRCIKAILDGTEGTAT MPSQGIMGSYWRQVMTQTSPTYSGTNTTF RTEHPLEGVWSPITLGDLRVHRVSLTKSPGPD GITPRTVRSIPSGVMLRIMNLILWCGKLPVSIR QARTIFIPKVGNASRPQDFRPITVQSVMVRIL NAILASRLTSSVDWDPRQRGFLPTDGCADNT TIVDLILRDHHKRCKSLYIATLDISKAFDSVSHA AVSATLTAYGAPKEFVDYVQNSYEVCGTTLN GDGWRSEEFIPARGVRQGDPLSPIIFNLIIDQL LRSYPNEIGATIGDHTTNAAAFADDIVLFAETR LGLQTMLDTTVDFLSSVGLTLNSDKCFTVGIK GQPKQKCTVVIPETFRIGSRSCPALKRTDEWK YLGITFTAQGRTRYSPADDLGPKLLRLTRSPLK PQQKLFALRTVLIPQLYHKLTLGSVMIGVLRK CDILVRSTVRKWLGLPLDVSTAFFHAPHTYGG LGIPSVRWVAPMLRMKRLSNIKWAHLAQSE AASSFLTDELNKARGRTLAGLNELTSRTEIETY WANRLYMSVDGRGLREAGLFRPQHGWVCQ PTRLLTGQDYRNSIKLRINALPSRSRTTRGRNE LERQCRAGCDAPETTNHILQNCYRTHGRRVA RHNCVVNNLKRILEEKGHTVHVEPSLQLETSV SKPDLVCIRDNHACVIDAQIITDGLFLDDVHH RKVEKYKRPEVISALRREFGVSGNVEVLSATL	cd01650	411	673

	TART-1_DWi	GCF_000005925.1	Drosophila willistoni	AATATTGGCTTATT TTACTTAAATATTT GGTCTCTTTATGTT TTTCCCCACTAAAA GCTAATTGCTCTAA AACTTCTAGCCTAT AAAGCTAACTCAAC TAAAAGCTGACTGC TCTAAAACTTTTAG CCTTTAAAGCCAAG CCTTGCTCGCAATT GATGCGCCAACAG CCAAGCTAAAAAAC TAGTGTTACTCAGA CCCAAAATATGGGT CTCCCTACTTTTTTC TCCAGTGTGGAGG AAAACCAGTTTTTT CCCCGACATTAATC GCCACGTTTTTGTA GCCGCCTTAGCACC AAATTCTTATTCGC CAAGTTTATTATTA AAAAATTGCTACCT ACTTGACTACTTAT GCTAAAAGTAACAT AATCTTAAATCTAT CATACTACAAACTA TCTTACCTAAATCG CTAAAAACTAAATG TTAATCTAATGTCT TATTGTATATTTTTT ATTTAGTATTAATC TAGTTTTAAGATAA TTTTAATCCCCCCG CCACTGCTGCTGTA	TGAGGAAAAAATCAGGGA CAACCACTTGTCATTCAAA GGCTATCCTGCCCCCCTTG GGCCAAACGAACAAATTC AATACACGGAAAATGTAC TACTCGCTCTAAACATGGC TAGTCTGGAATCTGAGTTG CTCAAATTAATTGATAACA GCACGAATAGCACTGATG CTCTAAAAAAAAAGGAAA ACTGATACTGCACCCATGA TGGCCCTAACTAACAATCA CTCTAACACCAATACTAAT ACTTCTGCAAACATTCACT TACCTAACCATAAAGAAAT ATGATTTTAGGCGGCCCCT GCAAAGCCAGAATCCTGC CCACTATTCAGCTGCCGTC CTGGCCTGGCCAGAGGGC CAGTCCTCGACAGCTGTGC CTCCAGTAGACGCGCCTCC TGCACTCTCTGCCTCGTTC TCAGCGTCTGCACCATCTC GTCCATCGTCCTGTCCAGA TTTCCGCTCGCCTGATGCA CCAATACCTGAAAATAAAA CAAATCTAAAAATAAAACC AACTAAAAAGTACAAATA CTCACCTGAAAATACACAA CAGCAAACAGTAGCTCCTT GTCCAGCGTCGTATGCATC GTCTCATCTGCGTCGTCCA GAATCTCGTCAAGCTAGTC CGTCTGCAGCTCCTGTATC TGAAAATATATTAAAATTA TTTTCTAAATCAGCAATAA	NWRGIWSNQSVRRLIAKGLISSGDSNVISARV VTGGLYCFRQFMYLAGYTRDWT MNILNKEISLSKNKHLNKKLCKIDSSDRYRPHK LWQLTNKLKRQPLPNWPLRIPSGVDKKHAW TRSPEGKMELFTDQLEKRFSPNLTNSSLDRAK INEETNAILAQKMLMEYACKTHHSIASSSANF DSSNSPFVKPITLSELKKEIDSLINKKAPGYDRI NNALIKVLPLKAIIFLLKIYNSILRFGHFPLVWK KAIVVMVKKPKKTISDASSYRPISLLPAFSKILE SLIAKRLLKFNRLSQSIPNHQFGFRKFHGAEH QLARVTQFLQRAFEDKLYGSPIFIDITEAFGKV WHCGLRNKLIKLLPINLYKVLDSYLKDRTFSVR GNHGIMSRDGKILAGIPQGSVLGPLLYNIYSS DMPIPKFIIEHETLIEYYPTYKHGDTLLSTYADD TIFISTDKNALVAMNKTENYLHKFSTWSNKW CFTVNVSKTVHVLFSRRSTSPAVKAATVKFNN KPIKNTSSCRYLGVLVDSKLLFKTHIAKLCAQA KSTYTRYKWLGGAKSHLSTRCKVTIYKQILVPI WTYAIPIWGSLISNTLFKKLTGTHNKLLRKLTN ANIYCPNSEIQEKHNAKPLETIYNDACLRFFNS MSDHETKEIEKLYLKPFIPKRLPWPRYSKQIDK LVASRTIKKAQTPPLMPPLNPSVTVTNIKLRIE SKLMEFFSKAPALIADNCHQTYNRITEKQDVL RAKEASIARKNRQKIQSGH	cd01650	195	471



				ACGGCCTTTTAAAT	AACACAAAAACTCACCTTA	
				TTAAATTTTTGCTAT	AAATCTACGTCAGAGAGC	
				ATACCCTTGTAGGT	AGCAGCATCGTCTGCGAA	
				GAAAGGGCAAAAA	TCCTCGTCCAGCATCTGGC	
				ATAAAAAGCGGCC	CAATCATCTGGTCCATCTC	
				TCAATCGAAAGACT	GTCCTTCACTAGCGCCTGC	
				GAGGCCAGGCGGA	ACCTGAAAAAAAAACACAA	
				TCAAACGCCTTACC	ATCGTCAGAAAAATCTCTC	
				TTAAAAAGTGGGG	AATAATCACGTAAACATAC	
				TAATTCAAGGTCAT	CTGCAAATTCTCTTCAGTC	
				GTAAACGCGGCAG	ACCTGCAGCAATACCAGCT	
				AACTCTCCTTGAAG	GGTCCTTCTGGCCCTCTAC	
				TGATCGGCTGACA	CTAAGGCTGCATCTGAAA	
				AGAAACTGGACAT	AAAAACAAAATATCAACA	
				CCTGGCGACGGGT	AAGAATCAAGCTAATTTAC	
				GGTCCTACTCATTT	CTAAATACATATCTAAAAT	
				CCCACACGACCGGC	CCAGCCGCAGCCGCTAGC	
				GAAAAAATCCCTCA	AGTTTCAACTGGAGCTCCT	
				GCCATTGACTTTGG	CCTGCATCTTCTGCTGCAC	
				AGTTTATTCTGGAA	CACGTGGCCCGTCATCTG	
				TAATCGAGGACAG	GTCCTAAACGCCATAAAGC	
				ATTAACTTCACAAA	TGCTCCTACACCTGGAAAA	
				CTATATGCGAACTA	AATATAAAAATGCTAAACT	
				TCCTCTGACCACTT	ATGCTACTAAATCAACAAA	
				ACCTGTTTATGTAC	TTACTCACCTGCCATGCAA	
				ACATCACGATGGAC	ATGCCGCCAGTAGCATCG	
				CAGCACCCAAACTT	CGTGCAAATCTTCGTGCTG	
				ACTTCCTGACAATC	AATCTGGCCAATCTCCTGG	
				GAAGACTTTTGCCT	TCCAGCTCGTCCTTCACTA	
				GTAGGATCTAACAT	GCGCCTGCACCTGGAAAA	
				TAAACTGTTCCAAA	GAACACAAATCGTCAGAA	
				ACACCCTCAGCGAC	AAATCAATCAATAATCAGG	
				TCCCTCCACCTAAA	TGAGCTTACCTGCAAATCC	
				TATGGAAATCGAAT	TCTTCAGTCACCTGCAGCA	
				CTGGATTAGACATT	ACACCAGCTGGTCCTTCTG	
				GAGGACGCGATTA	GCCCTCTACCTAAGGCTGC	
				ATATCTTCACAAAC	ACCTGAAAAAAAACAAAA	
				AATAGTCACCATGC	AATCAACAAAGAATCAAG	

								816
							619	
						SSF5 6672		
				TTAAATTAACTAAATACAT ATCTAAAATCGAGCCGCA GCCGCTAGCAGTTCCAACT GGAGCTCCTCCTGCATCTT CTGCTGCACCACGTGGCCC GTCATCTGGTCCTAAACGC CATAAAGCTGCTCCTGCAC CTGGAAAAAATATAAAAA TGCTAAACTATGCTACTAA ATCAACAAATTACTCACCT GCCATGCAAATGCCGCCA GCCGC	TAAGTGTATTTCATAGTTT ATCAATGTTTATGTTAATA ATAAATAATAC	MPLFKSLDIWKYVNLTLLPPLPPSSSSPPLPPS SSPLPPSSSPLPPSSSPLPPSSSPLPPSSSPLPPS SSPLPPSSSPLPPSSSPLPPSSSPLPPSSSPLPPS SSPLPPSSSPLPPSSSPLPPPSSSSSRPHPGPRK KRNKQKKRTGPRKCPPPTSSRPLPHTTEGLNP VSDDLFCPVSGCPGPSSSATDEWIECSACLLW YHTSCLLLQAIPEDDWFCANCPQIHTQIGPQ VNSNTPSPANQSTLGTHSQIIQSSPCTNPRAS QPISTLSTPLSNFISSKATCTSLTLEWLNNNIVP GGIIEKNLLHEAYRQDSLGIFKPLNDAEIGKLIK NSFSNIRTKRLRTEFHGASKYHYVGIQLKSEQ VPETVFDFKSIISSLRQNCPTLKQCPKACRIPV AQSIKSAIDKCITENTPEAWHLLLTFPYVTLQC DSRANIKNGQSLNNKIRNNLSIFESLDKNTYQ DYLISLSSKNSVFDKSGSKINDIGPKVISKVSEG NIRGAIRLLTSDDVVAPNSSATLDTLVEKHPQ QSLDNFTVLDEPNIPPFTTTTEQVISAIFSFPSG SAGGLDALRPQILKDLVSQHTGDVGTQLLESL TKFCNFMLAGKVPLDICPFIYGASLTALKKKSG GIRPIAVGCTYRRIVSKICVKVMNHVFQDYFS PVQVGFNVKNGAEAGAHAARRFFSFPHQSP TAFLKIDFKNAFNMVDRGKLLSSVHELVPTLY PYINQCYRFSSSLVWNSSLIDSCRGVQQGDPL GPPLFCLTLHPVISNLGSEFNMWYLDDGTLG GNPTTVLDDFQHIISCANEVGLELNFDKCEVS	816
	CRE-1_FoCa	GCF_002217175.1	Folsomia candida	TGCAGCCCTGGCCA CTAAACCTCTACGA AGGGACCTGCCAA ATACGAATCGTACC GAATCGGAGGAGC CATACCTATAAATA AAAAATTGCACAAA TTATTAAAACAAAA AAATTAAATAAAAT TGATATAATGCCTC TCACGAAACTTTGA ACGTAAAAAAACAT TAA	GGTAACTTACACCG GAGATACTACTTAT CTATCGTTGGAAGT CCCTCTACTTTCACT CTTGAATTCCGCCC CCGTCTATATATTTT AAGCGTGGAAGGA AATTGCCAGTATTG TTGGATACAAGTCG TCAGACCACG			

					VLGPQTNTQCILESFQSVAPGISCMTQENAEL LGAPLTLESIEHVFRKKIHAFTNMAGRLHSIPS HIAYYLLKHSLAIPRLVYLLRAAPIFKSPHLLEEF DSLLHNALEQITNCQINPLSLEEIALPIKLGGLG IRQSKTLATPCFLASSFATTDLVSLILPSYVTALS DSLVTEAQDLWKSISCEELPRSTITSFQSSWES PIFNYVSSRILESSTTPINTARILASRSKESGAW LQALPSRNLGTLLDDNSFRIAIGLRLGLNLCFP HKCVCGAQVDELGLHGLSCKKSAGRHSRHT QANDIIKRALISADIPSIREPLGCCRTDGKRPD GLTLIPYKQGRSLIWDFTCTDTFAPSYLSHTLR HAGTAAKLAETKKHKLYSELKTNYIFVPVAIET SGVWGTEGLKFIQDVGRRMKDVTGPMSTTF LIQRISLAVQRGNAASIMGAVPSQRGLEELFYI LNDRSFYLQKSCKGPNGDLGRLSGSSKGFPN PPTSGFSSTPQAN	MIILSYNSRGLGRGVKWAAVRRLIVKHKVDL VCLQETKRENFNKSICQAIWGDSTGHWDFV PSVQAAGGLLCMWNNSTFVVERKEKGRSFL MLEGTCCSNNQKLIFVNVYAPCDLVGKKAL WDELRQLKASNPSGMWCFLGDFNSIRSSQE RISLSQRRADPYDIAAFNQWIDDMELQEIKCY GNSFTWIRPNGCVKSRLDRFLVSQNWLSLW PESCQTVLQRNLSDHCPSILQTNMVDWGPK PFRVFDWWLQQKGYQKMVREAWKKDQQ GGWGGIALKNKLKNLKSVIKCWSKAEGNINP KKVLNIQQKLNEVEDLAAHRNLTDQEIKDRTL LQQELWNASNAFESLLRQKSRARWLKEGDY NTAYFHKYMNFRRAYNNIPGIFIYGEWIQQP DAVKNEAANFFHNRFTEQMHSRPTLDGVHF KSISQRQREQLTVPFSDQEIKQAVWSCGGDK CPGPDGLNFNFIKEFWDVLNPEFFRFVDEFY AHESFPRGSNASFVALIPKKSHPQSLNDYRPIS LIGCMYKVIAKLLSNRLRSVMDELIDERQSAFI KERHILHGILILNEVVEEAMRSKKPAMIFKVDF EKAYDSVSWSFLDYMLLRLGFCQKWRRWIS ACLHSATISVLINGSPSKEFNPSRGLRQGDPLA PLLFNIIGEGLTGMMREAVHKNLYRSYLVGKK	cd01650	497	762
	L1-16_GMa	GCA_012273815.1	Glycine max	GTTTTGGGAAGTA ATCCCAAGGCCCAG CAGAGTGACAAAT CAAAGGAGATGAT GAGAAGTAATACC AAGGCCCAGCAGA CTGAGGAACCTTTC GTGGACTCCAACCC AGCTTCTAAAGGTC ACCCTCTGTCTTCTT TATACGAGGACCCT ATACAGCTGCAATG TGACATCTTGAAGG AAATGGGCTTGTCT TATGAGGAAGAAG ATAGCAAGGTTAA GGGGCTGCTGCTG GATATGGAGAATA AAGATGATTTGATG GCAGCAGAGAAGG GAATTAAATCTCAG CAATT	GC			

					REPVNILQYADDTVFVGEVDWQNVHAIKALL RGFELASGLKINFAKSQFGIIGGGVNWGLAA ANILNCSQMNYPFHYLGIPIGANPSSLMVWE PLITKFKAKLSKWAQKDISMGGKVTLINSVLN ALPIYLLSFFRIPQRIVQKLISLQRKFLWGGDN THNKIPWVKWADICLPKNEGGLGIKDISKFN VALMGRWIWAFASGQQQLWVRVLISKYGG WPEFQHDTDKRGFSHWWRDLRKIFHQSDH NIFKHQLTWKVGCGETIKFWTDKWLGEDYTL EQKYNQLFLISRQQNSLISNMGEFNHDSWE WNLRWRRNLFDHESDL	MDRSLKILQHNLNGLRARLTELIKFLSTARYDV LCFQETKLYPGVKDTILPGYRLSRLDRPGAHSK CFGGGLAIYVRTELSYQTLVLHSTEFEVNGVSL SLADGSKIAIVNVYLPPGKNVIEPSSYRNLCEQ ISHRNVLFCGDFNSKHNAWGSYKQDERGEG LLQFIEDNDMCILNDGRPTLIHNSSGVMSHL DLTIATSNIASNCEWQVSSTETLGSDHLLIDIT YRGQIKREGTVGKPKWRIAKADWIKFTENCS QWDVKDIQDADVDVYANNVTMVLMEAAK DAIPRTSGKNRACPKPWWSDACQNAVKLRK KALRRVQRNLNTETKRTYLEATENARKVISEA RTSHWHDFCQTLNQHTQIRDVWAKIKSMN GTKSTREIPVVTYGDRVGESDRQKADLMAES FADASSDKHHSPEFMNRKTNFAVDHSNVLT HIDENDPLNLPFTSDELQKALSSKKDSAAGED EVKYSMLKKLPPEASKILLDFFNESWTNGHVP TKWKTASIIPILKAGKDPNKTESYRPVSLTSTLC KILETMVTHRLRYYLESNNLLDKNQSGFRKLR STADHIVRLQHAMSKTLNSSGFLCAVFVDLTK AFDLVWTDGLMHKLRQLGIGGRMHNWIAS FLADRKIRVTIGKSLSEEFSLENGTPQGSVISPT LFSVMVNDIRNPKYLSNKCQIAMFADDSTFW GNGTLYTNLQMRIQTCLDRFAKWANDWGF RISKAKTVGMLVCKDPQLTDMRLVLDGERIQ FVSETKLLGVTFDRKLSWEPHIKNITNRCQIVL NLMRNISGNEWGANVVTLMMIYRALIRPILD YGSQAFMTARPMLLHKLDVIQHKALRLATGA	cd01650	506	765
	Nimb-3_LG	GCF_000327385.1	Lottia gigantea	GCCTCGAACGGAA CAGGATTAGACTC GTATCTTTAGTACT GTCGCA	TAAAGTAGCACATCAGAA GAGCGCCCCGCTCCTCTGT AGGTGTGGCGATCATTGG GGACTTAGTTCTCGATAGA TCGCCAAGAAGAAACACA GAAGAAGAACGCAGTATC AAGACTACAGAAGAAGAT CAAGAAGAAACCATGGGA GCAACCCGTTCCCTACTCA ACGTCTGCGATATGGAATT TCTCCGAAGCGGACACCAT CCACCAACAAGAAGAAGA				

					GIGIKTESLLQDVGEVPLQIRREELALRYWSKL KHLGESSPTYLETARFRHAKKTISFGQSVRDSV RKYGLDRLPIKPLVFSQVAPWTVSPPIVDTSLA SRLDKKRDDPCIARMIALTHMDARWSESIRIY TDGSHNPDDGTTGLGIYVPSSSTRISKRLPNN CSVYTSEMLAILYAVKYSAGHYHSNYVVLSDS LSSLTSILNDSSNSQAELVDEIRQACSLVHSQG KSIELAWIPSHVGIQGNEIADELARIGLGLSEID RIDYSLSEIRSMITKSIKSTWEATLIDANAFRFR LNDSISSKVFLHHPNRQICRTVARLRMMGEV RNHKYELKLCPCGEAGTAEHLLIHCEQLRELR KSLVEIMKRYRLEYTLKNILSHTGEGMSTICET VYNFLSESRMIT	cd01650	711	962
	CR1-6_NVi	GCF_009193385.2	Nasonia vitripennis	T	TCAGATATATGTATCTCCG TATTATTATTATTTTTGTAC ATTCGAATCTAATGAACTA TCTGTCACTTGTTGTATATT TTTCAAAACTTGTATTTATT ATAGTATCTTTCTAAAGGA CGCGTCAGCCGTTAAATTA TTAAATCAATAAATATATA AATAT	MTYPREGCTGPRDGLSKQEELTAGYSGFSNA WTLTRIIMSHRAWTVELFLLNCTIGPIDPSTSL CTTFLSKATTSRCHYWSVPLLVLMLMKSLDA VFPGPQETALFLVRFGSSDDASRVLKNRKLLP SSVEATADHTAYQRDLYRKLKLKASLHNTSHP DSLKKVVYVKGTPTLVDLRRTSTAPLSEKPSKN LQASETFKKLAGILKPVNKLSLYYQNAYSMYG KLHLFKAKLFQLSILPDIVVITETWLQDCVLSSE LGMQQFDVFRCDRDLSSHGVQRGGGVLIGV KKFLQAELVHVDNVIEQLIIRISLPSHKLIVAAS YLRPDSSVQVFSSHVSNLEDANIKFPDHRMV LLGDFNLHHITWSSDTSAYSPMAYLSPALRSA ADIISGYSTSMDLVQHFKHHPKKGYSLDLLFA QKDFISPMDFNEDILPCDDHHVPGMFSCNIS VASSASCNLKFCKRKFYSADFEAIRARLSEED WATVLYSFSDDPDSGIARFYEILNGVIDQHVP ITKQSLSSYPKWYDSELINAICKKKIAHSTWKS SNSKRDEIEFKRLRAVCVRLSRSKYRAYISSVEL GLKRNMRTFWSFIKNLRNDSSIPSKMYHGPT RAASDVEVSNLFSGYFSSVYRTDSLREFESND EPLVSLSNIVLSIETLRSIVANLLDDNINGGPDGI PPYFIKRCWSLLERPIVHIFESMLRSGYFPTFW KYSFVVPIYKNGDRHSVANYRPISTLSCIPKILD AYLASELSNSLLSKVAQQQHGFISGRSTLTNLL VFNDFLSESLCNHEQTDAVYTDMSKAFDSVN			

					HSRLISKLFNFGIRGKLLDLLRSYLTNRSQAVR VNGAVSSPVDVPSGVLQGSHLGPLLFCLYVN DLVPKLSAAQVLLYADDIKLFMRVRSADDSVK LQRDLDTLVAWSVENGLSLNANKCSVISFSRS KSLLLFNYCINGSRLQRVDVVKDLGVIFDSSLT FITANQQTDVVVSRSLRMLGFIKRSTVDFSDP SAIIYLYKTLVLPHFLYCPQIWRPHSQYLINKLE SVRHSFLRYIAFKMGKPLHRFCHDYSEIAEGC DILTVDSTFYLHDCVMTFKILRGLTNCERLSDL FVSRTVAYPLRNLRDLREGTFSSNLGFYSSINR MKRQWNSIPNNVQSKESVGQFKSSIRSLSLA Y	cd01650	412	668
					MESLFIEIINEKSKNTLIGVIYRPPHSNVNDFLN EMELLTTKLNTENKKCYILRDFNLNLINIEYHEL TRQFLDIMYSSSMIPTISKPTRITSHSATLIDNIF TNNINMQDSFSGLLCADISDHIPIFHITKIGKT KQNKEKRNMYTYKRIMNDSCIDNFLGEIDQI NWDNVVLKENVNDSYNSFLQNITYIYETCFPL KKFRNKKINSKPWITNGLLKSIHKKHKLYKKSQ ELPSESNFNKFKQYRNKLNSLLRIAKRDHFTDS LEKSKMDAKNTWKILNSLLKKNIKKPNYPDIF KNNNQEISNPKDIADKFADYFTNIGPNLSKNI ENTNIDIKSYLKGNYSNSLLFRLTTPDEIKEITLS LKNKNSCGFDEINIRVIKAAISILSNPLSTLINQS IMAGIVPDKLKIAKVIPLFKSGEKELFSNYRPISI LPCFSKIYERAIYNQLISFINKYNILNNNQYGFR DKHSTVMAVCDFIEKISSSIDKGYNTIGIFLDLS KAFDTINHNILLQKLQFYGIRGIPLLWFQNYLS NRKQFVNYKDTKSEEYDITCGVPQGSILGPLL FLIYINDVPNATELFSFTLFADDTNLIFKSKSAE DMDEILNSELFKLTEWFKTNKLSLNTKKTNYII FRNSKRKTKHTTKIEINNEPIQMVNETKFLGLII DSQLNWKSHINHVRTKITKNVGIMNKLRFYL PKFSLKTLYFTLVHPYIHYGNINWASNYDTSLK RIFLLQKRCVRIITNSHYLSHTEPLFKEERLLTVY DINTLETALFMYKFDNNQLPGTFDNFFKRNN TIHSYNMRNSCDYRTTLRKTRRGQFSISYRGP			
	Daphne-15_NVe	GCF_000209225.1	Nematostella vectensis	ACATAAACTCTGAA ACTACATCT	TAGATATGATTCTACTTTA TTTCATCAAATCAGCCACA TGTAAATTTAATAATAAGG ATTCCTTAGTCAAATCAAG AACATTCATATATATTTTA GTAGAGACGCTACTTAGTC TGCGACAGGGGTCTACAC CTCATAAGCATTGTGCTTT TAGTAGACTCCTCATCATC ACTATTTTTTGCATTATTAT CATTATACTATTATTTGGT GAAATAAACTTGATTGATT GATTGA			

					KIWSNLTCQAKQCRSLNMFKKIVKDNILSQK AAIDL			
	L2-8_OL	GCA_004348135.1	Oryzias latipes	T	TAGTGGGACTTGTTCATTT TATTAATTTAATTCTTTTAT TGTTTTTTGCTTTTATTGTT CAATTGATTTTATTGTTTTA TTGTGATTGTTTCTTTATTG CATCATATTGTTTGTGATT TTATTGGATTTTATGTTAA GCACTTTGGTTGCCTTGGC TTTTATAAGTGCTATACAA ATAATAAATTGATTGATTG A	MRKYNVSFHCYADDIQIYLPLKSGGAEGLDRL QNCLADIKSWMELNFLCLNKNKTEMFVFGK SDLLTGCCSAVGPLDPSDAPFIRNLGVNLDRH FKFDKQVSLVVQSSFYHLRLISKVKPYLPPHAL QQVIHALIFSRLDYCNSLYSGMDQSLLRRLQL VQNSAARLLTGTRRCEHITPILRSLHWLPVSFR ISFKILLLLVFKCLNGLAPQYLSELLHVYRPCRVL RSEDQMLLVVPRTRLVTRGDRAFSVAGPKL WNALPYSVRSARTVSNFKTLLKTHLFTQAFNS SERI	IPR000477	3	89
	Rex1-5_OL	GCA_004348135.1	Oryzias latipes	G	TAACCTTTGCACATTTTTTT TATTATCTCTTTATTTAGTT TTTTCTATTTTGCACATTTC TTTTGCACATCTCTTTACCT GCACTGTTCTGTAAATAGC TCTTTAAATTATAATACCTT GATGTCTGCATTTATCCTA CATTTTCTTGTTCTTATTTT TATTTATATTATTCTTACTC TTTTTTTCCCTCTGCTGTCA GTTGTTATGGTAACGAACA AATTTCCCACGTGTGGGAT CAATAAAGTATTTCTGATT CTGATTCTGAT	MKLNQHKAAGPDGVSPRTLRTCAHQLSAVL QHLFNLSLRLERVPLLWKTSCLVPVPKKTSPS HPGDFRPVALTSHVMKVLERLVLAQVRPVVS SSLDPLQFAYRPRLGVDDAVIHLLQRTHSHLD GGGGMVRITFFDFSSAFNTIRPSLLGDKLRLM GAGASTVSWITDYLTDRPQFVRLGGVLSDVV LSGVGAPQGTVLSPFLFTLYTSDFQHNSESCH LQKFSDDSVVVGCISGGREEEYRRLVDGFVE WAGENHLLLNVDKTREMVIDFRRKRTVLQPL CILGRDVEAVEEYKYLGVTIDSRLSWRSHSTA VYKKACSRLYFLRKLRSFNVCSKMLGIFYQSVV ASTLFSSVVCWGSSIAVSDTNRLNKLIRKAGS VIGCKQETFEAVMERRSLNKLLSIMDNPEHPL HPVLDKQRSSFSRRLRQLRSKTARYQKSFIPN TITLFNNSQCRR	cd01650	51	301
	RTE-15_PSi	GCF_000230535.1	Pelodiscus sinensis	CACTACACCTTTTTT TGTCAAGGAAAAC ATAGTGATGAGAA ACGCCTGCATGGA GTTGGTTTTGCCAT TCGCAACTCGCTGC TCAAGTCAATAGAC	TAATGCTGCATCCCACCGT CTTTTGAGACATATGGAG GCCATGA	MTPSAESKRIMTLRLFTESGYINIISAYAPTLTT NSADKNHFYESLQKVVDSISSDEQLFILGDFN ARVGIDTSSWPICLGHHGIGKMNENGQRLLE FCSINELCITNTFFPGKECHKVSWHHPRSGH WHQLDFVLVQRKDLNCVKSTRTYHSADCDT DHSLIVSKTKLKAKKFHRTTLNKKPKINCMNT RNSAKCEDFAKAFAKKWEENAHVESVEETW QTLCSAIYEAAREAFGEKKSSNKDWLEENAD RLMPLIKQKNEAYLEYIKNPSDSAKAKLQHAK GLLQREIKHCINQYWIDICSNIQVASDTGNAR	cd01650	445	722

								763
							499	
						cd01650		
				AMYEGLRKALGPNITKIAPLKTKEGISITDKSK	MDVDGGPAMPEPWVLRFDGACRRNPGPG			
				QMARWIEHYSELYSEERNIMSETRDFIPTMEL	GAGAALFKPCGTVAWTCSHYMPNSSETNNT			
				MQELDAVPSKQELIDAIDALSSGKAPGTDCIP	AEYTALLLGVQSAVHHGASHLEIEGDSHLVVA			
				AEVLKHNKDCLIPPLYDLLLKCWTKGEIPQDM	QVKGTFACRNPRLRQLRNRVRHALRAVTSLT			
				KDANIVTLYKNKGDKGDCNNYRGISLLSITGK	LKHIDRKANAHADRLANRALDLKRSLAECGE			
				AFAKVILRRLQLLANSIYPESQCGFCTGRSTID	HQGAMESCLHMNPAAQRQREQPAPPARPA			
				MIFSLCQLQEKCREQNMPLYISFVDLTKTFDT	CAPTRAEPASDHDEDIDAEIAARDGGEAFPTL			
				VSRQGLYLLLERIGCPPKLLQLIIAFHEGMRATI	PIGPGTAPARQPRLRLRQLTEDEQEAAASAL			
				QFDGYTLDSFQVNTGVKQGCVLAPTLFGIFFS	QAMAEELACKIEDADSWTSGDGYISAIPGRIR			
				VLLSCAFKNSHGSIYLHTRSDGSLFNLARLRAK	QVLQPFTAAQPHPRPTLQQQRQRPPRVTRT			
				SKIKKVLIRELLFAADAAIVAHNETQLQDLMA	QREHRLDEALDEMAAVQQERPTSRSAVRRA			
				QLSRTCQQFALTISIKKTVVLAQGNTTSPCIML	RRRVGRIRASMRQQQLRHDFARNESKCVEDI			
				NNNPLEVVLKFCYLGSTITSNLSLDAEINIHIGK	LRAASAETAAEEHPETCPIDSGTLHEHFTAVN			
				AATTFSRLTKHAWANSKLTIRTKILIYQACVLSI	SPRINFLPDEACGALFREAMADVGTPQERRS			
				LLYGSETWTTYASQEKQLSSFHLHCLHKILNIK	ALTDELTVDEVEDQLMQAATNSSPGHDGVG			
				WQDKVPNEKVLERAGMSSMGAILHNRRLR	YDIYKKFAAQLVPLLHAAFQFCWRHHRVPAL			
				WLGHMKRMDNNCIPKNILYGELAEGTRPKG	WKVGFVRLIHKKGDPTDPANWRPICLQTAIY			
				RPKLCFKDYCKATLVKFQIDISWENLAVDRN				
				TWHQHLRQGQKHYDKTLAENQNKKRMKRK				
				EQMITTNTTTAWSCIYCSRTCLSHIGLLSHTKH				
				CKKRPV				
					TGAAGCTACACAAGCGCG	AGACGAGCAACGC		
					AGGCTCGACAGCTGGACC	GCTGGGGCCCAAG		
					TTCAGCTGTCGAGCCACTG	ACCTGGCATGAAC		
					CATCCGCGCCAGCCACCGC	GACACTGCCCAGG		
					ATCTGGGGCTGGCACTGC	CTGTGAACGACGA		
					CGGCGCCACCGCGAAGGA	GCACGCTGCTAACC		
					CAACGGAGCGGCAACGCC	CGGCAGCGCACCG		
					TCGCGAACGCCGCGTGGG	GCCTCTGGGCTCCG		
					TCTGGGGGGGCCCCGCGG	CTGCACCGGTCACC		
					CGCACATCGCAGGCTCGG	GGTGCAACACGGT		
					GCACAGCGGTAAGCTGGT	GGGGTCCACGTCA		
					CATTTGACCAACAGGGCA	CGGCGCGATCGGG		
					CTACCCAGGTAGGGAACC	GCAGCGGCTGTAG		
					GCCCTTCAAAAACCCAGG	CTCGCCTGCTCACA		
					AAGACACAAACACCCTCCC	GGCCTACCCACGG		
					TTTAGTGACATGCATATTT	CACCAGCACCAGC		
					TAGCCTACATTTCAGTTAC	GACGCGCCGGCCT		
	Utopia-1_PS	GCA_009848525.1	Phytophthora sojae					

	Utopia-2_PS	GCA_009848525.1	Phytophthora sojae	GCTTCGGCTCGGC GCTGCCCAGACCC GCCCGCTTCCCCCG CGGCCACGACGAC AGCCCCCGATGCG	GGAGAGGTTACTAACTGG TAAACACGAACACACAT	KLYSGLLARRLSTYLEANGLLPMAQKGFRAYN GCHEHNFVATTLLDQTRRMRRRLYQVWYDL RNAFGSVHQDMLWYVLRLLGVERAFVERCD DIYEDSYFVVGNAADGATEPVRQEVGVYQG CPLSPLLFIAALVPLLRALEKLDGVGVALADGV RPCTTAYADDLKVFSDSAAGIARCHAVVEQFL EWTGLQANPGKCAFLAVTRNARGNPAHDK DMKLSLHDEEVSSIKLHDSYRYLGVGDGFDH VRHRLQLEPKLQQIKREAVALMQSGLATWQ VVKALKTYVYPKVEYALRHLRPLQSQLQSFDR VVAKGLRHLLRLPRSATNEVLYAPTSSGGLGL QPLVEMHRALQIAHAWQMLHSKDPAIREVA RAQVCQVARKRHRLREEHWRGRDDELVRCF LNSELAASPHAEALRRHGDIGSLWSDVQRWL RIYHLSLVVQDDRNGLDPLGLRVPHHAKWLD HKSVLRHVKLHLKIRHQTRWKGLADQGKTVR AHGGVGAKFVSTGAGLSDDDYRFGVKARLN QIDTNAVLKRKRLRSHKTCRDPTCSSAETLAH VLNHCEPNMDAIRQRHDDALEQIGSKIRNAL KRGESTAELRLNQTVPEYTGAALRPDIVLRNV AAKKMVIADLAVTFEEHAAGARHSSLQLSHD HKTLKYQPIVAELQLKGWQVQTAAIVYGSLG SVQPSNSTPTRRS	MDVDQGRVRPRQRRPSPRRPSERRPDGKRR RLNSEDDADAQKPAEHLQLADEDGAENHTP VHTPSAASAHQAPVLAVYVHNAERFTCTLCT FTAASFSSLQRHRTTRHRRAVLQDKFLAGCA CGTPIASRPAAAKHAQACAGRSTALPATAPP VAGTSSPTDAAATATVAAATVTPDSPQPATP ELAAPPPLASPTHPEVVEVQEQPSQGRWGP ALPRALVAERVAARLGAIPAPRWEPPLPRELV STRIAARLGEVPAPRWGPPLPRRMVASRIAA RLLPPEPTADGETKEDTPADGADLTARDANQ ENNPAAAPDRTATAAAMDMDSGADGEWL LRFDGACRANPGPGGAGAALFKPDGSVVWT CSHYMPSSSETNNTAEYSALLLGSRAAADHG ATSLRVEGDSTLVIQQVRGIFAARNAILRRLRD QVKVELARVGSFSLHHIDRQANAHADRLANR	cd01650	912	1175

ALDLRRTMMECGAHTAGDGCTNTTLDPMP VLPPQPTPRAPPPTPAADTNIEMQEPEDEDL ADIDDGEVYAAMRIGPDAVPQRRPRLRLRQL TEAEDDAAREMVERLGATLAAKITDASDWES AEGYITALPHLLYDKLQPYSQAQRRPEPRPQR QRQQHQGSQCPRQQRQQGEQRRQRPLHD QGQDRRSLEGDAHGDNERDGDASQQDSAR RRQCRRRGRSGAAKRRRRRRPPRVTRHHRE HRLDEALDDLHAVEQASPGDRTAVRRARRR VGRVNSAIDQQHLRHRFDTDEKACVEDILAK ARAQQETPRTAASSAGTASDPAAPAAGATD DGTCPIPGDELWRYFDGVNTPSQEFAPDAPA GAAFRSAMTHLPAASRFRELLTEAPTQDDIES QLQHAHGTSSPGLDGVGYDVYQHFAAQLLP VLATAFQKCWEAKRVPQSWKLGVVRLLHKK GAREDPANWRPICLQQAIYKLYAGVLARRLV RWLDANDRHAPGQKGFRAVNGCGEHNFLA ATLVDQARRKRRQLYEVWYDFRNAFGSVPF ALLWDALARLGVPEDYVTMCKGLYESAAFVV GNAADGTTAPISLRVGVFQGCPLSPQLFNAAI SPLLFALKRLPATGVPLSGDDCPGASAYADDL KIFSGTEDGIKRQHALVADFLRWTGMAANPS KCSAMSIQRDRRGVLKADDLKLELDGSPIPAL SMSASYTYLGVGDGFDHTRRRVELAPTLKQL KDDATALLQSGLAPWQVVKAVKTYLYPRVEY ALRHLRPFQQQLAGFDRHLARGLRHLLRLPA SATAQCFYAPVSRGGLGLLPLTELHAALQVAH GWQMLNSKDPAIQRIARVQLRQIADARHRL DTQAWAGREEELCELLLNSQLGTSPDAPPKR RNGDIGSLWVDVQRHLRTLGLQFATAPACA DTGTAAATLQLRVPHHDKWLDHRTVLRHVK LHIKHRHWTKWAAMRDQGRTARAHGGVG SSFLTRPRGMCEADYRFAVAARLNQLDTHSV LKRRRLRAHDRCRQPGCSRVETLAHVLNHCP GTMDAVRGRHDDALKTIERALIASSSGDRQD RVELRVNQTVPSLAGPALRPDLQLYNHTKKT VAVVDLAVAFEEQVGEDPDSSALARIAAHKR AKYDRIKRHLERQGWKVHLSALVYGSLGAVA	ACTACCAGGACGCAACGG ACTAACCGGCAGGTAATA CATTAACCCTT					

					GGNRSVYTEHLGLLKRDAKRLDWQKKEEDRE DKCSSWPHEGWVKAPGDRTAIG			
	L1-2_SP	GCF_0000022 35.5	Strongylocentrotus purpuratus	TTATAGTAAGTTGT ATACTAAAGAAGAT ATTGATGTTAAAAT GCAGAAGATTATAT TTTTTCTCAAAATAT TTCATCTCTCTCTGA CGATGAAAATAAA ATATGTGAAGGTTT ATTAACACACAGAG AATGTAAACAAGCT GTATTCGCA	ATAAATAACTGTGTATTAT TGATGTTCCATTGATTTGC TAAATTGTAAGCAAAATTA TGTTACTGTTGAATATTGA ATAAACACCTCTAAGAGG CAAAGAGAAAAA	MKKNKTPGSDGITAEFYQVFWNEICDILVAA LNECYANEKMSDTQRKGIITLIYKKGDHRELK NWRPITLLNSDYKIIATVLAKRFQKVIGKIIHEN QSGYIRGRLAACNVRLTKDIIEYMTRNGKNG ALMMVDFSKAFDVIDVTFIKKCLSKFNFGPVF QKWVTILYTNISSSVIVNGWLSETFSVLRGIRQ GCPLSALLFVMAAELMANKVRENKEIKGLDI DDKGIDLKLLQYADDTLFYVSDINSFENILIELE VFGIVAGPKLNREKTVLIWIGTACQRWSLSCF GLHWTDEPVKYLGHHIAPDLQCALKIDWEN KLEKMQKLLDNWRKRNLTLFGRLTIIKSLAISQ VIHLMIVDTIPSTFLIKLNSLIFKFLWKSKIEKVK TFRLKWLGMFLDNTTGNWKDMCKHWFDC LGGLHLLLNCNYTDYTLNLLKLRNVPHFYIEIL HAWYLINKYTSAECQTQKHVDENILWYNQDI VFEKKMLFYEDWYKSGIIYLNDIFKNGTFVSV NELFIRLKTRRSKQFLIFDYTKLQKSIPKIWIND KHRVPLVFKDHNFSIPHFVFGKSVKPVSLLSSK QWYNMISFEQIKINKCCLYWEDVLQKDVKW SSVFNRNLIGIKDFRLREFNFKLLYNMLPVRN NLFKWGLSYDDLCLKCKKTEDVLHAFISCKAN ETFFAYLKDLVLRAFNLRIEIDIVLLLKMNSQTE IDLIMTIAFWSIYKMIIFRNKYGYDNRDLSLIYL FKREMTNRIEINEHLKVQNMKFPANLLSYI	cd01 650	47	306
	RTE-3B_SP	GCF_0000022 35.5	Strongylocentrotus purpuratus	GGGACAACGGCGT CTCTCAGAACTGGT GTTCACATGGATAT ATTCATGGTCACCA CACTGCGGTGAGG AAGGCAACGGGAA ACCACCTTCACTAT TCTTCCCAAGATAT ACC	AAGAAGAAGA	MESHNSKSKRNGVKGTSRNRGASLGRPSHV RQVGPDRHSASVRNRNIMRIGTWNVRTLYQ KGKLDNVVQEMNRMRINILGISETRWTGSG EFTVEKHRMIYSGGIRHERGVGVLLDPFTANC VLGFWAVSDRVLLVKLQGKPFNMAIIVVYAP TAESTDEDIDIFMNSQITMTAKSQCKSNEIVIA MGDMNAKVGRGRYGGTVGPYGIGDRNERG NRWVEWCEAKGMVIVNTWFKEHPRRIYTW KSPGDRARNQIDYITINERFKRGVKQAKTYPG ADCRSDHIPVVATLQCKLKKMTKAKSIQKVDF EQLQNRDVRLRYTVKVQNRFDSLTDQGEEVT WESMRDILVETAEETLPKKEKKSKQRWMTEE	cd01 650	549	830

					ILSKMSKRQKIKDRDGIEYRELDREIHDHCTKA KEEWLNKQCAEIERSFNANDPRVFKKVNDLS GKKLRSSKLGCIKAKDGSVLVEKEEILNRWTEY IGELFDDVRKEPPHCAGSIEGPRILLSEVRSAIS MMKRNKAAGPDGVVMEMVEALENYGVM KLTEVINKIYDDGKFPDDLCKSIFITLPKKPGAV ECGQHRIISLMSHITKIILRVLLQRARNHLTPEI SNVQFGFVKDAGTRNAVFVLRSIAERAVEM QKDVFLCFIDYTKAFDKVRHDQLFEDLTKIGL DTKDVKLLQNLYWEQTACVRVETELSDYINIK EGGSARMCHVSRSLQLLQVTTSTLKRGVRQG CVMSPDLFNYYSELILRNLEQERGLKVGGHNI TNIRYADDTILIAESEKDLQRLLDVVVRESDKR GLSLNYKKTECMVISKKGGIPCSLKINNNQIKQ ATSVNYLGTTITDDARGEQEIKRRIALAKSTFS KLENILKNKKLKMKTRLRVLNCYVYPVLMYGC EAWTLTSDLRKRLESCEVWFLRRMLRISWTE RVSNDEVFAMAGTRRKLLNNIRGRQLSFLGH VIRKNGLENLALTGRIEGRRSRGRRRILWMSS LRSGISECGIDLGENELLHRIWDRELWHSMV ANVSGYGT	MSGETRYIYDKADYDAMYDKIMEIDWNNEF WEMSMEKSWSFFKTTLLDIFNRYTPKKTLNV SQRKKHWMNRETLEVVREKHRHFRKWQEM KSEKNREKYSSAEIEQQRRCYSRSSNQARWQ CRKAVSSYEKYIASGVKKNARAFWSYVNSKIT TRESVCNLTKPDGSKTETDQEKAEVLQDFFSS VFTQEDLQDIPLPLDVEFDTPLDDIQITSGDVF ELLRRLKPNKSPGPDKIHPRLLKELADVLALPIS LLFQKSLESKQLPQDWKSANVTPIFKKGNKSD SKNYRPVSLTSIICKMMENLVRKQLLDHLMT NDLISVHQHGFVVGRSCITNLLEVLDQWTQIL DDGGTVDAIYLDFMKAFDTVPHHRLLLKLESY GVRGNVLAWLESFLSGRQQRVVVRGKQSK WGKVLSGVPQGSVLGPCLFLLFVNDLPKVVN SNISMFADDTKLYVRADTADLRKQLQEDLSR MEEWAEKWQLCFHPDKCTVLRLGKLPAEAS TDYCMKKGTDSVTLSVTKVEKDLGVLVDSQL	cd01650	270	526
			A	TAGGTGTTGTCTATTGTCT CATGTTATTACACAACCTC CTCCCTCCACAAAATAGCA GTAATCACGGAAGAATCC TGGATCGATGAACAGGCT TAGACCTACAACATCGTGT TTTAAAGTAAAGTAAGTAA GTAAAGTAA			
			Strongylocentrotus purpuratus					
		GCF_000002235.5						
	CR1-25_SP							

					SFKDHISQTVLKANRVLGVIRRTFMYLNTDT MPFLFKGLVRPILEYGQAAWSPYRLGEQRML ESVQRRATKIIPGLRNLTYQERLTKLKLPTLVH RRKRGDMIDVYKYVNGLYTVQEMPFKMSTE DRTRGHSRKLEKKYSRLDIRKYFFSNRVVELW NGLPENVVAAGSVNSFKNRLDSHWSKSVPV SLLYEPV	cd01650	26	284
	Crack-8_SP	GCF_0000022 35.5	Strongylocentrotus purpuratus		TAAATATAACCCAAGTACT CGCAAATAAGTAGCCACT ATCACAAACTTCAATATTT AGTAAATTACCAATTGCAG ATCACTGTATTTCTCATAA ATATCCTACAAATTACAAA TAATTATGTTTAGACACAC TGTCACCTCAATATAATAT TTTTTCTTCTTCAAAATTTG TTTATGTATGATTTCTTGAT GACTTGATTTTTTTTGGATT TTTTTTGTTTACTTGCTGAG TTTAAATTTACATTTGAGA TTTAATCTGTAATTCTTTCT GGTATGTTTACTACTCACC ATGTGATGATGTTTTCATG TATATGTATGTCTTATATA TCTATTTTATGTTATGTAG GGCCCCCAAGAATAACAG TATTTTAACTACTGGTGGG GCTACCCTGTTGAAATAAA ACGAAATAAATAAATAAA TAAATA	MIISKPITRLFNLSLRLGVVPQEWKQARVTPIH KGGDKDNVNNYRPISILPTIMKVLEKAVHSQL HDHLLEFDLLNAQQSGFRKGYSTNTLLTYVSD NLHRNIDSGKLTGIVFLDIRKAFDSVDHSKLLE KLYTYGVRGNTLSWFESYLRGRYQCTLVQRA RSDLQVLQSGVPQGSILGPLLFTLFMNDLPNS VEHSKVMLYADDTAIMFSALNPKEIESTMNS DLDRVKSWLCHNKLSLNTDKTNFLLVGTPQK LRHIEAPIDIKIDGLQIRQVEHVKYLGVWLDQ HLNWNEHMKHTSSKIASRLALLRRVRKYLTID SAKLLANALVLPYFDYCCTSWSNCNQGLRDT LIKLHKQMARIVLQMNVLTPSSEMFSKLKW MNMEKRWQFQKCKLMFDIVNGYSPSYLQD MFTQSAATHRYSTRIADTRGLVIPKVKTENGR RAFSFEGAMTWNNLPSELRNAQTKQSFCTSY FRNM	cd01650	550	804
	Daphne-19_SP	GCF_0000022 35.5	Strongylocentrotus purpuratus	ACCAGAAGGTGTT AGCTGCATGGACTC GTGATGGTATGATT ATTGCCTCAATTCA AGCCTCTGGTGGA AAATCGACCTCTAA AGTTATACGTTCAG ATATCGACCTCGAA	TAACTCACCAGGCTCTTTA CCCCCCCCCCTTGTTTTCCC TCTTTTTGTCATGCTTCCAA TCGTACAACAATTAATTTA CCTTTTATTTACATTATACT GAGGGATGGGCTCTAGCA TAAGGACCACACTCCTTTT TAGCTCATCCCCTTCACTCC	MPGSTVNSIINDQSFVEFLSPRNGENDFYGPS HGSYTCNEFNDLVNDSNDGLFSILHINCRSLS KNYDNISTFLSSLTVQFTAIALTEIWIKEYNSYY DIHNYTLVLKGRQNKKGGGIGFYVRNDVKFN RRLDLDINTPYIESLFIEVVTANISAILGVVYRPP DQPIQSFLESVENVITTVNTENRIFYMIGDYNI NLFKVNCQSHVKDFIEMFMAVGMYPLVHGP TRVVNNSISLIDNIFTNNLENCSSGTILTDISDH			

	Daphne-26_SPur	GCF_0000022 35.5	Strongylocentrotus purpuratus	AACCTTTGATAGTT ATGTCACGTTAGCG ATATTATGAACGAT CGAGCTGACATTTT AATGACTTATATTT CTCCCTGGAGTGA GAAGTTTTTTACAT TTTATCTTTTTGGAT TTTTCTAATACTCG AACTTTTAAATGTA AAATACCTTTTGTA TTTCTAATTTGTTTT TTGTATACTGCTTG T	CATTATGTTATGTTATTTTG TTGTTGTTGAATGTATGAT ATTTTATGTGTGATTATTC CTCATTTAATTGTATATGTT CTTGTATTTTTAATGATGA GTGTAAATAAACGAACTT GAACTTGAACTTG	FPIFCIYSQSTQTVTETCITMKRQFSAKNINTFI TNIGNTHWILKENDANSSYDCFLKQFQHVYEI CFPMKCQSKKKKRFDKPWFNDELRQLNRKK SLYYKKYLNNRTAYRLDAYKKCRNKLTNRIKYV KYKYFKQSFVDAKGDTKQTWKIINGLLGTKSK KNEIPSTLQYNNETIDNDTDICESFNEYFINIGP KLTSTLDIPTGDFKQYLKKMNLLLFYFFFTPV TPREIINVDKLKSNKCSGFDSVDVKMVKRVIH LICYPLCAIFNSSLETGIVPDQLKIAQIIPIFKSGT KTEMSNYRPISLLPVFSKILERAVYNRLYSFLNL HNILNNSQFGFRTGHSTSSASVDFTCQVSQV FDKSEVMMGLFLDLSKAFDTLDHNIILAKLFN YGIRGVLLNWFKSYLSNRKQFVKIKGNTSSM QTVVCGVPQGSILGPLLFILSINDMCNISDKLK YILFADDTSVFMSHKNIQSLQQDFGNEINKLL HWLSLNKLILNVNKTNYMVFTKQQKIDFNEV DLQVNNHEFKRVSVNKFLGIEIDSKLTWNEH TGAVCNKISKNTGILNKLRFLPKCVLRMLYNTL ILPHLYYCSEVWAMTSQTNLNRILKMQKKAI RIISNSHYRAHSKPIFYKLKLLNVFDICEYHIAVF MYLCRNESLPCALLKYFNLNSTFHDYETRNAS NFHLPQIKTTLMKKSISFDGPKIWNSIPINIRN CRSLQTFKRKYKMYLIESYVL	cd01650	424	681

				TGAATGTGGAAATGTAATT TAACTTATGTTTTATACTCA TGTAAAGACATTGTCACAT TATTGTTATAACCTATGAT GATTCAGTAACCAATAAAT ATTTCAACTATTTCAACTAT TTCAACTA	DLALAQLYDKVSNALASRKHVIGVFMDLSKA FDTLDHKVLSSKLHYYGVRGVALDWFNDYLS GRKQYVSYESVDSSLLPMKCGVPQGSILGPLL FLLYVNDISNSTTSLQYIIFADDTNVFCSNSDF RSLQRILNTELPKLSLWFRSNKLSLNLKKTNYI HFQSRKNATDKPIPEIVLDGEYLEKKQNTKFL GTIINENLKWSDHVHYVSKRVSRTVGLLYKLK HYVPLNILYMLYNSLILPYMSYCNILWADSKV YTDKMLLLQKKAIRVCTNAGYRDHTDPLFAR LKTLKVTDINFLQKALFMFRLHNDSLPSHFPP MFCLNRNIHTYPTRQSNKYHLCNPKTTMAM KSIRHSGPDIWNSLSVDIRNLKSVYTFKKSVKS ILLSKY	cd01650	746	998
	CR1-30_SP	GCF_000002235.5	Strongylocentrotus purpuratus	GTGAGGTGATGGC TACAGGGAGCAGA CATGGTCTGGTTTT GGTGATATAAAATC CACTTAATCTTGCA CTTTTAACTGCAGA TCGTGTATTTTATTT ACCTGATTTTGCTT CTTGAACATATCTT CAAGCTTTTGATAT GATTATTTTGTTCTT TGATACCTTCTGGA GGTAGAATATTGT GATAGTTTTACAGC CATTTTCTGCCAAG GAATTTGAACACAC ACCTGTAGTCCCAT ATGGAGGATCCTG AGTTTGCTTATTTT GCCCAGCCTTCTTC ATACCTACACTACA GGCAACAATAACT GTAATGTATATTTT CTGACTGAATTATG CTTATATTATTCTTA	TAAGCTTACTAGCTTGGTT TTTACTTGCACTGGACTTA GCATATTTTACTGTAAATT GACAAGTGCTTTGTCCACA CACCATTCCAGCAATTCAT GACACTCCTTGTAGTCACT TGCCTCACCTACCATCCCA GTATAGCTTCCTGAGTGA GCTGCCAAAGGGATTACC CAACCAAGTACCAAGTAA	MKNYFLPPSSSRSHPLLLAILSIVCALKLLKSGN LYTQIADHRSASLPFFDLHATLCEPRSLHQPTR RRTYTADIDNCKLLLFISSCNKWNRNTKLCPL DFRSTSLSHLTILLLLSGQVETNPGPSTSAPETF PCAICGDEVRDNDPALLCDHCDCWCHISCVG ISPDSYDILTKKSRSFAWVCCQCSSTNIGSSSSL QGLNLSQTNRFSPLHHDSSNEEILPSINPLTST PTSSPGGRRQPRFHPPSRNRKKAFKVMVVN CNGLKGNAKKSSFHAAIAHHTPDFVFGCESKL DDSLSSYSIFPSNYSIYRKDRNIHGGGVFIAIID TLIVTECPEFDSDCEIQWCNVQLANAKPLYIG SYYCPPNNRQQGLEGLHQSLSKIMTRHRRSH PNILIAGDFNIHPDINWDDQSTTNPATSTCHQ KLLDILLHNSLSQTVREVTRPSSNNILDLVVTS NPALVENVCVKSGISDHNIVTFTLAANPKISVK PLRKIYQFHKADEQQLRDAAAEFATEFLQSDP ERRSVETNWSIIVNFLNRCLSELIPSKMSKGRK HLPWITPPLKRQMRKRDSLFKKAHSQSCSTA WTAYRQFRNQVAKNVHRARCEYVNNIIGAS LKDNPKSFWSYIKTCRSENIGIPTLRTPSKLCAS ALDKAETLNDYFQSVFTPHQLQPLVQKGPSP FPTIGHLHVHRTGVEKRLQQLDPSKASGPDEL PPKLLKLVAREIASPLSFLFQQSYNSGIVPAQW KQALVTPIHKSGEKCDPSNYRPISLTCVCCKV MEHIILSHVSKHVAVNNILTDAQHGFRRGLST			

			TGATATAGAAGCA ATTACTCACTGGGA TTTAAATCATTTAC CAGATGATTTAGCT AACCGTCTACTACT ACTAATAGCTAACC TACATAAAATGGGT GTAGCTGAGTGAG TGAGACTGCTGAG CTTTTCTGACTTCCC TTTGTTTGTTCTGA ATACGTATTGAATA TTGAATACTGTCCT GCCAAATTTAATCA CCGGTATTCATCAC CGTTCGACTTGCAA ACGCCATC	AG	TACCTGTTCATAAGAGAAT AATCTATAAAGTTTCACTA TTTGTTTATAAATCCATGT ATAACCTTATCCCATCCTAT ATTTCTAGTCTGTTTTGTCC CATCCAGCAATCAAGATAT CCATTACGTTCAAGGTCAG ACCCCCTTCTTCTGCAAAT CCCCCGCATCAATTCAAAG AAGGGTCAACAAAATATA ATTTTTTCTGGTCCCACTAT TTGGAATAGTCTCCCTTTA AGCATTCGTTCTTGCACCT CAGTCGCATTATTTAAGAA ACATTTAAAAACCCATTTG TTTTCTTTGTAATTGT	TTQLTTAVHDWSSILQQRSQADIIFLDFKKAF DRVPHLLLSTKLQYYGISGDSLQWIMSLLTDR QQAVVVDGSKSTWRDVTSGVPQGSVIGPTL FLLYINDIQDNIQSTMRLFADDSTLYREIRCPE DHQILKDDLQQLSRWSSNWMMDFNVKKCA ILTITRKRTPSIREYHLSNGLIPRVKEYKYLGVTV AADLRWNKHCQTIRHKASRTLGLLRRTLSPCT RDVKSRAYTALVRPQLEYASEAWNPHSVTVT KMLEQVQRTAARFVYRDYRYTTSPSALVAAL GWDTLHTRRILDQCTLLYKIHHRLVSIPAPTIVI PATYFGRHDHNLKYVIPVATIDSFKFSYYPRAI RIWNHLPGSAVNATGITNFKEAALPSIRMMQ PPVGSHII	MHGHTLDLVITRCADNILHNCMVLDKGISDH YVISCDLNLQKSDAKRTLITSRNYRLINQESFG NELVAVLQSISQCGDVNEQVEFINKAICDVLD RHCPVSVRMKRDRPCGPWYDDNIHLLKRKK RKSERKWRKSNLRADRDIFVCHKDELNAAIV KSKSSYFTNLLLEANCKTMYATLHLLLGSSKKV LPYCDDQTDLANKFAGFFDDKISKIRTVLDER EQGLQTSGIVQSMDKHQCETGFSTFVDVSPS DIKMIISKSANKFCSLDPMPTWLVKDNIEVLV SPFVSLINSSLRSGVFPDKFKDAIISPCLKKSTLD SLVLANYRPVSNLAFLSKVLETIVSKQLVKHLD RCDLHEVLQSAYKANHSTESALLRVKHDLLLAI DKREGVIMVMLDLKSAFDTIDHSVLLRRLHSD FFLSGTVLDWFQSYLEGRSFAVSIGNHLSHKL SLKYGVPQGSVLGPLLFTMYLRPLGDILRNNG VEFHLYADDVQLYYSFNPKSSSSFQNAILLIEK CVSEVNVWMIKNKLLMNEQKTESLLIISSRLR HYLPDHVHLLLGDAFVEPSESVRTLGVVLDSS LTMSQQVTSLCKSLNFHLYNISRVRRMLTDE ACHHAIRSLVLSRLDNANSLLININVSDIDRLQ RIQNRAARIISNARKYDHVSPFLNNYTG	cd01650	306	575
	L2-1_SP	GCF_000002235.5	Strongylocentrotus purpuratus				

	CR1Y_SP	GCF_0000022 35.5	Strongylocentrotus purpuratus	TGTGTGTGTGTCGC GGGAGTCGGATGT ACACTGCAAATGCT TCCTGGAAACTGCC TACATCGACACCCT GGGATGTCTTTATC TGGAATTTCTTCAT GAATTTAGGAATTT CACCCTTCCTTCTCA TGGCAGATTGAATT TTATGTTCTTTTGG ACCAATTCTGTCAA CTGAAATCTAAGCT AGATTTTGCCACTC ATTTCATTAAGAGT GCCATCAAAATCTG TAGTCCCACAGCAA GACTCTGATGTCTA CAGCTTTCACATGG GCTACAGTCATACA TGTACAGACATGTA CAGGGAACCACCT AGCTTCTAGTTCTA CACTGTACAACTGG CAACCATGGTCAAT ACTGTAATGTATTA TTATTATTACATTCC TTTCTCCACTATGA GTTGATTGAATTAA CATGCCCAAGAAAC CTAAATTTCCATGT GGTTCATGTGGGA AAAATGTGGCTTG GAAGTCAAGGGCA TTCAGTGTGAAGGT GAATGTGAACAGT GGTTCCATGCAGA GTGTATTAACTTTA	CAAGACCCTCCAGGACCT GCCCATGCACATAGCTACC AAGTCCTCCGTTCAAGAAC TGAGCAGCATAAGCAGTC CTTCTTTCCCAGGACTGTG AGAGTGGAACAACCTCCA TCAGAAACAGTATGGCCTC ATCCGTCAATGCCTTCCGT CAATGCCTGACGGGAGGC AGCAACTAAGTCCGGGAA CTTTTTACTTGCACCATGT AAATAGCATTGCACCATGC ACCTTGCTCTCGTAAGCAC ACACACACTCATCAGTCTG CAAGAATCTTCAACTGTGA AGGAGGTAGACTTAACCC AAAGAAGAAGAAGAA	MAPAQLCGSATSVEYTISAGQIYSFPVPSVAT SAAKICHLNHTDIEPCTTNPDSKDEGESDDSL CLGSPKFSSSPRTLKKKPINPTKPKAKVLTILNV NCQSIRAKKQSFAILIHQHNPDVVIETESWLT EDIRSSEVFPTDHYNIARRDRPNDLHGGVFLA IRKDLLASREIELETDCELIWCKISMPGAKTFH VAMKLRAHLVQDQHARSQDFPCCDEMSLTE LERSLSRLSPTHSVVIGGDFNLPGWDWRNQ QVGNCRYPNLHRRLGSLLDDHGLTQVVTEPT RYQNTLDLLLVSNPTVVNEVTVVPGVSDHDC PLVKLNVRPTRPDEEESFTLQQSSVGEVWDQ PGQMRRKVLLYSKAQWEKFGESIEAVRDEIE AKVTYASVNELWLIFKTGIEKGINDFIPQKLTP HKANLPWVTNGIRRLIKKKRDRLSNECRHLRKR EGLIPYVLEERLRDLKRTLQFALRKAYWNYVE SIFTVEDDDEGPGKYNCMKRFWQFIKNNRN DSAGIAELKSDGVLFSEPAEKADALNRQFKKV FTQEIPIPPHLLPDTSEHESMPDIVITLKGIEM MLKKLKPHKAPGPDGITPRILKQLSRVIAPILRI IYRLSYDTGEIPDDWHKANVVPVYKKGDRTD PSNYRPISLTCICCKMMEHIFASSIMRHGDEK DILCPLQHGFQAKKSCELQLTRLCHLGFVTDLL NNMELSQQTDVLITDFSKAFDKVGHQRLLKK MEYYGVRGKSIDWISSFLPNRKQQVVLDGHT SPTVEVQSGVPQGSVLGPCLFLFYINDLPNQL STKVRLFADDTVLYHTIQETHEDAELLQQDLD QLCGWSSKWMMALNPKKCQVITVSRKKSQ VCPSVFPQWRNHWSLCQLPNTLALLSLLTST GTQHVSATCKKANNTLAFLRRNLRVNSTQLK TTAYQALVRPLLEYGSAVWDPHTIRCVNQLE MVQRRAARYVLNRYHHTSSVSDMLSQLGW TSLQKRREHQRLTLLYKMHYSLVDFDIHHPTS NRQQS	cd01650	621	854

			ACTCTACAGACTAC AGTGCCCTAG				
CR1-7_SP	GCF_0000022 35.5	Strongylocent rotus purpuratus		TGATGTTACTTTTAATAAT TACTGTCGATGTTCAAACT GAATGTTTCTTTTAGTTTG CTGTTGTTACCCTGTGAAT TGCCATTGCGTACGGGTA GTGAAAATGTTATACAAAT ATTGCTAAGTGTATGAATA TTGTTTATTTTTCTTCATAA TAAGTTTCAACTTAATTAA ATTACGTTTTTTTAATATAC ACAAGACTATGTCATTGTT GTAACTTTCGGAGCCTATG ACCTACAAGCATTGCTTTC TCTTTAGGCTCCATCATAT TTCTTCCCTTATTCAACTTA GTCTTATGTATCATTCTTTT TTTTTTTTTGTATAAACGCA TCCTAAAATGAGTTCAAAA TGTAACATGCTATATGTAA TGATGTACTATACAACATG TGAGAAATATGAAAATAA ACTGAAACTGAAACTGAA A	MDFNLNLLQCDENTSCQDFLDLMLSKSFAPT ITKPTRITEYSATLIDNIFVSGTPIVKSGVIVTDIS DHFPIYTVISNVISHSNPNTNKCSGIRYLSDPD LRRLKEALTITDWYAVYNCDDVDTSFDNFME ILMLTLNSFPTRQQNCSNYKKHPRLAWISKSL LKSINRKNTLYYKYKKNHTVKSKLKYSYRYRNILT TSLRLAKKNYFTKQFNVNKGDIKNTWRVINN VLKSRRVSTHISNIMINENRIQDPLVIAEEFNN YFANIGPNLAQNIPVSDLHFQYFLETKNTNS MFCTPILEHEVIDIVKNNMKTKKSAGNDGITNF ILKEIIGEIVSPLTHLLNLSIMNGIVPKKMKIAKV IPIFKKGNAQDIGNYRPISLLTSFSKILEKLMYS RTMQFLIDFDILSDTQFGFRQKHSTTHALLMF LDKVAHAIDDMAHTVGIFLDFSKAFDTINHDI LLYKLSHYGIRGKALEWFRNYLTDRKQYVSIN GHDSYQRSVSCGVPQGSLLGPLLFVIYINDFR NSSEIMSFILFADDTNIFYSHRDPQYLLRTVNS ELKRIQAWINANKLSLNVSKTHFMLFSNCIKS LPGNILIDNTNLNQAESTKFLGIYIDSDLSWKI HVNYLCKLISRNTGILNKLKHEFPTHILLSLYNT LILPYLNYGILAWGNGTKAQLEKLLIIQKRAIR NINHAGFLSHTSQLFFNNKVLRIGDVYVYSVG VFMYQLTINELPAVICALFVRNNTIHSYSTRQ RNLFHLPLTRTLFAQKNITFTGPKLWNDLPIEF TESTSLYSFKRKLKCHLLSNYALMVEEA	cd01 650	356	610
Hero-2_SPur	GCF_0000022 35.5	Strongylocent rotus purpuratus	ATATATCTTCACTCT CGATATATCGGAAC TCTCGATTTTTCCTA CCCTACAACTACCG GCGACGAATCACCT AACAACCGCA	TTTATATATACATTAATTTT TCATATTTGTTTTGAACTT GTAACCATTCTGTACCAAA ATGACAATATTTAACGTAA TACGCAATTTATCTCGATT CAATTATACACTATTCATTT ATGTATCATGTATGGGGGA TCTTCTAGCTTCCTCACCAT AATGTTACAAATCTTCTTCTT TTACCCTGTTCTGTACTAT ACCACTTCTGTTTTCCTCTC	MPPYVLIGSLLLFFTFQANCLRFNVPNLRATE WQPHNAAIKDRLTLLHHDLGAGIVTPKSAAD DLTSILTEYFSGQDDFVDKKLSDTYIEHEPKSLI QARKCKNVLRKRAFGKNGTAEDRKAFRAAV KTVSYLRKQDKQKKCDKSRSHMEKYRNNFY EFSKAACSGSLNESPTKPLFSEDVANNYFGDK YSTPNVINIGSLNWFPYVNINTPDRFEFDMTP IKPKDVKNVLRRKKAMSAPGNDGLMYGMIK HLDSVHHFLATLYNKILDSGDPPESWSESKVT LIYKSGDTDKPENFRMISLTSCISKVFHQIIAER SMSYLTENNFIDKTVQKAFIDGVNGCIEHNIV	cd01 650	280	545

				CTTTTCTCCTTTTCTCCTTTT CACATCCTTTTCCCTCTCTC TTTATCCCGGTGCTCCTTG TTATCACATATATTAT	LHEILSDAKCKKRSLHTTFFDLADAFGSVSHQL ITSSMERYAIPTPVARYVATLYSRLNGRVRGP GWITKPFEFKRGVFQGDPLSPTIFLMCFNPVL EYLKEQQDQFGYKWGDRKIITTPYADDFNLIT GNKLQHQKLMNGINDRLKSMGLTLKPRKCS SLSLTSGKPNIVDFTIGSDVISSVHGHPVKFLG SLITFSMKTSEIFEYLKGKIESTLENINKCNIRNE YKLCILTRYALPSLRYMLTVHDVHASHLKSLDII LDRYIKMWLGIPSRGANLAVVHDPDGLNIPSI SDLYRTCHCLAYSRSRVKGDDVVNHALDRRL ERERGWSRQTSFSLEAHEIHSKIPDSVSWNQ TKTVIKKAVQDVCRENWKRVIEPLLVQGEFM KVVHMENTDVTWKSIIYNLPQSVMKFSVNA AIDSLPTNKNLYRWGKRISDKCNLCNCTGTLH HVLNNCERMLDRYTWRHNNILRIMLKHIKNT SNNSTKVYADIEGESINGGTIPPNIIRTQHKPD ICHIDNKNKDIIIIELTVPFETNIDKAHQRKQDK YASLTHDITSTGWKSKLICIEIGSRGLITPQNKY RLKEILKITNSTRYPESRDEILKAVLFSSYVIFQG RREPAWQFDQYL			519
							263	
						cd01650		
					MHRKQTKHNIESLNFDLAKEDWVEVLCVSD VNVAYEKFLEIVLHYYHKNIPLVKIKKGKKAKQ PWITKGILRSIVARNKLYKEALTSQCNDKLKKY KTYRNHLTSLIRLSKKLYYSNNIESNKSNNNLL WKTVKELLGKKKTESNKCFLIDEQQITDPIKIS KAFNEYYIDVGPNLAANIDSGDMDYNKSLNF PRENSLFLTPTNDYEIINVVQHLNSTKSSGHDS ISVHLLKQIIFNMVTPLTYIFNLSISTGVFPHLLK IAKVIPIYKKDDSSVISNYRPISLLPSISKILEKIIYK RLNTFLQKNNSLIPNQFGFRIGHSTDYAILHM YDKIIEAFANSEHIVGVFMDLSKAFDTIDHKILI HKLQYYGVRGQALSWFTNYLQDRQQYVSFQ NHESTKLTIKCGVPQGSILGPLLFIIYINDIIETSS LLNFIIFADDTSISYSHKNFDTLITRLNCELMKIS RWFKSNKLSLNLSKTNFIHFKQTNVHDMRNC NIEIDGLPLIEKQSVKFLGVSIDSHLNFNEHIRN IHTSISRNIGILYKLKQILSERSMIILYNSLILSHVI YCNIVWGNCSLTRINSLFLLQKRAIRMITNSH		
	Crack-2_SP	GCF_000002235.5	Strongylocentrotus purpuratus	GGTCTGTTGTATTA CGATATTTCTGACC ATTTACCGATATTT ATTATTTCTAAAGT TCACCTTGTTAAAA AAATTACTAAAAAT AATTGTGCA	TTTCAACATGCAGTAGGCA ATTGTCAAATTTCGGTCAC ATTGCAACTCAGTTTTGCA AATTTTGTTTTCTATTTCGT ATTTCTTAATTTCCTTTATT GTAAAGGCTGTATTTGTTT CTACTCGACTATACCACAC AGAGCTACGTACACTTCCA CTATTTGTTACAAACACAT ACAAACACGCACGCTCGC GTGCACGTGCACACACAC ACACACACACCCACACACG CACACACGCACACACACA ATATATATATATATATATA CACTTGCACAACACACTCA CCAACTCATACACAAAGAT CGCTTGAACACTTTTTCAT TTTAAGTTTTACCTTAAGA			

	RTE-11_SP	GCF_0000022 35.5	Strongylocentrotus purpuratus	ACAAATGGATACA AGATGCAAAAAAG TGGAAACCTCACGT CAGGCTACCAAATC CAAGGCAGTCTACA CAACAATCAAACAA ATCACTAAGAAATC TAGCATCAGAATGC AGGCAGTGAAAAG CAAAGATGGAGAA ATCCTTACCGAGAT GAAAGATGTTAAA GAGAGATGGAAAG AAAGCTTTGAGGA ACTGTACAATCACC AGAACCAAACAAA TGGAGAAGCACTA CTCAGATGGCCAG CCTCGAACCGGTTC CAGACATTCTGCGA GATGAAGTAGAAT	TTTTTTTTTGTTTTGTTCTT GTTTTTTTTTAGGCCGCCA TCCTTGTCAAGCTTTGCTT ATGATTGCGGTCTCCTGCA CTTTTCCACCAGAATAGTT AGGATAATACTTTAAATTT ATGATAATATGTTATGGAA ATAATAAGTTTCGTTATTA TAATTATGTACATATATTA TGTAAAAAAACATTTTGAC ATGAAGAAATTATTGTTTG TTTATGTTGTATTATACCTT ACATGTATATTATGTTATT TGTTATTCAGTTATATTTTA TTTGGAAAAAATGCAGAA ATAAAATCAAATCAAAATC AAA	GTCAAGTAAGTCAATTA	YISHTEPLFQRLKTLKIQDIHIFQTCIFMYKYSH NQLPSLFHNIYSLNSNIHTYPTRCSTNYHLENP KILLAQKSVRHHGPDVWNSLPNSIRHCSSLYS FKASLKKTLISKYLHN	MPAKPCRESIRIILQRRIMKKTEEVLSEAQAGF RPGRSTMDQLFTLRQIMEKRLERNQSLLCCFI DYEKAFDSVWQEGLWKAMNFFGYSSKYIHLL QALYQQSSSAVRVNGELTDWFKTTVGVRQG CALSPQLFNILLEAVMLYAIHDTKIGVRVQGQ LINNLRFADDIVLIADNDRDLQTIVDLVHQSSS NFGLKINITKTEVQAISKQPQLLNINVGGEQLK QVDKFTYLGGTISQTGSCSEEVKHRIGKAIGA FQALKKIWASQDIRCSTKVELYKVLVLSILLYG AETWTLKKVDENRLHTFEMACLRRIMGVTRF DRLRNTHIRTQLNMEETIIDRVATKRLRYFGHI NRMNSKRYPHILLNGNIHGKRPRGRPAKRW TDCIKADCKNRQVDSLTKATRLTEDRKVWQA IVKQKPSYIPVMVWTA	cd01 650	12	235

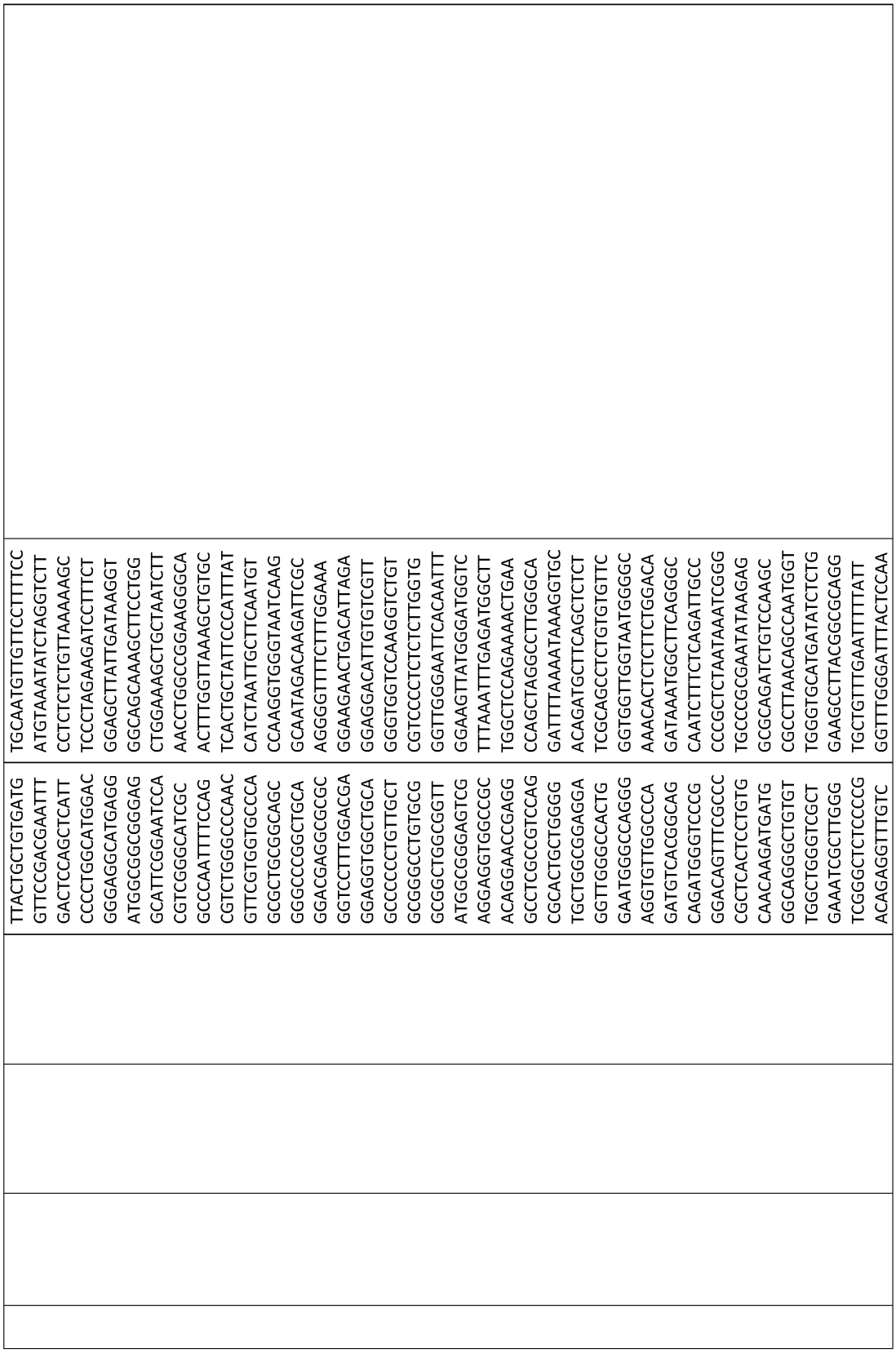
	REX1-2_XT	GCF_0000041 95.4	Xenopus tropicalis	TTGCTGTACGAAAA CTCGCAGAAGAGA AAGGATTTGACAG CATCACAGGAGAG GAACTCAAAGCTAC TGGGCCAACAGGA ATCACCATACTTCA CAAAATATGGCGG GAAGAAAGTTTTCC GGCTGACTGGGGA AAAGCCATAATAAC ACCAATCTACAAGA AAAAACAAATCAAC TGTGGAAACTACA GGGGTATTAG	GATTCAAG	TGACATTCTGCCCAAACCA ACACACACTCCCCATAACT ACTGGACTCTTCGCAGTGT CACTTTAAGCAAAGCCACT TTAAATCCTCACTGCACAA TTTCAAATATTGCATTGCA TTTTATTATTGTAAATACTT GTACCTTTATTGCACACTT TTATTATATTTAATATTTGT ATTTTTTTTTTTTTTGTTTTT TGTTTTGTTTGTCTATGTA AAGTTGGGAGGAACACGG GTCAAAAAAATGTCATTAC AGTAATGATGCTTCATTGT TATTTGTATATGACAATAA ACTTGAAACTTGAAACT	MAPVRMAAVAQCPPQLCFFLMFVVLFFSFF ATCNALISYDRQTLLDIRQNSQSGPSQTEFFM NRGDLPWLPWQKKHRRKRGKRAGILVRLRN RQHRAPLPSILLANVQSLENKLDELRARITFHR DIKNCNIFVFTETWLVPSIPDSAIVPEGLSIHRR DRTIDSGKCRGGGVCVLTNKRWCTDVQTVS SGCSPDLEYLMIRCRPFYLPREFTAIFCTAVYIP PHADTTRALDELFDAIDRQETLHPEAAFIVAG DFNKANMKKVLPKFYQHVNFPTRGENILDH VYTPYAHAYKARPRPAFGKSDHSSVLLLPAYR QKLKREQPMMRSVQRWTDQSDSTLRHCFNI TDWEVFRTAADNNINNYAEYVMDFIHKCIN DVVPRVNVRTYPNQKPWVNGEVRAALRART AAFKSGDPSEYKSSRYALRKAIKTAKGQYREK VESCYSNSNIRNMWSGLKTITDYKGKSCRVG EVTASLPDELNTFYARFDKHLSQELLLEEDGRP YQISQADVCKHFRKVNTRKAPGPDNIPGRVF KACAHELADVFTDIFNLSLLQSAVPTCFKRTTII PVPKKMNVSCLNDYRPVALTSIAMKCFERIVK AHITSLLPASLDPLQFAYRSNRSTDDAIAIALH TALSHLDQRNTYVRMLFIDYSSAFNTIVPSRLV MKLRDLNIGSSLCSWILDFLTNRPQVVRIGNI TSSTLTLSTGAPQGCVLSPLLYTLFTHDCTATH	cd01650	564	818

					SSNTIIKFADDTTIIGCISDGDESAYRAEVRALT SWCRDNNLLLNVSKTKELIVDYRRLQGGHT PIHIEGAEVERVSCFRFLGINISEDLSWSHHVG VITKAARQRLFFLRRLRRFGMDSRILTNFYRCT IESILSGCITTWYGSCNALDRKALQRVVRSAER ITRTELPAMQDLYRQRCRRKMQRILSDPSHP SHRLFTLLPSGRRYRSIQTRTSRYRDSFYPQAI RLLNC	cd01650	1115	1378
	LINE1-38_ZM	GCA_902373975.1	Zea mays	GATGGCCCTCGTCT CTCCCGCCCCCGCC CTCTCGGCTCCCGA CGCCCCTCCCCCTC CCCAGCACCAGCAC GCGCCGGTTCCCAT TCCGGGGCCGGCG GCCAACTCCCTTGC TGCCGTCGCCGCA GCACCCTCTAGCCC TGGCGCGTCGGCTT CTATCCCCTCCTTGT CTCCTCCCGTGTCC CATGTAGCCAGGG GGCGCTCCAAGTCT CTAAGGTGGGGTG ACCTCAGTCCGGCC TCTTCGGATGGCGT GACGTCGCTCGACT CCCGGCCTTCCTTC AAGGACGTGCTGG TGAGACCAGCGAA TCCTTTGCCGCCCA GCCGTGTGGTAGT CCAAGAGTCAGCCC TGGCAGCTGGAAC GTCCTCGGCGCTG GGGAAGGTGAACC ATCGGCCGCCTTGC TCTCACCCCCTGTC	TGAGCTCAGGATGTGGTG TAGTGCCGGGGCAAAAGG GTTGAACCTGCTTTCAACT CTTATGTTTATGCTTGACC CAGGGGGATAGTCTAGTG GCTTGTTCAGAGTGGTTGT GTGTGGTTGCTGTGTTGC GTCTTTTGGTATGTTGTGT CTTGTGATGTGTGTCGT GTCAGGTGGTATGTTGTG TGCGGCGTGTGTGTGTGT GTGTGTTGTACCTTCTTCT ATCTATCTATTAATGGAAT GATACGCAGATCTCCTGCG TGTTCGAGAAAAAAAAAA	MLGVSAAQGNGGPASSRRKRWRKKKRLERE DEPTLEIVAPPPPVQMLPASDATACVIDWSV QMSIAEDNLKTAAVITVVCSRDISANEIVSVLA PRLEMEEGGLVLRQLSQSSFLAVLPSRVQLDT LVDSFPIIREENFMLSSRRWTRFGGAGGNSLK HLVELELRGLPVHAWETSVVQQLINPHACIS QVLPDTLEVRSLEAFRCLAWCSDPERIPATKD LWITEPDQASAAEGKKALVYPVQIRWVPADL SLLTGDPPLLIGGDDQDKSDTPPSRRRCSDNP PRTQHDAINDVPGNKTHTQRRPVHDRLGPL RRSDSPRELSHPRVQIDFEARSSQPCVLSDAV ADSCVGNPPTPVPDGVQCLAAGDVLLRSEEG CEDAEVIPMTPMAVSSPCPSSGGRRRLLLKTA TKQAKFRNPPIAFFRRRRRRAPRTPEGATLPPC SSGDRGAGSSLGALPPAPAAGVAASLLPHGS PPLVESHGSPGQQDFLRRVSKAVGAVLPLPK AVGGGRRRRQPTSLPDAIRRSRRVAGLGVEQ VQHIPKARRTIMKSILAEPDSEHFSQQNLDAY AKLFQAPLSRAQVEALAALFGWVMPDDEGE VVPPGTATLVCLQPTVAYYSNCCPMDPSLILF WNVRGLNSAARQDVVRSLVDSLRVDVVCLQ ETKLVNCDRRTVLSMLGADFDNNFICLPSVG ASGGVLITWRAKLGRIVASRVDSFSVSVKFNP ADGVAWWLTCVYGPQGHDRKIAFINELRAIR AACTGPWLLGGDFNLIFRDEDKNNININRPM MGRFRRCINDLSIKDIDLLGRKYTWSGANNV QVKLDRVFCSSDWEEKFPNYLLQSSASEDSD HCPLLLGLKDNHTGKRRFHFESFWPRFDGFL EAVSNTWGSVQQSSCPLKTLSLKFKATTKGL QSWSDRKVGNFKLQLELAREILLQLEVAKESR			

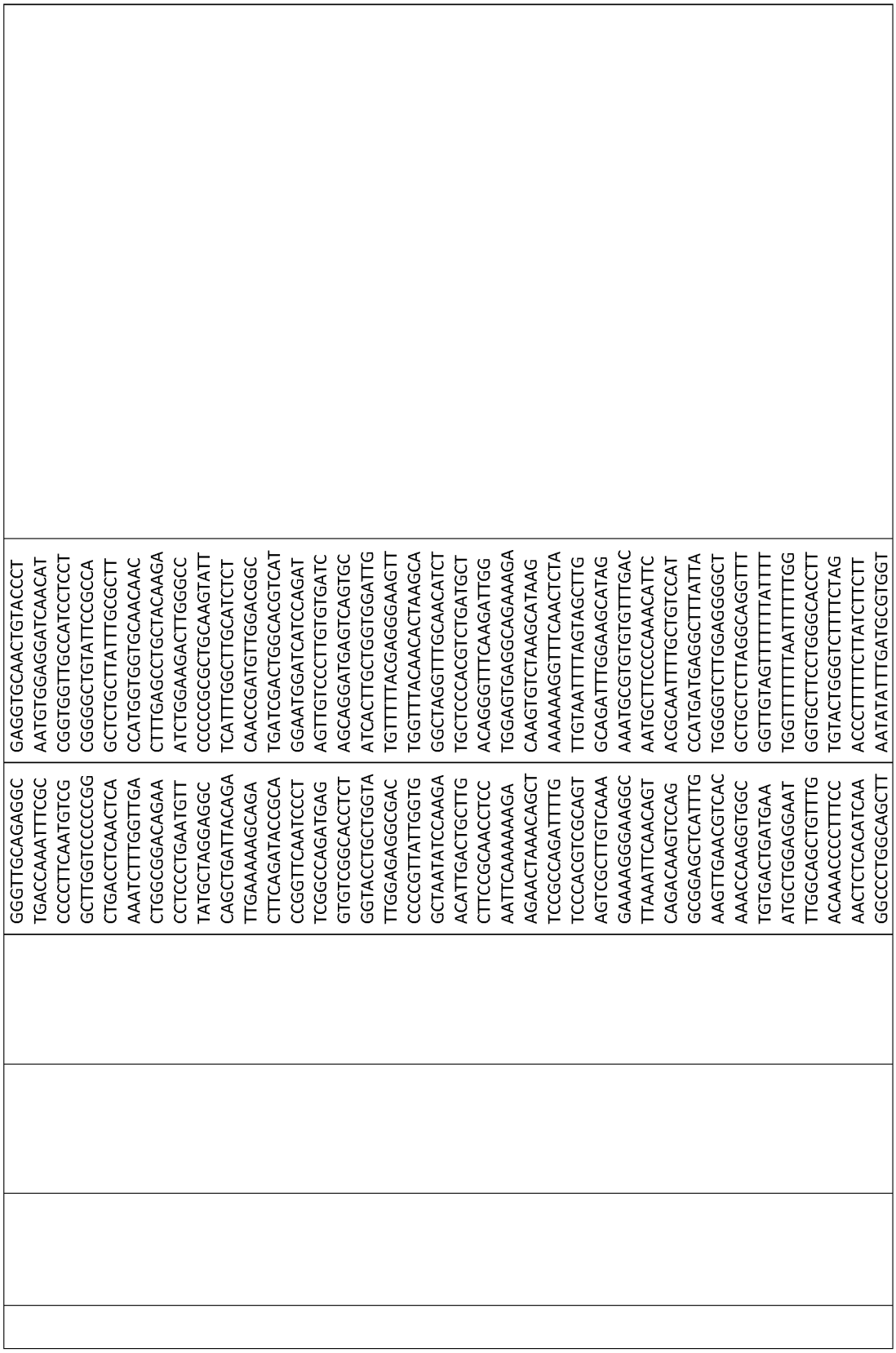
	LINE1-24_ZM	GCA_902373975.1	Zea mays	CCTGCATGCTTGAG	TAATTCTCCAGGGGGACTC	MNLSKILIWNVRGLNRKARRDAVRMIMVESTRPDIVCLQETKKESISRFM		
				TTCTCCACTGCTAT	TTCCAAGCTTTATCTCCTTA	VFSLLGNVFDQFTFLPAIGTRGILLAWKGSVCSALQTRVDEFSISVQFNQ		
				GGCTCGGGAGGAG	ACTCGGCGTATGTGACTCT	NDSLPWWFTGVYGPQSDEKKLQFLQELRMVRSVCVGPWAVGGDFNLI		
				GCGGCTCTTCGCCG	CTTACCAAAAAAGCTGAG	YRAEDKNNNNLDRAMMGRFQRIINDLQLVELPLLGRKFTWSNERSSPTL		

				TACAAAGCGGCGG		ILNVLETWLFGRLKKCSLSLASLIRSMARTRSRI		
				AGTCCCAGGCAGG		AWLKDGDANTRLFHSWSRYRKRKNFIAKIQD		
				CTCCCCACTCTACA		GEQLVTSHSDKAEVLRSFYSDLIGTGQQRSRT		
				CCGGCTTTCCGCCC		VNLDTLEMQHTDLQGLEQPISEEEVWNTIKQ		
				AGCTCCAAGTTTTG		LPPDKAPGPDGFTGRFYKVCWSIIKKDIMAAI		
				ATGATCAAGGATG		MAIWNKDFRNFRYLNTAFITLLPKKEDAILAK		
				GACAACGGTGGAA		DFRPISLIHSFAKLITKIMASRLSGYMDKLVSR		
				TCCAAACAAGGTA		NQSAFIKGRFIQDNFMLVQQTARLLHAQKQ		
				GAAGACGCCGGTT		PRILLKLDISKAFDSVSWSFLLEVLEKVGFGPI		
				GAGGGCTTCACGA		WRDVISGLLMTSSTQILLNGVPGEFIQHLRGL		
				CGCCCGAGTCGAG		RQGDSLSPLLFIIVMDVLNGLIVKAADEGLLQ		
				TCCCTGCTGAGCTT		PLSTRSIQHRISLYADDVVIFLRPAASDLHLVTR		
				GTTGGAAGGTGCT		ILELFGDASGLKTNVQKSSVAPIHCSPVDME		
				TCAACTGCTTGGGC		MVTQIFPCKVETLPLKYLGLPLSLRKLSKEQLQ		
				TGCGATCACAGGG		PLIDRLADLLPNWKADLMTRAGRAIQVQHVL		
				CTGCGGGGTGTGA		TATVIYHAMALDLPKWAIKAIDKIRRGFLWK		
				CAGTGAGCCTCGG		GRKDVLGGHCLVAWGKVTRPKELGGLGISDL		
				TGCTTCCTCTGCTC		QKLNWALRVRWLWLEKTDPSKSWADLPLQS		
				GGAGCTTGGCCAC		SSILKSILSTALITEIGDGRNTLFWRDRWLMGL		
				AAGTCTTCCTCCTC		RIEDLTPHIFAIVPTRITNRRTVAEGIVEMAWI		
				CTGCCCCAAGACCC		RDLRGTITWDVLSDFLTLSEVIADYSINVGVPD		
				CGGTGTGGAAGCG		KHIWKFSPTGQYSSKSAYDMLFLGSTSLGAYE		
				GCTCGGGCGCCCC		RIWESWAPSKCSFFLWLVMHNRCWTADRL		
				AGGCCTGTGAAGT		ARRGLPHPSVCPLCDQEEETIQHLLISCVFAR		
				CCCGGTCTGGGAG		QFWYSLLSRSGLPNLSPDNASCSLDDWWSR		
				CGGCTTTCACCGGC		AVEKIPRDLQSGFNSLVILGAWSLWRHRNDC		
				AGTTGAGGGGAGA		VFNGEIAKPGECSHPRWR		
				GTCCAACGAGCGG				
				CAGGCCACCGTTCG				
				GTTTGGCAGCGTAT				
				CGCTACGCCCGAA				
				GCTGTTCCGATCGC				
				CCCTTCTGTCCTGG				
				TGGGAAGC				

		GGCTTTGTTCGTCT	GCTGTGGAGGTAAAAGAC	VRLDRFFYSRDWEDLFPDCLMQSSSSSISDHCPLVVGLHEFTQGKRRFHF
		CTATTGTTGGCAGC	TTCAGACCGATTAGTTTAA	ESFWTKLEGFMTEVENSWCQPIAVGCPLQRMADKLKRLSKHLQSWSQK
		AGACCGGAGATCA	TTCACAGTTTTGCCAAGAT	RVGHVKRQLVMTNEILHRLEVARDSRDLSYHEEWLRRKLKHHSLGLASLE
		GGGAGCTGAAGTG	TCTCACCAAGATTCTGGCC	RTMARHRSGLNWLKEGDANTAYFHHHSRYRKRKNFIGKLKVEDRIVVD
		CGGGATGAAGTGG	AATAGGCTTGCTGCAATCC	QQEKKHAVWNFYNNLLGVAASREFTLNLGVFHRPASNLEVLDQIFSEEE
		CTCGGGCTTTTGGA	TCCCTACTTTGGTCTCCCCC	VWNTIKSIPSDKAPGPDGFTGRFYKMCLDGCQRGFLGSCQ
		ATTGACGTTAATTC	AATCAGAGTGCATTCATCA	
		AATATCGATTCATC	AGGGAAGAAGTATTCATG	
		AGTCCTTGCCAGAA	ACAATTTTATGATGGTTAG	
		GATTTCCTATTGTT	AGAAACTGCAAAAGCGAT	
		GTTACCGGATGAA	TCACAGGCAGAAAATTCCT	
		AGTGCTGATAGAG	AGAATCTTGCTGAAGTTAG	
		TCTTCAGCGGAGG	ACATTGGCAAGGCCTTCG	
		AAAATTGTTTAGGG	ACTCAGTTTCGTGGCCTTT	
		GCCCTCTTTTCAAT	CCTCCTTGAAGTTCTGGAA	
		CTTCAATTCAGACG	CATGTGGGATTTGGTCTAT	
		GTGGTCGCGGTTG	TCTGGCGTAATTTAATTGC	
		GCTCATGCTGAGG	TAAGCTATTGTGTTCCTCC	
		CGGCTAGGTTGCC	TCTACTCGAGTGCTAGTTA	
		AGCACTGGTGGAA	ATGGGGAACCAGGAGATC	
		GTCGAGGTTCATG	TCATTTGTCACCAGCAAGG	
		GCATCCCAGCCCAC	CCTGCGGCAGGGAGATCC	
		GCATGGGATCGTTC	TCTCTCCCCAATGCTCTTTA	
		TACGGTGGAGTAT	TATTAGTAATGGATGTGCT	
		CTGCTCCGTGACTC	GAACTCTCTTTTTCAGAAA	
		TTGCATTATCACAG	GCAAATGAGTTGGGCTCT	
		ACATTCATCCGAGC	TGGGACCTTTATTAAGAAG	
		ACGTCATTGAAGAA	GGGAGGGGGCCAGCGGG	
		TGATCTGTCTTCCTT	TCTCTCTTTACGCCGATGA	
		CAGGCTTCGTGCAT	TGTTGTATTATTTGTTCAG	
		GGTGTTTGACACA	CCTAACAGAAATGAGCTG	
		GATTTGTTCCCTCG	AGTGTGGTCAAGGAGCTG	
		GCCGATGAAGCTC	TTGAGGGTGTTTGGCAAG	
		CTCATTGTTGAGCC	GCATCAGGGCTGATTACA	
		GGGGACGGATGTG	AACATAAATAAGTGCAGT	
		CAGGAGAAAAGAT	ATGACACCTATTCAGTGCA	
		GTCTTTCCTATGAT	ACGAGCATATCATTGCTGC	
		ATTGAGGTGGCTG	TGCTCGACAGATGTTACTC	



			TTACTGCTGTGATG	TGCAATGTTGTTCCTTTTCC	
			GTTCCGACGAATTT	ATGTAAATATCTAGGTCTT	
			GACTCCAGCTCATT	CCTCTCTCTGTTAAAAAGC	
			CCCCTGGCATGGAC	TCCCTAGAAGATCCTTTCT	
			GGGAGGCATGAGG	GGAGCTTATTGATAAGGT	
			ATGGCGGCGGGAG	GGCAGCAAAGCTTCCTGG	
			GCATTCGGAATCCA	CTGGAAAGCTGCTAATCTT	
			CGTCGGGCATCGC	AACCTGGCCGGAAGGGCA	
			GCCCAATTTTCCAG	ACTTTGGTTAAAGCTGTGC	
			CGTCTGGGCCCAAC	TCACTGCTATTCCCATTTAT	
			GTTCGTGGTGCCCA	CATCTAATTGCTTCAATGT	
			GCGCTGCGGCAGC	CCAAGGTGGGTAATCAAG	
			GGGCCCGGCTGCA	GCAATAGACAAGATTCGC	
			GGACGAGGCGCGC	AGGGGTTTTCTTTGGAAA	
			GGTCCTTTGGACGA	GGAAGAACTGACATTAGA	
			GGAGGTGGCTGCA	GGAGGACATTGTGTCGTT	
			GCCCCCCTGTTGCT	GGGTGGTCCAAGGTCTGT	
			GCGGGCCTGTGCG	CGTCCCCTCTCTCTTGGTG	
			GCGGCTGGCGGTT	GGTTGGGAATTCACAATTT	
			ATGGCGGGAGTCG	GGAAGTTATGGGATGGTC	
			AGGAGGTGGCCGC	TTTAAATTTGAGATGGCTT	
			ACAGGAACCGAGG	TGGCTCCAGAAAACTGAA	
			GCCTCGCCGTCCAG	CCAGCTAGGCCTTGGGCA	
			CGCACTGCTGGGG	GATTTTAAAATAAAGGTGC	
			TGCTGGCGGAGGA	ACAGATGCTTCAGCTCTCT	
			GGTTGGGCCACTG	TCGCAGCCTCTGTGTGTTC	
			GAATGGGCCAGGG	GGTGGTTGGTAATGGGGC	
			AGGTGTTGGCCCA	AAACACTCTCTTCTGGACA	
			GATGTCACGGCAG	GATAAATGGCTTCAGGGC	
			CAGATGGGTCCCG	CAATCTTTCTCAGATTGCC	
			GGACAGTTTCGCCC	CCCGCTCTAATAAATCGGG	
			CGCTCACTCCTGTG	TGCCCGCGAATATAAGAG	
			CAACAAGATGATG	GCGCAGATCTGTCCAAGC	
			GGCAGGGCTGTGT	CGCCTTAACAGCCAATGGT	
			TGGCTGGGTCGCT	TGGGTGCATGATATCTCTG	
			GAAATCGCTTGGG	GAAGCCTTACGGCGCAGG	
			TCGGGCTCTCCCCG	TGCTGTTTGAATTTTTATT	
			ACAGAGGTTTGTC	GGTTTGGGATTTACTCCAA	



			GGGTTGCAGAGGC	GAGGTGCAACTGTACCCT	
			TGACCAAATTTCGC	AATGTGGAGGATCAACAT	
			CCCCTTCAATGTCG	CGGTGGTTGCCATCCTCCT	
			GCTTGGTCCCCCGG	CGGGGCTGTATTCCGCCA	
			CTGACCTCAACTCA	GCTCTGCTTATTTGCGCTT	
			AAATCTTTGGTTGA	CCATGGTGGTGCAACAAC	
			CTGGCGGACAGAA	CTTTGAGCCTGCTACAAGA	
			CCTCCCTGAATGTT	ATCTGGAAGACTTGGGCC	
			TATGCTAGGAGGC	CCCCCGCGCTGCAAGTATT	
			CAGCTGATTACAGA	TCATTTGGCTTGCATCTCT	
			TTGAAAAAGCAGA	CAACCGATGTTGGACGGC	
			CTTCAGATACCGCA	TGATCGACTGGCACGTCAT	
			CCGGTTCAATCCCT	GGAATGGATCATCCAGAT	
			TCGGCCAGATGAG	AGTTGTCCCTTGTGTGATC	
			GTGTCGGCACCTCT	AGCAGGATGAGTCAGTGC	
			GGTACCTGCTGGTA	ATCACTTGCTGGTGGATTG	
			TTGGAGAGGCGAC	TGTTTTTACGAGGGAAGTT	
			CCCCGTTATTGGTG	TGGTTTACAACACTAAGCA	
			GCTAATATCCAAGA	GGCTAGGTTTGCAACATCT	
			ACATTGACTGCTTG	TGCTCCCACGTCTGATGCT	
			CTTCCGCAACCTCC	ACAGGGTTTCAAGATTGG	
			AATTCAAAAAAGA	TGGAGTGAGGCAGAAAGA	
			AGAACTAAACAGCT	CAAGTGTCTAAGCATAAG	
			TCCGCCAGATTTTG	AAAAAAGGTTTCAACTCTA	
			TCCCACGTCGCAGT	TTGTAATTTTAGTAGCTTG	
			AGTCGCTTGTCAAA	GCAGATTTGGAAGCATAG	
			GAAAAGGGAAGGC	AAATGCGTGTGTGTTTGAC	
			TTAAATTCAACAGT	AATGCTTCCCCAAACATTC	
			CAGACAAGTCCAG	ACGCAATTTTGCTGTCCAT	
			GCGGAGCTCATTTG	CCATGATGAGGCTTTATTA	
			AAGTTGAACGTCAC	TGGGGTCTTGGAGGGGCT	
			AAACCAAGGTGGC	GCTGCTCTTAGGCAGGTTT	
			TGTGACTGATGAA	GGTTGTAGTTTTTTTATTTT	
			ATGCTGGAGGAAT	TGGTTTTTTAATTTTTTGG	
			TTGGCAGCTGTTTG	GGTGCTTCCTGGGCACCTT	
			ACAAACCCCTTTCC	TGTACTGGGTCTTTTCTAG	
			AACTCTCACATCAA	ACCCTTTTCTTATCTTCTT	
			GGCCCTGGCAGCTT	AATATATTTGATGCGTGGT	

			TGTTTGGATGGTCG GTCCCAGATAATGT GCAAGAATGCTTG GTGTTCCTTTGCTA GGGAGAAGCATTG AGTGTGCTTGCTGC TGTAGCCTCACTTA GTGGTTGTATTTCT GTTATCA	TCTCCCACGTGTTCGAGAA AAAAAAA			
	LINE1-41_ZM	GCA_902373975.1	Zea mays	G	TAATTACGTAACCTCTATG TACCCTTCCTGCGGTGTCT CTTGGGCACTGCTTCACTC TCTTCTTTATTAATAGAAA CGTGCACAATCTCGTGCCG TTCGTTCAAAAAAAAAA	MADVMAVMNSLHDQRFLNFDLLNRANIVLL PKHDGAENLSSFRPISLIHSVAKLFMKLLSMRL APTMRDIISKCQTAFIKGRSIHDNFLFVRNMA RRYHRNGTPMLLVKLDISKAFDSVRWDYLLA LLAHLGFPPKWRDWVASLLATCSSQVLLNGI PGQSITHGRGLRQGDPLSPFLFVLAIDPLQRLL HLATEAGILSRVAKARTRLRIAMYADDAIIFVK PEKRELDSFATLLHQFGEATGLRTNIQKSSVV PIKCAGLNLDEILAGFPASRSQFPIKYLGIPLSV ERLKKVDFQYLLDKVSSKLTCWQGRNFTFAG RLVLVKSVLSAQPVHTLTTINVPKVVLEKIDKF RKRFLWAGNENLTGGKCKVNWPTVARPLDL GGLGVLDIHRFARALRLRWLWREWEVPEAP WAGLQTPCNDTDKLLFAAATRIAIGDGAKTS FWHCAWVNGRRPKDIAPDIFAASSDCRLSVR EALTDSTWVNRIDLHAIHTADHLKQFVELWS FAHSVSLVQNTEDRIDWIYNASGEYSAASAYR IQFMGHTCSNLKALIWKPWGPSKCKLFAWLI IRNRVWTSDRLAARGWPHSPSCPLCKRHPET AHHIIAGCRYSRRIWDALASWLSIGPLAPAT WPDSTSIKQWWFMVGSVAVPKKGTRSLLLL VVWHIWLERNNRTFQRQERSVADLLAAIKAE ARLWGVAGAKHLLALLPML	cd01650	23	286
	LINE1-35_ZM	GCA_902373975.1	Zea mays	ACAAAACATGGTAC TGCGCTACGACGG AGCCGACGGGTTG CAGGGGCAGGGGT GGAATTTAACATGC AGGATCTGGAGAA	TAGCTCTGGTTTTGTTTTC AGGATCTGTTTGCTCTTGT TTTTGTTTTTTTCCTCTTCTT TTTTCTTTGGTTTTTTTATC CCCTCCGGGGAGGTCTTTT GTACTATGGTCTATTTTGG	MDPSKIICWNVRGLNSADRQNSVRNLVLSSK ADVVCLQETKMTVISTRTILAALGSDFCQFSFL PSCGASGGVLVAWKRHLGVTGLSRIDRYSISI QFCNAHNRCAWWLTCVYGPQSNSDKMLFL QELRDVRAACTGPWLIAGDFNLIYKASDKNN SNINRAMMGRFRKLVEDLSLNEVPLFGRKYT	cd01650	498	760

	LINE1-2_ZM	GCA_902373975.1	Zea mays	GGACCCTGTTGATG AGCCAGGGACCGT GGAGCAGCTGGCG CCCAGTACTGAGG ATTCTATAGCTGTC ACCTCTATTCCTTCC	CAGAGCCACAAAA CAAGTAATGCGGA CTCTGAGAGTCATC ACAGAGAATGCAA GCATCAACCAGCAA GCTGTTAAGGATTA TGAGAATCTTTTTC AGCATCCTTTAACT CAAAGTCATGTTGA AGCGTTAGCGGCT CTCTTTGGATGGGA GACTCCGGGGAAT CTCTTCCCGTAGTG TGTTTTTCTTTCTGG GTTGTAATTTATAA TTGTTGTA	ACCCTCTTTTTCCTCTACTT AATATAATGATGCGCAGTT CTCCTGCGCTTTCGAGAAA AAAAAA	TAGCAGTTTTCCTGCTGGT AGTCGTTTTTGGTCGGGTT GTTACCTTGTGTGTGTG TGGTGCGTTGTGTTATTGC CGGAGGGGTCTCGGCCTA GGGGCCTTTGTAATTAGG	WSNQQDSPVLVKLDRVLCSVEWEDNFPNVL LQSAASEGSDHCPLLLGIRDNKLGKRRFCFEA FWPKLEGFQDVIQEVWDSVAAVNCPFMTLN LKLKATARKLQSWSGKHVGHVRTQLALAKEIL HRLEIAQDCRALSLAEIWLSNLLKKHSLFLSSL QRTIARSRSRISWLKEGDANTKLFHMHSRHR KRRNFVAQLVNGDQILTNHDDKAVLVDNFY ARLIGECGDREQTINLEALHLPTFNLAHLDEP VTELELWNTIKGLPNDKAPGPDGFTGRFYKS CWNIIKMDLMAVVAAVWNRKFDNFCLLNS AFITLIPKHDGAVHVKDFRPISLVHSAAKLITKL LANRLSQKLNDMVSPIQSAFIKGRFIQDNFM LVQQTARLFHQQKRPRLLLKLDITKAFDSVSW PFLIEVMKYLGFGQIWRDIICGLLSSSSTQVLL NGIPGGHIEYRRGLRQGDPLSPMLFILVMDV LGLLFSKAEHAGLLLDLSPRRPLHRVSMYADD VALFLHPSAADIAITLDILHLFGEASGLRNNEQ KSNVYPIQCGGVELEMVQDLLPCGISTFPCKY LGLPLSLKNLTKVQVQPIVDRIADQLPGWKA DLMTRAGRKVQVQHVLTGMLIYLAMAVNIP SWAIKAIDKIRRSFLWRGRKEARGGHCLVAW GRVCRPLQLGGLGISSLKELIWALRMRWLWL QKTEPNRPWCSLPIKVPDKARAFFSMVLTSK VGNGANTLFWTDKWIHGQKIADLVPRLFDII PKKITNRRTVQEALTDRRWIADITGSLSVGVL LDYLHLWNLISDFELLPDIEDKHTFSIAANGTY SAKAAYDGLFIGAVQFGHYERVWKTWAPSK CHFFLWLSVLNKCWTADRLAKRGLDHPERCP LCDQESESLDHILVSCVFAREFWFRLMCIFRL QNLAPTPSTVSFTGWWEMIDNGSGDMIMK GVNSLIALGAWIIWNHRNRIVFDGISPSVSAA LCQAREEQQLWEMAGAKGLSFLAATIRGT	MNLSNILVWNARGLNNKARRDYVRDTILSSSK ADIICLQETKVASFSSHLLLSVCGSDFDKFLTLP AVGTRGGILIAWKGAVCQVISSRIDNFSVSVQ FAGMDGMNWWFTGVYGPQEDDRKIQFLQ ELLDFRTLCSGPWLLSGDFNMIYQAADKNNA NLNRALMGRFRRFLDDCLLKEIPLHGRKFTW	cd01650	496	759

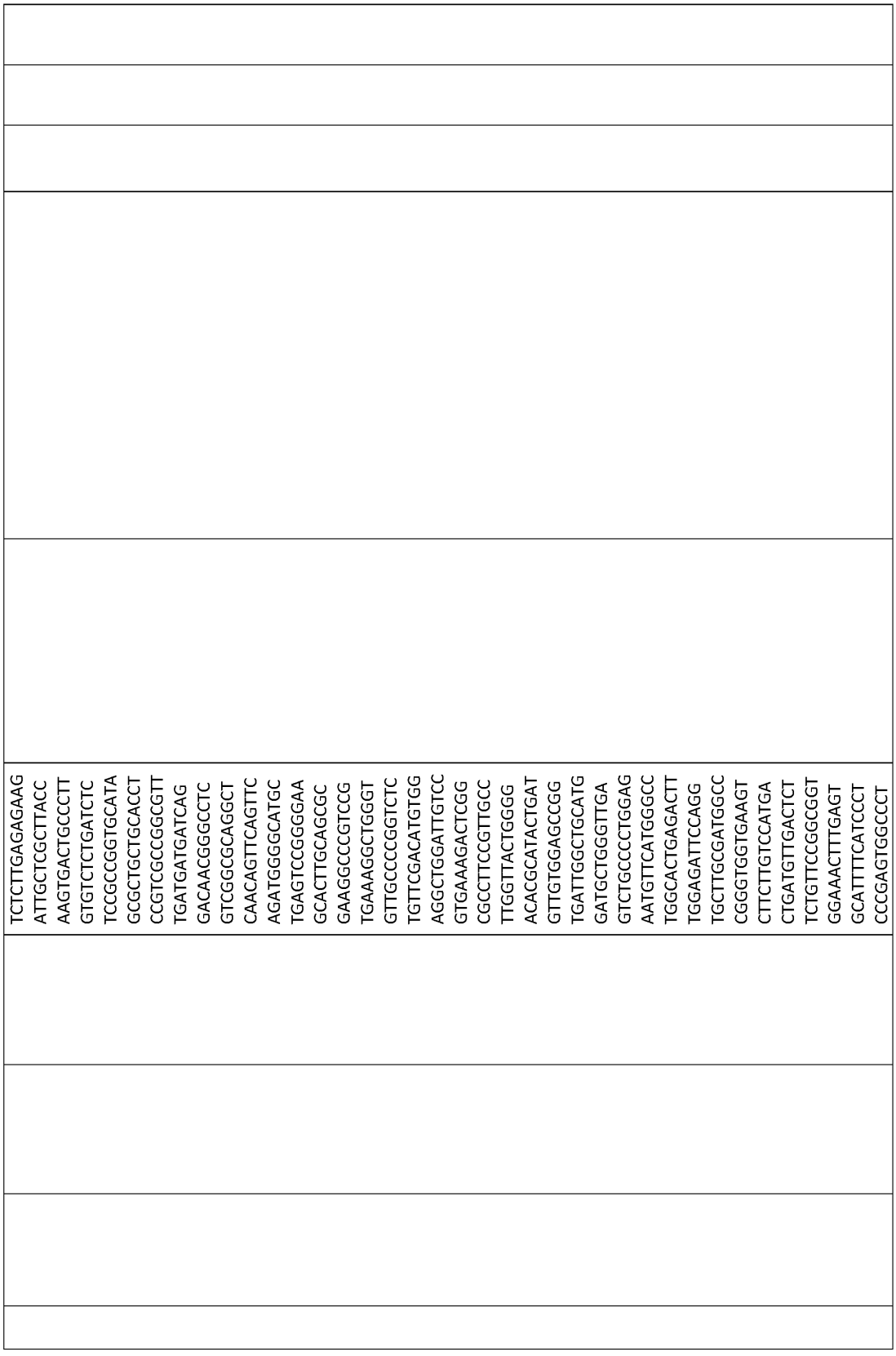
		GCGAATTTACCTCT	GGTTGACAAGTCTTGTCTC	SNERSSPTLVRLDRVFCCNDWEDFFPNSLLQS	
		GCTTACTTGGGAA	TCCTATGTTCTTTTTTTCTT	ASSGVSDHCPLILSLNVQTHKKRRFHFESYWP	
		GAGGATAGACTTG	CTTAATATAATGACACGCA	KLPGFLEAVSHSWNAPVSTACPVEGIFLKLQR	
		GTGCTGCTGTGGTT	GCTCTCCTGCGTTTGTTCA	LSKGLQKWSHHKVGNVKLQLATAKEILHRLEI	
		CCAGACATTGTGTC	AAAAAAAAAA	ARDNRVLSPGEEWLRKKLKLHCLGLASLERTI	
		TAACCAGGGTATTC		ARLRSRILYLQEGDANTSFFHQQAARYRKKKNF	
		CTCACCCTGTTGAG		IAKFQVEERIVTSQEEKQQAALDFYDNLLGTA	
		AGGCATAGGTCTG		EHRDYTLDLHSLGIQQHDLGDLELVFSMEEV	
		TTGAGCCGCAGGA		WSVVKDLPLDKAPGPDGFTGRFYKVCWDIIK	
		TGCTGCCATTAATC		EDMMGALLAVQCGHVSKLKLLNTAFITLLPK	
		GCAGAGGTGGCCT		KVDALLVKDYRPISLVHSFAKLVTKVLAARLAP	
		CCCTTCCAACCTTC		HLPSMVSINQSAFIKGRSIQDNFLLVQQLTRS	
		AGGTATATTTTAGG		LHHSKEPHILLKLDISKAFDSISWPFLLDMLQH	
		CGAAAGCGTCACTC		LGFGRSWCNLISLLLCTSSTRILVNGEPGEYILH	
		TAAGAAATGTTCTG		HRGLRQGDPLSPMLFILVMEALNAMVSYAF	
		CGCCGGATGCGCC		REHMLQPIANQHAKHCISFYADDVVLFMRPS	
		TGCTGTTTCTTTCA		HQDISAIISILDIFGHATGLKTNISKSSVTPIQCG	
		ATGACCTGGAGAA		ANELTTISALLPCATKNFPCTYLGIPLSIRRLAN	
		AATCATCAAGCCAG		SDLLPLVDKVADKLPGWKANLLSRAGRLVMIV	
		TGGTACAGCTGCTG		KTVLSSVPIYLMLALELPKWVFKAIDKRRRGFL	
		CCCATGCCGGTTGC		WKGQGDANGGNCLVSWEAVQRPLKYGGL	
		CAAGATTAGAAGC		GILNLEMFGWALRARWLWLQKTDASRPWA	
		AGTAAGACTGTTGT		GLPIRVHRNAKALVDVAITSVVGSGESVKFW	
		TGCTGTGCCTCGTC		SDRWLLGKTVAEHCPTLIQVISRRALKRRTVA	
		GGAGTAGGAGGGT		EGLTNRQWVSDIRGGLSVTVLVEYLQLWNLV	
		GGCTGGTGTCAAG		DGVVLQPDIADQHIWRLSAHGTYCSKSAYDA	
		CCTTGTTCTCCAGG		LFVGSIPFGPWRRVVWKTWAPLRCKFFVWLAI	
		TCCGATCATCACTG		KNRVWTADRLAKRGLPHPVACPLCDQAEETI	
		AAGCCCAAAAGAA		QHILVSCVFARQIWTSMLHNLGLLAIVPQHG	
		AGTCATCAGGAGTT		CTRFSNWWCQSIKKVEKSLRKGLNSLIILVAW	
		TGGGTCTCGCAGCC		EIWKHRNACVFEGVVPCTQRVQSAVIEEGNV	
		ACGGAAGAAGCCA		WCLAGASALQDLLLRQLPRGP	
		TTGATCAAAAAGCT			
		CAAGACAACTATTT			
		TAAGCTGTTCGAG			
		GGGCGCCTGACGG			
		ACTCTCATCTTGCT			
		GCACTTGCTGCTAT			

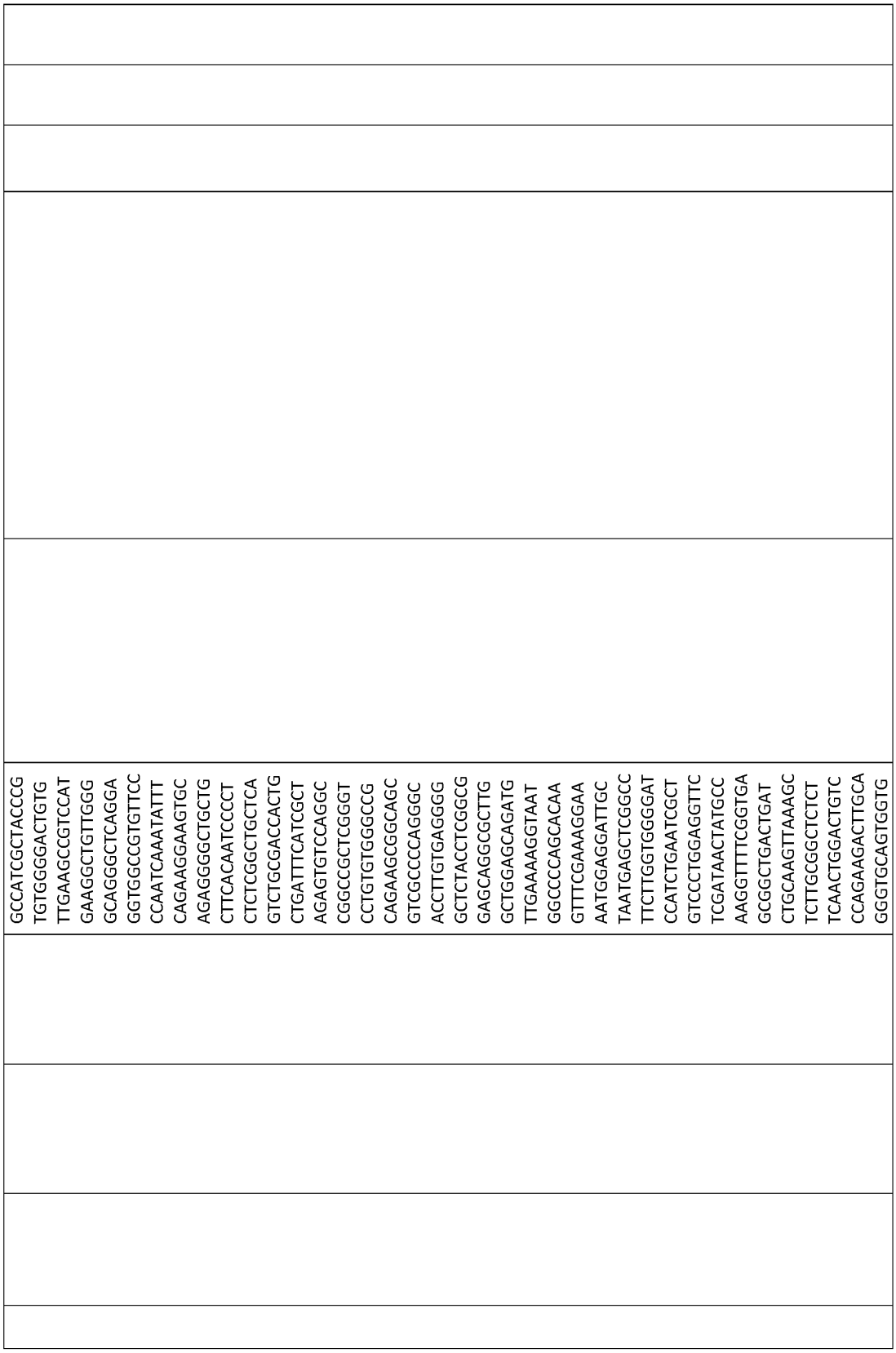
			TTTTGGCTGGCAGG CTGAGCGATTTGAT GATTCCAGCAGCCC AGTTCTCCTGGCAG GCTGTTAATTCCTG CTGTCCTTGGTCGT GTTTATTTGTTTTCA	TAACGTTCCTTTTTCTATG GAGGTCATTATTCTCATGA CATGGAGCATTTGGAAAA GTCGAAATGAATGGTTGT TTGAAAACAAAGACCCTTC AGTGCATCACTGTACGAAT GAATTCAGTAAGGAACTA CGGTTGGTTATTCACAGA GCGCGAGGGAAATACGAC ATCTCCATTCCAAGCTGGT TGGATCAGTGGCAGTCAT AAGATTTGTTTTGCTCTGA CCAATCTTTGTAACTTTTCA ATTAGATCTTACCAGCTAA CCTCTAACTCTTGTAATCA AGTCGAGTTGTACCTTTTC ATATTTAATAAAATTTTAC AGTAGGGGTTTCCCCTCCT GATCTTTCAAAAAAAAA	MPSDKLWLLVGDFNLIRRPEDRNKAGGDINL MLKFNEAISNLDLVEIPLHGLSYTWSNRQREP LLQRLDWFFISQEWSIVYPDTCAKTMPRDISD HVPCLVSFKSKIPKPKIFRFENFWLEFEGFMNI FQTTWLGLPCLPDKAKNLTAKFKIVRKVLKD WQRSLPKIDKTVRDIKLLIEFIDNIEEHRDLSIE EWNFRDIMQQKVAELLNIQKIYWKQRASIK WITEGDICSRFFHAHATVKHRRNTIASLTDEN GIIHFEHDHKSNLLWEAFKSRLGSSDFLGIDFD LSELLIRNEDLQGLDSPFTKQEIDNIIKSLPSDK SPGPDGFNTNFIKKCWHIIAQDFYDLCEKFYQ EEVCLRSINGSFIVLIPKKENPQKVGDYRPISLL NNNMKILTKLLANRLQSFMTRLVHRNQYGFI KGRTIQDCLAWAYEYIHLCHSSKKEIIVLKLDF EKAFDTLEHELIMQVMTHKGFGSKWINWVR NILHSGTSSVLLNGVPGKSFHCKRGVRQGDP LSPLLFVLAADLLQSIINKARHLNLLQLPLPDIC EQDFPIVQYADDTLLIMEACPKQLFFLKAILNS FASSTGLRVNYNKSNMYPINVNPEKMEILSRT LNCQIGAMSFTYLGLPLGLAKPRIQQFLPLIQR IEKRLSCSSNFLSQAGRLELVNSVFSALPTFFM CTLKIPKSIIKQIDIYRKHCLWRGNDTSSKKPPL VAWSTVTQRKVNGGLGVLRLETQNDALLLKF FHKFFNSYDLPWVNIIWNKYYTPNRLPGHRRI GSFWWKSMLKLLHNFKGLAHPIIGNGRSILF WEDQWNKSIPRQQYPELFSFTKSTKLSIAEVK EQDHLFRIFQLPISEQAYEQYLELQVVWEQIV LSNVKDRWRYIWDSDSYSTQKAYRHLMGQT QVHSIYKCLWKSKCQPKHKVFYWLWLKNRL NTRNMLRRKNMTLESYSCENCLWQREETLY	cd01650	365	629
	LINE1-18_ZM	GCA_902373975.1	Zea mays					

	LINE1-28_ZM	GCA_902373975.1	Zea mays	GGCGGTCATTTCG	TAATGTTTTTCCTATCCGAT	HLFLRCNFAKACWNSIGIPPTQNCSSRGGLG KPQTTT	cd01650	506	704
				GCGGACCGTTCGG	GTTCTGAGGCGGATCTTG	MVVFSMFFQMDPSKIVCWNVRGLNGSARQ			
				CTCCCCGCCCTCCC	CTTGTGTTCAGGAGCTCCT	DVVRCLVDSVKADVVCLQETKMSVFSRLFVF			
				GCCTTCCCTGCATG	TCCATGTTCCATTTCGGAA	RLLGSEFQNFVFTPSVGASGGILVAWKNNISS			
				CCCTTCCGCCTCCC	TTTCCCTGTCAGTACCTTG	CGSRIDDHCVSVNFSLETGSSWWLSCVYAPQ			
				CCGCCCATCGCTCC	GCCTCCCGCTTTCTCTTGG	GNHEKIQFLQDLRLIRTHCSGPWLVVGDFN			
				GTTAACGGCGCCC	AAAGCTTAAGAGAAGTCA	MILCDEDKSNSNVDRAMMGRFRRWKDDLA			
				GCCCCACCCTCCCC	AATTCAAGCAATTGTCGAT	LLELPLIGRRFTWSNGHACPTLVRLDRMFCTA			
				GCACCACCTAGCCC	AAAACTGCAGCACAGTTA	DWEDLFPDCLLQSAATQDSDHCPLLLGLCDI			
				CTCGCCGGCGCTCA	CCTGGGTGGAAAGCTGAT	NKRKGRFHFQAFWVKLEDFQDVVANAWSSI			
				CCCCGACTCCGGCG	CTGATGACAAAGGCTGGT	PAGQCPLVSLSKKLQATAKALQRWSDNKVG			
				ACCACTTCCTCCTC	AGAGCCATTTATGTTCAAT	DLKLQLEMARELLHRLEMAHDSRTLSPSEDW			
				CCCCTCCCGTGTGG	TTGTGATGACATCAAGAAT	LRRTLKKHILVLSSLSRSIARLRSRINWLKEGDA			
				CGACTTTGTCTCCC	GATTTATGCTGCTCTGGCG	HTSLFHAQARIRKKKNFIGSLSTSEGLILTSHED			
				GCGGCTTGTCCGTT	TTGGACATCCCTTCTTGGG	KADELFHFYDALLGTALGRENSLNLDALEMPS			
				CTCCCCCGCTGGGC	CTTTTAAAGCAATTGATAA	FNLNELDVPFSEEEVWGTVKSMPSDKAPGP			
				GTTCCAAGGCGCA	GCTTAGACGAGGCTTTCTG	DGFTGRFYKVCWVTIKADVMAAISCVWARK			
				GCGGTGGGCGTCT	TGGAGAGGGAGGAAGGA	FRNMGLLNSAFITLLPKVEGATQVKDFRPISL			
				TCCTGCCTCACCAG	GGTGAATGGTGGCCATTG	VHSFAKLVTKMMANRLARHLNDMVSPNQT			
				CCCGTCCTCATCGG	TCTGCTGACTTGGCCCAAA	AFIKKRFILDNFMLVQHTARFLHRQRQARILF			
				CCTTGGGCGTTGGT	GTCACTAGACCCAAGGAC	KLDISKAFDSVSWAFLLEVLQKLGFGQIWRDII			
				GCCGTTGGAAGCG	CTGGGGGGTCTGGGTATT	CGLLASSSTRILLNGVPGDYIAHQRGLRQGDP			
				GTGCCTCGCAGAG	AGTGATCTTCGGATCATGA	LSPMLFILVMDTLNLLIQRASDEGLLHSLSTNS			
				GACTTTCAAGGAC	GCTGGGCACTGCGGGCTA	LQHRFSLYADDAVVFLKPAAFRHKSDSGNPS			
				GTGCTGCTGCTCCC	GGTGGCCTTGGCTGCAAA	DFWGCVRSKDQRSEE			
				TCCGCCCCTGCCGT	AATCTGATAGGACTAAGC				
				CTGCCCGGCAGTCC	CATGGTCGGACTTCCAGAT				
				CTCCTCGACCTCCG	TCAAACTTGCGACATTGTG				
				CACAGACCGAAGTT	CAAAGCTTGTTGGCTGCG				
				TGCGGTGGTGCAA	GCTGTTATCACTCATTTGG				
				CGTCCCTGGAACG	GAGATGGGTCTGATACTC				
				GTGCGCAGGCCTG	TGTTTTGGAAGGATCGTTG				
				TTCCTGACCGGGA	GCTGAATGGCAAATGTATT				
				GGGCTGGGTTAGG	AAGGAGGTGGCACCTGCT				
				GTCGAACCTCGTCG	GTCTTCGGTATGGTGCCCA				
				ACGTCAACTCCCCC	AGCGTATTGCTTCTAGACG				

				CTTCGCCACCTCCC CGTCGATCGATCCC TGAGGATCTCCGG GTAGGTGTTTTAAC TGTCTGGCCTCCTC TCATCGGGCGGCG GATTGCCGACGTTC GGTGCGCTGCTTCA AGTGTTGGGGGTT CGGACATCGTGCA GTTCAGTGCAAGG TTCGTGCTGCGCAG GTGATCAAAGCTTC AGTTCGGTGTGGG ATCGTCTTGGCCCT AAGCTGGTTCACAA GCCTTTGGACTCTC CTAGACGCGCCAT GGAGTGGCGCAGG GTCTCTTCGCCGAT CCCGGCCATCCAGA GGTCCGGTGTGGC TGATCCTACCTTGG GTGGCACGAAGAA ACGCCGCCATCGTA CGCGGGGTCGAAG GCCTAAGACCGCCC CCGTTGCACCTTTG GAGGAATCCAGTG AAGTTGCACCTCCA CCCGGTTCCTCTTC CCTAGAGCCGGTC AGGGTGCATGCCC CTCAACCGGTTGTT CCTGATATGGTGG AAGCCACACATCTC TGCATTTTGGACCG TGCGGATGCGCTT	CCTGGTCAAGGATGCCTT GCCTGATCTCAATTGGTTG GCTGATCTGAGAGGGGCT ATTTCTGTTACAGCCCTTG CTAATCTCATGGACCTTTG TGACTTACTCGAGGGGAT GGTTCTGCAGCCGGATGT TCCTGATGTGCATACCTGG AGGTTCAGTGAATCGGGT GAATACTCGGCTAGTTCAG CTTACAGGGCTTTGCTTCT TGGCTCTGTTAGCTTTGAG CCTGCTGAACGCATTTGGA AATCCTGGGCTCCAGGAA AGTGTAAATTCTTCCTCTG GCTTGTGGAGCACGATAG ATGTTGGACGGCTGACAG ACTTGCTAAGCGTGGCCT GGATCACCCAGAGTCCTG CCCTTTATGCGACCAGATG GAAGAAACTATCAACCATT TGCTGGTTCAGTGCAGTTT TGCAAAACAGTTTTGGTTC AGCCTTCTGCGGTCGGGT GGTTTCCAGGAGCTCTGCC CTTCCCATGATGACGTCTG TTTTGAAACCTGGTGGAG GAATAGTGCTGCCAGGGT GGGCATGCCTTTATCCAAG GGATTCAATTCTCTTGTCA TTTTGGGTGCATGGACTTT GTGGAAGCATAGAAATAG GTGTGTTTTTGATAACTGT AACCCTAGCGCTGCGGTG GTGTTGGCAGCTGCAAGG GAGGAGTTGATGGGTTGG CGTATGGCTGGAGCCAAA GCGCTCTCTTTCTTTCATTT				

				AGAAGAGGTGTCTTAGGT GTCTTAGTTGTTTGGGGA GTGTTTTTTGGTTTTGGTTT GCCAAATGTAACACCTTTG TACTGGGGTTTTTTCTCCCC CATTTTATTCTTCTTCTTAA TACAAAAATACGCAGCTCT CCTGCGTATTCAAGAAAAA AA	
			GCTAGAGAAGAAC TGCGGTTGCGACG GTGTGCTCTGTTCG TTGTTGTCACTGGT TCAAGACCGAGGG TGGCTGCTGAGACT CTTGCTGATGAGGT GGCTGCTGGTTTCG GGCTGGATGCGTC GTCCTTCTCCGTTC ATCGTTCGTGCCCA GAAGATTTTGTCTT AATCCTGAATTCGG AAGAGTTGGCCAA TCAGGTGTACAACA ATGGTGCTGTTTTC AATTCTCTTTCTGG GTCTTTCAAGTTCC TGAGATGGTCTAG ACTTGCCCATGCCG AGGTGGTGGCCTTT CCATCTCCTGTTCTT GTGGAGTTCAAAG GGATTCCAATTCAT GCTTGGGATTTGGC TACTGCTCAGAATT TGTTGAGGGATCA CTGCACTGATGTGG AGCTCCATCCTGAC ACGGTGAATCGAC ACGACTTCTCTTCC TTCAAGGTTTTTGG ATGGTGCCGACGT CCACACCTCATTCC AAATTTGGTGGAG CTCCTTGTTCCGGA ACCGTTGATGATGA CGCCTGGCTCTACC			


TCTCTTGAGAGAAG
ATTGCTCGCTTACC
AAGTGACTGCCCTT
GTGTCTCTGATCTC
TCCGCCGGTGCATA
GCGCTGCTGCACCT
CCGTCGCCGGCGTT
TGATGATGATCAG
GACAACGGGCCTC
GTCGGCGCAGGCT
CAACAGTTCAGTTC
AGATGGGGCATGC
TGAGTCCGGGGAA
GCACTTGCAGCGC
GAAGGCCCGTCCG
TGAAAGGCTGGGT
GTTGCCCCGGTCTC
TGTTCGACATGTGG
AGGCTGGATTGTCC
GTGAAAGACTCGG
CGCCTTCCGTTGCC
TTGGTTACTGGGG
ACACGCATACTGAT
GTTGTGGAGCCGG
TGATTGGCTGCATG
GATGCTGGGTTGA
GTCTGCCCCTGGAG
AATGTTCATGGGCC
TGGCACTGAGACTT
TGGAGATTCCAGG
TGCTTGCGATGGCC
CGGGTGGTGAAGT
CTTCTTGTCCATGA
CTGATGTTGACTCT
TCTGTTCCGGCGGT
GGAAACTTTGAGT
GCATTTTCATCCCT
CCCGAGTGGCCCT


GCCATCGCTACCCG
TGTGGGGACTGTG
TTGAAGCCGTCCAT
GAAGGCTGTTGGG
GCAGGGCTCAGGA
GGTGGCCGTGTTCC
CCAATCAAATATTT
CAGAAGGAAGTGC
AGAGGGGCTGCTG
CTTCACAATCCCCT
CTCTCGGCTGCTCA
GTCTGCGACCACTG
CTGATTTCATCGCT
AGAGTGTCCAGGC
CGGCCGCTCGGGT
CCTGTGTGGGCCG
CAGAAGCGGCAGC
GTCGCCCCAGGGC
ACCTTGTGAGGGG
GCTCTACCTCGGCG
GAGCAGGCGCTTG
GCTGGAGCAGATG
TTGAAAAGGTAAT
GGCCCCAGCACAA
GTTTCGAAAGGAA
AATGGAGGATTGC
TAATGAGCTCGGCC
TTCTTGGTGGGGAT
CCATCTGAATCGCT
GTCCCTGGAGGTTC
TCGATAACTATGCC
AAGGTTTTCGGTGA
GCGGCTGACTGAT
CTGCAAGTTAAAGC
TCTTGCGGCTCTCT
TCAACTGGACTGTC
CCAGAAGACTTGCA
GGGTGCAGTGGTG

			CAGGCCTGATCGCC TCTGTGTGTATTG					780
	L2-2_SSa	GCF_000233775.1	Salmo salar	TACATTACAACGGC TTGATTTGTTTGAT CTGAGCAATTCTTC TTAGCTAGCTACAT AGCCGTCTTTGTAT CAAAGATAATTGTG TAGTTTAGAGTAAT TAGAGTAATTATCG AGGCTAGCTATCTT CGTCCTCCTAACGT AGTCAACACCGCTG CTAGCTAGCTAGTC ATCACTGCTAGCTA GCCAACTTCTACCG ACTAGCAGCACTGT AGAAACTATTACAT TACAACGTAACGAC TTGATTAGTGTGGT GTTAGTGTTAGTTA GCCAGCTACATAGT TGTCTTTGCTGTCT CTGTATCTAAGATA ATTGTGTAGTTTAG AGTAATTATCGAG GTGAGCTAGCCAG CCGCGACGCGGCG CCAGACTTAGTCAA CACACCTAGTCATC ATTAACCCACTGCT AGCTAGCCAACCGT TACCGACTAGCAGC GCTGTAGATACTAA TACTTTACAACGGA ACGATTTGACTAGT GCAGTGTTAGCTA GCTAGCTACTAGTT	TAGGATAAAGTAATCCTTC TAACCCCCCCCCCCCTTAA AAGATTTAGATGCACTATT GTAAAGTGGTTGTTCCACT GGATATTATAAGGTGAAT GCACCAATTTGTAAGTCGC TCTGGATAAGAGCGTCTG CTAAATGACTTAAATGTAA ATGT	MCLSIPVLSPLHRPYKRSTPRGRCHSNLVVPA RTTHVEFQVSGSLWNCRSAANKAEFISAYATL QSLDFLALTETWITTDNTATPTALSSSAYVFSH TPRASSQRGGGTGILISPKWTFSLSPLTHLSISS FEFHAVTVTSPFKLNILIIYRPPGSLGEFINELD ALISSFPEDGSPLTVLGDFNLPTSTFDSFLSASF FPLLSSFDLTLSPSPPTHKAGNTLDLIFTRCCSS TNLIATPLQVSDHYLVSFSLSLSSNTSHSAPTR MVLRRPNLRSLSPATLSSSILSSLPSAQTFSNLS PDSASSTLLSSLSASFDFLCPLSSRPARSSPPAP WLDDSLRAHRTGLRAAERKWRKTRLPADLA SFHSLLSTFSSSVSAAKATFYHSKLQASASNPR KLFATFSSLLNPPPPPPPSSLSADDFVNHFEKK VDDIRSSFAKSNDTAGPAHTALPCALTSFSPLS PDEISRLVTAGRPTTCPLDPIPSSLLQTISGDLL PYLTSLINSSLTAGYVPSVFKRARVAPLLKKPTL DPSDVNNYRPVSLLSFLSKTLERAVLGQLSCYL SQNDLLDPNQSGFKTGHSTETALLCVTEALRT AKANSLSSALILLDLSAAFDTVNHQILLSTLSEL GISGAAHAWIASYLTGRSYQVAWRESVSAPR ALTTGVPQGSVLGPLLFSLYTKSLGSVISSHGL SYHCYADDTQLIFSFPPSDNQVANRISACLAD ISVWMTDHHLKLNLGKTELLFLPGKDCPFHD LAITVDNSLVSSSQSAKSLGVTLDNTLSFSTNIK AVTRSCRFMLYNIRRVRPCLTQEAAQVLIQAL VISRLDYCNSLLAGLPACAIKPLQLIQNAAARL VFNLPKFSHVTPLLRSLHWLPVEARIRYKTMV LAYGAVRGTAPPYLQALIRPYTQTRALRSSTS GLLASLPLRKHSSRSAQSKLFAALAPQWWNK LPHDARTAESITTFRRHLKPHLFKEYLG	cd01650	519	

	Dong-1_MMa	GCA_000484575.1	Mesobuthus martensii	TAACCCCTCCTGAA TCCTGCTAAGCTGC CTCAAAGTGGCGG AGGGGTGTACGTG GGCGGAATAGGCG	GTCTTTGTCATAGC TTGATAATTGTGTA GTTTAGAGTAATTA TCGAGGTTAGCTA GCCAGCTATTTCCG TCCCCCGCGACGCC ATTTTTCCTAACCC AGCCAACTATTACC GACGAGCAGCATT GTAGAAACTAAATA CATTACAAAGGAAC GTCTTGATTAGTGT TATGTTAGCTAGCT AGCTACAAAGTTGC TTTGTATCATGACA AGGTGTAGTACTG AAACTATCGAGGTT ACCTAGCCAGCTAC ACGTTCAAACAAAG TCAACAACGCAGCC ACTGCTAGCTAGCC TACTCCACCAGCCA GCAGTACTGTATCA TTTTCGTCATTTTAG TCAATAAGATTTTT GCAACGTAAGCTTA ACTTTCTGAACATT CGAGACGTGTAGT CCACTTGTCATTCC AATCTCCTTTGCAT TAGCGTAGCCTCTT CTGTAGCTTGTCTA CT	TAACTCACAATTAAACATC ACAAGTAGCTATGTCTTGA CTCTTGTCCGACATAGCAC TTTGCCCCGGAGTTATTCC GTGCTCACATTGATGTACC	MLAGVPTLVVFFQYAMAQSRKRKHTGSDVS RVNKVRVCELGRGSHSESLLQSPLATGGESST NPSLATAIQIPRVGNGSPHQSPTSQSVSLETQ PETTKAGKPRQRMKWTTDMNEFIMRTYYTI TKLETDITGYRRELYDKFTSRYPNITVTEQRIAD	cd01650	534	805

				ATGCAAGTCAGAAATTATTTTCTAGTAGGGTCACCCTGGGCTAGAAGGCTATTCCATTATGCTCAACTAA	ATAATATAATAATAATAATAAATACA	QRRTIIHRNLIPHSTLESIKTQVAQQISQHETDSHTQIEHDVEQTPEITYIDDNTQPINNTDTDLINKLTDTLQETLTKFTGTDPTQRPRIPKLHNSQQLNNTLHCMNTHILPKLINAETTLIEIHTLIYAAAVTVTTHLGLKTLPNKIGTNTKHHLKPAWQKRLEKQINEIRTELGRLTQYTKNNTSTKLMKSIRKIFRKLKTHTRHENHNTHIKEYIDTLKQKLAAKSQRLKRYKTSLARKLDNTLFTNNEKTFYRNINTKDPKQTQNREIPTQQSLTSFWSALWAVPVTHNQRASWIAEEETRTLDVPEMPYTTITTEELTNTIRYTHNWKAPGVDGIQNFWYKHFTCLHDLLAKHLTHLIHEPKDTPSFLTHGITFMKPKSTNTGNPSQYRPITCLPTLYKILTSCITNKIYAHIEENNILTEQQKGCRRSHMGCKEQLIIDTTILKQVHRQHRHLHAAYIDYKKAFDSIPHSWLIQTLQLYKINTKYITFLQNIMQHWRTTLHITINNNHVATEPIHIRRGIFQGDSLSPLWFCLALNPLSNKLNTSNCGYDIRQNQTTHYIHSHLFYMDDIKLYASTQTQLQQLLQITHTFSKDICMEFGLDKCKTLHIEKGVLKKGGFTLENGDTIESMEEGEVYKYLGFQQARLTEQKKTKQTLSKLYTNRLNNILKTHLNGKNTVKAINTFAIPILTYSFGVIQWTRTELESLQRKTRTTLTKYQYHHPKSAIERTTLPRKEGGRGLIDIPHLHDRQIYNLREFFHKKQHTSQLHKAIVLADRNITPLNLQNNTPCQIDNTAHIQEKKTSWCKKELHGRHPHDLQKHFIDAKASNSWLVRGDLFTETEGFIIAIQDQVINTKNYRKYILKDSTLTNDTCRKCHIAPETIQHITSGCKLLAQTDYLHRHNQVANIIHQKLATKHKLINTLTPYYKYIPDTILENSTHKMYYDRTILTDKTIHHNRPDITLIDKQNKVTYIIDISIPDTHNIQDAFTQKISKYTELSQEIKRLWGMQKVVTVPIILSSTGVIPHTLHKSLKTLDLHPKTYITLQKASLLNTCNIVRKFLSTDTDS			
	SR1	GCF_000237925.1	Schistosoma mansoni		TGATTTACCAAATTCTTTTTTTCCCTCTATTATCGTTCATATACCTAGGTTCTTGCCTGGAGGTATTGGTGATCCACTGCTACCAGACACGGAAG	MEFLVDDLSILAPLGKSDHAVLSFSFVSKTELRYPTSNKRWNFKRLNVLALQDYLQQVDWDVHPQLEVDAHWDFLLHTILCATEHSVPKMVPKSYKQPPIIKNRTRRLLSRKRHCWAEYKRTDNNGAYRQYKHIRNICTKAIREDRLQFQTKLIDKFV	cd01650	288	537

				CCCGTTAAGCGAAAGCTTC TTCTATTCCGTCCTCAACC ATTTGAACCATTTG	SNPKSLFSYAASLRQGKTGVSQLLGPNGPTN NDSDAANLLAEQYSRTFQLTHINHTDESFTCT CTGLSEVDLSADLVLRKLQHLRKDTSPGPDM VHSAVLREAASILATPLSVMFAHSLSRGKLPEI WKLAHITPIFKGGRRSEPSSYRPVALLSIPSKI MESLVYDGILEYLSSSKFFSPQQHGFRKGHSC MTNLLTAVDRWTTILDRKGKVDVIYLDFSKAF DRVNHACLIKKLGRLGIKPPLIDWLSSYLENRH FKVRVNFTLSQAMECPSGVPQGSILGPLLFLIY INDLPQQVSSDLLLFADDVKLWREIRNHNDIL VLQEDLTRLQSWADDNGLTFNTSKCKVVHLR HVADYSYNLGNSPLEVSQVEKDLGVLVPYDL KSYANCDKNASQANLALVTLKRIFGQFDGRT FHIIFNSFIRPHLEYGNIVFPPSLQKDKDTLERI QRRATKSVRGLKFKPYEERLKSLNLYPLEYRRL RGDLLMTYSILNTSGHPLKHLLKLSHNTNLRG NTQKLETLHSRTDCRHNFYSVRVVKCWNSLP TELVQATSQESFKRKLDLFLRTKDNILL	cd01650	1008	1259
	CR1_AC_1	GCF_000090745.1	Anolis carolinensis	TTGATCCCAGAAG GGCAGTTGGTTCTC GCAGAAGCAGAGG CAGCAGGAGGACA TTTTTGCTCCCTGG CTTCAGGTAGGATT TGGCTAAGATTTGG CTAAGATTTTAAAC TGAGTGTGGGCTT GGCTCCTGTGGTCA AAGTCTAACTGTTG TGTGACTGTTGTGT TGAAAGAGAGCCA GGTACTTAAGTTGA TTGACAGCAGGCA TTGTCCCCTGTTAA TTGAGTTTCTGGGA AAGCTTTATTTGCC AGTGTGAGTTTCTG CCAGAGCCTGATCC	TGAATGCGATTTCCTGCTT CTTAGCAGGGGGTTGGAC TAGATGGCCCATGTGGTCT CTTCCAACTCTACTATTCTA TGATTCTATGATTCTATGA TTCTATG	MQQKQNIPITCTSCGMFRFFTQKLDNYICSK CKQMTLMEQRIQQLEHRIKTLKDIQALELFLD TAQHAAVDQQPASHQHHQDHDQEPYGEIN SKVDNPQAWKEVTHRRRRRTRQPPQNSSA QLHLHNRFEILTSLTYNQETHLVEDHSFLDTT QWTTLDQCAGDSPDGDSASPQSHLCNHAN APSPIIDQEQQEHPLGETRSNRNIPWESNGLS DVSQWIVIDECTGDVEEEDNTLHLHDSLQQE HSSGDIHTVLHKRDPVNPQRKQVLVVGDSLL RGTEAIISRPDGMARETCCLPGAKIHHITQRLS RLLKPHHPPHLMLIHVGTNDTARHTFQKITN DFRALGTKLKLYNVHVIFSSLLPVVGHGSTRA GKIVQVNNWLRKWCQEEHFGFLDHGLLFQE DGLLASDGVHLTQVGKHLFAHRLTNLIRRTLN PWSTLPRGWTTGKRWGASHTSRKTSFCTQT HKPHQAHFKLDPLGEGNNSLANTILPTTTGN RRKAKRRAAQTQQGPSTESTIIPNKQFEGRS QGLTCLYTNAQSMGNKQDELQLLAQHHTYD VIGITETWWDDSHHWNLTIEGYNLFHRNRT KGRGGGVALYVKNSYVAEEMQDCNPGNQL			

				ESIWIRIKGTGTQKDLVVGVYYRPPSQDEGLD EAFCQQLTKQAQRRDIVVMGDFNYPDICWK TNSAKSTKSNKFLTCLADNFMVQKVEEATRG SATLDLILTNVEDLINTVEVVGSLGASDHVLLQ FAIQRNAETKTSQTRILDFKRADFQKMKELLS GIPWTPILKNKGVKDGWEFFKSEILKAQMQT VPTKKKNKTSAKKPEWMSKELLTELKLKSDM HKKWKRGEITKEEFKRIANTCREKVRKAKAQ NELRLARDIKNNKKGFFAYVGRKRKKKEAIGP LQGEDGVMATGDREKAELLNAFFASVFSEKE SHLQPQQHGMDEGLGEIQPQIGKQVVQEHL ASLNEFKSPGPDQLHPRVLKELAEVISEPLAIIF ESSWRTGEVPADWRRANVVPIFKKGKKNDP NNYRPVSLTSIPGKILEKIIKEVVCEHLETNAVI ANSQHGFTKNKSCQTNLISFFDRVTSWVDTG NAVDVAYLDFSKAFDKVPHDLLANKLVKCGL DKTTVRWICNWLSERTQRVLTNASSSSWKEV TSGVPQGSVLGPVLFNIFINDLDEGLEGTIIKF ADDTKLGGIANTPEDRSRIQNDLDRLERWAK TNKMKFNRDKCKILHFGRKNGNQRYRMGD AWLDSSVCEKDLGVLVDNKLNMSQQCDAA AKKANGILACINRGIASRSREVMLPLYSALVRP HLEYCVQFWAPQLKGDVDKLESVQRRATKM IKGLENKPYEERLKELGMFSLQKRRLRGDMIA MYKYVKGSHREEGASLFSAALQTRTRNNGFK LQERRFHLNIRKNFLTVRAVRQWNSLPGAVV EAPSLEAFKQRLDGHLSGVL		
			CAGAAGGGCAGTT GGTTCTCGCAGAA GCAGAGGCAGCAG GAGGACATTTTTGC TCCCTGGCTTCAGG TAGGATTTGGCTAA GATTTGGCTAAGAT TTTAAACTGAGTGT GGGCTTGGCTCCTG TGGTCAAAGTCTAA CTGTTGTGTGACTG TTGTGTTGAAAGA GAGCCAGGTACTT AAGTTGATTGACA GCAGGCATTGTCCC CTGTTAATTGAGTT TCTGGGAAAGCTTT ATTTGCCAGTGTGA GTTTCTGCCAGAGC CTGATCCCAGAAG GGCAGTTGGTTCTC GCAGAAGCAGAGG CAGCAGGAGGACA TTTTTGCTCCCTGG CTTCAGGTAGGATT TGGCTAAGATTTGG CTAAGATTTTAAAC TGAGTGTGGGCTT GGCTCCTGTGGTCA AAGTCTAACTGTTG TGTGACTGTTGTGT TGAAAGAGAGCCA GGTACTTAAGTTGA TTGACAGCAGGCA TTGTCCCCTGTTAA TTGAGTTTCTGGGA AAGCTTTATTTGCC AGTGTGAGTTTCTG		

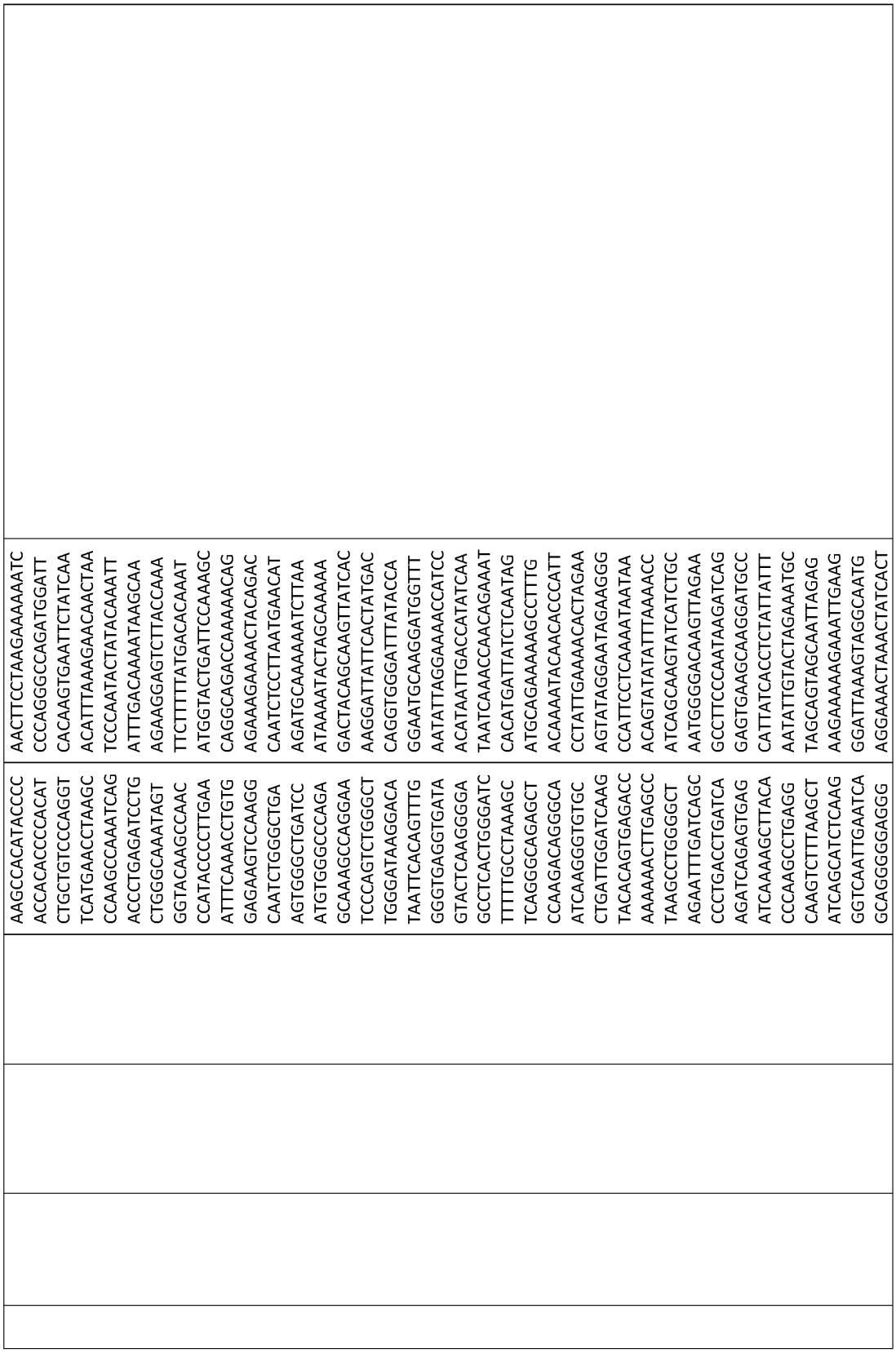
			CCAGAGCCTGATCC CAGAAGGGCAGTT GGTTCTCGCAGAG GCAGCAGGAGGCA CCTTTACTAACTAT CCTTTGCAGTACAT ACAGGAAACAAAG CAACATAAACAAAT ACACAAAGAGGTG GTTTGTTGCAGTTT GCACATAAAGTGC GTGCATATTACTTA ACTGTACAAAGATC CCACTTGGCCACCA CGCTCACCAAGAG	AGAAACTCGCACA GATACAGACAGAC ATCATCTTTCTATCC AAGTGCAAGAAA	TAACAAGTCTTTGAACTTT CTCACCCATTATCAAGACA GTTTCCCCAATTATCACCT GTAATACCATTAACTCACA AACATCCCACTCTCCCTAC CTCTAATATCATCAATTCA CAGACACTTACCTTCCTTC CTCCCCCTTCCCACCCCCCC GCCTCCTCCTTCTGTTCTGC AATGTGATTTGTCCTTTTC ATATTTGTTCATTTTTTTTA ATTGTATCCTTTGGTATAT ATGGTTGTGACTACTTTCT TCCACTATTTGATCTGAGG AAGTGGGTCTGGCCCACG AAAGCTCATCATCTAATAA ACCATCTTGTTAGTCTTTA AAGTGCTACATTGTCCTGC ATTTTGCTTCAACTACCCC AGACTAACACGGCTACATT TCTATC	MDIVPKGLKVNNPLRSTYFTDYSERLCLTLSKK LRNHLISILYKKQKTIKNDISNLDRLISRQTSTPT DTLQDFCSTRQEIYYTHFYSLQQRKDNKFSNL LHTTGYNSNCLNPPDNIVNLSSYKLNPAEESV LSRGLSFCPTSPTNWIQFCGDLEAFFRRLRLRE YFQTTTEHQSDLPDPPYQHQRKNNSIWTPP DGRNYNLDFYIHSFRNNAQTDIIRKRQQATH NLSRAEHHAIQSLKNNLDIVIKPADKGGAVVI MNKADYNQEATRQLSNTTFYKPLPSDPTLDY QKKLHCLLKSLPTATREQILTESPSDPRPGLFYL LPKIHKPGHPGRPIISGIGTLTTGLSSYVDTLLK PFVTNTPSYLRDTTDFLRKLQNIDNLPNNTILA TMDVEALYTNIPHEDGLQAIRNTIPEDTTANL IADLCNFVLTHNYFRFENNLYLQISGTAMGTR MAPQYANIFMADLEQRFLNSRPLSPLLYLRYI DDIFMIWTHGQETLEIFHRDFNNLHPTINLSL DHSTREIHFLDTTVQINNGKLDTTLYRKPTDSY SYLHASSSHPEHTTRSIVYSQALRYNRICSNPT DRDQKLQDLYQAFINLNYPPRETKKQIERARR IPRNHLLQDRPKKTNNRTPLVITYNPQLKPVQ HIINKLQPILEQDTILKEALGDRPIVSYRQPPNL KMILTNHHRTYHTNTNPGTFPCNKPRCQLCP HIHSADTIIGPNQVSYKIKNTYSCASRNIIYAIM	IPR00 0477	305	558
	Penelope- 2_PSi	GCF_0002305 35.1	Pelodiscus sinensis				

					CRKCPSAMYIGQTSQTLRQRINAHKTDIRQD HKEKTVSCHFNQKGHSLNDLATCILLQRPFTS ALERESSELSFMLKFDTYQPGT			504
	Penelope-2_PSi	GCF_000230535.1	Pelodiscus sinensis	AGAAACTCGCACA GATACAGACAGAC ATCATCTTTCTATCC AAGTGCAAGAAA	TAACAAGTCTTTGAACTTT CTCACCCATTATCAAGACA GTTTCCCCAATTATCACCT GTAATACCATTAACTCACA AACATCCCACTCTCCCTAC CTCTAATATCATCAATTCA CAGACACTTACCTTCCTTC CTCCCCCTTCCCACCCCCCC GCCTCCTCCTTCTGTTCTGC AATGTGATTTGTCCTTTTC ATATTTGTTCATTTTTTTTA ATTGTATCCTTTGGTATAT ATGGTTGTGACTACTTTCT TCCACTATTTGATCTGAGG AAGTGGGTCTGGCCCACG AAAGCTCATCATCTAATAA ACCATCTTGTTAGTCTTTA AAGTGCTACATTGTCCTGC ATTTTGCTTCAACTACCCC AGACTAACACGGCTACATT TCTATC	MDIVPKGLKVNNPLRSTYFTDYSERLCLTLSKK LRNHLISILYKKQKTIKNDISNLDRLISRQTSTPT DTLQDFCSTRQEIYYTHFYSLQQRKDNKFSNL LHTTGYNSNCLNPPDNIVNLSSYKLNPAEESV LSRGLSFCPTSPTNWIQFCGDLEAFFRRLRLRE YFQTTTEHQSDLPDPPYQHQRKNNSIWTPP DGRNYNLDFYIHSFRNNAQTDIIRKRQQATH NLSRAEHHAIQSLKNNLDIVIKPADKGGAVVI MNKADYNQEATRQLSNTTFYKPLPSDPTLDY QKKLHCLLKSLPTATREQILTESPSDPRPGLFYL LPKIHKPGHPGRPIISGIGTLTTGLSSYVDTLLK PFVTNTPSYLRDTTDFLRKLQNIDNLPNNTILA TMDVEALYTNIPHEDGLQAIRNTIPEDTTANL IADLCNFVLTHNYFRFENNLYLQISGTAMGTR MAPQYANIFMADLEQRFLNSRPLSPLLYLRYI DDIFMIWTHGQETLEIFHRDFNNLHPTINLSL DHSTREIHFLDTTVQINNGKLDTTLYRKPTDSY SYLHASSSHPEHTTRSIVYSQALRYNRICSNPT DRDQKLQDLYQAFINLNYPPRETKKQIERARR IPRNHLLQDRPKKTNNRTPLVITYNPQLKPVQ HIINKLQPILEQDTILKEALGDRPIVSYRQPPNL KMILTNHHRTYHTNTNPGTFPCNKPRCQLCP HIHSADTIIGPNQVSYKIKNTYSCASRNIIYAIM CRKCPSAMYIGQTSQTLRQRINAHKTDIRQD HKEKTVSCHFNQKGHSLNDLATCILLQRPFTS ALERESSELSFMLKFDTYQPGT	IPR000477	327	504
	REX3	GCF_002775205.1	Xiphophorus maculatus	C	TAAGCGGAAGAAGATGGA TGGATGGATGGATGGATG GATGGATGGATGGATGGA TG	MTLGRSSMIDFVIVSSDLRPYVLDTRVKRGAE LSTDHYLVVSWLRRWGTWSLSGPCSVPPLSR RPIGAVAARLSVPVAAATLEPAGGHLLVRDA VRLKKESYRAFLACGTPEAADGYRRAKRHAA RAVAEAKTRAWEEFGEAMEKDFRTASRRFW STIRRLRGGKRCSTNTVYSGDGVLLTSTRDVV GRWAEYFEDLLNPTNMPSTEEAEPGDSGLGS PISGDEVAEVVKKLLGGKAPGVDEIRPEFLKAL DVVGLCWLTRLCNIAWTSGAVPLDWQTGV	cd01650	277	532

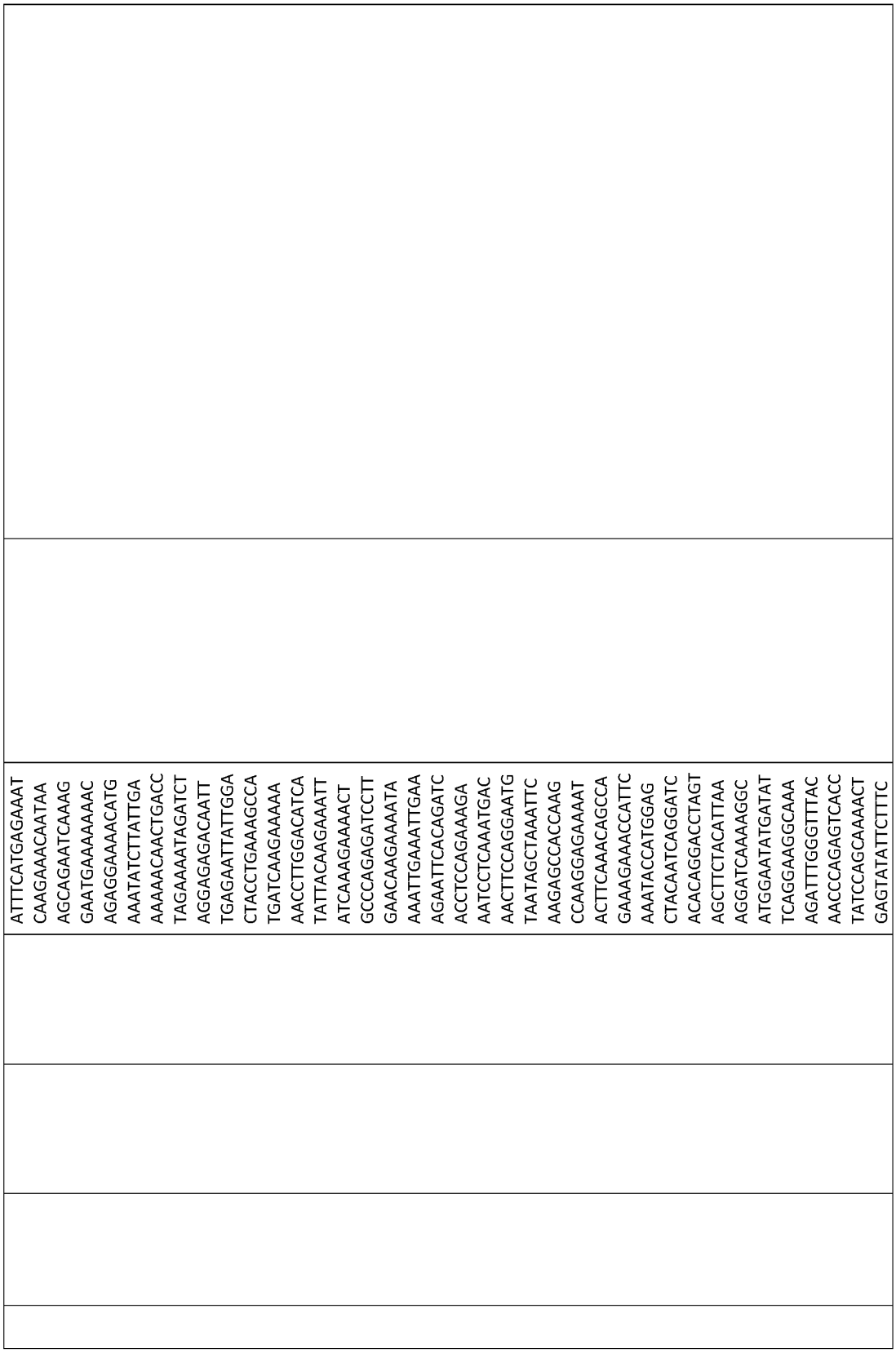
								680
							415	
						cd01650		
				VVPLFKKGDRRVCSNYRGVTLLSLPGKVYSGV LERRVRRIVEPRIQEEQCGFRPGRGTLDQLYT LGRVLEGAWEFAQPVYMCFVDLEKAFDRVP RGALWGVLREYGVPGPLIRAVRSLYDRCQSL VRIAGSKSGSFPVRVGLRQGCPLSPILFITFMD RISRRSQGVEGIRFGGLRISSLLFADDVVLLASS DRDLQLSLERFAAECEAAGMGISASKSEAMV LSRKRVECLLRVRGGVLPQVEEFKYLGILFTNG GRRFREIDRRIGAASAVKRALYRSVVVKRELS QKAKLSIYRSIYVPTLIYGHELWVMTERTRSRI QAAEMGFLRRVAGLSLRDRVRSSVIREGLRV EPLLLHIERSQLRWLGHLVRMPPGRLPGEVFR ARPTGRRPRGRPRTRWRDYVSRLAWERLGIP PEELEEVAGEREVWASLLKLLPPRPDLG	MVMVNKKWNIKNSKKITFSRAEAVQIDIENE KEQIQILAVYRAPQGNIAEFIDDIDNWIKVEN SKNLIITGDINIDITKQSNLVDKYLNCLAEHGL QSMIKDITREEILGLQLTRSCIDHICIRSKKYCK AHVIAEKLADHYFVGIAIPELGIPESERDQVTT YINNAKVDKYVREINWESIANEEDPEIIYNKIK NAFTDIYKKSSEQVTVKKRKTGNQWVTKEID QLIRKKNKAWRLLKSSPESSQRKEEYRSIRNKV TSIIKKEKRNYFWKKFEDTKDPSQCWSVINQV LGRNTRLSIDQTIKKNFGTYCNDEKIANDFNN EFIQSIELLCNHGYQNETETRARIGNSMYLPE MTEEDLVEIVSNMKTNKAVGIDSIRPRDIKEN VNIIKTVLLALYNKSLAIGKVPSGLKTAVVRPIY KKGDKKTLLNYRPVCILPVLNYIMEKYVHKAL NGFINKYNLLTKTQYGFKTNSSTITLLEDFSDF VNKKLDEGNYVLALFIDFSRAFETVRHETLIDR LFSLGVRGKLLDWFKDYLSGRTQVVKIASSFS EKRLVKQGVPQGTVLGPILFNIYINDIVNVHL NSVLFQFADDTVILYADKDYTTAENKIQEDVY KLVKWYKENQIIINTKKTELMCFRSPQRRTQ MEGNIKIHSDSCKNCNCEPLKLTTSVKYLGLYI DEHLKWNQHVEHICKRLRSVAAQLYHLRSCL PQPILKMIYHSLGKSVLVYGITIYGHCAKYLQD KINSILNKVVKNLYVDENRPLHYMYQHLDIQK				
			TCTATTGATTGCTT AGTCATATCTGAAA CGAACATAAAGGA CTGCGACGAAGCA TTATTCCAAAATAA AAGTTTTGAAACGA TTACAAATAACAGG AAGAACAAGCGAG GAGGTGGAATC	TGAACAAGTGTAAAAAAA GGGTTTGTTATAAGTGACT AGAGCTCTGTCGCTGCCCT ACGCACAAGCCGCTTAGG CTTATGTGGGGTATGAATC TCTGTAAAATTGTATGTGG AAAAGATGAAATAAAGAT TATTATTATTAT					
		Ixodes scapularis						
	GCF_002892825.2							
CR1-3_IS								

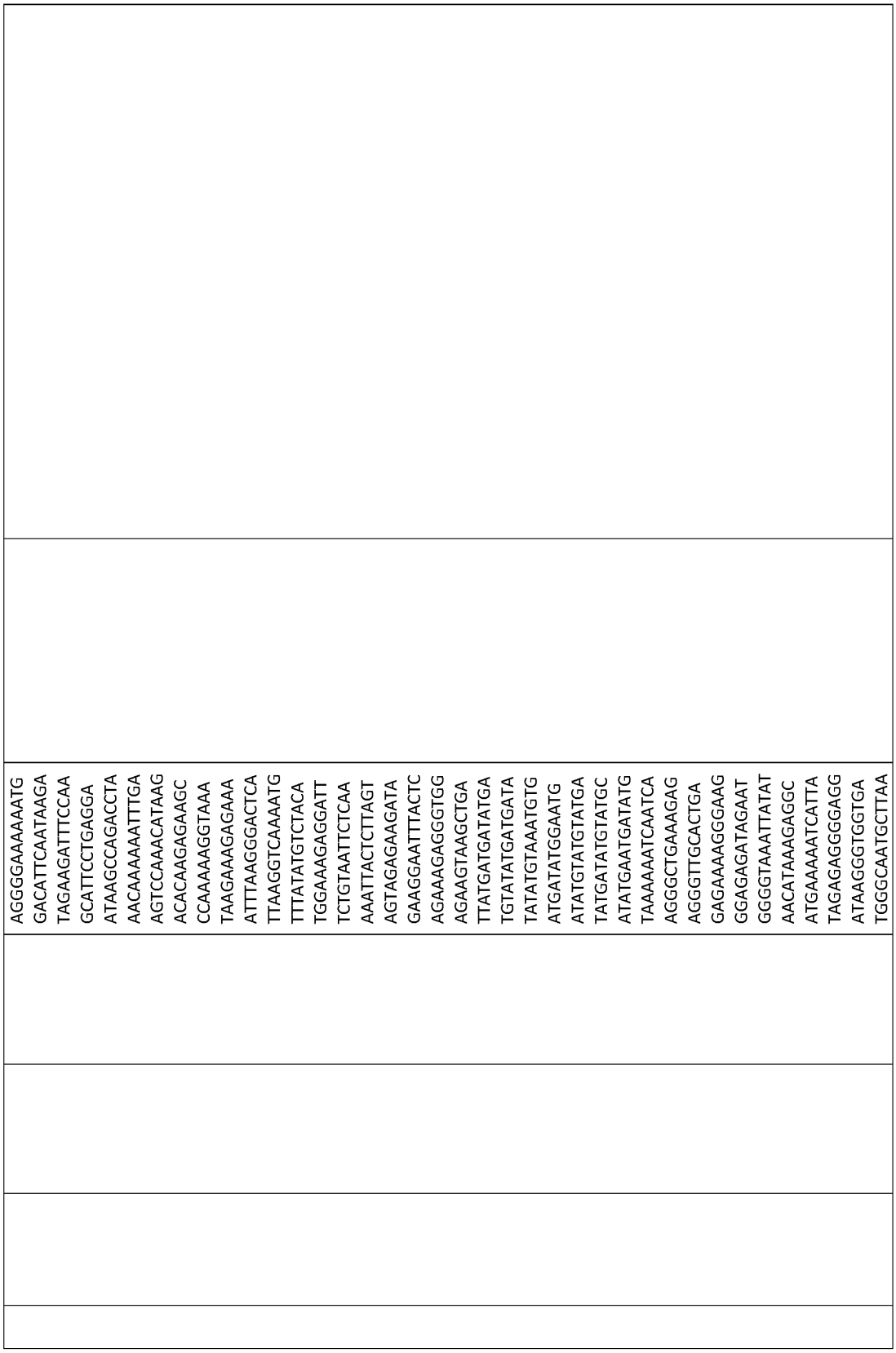
	RTE-1_PM	GCF_010993605.1	Petromyzon marinus	GCCTGTGCGTGTGT GTGTGTGTGGC AAAACAACCCCCTA CCTCCTCGCGGTGC CGGACGGGCAACT CTGTCGTGGACAGT GCTGTGGGGGGAG GTCCAATTTTGATT GGCTCCTCCCCTCA TGGCATATGTGTCC ACCACAGGGGCTA AACGGGGGCTCTG GACGGGTGCGAGG GGCTGGTAACCCCT CCACCTTCAATTAA TTACCACGAAACCA CGAATTAAACTGAA ATATGTAACTAATA AGCTTGTGCCGGA GGCCCCGGGGGGC CAGGCGTCT	CG	LSDLLELNIIIIQHQFSDKYKTKYIPIRNLRPKNLF QEPKVKTKYGKAVRASYVPRVFNKLPPRIIEIT KISELKRELKQWFINRKELQLAEEKSTRGWNL NCVVLCVLFCRELWSKGVYMSCFHFFFLFAR N	MTSPGQTPIGGARRLNLYPRGSSRNNEKPLR VATWNVRTLTAEERLPLLCRELDRYRLDLCGI QEVRWTGSGSCHHEGWTVLYSGQQTAKQQ GVALLLNKRMKGAWERGGEVWEAVSPRLL WARLPINRRRGMVVVVTYAPTEDEHPWRD AGKADESDAFYDLLSQVLAKVAPRDRVIMLG DFNARVGHDAGTWSGVIGGHGPDQLNNN GRRLLDFCAAHGLVITNTLFRHRRIHTTSWM HAGSKQEHLLDYIVTDQRMRAHVLDTRTYRG ADLPSDHRLVICKMRLPFFHSGGGCTPRTPF QPKVAWSQLSDESCKREFHHRLRLGLASSPT PPDVEGKWARFRTVTHAAAMAALGTRSRPP QRPWLTEEVFRMVEKKRQAHCHRLQNPTSE NEEAYRSLRSEVRRAVRRCKDGWWSRVAEE MESASVARDHKSLFNSIKKVTCSATPITIIRGK NGDPISDPQEQVCRFAEHFCEVLGEGKSLSQ ENIVAQSEDDPAEAAVPQILEEVTPCEWLAE VPSVEEVLKAIRSLRPGKAAGRDDLPSEMLRE GGVCAQTALYDIITTVWTTGRVPQDWKDAL VVPLFKKGDRTDCNNYRGISLLSVPGKVLARII QHRLARHAEERLAEPQCGFRKGRSCTDAIFSV RLLQQRCREFQQDLHLCFIDLVKAYDTISRPG LWVVLRAFGVPDPLLTIIKDLHDGERARVKLR GRESEPFPVHLGVRQGCPLAPTLFNIYFDHVV REAFNSCTGGISIWYRYNGRLIDARVRSRDGS LLLNHVMYADDLVIAAHSEEELETLLLRLERAT KRWGLTISPTKTKHLPGYFVPSDTPPPQLPIG DGAVVERVESFPYLGSVLMTGSSLTAEVSLRIS RAAISFHRLRNAVWKLPGLSLRTKLQVFSATV IPSLLYACETWTPLMEHLRRLETFRLACLRSIL GLTRLDCVRSSQILERSGQGPIGELLRQARLR WLGHVARMPSHRMPKQLLFGRIEGAKRARH GLEKRWSDVVQEDVELSGLADDWYQRCQD	cd01650	578	856

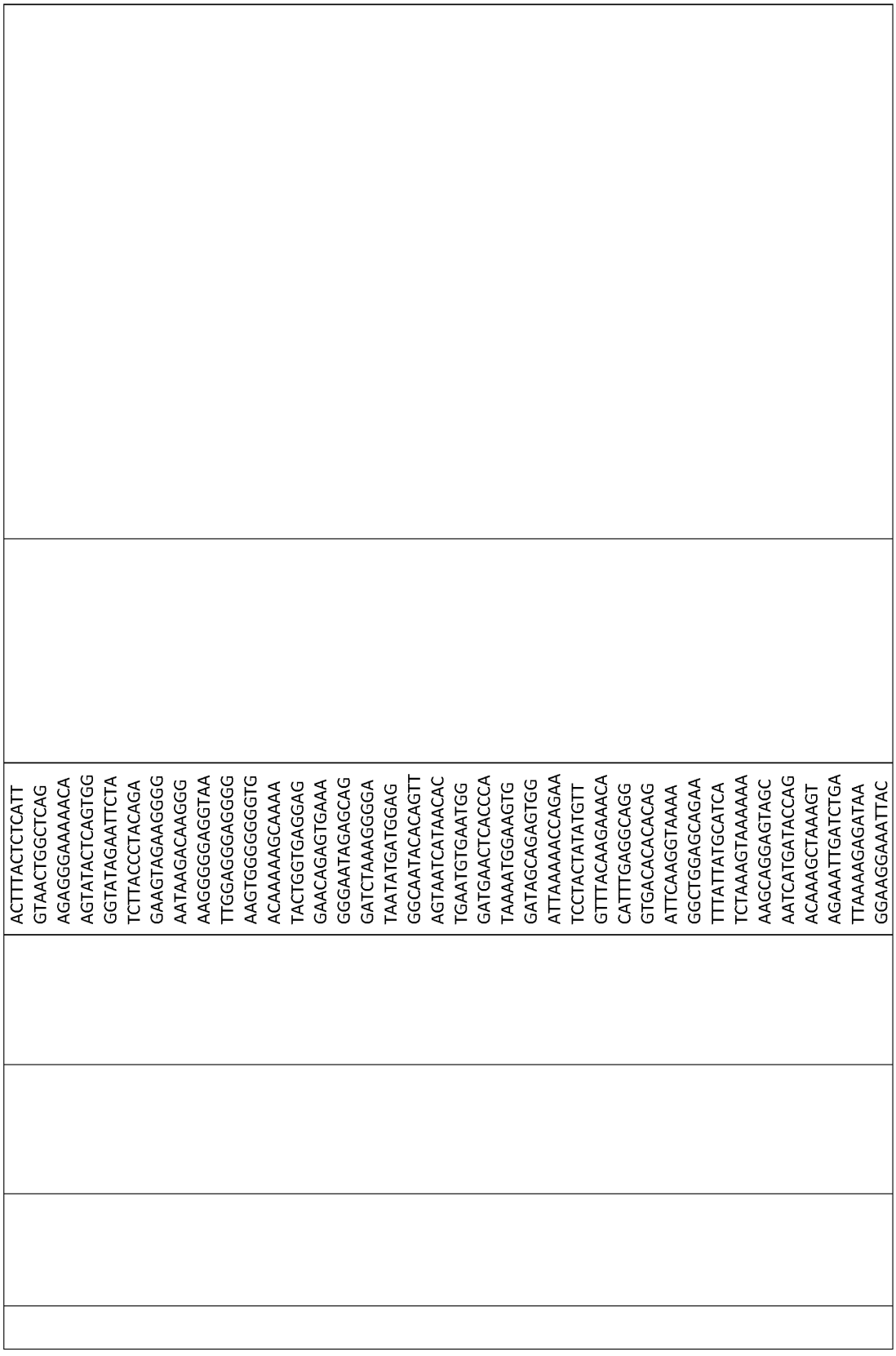
						RPLWRRLVKDATGQLEAVALQQQRRAEERR SQQRAERRVAKAQQVGTVGGTGGASASHLS QSTRGTWVCPVTSCQRAFDTS			
	Perere_Sm ed	GCA_002600 895.1	Schmidtea mediterranea		TTTACGTATGTTAAATCAT GTTTGTTTTTTTAATATTAT TAAAATAATTGAAAATTTT AAATTGCATTAATTTTTAA AAAATTAGAATATAAGTA AAAAATTTAAAATTTAAAA TTAATTTAAAGTGAATTGA ATATATTTTAAAAATATTT ATTTATTAAAAAATTAAAT TAATAAAAAAATAATTTGTT TAAATTAAATGTTAATTCA TAATATTTTTAGTTGGTTTT GAGTACCTGAAG	MIAVLSKYIVRKVFYWRYRLIKSLPNLFNYLFIN HAAAVPIFKRFMKVYNSLKRKELDISFLETCLE EKMIPSFIYKMKLSDCLSSKDVQLIHRRTLRR MLESEKIRRTSIDKKLTNIGLMLNNAATNEFCL QVCVEWCIKNSIVQIKSKENAHIKKLINLRKIF GISKFSVENTVINLSKIVLNDDEERVLKFGPNH CIKNRLDETKLCAKIESLYFHIARSMKGINGNY VKSSLSASTMEYIDKNKKQNIRDFMTLKNLAK KNINICKFDKGNGIVILDHDVYVNKLNDLLNS PQFVKLPKGRKNGKLPPLKDEEFVQKMLNGF IEKKLIDEEVGKSIRPIGSQPAKLYGLPKIHKDG VPMRPVLSMVGTAQYKIAKFLDGLLKPLMRS EFECKDSFEFVSCIAKLQKRSMNDVMVSFDV CSLFTNVPLVETIDLCCALWNENDSEHHILDR RAFRKLLEFATSNVNFLFNDEWYQQIDGVA MGSPLAPTMASIFLASLEKKIASFQFEKPFVYK RYVDDIFLIFENQKHVEPFLQFMNSLHKNIVF TCETERKSSIAFLDLLIQRKDKQYETEIYRKPTD TGLYTSPESFCEFKYKRNMVKGLIYRSWALSS TFANSVKSVDKLVELLIKNGYSKSFLLLMIKETV DKLIKNPKDPCCGKSECNLSDGCELSVDLGDC KNGYKKIEPKYVVLVLPYSEGFTNYKRKLGKLIG NLEYKIVSNSCKVRNMFINKSKTPVGLCSDLV YQFTCNGCNATYIGETSRHLCTRVLEHCRLKG LTNISEHNRGCKSDIGMSDFRILLRSFNSYWE RVICEALLIRSLDPKINVQSAVTTNVLNVFK	cd00 304	474	565
	L1-3_MD	GCA_000002 295.1	Monodelphis domestica	GACAGAGTGAAAC AACTTTCAAGAAAA GAAAAACCCAAAA GGTAGGCAGAGAA CATCTGGGGCACT GGGGTGAGAGGA GAGCAGAGCCAGC AAGAGGCATAGCA AGAAGTGAAGAGC	TAATTATTAAGAACTATTT TGCCCAATTATATGGCAAT AAATATGACAATCTAGGT GAAATGGGTGAATATTTA CAAAAATATAAATTGCCTA GATTAACAAAAGAGGAAA TAGAATACTTAAATAATCC CATCTCAGAAAAAAGAAA TTGAACAAGCCATCAAGG	MRKKLKELKEEIDSKTILVGDLNLPLSELDKSNQKINKKEVREVNEMLEKLE LIDIWRKLNRDKKEYTFFSAVHGTYTKIDHVLGHRNIANKCRKAEIINATFS DHNVIKIIINKGPWKGKFKINWKLNNLILQNWWVKKEIIETIMDFIKEND NEETTYQNLWDTAKAVLRGKFISLSAYINKIEREEINELGLQLKKLEKEQIK NPQIKTKLEILKIKGEINKIESKRTIELINKTRSWYFEKTNKIDKVLVNLIKKRK KRIKLTVSKMKRVISPLMIKRKLR			



			GGCTCTCAGAGCTC	CTTTGCAGATGATATGATG	
			TCAGCCCTCAGACC	GTATACTTAAAGAATCCTA	
			ATCAGGTAAACTG	GAAAATCAACTAAAAGGC	
			GTAAAAACCCAGA	TAGTGGAAATAATTAACA	
			AAATAAGCTGAGA	ACTTTAGCAAAGTTGCAG	
			ATAACATCAAGGA	GGTACAAGATAAACCCAC	
			AGGACTGAAGTTT	ATAAATCATCAGCATTTCT	
			AAGAGGGTACCCC	ATATATTTCCAACAAAACT	
			AACTTCCCAGAAAA	CAGCAGCAGGAGTTAGAA	
			CAGAGCCTACCTAT	AGAGAAACTCCATTTAAAA	
			AATTAGATAGGAA	TCACTCTAGACAATATAAA	
			AAATGAGCAAACA	ATATTTGAATCTATCTGCC	
			ATCAAAAAAGCACC	AAGACAAACACAGGAATT	
			TAACTCTAAAGAAT	ATATGAACACAACTACAAA	
			TTTTTATGGAGACA	ACACTTTTCACACAATTAA	
			AAGAGCAAGGTAC	AACTAGATCTAAATAATTG	
			AGACACAGAAGGG	GAAAAACATTAATTGCTCA	
			GACAGTGAAAGCA	TGGGTAGGATGAGCTAAT	
			AAGGAAACACATG	ATAATAAAAATGACAATCC	
			CAAAGTCCAAAAG	TACCCAAATTAATCTACTT	
			AAAAATATGGATTG	ATTCAGTGCCATACCTATC	
			GACACAGATTCTG	AAACTACCAAAAAAACTTTT	
			GAAGAGCTCAAAA	TTATAGAATTAGAAAAAA	
			AAGACTTCAAAAAC	ATATAACAAAGTTCATCTG	
			CAATTAAGAGAGG	GAAGAACAAAAGATCAAG	
			CAGAGGAAAAGTG	AATATCAAG	
			GGGAAGAGAAATG		
			AAAGTGATGCAAG		
			AAGAAAAGAAAGT		
			GATGCAAGAAGAA		
			ATGAAAAGTGATG		
			CAAGAAGAAATGG		
			GCCAATTGAAAAA		
			GGAGAACCAAAAA		
			CTGACAGAGGAAA		
			ATCAGGTCTTAAAA		
			ATCAGAATTGACCA		
			TCTAGAAACTAATG		


ATTTCATGAGAAAT
CAAGAAACAATAA
AGCAGAATCAAAG
GAATGAAAAAAAC
AGAGGAAAACATG
AAATATCTTATTGA
AAAAACAACTGACC
TAGAAAATAGATCT
AGGAGAGACAATT
TGAGAATTATTGGA
CTACCTGAAAGCCA
TGATCAAGAAAAA
AACCTTGGACATCA
TATTACAAGAAATT
ATCAAAGAAAACT
GCCCAGAGATCCTT
GAACAAGAAAATA
AAATTGAAATTGAA
AGAATTCACAGATC
ACCTCCAGAAAGA
AATCCTCAAATGAC
AACTTCCAGGAATG
TAATAGCTAAATTC
AAGAGCCACCAAG
CCAAGGAGAAAAT
ACTTCAAACAGCCA
GAAAGAAACCATTC
AAATACCATGGAG
CTACAATCAGGATC
ACACAGGACCTAGT
AGCTTCTACATTAA
AGGATCAAAAGGC
ATGGAATATGATAT
TCAGGAAGGCAAA
AGATTTGGGTTTAC
AACCCAGAGTCACC
TATCCAGCAAAACT
GAGTATATTCTTTC


AGGGGAAAAAAATG
GACATTCAATAAGA
TAGAAGATTTCCAA
GCATTCCTGAGGA
ATAAGCCAGACCTA
AACAAAAAATTTGA
AGTCCAAACATAAG
ACACAAGAGAAGC
CCAAAAAGGTAAA
TAAGAAAGAGAAA
ATTTAAGGGACTCA
TTAAGGTCAAAATG
TTTATATGTCTACA
TGGAAAGAGGATT
TCTGTAATTCTCAA
AAATTACTCTTAGT
AGTAGAGAAGATA
GAAGGAATTTACTC
AGAAAGAGGGTGG
AGAAGTAAGCTGA
TTATGATGATATGA
TGTATATGATGATA
TATATGTAAATGTG
ATGATATGGAATG
ATATGTATGTATGA
TATGATATGTATGC
ATATGAATGATATG
TAAAAATCAATCA
AGGGCTGAAAGAG
AGGGTTGCACTGA
GAGAAAAGGGAAG
GGAGAGATAGAAT
GGGGTAAATTATAT
AACATAAAGAGGC
ATGAAAAATCATTA
TAGAGAGGGGAGG
ATAAGGGTGGTGA
TGGGCAATGCTTAA


ACTTTACTCTCATT
GTAACTGGCTCAG
AGAGGGAAAAACA
AGTATACTCAGTGG
GGTATAGAATTCTA
TCTTACCCTACAGA
GAAGTAGAAGGGG
AATAAGACAAGGG
AAGGGGGAGGTAA
TTGGAGGGAGGGG
AAGTGGGGGGGTG
ACAAAAAGCAAAA
TACTGGTGAGGAG
GAACAGAGTGAAA
GGGAATAGAGCAG
GATCTAAAGGGGA
TAATATGATGGAG
GGCAATACACAGTT
AGTAATCATAACAC
TGAATGTGAATGG
GATGAACTCACCCA
TAAAATGGAAGTG
GATAGCAGAGTGG
ATTAAAAACCAGAA
TCCTACTATATGTT
GTTTACAAGAAACA
CATTTGAGGCAGG
GTGACACACACAG
ATTCAAGGTAAAA
GGCTGGAGCAGAA
TTTATTATGCATCA
TCTAAAGTAAAAAA
AAGCAGGAGTAGC
AATCATGATACCAG
ACAAAGCTAAAGT
AGAAATTGATCTGA
TTAAAAGAGATAA
GGAAGGAAATTAC

			ATTTTGCTAAAAGG TACTATAAACAATG AAGTAATATCATTA CTAAACATTTATGC ACCAAATGGTATAG CATCTAGATTTCTA A	TGATTGATTGTTCACGCTT ACAGTGGCTTTCAGACAG CATACATAAAAAATTTCGT CTCCCCCCTCCTCTTTTCAA AAAATAGATGGACTAGTC CTGCTGGGCAAACTGCCA AATCGTTGCGCGTGACGC AAACGTTTTTTTTAAATAA ATAAATATGCCGTTTTCAA ATAAATAAATTTATGCAAA CCGACACACCTGATGATG ACCGTTTGAGTTATGGTCG AAACGTCGTGAGAATTTTT ATTGTTTTATATATATTTTG TTATTTTGTTGTTTCTCTTC AACGTGACTTTACCGAAAA AACTAAGAATATGAATTAC TGCAGTCACGAAAGCTTTA AATCAAAA	MNRLYGEGFHITLRSFNKLRDKKAQLIESLTFL KRCRDSGVIPIAFTLTSATVTSSNAKRILFKTSL ALISEKIRLIRTELDQLDKTLYKSHLSLTAIMSTS DWEHVDNMTASSAATMSTKCKENLISKFNLL FNKRPDPHTKKLDVNKVVINRSNSSISDGLKS VLAKGLNFAVTPKHIPVNDIIAQVDVALRGLG HIAADVARSEIAATLKNAKMPASNLSKDERFE LYKLKKRNDIVVLSADKGNATVIMDKIEYNNK IKELLDDGSYTKIKRDPTNSLVTQTNKIISDSTA LIDQSKQFSLKPSAARPPRLYGLPKIHKVDTPL RPIVSTISSPTYKLARFLSEILRPFTGKTTSAVQ NSTHLVSTIKGISLSPDDLLVSFDVVSLFTKVPV PDTIAIIEDLVHQGLNTEVPPLVRHCLTNTYFT WNGTFYKQTEGTPMGSPLSPVIANIFMEKFE TDAIQTATLRPTLWKRYVDDTFVIWPHGRPA LDQFFDHINSIHPAIRFTMEVEIDGKIPFLDVLI KKRPDGSMSHSVYRKATHTDAYLKPDSHHH PSQKNSLIKTLYHRAQCISDQEHISSELKHVHQ ALRANGYSDYTIRKAITQTMKKNPLEEPFKTIT LPYIAGITKKISKILNKHKIQVRFNTLHKIRDLVP SVKDVIPDIQYPGVYKLSCLCGSSYIGETKRHIS DRIREHKADLKHGRTNTSAVADHCSNSVGSH AIDFSKTEVLCKPSGFHERFVHEAIHIYKHRDN FNREDGFYLSNHWKDLINHFSS	cd00304	440	524
	Poseidon-1_PM	GCF_010993605.1	Petromyzon marinus	AGCTTAGCTGATTG GAGAATATAGAGT TTATTGATAACATG AAGGTAATTTATAT TTTAGTTTTTATTG GTTCACTATATAGT ACTGTGAATAGCTC	TAATTAATTGTATTTTACG ATGACCTTCAGACCGCAGC TCTTTCGCCAGCATTGCCT CTTAATACCTCTCACACCA CGGTTAATGGTATGGTTTA CAACATGGAGTATACTAG GGCTTCGCAGCTGTCTCTC	MNHCSHESFKSKFNRLYGEGFHTSLRSFNRLR DKKAQLIESLTFLKRCRDSKVIPTAFTLTSAVAT SLNAKRILARTSFALISEKIRLTRTELDQLDKTLY KSHLKLTATMSTSDWEHVDNMTTSSAATMS NKCRDNLISKFNHLSSKKSDPHTEKIDVNKVVI NRSNSSISDGLKSVLAKGLNFAVTPRHLPVND IIAHVDVALKGLSHISADVARSEIAATLKNAKL	cd00304	453	537

		AGATGTGTAGGCC	TCCCCCTCCTCTGTTCCGA	PPSNISKDERFELSQLKKRDDIVVLSADKGNA	
		TTGACTTCAACATA	CCATGGACGAACTCGTCC	TVIMDKTEYNNKIQELLNDGSYTKIKRDPTNS	
		GTCATTTCTGTGAC	GTCTGAACGGGGCCCGTA	LVTKTNKIISDSTALIDENKQFSLKPSAARPPRL	
		CAGTCACTAGCAAA	GCTATCCGTGGTACGAAAT	YGLPKIHKEDTPLRPIVSTISSPTYKLARFLADIL	
		TCACCCAGTTATTT	TTGAGGGTAAGAATCCAC	RPFTGKTTSAVLNSTHLVSTIKDISLSPNDLLVS	
		CAATGCTAAATTCG	CCATTTTTTTCGCCATATA	FDVVSLFTKVPVPDTIAIIEDLVHQGLHTDVPP	
		CCTAGTTATTTCAA	GCATTGTGATATATAACGA	LVRHCLTNTYFTWNGTFYKQTEGTPMGSPLS	
		TGCTAAATTCTCCT	CTTATGGTTGGTAGCTTAG	PVVANVFMEKFETDAIQTATLRPTLWKRYVD	
		GTTTGATTTGTACT	CTGATTGGAGAATATAGA	DTFVIWPHGQPALDQFRDHINSIHPAIRFTM	
		AAGTGCCTCGGTG	GTTTATTGATAACATGAAG	EVEKDGKLPFLDVLIKRRHDGSMSHSVYRKAT	
		AATGTGTATGCTCT	GTAATTTATATTTTAGTTTT	HTDAYLKPDSHHHPSQKNCLIKTLYHRAQCIS	
		GACTTCAACGTAAC	TTATTGGTTCACTATATAG	DQEHITSELKHVHQALRANGYSDYTIRKAITP	
		TATTTCTGTGAGCA	TACTGTGAATAGCTCAGAT	VMKKDPLVEPLKTITLPYIAGVTKKISKILNKHK	
		GTTACAAGCTAATC	GTGTAGGCCTTGACTTCAA	IQVRFNTVHKIRDLVPSVKDALPDIQYPGVYK	
		ACTCAGCTATTTCA	CATAGTCATTTCTGTGACC	LKCLCGSSYIGETKRHISDRIREHKADLKHGKT	
		ATGCTAAATTCTAT	AGTCACTAGCAAATCACCC	NSSAVADHCYNSVGSHDIDFSKTEVLCKPTGY	
		GCTAATTCCATTGT	AGTTATTTCAATGCTAAAT	HERFVHEAIHIYKHRDNFNREDGFRLSNHWK	
		CAAGGGGTGCCTC	TCGCCTAGTTATTTCAATG	DVLNHFCS	
		TGCGGATGTGTAT	CTAAATTCTCCTGTTTGATT		
		GCCTTAACTTAGCT	TGTACTAAGTGCCTCGGTG		
		ATTTCAATGCTAAA	AATGTGTATGCTCTGACTT		
		TTCTATGCTAATTC	CAACGTAACTATTTCTGTG		
		CTTTAACAATGGGT	AGCAGTTACAAGCTAATCA		
		GCCTCTGCGGATGT	CTCAGCTATTTCAATGCTA		
		GTATGCTTTAACTT	AATTCTATGCTAATTCCAT		
		TTAACATAGCTATT	TGTCAAGGGGTGCCTCTG		
		TCTGTGACCAGTCA	CGGATGTGTATGCCTTAAC		
		CAAGCTAATCATTC	TTAGCTATTTCAATGCTAA		
		AGCCATTTCAATGC	ATTCTATGCTAATTCCTTTA		
		TAAATTTTATGCTA	ACAATGGGTGCCTCTGCG		
		ATTCCTTTGTCAAG	GATGTGTATGCTTTAACTT		
		GCTCTTTGTTCCTCT	TTAACATAGCTATTTCTGT		
		CACACAGCTTAAAA	GACCAGTCACAAGCTAATC		
		GCCGTGATCTGAAC	ATTCAGCCATTTCAATGCT		
		GTGGCTCTCTCACT	AAATTTTATGCTAATTCCTT		
		CCAGCGCTGTCCTA	TGTCAAGGCTCTTTGTTTC		
		CCAGACAGTCTCGC	CTCTCACACAGCTTAAAAG		
		TCCCTTCACCTGAT	CCGTGATCTGAACGTGGC		

	Poseidon-2B_PM	GCF_010993605.1	Petromyzon marinus	GAAGGCCGCTTGT GTTGTGGCCGAAA CGTCGTGAGAATAT TATTGTGTTATTTTA TTTTATCAATTTATT ATTTTATTCTCATTC TACGTGACTTTACC GAAAGAACCAAAG AAG	TCTCTCACTCCAGCGCTGT CCTACCAGACAGTCTCGCT CCCTTCACCTGATGAAGGC CGCTTGTGTTGTGGCCGA AACGTCGTGAGAATATTAT TGTGTTATTTTATTTTATCA ATTTATTATTTTATTCTCAT TCTACGTGACTTTACCGAA AGAACCAAAGAAGATGAA TCATTGCAGTCACGAAAGC TTTAAATCGAAA	MELDTLDRTLLQAHLMLTQKMSPDWNHVD NMTASSAESLSVKCKANLISKFNHLSKKNPVP HITKLDVNKIVINQSSTPISDNLKTVLARGLNF AVTPKHIPVDDIISQVDVALKGVGHIQADVAR AEIATILKNAKKPISNISKNERSELANLKNRSDI VILSADKGNATVIMDKDSYNNKIQNLLNDGS YSKIKNDPTNSLAIKTNKIITSSTSLIDPTKRPSL KPSAARPPRFYGLPKIHKEGTPLRPIVSTINSPT YRLARFLSDTLLPFTGKTASAVLNSAQLVSMV KDVVLSPDDILVSFDVVSLFTKVPVPDTIEIIEN LVQHGLDPDTPALVRHCLNNTYFTWNGSFY KQNEGTPMGSPLSPVIANIFMEKFESDAIQTS TLRPSLWKRYVDDTFVIWPHGLPALNLFLHHI NSRHPAIKFTMEVEKDGKLPFLDVLFERRPDG SLGHSVVRKPTHTDSYLKPDSHHHPAQKNSLI KTLYHRAHCISDQHHLTSELKHVHLALKRNGY SDRTIRQAVSPSIQKQPLDTPVKTITLPYIAGV TKKISKILNKHKIQVRFNTVHKIRDLIPSVKDVI PDLQYPGVVYKLSCTCGDCYIGETKRHVSDRIR EHKSDLKHGKTNSSAVADHCYNAVGDHDIIF SDTKVLCKPTGFYQRFVHEAIQIYKHRHNFNR EDGFKLSNHWKDVLNRTKH	cd00304	361	445

				TTCATCATGCTAAATTATA TGCTAATTCCTTTGTCAAG GCGCTTTGTTTCTTCTCAC ACAGCTTAAAGGTCGTGA TCTGAGCGTGGCTTTTTTC ACTCCAGCGCTGTCCTCCA GACAGCTCGGTCCCTTCAC CTGATGAAGGCCGCTTGT GTTGTGGCCGAAACGTCG TGATTTTTGTTATTTTATTT TTTACCTACGTGACTTTAC CCAATAGAACCAAAGAGT ATGGATCCTTGCAGTCACG AAAGCTTTAAATCTAGAA	MGSPLSPVVANVFMEKFETDAIQASTLRPSL WKRYVDDTFVIWPHGLPALNQFLNHINSIHP AIKFTMEVEKDGKLPFLDVLIEKRPDGSMGHS VYRKATHTDAYLKPDSHHHPSQKNSLIKTLHH RARCISDHHHLASELEHVHQALKMNGYSDYT IRKAITPTVGKEPSEDPIKTITLPYIAGVTKNISKI MSKHKIQVRFNTVHKIRDLIPSVKDVIPDMQY PGVYKLSCVCGSSYIGETKRHVSDRVREHKAD LKHGRTNTSAVADHCFNAVGSHEIDFSKTKV LCKPTGYQQRFVHEAIQIYKHRHNFNREDGY KLSNHWKDVINQFKS	cd00304	1	83
	Poseidon-1B_PM	GCF_010993605.1	Petromyzon marinus	CA	TAACTGGCTGTGATGTTTC CATCCTACCTACCGATGTT AGCAGTTATTTTCGACTCT CTGAGTTACATTCACTAAC ACAGCCAACCGTTAGCTGT GTTAGTAAATGGCGGCTT GCCACTGCCTCGTCCATGC CCTGGCCGAGCTTGGCCC CCTCCGATCTTACATTTAC ATGCAAATTTCTCTCCTAC TAGCCCCGCCCACTCGCTC CTACATGACCCTTGGCTGC TACACCCATCTCTAAACCT TTATAAACAGCCTGCCTCA GCAGCTTTCCCTCTTTTTTC ACCTGATGACGATTGCTTG TGTTGCGATCGAAACGTC GTGATTTTTAATTTTTATTA TTTTTTCTATCTACGTGACT TTACCGAAAGACCCAAAG AATATGAAT	MPIKMVRPRRLMDPDGFLRSLGDLPASSRG DPVDALVTRYNSEMARAIDTIAPERPLSLRRA KPAPWFTEELAVMKRERRGLECVWRKLRSES	cd01650	257	517
	L2-24_ACar	GCF_000090745.1	Anolis carolinensis		GATACGGTCGATATAATCT GATGTATTCGCTGGATTAA TGTGTAATTGAATACAGAC			

	Jockey-8_DPer	GCF_003286085.1	Drosophila persimilis	GGCTAATGCAGCA CACAAATTTACTTG CCTGTGAGATCGCA ACGCACCCATACAT TCCAAGAAGGCTG GCCAGGCCAAGGT ATGCCGCAATGATA GAAGAGCACAATC CAAACAGCAACAG AAGCCTGACAGCC ATCCAAACAATACT GCTCCAAGTCAGAC GCGAGAACTGGCT GCGCAATCCATTTG CAACACACAATATG GAAATCAGCCGCA GTAGCGCTAATGC AGCTGCTACGCAG	C	TGGCTCTGACTTTTGGCCC AATGCTACTGGCCCTGTTG TAATGGCCACACAGTGCTT TACTGTTTTAAATGGGGTA TAATATTGGTTTTTAAGTG TGTTTTTACGATCGTTTGTT TTACTATATTTTGTATTATA TGTGGCATTAATCTTGCCA TTATGTAAGACCGCTCTGG GTCCCCCTGGGGGAGAAG AGCGGTATATAAATTAAAT AAATAAATAAATAAATAA ATAAATAA	DRARAKALRKAYSVALRAAKKSFQTARIAAAK NRPSELFRVVGELLNGTPQTGPPDDLAENCA AFSHHFADKVAQIRSELDATVDAVSEEVARA PVCPILMDSFQLVQPDDVDKILGEVRATTCAL DPCPSWLVKEARGGLVEWVRVVVINASLREGI FPSCLKQAVVKPLLKKPSLDPTNWNNYRPISN LPFLGKVLERVVASQLQGFLDDTDFLDPSQSG FRPGHSTETALVALVDDLRRELDRGSVTLLVLL DISAAFDTIDHGILLGRLLGMGLGGTVLQWL QSFLEGRSQMVKLGDTCSDPWPLTCGVPQG SILSPMLFNIYMKPLGEVIRSFGVRCHLYADDT QLYYSFPPASKEAPQILNQCLAAVSDWMRA NKLKLNPDKTEVLLVSRKSDRGIGWQPVLDG VALPLKAQVRSLGVLLDSSLTLEAQVSAVAGR AFAQLKLVRQLRPYLVKSDLATVVHALVTSRL DYCNALYVGLPLKTARKLQLVQRSAARLLTGA NYRERTTPLFKELHWLPVIFRAQFKVQVITYK ALNGLGPTYLSDRISPYEPARRLRSSGEALLSV PPPSQVRLVGTRERAFSAVAPRLWNSLPREIR QAPTLLSFRKQLKTWMWDQAFGPS	cd01650	470	731

					KVKTIIPDNLYILLTSYLSDRSFHVVCCNNITSSL KHITAGVPQGSVLGPVLYNLYTADMPQPPRH ANSTQMLATYADDAVYLVSALSPKQANENV ERLLADALSWSKKWGITISLAKTAHVLHTLSP VQQHTPAPRLAGQLLAKKSVMPYLGIHMDT KLNFNKHVAAIVTKTRARIKKLDWIIGRNSKLP LPPKVLLYKQLIAPIWQYALPVWGPLTSSMQ MAKLQTIQSKALRAMTNAPRYVSNNALHAD LQVATVEEVYNRLCRNCYT		
				GCAAATGACGGCC CCAGCAGAGAACC TGATGATGGACGG CGGCAGCTCCAGA ATCCAGATGATGG CGGCGGCTGAAAA GAAACTAAAAACCT TTTTATTCTTCCACA GATACGAAGCCACT TCACCTGCACAGAA ATTGGTAATCCACA TCCAGAATGACGAT GCACTGGACCTCG GCCAGCTTCAGCCC TGTTGATGCTGTCC TGCAGTCACAGTTT AGCGCCGACACTC GCAAGGATTCCACT TCGCAGCCGACGA CGAGACCAATAGG CGTAATGATGTTCT TCACGGCTGCATAC TCCACCTGAAATAA TTAAAAACAAAAAC AATGTAAACAAAGT AGCAATAAGTTTAG TGTTACCCTGGCAT TTAGCCAAATTATC GATAGCTACTAAAA AAAAATAAATACCA AATAAGCCAATACC GTAAGAAATGTTAC CCTGCCAATTTACA GCCACTTCTGCAAC TACATGTAATGTAG ATAAGCAGTTGGC GCGAAACAATTTGT AAAATATGCTAATT			

			ATAAGAGTATAAAC AAATAAAACGGTTC TCTCTTC				
	R4-1_PH	GCA_001587735.2	Parhyale hawaiensis	CCGACCGCCAGCG GGATAACTGGCAA ACCCTGTCTCGACC ACCGGCCCGTGAA TCCATCGGGGCGT ATGAGTCAGACAC AGGGGGGTGTTTA AGGTGACCCTGTTG CGAGGAAATGCGC AGCAAAAGCCGGA TGAGCCTTAGAACA TCGAAGGCCAACG ACAACCTGCCGAAA GACTGGCGGCTAA TCCAGTCAACCTCC TGTAGGTCACCGG CTGGTCATGCTGAA TCTCAGCTTAACCA GGCGTACGAACTT ACAGTTGGAGGGT CGAGCACCCCTGAT GGCTGAAAAGGAC CATCAAAGTCGAA GGTAGCCAACGAG AAGACCAACGGCT GATTCAGGCGGAA GAGTCAACTCGTTG AATGCGTTCGACA GCTTGGGGTAGAT GGAACTCCTAAGCC CTGAAAGGCAGTC CATCTTCGCAGACG CTAAGGTGCCCCGC CGCCTGAGGGTTAT CAGGCCCCGCCGC	TGAGCCTTAGGTCGCGGG ATGTGACCCGGCGCCAGA GTGTAGAGCTGAACATCG CTCAACCGATCCAATTTGG GTCGTGAAATCCCCTCGAT AATAATAATAATAA	MKMSHNRDTPSNGVKGTSVRLGTSLVRSPV GEAGAVRERGTHPSESVSQDSDASVNATGE GSVREQAPLSPPGAEEATVPTQRRTRHKWSK EDRVVLWECFVASKREGPGYLKRLKQLWDER GIPGNFPQASLSGQIRQICSKNLLSEEERLQIA ARMEAQVASPSADEPARQVPTRPVTPPRSPP VEPARRPSIPSEETPDLGAVPSEIDSADPNRSP SRGPRHLPAHNMSQSESEDDVTDPDVSDQQ RSDSLEPRDLLRNSSVESTPGHPNQELSDTLLS NYVPSEIDSDDPNQSPRRGPRHLPAHDMSLS DSMDEETEPDLSDQQRSDLLELRDLLRNSSVE TTPKGHPSLRHLPEPKIRAAAFRVNSVIGKIHT NNITETNALIKAGADLAVRILEVQPRPQRTQR KKDPPWKHRLEKNIAEIRKHLSWISEWRRGN LHDEEKKTLLESRYRCLEVGLTNLEDTLKQRLS AKRSKVRRFEARVAGFHQNQLFNTNQKRLY QTLRGEETSSDSPNAEESIRFWSDIWSKEVRH NNAAEWLHDVKEKNVAADPDLTITSQQLKK QLSKTKNWKAPGPDMVQGYWIKTFTSLHSRI AAQLNHCLQRGTVPTWMTTGKTVLIQKDKA KGTEVSNYRPITCLPLMWKVLTGIIYERVYQH LDSKKLLPDEQKGCRRNTRGTKDHLLVDKLLT KDARSKKKNLSMAWVDYKKAFDMVPHSWI LECLDIYGIAGNIRNLIATTMPNWKTQLTSAN KHLGEVSIRRGIFQGDSLSPLLFVLTMIPLSETL NKAGQGYNYSRTMKLNHLLYMDDLKLYAKS KDQVEQLLNIVHQYSQDIKMQFGISKCGVLNI ERGEVTASEGIIIEEGTIKDIEETGYKYLGVMEY NTILHRTMKDSIRKEYLTRLRLILKSHLNGGNTI KAINTWAVPVVRYSAGIINWTKKDCTDMDIK TRKLMTIYRALHPRSCVDRLYINRREGGRGLIS VEDCVEAEKRALSQHFRESDDPWARCLIEAK LLKETETADQFKERRRLDRTNKWKSMKMSG QYLEAVQDKIVPDSWNWLLRGELKRETEGTI LAAQEQALRTRYIQNKIDKRNVPSTCRICRSS	cd01650	616	879

			CTGAGGGTTACCA GGCAACC		DETINHVISECGVLAQKEYKRRHDKVARHLH WTLLRIHNFPVSERWYEHEPAPVVENEAVQI YWDKRMETDRVLHANRPDIVVKDKQEKSAK LIDVSIPFDSRIIDKEAEKKEKYRDLAIELQRLW QMKVDVVPVVIGALGAMSKNLKTALRELKC GHLHPGTLQKSALLGTAHIIRKVL	cd01650	85	341
	Proto2-4_CTe	GCA_000328365.1	Capitella teleta	GATGAATTGTATAA TTGTGTTGATTATG ATCGTGATTCA	TGATTAAGCTACCTTTATG TTGCAGTTTTGATTTGTTT GTTCGTTTGTTTGATTTTCT TTCACCTACTATAATCCTG TATATTTATTTATACGAAT AAATTATTATTATTATTA	MNELIAFFKISVQNVCDQGSCYSSHQFTAQG VRSAVSKLKDKSDCKDNFMSNHVIHGGDTLF VHLSFLFNCMISHTCDPKQLRMSVLIPIPKNR RKSLSVSTNYRSIALSSIVGKVFDRMILTEHVNI FSTSNLQYGFKPKHSTVQCTFVLQNVINHYKS RGSPCYVMLLDASKAFDRVQYVRLFYLLRERG LCARIFMLLALLYTNQSMCVRWKNCISDNFS CSNGIKQGGVLSPTLFCIYFDELLIRLKHCGSG CHLGHLFTGALSYADDVALLAPTLTAAKVML NVCEQFANNFNVLFNADKSALLVCGRPIRKE VNLRLNGRIIQQNSTAVHLGTTIGDNSSTRNV KKAASDLYCLTNVLMCKFGHCSIDVLSALFFFF FFYCSSFYGCPLWDLRNIEPILTAWRKCLKRI WKLNIRTRTKYINNLPKGNLISALLIRSLSFYM QCLSSSNAIVKLASNYCFYGCMYSVISENVHL CSVLSSVPLRFIFDDFENVKSIIQRGMYDEATI AESLALKDLCYLRDGLMFLNFEITSEELLCAFN HFTI	cd01650	85	341
	R4-1_AC	GCF_000090745.1	Anolis carolinensis	CCCCCCCGCAGGTG CCACCTTGACGTGG TGGGGGGGCTTAG CTGCTCCAATGATG ACTGAGGGCTGTG CTGGCGGTAGTGT AACTACCGGCAGG TCCAACCAAGCCAG AGAGGCCTCAGCT GAGGAGCACAACC AAGAGCACCTAACC CATTAGCATT	TAGACACTTGGGAAGTGT TCGACTTGTGGTTTTGTGA AACGAAATCCAGCATATCT ATCTTGTTTGCTGTGTCAT ACAACGTCGTTGTGTTGAT	MENQTQTKSILARSSPRIKRTAVDAADSGHP STSGLRNHQNPNEQSQKRQKYTMPENRTIM KCYYKSEPQRRGYQKRMHQLWKQEYPDSQI TEPRLADQRRFIIRNKVFSEVELEEIQKICKVNY HQTTAQTAAETPATLGMVEQIEPEESVELLQ EFEEPALETSVEPPGTLTARQQELKDKIMAHA AANAIRKRLPTLKTVPKRHLAPLMKDVNAALS TVQITSIEQTNQYAYSAAVIVTEELGLLQPRQP QRKSTGKPKWKVRLELKIKKLRSDASNLKNM KEKKLKNDKIKQYLIRKYWLNTRKIEEALEIVKE QITATARKIERYEARIIQYRQNQLFQSDQRRFY QSLNQTTDTVTIKPEKTATTKFWKELWENNK NYNKNAGWIKEFEGKFSQNKMELMEITTEM ISKRVQKVKNWTSPGSDQLHGFWLKHLISLH	cd01650	465	731

	CR1-66_AAe	GCF_002204515.2	Aedes aegypti	ATCGCTGTTCGTACCTGA	TAGATTTTTATGTTTTTTTTTTCATTAAGACAAACAGATGTCAGATGAATAACAACTTCAATAAATAAATAAATAAATAAAATAAATAAATAAAT	GKMAQQFNEMLQKGSISEWLTTGRTYLIQKDPAKGAAPGNYRPITCLPTMFKLLTGIIADRIQDYLEEKNILPDEQKGNKRKSRGTKDQLLIDKMILENCKSRKANLHMTWIDYKKAFDSLPHSWIIKCLDAIGICKNVGTFIENMMEHWKTELFVGNESYGLVNIRRGIFQGDSLSPLLFIIAMIPLSTILQKTNLGYQTSKNSHKISHLMYMDDLKLYGKTETEIQSLTNTVRIFSTDINMEFGLDKCSTVALKKTEIQSSEGIINMPNGQTIKCHQPEAYKYLGILQLDNIKHEHVKTVVSKEYTQRVRKILKSKLNGGNTIKAINTWAIPVIRYTAGIINWTQVELDNLDRKTRKLMTIHHSLHPRSDVDRLYLPRRSGGRGLLQVKQAVKEEEHALAEYVKNQKLLKTQQTKNQYKKTALQTRADSWHNKTLHGKFLDKIEGKADKEKTWLWLTNGTLKKETEGLILAAQEQAIRTNAIKAKIEKSADDPKCRLCKETDETIDHILSCCKKIAQTDYKQRHNYVAQMIHWNLCLKYHLPAAKNWWDHKPAKVLENEHAKILWDFRIQTDKVLEHNTPDITVVEKNKVWIIDVAIPGDSRIDEKQQEKLSRYQDLKIELQRLWDVAIPVVMGTLGAVPKDLSRHLETIDIDQKPVQVVPVVMGTLGAVPKDLSRHLKSKLNGGNKITICQLQKATLLGSARIIRKYITQSKITICQLQKATLLGSARIIRKYITQS	MYSKRFQPSNETLAISHDVQPGCTPVSLVEASYPLVTVEPILPATSSHPGPVYEYGEEVFQTLFAGKYSCDSSNFSPVTSVASSDIPTRKSDTKSAAPDDFASDNPANRLPNVITDDLNNNQSCHALRFYYQNVRGLRTKVDSFFLAVSECDYDVIVLTETWLDDRFQNAQLFGSKIIVYRTDRNAANSSKRRGGVLIAVSTHLNSFIDPTLTSIALEQLWIKISTPSRTISVGVIYLPPDRRDNMDYIRDHLNSMNSVLSNLALNDCVLQFGDYNQPNLLWEASDDGFLRVNSQSILSTAGSTLVDGFHFHGLTQLNGIPNRNSRFLDLVLGNDEARYSCNVTEAIEPLSDLDVDHPALDVILNLPDSITFEDCSDNSQMYDFRQADYAGIHAAISRIDWQPIDDANNIDDAVTYFNHVVQDIIAQNVPLCRPKPKPPWSNSHLRQLKRNRAKSQRHYCRNRSQFTKQEFNRASNAYRTYNRLLYRRYVVVRTQTSLRRNPKRFWS	cd01650	638	885

						FVNSKRKEVGLPRSVHLDGQTATSEQEKCFLFAQHFQKTFHDSPATADQIAAAIEYTPRDVFDFRSFTIDEQTIQAAFRKLKMSHAPGPDGIPSCILKNCSHSLATPLAKLFSRSLQEGKFPDEWKVSSMFPVYKKGDKKDVHNYRGITSLCSCSKLFEILINDALFNSCKSFISTAQHVFFPKRSVTTNLTEFVSLCLLTLDSGGQIDTIYTDLKAAFDRVDHGILLAKLDKLGVSHSLVRWFESYLTNRLLYVKIGSSQSESFTNLSGVPQGSNLGPLLFLLFINELSQLLPPGCRLFYADDVKIYMIIDSTADCELLQSRVNIMANWCSSNFLTLSVNKCNVISFHRRQKPIIFDYRICDTVLNRVDQVKDLGITLDQEMNFKPHYNDVIAKANRQLGFIFKLSDDFRDPLCLKGLYCSLVRSILEFSVVVWCPYQSCWIARLESVQRRFLRYAFRHLPWRDPLNLPAYEDRCRLLGIETLEQRRRRAQAVFAAKLIVGDIDSPELLRQLNIYAPEQVIRQRNFLQLVPCNRLYGMNEPLRASADALNEAFTAFDFNLPLISFIRRLSLS			
	DongAa	GCF_002204515.2	Aedes aegypti	GGTGGGAGCTCAGGCAACATAACTCCTGGGCGTCCATGAAAAACCACCATCTTGAGTAGGTGGGAGCTTGAGGTAAAATATCCTGGGCGTCCATGTAAAACCACCCTCGGCGAGCAGCGGCGCTTGAACCAAGCTATTAAGCACTGTAATTGGAGGAAATGCCAATGCAGAACCCGTAGCTTCCGGGA	TGAGAAGCTCATCCAGCAGACTCATTAGGATAAGTTAAACCGACCAATGAGATCCACAGAGCCTAATCCCCTTTGGCATACAGATTGCCCGGGGTAGGTGAAAGTTTCCAGCATTTTTTGCTGAGAAGTGCCAAAACTCTATAATAATAATAA	MKANPAMETRSMRNRRMRSPEEGAPTGAGPGTGDRASGQRLEEDNVQERPSSQRAQPVPRTRNRNGSSRNHQGHAAPTDVAVADRRQSLTLAGGRRQRIMWTREMNMYVIRCYYVCTRMETDMSGRSGMLEMFNERFPRFARQLDLNKLYTRQRAIMSHNMLTAAEVEYIKLEVQREIGEEGTRSSDTSRRSSVGLDAPITSESRDSPAPTAPGPTADMQRQQLRDELALHMGTAVTQFRGTDPMSRHRIPKLRYSYRLTSAVSILNQDILPQYLEAVENLEDLQFIVYSAAIAVVKTLGLRTRPQGEDGTRPNAQKPVWMRRLESRIATLRSKVGRLTQYKQGNRSTRLVRHVAEIVRPAELRDLREVDITEILDTHVQRLSALAKRLRRYGECSKRKEQNRMFNINEREFYNRIRNDKVDFGEGLPEIGDVTQFWANLWENPTEHNGDGMWLAEEERQCDGIGDMTAVVVTAQDIREATRYTRNWAAPGPDFVHNFWYKKLTTIHGRMAECFNTVLGDPTQLPEFITRGVTFLLPKDQHTADPAKYRPITCLSSLYKVLSSVIARRVQAHCDANNVMTEEQKGCRRNTQGCKDQVIVDAVIVGQATQKQRNL	cd01650	530	802

					SMAYIDYKKAYDSVPHSYLLKVLQLYKVDGNV IRLMQHAMGMWSTSLHITDGTAVLRSRTLSI RRGIFQGDTFSPLWFCLAMNPLSKALNQCNY GYQLKSGERSTRVTHTFYMDDLKLFAESVQR LHQLLQLVTTFSNDIRMEFGIDKCRSIHLRRG QVMDASCFRVNEQEEIRNMVEGETYKYLGFL QLRGIRHTMIKKELQEKFLSRVNCILKSFLSAG NKVKAINTFAVPLLTYSFGVVKWTKTDLEAIE RAVRVAFTKHRMRHPKSSIERVTLPRAAGGR GVTDIQALCVSQIQQLRAYFVESQNRHEIYRT VCEADHGFSALHLAQEDYQLNCDIKTVDEMI AMWKQKELHGTHPHQLELEHIDKVASNTWL VRGDLFSETEGFMVAIQDRVIATKNYRHYILH EDVEDRCRKCNSVGETIEHVVAGCSVLAGSA YLDRHNEVAKIVHQQLALKHNLVDRFVPYYK YLPDPVLENSCIKLYWDREIITDVLIRANRPDIV VYDKRMKRVTLIDIAVPLDHNVQSTFSNKIAK YHDLAEELKQMWHLEDVRIVPVVLSATGIVP KSLLRSLDELELKKDLHSIQKAVILGTCSIVRRFL NHHN		cd01650	589	836
					MNQRHYCAKARSLPGSNAPYGSSTVSNEQE PLHRCPDVVLNEQGFSRFRERTRVNKLVHKV KGKGRSDRRLRFGTWNIGTLTGKSMEVVDT MTRRKINIMCLQETKWVGAKARELDTSGFKL WYTGKVKNRNGVGIIVDKQWKKDVVDVKR VGDRIISIKLVVEGGAFHVISAYAPQVGSDEQ HKIRFWEDLESLVQGIPLGDKIFLGGDLNGHV GREVTGYGSIHGGHGFGVINAEGKTILDFSST FDLLIANTCFKKRDEHLITYKSGMTSSQIDFFLL RRVDRKFCINCKIIPGESLTTQHRVLVMDFRV EQKLRKRHHTKNPRTRWWRMKGEEQRSFL RRVGEEAKWDGNGSAEEMWREMAEVIRRT AKESFGESKGIGPRDKESWWWNASIQEKIKI KRECFKEWSLCRNADNWEKYKAAKKETKVA VSEARTRAYEGLYQSLGTKEGEKGIYRIAKSRE RRTRDLDQVKCIKDKDGEVLAQEEKINERWK SYFYELFNEGQKTLPSLGRLCTREEDQNFDYY RRIRDFEVKEALKQMKNGRAVGPDNIPIEVW	TAGTGGGA	ACTCCGCCTATGTT ACACTTGCGGTACA TAGCCGGTCCCAA GCCCGGATAAAGG AGGAGGGTTGTGT TAGGTCTTCGGCAA CCAACATAAAAATA TAGCCGAACCCCCA TGAC	Arachis ipaensis	GCF_000816755.2	RTE-1_Alp	

					KGLGEKGINWLTKLFNEILRSKKMPDEWRKS			
					TLVPIYKNKGDIQSCGNYRGIKLMSHTMKLW			
					ERVIERRLRKETQVTENQFGFMPGRSTTEAIY			
					LLRRMMERYRSNKRDLHMVFIDLEKAYDRVP			
					REVLWKVLEKRRVRIAYIRAIKDMYDGATTSV			
					KTQGGVTEEFPIGIGLHQGSSLSPYLFTLVLEV			
					LTEHIQEPVPWCMLFADDIVLMGESREDLNK			
					KLELWREALEVYGLRISRSKTEYMECKFSLRRE			
					NPNIEVKIGENILRKVKSFKYLGCIIQDNGEIEQ			
					DVNHRIQAGWSKWRSASGFICDKKVPLKLKG			
					KFYRTAIRPAMLYGTECWAAKGEHEHKLSVA			
					EMKMLRWMSGHTRLDKIRNEDIRERVGVAP			
					IVEKMVESRLRWFGHVRRRPIEHPVRRVDE			
					MEDGQRAKGRGRPKKTIHEVVKRDLHVNGL			
					SVDMIHDRAQWRRLIHVADPT			
				TGATGCGATCGAGAAGAA	MTLEEGNVRYGARPVRTHQAGAVPGPKPRR	cd01650	548	804
				AAAGAAGAA	GPPQRLEGLPSSWSARRRGRKATGNTNTKQ			
					VKKSLNIMHWNAEGISNKKTELENFLKDNEIH			
					ISCIQETHLQEEASFKVRGYQCFRNDRQGRKK			
					GGVLTLVRNNLNTVEIKKYTGEAEYIHNRITTS			
					RGNLDLVNYYCPNDTNLVLETIEVPDTDFIIAG			
					DFNSQSQSWGYNRLDRRGEEIEAWQDENHL			
					LLVNDPLDQPTFYSRAWHSTTTPDLALCTENL			
					HRHLSRKVETQLGGSDHRPVLLTIVESLPEDS			
					PKRPRWNYKKADWRLFSIRTNELTKNIAVFG			
					KNINNAIREFNENILKSAKETIPRGVRKNYTPY			
					WTPELQKTHNDLIKAREEAETHPGRENNNKL			
					QRCKAHYLKTKLECTRKSWRDKTHKLNLEKD			
					TTKLWRLVKGLSDEGNKNDKISLEVEGEMEM			
					EKRAAHTFVKAYAQESNIKIPRERQKEVRQEE			
					RRKRSTNNIPEAMEKDITMEELKKAINQLKQR			
					KSTGPDEITNEMLQHLGNTALQTLLDIFNLS			
					WKIGQVPQCWKEATMILILKPGKSKAQPLSY			
					RPISLISCVCKTMERIIACRMQWYLETQHIIVP			
					EQAGFRQFRSTEDQTIYLSQVIEDSFQAQKV			
					MLATFVDLQKAFDKVWKEGLLVKLLRSGIQG			
					NMYQWTKSYLHNRRARVLVDGYQGRKMLL			
					RQGVPQGGVMSPLLFILFINDLMNELPKGIH			
	Ingi-2_AC	GCF_000002075.1	Aplysia californica	CCGGAGTGGGGCG				
				ACCATCCCCGACCC				
				GGCGGGTCCCAGG				
				TGGCGGATAGGGA				
				ACGGCCACTGTTAT				
				AGAGGTTAGCTGT				
				GAATATATTGAATA				
				AGCAGCCCCGGAC				
				CAACTGGCGATGCT				
				CTTACCAGGACGG				
				GGTGAGGTGGTAG				
				GTGGCACTACCAC				
				GTCATTAAATGCCC				
				TAGGAGTCCTCTGA				
				CGGGAGCATCATG				
				AGGAAAAGGACCG				
				CATTAAGTCCTGGC				
				TACGGGGGGGATC				
				CGGTAACAATTCCT				
				CCCTGTGCTCCCAG				
				GGTAGGTCCTTGG				
				ACCTGGAGCTGGA				
				GTGCTGAAATCGG				

	JAM1	GCF_0022045 15.2	Aedes aegypti	TTGGCGCGGTAAA TGGCTGGGCATGG CGTACCATTGGTAC CTCGCGTACCTGCA GGAATAAAATAGA CCCCTTTGTGCGGT CCTTAGCCTCTTGC CCAGCAACTCCTAT CCCTACCTCCTCGT GGTACTGGCCGGG GTACGAGTAACCTT GGGGAAGATCGGG TAACCAACCCCCGG TGGGAACTTTGGTC GTATGCTGACAGG GAAGGGGGGGTTT GCTTTTGCAAACCT GGAGCGTCTGTACT CCACGTTAGGAGC GGCTCACAACAGC	AGGTATTGGGCCG CAAGGCGCACTCTA GAGAACCACAATA CCACGTAATACTGC AATAT	TAATTGATGTAAAGCCAAA TAAGTAAGTAAGTAAGTA AG	AALYADDLVLWCSEEFVTTATYRMQLALDKL TTWTKTWCVTANKEMSSATLFTLSQAEPKHL TLDKTPLKFENQQKYLGIIFDKRLTWKQHIQN AENKARKKLNIMRKLAGTQWRANEKTLRQV YLGAIRPHLEYGSSAWMTAAQTHKYTLEKVQ NQALRIITGAMRSTPIEKMEQKTSIVALQRR WECKALVQYTKATAMEDHILHTRTERSNTAR LKRSNFIKESRKIQRSLQDHLPATVNMLNHTP EAPSWEKKNTNISIHTSIPMFTTKDDHSDLSK KSLTLSMLEERYPREAWIRIYTDGSATDAVKR GGAGVYVQFPNRRRQAEAVPTGRHCTNYRA ETEALVQAANIIKNVVDPQSQVVFLTDAMSV LQAINSGKLPRLRKEIEDIPCLRIMLQWIPSHC GVLGNEQADRMAKLGAQKEQPDNSVNYTEI KTIINTLFKAPQTPDGYHELSRQEQVIIFRLRT QHTRLRQLMYKCFKIVQSPKCQCGEADQTVE HILQHCKTHRALRQETWPTPLTLEEKLYEPPD ELTRTAQFMTDTGLQV	MGDMQRRVIGWWPINERMCKLRIKGRFFN FSIINVHSPHSGSTDDDKDAFYAQLEREYDRC PSHDVKIIIGDLNAQVGQEEEFRPTIGKFSAHR LTNENGLRLIDFAASKNMAIRSTYFQHSLPYR YTWRSPQQTESQIDHVLIDGRHFSDITDVRTY RGANIDSDHYLVMVKLRPKLSVINNVRYRRP PRYNLERLKQPDVANAYAQHLEAALPDEGEL DRAPLEDCWRTVKAAINDAAESVVGYVERSS RNDWFDEECQEVLEEKNAARAAMLQHGTR QNVERYRLKRKQQTRLFRDKKRRLEEVECQE MELLYRSQETRKFYQKLNTSRKGFVPRAEMC RDKDGSILTDGREVIERWKQHYDEHLNGAEN TGTEGQDSEGDGYVSTADSGNQPAPTMGE VKDAIQQLKNNKAAGKDGIGAELIKMGPDRL AACLHRLIVRIWETEQLPEEWKQGGVICPIYKK GDKLECENYRAITILNAAYKVLSQILFRRLSPIA NEFVGSYQAGFIDGRSTTDQIFSVRQILQKCR EYQVPTHHLFIDFKAAYDSIDRIELWKIMDEN SFPGKLTRLIRATMDGVQNCVKISGEHSSSFE SRRGLRQGDGLSCLLFNIALEGVMRRAGLNS	cd01 650	457	716

	RTE_Ele2	GCF_0022045 15.2	Aedes aegypti	TGGGTTGTAAAAT GGTGAGTCGACAA CCAGGAAGGAGCG TCCAACATAGCTCT	GTCTGTTCCCCATG TCAGGGGCGGCTG ATCATCGTCCGAGT GCCAGAGAAGGAC TCTAAGCTAAACTG CGCACTATGGCCCT CCGAACATTTAGG GGGAATGGTCCTC CGGAAATCTAGGG GGTTGGTGTCAGG CCCTGCAAGCCAAC CGTAAAAACACATC AGCACAGGAACGT CAACGAGAGAATA CGGACCGGAACAA TCGGCAAAGACCA CAGCGACGAAAAT GGACTAGCGATTG GAAACTCGGTACG TGGAACTGCAAATC TCTCAACTTCATTG GAAGTACTCGCATA CTCTCCGATGTACT GAAGACCCGCGGT TTCGACATCGTAGC GCTGCAGGAGGTG TGCTGGACAGGAG CATTGGTGCGAAC GTTTAGAGGTAATC ATACCATCTACCAG AGCTGCGGCAACA CACGCGAGCTGGG AACAGCTTTCATAG TG	TAGATTCAACGCAATAGCA AGGAATTGTTGCCCATCAA GTATCAAGTAAGTAAGT	RGTIFTRSGQFVCFADDMDIIGRKFETVADLF TRLKREATRVGLMVNASKTKYMILVGGTERD RTRLGSSVTIDGDTFEVVDEFVYLGSLLTADN NVSREIRRRIISGSRAYYGLQKKLRSRKIHPRTK CTMYKTLIRPVVLYGHEAWTMLEEDLQALGV FERRVLRTIFGGVQENGVWRRRMNHELAQL YGEPSIVKVAKAGRIRWAGHVARMPDNNPV KMVFATNPVGTRRRGAQRARWIDQVHQDL ESVGHSRGWREAAMNRGNWRNIVGEALSR	MGSIPNFSPTTKNGERRNNERYFRQPTWQR NKDYEWKLGTWNVRTLNEPGRASLLARELQ KVGVGVAAIQEVRWPRSGEREFRAVDPIANT TFKYNIYHSGGDKAEHGVGFIVIGKQMKRVM	cd01 650	561	821

	RTE_Ele2B_AAe	GCF_002204515.2	Aedes aegypti	TGGGCTGTGAAAT GGTGAGCCGACAA CTAGGAAGGAGCG TCCAACATAGCTCT GGTCCTCACAAGTC CCTACCTCACGCTT CCACGGGTCAAAT GATGACAAAGACC GCCAGCTAAGGGT	GGTCCTCGCAAGTC CCTACCTCACGCTT CCTCGGGTCTAACG ATGACAAAGACCG CCAGCTAAGGGTT GCGTACTTAGCTGG TAGTGCAACCTGG GCACTGTTGTCCTT CTGACATCAGCTAG AGTGAGGGGGTGC GTCCCGAGCGTCTG TTCACCAGGAGGT GCGGCTCAAACAG CGTCTGTTCTGGTA TCCAGCGGCTGAG TATGAAATGTTGTA TCACGTCAACTAAC CTAAGGTGGCAGC CCCATCAACGTGAT GTAGGTAGCGCAA CCCCGGTAA	TAGATTCAACGTAGCGGA TTGTAGCCCATCAAGTATC AAGTAAGTA	KWKPISERICVLRIRGKFFNYSLINIYAPTNDKP DDVKDAFYEGLDKAYGECPKHDVKIVIGDAN AQVGREDFFRPIIGKESLHSVTNDNGLRLVTF AAARGMAICSTYFARKDIRKHTWRHPNGDIC NQIDHILVDGRHFSDVIDVRSFRGPNIDSDHY LVVSKIRARLSTVANSRPQRTMRFNIQRFSAD GVTAEYHQKLDERISEVNVSENLNNLWDAIH GVVSETAREVIGTARRRPRSGWFDEECQRVT DEKNVARRRMLVSGTRLNRERYREARSAEKR IHRRKKKEFDENVIAEAQNCMEQNDMRRFY ETVNGVRRKSAPSPVMCNDREGNLLTDKTM VAARWKEHFETLLNGGNDRASENRMNIDN DGQAVEPPSLDEVKKAIKELKNNKAAGKDEL PAELLKHGSEQLHRIIHHILLRIWEEEELPASW MDGLICPLFKKGHRLECANYRGITLLNSAYKI MSRILFNRLRPLEESFVGEYQAGFREGRSTSD QMFTLRQILDKFREYNLQTHHLFIDFKAAYDS VKRNELWEIMMEHGFPTKLIRLIRATLEGSKS SIRVADEISTSFVTLDGLKQGDALSNLLFNIALE GAIRRACVHRNGTIITRSYMLLGFADDIDIGI DRRAVEEAYVPFKRETTRIGLIINTTKTKYMVA GRHRGPIGDVGSEVVIGGDIFEVVEEFVYLGT LVTCDNDVTREVKRRIAAASRAFYGLRNQLKS RSLQTKTKLALYKTLILPVALYGHESWTLKEVD RRAFGVFERKVLRTILGGKLDNGIWRRRMNH ELYQVYNEMDIVKRIKHGRLRWAGHVARMS EERQAKQIFSREPGRGRRLRGRPRTRWLFAV EEDLRSLNVQGDWKRLAQDRVQWRRLLHS A	MGSISNFLPTTKNGERRNNERYFRQPTWQR NKDYDWKLGTWNVRTLNEPGRVSLLARELQ KVGVSVAAIQEVRWPRTGEREFRAVDPVAN TAFKYNIYHSGSDKAEHGVGFIVIGKQMKRV MRWKPINERICVLRIRGKFFNYSLINIYAPTND KPDDTKDAFYDCLDKTYGECPKHDVKVVIGD ANAQVGREDFFRPIIGTESLHSATNDNGLRLI TFAAARGMAISSTYFARKDIRKHTWMHPNG ELCNQIDHVLVDGRHFSDVIDVRTFRGPNIDS	cd01650	561	821

								830
							570	
						cd01650		
			TGCGTACTTAGCTG GTAGTGCAGCCTG GGCACTGTTGTCCT TCTGACATCAGCTA GAGTGAGGAGGTG CGTCTCGAGCGTCT GTCCACCAAGGAG GTGCGGCTCAAAC AGCGTCTGTTCTGG CATCCAGCGGCTG AGTATGAAATGCT GTATCACGTCAGCT ACACCTAAGGTGG CAGCCCCACCATCG TGATGTAGGTATCG CGACCCTGGTAA	ACCGCTAGCGCAT GACGGCTCACCATA GTAGCATGTCCGA AAGAACTTGAAACT ATTGAGAACCAGA GCTACAGGTCCAAA GCCCTGATAAAGG TGGATGGTTGTGAT GGCTATGACCGTG GTCATCATGTAAAG AACAAGCGGACGA TGCTGTATCGTGTC AGCATCGAGGAGG CAGCCCCTCTAG	TGATGCTACGCTGTAGCCA TCAAGGTATCAAGGTAGG TA	DHYLVVSKIRARLSTVANTRTQRPLRFNIQRLS TEGVADEYRRKLDERIRGYNVGENLNNLWES IHGAMSTAAREVIGTAQRRPRNGWFDEECQ NVTDEKNVARSRMLVSGTRQNRERYKEARA AEKRIHRRKKRQYEEDIITEAQNSIEQNDMRR FYETVNGVRRKTAPSPVMCNDREGNLLTDKI LVAARWKEHFEQLLNGENTSASENRLIISNDG QAVEPPTLDEVRRAIKELKNCKAAGKDELPAE LFKHGSEQLYEILHRTLLRIWEEEELPASWLEG LICPLYKKGHRLECANYRGITLLNSAYKIMSRV LFNRLRPLEESFVGEYQAGFREGRSTTDQMF TLRQILDKFREYNLQTHHLFIDFKAAYDSVKR NELWQIMVEHGFPAKLIRLIRATLDGSKSSVR VADETSTSFVTLDGLKQGDALSNLLFNIALEG AIRRAGVQRSGTIVTRSHMLLGFADDIIIGID RRAVEEAFVPFQRETARIGLTVNTSKTKYMIA GRQRGSNRDVGSEVVLGGEKFEVVEEFVYLG TLVTCDNDVTREVKRRLAAANRAFYGLRNQL KSRSLQTKTKFALYKTLILPVALYGHESWTLKE VDRKAFGVFERKVLRTILGGKQENGIWRRRM NHELYQVYKEVDIIKLIKHGRLRWAGHVVRM PEERQAKIIFSREPGRGRRLRGRPRTRWLFAV EEDLRALNVQGDWKRLAQDRAQWRRIIHSA	MHDVGSATLVRWPIEDSSPTKKGKSRANGLI LRQQTRQRIKDNDWKVGSWNVRTLNEPAR VGLLARELRNVGVNVAAIQEIRWPRTGEREF RAVDPIANTSFKYHIYYSGGDRAERGVGFIVIG KQMKRIIRWKPVSDRICVLRMKGKFFNYSLISI YAPTNDKPDDMKDEFYESLDKAYGECPKQD VKIVIGDANAQIGKEDFFRPVIGTESLHSVTN DNGLRLVTFAAARGMAISSTYFARKNIRKHT WRHPSGDACSQIDHVLVDGRHFSDVIDVRTF RGPNIDSDHYLVVAKIRARLSSVTSSATHRTLR FNIQRLSTDGVAAQYRQQLDEKLGRVNVAG DVDSLWDPIHEAVTTTAREVIGTGQRRRRND WFDEECQRVTDMKNVARSRMLVAGTRQNR ERYRAARAEEKRIHRRKKRQHEESVVADAQE SMNRNDMRRFYATVNGARHNTVPVPAMC			
	RTE_Ele5	GCF_002204515.2	Aedes aegypti					

	RTE_Ele2C_AAe	GCF_002204515.2	Aedes aegypti	GGGCTGCAAAATG GTGAGTCGACAAT CAGGAAGGAGCGT CCAACACAGCTCTG GTCCTCACAAGTTC CTACCTCGCGCTTC CACGGGTCAAATG ATGACAAAGACCG CCAGCTAAGGGTT GCGTACTTAGCTGG TAGTGCAGCCTGG GCACTGTTGTCCTT CTGACATCAGCTAG AGTGAGGAGGTGC GTTCTGAGCGTCTG TACACCAGGAGGT GCGGCTCAAACAG CGTCTGTTCTGGTA TCCAGCGGCTGAG TATGAAACGCTGTA	TAGCCCATCAAGTATCAAG TAAGTA	MGSIPNSLSTTKNGERRRNERYFRQPTWQR NKDYDWKLGTWNVRTLNEPGRVSLLARELR KVGVCVAAIQEVRWLRSGEREFRAVDPIANT AFKYNIYHSGGDKAEHGVGFIVIGKQMKRVI RWKPINERICVLRIRGKFFNYSLINVYAPTNDK PDDVKDAFYECLDKTYGECPKHDVKIVIGDA NAQVGKEDFFRPIIGKESLHSVTNDNGLRLVN FAATRGMAISSTYFARKDIRKHTWQHPNGEL CNQIDHVLVDGRHFSDVIDVRTFRGPNIDSD HYLVVSKIRARLSTVSSSRNRQSLRFNIQRLSA DGVAADYHQKLDERISEIDESVNLSDLWESV HGAVSTVAREVVGTAQRRPRNGWFDEECQ RMTNEKNLARSRMLVSGTRQSRERYKEARA AEKRIHRRKKKEHEEAVIAQAQEAMEQNNM RRFYESVNGVRRKTAPSPAMCNDREGNLLTD KTMVAARWKEYFESLLNGDNGSGSGSRIQID DDGQAVEPPTLDEVKKAINGLKNNKAAGKD ELPAELLKHGSEQLHELLHRIISRIWEEEQMPT SWLEGLISPLYKKGHRLECANYRGITLLNSAYK IMSGILFNRLRPYEESFVGEYQAGFREGRSTT	cd01650	561	821

	JAM1C_AA e	GCF_002204515.2	Aedes aegypti	GGGTGGTAAATGA CTGGACAATGCGT ACCATTGGTACTTC GCGTACCTGCAGG TATAAAATAGACCC CATTTGTGGTCCTT AGCCTCTTGTCCAG TAACTCCTATCCCT ACCTCCCCGTGGTG CCGCCTGGGATAC GAGTAACCGTAGG GAAGATCGGGTAA CCAACCCTGGTGG AACCTTGGTCGTAT GCTGACAGGGAAG GGGGGCTCCTCTCT TCTGAGGGTGTAG CTTATCAGAGCGTC TGTTCTCCATGTTA GGGGCGGCTCAAA ACAGCGTCTGTTCT CCATGTTAGGAGC GGCTGATCATCGTC CTAGTGCCAGCGT GGGACTCTAAACA	TCACGTCAGCTACA CCTAAGATGGCAG TCCCATCAACGTGA TGTAGGTAGCGCG ACCCCGGTAA	TGAAGGACGTAGAGCCAG CAAAAGTAAAGTAAGTA	DQMFTLRQILDKFREYNLQTHHLFVDFKAAY DSVKRNELWQIMSEHGFPAKLIRLIRATLDGS KSSVRVVDEISSSFVTLDGLKQGDALSNLLFNI ALEGAIRRAGVQRSGTIITRSHMLLGFADDIDI IGIDRRAVEEAFVPFKRETARIGLTINTTKTKY MIAGGQRGSGRVSGSEMVLGGDKFEVVEEF VYLGTLVTCDNDVTREVKRRIAAASRAFYGLR NQLKSRSLQTKTKLALYKTLILPVVLYGHESWT LKEVDRRAFGVFERKVLRTILGGKLENGIWRR RMNHELYQVYKEVDIVKRIKHGRLRWAGHV ARMPEERQAKIIFNREPGRGRRLRGRPRTRW LFAVEEDLRALNVQGDWKRLAQDRAQWRR LIHSAQIHRSEL	cd01650	550	810

	JAM1B_AAe	GCF_002204515.2	Aedes aegypti	GTGCTGTGCACGAT GATCCTCCGGCGA GACAGGGGGTTGG TGCAGGCCTTACAA GCCAGCCGTAA	TAGATGTTAACTAAAATAAA TGAAT	RARIGRRVTIDGDEFEVVDEFVYLGSLITSDNN CSREIRRIIAGSRAYYGLHKTLRSGKLHFRTKC TMYKTLIRPVVLYGHETWTMLEEDLQALGVF ERRVLRTIFGGVCENGVWRRRRMNHELAQLY GEPSITKVAKAGRVRWAGHVVRMPDNNPA KMVFNSNPAGTRRRGAQRARWFDQVEQDL GSVGRSRNWRLAAMDRVSWRNIVAQVMS	cd01650	553	795
				GGTAAATGGCTGG GCATGGCGTACCAT TGGTACCTCGCGTA CCTGAAGGAATAA AATAGACCCCTTTG TGCGGTCCTTAGCC TCTTGCCCAGCAAC TCCTATCCCTACCTC CTCGCGGTACTGGC CGGGGTACGAGTA ACCTTAGGGAAGA TCGGGTAACCAACC CCCGGTGGGAACT TTGGTCGTATGCTG ACAGGGAAGGGGG GGTTTGCTTTTGCT TTTGCTTCTGCAAA CCTTGAGCGTCTGT ACTCCATGTTAGGA GCGGCTCACAACA GCGTCTGTTCCCCA TGTCAGGGGCGGC TGATCATCGTCCGA GTGCCAGAGAAGG ACTCTAAGCTAAAC TGCGCACTATGGTC CTCCGAACATTTAG GGGGAATGGTCCT CCGGAAATCTAGG GGGTTGGTGTCAG	MKNQATNNQRENTNRDNRRRPQRRKGTSD WKLGTWNCKSLNFIGSTRILADVLKDRGFGIV ALQEVCWKGSMVRTFRGNHTIYQSCGNTHE LGTAFIVMGDMQRRVIGWWPINERMCKLRI KGRFFNFSIINVHSPHSGSTDDDKDAFYAQLE REYDSCPSHDVKIIIGDLNAQVGQEEEFRPTIG KFSAHRLTNENGLRLIDFAASKNMAIRSTYFQ HSLPYRYTWRSPQQTESQIDHVLIDGRHFSDI IDVRTYRGANIDSDHYLVMVKLRPKLSVVNN VRYRRPPRYDLERLKQPDVAAAYAQQLEAAL PEEGELDEAPLEDCWRTVKAAINDAAESNVG YVGRSRRNDWFDEECQEVLKEKNAARAVML QQGTRQNVERYKRKRQQQTRLFREKKRRLEE TECEEMEQLCRSQETRRFYQKLNASRNGFVP RAEMCRDKDGSILTDEREVIERWKQHFDEHL NGAESTGNEGRDNGGNAFVSTAEDGNQPA PTLREVKDAIHQLKNNKAAGKDGIGAELIKM GPERLAICLHRLIGTIWETEQLPEEWKEGVICP IYKKGDKLDCENFRAITILNAAYKVLSQIIFRRL SPVVNEFVGSYQAGFVDGRSTTDQIFTVRQIL QKCREYQVPTPLGIMDENSFPGKLTRLIRAT MEGVQNCVKVSGEHSSSFGSHRGLRQGDGL SCLLFNIALEGVMRRAGLNSRGTIFTRSSQFV CFADDMDIVGRTFEKVADLYTRLKREAAKVG LVVNAAKTKYMLAGGAERDRARLGSSVTIDG DTFEVVDEFVYLGSLLTADNNVSREIRRRIISG SRAYYGLQKKLRSKKIHARTKCTMYKTLIRPVV LYGHETWTMLEEDLQALGVFERRVLRTIFGG VQENGVWRRRMNHELAQLYGEPSIQKVAK AGRIRWAGHVARMPDSNPAKMVFASDPVG			

	Lian-Aa1	GCF_002204515.2	Aedes aegypti	GCCCTGCAAGCCAGCCGTAA	TAGGCTCTCTTTACGCTCATGCCGTTTTGCAGTGCCCTTTTTAGGGCGCTGTTCGAACCCATTGTGGTATGGAGCTACATGCTCTCATTTCGCTTATGCGATCTTCCCTCTTCAAGGGACCCCACTCCTATTTCCTCCATCTTTCCCTTCCCTTTCCTCTCCCATCGGGTAGATGAAATAGGCTCAAATATGGCGATGGCACACAATCTCCCAACTGGTGGGGGAACGTGCCTTTGGAGCCGGCCTTCTGATACCTGATACCTGA	TRRRGAQRARWADQVRIDLASVGQNRGWRDAATNRVLWREIVDSVLSV MQNGAAQVALVQEPYFRRGNFYLGNLVDPVFATFSKLEMANSRSMPRACVLVNKAIVATLISELTTRDVCAVTIDVSVGDLNRKYVYCSVYLPHDEPSPTDDFKRVVVHCVTKGLPLIVGSDANAHHIIWGSSDINLRGSSLMEYILSSTDLGLLNIGNRPTFMVSNREEVLDITLCSNRISHELTNWHVSDEESLSDHRYIFFEHSNVTAQTLRFRNPRSTNWELYIELVATKFHGYSPSIENPSDLDDAVDTTTSYIMEAFEEACPLRSVKTTRGTPWWNSDLTTSYIMEAFEEACPLRSVKTTRGTPWWNSDLTRLRKQCRRSWNRRRSAGSESFKSARKAYKKALRSAERSGWKNLCTNVSSLSEVSRLNKILAKALRSAERSGWKNLCTNVSSLSEVSRLNKILAKSKDFQVNEIRLPNGDFTSSDEEVLECLFNTHFPGCVDIASTDEPNVFSCSYESLASARSIVTTESIQWALNSFAPFKSPGADGIYPVLLQKGFEFIKHVLKKLLVSSFATGYIPKSWRDITVKFIPKGGRASYEEAKSFRPISLTSFLLKCLERIIDHHIRDVYLANMPLHVNQHAYQSGKSTVTLLHKVVYDIEKAFAQKQSCLGVFLDIEGAFDNVSFDAILEAARNHGLPTMITNWIHQMLKNRHLFSTLRQAAIRKLSVCGCPQGGVLSPLLWNLVADTLLRQLNNCGFPTYGFADDYLALIVGMCISTLFDLMQSALQVVESWCRQYGLSVNPNKTSIVLFTERRNRDGIRPLRLFGTEINVTDQVKYVGVILDSKLSWTPHIDFRVKKACMAFGQCRRTFGKTWGLKPKYIKWIYTTVVRPILAYGCLVWWQKGEVRTIQSKLGHLQRMCLMAMSGAFSTTPTAALEALFDVAPLHIYLKQEALSCSYRLWVLDLLEKNPVNRRSTHTSLFPLLVNWDKIVLAPSDLTIACNFPYRTFTTQFPSREEWTSGYLERSISNNIVCYTDGSLLEGRAGAGVYSRELRLNQFYSLGRNCTVFQAEIFALMCGVQSALQQRVMGKVIYFCSDSQAAIKALASANSRSKLVIACRTQIEELNSVNSVNLVWVPASHIGIEGNELADELARDGASHDFIGPEPAIPISKCWVKLQINSWAATQHKQYWNSLESCRQTKLYITEPSPKVAKYLTNLSKQNCSLLVRALTGHCRLYHMANIQRADSFVCDSCDSDYGTSYHLCRLNYHMANIQRADSFVCDSCDSDYGTSYHL	cd01650	468	725

					ICNCPVFAQMRFQLLGKHLLSETEYRSLNLQD ILLFLTRCGNEL			
	RTE_Ele4	GCF_0022045 15.2	Aedes aegypti	CCGATCTGGTTTTT TCGTTGCTGGTCTT TTTCGTGCTGAGAT TGCAAAAAGCTTAA TCCTGCCTAGTTTG GGTAGTGGCTACG GTTAGGTTAGCTCA GATCAATCTTCAGC ATAAAAGAACAGC AACGATCAATCTTT GCAGACTC	GGGCTGCAAAACG GTGAGTCGACAAC TAGGAAGGAGCGT TCAACACAGCTCTG GTCCTCACAAGTTC CTACCTCACGCTTC CACGGGTCAAACG ATGACAAAGACCG CCAGCTAAGGGTT GCGTACTTAGCTGG TAGTGCAGCCTGG GCACTGTTGTCCTT CTGACATCAGCTAG ATTGAGGAGGTAC GTCTCGAGCGTCTG TTCACCAGGAGGT GCGGCTCAAACAG CGTCTGTTCTGGTA TCCAGCGGCTGAG CAAGAAATGCTGA ATCGCGCACAGCTA AATCCAAGGTGGT AG	TGACTGGTAAGGTAAGTA	MPHQRDRPSVSWDVKQSWHDGAPARQEC WRRPNKPPVKNLIANNIGDNTTRNNRQRPT RRNKDYDWKLGTWNCKSLGFAGCDRIIYDEL HPRNFDIVALQELCWTGQKVWKSGHRAATF YQSCGTTNELGTGFIVLGKMRQRVIGWQPIN ARMCKMRIKGRFFNYSIINVHCPHEGRPDDE KETFYAQLEQIYDGCSPRDVKIVVGDMNAQV GREEMYRPVIGRNSLHAVSNDNGQRCVNFA ASRGMVVRSTFFPRKDIHKTTWRSPDNLTEN QIDHVLIDGKFFSDITNVRTYRSANIDSDHYLV AVCMRSKLSTVTNSRRSRTPRLNIERLRDSEV AQEYAQQLEVALPTEEQLGAATLEDGWRHIR SAIGSTATAALGPAAPNQRNDWYDGECEQL KNEKNAAWARMLQHRTRANEARYKQARNR QNSIFRRKKRHQEERDREAMEELFRAKDTRK FYEKLNRSRKGFVPQADMCRDCDGNLLTNE REVVERWRQHYDEHLNGDVASSGGGTEITL GARADDERLPPPDLQEVETEIRRLKNNKAAG VDQLPSELLKYGGDALARALHWVISKIWEEE VLPEEWMEGVVCPIYKKGDKLDCANYRAITIL SAAYKVLSQILCRRLSPIAREFVGQYQAGFMG ERATTDQIFAIRQVLQKCREYNVPTHHLFIDFK SAYDTIDREQLWQIMHEYGFPDKLIRLIKATM ERVMCVVRVSGTLSSPFESRRGLRQGDGLSC LLFNIALEGVIRRAGIDTSGTIFRKSVQLLGFAD DIDIVARNFETMADTYIRLRAEARRIGLNINVS KTKYMIARGSREDTARPPPRVHIDGDEIEVVE	cd01650	588	848

					EFVYLGSLVTADNDTSREIQRRIVAGNRAYFG LRRTLRSSKVRHRTKLTIYKTLIRPVLLYGHET WTLRAEDQRALGVFERKVLRTIYGGVQMED GTWRRRMNHELHQLLGEPTIVHLAKIGRLR WAGHVTRMSDTNPVKMVLENNPTGTRRRG AQRARWVDQVEDDLRTLRRMRGWRRAAM DRVEWRRLLRTAEATQALA	cd01650	539	790
	CR1-10_AMi	GCF_000281125.3	Alligator mississippiensis	CT	TGACCCCCAGTTGTCTTTC CTGCCTAGTGCAGGGGGC TGGACCCGATGATCTTCCA AGGTCCCTTCCAGCCTTAC AATCTATGAATCTATGAAT	MGQGFKLRLLGDGETLIGACPGTNTQESPVD RDTRRATLAGAEGIVLPQHDRKSRVSVGGLR CLYTNARSMGNKQEELALLLSSGHYDLVGITE TWWDSSHDWSVAIEGYTLHRRDRVGKKGG GVALYVKEQYTSLEADIGSKEGQLETLWVKL WGGGGERDLTVGVYYRPPNQGEELDLEFSR ELTEAAHARDMVVMGDFNYPDISWEEHSAK SDQSRSFLTCIEELFLTQEVHGPTRGNALLDLV LAKGDDLVTDLNIEGSLGDSDHETIMFTVHH KADKSDSRIEVLDFKNANFNKLRSLVGEVLW DQGPKVKGVHEEWLFLKDTILEAQRKSIPMQ RKGSRRAGKPSWLNREIMVLLRKKKEAYKQ WKLGTVPKEDYTTMAHTCREQIRKAKAATEL RLATGIKDNKKSFFRYVGNRRKTNGSVGPLLN NSGQLVTDTQEKAELLNDYFASVFHQTKGKI MPGRVQDEHGRNDHLPTIAVELVYDQLEKL DIYKSAGPDELHPRVLKELAEVIAEPLAKLFQN SWCSGEVPEDWKRANVVPIFKKGRREDPGN YRPISLTSIPGKILEKFIKESICDKLAEGKILNDSQ HGFIAGRSCLTNLISFYDQVTHHLDKREEVDIV YLDFQKAFDLVSHDVLMRKLEDCGLNCETVR WVGNWLQGRTQRVLVNGSVSSWREVASG VPQGSVLGPVLFNIFINDLDVGVKSSLAKFAD DTKLWGSVVTLEDRLQIQADLDRLASWADQ NQMKFNVEKCKVLHLGTNNPQHTYRLGGTK LTSTMNERDLGVIIDHSMNMSHQCDAVASR ANNTLACINRCISSKTKEVILPLYSALVRPQLEY CIQFWAPHFNKDVEKLERVQRRATHMIRDL QGKPYEERLRDLGLFSLKKRRLRGDLVAAYHY IRGVHQGLGEQLFTRAPLGKTRSNGHKLLED			

	CR1-2_ASi	GCF_000455745.1	Alligator sinensis	CCTGACCGTATTAT GAGGGACTTCGAG GCTCTGGGGACCA GGCTGAAGGAGAC GGGGGCGCAGGTG GTTTTTTCCTCTCTC CTTCCGGTGAGGG GGCGGGGACAGAG GCATGGGACCTGC ATTGGTGAGGTCA ACCAGCGCTTACGC AGCTGGTGCCGCC GGGCAGGCCTTGG GTTCCTTGACAACG ACCCATGCTTTCGG GAGGGGGGATTGC TGGGGCGGGATGG GCTTCATCTCTCCC CTAATGGCAAGCG TATCTTCTCTTCCAG GTTGGCTGATCTCG TACGGCGAGCTTTA AACTAGACTCGCCG GGGGAGGGGGCT GCAGGAAGACGAG GAGACGCCATCAG GGCAAGGGGTGCA AAGGTAAGTAATG GGGGTCTTGCCATC AGGGCAAGGGGTG CAAAGGTAAGTAA TGGGGGTCTTGCC ATCAGGGCAAGGG GTGCAAAGGTAAG TAATGGGGGTCTT GCCATCAGGGCAA	TGAGTCCCGTTTTTGTTTTT AGGGTGAGTGGTTTTGTTT GTGTTTTTGGGTTGTTCTC CCGCCCTGGCGGGGGGGTC GGGCTAGAAGATCTATTA GGTCCCTTCCGACCCTGAA ACTATGAAACTATGAAA	RFRLNIRKNFFTVRVSRLWNKLPPEVVQSPTL EIFKRRLDSHLAGVT MGNKQEELALRIASSNPDVVGLTETWWDSS HDWAVNIRGYRLYRRDREGRKGGGVALYVK EHYTSSASSMGSGEGRAEVLWVKIQGGQGE RDLTVGVYYRPPNQEEELDREFSGQLAETLKA RDVVVMGDFNYPDICWEEQSAGTDRSRRFL AELQDLHLIQEVHSPTRGDALLDLVLAAGDDL VRGLQVLDHLGNSDHRVLEFTIQRRVSRARS KVVALDFRRANFSEMRRLVGEELGSLRAEELS VQEEWSFLKETILGAQGVSIPTRSKAGRSVQK PPWLTKGVRECLLAKRVAYARWKGGVISKEE YTSSARACRRAVRKAKAEIELGLASRIKDNKKS FFKYIGRMKKAPGNVGPLQDMRGNLVVASE DKADLFNQFFTSVFLCRDRDFPAVIEAGREGI AFRPQVGEDRVRELLEGLDAFKSAGPDALHP RVLRELAGVIAGPLARLYERSWCLGQVPDDW KVANVVPIFKKGKREDPGNYRPVSLTSILGKLF EKIIKEHICDGPASEMMLRGNQHGFIRGRSC QTNLIAFYDQVTKAMDAGVAVDVVFLDFSK AFDTVSHPILINKLGDCGIDAYTVRWIADWLK GRTQRVVVDGSYSTWGEVGSGVPQGSVLGP ALFNFFISDLDVGVESNLFKFADDTKIWGEVS TLAGRERLQKDLDRLQGWANRNRMCFNTD KCRVLHLGSRNQQHTYTMGSSLLESTEAERD LGVIIDSKMNMGRQCEVMVGRANRTLSCIH RCISSRAREVILPLYATLVRPQLEYCVQFCAPH FRRDVDNMERVQRRATRMIRGQQGRPYSE RLRDLNLFSLHKRRLRGDLVTVYKLTRGNQKG LGETLFPLAPAGITRNNGHKLMERRYRLDIRK NYFTVRAARSWNQLPREVVLAPTLGIFKKRLD TYLAGLI	cd01650	470	720

			GGGGCTGCAGGAA					870
			GATGGGGGGTCGC					
			CATCAGGGCAAGG					
			GGCTGCAAAGGTA					
			AGTAATGGGGGTC					
			TTGCCATCAGGGCA					
			AGGGGTGCAAAGG					
			TAAGTAACAGGGG					
			TCCTGATGGTCTGG					
			CTAATCAGGGGGT					
			GGGGGGGCACACA					
			GTCTTGCAAGGCCT					
			TAAATGCCTGTACA					
			CCAATGCTCGTAGC					
	L2-18_ACar	GCF_000090745.1	Anolis carolinensis	GAGAGGCCTGCTG	TAAGAATATATGACCTCAA	MLLMHPLHFTTAFALCIVGGGWGRLEGYGD	cd01650	610
				TTGTGCTGCAGAA	CTGACAGACCCAGTCACAT	CRVGEGNGCERMPIYNRKLVNQITSFGPGEE		
				GAGGCCAGAAACA	GAACATGGAAAGCGCCTT	KGKGQCDRGGGCSDGWGAPIEVVSGRGRY		
				TCTCCAGCCAGCCT	AAATGTATATTTAAATGTA	GKGRPNHLIRPRVFGRTLIVPKRSPDIANLGS		
				GCCTGTTGCTGCCT	TAATTATGTATATTTTAAT	QDGGPSRLKVVLFNARSVNGKTSFIQDLILDE		
				CTGGGACTCCTTGT	GGCTATTCTTAATTTTATT	HVDLACITETWLDEMGGVNLTQLCPPGYSV		
				GAGTAGACTGCAG	GTAATTTTTAATCTTGATG	QHQPRSGGRGGGVAVVYKDSIPLTRCPVPQS		
				CCCCAGGCCCTTTC	GATTTTTAAATTTGATGAA	TMFECVHLRVGDRDRLGILLVYRPPPRCTTVSL		
				CATCTGGGTGTGAC	AACTGTTTTTAATGTGTAT	PELAEVVSGLALESRRLIVLGDFNVHAETNLS		
				TCAGCACTTCCTAG	GGTTATACTGTAAGCCGCC	GAAQDFMAAMATMGLSQLVTGPTHSAGHT		
				GAAGGACCATCTAT	CTGAGTCCCCTGATGGGT	LDLVFCQGWEEGGGVEELTIAPLPWTDHHLI		
				AAAGCCTGAGTCCA	GAGAAGGGCGGGGGTAGA	RFRLTAPPNLCRGGGPKMVRPRRLMDPEGF		
				CAAGGCCCTGGCA	AGTGATGTAATAAATAAAT	LTTLGEFPASSAGDPVDALVSLWNGEMTRAI		
				GGGCTGAGAGGCC	AAATAAA	DTIAPERPLSSNRAKPAPWFTEELAAMKRKK		
				TGCTGTTGTGCTGC		RELECVWRSGYNESDRTRLGSYLRAYAAAIKA		
				AGAAGAGGCCAGA		ARRAHLAANIASAQNRPAALFRVIRGLLTPTN		
				AACATCTCCAGCCA		QGGSPDNSAARCEAFARFFADKVALIRSDFD		
				GCCTGCCTGTTGCT		TILTAVSEDVTSAPACPVLLDSFQLVELEDVDK		
				GCCTCTGGGACTCC		ILGEARPTTCILDPCPSWLVREARGGLAEWVK		
				TTGTGAGTAGACTG		VVVNASLREGKFPASLKMAVIKPLLKKPSLDP		
				CAGCCCCAGGCCCT		TQFGNFRPISNLPYLGKVLEHVVASQLQGFLV		
				TTCCATCTGGGTGT		DTDFLDPAQSGFRPGHGTETALVALVDDLRR		
				GACTCAGCACTTCC		ELDRGSVSLLVLLDLSAAFDTVDHGILLGRLAG		
				TAGGAAGGACCAT		MGLGGTVLQWLHSFLEGRSQMVSLGDACS		

				APQPLTCGVPQGSVLSPMLFNIYMKPLGEIIR SFGVRCHLYADDVQLCHSFPPVTKEAVQVLN RCLAAVSDWMRANKLKLNPDKTEVLLVSRKA EQGTGLQPVLDGVTLPLKAQVRSLGVILDSSL SLEPQVSAVARGAFAQLKLVRQLRPYLGKPDL ATVVHALVTSRLDYCNALYVGLPLKTVRKLQL VQREAARLITGAAYRERTTPLLRQLHWLPISY RAQFKVLALAYKALNGSGPAYLSERVSRYDPP RSLRSSDEALLAVPSASQVRLAGTRDRAFSVV APRLWNALPIEVRQAPSLLSFRKRTKTWLWD QAFEHESAAKMR		
					CTATAAAGCCTGAG TCCACAAGGCCCTG GCAGGGCTGAGAG GCCTGCTGTTGTGC TGCAGAAGAGGCC AGAAACATCTCCAG CCAGCCTGCCTGTT GCTGCCTCTGGGAC TCCTTGTGAGTAGA CTGCAGCCCCAGG CCCTTTCCATCTGG GTGTGACTCAGCAC TTCCTAGGAAGGA CCATCTATAAAGCC TGAGTCCACAAGG CCCTGGCAGGGCT GAGAGGCCTGCTG TTGTGCTGCAGAA GAGGCCAGAAACA TCTCCAGCCAGTCT GCCTGTTGCTGCCT CTGGACTCCTTGCC TACTGCCCCCCTCC CTTTTCCCAGCGC TGGTGAGTGGACA TAAGAAGGGAGGG AAGACTTTTAAAAC TCCAAAGCCATCCG GAGGAAGAAGGCA GAGGAACTGACTTT AAAATCATTCTGTC TCTTGCTGCAGGCC TACTGCCCTCCTCG CTTTCCCCAGCATT GGCTATTGAACTAA GAGTCTGTCTTACT GAGGTCCACTCATT ATTTATATATTGCT			

	RTE-1_AC_1	GCF_000090745.1	Anolis carolinensis	CCCCCTCGATGGAT TGTCACCTTGCCGT GGTGAGGGGGGCTT GCGTGTTCCGATGA ACCTGTGGGCACA ACCACTGGAGTGA TGCACTCCCAGGA GTGGCCGCAGGGG AGGTTCCAGACCAA GCACAATCCGAAG ACCCAAAGACCTCA ACGGCGGAGCAGG CGGAGGATAACAT GGTACATGTTACAA CGGCTGTGAAGGC GGAAGAAGGCTGC AACAGACTGAGAA GCCACTGTCGTTGT GTTAACCACACCAC TGCTGGAACCTCAC TCTGTGAAGACTGT GTGTTGACCGGCC GTGCACCGACCTCC	GCTATAACAATCTT TAAATGCTGTTATA ATAACTATTATGCA CTGTGAAGCCTGGT CCCCCTGACTAGTT TACTAAGTCTATCA TCTCTAATTGGTCC TGCCCTGGCACCTT ATAATACCATATTG CACCATATTGAATT AGTTCACCATCATA ATCCCCACTGCCTG CCTTGCACTTTAGT CAGCTCAGACCCCT ATACCT	TAAGTAA	MLTKTSRKIFYCQLDTILSEIPKEDKILLGDFNA RVGRDSDLWPGIIGKDGVGNSNSNGILLLTK CGEHNLVITNTLFRQKNKLKTSWKHPRSKHW HLLDYVITRARDRRDVLLTRAMTGTDDCWT DHRLIRSTMAIKIVPKRRLQGRKTRRKMNTQ ALQEPSKRALLQTTLKDHLPTEHPENVEEHW NKLKTSIITACEESIGYLTKKHQDWFDDNDKEI QQLIDNKRKAFQTWQRDTNCAAKKKIYASAK AEVQRRTRELKNIWWTKKAEEIQHLADTHD AQGFFKATKIIYGPRNHGIQPLRSSDGTKILKD KTSIALRWKEHYQNLLNRSSNVAEETLSQIPQ QQTRDELAALPSLEEVSNAISQQKNNIKASGP DGIPAEIFKEGGPELIQQLHQLIEKVWMTEKIP ADFKDATIITLFKKGDRTDCGNYRGISLLTSAG KILARILANRLLPVSEDTLPESQNGFRPSRGTV DMIFTARQLQEKCREQNQPLYMAFIDLAKAF DTVNRSALWTILQKIGCPDKFVNILRLLHDD MMATVLDSNGSQSDPFKVESGVKQGCVIAP TLFSIFIAMILHLVDGKLPTGVEIIYRTDGKLFN LSRLRAKTKVTTTSVIELQYADDNVVCAHSEE DLQATLNTFAEAYEKLGLSLNIEKTKVLFQQA PANPSAKPGIQLNGVTLENVDHFRYLGSHLST KVNIDTEIQHRLSSASAAFFRMKQRVFDDRDI	cd01650	417	698

					RRETKVLVYKAIVLPTLLYACETWTVYRRHTKL LERFHQRCLRKILQJSWEDRRTNVSVLEEAKT TSIEAMLLRHQLRWTGHVVRMPDHRLPKQL LYSELKNGKRNVGGQEKRFKDGLKANLKNCG IDTENWEALALERSNWRSAVTSSAAEFEEAR TEGLREKRAKRKERQANPDRDRLPPGNRCPH CGRICGSRIGLFSHLRTHAQDTKVGRPSSSNY EGSPK		
				ACACATTAAAAAAA AATCACGCACAGG CGTCTTCCAACAAA AATAAAAACAAACC TACAAAAGTCCCAT GGCGATCAGCGAG TGGCGACGGGGGC AGGACTGTGAAAT CTGGAAGCCCCTA GTCACAGACTGGC ACATGGGCGGTGG ATATGGACTCAGTC GTTCTACCTCAAGG ACCGAGGCAGTTG AGTAGTCCGGCAG CTGTATCCATGACT GAGCAGCCCTTTTT AGGATCCGCTCTGC TCACCCCACACGGG GAAGGGGCTAGAA AAGGTGCCCTAAAC ATAGTCTGCCTCTC CTACCCTGACTGGA CTGCCGCGTCCAGA GGGGTCACCACTCT GCGGCCAAAAAAG AAAAATGAACTTTG GAACATGGAACGT ACGGACATTGTTAG ATAACACTGACAGC GAACGCCCCGAAC GCAGGACTGCTCTC ATTGCAAGGGAGC TGGGACGCTTTAA GATCGACATAGCA GCCCTTCAGGAGA CCCGGAGAGCAGG AGAGGGACAGCTG			

			AAGGAAGAAAAGG GAGGCTACACCTTC TTCTGGAAGGGAC TGCCTGAAGAAGA GCGAAGAATACAC GGAGTTGGCTTTGC TATCAGAAACGACC TGGTGAAGCACCT GACTGAAGCACCC ACTGGCATCAACGA ACGACTTTCAACCC TCCGAATTAACCTT GCCAAAAACCAAC AGGCAACCATCATA TGCGCCTATGCACC AACTCTAG	GGGGGACACGGAA AGAGCCTCCCCGAA GATTGAGTGAATTC AGTCGGGCGTCCC CTGGGCAACGTTTC TTGTAAGCGGCCG ATCTTTCCACCCCA AAAGCATTGGATG A	TAGTTGCTTGTGATTTCTTT TCTTTTTTATTTTATTTCCA TTATTTGAAATGTATTTGT TGTAGCAATGCTTTTGACA CGAAATAAATAAA	MDEYQRSLSRPLLTIMSINIEGLSLAKEELLAK MSEDISCDILCIQETHRDITMRRPKILGMQLA VERPHRQYGSAIFVRSGVAISATSLTEVNNIEI LSVELDSCTVSSLYKPPGADFYFTPPTSCHNHE AHFVVGDFNSHSCVWGYDEDDRNGEAVLT WADNSRMSLLHDSKLPPSFNSGRWKRGYNP DLIFVKESISHQCTKRVLNPIPNTQHRPICCVA YAAVRPKSVPFRRRYNFNKANWTKFTETLEA AISDIEPSIENYDLFVEAVKRSSRLSIPRGCRTSY LPGLNEESLNQLQEYLRLFQENPYSDGTIAAG QKLSTALANAKKDRWIELLENLDMSKSSRKA WQLLRRLDSDPLVNPGHANVTPDQIAHQLI QNGKTNCSRIKMKINRVPELETHQLSSPLNLK ELREAIKRCKTGKAPGLDDLMMEQIKHLGPK AENWLLKFYNQCLAHKQIPRAWRKTKIIAILK PGKDASNARNYRPISLLCHLYKVYERMLLNRL GPVIEPKLIAQQAGFRPGKNCTGQILHLTEHIE EGYEKGCITGTVFVDLTAAYDTVQHRKMLHK VYHITRDFDFTKTVQTLLENRSFYVEFQGQKS RWRRQKNGLPQGSVLAPTLFNIFTNDQPQP PLTKSFIYADDLGLTTQAKDFETVEKQLTNALK DLSSYYKENHLKPNPAKTQVCAFHLRNREAN	cd01650	469	724
	Vingi-1_Acar	GCF_000090745.1	Anolis carolinensis					

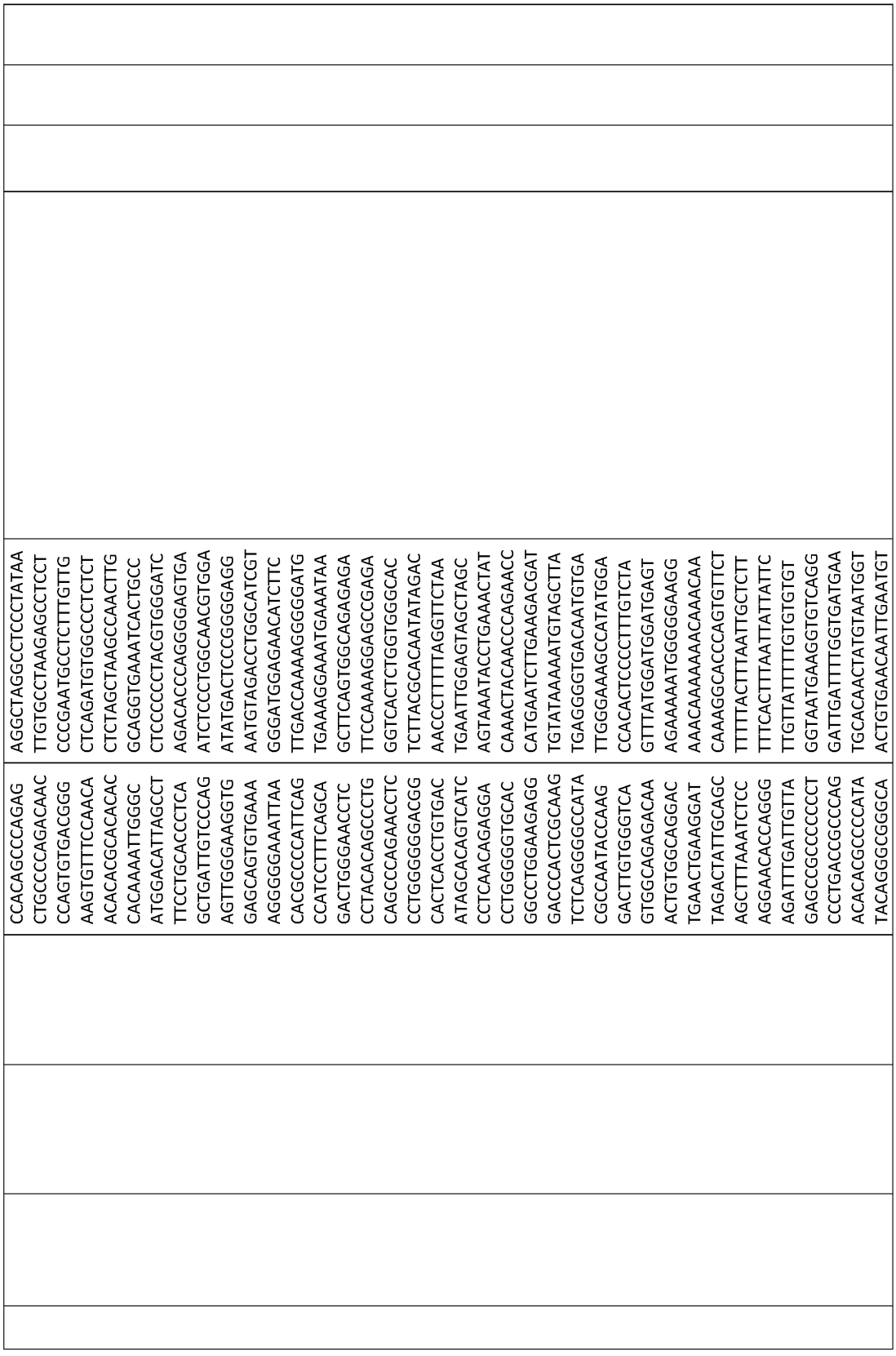
						RKLKVTWEGQELEHCFHPKYLGVTLDRTLTYRKHCMNTKHKVAARNNILRKLTGSAWGADPQVIRTSALALSFSTAEYACPVWHKSAHAKQVDIALNETCRIITGCLKPTPVDKLYKLAGIAPPDVRREVAANGERKKVEHCESHPLHDYHPPPTRLKSRKGFMRTTTPLDVPPATARVSLWAAKPGNSNWMAPQEGLPPGANQEWATWKSLNRLRSGVGRSKDNLARWHYLEESSTLCDCGAEQTTQHMYACPQCPASCTEEELFKATDNAVAVARFWSKTI			
	RTE-1_AG	GCA_001542645.1	Anopheles gambiae	TGCTCTGTAATGGGATGGGATCCGAAGCCCCATTGAGGGATAATACAGGCTCTCCCATCCAACTCCTATTCCGACACGTCCTCGTCGTGCAGAGTGGTAACATGTGTCACCACATATCCAAGCTTGGGTACACGGGCTTGACCCAACCCCTTGGGCGGATGGTGGCACATGGCGAACCAGGAGGGGGGTGGGTATCCCGGGAAACTGGGTGCCTGAACGTCGGGGGGGCAACCCGACGCTAAACAAAACGGTCACGTGGGCGCAGGGCCAGTCACCCAGTCCCATGAATCAACGACTACGGAAAGCACCAGAAATATTCGAAACGGACTTAACGATACCGACCCTACGCAATGCCCCCGGACTACTCTGAAA	TAGCGCCGGATAAGTAAGTAAGTAAG	MGSWNVRTLSEAGALKKLDDALATLSMDLVALQEIRWLGNGVHNRRGKHCYDIYYSCHDRHRVLGTGFAVGPRLKPAIMDFKAINDRLCTLRMRGKFFNISLINVHAPTEDKEEEEKDLFYGRLARIVDACPRHDLIIILGDFNAKVGREPMYRQYTGCHSLHEHSNDNGSRLVQFAAANNLVVGSTKFARKKIHKITWAHPGGESFNQIDHVLISRRRQSSLLNVRTYRGANIDSDHYLVGLVIRCRIARPRANGGGENTQARLNTDSLRDIAVQQEFKTALEESLLPEDRYETTSERWNALKTKIINCARNILPPRRGNTKSGWFDDECRQVTERKNTAYRAMQQRHRTRACAEEYSRLRREEKRVHRSKKHALEEQNMRELEQTREAYGPTRKFYQAIAGHRNNVVPKVTCCRNKDGDLVSNQPEVLSRWAQYFDELLNDQFSEQLEAPLADNVMLLPPSIEETRKAIRRLKNNKAPGTDGIAAELVKNGGARLENEIHQIVTEVWDSESMPCDWNLGIIYPIYKKGDRLDCNNYRGITVLNTAYKIFSLILQDRLVPHVEEIVGNYQRGFRNGKSTTDQIFTMRQILEKMAEYRNDTYHLFIDFKAAYDSIARVKLYDAMSSFGIPAKLIRLVRMTMTNVTCQVRVDGKLSGPFATTKGLRQGDGLACLLFNLALERAIRDSRVETTGTIFYKSTQILAYADDIDIIGLRLSYVAEAYQGIEQAAENLGLQINEAKTKLMVATSADLPINNPNLRRRDVQIGERTFEVVPEFTYLGSKVSNDNSMEVELRARMLAANRSFYSLKKQFTSKNLSRRTKLGLYSTYIVPVLTYASETWTLSKSDEALLAAFERKMLRRILGPVCVEGQWRSRYNDELYE	cd01650	521	783

					MYGDLTVVQRIKLARLRWAGHVVRMETDD PARKVFLGRPQGQRRRGRPKLRWQDGVEAS AIKAGITDWQTKARDRERFRTLLRQAKTAKRL	cd01 650	557	811
	RTE-5_AC	GCF_0000020 75.1	Aplysia californica	TGGCTGGAGCTAA TGCTCCTGGTAGG GCCACCCATGCCAG ACAGGTCGACGGG TAGAGGTCAGACA TAGCGTGATCCACC GGCCCTCCAGGTTT CGGAGGTTCAGCT CAGGGCTAGCGAC CCAGACTGGTAAA AAGCATATGTTACG GAAACAGCAACAA AGAGAAACCATATT TCTGTGTGTGTTGG ACCTTCAGAGTCGC TTGTGACTCGT	TAGTAAGTAAATAAGTAA	MNAGSESRQEAVCRKIEDLNTAKTKVIIGLW NVKTLYGTGKSAQVTAEMRRYGLHVLGISES QWTGSGWLRTATGETLIYFGREDNQHREVV ATILNKGLEKCLMEWKPINSRMMTLRLKGKH TNTSIIQCYAPTNDSEDESKDYFYEELQRGLTG IPRHDLIVVMGDLNVKVGEDNTNYERTMGQ HGMGTRNENGERLVDFCALNNLVIGGTLFP HRTIHNLTWNSPNGRDKNQIDHLMIGGMLR RSLQDVKVRRGADVGSYHHLVTSRIKFKLKKT FTREATRTQFNIQKLSDGNVTSAFRLELKNRF QALQDLEETTPNTDEVNTQWKLIVEAYTESS KTCLGVKKKNASNEWIQQGTWNIIEERREIRK RELAAKSIRLQEKYAEDYKEVCQKVKRLVRED KREAMENLATETEEAAAKGEQGNVYKITRIIC GKYHGGMSGPIKGKIGRLLTIEKEQEARWPE HFREVLNQPPPEDTPDIQEATLNLDISSDPPQ RHEVVKAVKSLKEGKSPGMDNLNAELFKVDP ELTTTILQPLFKTIRKGEQITEDWRTGVIVKVP KKRALGDLNNWTGITLLSVASKILAKIIINRMS DAVDICLRKEQARLRKERRCTDQIFALRNITE QCTERQRQLYVNVVDFEKVFDGVHRESLWCI LRSYGIPGKIVEVIKSFYQSFKHVVGNRTTAFE VKTGVRQGCNMSAIMFNLAIDWVMRKTTE DTSRGIRWTLFSTFEDLDLVDDLAQLSHTNQ HMKEKTNPLHKYERHIGLRISQKKTEMMTLN VSSPTLIMENNAHLPITDKFTHLGSINKTDGG TKEDIMNRLGKERSVHRSVINIWKSEQFSIQT TLKQYNSCVLSTLLYGFECWRKTGHDRAKQS SFHTTSLRQILRIFWPDNISNTDLLQKWRQED RTTIIKRRRWRLIGHVLRREPDSVVKTAFHWT PEGHRKRGRPKVTWRRKVESEMKSWNHSW GTLTKLGQNRQSWRSFVAARNSRWRNG	cd01 650	610	861
	BovB	GCF_0022637 95.1	Bos taurus	GCGCAGGCGGCTG CGACGGGCGCACT AAGCGCGGCCGAG	TGAACTGAACTGA	MKLSHARGATQDGRVMVERSDRMWSTGE GNGKPLQYSCLENPMNSMKRQNDRILKEELP RSVGAQYATGDQWRNNSRKNEGMEPKQK	cd01 650	610	861

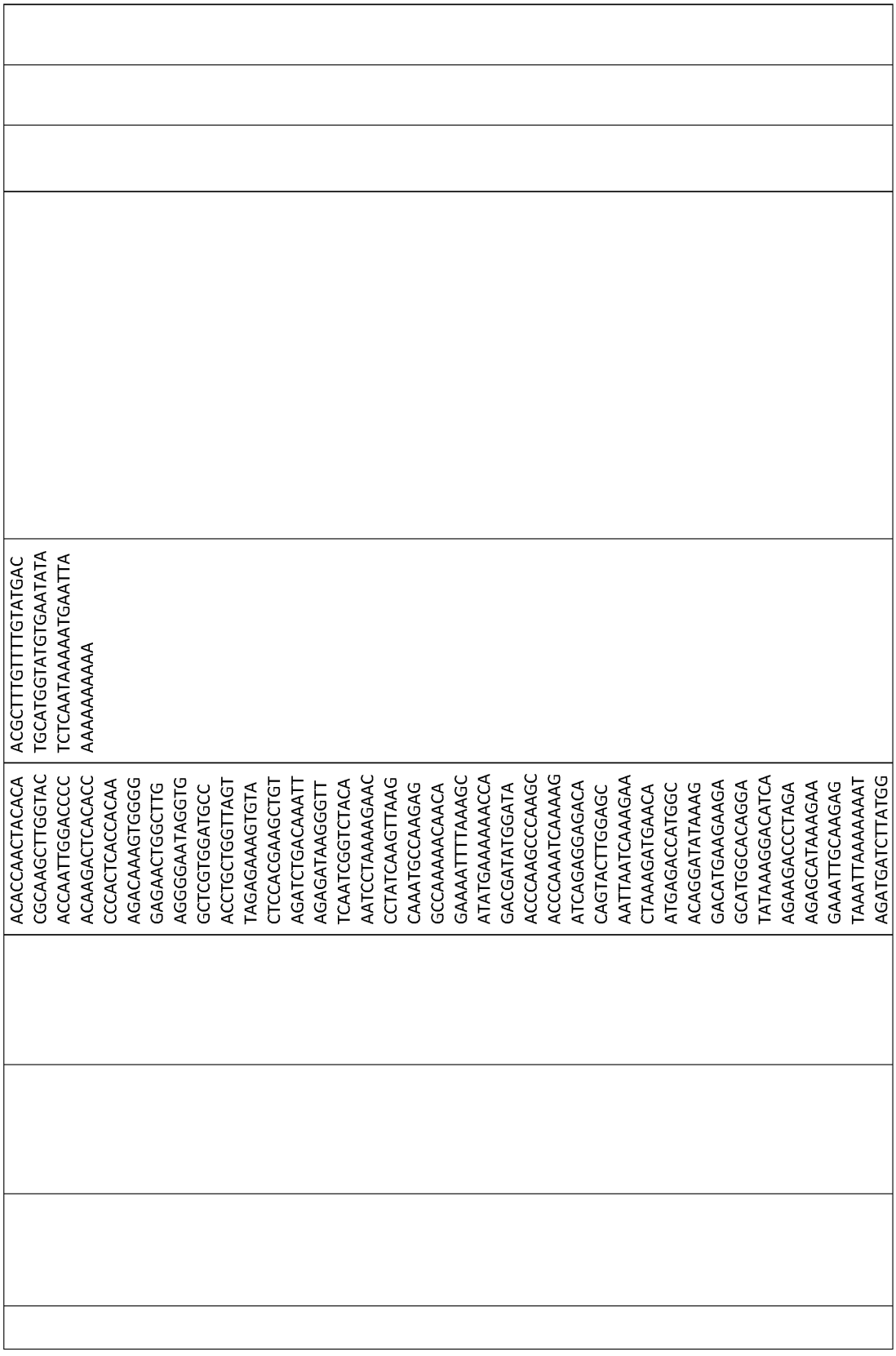
					QYPAVDVTGDRSKVRCCKEQYCIGTWNVRS MNQGKLEVVKQEMARVNVDILGISELKWTG MGEFNSDDHYIYYCGQESLRRNGVAIMVNK RVRNAVLGCNLKNDRMISVRFQGKPFNITVI QVYAPTSNAEEAEVERFYEDLQDLLELTPKKD VLFIIGDWNAKVGSQETPGVTGKFGLGIRNE AGQRLIEFCQENALVITNTLFQQHKRRLYTWT SPDGQHRNQIDYILCSQRWRSSIQSAKTRPG ADCGSDHELLIAKFRLKLKKVGKTTRPFRYDLN QIPYDYTVEVRNRFKGLDLIDRVPDELWNEV RDIVQETGIKTIPMEKKCKKAKWLSGEALQIA VKRREVKSKGEKERYKHLNAEFQRIARRDKKA FLSDQCKEIEENNRMGKTRDLFKKIRDTKGTF HAKMGSIKDRNGMDLTEAEDIKKRWQEYTE ELYKKDLHDPDNHDGVITDLEPDILECEVKWA LESITTNKASGGDGIPVELFQILKDDAVKVLHS ICQQIWKTQQWPQDWKRSVFIPIPKKGNAK ECSNYRTIALISHASKVMLKILQARLQQYVNR ELPDVQAGFRKGRGTRDQIANIRWIMEKARE FQKNIYFCFIDYAKAFDCVDHNKLWKILKEM GIPDHLTCLLRNLYAGQEATVRTGHGTTDWF QIGKGVRQGCILSPCLFNLYAEYIMRNAGLEE AQAGIKIAGRNINNLRYADDTTLMAESEEELK SLLMKVKVESEKVGLKLNIQKTKIMASGPITS WEIDGETVETVSDFIFLGSKITADGDCSHEIKR RLLLGRKVMTNLDSIFKSRDITLPTKVRLVKA MVFPVVMYGCESWTVKKAERRRIDAFELWC WRRLLRVPWTARRSNQSILKEISPGISLEGM MLKLKLQYFGHLMRRVDSLEKTLMLGGIGGR RRRGRQRMRWLDGITDSMDVSLSELRELVM DREAWRAAIHGVAKSRTRLSD		
			AGGAGCTACCCCAC GTCCGAGGTCAGG GGCAGAAGCCGGG AGGACCCCATGCCC GAAGGGCGGCGGC CAAGAGGAGTTAC CCCACGTCCGAGGT CAGGGGCAGCGGC CGAGAGTGCCAGA CTGCGACGGCGCA GGAACGGCCGAGA GGAGCTACCCCAC GTCCGAGGTCAGG GGCGGCGGCCGAG AGGAGATACCCAG CGTCCGAGGTCAG GGGCGGCGACGAG AGGAGTTACCCCG CGTCCGAGGTCAG GGGCGGCGGCCGG GAGGAGCTACCCC ACGCCCCTAAGCCC GAGGCCAGGGGCG GCGGCCGGGAGGA GCAACCCCACGCCC GAGGCCAGGGGCG GCGGCCGGGAGGA CCAACCCCACGTCC AAGGAGCCGTGGC TGCGCGGGCGCAG GAGGGCCTAGAGG AGCTATCCCACGTT GAAGGTCAGGAAG GGCGGCGGTGAGG AGATACCCCTCGTC CAAGGTAAGGAGC AGTGGCTGCGCTTT GCTGGAGCAGCCG			

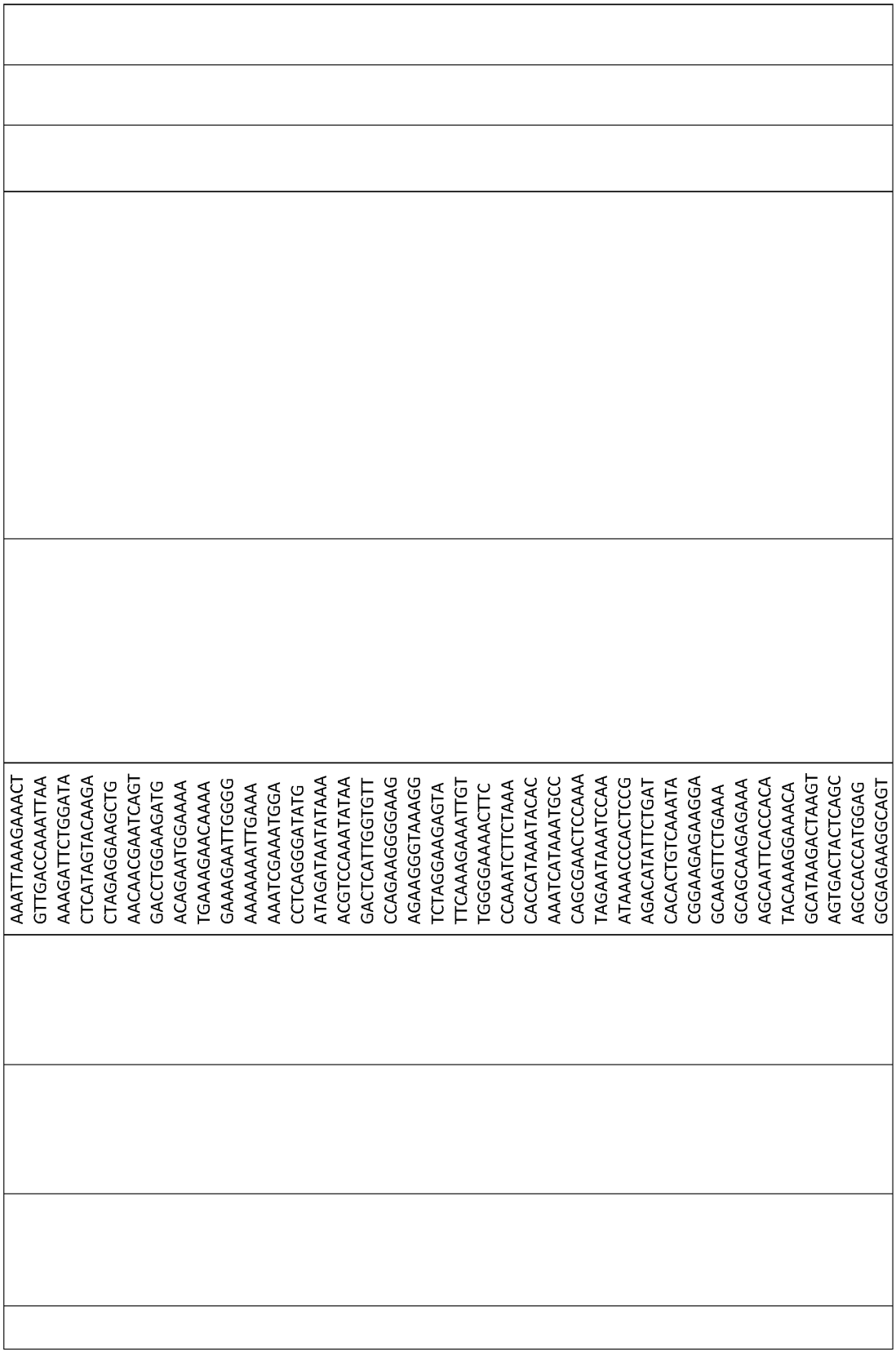
			TGAAGAGATACCCC ACGCCCAAGGTAA GAGAAACCCAAGT AAGACGGTAGGTG TTGCAAGAGGGCA TCAGAGGGCAGAC ACACTGAAACCATA CTCACAGAAAACTA GTCAATCTAATCAC ACTAGGACCACAG CCTTGTCTAACTCA					
	L1-1_Cho	GCA_000164785.2	Choloepus hoffmanni	GAGGCGGGGCAAG ATGGCGGAGTGGT GAGGTGTATGTTTT AGTTACTCCTCCAG GGCAGTAGGTAGA AAGCCAGGAACTG CGTGGACTGGACA CCGCAGAGCAATTT GACTTTGGGCATAC TTCATACAACACTC ATGAACACGTGGA ACAGCTGAGATCA GCGAAATCTGTAA GTTTTTGTGGCCAG GGGACCCCGCGCCC CTCCCTGCCAGGCT CAGTCCCGTGGGA GGAGGGGCCGTCA GTGCTGGGAACGA GGAGGGAGAACTG CAGTTGCAGCTCTT ATTGGAAACTCATT CTACTGATCCAAAC TCCAAACATAGATA GACTGAGACCAGA CACCAGAGAATCTG GGAGCAGCCAGCC	CAATAGTACAGACAGCCC AATTATAAAATGGGCAAA AGATATGAAAAGACAGTT CTCTGAAGAGGAAATACA AATGGCCAAGAAACACAT GAAAAAATGTTCAGCTTCA CTAGCTATTAGAGAGATG CAAATTAAGACCACAATG AGATACCATCTCACACCAA TTAGAATGGCTGCCATTAA ACAAACAGGAAACTACAA ATGCTGGAGGGGATGTGG AGAAATTGGAACTCTTATT CATTGTTGGTGGGACTGT ATAATGGTTCAGCCACTCT GGAAGTCAGTCTGGCAGT TCCTTAGAAAACTAGATAT AGAGTTACCCTTCGATCCA GCGATTGCACTTCTCGGTA TATACCCGGAAGATCTGA AAGCAGTGACACGAACAG ATATCTGCACGCCAATGTT CATAGCAGCATTATTCACA ATTGCCAAGAGATGGAAA CAACCCAAATGTCCTTCAA CAGATGAGTGGATAAATA AAATGTGGTATATACACAC	MADSRNAFTVITLNVNGLNSPIKRYRFAEWIK KNEPSICCIQETHLRHRDTKKLKVKGWKKIFH ASYSQKKAGVAILISDKIDFKCKDVMRDKEGH YILIKGAIQQEEITIINVYAPNQGATKYMRQTL AKLKEAIDVSTIIVGDFNTSLSPIDRSTRQKTNK EIENLNNLINELDLTDIYRTLHPKSPGYTFFSSA HGTFSRIDHMLGHKTSLNKFKKIEIIQSTFSD HNGIQLEVNNHQRLRKFTNTWRLNNTLLNN QWVKEEIAREIAKYIETNENENTTYQNLWDA AKAVLRGKFIALNAYIKKEERAKIKELMDQLKK LENEQQTNPKPSRRKEITRIKAEINDIENKKTIE RINNTKSWFFEKINKIDKPLARLTKSKREKTHI NKIMNEKGDITADPEEIKKIIRGYYEQLYANKL DNVEEMDNFLETYEQPRLTREEIEDLNQPITS KEIQSVIKKLPTNKCPGPDGFTGEFYQTFQKEL TPILLKLFQNIEENGTLPNSFYEANINLIPKPGK DATKKENYRPISLMNIDAKILNKILANRIQRHI KKIIHHDQVGFIPGMQGWFNIRKSINVLQHI NKSKEKNQMIISIDAEKAFDKIQHPFLIKTLQK VGIEGNFLNMIKSIYEKPTASIVLNGERLKAFP LRSGTRQGCPLSPLLFNIVLEVLARAIRQDKEIK GIQIGKEEVKLSLFADDMILYLENPEKSTIQLLE LINKFSKVAGYKINAHKSVMFLYARNEQTEET LKKKIPFSIATKKIKYLGINLTKDVKDLYKENYIT LLKEIEGDLKRWKNIPCSWIGRLNVIKMSILPK LIYRFNAIPIKIPTTYFADLEKLVIKFIWKGKMP RIAKDTLKKKNEVGGLTLPDFEAYYKATVVKT	cd01650	512	774

				CAGCAGAGAGGAG ATAGGGATAGCAA AAACAGCAAAGAA AAACCCAAAAATAA AAGCAGAGACTTTT TGGAGTTCTGGTG AACATAGAAAGGG GAAGGGCAGAGCT CAGGCAGAGTCAG GCTCATATGCAAAT CCAGAGGGCGAGT TTCCTTGTCTGAAA AGCTGGTTTCTCTG CTCTCTGGATCGTT CCTTAATGGCCCTA ATCTCCTTGTCTCTT AGCATTTCAATAGC CCATTAGATCTGCA AGGAGGGCCCTTC TTTTTCTTCTTTTTTT TTTTTTTAATCTTTT TCTCTTTTTCTAAAA CAATTACTCTAAGA AGCCCAATACAGA AAGCCTCAAAGACT TGCAATTTGGGCA GGTCAAGACAAGA GCAGAACTAAGAG AGCTCTGAGACACA AGGCAATAAGTCC AGTGGCTGAGAAA ATTCACTAAACACC ACAACTTCCCAAGA AAAGGGGGGCGTC CCCTCACAGCCATC TTCCTGGTGGGCA GGGAACACTCCTG CCCGGCGCCAGCC	GATGGAATACTACGCGGC AGTAAGAAGGAACGATGT CGTGAAACATATGACAAC ATGGATGAACCTTGAAGA CATAATGCTGAGCGAAAT AAGCCAGGCACAAAAAGA GAAATATTATATGCTACCA CTAATGTGAACATTGAAAA ATGTAAAACAAATGGTTTA TAATGTAGAATGTAGGGG AACTAGCGATAGAGAGCA ATTAAGGAAGGGGGAACG ATAATCCAATAAGAACAG ATAAGCTATCGTGGGTAA ATTTAACGTTCTGGGAATG CCCAGGAATGACTATGGT CTGTTAATTTCTGATGGGT ATAGTAGGAACAAGTTCA CAGAAATGTTGCTATATTA GGTTACTTTCTTGGGGTAG AGTAGGAACATGTTGGAA GTAAAGTAGTTATCTTAGG TTAGTTGTCTTTTTCTTACT CCCTTGTTATGAGTTTGTT TGAAATGTTTTTTTATTGT ATGTTTTTTTTAAAATTTTT TTTATATAGTTGATTTAAA AAAAAAGTTAATTTAAAAA AAAAAAAAACAATGAAAA AAATATGCAGAGCCCCCTT GAGGAGCTGGTGGAGAAT GCAGGGGTATTGGGCTTC CCCACCTCGATGGTTGCTG ATGTGCTCACAGACATAG GGGACTGGTGGTTTGATG GGTTGAGCCCTCTACCACA GGACTTGCCCTTGGGAAG ACTGTTGCTGCAAAGGAG	AWYWHKDRHIDQWNRIENSEIDPQIYGRLIF DKAPKVTELGHNSLFNKWGWESWISISKRM KEDPYLTPYTKINSKWIKDLNIKDSTIKLLEDNV GKHLQDLVLGGHFLDFTPKAQATKEKIDKWE LLKLRSFCTSKEFVKKVKRQPTQWEKIFGNHV SDKRLISCIYKQRNPTTQ			



		CCACAGCCCAGAG	AGGCTAGGCCTCCCTATAA			
		CTGCCCCAGACAAC	TTGTGCCTAAGAGCCTCCT			
		CCAGTGTGACGGG	CCCGAATGCCTCTTTGTTG			
		AAGTGTTTCCAACA	CTCAGATGTGGCCCTCTCT			
		ACACACGCACACAC	CTCTAGCTAAGCCAACTTG			
		CACAAAATTGGGC	GCAGGTGAAATCACTGCC			
		ATGGACATTAGCCT	CTCCCCCCTACGTGGGATC			
		TTCCTGCACCCTCA	AGACACCCAGGGGAGTGA			
		GCTGATTGTCCCAG	ATCTCCCTGGCAACGTGGA			
		AGTTGGGAAGGTG	ATATGACTCCCGGGGAGG			
		GAGCAGTGTGAAA	AATGTAGACCTGGCATCGT			
		AGGGGGAAATTAA	GGGATGGAGAACATCTTC			
		CACGCCCCATTCAG	TTGACCAAAAGGGGGGATG			
		CCATCCTTTCAGCA	TGAAAGGAAATGAAATAA			
		GACTGGGAACCTC	GCTTCAGTGGCAGAGAGA			
		CCTACACAGCCCTG	TTCCAAAAGGAGCCGAGA			
		CAGCCCAGAACCTC	GGTCACTCTGGTGGGCAC			
		CCTGGGGGGACGGG	TCTTACGCACAATATAGAC			
		CACTCACCTGTGAC	AACCCTTTTTAGGTTCTAA			
		ATAGCACAGTCATC	TGAATTGGAGTAGCTAGC			
		CCTCAACAGAGGA	AGTAAATACCTGAAACTAT			
		CCTGGGGGTGCAC	CAAACTACAACCCAGAACC			
		GGCCTGGAAGAGG	CATGAATCTTGAAGACGAT			
		GACCCACTCGCAAG	TGTATAAAAATGTAGCTTA			
		TCTCAGGGGCCATA	TGAGGGGTGACAATGTGA			
		CGCCAATACCAAG	TTGGGAAAGCCATATGGA			
		GACTTGTGGGTCA	CCACACTCCCCTTTGTCTA			
		GTGGCAGAGACAA	GTTTATGGATGGATGAGT			
		ACTGTGGCAGGAC	AGAAAAATGGGGGAAGG			
		TGAACTGAAGGAT	AAACAAAAAAACAAACAA			
		TAGACTATTGCAGC	CAAAGGCACCCAGTGTTCT			
		AGCTTTAAATCTCC	TTTTTACTTTAATTGCTCTT			
		AGGAACACCAGGG	TTTCACTTTAATTATTATTC			
		AGATTTGATTGTTA	TTGTTATTTTTGTGTGTGT			
		GAGCCGCCCCCCCT	GGTAATGAAGGTGTCAGG			
		CCCTGACCGCCCAG	GATTGATTTTGGTGATGAA			
		ACACACGCCCCATA	TGCACAACTATGTAATGGT			
		TACAGGGCGGGCA	ACTGTGAACAATTGAATGT			


ACACCAACTACACA
CGCAAGCTTGGTAC
ACCAATTGGACCCC
ACAAGACTCACACC
CCCACTCACCACAA
AGACAAAGTGGGGG
GAGAACTGGCTTG
AGGGGAATAGGTG
GCTCGTGGATGCC
ACCTGCTGGTTAGT
TAGAGAAAGTGTA
CTCCACGAAGCTGT
AGATCTGACAAATT
AGAGATAAGGGTT
TCAATCGGTCTACA
AATCCTAAAAGAAC
CCTATCAAGTTAAG
CAAATGCCAAGAG
GCCAAAAACAACA
GAAAATTTTAAAGC
ATATGAAAAAACCA
GACGATATGGATA
ACCCAAGCCCAAGC
ACCCAAATCAAAAG
ATCAGAGGAGACA
CAGTACTTGGAGC
AATTAATCAAAGAA
CTAAAGATGAACA
ATGAGACCATGGC
ACAGGATATAAAG
GACATGAAGAAGA
GCATGGCACAGGA
TATAAAGGACATCA
AGAAGACCCTAGA
AGAGCATAAAGAA
GAAATTGCAAGAG
TAAATTAAAAAAAT
AGATGATCTTATGG
ACGCTTTGTTTTTGTATGAC
TGCATGGTATGTGAATATA
TCTCAATAAAAATGAATTA
AAAAAAAAAA


AAATTAAAGAAACT
GTTGACCAAATTAA
AAAGATTCTGGATA
CTCATAGTACAAGA
CTAGAGGAAGCTG
AACAACGAATCAGT
GACCTGGAAGATG
ACAGAATGGAAAA
TGAAAGAACAAAA
GAAAGAATTGGGG
AAAAAAATTGAAA
AAATCGAAATGGA
CCTCAGGGATATG
ATAGATAATATAAA
ACGTCCAAATATAA
GACTCATTGGTGTT
CCAGAAGGGGAAG
AGAAGGGTAAAGG
TCTAGGAAGAGTA
TTCAAAGAAATTGT
TGGGGAAAACTTC
CCAAATCTTCTAAA
CACCATAAATACAC
AAATCATAAATGCC
CAGCGAACTCCAAA
TAGAATAAATCCAA
ATAAACCCACTCCG
AGACATATTCTGAT
CACACTGTCAAATA
CGGAAGAGAAGGA
GCAAGTTCTGAAA
GCAGCAAGAGAAA
AGCAATTCACCACA
TACAAAGGAAACA
GCATAAGACTAAGT
AGTGACTACTCAGC
AGCCACCATGGAG
GCGAGAAGGCAGT

								689
								431
							cd01650	
						MTTSSWCRLGSLDASLGELRVQQLQTGLTTL FGFNVLVTFDNVHYKTSGASAPVPTSTQERL VGLTCEECGKWCKSKAGLVAHHRVHDNDSV GTNMVAQLACADCSRLFPTKIGLSQHRRHA HPTQHNADKLSRVKHSGARWSQQESQSLLR LANNLYPSCETQTALFARLEQYFPGRSAISIKT RLRVLNWQAQQDESSSGGPDQTIGQIAAYSS EADDYSVWFKQTVDCAVSLLESHADSSLASV DLLAFARGLQSGIMTPEQVLSLLDLHASRTFP HTWKTVSRRRRQLAHRMPVNRKQIRRANYA AIQTLYHQRRKDAASAVLDGSWKDLYKGNC GLPPDAEQYWKQVLSAPKHVDSRPSRVVVP SDWSLIEPITGEEVGRTVRSMGNSSPGLDKLT PRMLRRFNANVLAGYFNLLLLAGGCPPHLCR	
					TGAATGTACATTCTTGCCA TTGAATCTCACACCAGACC CAGTATGACGGACACTTTG TGCCTGATGTGCGAGTCTT GACTGTGCTAGCGCTTACC GCGCCTTGAAGAGCATTC AGCATTGTTTGTCCTTTCTT CGGTTGTTAGACTTTACAC GCATGTTTCCTTACCAAAA TTCTTACGTACGTTGGGAT TCATCCTATCTGACTGGAA CTGTTGGTTGCATGACTTC GAATGAGACATTTCTTTCT TTTATCTCATATTCTCCTGT	GGCATGACATATTT AAAATTCTGAGTAG AGAAAAATTGCCA ACCAAGAATTCTTT ATCCAGCAAAGCTC TCCTTCAAATTTGA GGGAGAGCTTAAA TTTTTCACAGACAA ACAAATGCTGAGA GAATTTGCTAACAA GAGACCTGCCCTAC TTGAGATACTAAAG GGAGCCCTACAGA CAGAGAAACAAAG AAAGGAGAGAGAG ATACATGGAGAAA GGTTCAGTACTAAA GAGATTCAGTATG GGTACATTAAAGG ACATTAATAGAGA GAGGGAAAAAATA TATATCTGACAAAC ATAAACCAAAGGA TAAG	TAGTCGGCGAGCT GA	Clonorchis sinensis				
	R2NS-1_CSi	GCA_003604175.1						

				AACCCTTTCGCATTCATCG CTTGCATTCACTTTTTATGT CTGTGACCATGCTCTTCAA AAATAAACGATA	TAACATACTAATACACTTC TAATATCCAAATCCGTTAA AGGATTTTTAGGCTGCATT AATTAGGTAAACCGGAAC CGGGAACACTTCCCATAAC ACCCGATGTACTTGCTACA TCGTTAGAAGAATGGCAT CTACGCTCATATTAGTCTG TTTCTCTCTTATTCCGAGGT CACCGTAGCCACCAGATCC AGTCTGTATCCAAGTCAGA GGGTCACTGCAGTCACCC GGATCCAGTATGTATCCAG CCCAGATGGTGGATCAGC ACCTAGAAAGGACCTCTAC AGCCCTGAAAGACAGCGG AGACCAGGACAACTAGAG	ARITLVPKVPNPTSPDQLRPISVSSILVRCFHKV LADRWSRRLQLPSLQFAFLHRDGCLEATSLLH ALLRHSSATASNLSLAFVDISKAFDSVSHDTIV RSAEAFGAPSPLVRYIAQSYENAVAVFPSSEV HCHRGVRQGDPLSPLLFIMAMDEVLGLSMP QLGYQFHDTLVDGFAFADDWVVCAESQARL KEKLEAAAVELGRAGMKINARKTKAMVICGD RKHRATAVSVEPFCFAEELITPLGPTDTVTYLG IPFTFKGKGVFNHRQHLLKLLDEVTRAPLKPH QRMEITRNYLIPRLTYSLVLGQVHRNTLKRLD NYIRQSIRGWLRLPKDTPISYIHAGKQHGGLG IPSLSATIPMQRRVRMEKLLSTQCRVLRNVVN DSAFGKVVRDLSLPIRVHGSCVNTKEELVAA WGDSLHNSVDGRGLRELVASPLSNRWLVFP ERVFPRIFIRGIQLRCNLLRTRVRSARHGHGG QTILCRGNCGQPESLVHILQSCWITHDARCA RHNRVARELAKRLRRLGYTVFEELRAPTSTSFI KPDLIAVRERRATVIDVSIVSDGRGVTVWNEK KQKYGADEFSLAIISALRAIGCDVDFLVHQPM IISYRGICFPQSAKAVIGLGLSKVTVSDLCLLAIV GSLRTYDTFMRGTWR	MFVLATVYRPPGHHTDFIKEFADFTSELVLAA DKVLIVGDFNIHVDNEKDALGSAFIDILNSIGV RQHVSGPTHCRNHTLDLILSHGIDVDGVEIM QPSDDISDHYLVLCKLHIAKTVNSTSCYKYGRT ITSTTKDCFVSNLPDVSQFLSISKTSEQLDDVT ETMDSLFSSILNTVAPLRLRKVKENSLTPWCN EHTRTLKRAARKMERSWRKTKLEVFRIAWRE SNLSYRKALKTAKSDYFSSLLEENKHNPRYLFN TVAKLTKNKASTSVDISQHHSSNDFMNYFTS KVDTIRDKIVTMQPSATVSHQTVHYRSPEEQ FHSFSTIGEEELYKLVKSSKPTTCMLDPIPSKLL KEVLPEVIDPLLTIINSSLLLGYVPKTFKLAVIKP LIKKPQLDPKELVNYRPISNLPFLSKILEKVVSS QLYSFLEKNGICEDFQSGFRPYHSTETALLRVT NDLLLSSDRGCISLLVLLDLSAAFDTIDHTILLH RLEHFVGINGSALAWFKSYLYDRHQFVAVNE EVSYRSQVQYGVPQGSVLGPLLFTLYMLPLG	cd01650	385	648
		L2-2_CCa	GCA_004011595.1	Cyprinus carpio	CAAAAGGTAAAGG GGGAGGTGTTGCT TCAATTTATAACAA TGTTTTCAGGATTT CTCAGAGGGCAGG CTTCAAGTATAACT CGTTTGAAGTATTG GTGCTTAATATAAC ATTATCCAGAGAAA CAAATGTTAATGAT AAATCCCCTGTG					

					CCCCAGAGACAGATCCCCTGTAAAGACCTTGTCTCAGACGACCACCGGGACAAGACCACAGGAAACAGATGATTCTTCTGCACAATCTGACTTTGCTGCAGCCTGGAATTGAACTGCTGGTTTCGTCTGGTCAGAGGAGAACTGGCCCCCCGACTGAGCCTGGTTTCTCCCAAGGTTTTTTTCTCCATTCTGTCACCGATGGAGTTTTGGTTCCTTGCCGCTGTCGCCTCTGGCTTGCTTAGTTGGGGACACTTCATTTACAGCGATATCGTTGACTTGATTGCAAATTATTGCACAGATACTATTTAAACTGAACTGAGCTGCATGATGACATCACTGAATTCAATGATGAACTGCCTTTAACTGTCATTTTGCATTATTGACACACTGTTTTCCTAATTAATGTTGTTCAGTTGCTTTGACGCAATCTTTTTTGTTTAAAGCGCTATATAAATAAAGATGACTTGACTTGACTTGAC	DIIRKHGVSFHCYADDTQLYISSRPGETYQFEKLTECIVDIKNWMTSNFLLLNSEKTEVLIIGPKSSACNDLERCLSLDGCSVNSSSSVRNLGVLFDSNLSLESHVSSICKTAFFHLKNISKLRPMLSMSNAEMLIHAFMTSRLDYCNALLGGCSARLVNKLQLVQNAAARVLTRTRKYDHISPVLSTLHWLPIKHRIDFKILLITYKALNGLAPQYLNELLLHYSPPRPLRSQNSGNLIIPRISKSTAGGRSFSYLAPKLWNNLPNIVREADTLLQFKSRLKTHLFNLAYT			
	L2-2_DRe	GCA_008692375.1	Danio rerio	GCAGGAGAAGCACTTTAGCAGCATCTAGAACAGCAGCCTGTAAGTACATTTAAGATTTGTTTCAGTTGTTGTGTATGGTTTGAGGACTTGTTTCCAGCTGTTTGTGTAAAGTTGTAGGACATTTAAACTTGCTTTCAGTTGTGTATAATACTAGAAGTTGTTTAGC	TAATCTGCAATTGCCTCTTTGAATATCACACTAATTGTACACAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAACTACTAACACTTCCCTTCTTAGACTTTACAGACCTGAAACTTGCCTTTAGTACTTATTCATTGTTGCTCTTAGTTGTGTAAATTGCTTCCTTGTCCTCATTTGTAAGTCGCTTTGGATAAAAGCGTCTGCT	MCFLIPVVTNTRKTREVRCRRNPHNLRSIHVSTISQLSLSVGLWNCQSAVNKADFITSIATYSDYNLMALTETWLRPEDTATHATLSANFSFSHTPRQTGRGGGTGLLISKEWKFTLIPSLPTISSFEFHAVTIIHPFYINVVVIYRPPGKLGHFLDELDVLLSSFSNFATPLLVLGDFNIYVDKPQAADFQTLLASFDLKRAPTSATHKSGNQLDLIYTRHCFTDQTIVTPLQISDHFLLSLNIHITPEPPHTPTLVTFRRNLRSLSPNRLSTIVSDSLPPSRKLTALDSNSATNTLCSTLASCLDRLCPLASRPARASPPAPWLSDALREHRSKLRAAERIWRKTKNPAHLLTYQTLLSSFSAEVTSAKQTYYRLKINNATNPRLLFKTF	cd01650	508	768

		CACTGTTTCCTTGG TTACTATAAGAGCT TGTGTAGCGAACG CAGACGCGGTTCG CGTCGTCCGCCTCA GTTTCGGCCCTTGT TTTGACTCGGGAG GCGTGTCCAAACG CCAGGTAACCACTA TATAGTGAGCACG GTAGCATTAGTCG GCAGGAGAAGCAC TTTAGCAGCATCTA GAACAGCAGCCTG TAAGTACATTTAAG ATTTGTTTCAGTTG TTGTGTATGGTTTG AGGACTTGTTTCCA GCTGTTTGTGTAAA GTTGTAGGACATTT AAACTTGCTTTCAG TTGTGTATAATACT AGAAGTTGTTTAGC CACTGTTTCCTTGG TTACTATAAGAGCT TGTGTAGCGAACG CAGACGCGGTTCG CGTCGTCCGCCTCA GTTTCGGCCCTTGT TTTGACTCGGGAG GCGTGTCCAAACG CCAGGTAACCACTA TATAGTGAGCACG GTAGCATTAGTCG GCAGGAGAAGCAC TTTAGCAGCATCTA GAACAGCAGCCTG TAAGTACATTTAAG	AAATGACTAAATGTAAAT GTAAAATGT	SSLLYPPPPPASSTLTTDDFATFFCTKTAKISAQ FAAPTTNTQDTTPTPHTLTSFSQLSESEVSKLV LSSHATTCPLDPIPSHLLQAISPAVIPTLTHIINT SLDSGLFPTTFKQARVTPLLKKPNLDHTLLENY RPVSLLPFMAKILEKVVFNQVLDFLTQNNLM DNKQSGFKKGHSTETALLSVVEDLRLAKADSK SSVLILLDLSAAFDTVNHQILLSTLESLGVAGTV IQWFRSYLSDRSFRVSWRGEVSNLQHLNTGV PQGSVLGPLLFSIYTSSLGPVIQRHGFSYHCYA DDTQLYLSFHPDDPSVPARISACLLDISHWM KDHHLQLNLAKTEMLVVSANPTLHHNFSIQ MDGATITASKMVKSLGVTIDDQLNFSDHISR TARSCRFALYNIRKIRPFLSEHAAQLLVQALVL SKLDYCNSLLAGLPANSIKPLQLLQNAAARVV FNEPKRAHVTPLLVRLHWLPVAARIKFKTLMF AYKVTSGLAPSYLHSLLQIYVPSRNLRSVNERR LVVPSQRGKKSLSRTLTLNLPSWWNELPNCIR TAESLAIFKKRLKTQLFSLHFTS			

	RTE-1_DCa	GCF_001605985.2	Dendrobium catenatum	GTTGCCAGGTCCCA AGTCCTGGATGAG AAAAGTGGAGGGT TGCGTTAGGTAGTC GACAGCCAGCGTA	ATTTGTTTCAGTTG TTGTGTATGGTTTG AGGACTTGTTTCCA GCTGTTTGTGTAAA GTTGTAGGACATTT AAACTTGCTTTCAG TTGTGTATAATACT AGAAGTTGTTTAGC CACTGTTTCCTTGG TTACTATAAGAGCT TGTGTAGCGAACG CAGACGCGGTTCG CGTCGTCCGCCTCA GTTTCGGCCCTTGT TTTGACTCTCGAGG CGTGTCCAGCTGAA TTCAATCAGCTAGT GCTTTGGGGTTATA TAAACAACTAGTTC ACCGCGGCAGCGG TCGCGGCAGCCTC GTGTGAAGACCGA CGAGGGTAAAGAC CATCGACTCTACCT GCGCGACTCCACC GAGCAAAGACACC GACAAAGCACTTG AGTACTTTACTGTA TTGTTTTACTTTACA CTTATTTTTGTTGT CAGTGCACTTTTAT T	TAGTGGGATAAAGGCTTG ATGTTGTTGTTGTTG	MNPLRNCVNARASPGSDALYGLNTVTSVQG PLHRRPGAVLNRQGFPHQGRMWAKKLVHE TKIRIGTWNIGSLTGKLMELIDIMIHRKLNILCI QETKWVGEKAKEIDSLGFKLWYSGRSRTRNG VGVIIDKSLKDNVIKINRIGDRIIALKIVVDDKIV NIISAYAPQVGLDMSSKLNFWNDLEELIQNIP	cd01650	581	829

	Daphne-2_LMi	GCA_000516895.1	Locusta migratoria	AAACTTGTCGAACA ATAATGAC	TTTCACCAGAATAT TGGGGGACTAAAG AATAAAATAGATG AGCTTTTGGTTTGT TTAGATGATATAGA AACTGATAATGTAA TAGATATACTATGC CTGTCTGAGCATCA CATTGTCACTGGTA TGGAAAATGTGAA TGTCAGTGGATATA	TAAAGGAATTTCTGAATG GCAACTCCTTCTACACTTT AGATGAATTTTTAGATAAA AATTAGTAATTTAGGAAAA AAATTAAAAATAAAAAAA AAATAAAATAAAATATTAA AATATTGAGTGTCAAGTAA TAATATGTATAGACACATT TCACTAAACTGACACGTTC CACATCATTACGAAATGTC GTATTTATGATCTATGGAA	LEERVFIGGDLNGHVGREVDDFKSIHGGHGF GTTNEGGKTILDFCLAYDFLIANTLFKKRNEHL ITYKNGYSTTQIDYFLIRKLNRLDCKDCKVIPSE NLCTQHRLLVLDFKIKGHKIKNKVKQFSRIRW GHLKNINKNIFESKILDLKILSVHEDANIMWEK VASDIKNIAKDVLGVSKGKAPNSKNSWWWN PNVQEKIKNKKICYKNWFNFKNSDNWEIYKV ARKDSKKAICDAKTLAFSDLYNRLDTKEGERY MYKLAKQRDRSSKDLVQVKCIKDEDQKILFLE DDIQQRWMRYFAKLFNEDCENKIDYSDLSTL NENKDFCFYRRISKTEVREALKKMKGGKAVG PDDIPIEVWKCLGDEGISWLTKLFNTILKTKK MPDEWRMSVLIPIFKNKGDAQDCANYRGIK LMCHTLKLWERVMEHRLRRDTNVSQNQFG FMPGRSTMEAIHLLRGLMEKYRENKEDLHMI FIDLEKAYDKVPREVLWRVLEKKGVNNAYIQII KDMYAGAITCVQTQGGLTKYFPISVGLHQGS ALSPYLFALVMDVLTRHLQEDVPWCMLFAD DILLVDKTREGVEGKLELWRSTLESKGFRLSRS KTEYMECNFSSNRPSEGIVTLGDDQVINKSTRF RYLGSIVQSDGEIDGDIISRIQVGWLKWRNAS GLLCDRNVPLKLKGKFYKMVVRPAMLYGAEC WPLKEKHNTKLSVAEMRMLRWMSGFTLRD RIRNEHIREKVGVAPVEDKIRESRLRWFGHIK RRPSDDPVRRVEVLDLTYVKKGRGRPKKTWL ENIRNDLSLLDLNENLTFNRTQWRKRIHVAD PT	MCIMSEISTSNNKIKTIWNIITRETGQPRTQE DSITIKVNEKITNKKSEVENTFNNYFLNVVENI GSKCSLEEARQLMEQGVPRQFDTMEIPPTSQ SEIRKLINSLKSKSSHGIDGISSRVLKACSQEICR ILSHICNSSLEQGIFPDRLKYAIVKPLHKKGDTL DVNNYRPISLLTALSKILEKVMYSRVASHINKN KVLTKCQFGFQKGFSTENAIYIFTNEILNALDN RKSPVGIFCDLSKAFDCVNHGILLDKLKYCGM SGKVHKWFKSYLTGRVQKVDISNSHNIQKQP SPQTGELSSMGCHRVRYWDLCCSYTKTAESS NWGTIKHGVPQGSVLGPLLFLIYINDLPFYIHK	cd01650	153	443

			AATTAGCTGCAAAT GTAAGTAGAGATA ATAAGGTAAGAGG AGGAGTTGCCATAT ATGTCAAAAGCTAT CACAGTGTAAAAA ATTTAGAAACAAAA AAGTTTTGTGTAGA GCAACATATTGAAG CATGTGCCTGTCAG CTTAAACTCAATGA TGATTCTCTTATAA TTGTAACAGTATAT AGGTCCCCTTCAGG AAACTTTCATTATTT CTGGAAAACTTGG ATGCCTTGTTGTGC TATCTGTCAGATAG GGGGAAGCAAATT ATTATTTGTGGTGA TTTCAATGTTGATT CACTGAAAGAGTG TAATAGGAAGAAT GACCTGGAAGTCTT ACTTGGCTCTTTCA ATTTGACATCTGTT ATTGATTTTCCTACT CGGGTAGTAAAGG ACAGCAGTACATTG ATAGATAATTTTTT ATAGACCAAGATA AGTTTAACAACATA AATGCTTGTCCTGT TGAGAATGGTCTTT CTGATCATGGTGCA CAGCTAGTAACAGT ATATGACATAGCTC CTTTCAGTAATTCG	CAAGTATTAATCTAATCTA ATCTAA	DAKLVLFADDTSIAITPESQELTDEIVNEVFQRI IRWFSVNGLSLNFEKTQYIQFHTINGITPLINID FNQKSVAKVEYSKFLGLCIDEGLNWKKHIEDL LKRLSSATYAIRVIANFGDVQLSKLAYYAYFHS LLSYGIIFWGNSSFSKKVFIAQKRVIRIIAGAHP RSSCRHLFKDLGIFTIYIYYSLMKFVINNPTEFK CNSSVHSYNTRRKGDLHYSRLNLAILSQKGVN YSATKVFGHLPSSIKILTDSQLAFKKN		

	RTE-2_LMi	GCA_000516895.1	Locusta migratoria	CCACCCCCATGAAC CACAGACATTGCCG AGGTGGGGTGGCT TGCGTGCCTCAACG ATACAGATAGCCGT ACCGTAGGTGCAA CCACAGCGGAGGG GTATCTGTGGAGA GGCCAGACAAACG TGTGGTTCCTGAAG AGGGGCAGCAGCC TTTTCAGTAGTTAC	TGAAGACCACAACA	MMASSWVKYSGGKIVPHSDLRAGTTQEDVV IKKNKTGILRIGAWNVLTLNRVGKLENLKREM DRLKLDMVGISEVRWQDEQDFWSGEYRVIN TKSVGGNAGVGLIMNKKIGMRVSYYEQHSE RIIVVKIDTKPTPTTVVQVYMPTSSADDEDIEE MYEEIKEIIQKVKGDENLIVMGDWNSVVGKG REGNVVGEYGLGKRNERGDRLVEFCAEHNL MITNTWFKHHERRLYTWKRPGDTGRYQIDFI MVRNRFRNQVLNCKTYPGADIDSDHNLLVM NCRLKLKRLQKGRKVRRWNLDKLKEPEVVES FRGSIRQGLAETGESNTVEEEWVALRDRVVK AAEDHIGRRTRPNRNPWITQDILNLIDERRKY	cd01650	548	807

								420
							148	
						cd01650		
			AGGGGCAACAGTC TGGATGATTGACTG AGCTGGCCTTGTAA CATCAGCTAACATG GCTTTGCCATGCAG TTACTGCGAACGGC TGAAAGCAAGGGG AAACTACAGTCGTT ACTTTTCCCGAGGA CATGCAGCTCTACT GTATGAATATATG			KNSASETGKREYRRLKNEIDRKCRLAKQEWLE DKCRTVEECITRGKIDTAYRKIKEIFGDKKSSC MNIKNSDGKPVLSKEGKAERWKEYIEGLYKG NKLEGSILEMEDEVEEDEIGDTILREEFDRALK DLSRNKAPGVDDIPSELLISLGEPVMKKLFHLV CKMYETGEIPSDFKKNIVIPIPKKASADKCENY RTISLISHGCKILTRIIYRRRMEKLVESDLGEDQF GFRRNVGTREAILTLRLILEDRLRKGKPTFLAF VDLEKAFDNVDWNTLFQILKVAGVKYRERRV IFNLYKNQTAVIRVEGHERDAEVGKGVRQGC SLSPMLFNLYIEQAVKESKEKFGEGIKVQGEEI KTLRFADDIVILSETAKDLEEQLNGMDSVLKG RYKMNINKSKTKVMECSVRKSGDAEGIRLGN ETIKVVDEFCYLGSKITSDGRSKEDIKCRMAM GRKAFLKKRNLLTSNINLNVRKSFLKAFVWSV ALYGSETWTIRSSDKKMIEAFEMWCYRRMM KIRWVDHVTNEEVLNRIGEKRSFWHILTKRR DRLVGHILRHQGIVNLALEGSVWGKNRRGRP RDEYSKQIQKDVGCRTYSQMKRLAQDRVAW RAASNQSAD		
					TGAAGACCACAACAACAA CA	MAGVLEAKCKSVEECIKGKSRYRLQENKRTFG TRKQLYEYQELDGKPVLSKEGKAERWKEYIEQ LYKGNELRDDVVGREEEVDQDEIGDTILRDEF DRALKDLGRNKAPGVDDIPSELLKSLGEPAM TRLFELVCRMYETGEIPSDFKKNVIIPIPKKSGA DRCENYRTISLISHGCKILTRIIYRRRMEKLVEAD LGEDQFGFRRNVGTREAILTLRLILEDRLRKGK PTFLAFVDLEKAFDNVDWNTLFQILKSAGVKY RERRVIYNLYKSQTAVIRLEGHEREASVGKGV RQGCSLSPMLFNLYMSRRPFSDVIQFVHEQA VKESKEKFGIGIKVQGEEIKTLRFADDIVILSET AKDLEEQLNGMDSVLKRAYKMNINKSKTKIM ECGLRKSGDAEGIRLGNETMKVVDEFCYLGS KITCDGRSTEDIKCRMAMGRKAFLKKRNLLTS NINLSVRKSFLKAFVWSVALYGSETWTISSAD RKRIEAFEMWCYRRMLKIRWVDRITNEEVLN RIGEKRNFWHTLTKRRDRLVGHILRHRGIVNL	
	RTE-1B_LMi	GCA_000516895.1	Locusta migratoria	CAAAATAGTGAAG CAAGACCTACAAAA AATCCTTGGATATC ACAGAAATTTAATT AGAAAATATAAAA ATGCAGTATGAAG GCAAAATGAGTAC AGACGACTAAAAA TGAGATTGACAGG AAGTCAAAATGGCT AA					

	Daphne-1_LMi	GCA_000516895.1	Locusta migratoria	TAACTAAAACCTCT AGGACAGCAATTG ATAATATTTTTATA GACAAATCAAAGG AACAAAATCATATC ATAAAACCAATAAT AAATGGCCTATCAG ATCATGAT	TGATGCATGCACTCATACT ATATTAGTATTATTATATT ATCAGCAAAAAAAACCTA AAAATTATGTATCTGTATT TTATTATAAGTGTATGTAT CTGTATTTATATTATTAGT ATTATCTATTATTCAGCTT GTAAATTTTGTTTCTTGAC ATGTTCCACATCCACGAGG ATCTCCTCAACATGGATCT ACGGAACGAATAAACTAA TCTAATCTAATCTA	ALEGSVWGKNRRGRPRDAYSKQIQVDVGCR SYEEMKRLAQDRVGWRAASNQSPD	cd01650	288	549
	CR1-2B_MUn	GCF_012275295.1	Melopsittacus undulatus	C	TAGGTTACCCTGCTCTTGC AGGGGGGTTGGACTAGAT GATCTTTTGAGGTCCCTTC CAACCCTTGGGATTCTGTG ATTCTGTGA	MQIPVAPEGVRTKEEFALVDEDWVRDQLCN LDIYKSMGPDGMHPRVLRELAEVIARPLSIIFG KSWETGEVPEDWKMANVTPVYKKGKKEDP GNYRPVSLTSVPGKVMEQLILDSISRHIKDEG VIKNSQHGFMRGKSCLTNLIAFYEEVTRWRD DGRAVDVVFLDFSKAFDTVSHSILIDKLRKCGL DDQVVRWIKNWLKGRRQRVVVNGAESSWR SVTSGVPQGSVLGPVLFNIFINDLDEGTECTLS KFADDTKLGGVADTPEGCVAIQRDLDRLESW AGRNLMKFNKGKCRVLHLGKNNPMYQYRL GVDLLERSEGERDLGVLVEGRLRELGLFSLEKR RLRGDLISVYKYVKGGCQDDGARLFSVISSDR	cd01650	77	327

	RTE-1_MD	GCA_000002295.1	Monodelphis domestica	GGGTGTATGGGGT GCTCAGGGAGGGG TAGTATCTCTGGTA TGGAGGGCTTGTC GTGCCCTCCTAGGG CAGCTCTCCAGCCT CTGACCCCCACCTG ACACCCAGCTCTCA CTTGTGGCTCCCAG TAGCTGCTAGCATG TGGCAGCGGCCAC ACCCCGGGCAACG GCTTCGACAGGCC GGCTAAACCTTGTG AGGGTAGCCATCG GGTCATCGACCCCT GGTGAACCAGGGC TTTGCTCACCCAGC ATGTGAAGACTGCT TCGGCTGAACAGA CGGAAGAAACCAA TAAGAAGGTTCAAC GGCTGAGAGGGCG ACGCAGCAAAGCA CTGTGGAGTGCTTA GGGCGTGTTGGAG CACAAAGGACAAC ACGGCCATCCAATG CAGCTGAGGAAGT CTCCAGATGTAACA ATTTTTCGTGCCAC TGGACCCAGGCTTC CAACGCCGAGAGA GTGGGACTGTCTCT GTGCATCGGCTTTT CCACTTAAATCTCT	TGAAACTGCACAAAGACA ATAGTCATTCTCGATCACC GAGAGACTACCACT	TRGNGCKLEHRRFHVNIRKNFFTVRVTEHW NRLPRGVVESPTLEIFKARLDKFLCDVL MDSTAHPNQGRGLEKVSQTLPALQTPGQHT AAGGSSPLSGRNQRKNTKKLLLGAWNIRTLL DRENTPRPERRTALIGKELARYNIDIAALSETR LPEEGSLSEPTTGYTFFWKGRASNEDRIHGVG LAIKTSLLKQLPDLPVGISERLMKIRLPLSKDRY ATIISAYAPTLTSTEETIEQFYSDLSAVLHSVPT NDKLILLGDFNARVGQDHERWKGVLGKHGV GKMNNNGLLLLSKCSEFELTITNTVFRMANK YKTTWMHPRSKQWHLIDYIIVRRRDIQDVKIT RAMRGAECWTDHRLVRATLQMRIAPRHPK RAQTVRAFYNVSRLRDPSYLQTFQSCLDDKLS AKGPLTGSSTEKWNQFRDAVKETSKAVLGPK QRNHQDWFDENNTAIEDLLSKKNKAFMEW QNNPNSAPKKDRFKSLQATAQREIRKMQDR WWEKKAEEIQRFADMKNYKQFFSALKTVYG PLKPTTTPLLSSDGDTLIKDKKGISNRWKEHFS QLLNRPSSVDQSALDQIPQNRTIEQLDVPPSI EEVQKAIKQMSAGKAPGKDGIPTEVYKALNG KALQAFHIVLTSIWEEEDMPPELRDASIVALY KNKGSRAACDNYRGISLLSTAGKILARVILNRL LSSVSEQNLPESQCGFRPDRSTIDMVFTVRQ MQEKCLEQNLSLYIVFIDLTKAFDTVNRDALW VILSKLGCPAKFVKLIQLFHVDMTGEVLSGGE TSDRFNISNGVKQGCVLAPVLFNLFFTQVLRH AVMDLDLGVYIKYRLDGSLFDLRRLTAKTKTT ERLILEALFADDCALMAHQENHLQTIVDRFST ATKLFGLTISLSKTEVLFQPAPGRPTNQPCITID GTQLSNVNTFKYLGSTIANDGSLDHEINARIQ KASQALGRLRCKVLQHRGVSTATKLKVVNAV VLSSLLYGCETWTLYRKHMKQLEQFHQRSLR SIMRIRWQDRITNQEVLDRANSTSIEVMVLK TQLRWSGHVIRMDPQRIPRQVFYGELSAGLR KQGRPKKRFKDQLKSNLKWAGITPKQLELAA SDRSSWRTHINHAATTFEDERRRRLAAARER RHQATTAPPVTTGVPCPMCHKLCASAFGLQS HMRVHRR	cd01650	588	870

								486
							228	
						cd01650		
					MTNRIAYGGTSKKTVKEAANKVLGKKKKPKS KPWFDEECELWFERRKKAKLDSLQNRSDRTV EEYSNVRKQTSAIYRNKKREYQKNLIRRIETNS KENNPREMYRGINAIRKGFRSRAQLMKDEN GDLVTNDNELLSLWKNYFDKLLNVHENSEEL GDEIHTAELHVEEPSYQEVEAAIKKLKNNKAA GNDSIPAELLKYGGVELTFKIYKLVCAIWKNET IPENWKESIIIPIFKKGDKTDCNNYRGISLLATC YKVLSNVIQARLTPFAEDIVGDYQCGFRRNRS TSDQMFTIRQLLEKKWEFCETIHQLFIDFKKA YDSIKRSKMYQILVLLGVPKKLVRLIQICLNGST GKVRVGGNVSEPFMIRDGLKQGDGLSTVLF NLTLEYAVRKMQVSQLGATLNGTTQILGYAD DLDILGDCRETVARNAEIILKAVEYTGLEVSES KTKYMIVDKLGICRGEEDLRVGNFTFEKVSEF RYLGTTINDRNEINVEINKRLHSGNACFYAVS NLLKSRLLSKNVKIRIYRTIILPVVLYGCETWAL TKQADNRFRVFENKVLRKIYGPKKDEETGEW RRLHNDELHNLYASPNINRIIKSRRLGWAGHV ARMGDDRTAARVMKGRPMVTRPLGRPRRR		
				AAAAAA				
			TTCACGCACAAGTA TCTTTGTGCACACT CATCTATCCTAACC CCGTCCACCCTCTT CAAGACCTGCGGC GATGGGGGAGTGG CGACGCAACAGGT GGAGGTGACCACT GGCAGTTGTAGTC ACGATCCTGCACGT AGGCGGCCCACGG ACCAGTGGTCGCTC GGCCCTGTGGGCA GCAGGGACGTTCG GCAGCATCCTGGG CGACTGAGCAGCC CTCTCTAG	ATGGCGGAGTACT AGCATCGGGTAAC TTCACGTACCTGTA TGGGAGTGGGGGA TCTCTAGGCACCGG ATTTCTCGTAAGCA AAAGCATCATACAT TCGGTTAAAAGTTT CAAATCCGTCAATG ATAGGCTCTCGTAC ATCATTATCGAAGG TGAATGGTATAGAT ATGTATTTATTAAT GTACACTGTCCTAC GGAGGACAAAGAA GAAGAAGCTAAGG ATCTCTATTACGAA ACTTTAGAGCAGGT AATCGACCAGTTCG CGTCTTACGACACA AGAATAGTATTAG	Nasonia vitripennis	GCF_009193385.2	RTE-1_NVi			

	RTE-2_NS	GCF_0003936 55.1	Nicotiana sylvestris	TACTGTCAGGGCTA CACCAGTGGTAGC GGTGACAGCCCTTT TTGGTTTTTGGGTC	GCGATTTCAATGCT AAAATAGGTAGGG AGGAAATGTTTAG GCCTACTATAGGG AAGGAAAGCCTGC ACGAAGCCAGCAA TGATAACGGTATTA GGGTCATAAATTTC GCGGCGGCAAAAG ATCTTATAATCAAA ACTACGTGTTTTAA GCACAAGGACATA CACAAGGCAACGT GGACATCGCCGGA CGGGGCCACACAG AACCAAATTGATCA TTTCCTCATTGAAA AAAGACGTAAGGG CTTACAGAGGGGC AGATAGCGACTCG GACCACTTCCTAGT AGTAGCCAAATTAA GAGCTAGACTAGT AGCGAATCAAAAT AGTAAGCGAGCAA ACAAGGTAGAAAG CTTCGATATTGAGA AGCTACGAGATAG AACAGAGCGAATT AGGTACCAGATAG AAATTAATAACAGG TTTCAGGCACTTGA AGAAGCAAACACA TCGCCGGAGGGGA	TAGGGGAAGTTGTGAGAT ATTTCTACAGCGCACTAGA GTGAGACTAGCCAGTTAG GAGTTAGTCTTAGGATGCT	WEDNVKADLVEIGRVGVDRRGASWVGLTQ DRAAWKACVDEAMNFRVPNAT	MTIKLVVGECTLNVVSAYAPQVGLDEEIKRRF WEGLDDIVRSIPPSERLFIGGDFNGHIGSSAG GYTEVHGGFGFGERNGGGTSLLDFAKAFDLV IANSSFPKREEHLVTYQSSVAKTQIDYLLLRRC	cd01 650	438	687

				GTGGTATTTTCTGT GTTTTCAGGGAAAT TCTCGAAGTTGTTG GAGACAAGTTGGG CGCACGAGCACTA GCAAAGGAGAGCA ACCTGGACGAGCG TCAGCAAAGGGCG GTGGGAAGCGGTG CGTGAGCGTACAA CTACATCGAGTACA GCAAATCAACACA GGTTTGGTTCTCGG GAGTTGGTACCTCC CGTTCTACTGTTTC CCTTTTGGTGTTAG CTAAATTGATGTTA CTTTAGTTCACAAT AGTTCAATCATTCA ACTAGTGTACTTGT AGTCACTTGTTTCC TTCTAGTTTGATTC GTTGTCCGTTGTGC ACTAGCGAGTCTTC TTTAGATTTGTCGT AGGTGTAGTGGCT GCAGTCTGGGATG ATAGATTAAGGGC ATGTCCTCGGGTG GGGCAGAGTGTGG GGCCGGGGTCAGG GTGTGGGACGGGA GGCAGGGGAGGTA GAGGGGGCAAGG GAGCCTATAGGTT GAGAATCGGGTCA TGGAACATAGGTTC GCTAACGAGTAAG	ACTGGTCAGCTACTGATGC AGGGCTTTATCTGTTGGAT ATTAGTATACCTTCCATCTT TTTCGTATTTCCTATATTTC TTATATTGCTGTTATTTTGT TTTTATGTATTTTATGTTAT GTTATGTATTTTATGTTATT TTATTATGAGTCTATTGAT AGTACTAATATATCGTCTC TTGTTGCTTTCTTGAGCCG AGGGTCTCCTGGAAACAG CCTCTCTGCCCCTCGGGGT AGGGGTAAGGTCTGCGTA CATATTACCCTCCCCAGAC CCCACTTGTGGGATTATAC TGGGTTGTTGTTGTTGT	DRRLCEDCKVIPGETLATQHRLLVMDIGIMIR RKKRSVRGRPRIRWGALTKDKAQELEGRLSA MGAWRSSGDANTMWSTTADCIRKAAREVL GISSGRTGGHKGDWWWNAVVQGKVEAKK AAYLRLVGSTGEEEKRANSERYKVARKEAKM AVTEAKTAAFARLYEELGNKGGEKKLFRLAKV RERTARDLDQVRCIKDDDDKVLMGDDQIKR RWQTYFHKLLNEEGDQDIVLGELRNADSPHE LSYCRDIEVDEVMEAMRKMRRGRATGPDEI PVELWRCVGRAGLEWLTGLFNVIFKTNRMP EEWRWSTMVPLYKNKGDIQSCNNYRGIKLLS HTMKVWERVVEMRVRRTVSISDNQFGFMP GRSTTEAIHLIRRMVEQYRDKKKDLHMVFIDL EKAYDRVPREVLWSCLEVKGVPVAYIRVIKD MYDGAKTRVRTVGGDSEHFPVVTGLHQGSA FSPFLFALVMDAITHHIQGEVPWCMLFADDI VLIDETRGGVNERLEVWRHALESKGFRLSRTK TEYLECKFGAEPTEAGVEVRLDSQVIPKRGSF KYLGSVIQGTGEIDEDVTHRIGVGWMKWRL ATGVLCDKKVPPLLKGKFYRAVVRPAMLYGT ECWPVKNSHIQKMKVAEMRMLRWMCGHT RMDKIRNEDIREKVGVAPMEDKMREARLR WFGHIQRRSTDAPVRRCERLAVVGTRRGRG RPKKYWGEVIRQDMARLRITEDMALDRELW RSSIKVVG			

							879
							628
						cd01650	
					MPLVFQMPAGLHMVHRFQWSFQTKTDSEE GRGHPLLKELARENLMNSSETLSDTLLENFM NSPETLSDTVPEDEPLRLEGTQNMTGKELPH QSRTNLNDMDKVKLSGPPFADVARLKLRRNS CKHPLIIGTWNVRSMNLGKLEVVKNEMECIKI DILGISELKWTGFGHFESDNHMIYYAGNDKL KRNGVAFIITKNVSRSVLKYNAVSDRIISICLQG RPVNTTIIQIYAPTTKAKDEEIEDFYQLLQSEID QTCNQDVLIITGDWNAKVGNKEESPVVGKF GLGDRNNAGDHMIEFCKTNDLFIANTFFQQ HKWWLYTWTSPDGVHRNQIDYICGKRRWK SSISSAKTRPGADCGTDHQLLICKFKLKLKTIKT SPREPKYDLEYIPPEFRDHLKNRFDALNTNDQ KPEELWDDIKNIIHEESKRSLKRQERKKRPKW MSEETLKLALEWRVAKANGRNDKIKELNRRF QRAAREDKANYYNEMCKDLELENQKGRTCS AFLKLKELRKKFKPQVAILKDSMGKILNDAGNI KRRWKEYTESLYQKELVDVHPFQEVAYDQEP MVLKEEVEAALKELAKNKAPGIDGIPIEMFQQ MDAPLEALTQLCQEIWKTATWPTDWKRSIF VPIPKKGDPTECGNYQTISLISHASKILLKIIQK WLQQYIDRELPEMQAGFRRGCGTRDIIADVR		
					TAACAACAACAACAA		
			TCTATAGAGTTGGC GAAGATCCTTCAGA AGAGGAAGATTAA TATAGCGTGTGTCC AGGAGACTAGGTG GGTTGGATCGAGG GCGAAAAACGCGG ATGGGTATAAGTT GTGGTACTCTGGA GTCCTGAGGGGTA AGAATGGAGTGGG TATCCTGGTAGATA GCCATCTTAGAGA GTCGGTGGTAGAG GTCAGGCGGGTGA ATGATAGACTA	ACTCGTCTGTCAGT TTGTCATACTGTGG TGGCTTGCGTGTTG CTGTGAGGCTGGA GGCT			
		Orycteropus afer					
	GCA_004365145.1						
RTE-1_OAf							

					WILAQSKEYQKDVYLCFIDYAKAFDCVDHNK LWITLRRMGIPEHLIVLMRNLYIDQEAVIQTE QGDTAWFKIKKGVCQGCILSPYLFNLYAEQII REAGLYEEEHGIRIGGRLINNLRYADDTTLLAE SEDDLKHLLMKIEDYSLQYGLHLNIKKTKILTT GPISSITINGEKIEVVKDFILLGSTINTNGSSSQ DIRRCIALGKSAAKDLFQVLKSKDVSLKTKVCL TQAMVFSIASYACESWTMNKEDRRRVDAFE LWCWRRMLKVPWTAKRTNKSVLEEVQPECS LEARMARLHLTYFGHVVRRDHSLEKDVMLG KVEGQRKRGRPSTRWIDTVAATMGSSIARIV RLAQDQAVFHSAVHRVAMSQRRLDGT	TGATGGGAAAAGCCGAAA GAGAAAGAAGAA			MRIIHNTLDIPDRSRPGLTTTATDAVNLQGVG GNLTTVGRRKRGGRRVRGQREKGKGRNIGL RIGTLNVGTMTGKGRELADMMERRKVDVLC VQETRWKGSKARSIGGGYKLFYHGVDRKRN GVGVILKGEFVNSVLEVKRVSDRMMSLKLEIE GVMVNVVSGYAPQVGCELEVKERFWSELDE VIESIPRGERVVIGADFNGHVGEGNRGDEEV MGRFGVKERNLEGQMVVDFAKRMEMAVV NTYFQKREEHRVTYRSGGRSTQVDYILCRRG HLREVNDCKVVVGESVARQHRMVVCKMTLE VRKKKRGKTEKKTKWWKLQNEETCEEFRQKL RQVLGGQDELPDDWETTAEIIRETGRKVLGV SSGKRKDGKDTWWWNEEVQECVQRKRLAK RKWDVERTEEGRQEYKEAQRRVKREVAKAK QKAYDELYDRLDTKEGEKDLYRLARQRDRDG KDVQQIRVIKDRDGKVLTTQESVQKRWKEYF EELMNEENDRERREEDVVVVEQEVAEIGKDE VRKALKRMKSGKAVGPDDVPVEVWKCLGET AVEFLTSLFNRILESEKMPEEWRRSVLVPIFKN KGDTQNCSNYRGIKLMSHTMKLWERVVEAR LRKKVEICEQQYGFMPRKSTTDAIFALRMLIM EKYRDGQKELHCVFVDLEKAYDRVPREELWY CMRSSGVAEKYVRVVQDMYERSMTVVRCA VGQTEEFKVEVGLHQGSALSPFLFAMLMDRL TDEVRQESPWTMMFADDIVICSESREQVEE QLERWRFALERRGMKVSRSKTEYMCLNERD	cd01650	572	818
	RTE-2_OL	GCA_004348135.1	Oryzias latipes	ATCTAGCCACAGG GGTTGCAAGCCAG TTACTTCTGTGCCG GTCCCAAGCCCGG ATAAATAGAGAGG GTTGCGTCAGGAA GGGCATCCGGCGT AAAAAATTGCCAAA ATAACC					

					QGRSVRLQGAEVKKVQEFKYLGSTVQCDGEC GKEVKRRVQAGWSGWRKVSGVLCDRRVPA RLKGKVYKTVVRPALLYGLETVAVRQRQEAE MEVAEMKMLRFSLGVTRLDRIRNEYIRGTAH VACVSDKVREARLRWFGHVQRRDSGYIGRR MLEMELPGRRARGRPKRRYMDVLTEDMKL ANVRVEDVHDRVRWKRMIRCGDP	cd01650	613	868
	RTE-1_OL	GCA_004348135.1	Oryzias latipes	TGTGAACCGCCACC TTAACGTGGTGGG GGAGTTTGAGAGC TCGGGTGATCCCA GGAGCTAAGCTGT CTGGGGCTTAAGC CCCTGGTAGGGTCT CCCATGGCAGACA GGTCCTGGGTGAC GAGCCAGACAAAG AGCGGTTCAGAAA CCCCTG	TAAGCGGAGGAAGATGGA TGGATGGATGGA	MKRNNEDPITSPGLASPGPHSGARPRVGARR RAPGGRGSSHGPRSGTARRSDVGPLSHGPTT RRRPQKGPVQCGLGGSRGRVPRRPKPRSWN LAFGTWNVTSLGGKEPELVREVERYRLDIVGL TSTHSLGSETQSLERGWTLYHSGVAQGERRR AGVGLLVSPQLSRQVLEFTPVDERVMSLRLR VGDRSLTVVSAYGPNSSSEYQAFLVSLEGVLE SAPTGDSIVFLGDFNAHVGNDSGTWRGVIG RNAPPDLNPSGVLLLDFCARHSLSITNTMFEH KGVHQYTWHQDTLGRRSMIDFVVVSGDLRP CVLDTRVKRGAELSTDHYLVVSWIRWRGRRP ERLGRPKRTVRVCWERLAEPSVREVFNSHIRE NFKQVPREVGDIESEWTMFSTSIVSAAARSC GRKVSGACRGGNPRTRWWTPEVRDAVKLK KESYQAWLACGTPEAADRYRQSKRAAARTV AEAKTRSWEEFGEAMEKDYRLASKKFWQTIR RLRRGKQFSTGTVFSAGGELLTSTGDIVGRW KEYFEDLLNPTDTPSFEEAESEDFGVGLSITRA EVTEVVNELLGGKAPGVDEVRPEHLKSLDAV GVSWLTRLCSIAWRSGTVPLDWQTGVVVPL FKKGDRRVCSNHRGITLLSLPGKVYARVLERR VRPIVEPRIQEQQCGFRPGRGTVDQLYTLSRV LEGLWEFAHPVHMCFVDLEKAFDRVPRGIL WGVLWEYGVRGALLRAVRSLYDRSRSLVRIA GSKSDLFPVHVGLRQGCPLSPVLFIVFMDRIS RRSQGPEGIWFGDHRISSLLFADDVVLLASSN RDLQRALGRFAAECEAAGMRVSTAKSEAMV LDRRKVVCHLSVGGVSLPQVEEFKYLGVLFTS EGRSEREIDRRIGAASVVMRSLYRSVVVKREL SRKAKLSIYRSIYVPVLTYGHELWVMTERTRSR IQAAEMSFLRRVAGRSLRDRVRSSVTRGELG			

					VEPLLLHIERSQLRWLGHLVRMPPGRLPGEV FRACPTGRRPRGRPRTRWRDYVSRLAWERL GVPPEELEEVAGVREVWASLLSLLPPRPGPG			
	L2-6_OL	GCA_004348135.1	Oryzias latipes	AATTCACACATCAG TCACAAAACCTACG ACATGCACACATCA GATTTTACCCACAA CCATGAAGATGGC TCTTTTAAATGTCA GATCACTTTTAAAT AAGTCTTTTATCAT TAATGATCTGATTT AGACCATGATTTAA CCTGTTTGTTTTTAA CTGAAACCTGGCTT GGCTCTGCAGCTCT ACTGAGTTCTACCT TTTTAA	TGATTTTAACTGCTAATTT ATTCTATTAGTTTTAGTAT AGGCTATTTTATGTATTTA TTTTAATTTTATGCATCTTA TTTTCAATTTTATGTTTTTT TTTTCTTTTTATGTTTTTAA TTTTAGCATCTTATGATTTT AGCCCCAGTGTTTCTTGAG GGGATCTTTTCCTCCAGGG ATTGCCGGCTTGGTCAGC GGTTGTTGCTGCAGTTGTT GCCCTTGCTGGGGCGGCA GGGGTCTAATTTTACAGCT TCTTGGCTGTGAAGGGGA CCCTGGTGTTGTCCCGGTC TGGGTGGGCACTGTGGCG ATGTTCCGGTGGCCGGGC TGGCTCCTGGTATGAAAG GATTCCCAGATACTGTTTC CTCGCAGCAGAGCTAAGC CATACTTGTTTATTTATTTA TTTATTTATTTATTTATTTA AGTTACATAGTCATATTTC ATTGTACGTGGGAGCCAT GGGTCATGTGGTTTAATG GAGGGGAGTGGGAAGGG TGAAGGGTGGGTTGGGTT TTTATTGTAATGTTGTGTG TTTACTGTTTTTAGTGTTAA AGCGCTTCGAGCTGCACA AAGTGCATGAGAAGCGCT ATTTTATAAATTTGATTTG ATTTGATTTGA	MFFFSIRGGKKGGGTASIAHKTILSENILLSKYC SFEHHAFLFGDPPVLTVTVYRPPTPPSCFIEEF SEFLSSLHIRYNRILITGDFNLHVDKNDDIYACE FLNLLNSMDFQQHVTQPTHNRGHTLDLVITY GLSTGVSSVVDLAISDHYCVFFNVTSEIQQEA SVRTVRKRYLTPEVAASFIDIFKQTPAQILPAP CDFIVDHFNNRLQSSIDAVAPFKTKTLRTTPK MPWKTDDIKGLKQNCRRAERRYRKTKLIIHK DILKEQLKIYNNSVKRARTSYFSKIISENKNNPK FLFKTIDLINKDFNKSSMPSSNAACEDFADHF MGKINAIRSNISSMQYTDLNRPPALFLPEETL GSFVLVDAEMLGRVVSQVNTTTCLLDPIPTSL FKTFYGFFESELLNIVNCSLQTGVFPSACKTAV VRPLLKKSQLDPKNLDNYRPVSNLPFLSKINIE KSVATQLHEFLNAQKILEKYQSGFRKNHSTET ALLKIVNDLRLNYDSQKLSVLVLLDLSAAFDTV DHQILLDRLRSLVGLSGNVLKWFYSYLTDRHF YVSMDTCSSGIHEISCGVPQGSILGPILFNLYM LPLGEVIRRHGVDFHSYADDTQLYIAVSPDDL EPVNTLLNCILDIKSWMAENFLQLNQDKTEV LIIGSEDKREMILPHLHNFKPSQCVRNLGVLFD SELNFIPHIKNVVKTGFYHLKNIARVRPFLSLAS TEVLMHAFIFSRLDYCNALLSGLPKKSISNLQL LQNSAARVLTRTRGREHITPVLKSLHWLPVRF RIDFKILLMVYKCFYGLGPSYLNDILLKYEPSRT LRSSGAGLLVVPKVRTKTYGEASFCHYGPRL WNGLPENLRAAETVEVFKKRLKTHLFNLAFS	cd01 650	421	681
	L2-2_OL	GCA_004348135.1	Oryzias latipes		TGAAAGGTCCTATCTGCAC TTTTCTTTCCTTTCCTTCCTT	MSSFEYVSFKIELKQTPPLLCIVMYKPPQHSQ SFIDDFTEMLSVVCTDFDGLVITGDFNVHVD	cd01 650	387	648

								795
							536	
						cd01650		
				TATTTTTAAATCATTTTTCA ATTTTTTTCTTTTCTTTTAA AGTAAATTTTACGTTGATT ATTTCTATGATGTATTGTG ATTTTAATGCATTTTTCTGT TTTGTGAAGCACCTTGAAT TACTTTGTGTACGAATTGT GCTATACAAATAAACTTGC CTTGCCTTGCCTTGCCTT		NVNDRNAKELNAVLETFDLIQHVSCPTHSRG HTLDLLITKGVNIYNVSVVDVALSDHFCVFFDL SVAPKPPPGPAVVQRRQINESTRAQFVEMIN PENASGANVDEMLNSVTSSILNVLDAIAPMK VKLRKHRQKAPWRNDDLVRAQKQQCRRAE RKWRKSKLQVHYGMYKGELYAYNQILCRTRE RYFSEIIGSCSSNSRVLFTTVNRLTNPPTQLPIE LVCTPKCNEFAVFFNDKVQGIKKAINSTTHITI ERPPTHSKLTHFVPVTDQTVQETITRLSSSTCC LDVLPTRFLKSVLNSVLPSITQIVNMSLQSGIF PEALKTAVIKPLLKKSGLDPTVLNNYRPISNLPF LGKVLEKVVYQQLTDFLLLNNSFDVFQSGFRP HHSTETALIKVTNDIRLNTDDGKVSVLILLDLS AAFDTVDHGILLHRLQDWVGISGSALSWLKS YLEDRKYFVEIGSCVSDYMALTCGVPQGSILG PLLFNLYMLPLGQLIRSNNISYHNYADDTQIY VSLTTGEYGPVDSLSHCLQQISAWMQNNFL QLNSNKTEVIIFGPQKQREIVSSYLHSLSLKPS NQVRNLGVIMDSDLNFNSHIKSVTSAAFYHL KNIAKIKNIVSKPDLEILIHAFVSSRLDYCNGLLT GLSKRAVKQLQYIQNAAARVLTRTRKYDHISP VLRSLHWLPVPQRIDFKAALLVYKSLHGRAPK YISDMLVPYEPSRTLRTSGAGLLLIPRVRTKQG ESAFQYSAAKIWNSLPEVVRQASSVFMFKSR LKTFLFSRVYD	MLYTIFLVYGILCFFFVYFCIYLYITVYLCFLFFLF LLLCGDVESNPGPGRARGCRLLYCNIRGLHA NLAELDFVSRGVDVVCCSETLVAGRRHDAEL ALAGFQSPFRRLCGSGPGFRGMAVVVRSGFC AYRQSVHECACHEIIVVRVCGRLNNYYLFSLY RSPATDDSLYDCLLTAMASIQSTDPKAAFVFV GDVNAHHRDWLGSASPTDCHGVAALDFCTL SSCVQLVRGSTHIAGNCLDLVMTDVPDLMT VTVGSPIGSSDHSHLVVSLDLNQVVPVVDTR RVVFVKSRANWVAITRAVRTLPWRQIIHSED PVSELNDLTVSILERFVPKRTILVRSRDKPWFD DQCRLAFEAKQAAYRAWRRSRDRTLWQTYV DRRAEAKRVYEEAQRRLRQRSRESLLSIDHPH	
	CR1-1_PH	GCA_001587735.2	Parhyale hawaiensis	GCGTGGCCTGCGC GTTATCAGGCCGCC CCCATTGTCCGGCC ACCGGACGCCTCTT TGTTTGTATGAGCT GGCTTTGGGTTTGG GGCTCTGGGTAGC CCCTGAGTTGGAG GAAATGCCACTCTA AAGGCGGATGAGA CCTCTTTGGTCCA ACGGCATCCCTGA GAGGAGATGAACA	TAGCAGCGTTTTGGGCTTC CCGGTGTTGCAAGATGCTT CATTTGGGTTGTATTACCT GTCCGTGGGAGCTCAAGC GTGAAGTTTTAATAATAAT AATAA			

	RTE-1_PSi	GCF_000230535.1	Pelodiscus sinensis	GGCTGGTGCCATA GCTGGGTACTCAGT CTTCCGCAGTCTGC TCCTCTCTCTCTTCC ATCAGGTAGTACCA CCTGATGCTCTTTT CCTTTCCTTTCTCTT CTTCTTCTTTTCGGC CAAAGCCTGTACTG GTGATGGCAGGGT CTGGGGCGCCGTAG ATAGCTGGCCGCTT TCGTAGCTATCGTT ACTGTCGATTTGTT TTTGCTTTATT	TGGTACGCTAAGA GCACGGCAAGACA TGAAAGATGCCCT GGTCAACCACTGC GCCCAGCCCATCTC CAACCGTCTCGACT TTAGTCCTGCCATT GGAGTAAGATGGA ATCTGGGAAGAGA GAGTGAGGCTGAC TATGCGCACCTCTC CCTCACTTTAATCC AATTCACGCGCAAG TCTTGACATCTCGA CATCTCCTAAGTCC TGTGGCGACGGGC GAGCGACGAAGCA GCAGGTGTGGATA CACTGGGAGCTGT AGCCGCGGACCTG	TAGATGTCAAGGTCCTCTT CGAATACGATGGACGAAC AACAACAAC	RWWSELKGSVFGAEPSLPPLVGPGGGIITDPL ARAELLSAHFDGKQSRDVIALPHGCHPEPRLT SLAFRSGAIKVLLEGLDPYGGVDPVGMFPLFY KQLADVLAPKLAVIFRRLIRLGNFPRCWRTGN ITPIPKGPVSPYVANYRPITLTPILSKVFEKLIAG KLGRFAEVTGLLPAGQFAYRIGLGCCDALLSV SHHLQSALDCRSEARLVQLDFSAAFDRVNHR GLLYKLESFGVGGRVLSIIRDFVSERTQSVSVD GVLSASVGVVSGVPQGSVLGPLLFVLYTSDM FSSLENTLINYADDSTLMAVIPAPRLRDAVAQ SLNRDLSRISAWCSAWSMKLNASKTKSMIISR SRTLVPQHPQLEIDDTLLQESSSLEILGVVFDE KLTFEPHIRRLVSRASTKIGLLRKVNSVFGDSQ VARRCFYAFLLPVLEYCSPVWASAADTHLRLL DRLVSSASRLCADNDLVNLSHRRRVAELCMF YKVYNNERHWLYSSLPALKVFGRETRAACGA HSFTLEAVRCRTNQFCRCFVPWSAKVWNLLP ASAFGRIGLQAFKSAVNGFLLDFL	cd01650	488	770

	Poseidon-3_PM	GCF_010993605.1	Petromyzon marinus	ATATTGGTCGCGAA AGCCTACGTGTTAA ACTGACTAACCTCT ATGGGGAGGATAC TTTTCTTAAAACCA GAAAATTGGAGAA	CACACAGGCGGCT CAGGTCTAATGGTC GTCTTCTTGCTGGT GGAAGCAACAGCT GGATTCGGCGTCTA CCTGAGTGACTGA GCAGCCCTCTTTAG GATTGCACTGCTCA CCTCCGTAGTAAGA GGAAGGGGCTAGA AAAGGTGCCCTAA ACATTGTTTGCTTC ACAGAACCCTGGCC AGCTTACCGCGGCT GGAGGGGATCCCA TCTCATGCGGTCGA ACAATAAAATTTCG AACAAAATGCAAG GTGACACCACTCAC TATTGGCACATGGA ATGTACGCACGCTT CAGGACAACCCTTC AGCAGACCGCCCA GAAAGAAGAACAG CCCTGGTCGCCAGA GAGCTCACAAGATT TAACATCGATATTG CGGCACTGTGTGA AACTCGTCTAGCAA ATGAAGGTCAGCT C	TAATACACATTACATTTTG ATTTAAAGAATTACTAATC TACAACTCATTACTAATCA TAACCAATGACTGATTACT AATCGATTATATAATTACA TTACCCCCCCCTCCGCCTCT AATCACTGCACCTGAATGA	TNGVKQGCVLAPTLFSIMFSAMLSDAFQNSS LGIRLRYRTDGKLFNLRRLQAVTKIKETVLRDL LFADDCALNAGSEQEMQASMDKFAAACDN FGLLINTKKTEVMHQPAPKAQHQVPTITVKG QKLQAVDQFTYLGSTLSRSVTIDIEVNCRVAK ASSAFGRLLTKVWDRRGISLATKLKVYRAVVL TTLMYASETWTVYRRHARQLNHFHTICLRRL LRIRWQDKVPDTEVLSRAGLPSVYTLLMKAQ TRWAGHVARMPDHRIPKQLFYGELFQGKRS HGGQKKRYKDTLKASLKSLDIDTTSWETLAQ DRSTWHTLIYQGCQTSEARRTTIAQEKRALRK ARVAKAATVVPTHVCPTCGRTFRARIGLISHL RTHHEQSITC	MKLISSLAFLKRCRDSKVIPNAVKITHHIKSSKA NQILHKASELLLIERIQHNRSELSYISESLKQLHL DISNSVSMHYWNNIIDNITQQSADNVFKLTKT NQINKFSRLHNKQIPTENQKCKNKTVLNLTNT VLTGDMVSILSKGLNFAVTPRTIPVDEIIAEVE SAVRLLPHNEAEMTRHEVAKIENSKVPKVNI SSNEAKALRELRVNKDIQILAADKGRTTVVMS	cd00304	416	499

			GCTTCGCATTAAGCGG	GTTAAGTGATGACGTCACCCTGCTGCTCTATATAAAACCGGGAGCAAGCGAGGGCTTTACCATTCTGGCGAATTGCCCGATGATGCTGGAAGTATTTACCAGCGAAACGTCGTACTCAAACTGTAAATGTGGATTAACTCTTCAACACACCGGTGGGCTGAAGTAGTGAATTATTTGTCTTTTGTCAACGTGACACTTTTTTTGCTGATTGGCCGTGTCGGGTGGTGGAGGGTTACGACACATTTATAAATAAACGCGATCTAACCCAAAAACCTTTATTATGAATAATATTGGTCGCGAAAGCCTACGTGTTAAACT	RADYDNKIRNLLEESTYIKIKRDPTNTLERKTTAIIANSLIPKNEQYLIKPRASKPPRLYGLPKIHKIGIPLRPIVSTIGSATYQLAKYITKQIQPYIGKTSSFIKNSEHLVNVIKDIKLTPDDILVSFDVISLFTKVPVDESLKLLDTLVKNTGLQKDIPKLAGHCLNNTYFLWRGDFFKQKEGAAMGSPLSPAIANLYMENFEIRALASSTLKPSFWFRYVDDTLVVWPHENFEIRALASSTLKPSFWFRYVDDTLVVWPHGKDTLNDFLNHLNAQNPAIQFTGKDTLNDFLNHLNAQNPAIQFTGKITVRFGTVHKIQKLVSCTKDALPKLSTEGVYRINCSCGSCYIGQTGRNISVRLKEHQRNIKLSQTQLSAVAEHACQQGHSIDFDNVAVLSAAWKRIIK			
						cd00304	260	343
	Poseidon-3B_PM	GCF_010993605.1	Petromyzon marinus		MVDEIIAEVESAVRLLPHNEAEMTRHEVAKIIENSKVPKVNISSNEAKALRELRVNKDIQILAADKGRTTVVMSRADYDNKIRNLLEESTYIKIKRDPTNTLERKTTAIIANSLIPKNEQYLIKPRASKPPRLYGLPKIHKIGIPLRPIVSTIGSATYQLAKYITKQIQPYIGKTSSFIKNSEHLVNVIKDIKLTPDDILVSFDVISLFTKVPVDESLKLLDTLVKNTGLQKDIPKLAGHCLNNTYFLWRGDFFKQKEGAAMGSPLSPAIANLYMENFEIRALASSTLKPSFWFRYVDDTLVVWPHGKDTLNDFLNHLNAQNPAIQFTMEMEVNNTLPFLDILLGHGLNGKITHSVVRKQTHTDAYLNAESHHHPAQKSSLIKTLVHRAKTVSDGESLVREIKHVKDALRQNGYNNAQINRAILNAKSRSQSNTTESDQIVKTIKIPYISGVTDKLAATCAAQNITVRFGTVHKIQKLVSCTKDALPKLSTEGVYRINCSCGSCYIGQTGRNISVRLKEHQRNIKLSQTQLSAVAEHACQQGHSIDFDNVAVVCKTNYYWQRLIRESIEIERNKNNFNRSDSYRLSAAWKRIIK		

							826
						574	
					cd01650		
				MALMMVASAGRCALFCLVLSVNSVFCQDFR SSLTRDELLNIRETTPVDLIPTFLLPAAELVGILV NAALGFAQATKRRPRGKRSGALVRLRRRGIR TVLPSFFLSNVRSLYNKMDELQLLMVKNKDFL SASALCFTETWLCDLIPDTSLQLAGFHLVRSD RDIALSGKTKGGGICFYINSGWCVDVTVILQH CSPLLESIFITCKPFYSPREFASLILVGVYLPPCP QVKEAQRMLADQILSVERANPDSLVIVLGDF NKGNLSHELPKYIQFIKCPTREGNTLDHCYTT VSRAYHAVPRAALGHSDHAMVHLIPAYRQKL KLCKPAVRTSKQWTSEAVEDLRACLDCTDW DMFTTATNSLDELTEAVTSYISFCEDCCIPTRT RVNFNNNKPWFSAKLRRLRLEKEEAFRSGDR DRFKESKNRFSKAVREAKRLYSERLKHQFSTN DAASVWRGLRQITNYKPRAPHTTNDSRLAN DLNEFYCRFERQLDCPEPPLPTQEAPSPTNTT ETTISFQEGEVNRLFKRQNPRKAAGPDSVSPA TLRHCADQLSPVFTNIFNTSLETCHVPACLKSS TIIPVPKKLRPTGLNDYRPVALTSVVMKSFERL VLAHLKAITDPLLDPLQFAYRANRSVDDAVN MALHFTLQHLDAPASYARILFVDFSSAFNTIIP ALLQDKLFQLNVPYTTCKWITDFLSDRKQCVK			
				TAAAACCGGGAGCAAGCG AGGGCTTTACCATTCTGGC GAATTGCCCGATGATGCT GGAAGTATTTACCAGCGA AACGTCGTACTCAAACTGT AAATGTTGTGGATTAACTC TTCAACACACCGGTGGGCT GAAGTAGTGAATTATTTGT CTTTTGTCAACGTGACACT TTTTTTGCTGATTGGCCGT GTCGGGTGGTGGAGGGTT ACGACACATTTATAAATAA CGCGATCTAACCCAAAAAC CTTTATTATGAATAATATT GGTCGCGAAAGCCTACGT GTTAAA	TGTTTTTAACATTTGATCTA CATAACGACACCTTGCCGA TATTAACTATGTTACCCTA TTTGTATTTTTTATTTTTAA ATTATATCTTAAAGCTATT TTATATTCGATATTTATTTA ATTTTATATACTTCAGCCC CTTGGTGTTAGTTAGACTT TTAACCCTTATTATAGTAA GACTTGTTTAAAACTTATT TAAGATTTTAGATTCCTAT ATGTTGAATGTATGCACCA TCCTGCCAAAGTAAATTCC GTGTCGGTGTGAACTTAC ATGGCGAATAAAACCCTTT CTGATTCTGATTCTGA	AGATGGCGCCGAT GATGGCAGCCTCG GCCGGTTGGTGCG CCCTGTTTTGTCCA GTTTTGTCGTTTTG TCTTGAGAGGGGC CTGCTGTTCAACAC ACTCGTGCTGGTGA GAGAGGCCTGCTG TTCATCACACCCGT GCTGGTGAGAGAG GCCTGCTGTTGAG GGGCCTGCTGTTCA TCACACCCGTGCTG GTGAGAGAGGCCT GCTGTTCAGCACAC CCGTGTTGAGGGG CCTGCTGTTCAGCA CACCCGTGCTGGTG AGAGAGGCCTGCT GTTCAACACACCCG TGCTGGTGAGAGG	Petromyzon marinus				
		GCF_010993605.1					
	Rex1-2_PM						

				GGCCTGCTGTTCAG CACACCCGTGCTGG TGAGAGGGGCCTG CTGTTCAGCACACC CGTGCTGGTGAGA GAGGCCTGCTGTTC AGCACACCCGTGTT GAGGGGCCTGCTG TTCAGCACACCCGT GCTGGTGAGAGGG GCCTGCTGTTCATC ACACCCGTGTTGGT GAGAGGGGCCTGC TGTTCAGCACACCC GTGCTGGTGAGAG AGGCCTGCTGTTCA GCACACCCGTGTTG AGGGGCCTGCTGT TCAGCACACCCGTG CTGGTGAGAGGGG CCTGCTGTTGAGG GGCCTGCTGTTCAG CACACCCGTGCTGG TGAGAGAGGCCTG CTGTTCAGCACACC CGTGTTGGTGAGA GGGGCCTGCTGTTC AGCACACCCGTGCT GGTGAGAGGGGCC TGCTGTTCAACACA CCCGTGCTGGTGA GAGGGGCCTGCTG TTCAGCACACCCGT GCTGGTGAGAGGG GCCTGCTGTTCAAC ACACCCGTGCTGGT GAGAGGGGCCTGC TGTTCAACACACCC		LGTQVSGSRSISTGTPQGCVLSPLLFSLYTNSC TSNHLSVKLLKFADDTTLIGLISGGDESDYRW EADNLVTWCSQNNLELNALKTVEMVVDFRK NTAPLTPITLLDSPVKIVESFRFLGTILSRDLKW ELNISSLTKKAQQRMYFLRQLKKFNLPKTMM VQFYTAIIESILTSSITVWYAAATAKDKSRLQRI IHTAEKVIACNLPTLEDLHTTRSMRRARKIMA DFSHPGHSLFQLLPSGRRLRSIRTKTSRHKNSF FPSAAGLLNKAKGLH			

								819
							569	
						cd01650		
					MNSSQTLSDMVPESEPPRLEGTQKTTGEELP PQSRVNLNDVDGVRPLGSSFADVARLKMRR NSCKHPLIIGTWNVRSMNLGKLEIVKNEMECI NIDILGISELKWTGLGHFESDNHIVYYAGNEN LKRNGVAFIIKKNISRSVLKYNAVSDRIISIHLQ GRPVNTTIIQIYAPTTKAKDEEVEDFYQLLQSE IDQTCNQDALIITGDWNAKVGNKEGSVVGK YGLGDRNNAGDHMIEFCKTNDFFIANTFFNQ HKWRLYTWTSPDGIHRNQIDYICGKRRWKSS ISAVRTRPGADCGTDHQLLICKFKLKLKKIRTS PREPKYDLEYIPPEFKDHLKNRFDVLITNDRRP DELWNDIKDIIHEESKRSLKRQERKKRPNWM SEETLKLALECRAAKAKGRNGEVKDLNRRFQ		
					TCTGCTGCAAAGGACCTCT TCAAAGTATTCAAAAGTAA GGATGTCACCTTGAAGACT AAAATGCGCCTGACCCAA GCCATGGTATTTTCAATCA CATCATATGCATGTGAAAG CTGGACCATGAACAAAGA AGACAGAAGAAGAATTGA TGCCTTTGAATTGTGGTGT TGGAGAAGAATGTTGAAT GTACCATGGACCGCCAAA AGAACAAACAGATCTGTTC TGGAAGAAGTACAACCAG	GTGCTGGTGAGAG GGGCCTGCTGTTCA ACACACCCGTGCTG GTGAGAGGGGCCT GCTGTTCATCACAC CCGTGCTGGTGAG AGAGGCCTACTGTT CATCACACCCGTGC TGGTGAGAGGGGC GTGCTGTTCATCAC ACCCGTGTTGGTGA GAGGGGCCTGCTG TTCATCACTCCCAT GCTGGTGAGAGAG GCCTACTGTTCATC ACACCCGTGTTGGT GAGAGGGGCCTGC TGTTCACCACTACC CATGCTGGTGAGA GGGGCCTACTGTTC AACACACCAGTGCT GGTGAGAGAGTCC TGCTGTTCAGCACA CCCGTGCTGGTGC GAAG				
			GGTAAGCAGGGTA ACACCATCAGGTAA GGAGGGAGTGAGC AGCCCCCACTTGTC TGTCAGTGGAGGC TTGCTTGTTGCTGT GATGCAGGAAGCT ATGCCACTGGTATT TAGGTACCAGCAG GGTCAACCATGGT AGACAAGTTTCAGC TGAGCTTCCAGACT AAGACAGACTAGG	Procavia capensis					
				GCA_004026925.2				
					RTE1_Pca			

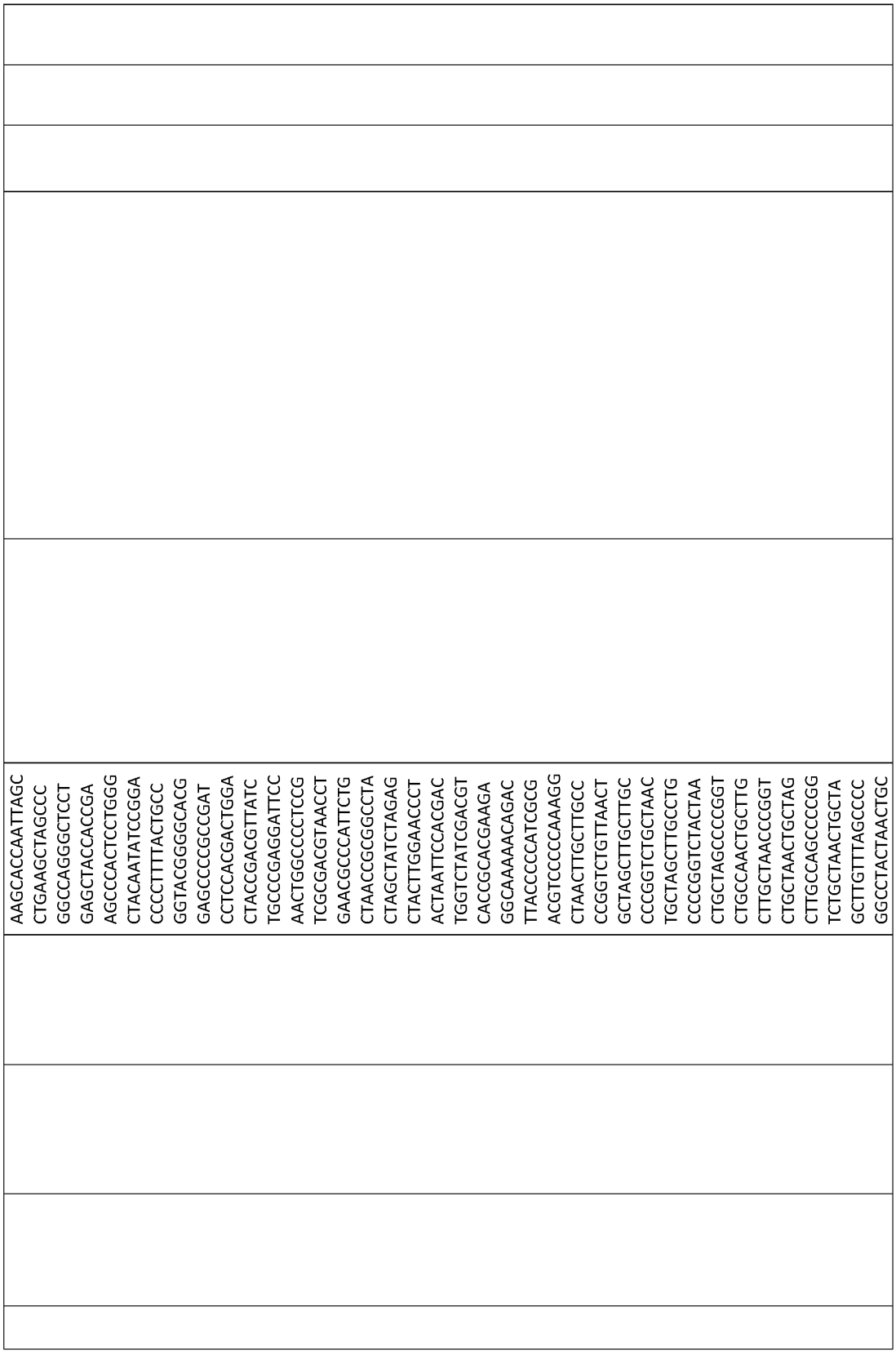
			AAGAAGGACCCAG CAGTCTACTTCTGA AAGGAATTAGCCA GGGAAAACCTT	AGTGCTCCTTGGAAGCAA AGATGGCAAGACTGCATC TTACATACTTTGGACATGT TGTCAGGAAAGATCAGTC CCTGGAGAAGGACATCAT GCTTGGTAGAGTGGAGGG TCAGCGAAAAAGAGGAAG ACCCTCGATGAGATGGATT GACACAGTGGCCGCAACA ATGGGCTTAAGCATAAAA ACGATTGTGGACATGGCG CAAGACCGGGCAGTGTTT CGTTCTGTTGTTCATAGGG TCACTATGAGTCGGAACCA ACTCGATGGCACCTAACAA CAACAA	RAAREDKRKYYNDMCKELEMENQKGRTHSA FLKLKELKKKFKPRVAIVKDSTGKLLNDSESIKR RWKEYTESLYQKELVDVQPFEEVAYDQEPM VLKEEVEAALKALAKNKAPGIDGISIEIFQQTD AALEVLTHLCQEIWRTATWPTDWKRSIFMPI PKKGDPTECGNYRTISLISHASKILLKIIQKRLQ QYIDKELPEIQAGFRRGHGTRDIIADVRWILA ESKEYQQDVYLCFIDYAKAFDCVDHTKLWLTL QRMGIPEHLIVLLRNLYMDQEAVVRTEQGDT AWFKVRKGVHQGCILSPYLFNLYAEQIIREAG LYEEEQGIRIGGRLINNLCYADDTTLLAGNEED LRHLLMKIKDHSLQYGLHLNIRKTKILTTGPVS NIVINGERIEVVKDFILLGSTINALGSSSQEIKR HIALGK			
							cd01650	647	905
	Crack-1_SSa	GCF_000233375.1	Salmo salar	TCCAAGATGGCGT AGCAGTCAGACGT CTTTGTCCTCCTCTT GTCGTGTCCCGTGT ATATATTTTTATATC TTTTTTCTTCGCATA TCTTTTTAATATTTT TCTTAACCTCAACT TCAAAACACTCTCC TGCAACCCGCCTCA CCCAATGTGGAATA GATATGCTATTTTA TTTACCTCGAACTG GAATCCACACAAG AAGCTAGCCAGCTA ACTAGCTACTAGCT AGTAGTCAGCTAG CCACGGTTAGCGG TCATCAGCTATCCT TTAGCTCGAAAAGC TATCGCCAGTTTGT ACAGCGCGACTCA	TAGCCCATCTATAATTAGC CCAAACAACTACCTCATCC CCATACTGTTATTTATTTAT TTTGCTCCTTTGCACCCCA GTATCTCTACTTGCACATT CATCTTCTGCACATCTACC ATTCCAGTGTTTACTTGCT ATTATTTCGCCACCATGGC CTATTTATTGCCTTACCCCT ATCTTACTCATTTGCTCAC ATTGTATATATAACTTTTTC TACTGTATTATTGACTGTA TGTTTGTTTACTCCATGTG TAACTCTGTGTTGTTGTAT GTGTCGAACTGCTTTGCTT TATCTTGGCCAGGTCGCA GTTGTAAATGAGAACTTGT TCTCAACTTGCCTACCTGG TTAAATAAAGGTGAAATA AAAAATAAATAAA	MITRLSMPLLNVNMPCPLLFWLVFIGLFHCR ASSPAHYTLSNLSVPPPTHAMTSPGFNDVSR DNISLIITQCLGLPPLYSHPTIPLSVHYTLNLFYR PQKPAPFTLCSGRARRPILIAFSRTLILLLLCSSG DVEVNPGPAVPSSTPTPQALSFDDFCNRKSL GFMHVNIRSLLPKFVLFTALAHSANPDVLAVS ESWLRKTTKNSEISIPNYNIFRQDRTAKGGGV AIYCRDSLQSSVLLSRSVPKQFELLLLKIHLSKN KSLTVAACYRPPSAPSCALDTICELIAPHLSSEL VLLGDLNWDMLNTPAILQSKLDALNLTQIINE PTRYNPKAVNTGTLIDIILTNLPSKYTSAVFNQ DLSDHCLIACIRNGSAVKRPPLITVKRSLKHFSE QAFLIDLALVSWKDIDLIPSVEDAWLFFKSAFL TILNKHAPFKKFRTRNRYSPWFSPDLTALNQH KNILWRSALASNSPRDMQLFREVRNQYTQA VRKAKASFFKQKFASCNTNSKKFWDTVKSME NKSTSSQLPTALRIGNSVTTDKSTIIENFNKHF STAGHAFHLATPTPVNSTAPPTATRPSLPHFS FSQIQSADVLKELQNLDPYKSAGLDNLDPFFL KLSAEIVATPITSLFNLSFVSSEIPKDWKAAAVI PLFKGGDTLDPNCYRPISILPCLSKVFESQVNK QITDHFESHRTFSAMQSGFRAGHGCTSATLK			

					VLNDIITAIDKKQYCAAVFIDLAKAFDSVNHHI LIGRLNSLGFSNDCLAWFTNYFSDRVQCVKSE GLLSGPLAVSMGVPQGSILGPTLFSVYINDVA LAAGESLIHLYADDTILYTSGPSLDTVLTNLQTS FNAIQLSFRGLQLLLNTSKTKCMLFNRSLPAP ARPSSITTLDGSDLEYVDNYKYLGVWLDCKLS FQTHIKHLQSKVKSRIGFLFRNKASFTHAAKH TLVKLTILPILDFGDVIYKIASNTLLNKLDAVYH SAIRFVTKAPYTTHHCDLYALVGWPSLHTRR QTHWLQVIYKTLLGKVPPYLSSLVTIAAPTCST RSSRYISLVTPKANSSFGRLSFQFSAANDWNE LQKSLKLETLISLTSFKHQLSEQLTDHCTCT		
				GACCAGAGCATAC CGGACCTATTTTTC TCCATATCCCCGGA TTTCTACCGCAAGC TCTGGACATTTACA CCTGGATCATTGCA GCTAGCTAGCTGCA TCCGAGTGACTACT ATTGGCTAACGTTT CGGTCCCGAAGCA AGCACCAATTATCC TGGAGCTAGCCCAT GCTAGGCCCATCTC CCGGCTAGCTGAA GAGTTCCATCAGCC ACTCCTTGGGCTAC AATACCTATTTTGC CAACTGACCTGGAC CCGTTTTACTGCCG TGCGGAGCCCCAC CGAGCCTTCACGAC TGGACTACCGACGT TATCTGCCCGAGG GAGTTATCCAACTG GCCCCTCCGTCGCG ACGTAACCTGAACG CCCATCTGCTGGCC TGCTAACCGCGGAT CGTTAGCTGTCTAG AGCATATCGGACT GCTATCTGAATAGG TCTATCGGACAATT TCTTGGGCCACTAT ACCTATTTTGCCAA TTGGACTGGTCCCC TCTACCACACGGAA CCCTACTAATCCAT CACGACTGGTCTAC			

	Rex1-2_SSa	GCF_0002333 75.1	Salmo salar	CTTCCGGCGCCGAC AGAG	CGACGGAAACGCA CGAGGTGGCTAAA ACAGACTTCTTCCG TCGCGACGTCCCTC TAAGGCCCTTCTGC CAGCTTGCTACCTA TGGCCCGCTAGCTG TCTGAATCGCCGTG TCCCCAGCCAGCCC AACTACTCACTGGA CCCCT	TAACATTGAGTGGCTGCT GCCAACATACTGACTCAAC TCCAGCCACTTTAATAATG GAAAAATTGATGTAATCA ATTTATCACTTAGCCACTTT ATATTATTTATATTAGATA ATGTTTACATACCCTACAT TACTCATCTCATATGTATA TACTGTACGCTATACCATC TACTGCATCTTGCCATCTT GATGTAATTAGATGTATCA CTAGCCACTTTAAACAATG CCACTTTATATAATGTTTTC ATACCCTACATTACTCATCT CATATGTATATACTGTACT CTATACCATCTACTGCATC TTGCCATCCTGATGTAATG TATCACTAGCCACCTTAAA CAATGCCGCTTTATATGTT TTCATACCCTACATACTCAT CTCATATGTATATACTGTA CTCCATACCATCTACTGCA TCTTGCCTATGCCGTTCGG CCATCACTCATTTATATATT TTTATGTACATATTCGTATT CATTCCTTTACACTTGTGT GTATAAGGTAGTTGTTGT	MAASLRVPRKLCSILFFYVLFLTLLPQVILGFITY SREELLDIRATSTHQHYDQEYDFPEADPLFGP PPRTMDRIPAGDPKQRRRRRRRGRRSGLLVRL RRRAHRAPLPSILLANVQSLDNKVDEIRARVA FQRDIRDCNVLCFTETWLTRDTLSESVQPPGF FTHRADRNKHLSGKKKGGGVCLMINETWCD HNNIQELKSFCSPDLEFLTIKCRPHYLPREFSSII ITVVYIPPQADTSTALKELHWTLCKLETTYPEA AFIVAGDFNKANLKTRLPKFYQHIDCATRAGK TLDHCYSNFRDAYKALPRPPFGKADHDSILLL PAYRQKLKQEAPALRSVQRWSDQSDSTLQD CFDHVDWDMFRIAANNNIDEYADSVCEFIN KCIGDVVPTASIKTFPNQKPWIDGSIRAKLNA RTTAFNQGKVTGNMTEYKQCSYSLRKAIKQA KRQYRDKVESQFNGSDTRGMWQGLQSITDY KRKTSPVADHDALLPDRLNNFFARFEDNTVP LTRPATKTCGLSFTAANVSKTFKRVNPRKAAG PDGIPGRVLRACADQLAGVFTDIFNQSLSQSA VPTCFKRATIVPVPKKAKVTELNDYRPVALTS VIMKCFERLVKDHITSTLPDTLDPLQFAYRPN RSTDDAIAITLHTALTHLDKRNTYVRMLFIDYS SAFNTIVPSKLVIKLETLGLDPALCNWVLDFLT GRPQVVRVGNNISTPLILNTGSPQGCVLSPLL YSLFTHDCVAMHASNSIIKFADDTTVVGLITN NDETAYREEVRALGVWCQENNLTLNVNKTK EMIVDFRKQQREQPPIYIDGTVVERVESFKFL GVHITDKLKWSTHTDSVVKKAQQRLFNLRRL	cd01 650	579	832

	RTEX-1_SSa	GCF_000233375.1	Salmo salar	CTATCAGTTAGCAT GTGGAACATTCAG GGCCTAAACTCATC AACCTTTGGACTGA AGAGTTTAGCACTG GAGTTCAACAAAA ATCTTAAAGATGTT GACGTCATCATTCT GCAGGAGACATGG TGTAAGGCTGACAT TGTCACTCACTGTC CCACAGGCTACAG AGAGGTCATTGTG CCATCACAGAAACA CAGCTCTGTCAATA GAGGCAGAGACTC TGGAGGATTGATC ATTTGGTACAAATC CGAACTACAAAATC TAATTGATCCCCTC AAAATTGGCAAATA TCACATTTGGTTAA AACTGAAAAAAGA ACTTGTACTGACAG AAAAAG	GGAATTGTTAGGTTATATT ACTTGTTAGATATTACTGC ATGGTCGGAACTAGAAGC ACAAGCATTTCGCTACACT TGCATTAACACCTGCTAAC CATGTGTATGTGACAAATA CATTTGATTTGAT	TGACAGACCAATCAACCTG CACATGTCCTCTACTGTAT GCTTATTGTTATTGTTGAA TGTATGGTTATTTTGACCC TTGGTTATTGTTGTTACTG TTGTCCCGTTGACAATATT TGATTCTCATTTTTATTTAT TTTTATATTGTAAATATCC AAAATAAGCTTTGGCAATA TGTACATTGTTACGTCATG CCAATAAAGCGAATTGAA TTGAATTGAATTGAGAGA	KKFGLSPKTLTNFYRCTIESILSGCITAWYGSCS AHNRKALQRVVRSAQRITGCKLPALQDTYTT RCHRKAKKIIKDNNHPSHCLFTPLSSRRRGQY RCIKAGTERLKNSFYLKAIRLLNSHH	MCSFAQYISPPQNPHITQRRSSPPRIPILLRGD ISPPQNPHITQRRYPPLRIPILLRGDIPPSESPYY SEEIFPTLEEETCHFQAQGNVLICGDTNARTG TLPDLTSTRGDSFITGHTVSNCLHLPHRNNSD STVNKNGRDLLQLCRSLGLYFVNGRLRGDSL GRFTYCSPLGHSTVDYMITDIDPFSLSSFTVKP LTPLSDHSQITLFLKRTDMETTTHSQPSKLYNI RNSYRWAQNSTEEYQIKATWNQNIQTLLDNF LDTTFTHSKEGINLAVKNINYIFRQTAKEAQLK LIKNKTKKITDDNWFDADCKIIRKKLRTLSNQK HRDPNNGELRLHYCETLKLYKRTLRTKKAQYN SKQLTLIEESINTNNFWQNWKKLKKSKQEELA IQNGDIWTTHFKTLYNTVQIDTNAEQRQIHE KLNGLEKAIKDNQNPLDSPITDQELYKKLQAL KFKKACGPDGILNEMLKLTSAKFQLAILKLFNL ILSVGYFPDIWNQGLITPIFKNGDKFDPNNYR GICVNSNLGKVFCSIINVLSKSQIGFIPKHRTTD HIYTLHTLIDKHVHQNNTKIYACFIDFQKAFDS IWHTGLFYKVIESGVGGKTYDIIKSMYTGNTC SIKIGKKRTEFFNQGRGLRQGCNLSPALFNIYI NELATILEKSSAPGVSLHNSEVKCLLFADDLCL LSPTAHGLQQSLDLLEQYCQTWALAVNPKNT CPREDPGPVRPVLQDLCLLSPTAHGLQQSLDL LEQYCQTWALAVNPKKTKIMIFQRRSRSQGI RPKFSIGTKYIEYCTHYNYLGLKISSTGHLNEAV NELREKARRAFYAIKKQIQIEIPIKIWLKLIEYVI EPIALYGSEVWGPLAKQDFIKWDKHPIETLHA EFCKILLHVQRKTTNNACRAELGQYPLIIKTQK RAIKFWKHLKYSDPLSYHYQALQCQELSKEKS PLIQLVLGLSSQTCSTNTLKPQDQNIQSIRINQ ITTQSKQNYIAYWETQAQTQSKMQCYLALN	cd01650	499	800

					RQYTVANYLTMVTDQNLRKTLTKYRLSEHSL AIEKGRHRKTWLPVEERLCNHCTTAEPETELH FLTKCQKYKTIRECHFPKFETLIQGFKDLSDED RLPVLLGEDAESCGLAAHYIAACHKMRDSV	cd01650	646	904
	Crack-1B_SSa	GCF_000233375.1	Salmo salar	CCAAGATGGCGTA GCAGTCAGACGTCT TTGTCCTGTCGTGT CCCGTGTATATATA TTTTTACATCTTTTT CTTCGCATATCTTTT TAAAATATTTTCCT AAAACTCAACTTCA AAACACTCTCCTGC AACCCGCCTCACCC AATGTGGCGTGGA TCTTTTTTTTTACTA AAGTATTTCTATTT ACCTCAGATCTGGA ATCCCACAACAGAA GCTAGCCAGCTAAC TAGCTACTAGCTAT CAGTCAGCTAACCA CGGCTAGCGGTCA TCAGCTAACCTTTA GCTCGGAAAGCTA TCGCCAGTTTGTAC AACGCGACTCAGA CCAGAGCATACCG GACCTATTATTATT TTTATTCTTTTTCTC TCCATATCCCCGGA TTTCTACCGCAAGC TCTGGACATTTTCA CCTGGATCTTCGCA GCTAGCTAGCTGCT ATCCCGAGTGACTA CTCCTGGCTAACGT TTCCGTCCCGGAGC	TAGCCCATCTATAATTAGC CCATCTAACTAACTACCTC ATCCCCATACTGTATTTATT TATTTATTTTGCTCCTTTGC ACCCCAGTATCTCTACTTG CACATTCATCTTCTGCACA TCTACCATTCCAGTGTTTA ATTGCTATATTGTAATTAC TTTGCCACCATGGCCTATT TATTGCCTTAACTTACCTC ATTTGCACACACTGTATAT AGACTTTTTGTTTTCTTTTG TTCTACTGTATTATTGACT GTATGTTTTTGTTTACTCCAT GTGTAACTCTGTGTTGTTG TATGTGTCGAACTGCTTTG CTTTATCTTGGCCAGGTCG CAGTTGTAAATGAGAACTT GTTCTCAACTTGCCTACCT GGTTAAATAAAGGTGAAA TAAAAAAA	MITRLSMPLLNVNMPCPLLFWLVFIGLFHCR ASSPAHYTLSNHLVPPPTHAMTSPGLNVSRD NISLIITQCLGLPPLYSHPTIPLSVHYTLKLFYRP QKPAPFTLCSGRHRRPILIAFSRTLILLLLCSSG DVEVNPGPAVPSSTPTPQALSFVDFCNRKSL GFMHVNIRSLLPKFVLFTALAHSANPDVLAVS ESWLRKTTKNSEISIPNYNIFRQDRTAKGGGV AIYCRDSLQSSVLLSRSVPKQFELLLLKIHLSKN KSLTVAACYRPPSAPSCALDTICELIAPHLSSEL VLLGDLNWDMLNTPAILQSKLDALNLTQIINE PTRYNPKSVNTGTLIDIILTNLPSKYTSAVFNQ DLSDHCLIACIRNGSAVKRPPLITVKRSLKHFSE QAFLIDLAGVSWKDIDLIPSVEDAWLFFKSAF LTILNKHAPFKKFRTRNRYSPWFSPDLTALNQ HKNILWRSALASNSPRDMQLFREVRNQYTQ AVRKAKASFFKQKFASCSTNSKKFWDTVKSM ENKSTSSQLPTALRIGNSVTTDKSTIIENFNKH FSTAGHAFHLATPTPVNSTAPPTATRPSLPHF SFSQIQSADVLKELQNLDPYKSAGLDNLDPFF LKLSAEIVATPITSLFNLSFVSSEIPKDWKAAAV IPLFKGGDTLDPNCYRPISILPCLSKVFESQVNK QITDHFESHRTFSAMQSGFRAGHGCTSATLK VLNDIITAIDKRQYCAAVFIDLAKAFDSVNHHI LIGRLNSLGFSNDCLAWFTNYFSDRVQCVKSE GLLSGPLAVSMGVPQGSILGPTLFSVYINDVA LAAGDSLIHLYADDTILYTSGPSLDTVLTNLQT SFNAIQLSFRGLQLLLNTSKTKCMLFNRSLPAP ARPSSITTLDGSDLEYVDNYKYLGVWLDCKLS FQTHIKHLQSKVKSRIGFLFRNKASFTHAAKH TLVKLTILPILDFGDVIYKIASNTLLNKLDAVYH SAIRFVTKAPYTTHHCDLYALVGWPSLHTRR QTHWLQVIYKTLLGKVPPYLSSLVTIAAPTCST RSSRYISLVTPKANSSFGRLSFQFSAANDWNE LQKSLKLETLISLTSFKHQLSEQLTDHCTCT			


AAGCACCAATTAGC
CTGAAGCTAGCCC
GGCCAGGGCTCCT
GAGCTACCACCGA
AGCCCACTCCTGGG
CTACAATATCCGGA
CCCCTTTTACTGCC
GGTACGGGGCACG
GAGCCCCGCCGAT
CCTCCACGACTGGA
CTACCGACGTTATC
TGCCCGAGGATTCC
AACTGGCCCCTCCG
TCGCGACGTAACCT
GAACGCCCATTCTG
CTAACCGCGGCCTA
CTAGCTATCTAGAG
CTACTTGGAACCCT
ACTAATTCCACGAC
TGGTCTATCGACGT
CACCGCACGAAGA
GGCAAAAACAGAC
TTACCCCCATCGCG
ACGTCCCCCAAAGG
CTAACTTGCTTGCC
CCGGTCTGTTAACT
GCTAGCTTGCTTGC
CCCGGTCTGCTAAC
TGCTAGCTTGCCTG
CCCCGGTCTACTAA
CTGCTAGCCCCGGT
CTGCCAACTGCTTG
CTTGCTAACCCGGT
CTGCTAACTGCTAG
CTTGCCAGCCCCGG
TCTGCTAACTGCTA
GCTTGTTTAGCCCC
GGCCTACTAACTGC

	Rex1-3_SSa	GCF_0002333 75.1	Salmo salar	TAGCTTGCTAGCAT CGGCCTGCTAGCTG TCTGAATCGCCGTG TCCCCAGCCAGCCC AACCACTCACTGGA CCCAT	AG	TAATCAAATGGCTACCCAG ACTATTTGCACCCCCCCCCC CCCCACGCTGCTGCTACTC TCTGTTATTATCTATGCAT AGTCACTTTAATAACTCTA CCTACATGTACAT	MAPEGMAAVLSALNQPCYFVCFFALFVTCFV HNVAATVSYDRKELLDIRTAITHLKLDKEFFFN ESDGRDILQTPDQAQIPVIRWRRKRRFRGKR SGCLVRIRRRVANLPLPSVLLANVQSLENKW DELKARISYQRDIKNCNILCFTESWLNDDIKNI QLAGYTLYRQDRTAASGKTRGGGLCIFVNNS WCTISKEVSRFCSPEVEYLMISCRPHYLPREFS SVFFVAVYIPPQTDAGTKTALNELYTAISKQEN AHPEAALLVAGDFNAGKLKSVLPNFYQHVKC ATRGKKTLDHLYSTHRDAYKALPRPPFGKSDH NSILLIPAYKQKLKQEAPVTRSIKKWSDEADAK LQDCFASTDWNMFRDSSDGIEEYTTSVTGFI NKCIDDVVPTVTVRTYPNQKPWITGNIRTELK GRAAAFKERDSNPEAYKKSRYALRRTIKQAKR QYRTKIESYYTGSDARRMWQGLQTITDYKGK HSRELPSDTSLPDELNNFYARFEASNTETCMR ASAVPDDCVITLSAADVSKTFKQVNIHKAAGP DGLPGRVLRACADQLASVFTDIFNLSLSESVIP TCFKQTTIVPVPKNTKVTCLNDYRPVALTSVA MKCFERLVMAHINTIIPETLDPLQFAYRPNRS TDDAISIALHTALSHLDKRNTYVRMLFIDYSSA FNTIVPSKLITKLRTLGLNTSLCNWILDFLTGRP QVVRVGNNTSATLILNTGAPQGCVLSPLLYSL FTHDCTARHDSNTIIKFADDTTVVGLITDNDE TAYREEVRDLTVWCKDNNLSLNVIKTKEMIV DYRKRRTEHAPILIDGAVVEQVESFKFLGVHIT NKLTWSKHTKTVVKRARQNLFPLRRLKRFGM GPQILKRFYSCTIESILTGCITAWYGNCSASDR KALQRVVRTAQYITGAKLPAIQDLYTRRCQRK ALKIVKDSSHPSHRLFSLLPHGKRYRSAKSRSK RLLNSFYPQAIRLLNS	cd01 650	581	833

	L2-2B_SSa	GCF_0002333 75.1	Salmo salar	GTTCCGGAACTCCA ATTTGAGCAGCAG CAGCTGAAAAATG AGCTCCTACAAATA TTGAAATTTACATG GTTTTTAGAGTGAA GTACATCGGAAAA AAGCAAAAGAAAA CTGCAAGACTGAAT AGGAGAAAAAACA AGCAACAAACAAA GAAGATAGGCTTTT GAAAAATAACTATT TTTCATAAAATTGT AGTTATCTTAGCTA GCTGAATTGTTTAC CAGTTGCTAAGCA GTTGCTAGGGACTC TTTTGGAAGAAGCT AGCTAGCTAACGA AGAACAACAAAGA ATATCTAGTTAACA GAAGAAAAGAAAG AGAAAATCAGGAC AGAAGAAAAAGGA TATCAAAGTGAAAC AGTTCTCTAAAGAC ACAGGAGACAATA ATACAAGAAACAA CACTTCTGTAGCTT GTCAACT	TAGGATAAAGTAATCCTTC TAACCCCCCCCCCCCCCTT AAAAGATTTAGATGCACTA TTGTAAAGTGGTTGTTCCA CTGGATATCATAAGGTGA ATGCACCAATTTGTAAGTC GCTCTGGATAAGAGCGTC TGCTAAATGACTTAAATGT AAATGTAAATGTAA	MCLSIPVLSSLHRPYKRFTPRGRCHSNLVVPA RTTHVEFQVSGSLWNCRSAANKAEFISAYATL QSLDFLALTETWITTDNTATPTALSSSAHVFS HTPRASSQRGGGTGILISPKWTFSLSPLTHLSI SSFEFHAVTVTSPFKLNILIIYRPPGSLGEFINEL DALISSFPEDGSPLTVLGDFNLPTSTFDSFLSAS FFPLLSSFDLTLSPSPPTHKAGNTLDLIFTRCCS STNLIATPLQISDHYLVSFSLLLSSNTSHSAPTR MVLRRPNLRSLSPATLSSSILSSLPSAQTFSNLS PDSASSTLLSSLSASFDFLCPLSSRPARSSPPAP WLDDSLRAHRTGLRAAERKWRKTRLPADLA SFHSLLSTFSSSVSAAKATFYHSKFQASASNPR KLFATFSSLLNPPPPPPPPPPSLRMTSSTILKRRL TISDPRLLSQTTPLVLLTLPYPVLEALCYLLLPPE SSSPSPPSSLSADDFVNHFEKKVDDIRSSFAKS NDTAGPAHTALPCALTSFSPLSPDEISRLVTAG RPTTCPLDPIPSSLLQTISGDLLPFLTSLINSSLT AGYVPSVFKRARVAPLLKKPTLDPSDVNNYRP VSLLSFLSKTLERAVLGQLSCYLSQNDLLDPNQ SGFKTGHSTETALLCVTEALRTAKANSLSSALIL LDLSAAFDTVNHQILLSTLSELGISGAAHAWIA SYLTGRSYQVAWRESVSAPCALTTGVPQGSV LGPLLFSLYTKSLGSVISSHGLSYHCYADDTQLI FSFPPSDNQAANRISACLSDISVWMTDHQLK LNLGKTELLFLPGKDCPFHDLAITVDNSLVSSS QSAKNLGVILDNTLSFSTNIKAVTRSCRFMLY NIRRVRPCLTQEAAQVLIQALVISRLDYCNSLL AGLPACAIKPLQLIQNAAARLVFNLPKFSHVT PLLRSLHWLPVEARIRYKTMVIAYGAVKGTAP PYLQALIRPYTQTRALRSSTSGLLAPLPLRKHSS RSAQSKLFAALAPQWWNKLPHDARTAESITT FRRHLKPHLFKEYLG	cd01 650	577	838
	SR2	GCF_0002379 25.1	Schistosoma mansoni	CCCAAATGCCCTGG TACGGCCGAGAGT GGGGAGAGTCCAC TCTCCCTCTCGAAA TGCTCTCACATGGC CACGCGTATATAGC	AGAGTGGTTCTAACTTTCT TAACTGAAAGAGTCTTCTG GTTGTACATTTCAGTCCGC CTTATCTTTTCATCCCTTCT CTTCCTACTTTCATTATTTT GTGTGGCGCATATGTATCT	MLPPIMNTEAVFLKPRTPFKLASFNVRTLMQI RQQIGLAMSLEGLNVDVCCLSETRIQDSSEVL QIRSPSVASKSLFHVRLSGDPVASSSGLAGVG VALSARAEAALIDWIPINSRLCAVRLESSIKVR RNRREKRCLFVISAYAPTDCSPDAIKDEFYHQL TVLLQKARSTDIVVLAGDLNAQVGRLGTEESR	cd01 650	549	808

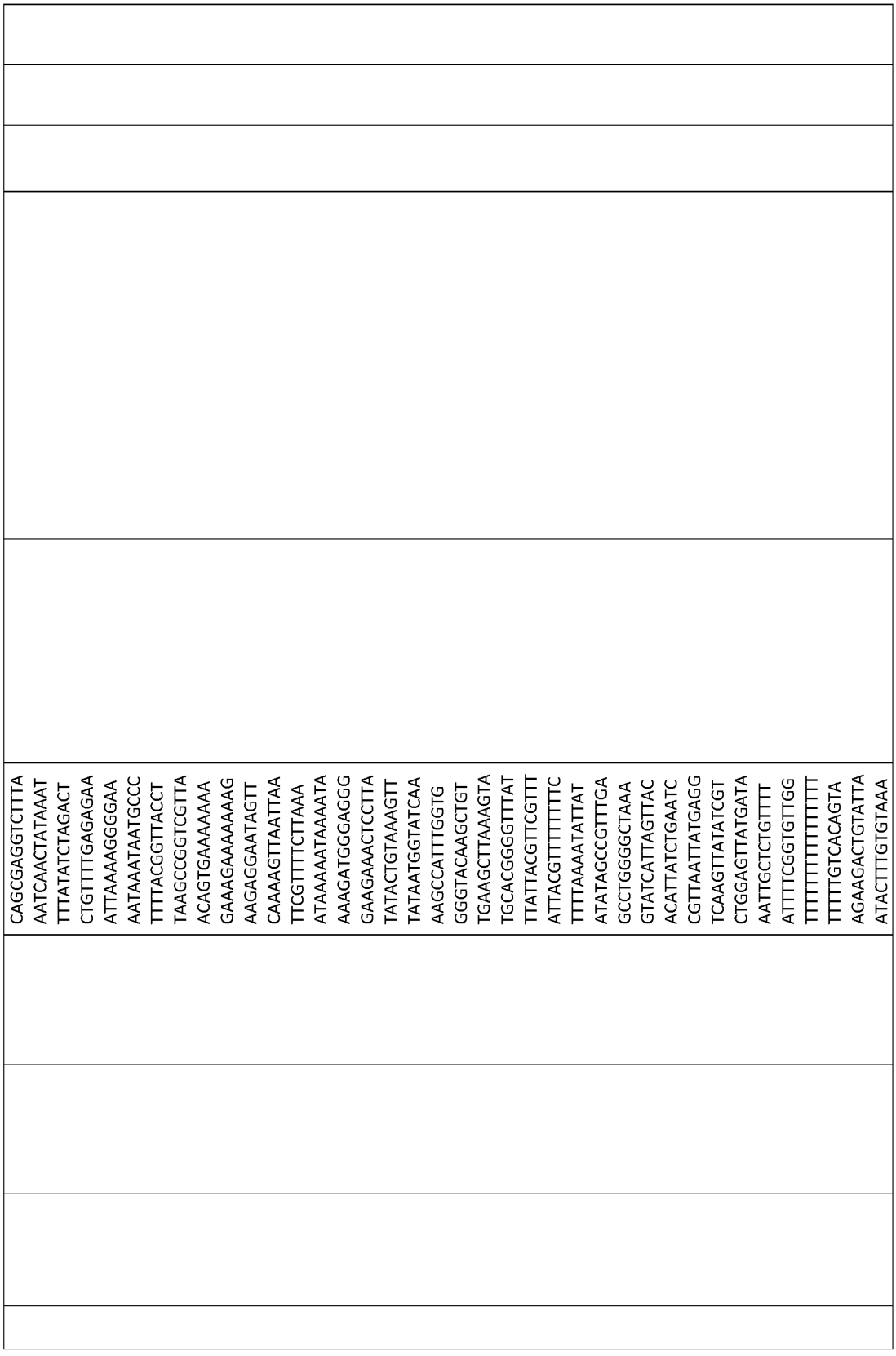
				CTCTGCCAGGGAA GTCCTACTCACTGC CTTCTCGTGGCGGG GGTGTTGTTTACGA AAGTGAGAGGACG AAAAGCGAATGTC CGGCGCTTTAACCG GGTTGGTGGACAC GGAGGGTCCACCT AGGGGAGTTGGAA AACCCTGATTCCAA ACCAATGGTGCACA TGGGCTCCAGTATC CTGATGGAACGAA TGGCGAATGAACC TGTCATTGGTCACC GGCTACCATGGGA CTGCATCTCCTCAC GATGCTCCACTGCC TTGTGGATCAGACC TTTAGGTCAAAGGC TCGGGGTGTGGCC CCCTAAGAAAACCA CCTGCTTCGGTTTG GGCACCCGGGCAG TATCACAGCCCACA CACAAATAAATGAA ATGATGAATTCTAT TGACGATACTATTC TCGCTCATACTAGA AGCAAGACAATTTA CATTATTATTATTAT TATTACCATTATTAT TATTATTATTATTAT TACTATTATTTACTA CGTCGCTAATTCAC CCCCTTTATCCCCA ATTTGTGTAACCTT	GGTGCCCTTTTGTACCAAT ATATATGTGTTTAAATAAA TAAATAAATAAATAAA	LGGRWGLVGRRTDNGDRLLQLCTDHNLFLA STNFRHSHRRCATWRPPSASQAWTQIDHIAI SYRWRGCVQDCRSFWSTYLESDHALVCANL TLLFSGRRIDHHQRIDVSKLAATSVASKYRAEL ASRLATIPPKSIDEHWLHLHDAMKMASKAAC GFAKRPAYKHWVSSGSLQLMEARRSTPGDR EFDHKRRLLRNEIGQSLRKDREAWWSERANE MEAAAASGNCRKLFQLIRATGSKKSGVSETIY EDDGMPITNISRRLGRWAEFFEGQFNWPAA PATSTSLSCPPWPVTTDPPNEAEVRKELQLLK RYKSPGPDDLPPALFKDGGDFLTKELTVLFAK VWEQESVPTSWNESIVVPIFKKGSRCSCNNY RGISLLPIASKLLASIILRRLFKTRERLTREEQAG FRSGRGCIDHIFTLRQMLEHRHTYRRPTIVVFL DIRAAFDSLDRTVLWDCLLKKGVPEKFINILKA LYTNTSGRVRAYNHLSPLFHSSSGVRQGCPIS PFLFNFAIDDILETALMDVSNGGVDMLPGER LLDLEYADDIVLLCDNAQGMQSALNQLAISV RRYGMCFAPSKCKVLLQDWQDSHPVLTLDG EQIEVVEKFVYLGSYISAGGGVSDEIDARIMK ARAAYANLGHLWRLRDVSLAVKGRIYNASVR AVLLYACETWPLRVEDVRRLSVFDHRCLRRIA DIQWQHHVSNAEVRHRVFGRRDDNSIGVTIL KHRLRWLGHVLRMSSQRLPRRALFADSGTG WKKRRGGQCMTWCRGMKESCKGLASVGP SRLPGWGPRDGATQWLETLSDMAQNRSQ WRSCCNLLLLST			

			ACTTTTCAACTTAT GCAGCAGCTCGCC CATGGTGGAAAAT CTGTCCTATCTAAA CGATCTCCAGTCAC TGTCTGGTTGAAAG TGA	AACGTT	TACTACCACACAAGACGGGT ATTTTGGTAAATCTTGAAA AATTTCGATTTAATTAATG TTTGAGTTAAAATTGCCAT TTGACTTGTGTATGTGATT TATTGTGTTAATTAATGAA TTCCACCCTCTTCCAGGGA AAGGTTTAACTTTGCATAA AATTAAACTCTCTGTCTAC ACTTAGTAGCAAATTAATC TAGAATGAAGAAGCCTGA TGGTATTTTTCCCAAATCT GCTGAAAACTTAGTTCTGA AGGATGGAAATAAAAGCA GATAAAGATTCAATATTCG ATAAAGAAATGGACCTCA ACAATCAATGTCAATTATC AATAATAAATCTGGGCAG ATGTCAGAAGGCGATAAT CAAATGCTGAATTCCAATG ACCGAATCTACTTTGATGT CTTAGATTTGTGATCAGCT TCCACTACAAAGGTGCTGC ATCAAAAGCCAATGGAGA TTCAATGCCGGATGAAAAT GCCGATGAAAGAGAATCA GAGCCGAGGACAATTTAG AGAAAATTCGATTGCAACT GCGTCAAGAGTTCCTCAAC AATAAGATGAAATATCGA AACTCAATTGGATCTTTTG	MKIEIGEFAALFQRFFEQLNAKSKQEEEFRNE SVPGSRVPSEDMPVFQINERFHALERSNELLN GKMDSIVNTLNYLISGLNGNFDNNNKLNCIC KTNHVKEPLSIVKTNLVSELPSALLVQNKVPCL KLSEEDLIRKIPTKIVPSKKVSKKTIEHTSRVNFS LFVRVLDSNERQQFFKLFPNLFALLKSKESEKK SEVLNIELTKISKKDNPLNIAGFKNFFFRNNNI CVTIDFYNSSKVKSKITKNLKIKTSLINSKIVKNY NTNKFIPASLSIKKFLFVLSDKLAETTNISNVTV REEQINKIITLLKIKNSLPDYLVKSQIKYKDGDFL MLDLLNSCEIEGELGQKITELSRSQIKELFNSFP EIQELFFIEKKANSGTFFHRKKSIDENPEKLINLI NKFKIEKTFPQQISFVGGKFVKIDKLNDSQFP QVESESDIMFNLGLFSLFGLSEYDNILNSLLCQ PKSPEKLVDIIFHLNYLNLKGNFDTSTNGTYEN YVQNYIVNCLLMHSILDLETCAFLNLDNIELW APKLIELIPSSSHYSSSCKEKLLKLKQWKLSGEI DRIHNIICPETIIEINNFLGVIDQSYTTKSDQLKF LTSLGEKINNIEIDNDVDDLVKAKRMLYFIIRD KKRFLNKDNFNYRIPFNCRKICTLKMSDFITKLI QSNELFEEREAVVDNNIDVSPNENLNILIVEP VNPTPDTVIVGQMDTFPTDIDETEPVTIDKPL KKITSKQAVKPKAQKSKREVPASVSISVPISEET DADSTIINYFFPERDPKIWFTDDDIDYYLESHI GNPNFAPVKCFIVEILSSNCKDEIFPMPDNLFK AQIILCPLNINESHWILFVYCKLTKISYFVDPILR KRHNLINKSSALKITMALNNNLFDLKAAVTFHP HDNLVYQDNNYDCGPYICAYAMLISQEWQT FPNNFIETIRKNVHRFQTSKFQNPPKVSGGGI TTLPIFKRFQNVTRPPKVALPYKASLQNENSC VPSLIKEWFSNKNYITVLSPELTLNLIAHNNQFI	cd01650	1533	1797
	LIN2_SM	GCA_002600895.1	Schmidtea mediterranea						

					CTTTTGGAGAGTCTCAGAA	RENVTFRQFTRVKYVIAVIPNNPHWCLYIFSP			
					CGGCAAGCTCCATGTAGA	VFEHSYILNFQGRVLNERLVSIGESMTDYLNK			
					GCTGTGTCCGATATGAAA	FFISTAKVNFIVNHNQTHFSFRSEDPHFFASAF			
					AATTCAAAATTGTTATATT	IEFLERFFFGNILSFTKINETLKSFKLQITNEICK			
					AATAAATTTTTACTATAGT	NLNLINFSINDEIITKYFKNLNLPSNFIFLNSAM			
					ACTACTGAGTGATTTTGTT	CTAIMDDCTRYLNDFLNYDNIIKAQVIFAIVAP			
					AAGTCTACGACTTAGTCGT	PKFREILFIIDYNTDEYYLLNPTTLLGNDNYLST			
					AGACGTAGTTGTAGGCGT	CKILRNKINEIRESPNRVLNLGECPHDIRGSGL			
					TGTTGGGCGCTGATGGGA	VSKILFCSYMSNYANDLPLVGLNLRQIGNVVS			
					GCTGATGGGCGATGATGG	SFLPIVSEIKKVKSLENNIINQNHKESENLKERK			
					GTGCGGATGGGCGATGAT	NKIDQLLIKVFKMNVDEIVDVILDQISIKTVPT			
					GGGTGCTGATGGGTGCGG	HKPYLGNKERVESQINKSNLIKNFNSDMKTTV			
					ATGGGTGCGGATGGGTGC	NTIINDVDVDVRPKYENIITNFEHKIPLDRTW			
					GGATCGGTGCGGATCGGT	DKVFKFCNRSDKKLKLKYLSNSEVLFELKKADN			
					GCGGATGGGTGCGGATG	TSPGVDGIKYRDLVMLDPEGKLLTFLFNKIISK			
					GGTATGGATGTGTGTGGA	KVIPDSWKSFKTLLIPKPDKGDKYDNVSSWRP			
					TGGGTATGGATAGGTACG	IALLSVIYKLFASCLANRLTFWINKNNLLHIGQ			
					GATGAGTGTGGTTCTAGA	KGGSRHDGCVEHNSILSSALEHSKYSKNSPLAI			
					ATGTAGGTAGGTGTGGAA	AWLDIKDAFGSVPHAYMWSLLRFVGVDEEF			
					GTTGGTAGGTGTGTGGAA	VTILERMYSDTSTFYSCGPILTPNIAIRQGVKQ			
					GGTTCTAGAAGGTAGGTG	GCPISMILFALAINPVLEAVSRSKCEPFMIGES			
					TGTGGAAGGTTCTAGAAG	AVKILAYADDIALVSKSAYDLQRVTNIAVAIAS			
					GTAGGTGTGTGGAATTTG	EIGFEYRPEKCGYIELPLIETESQILINDIKIKKLA			
					GTAGGTGTGTTGGTCCAC	SSEFYQYLGVPVGEETDQSPYDILKKVVSDTR			
					AACTACGTGACATAATAAC	KIADSDLLGWQKLKAYKIFLHSRLVFPFRTREI			
					AGTTAGGTAATGTGTTGG	KTSALAKSNANNNLTVNITSQLRNCFRKMMS			
					TCCACGTGACATAATAAAG	LPNHSEVCYFYNSTENGGANCVDLLDEYHTQ			
					TTAGGTAATAAGTGTGTCT	TITHFFRLFTSSCEFARKVNLDSLNFVAGPRLG			
					CCGAGTAGTGACTGACCA	IINPTLQERFDWINGKEPKPRHSGKKTRFQRA			
					GTAAGTACGTGACATAAC	RIAIAFFKKEHDINIMFIVTKGQPSLFITSEKRG			
					AATGGTGACATAATAAGT	TFILTPDLRKTTSKVLHMALCDSYLSKWEKSCT			
					GTGTCTCCGTGTAGTGACT	SNFIASAIKLSPKINKAIYRGDISEFAWHFIHRA			
					GACCAGTGGTCAAGTGAT	RTNTLSINAKPHNKGNKRLCRLCHTEDETMS			
					GTAATCATGAGGTCACTG	HIIQSCKIHLTLGSERHNDCLDLISSHLSKNSNL			
					ACCAAAGGTCAATTACTAG	IVVVDHVCSLVPESKERVDLMITDMVRKKIFL			
					GTAGCAACTGAGTGGACA	VDMKCPCDTINNFALVDLLNLQKYDSLKMQI			
					AAGGTCAATCTCCAGGTA	QDAKPDFKVELDTCIIGSLGSFPPKTPELLLKIG			
					GCAACAGCGCGTGCAGAG	VDRTHLKGLLKDCALSNISHSARRWH			

				GTCAAATGTTACGTAATAA TTACGTAAGCCAAGGGGG GTTGTATCTGACGGGCTCA CATCCTACTACT				774
	Tx1-2_SPur	GCF_0000022 35.5	Strongylocentrotus purpuratus	CAATCCCTTGATAT GTGTATTGTGGAA AAGAACTGTTGTG AGAACTTTTGTTTT TGCATTAACGAATT CTCCAAACGTCGAA AATTCAGGAATGTA AAGTGAAGAAAAA AAATTCCCCGCCAT ATCGTACTTGGAG GTATAGATTTTAAG ACTCAAGAATGTAA GCCTTGTGACATTC CATGAGATGTAATC TGGTTGGAAGTAC CATGCAATTTAAAG CACCAGTCAAAACT CATGAATGTAAAG AGAAAATAGTACCC GCCAATTTGTACTA ATATAGGCATAGG TTTGCCTTATTGCA TTCCATGAGATGTC TAGCTTGTTACTTA ATCTTTAAAATGTG TCATTAGAGGAAT GTGAACACTGTACA ACGCAAATTATAGA ACGGGTTCCTTTTA AAGATATTTACAAA TGGAGAGGGGGGA GTGGTTTGGTGTGC ATGTGTGTGTGTGT GTGTAACTTACATA	TAATGGATGTGGGTATTAT TATTATTATTAGTATTATTA ATTCTCTCTCGTGTACAGG TGGACAAACCGGAACAGG TTGCATGCAGTGTGTGTGT CGGCGTGAAGGGGGTATG ATGAGATGGGTGGGCTAG GGGGGGGGGGGGGGTAA GAGTTTGTGTGGTGTGAG TTGAGGGAACAGGTTGCA TGCAGTGTGTGTTTGTGTT TGTATGGTTGTTTATGTGT GTGAAAGGGGGTGTGGT GTGGGTGGGCTAGAGGG GAGGGGGGGGGGGGTGTG AGAGTGTTGTGATAGGAT GGTAATATGTAATTTTAAA CCTTCCTCTACTGTTATTTT GATTATTCACTGTTACATG CAAGACTTTTTTTTCCCCTT TTTTTTCTCTATGTTCATT GATTGATGTAACTTATGAA ATGTCTTGTATCAAAGATC	MSVNLKLLSINVRGLRDRVKRKKIFQWCKMS ANMIFLQETFSTEDIEYSWKGDWEGSMFFSH GSNHSRGVIVMFKPGITFVVEKTLIDEGGRYV IIKGKCQGEKLILGSVYFPTRNKENAQCDFLSD LNKIICSVFSQEYSLIIGGDFNVIMNHNLDYM GASNISKSKYRDSLLDFLDKLDLIDVWRKRHP NKKEFTFRQKQPLVQSRLDYIFISSKVDKLIRKS EILPSITPDHAGVRIEFNNLVDSFVYGKSYWKF NSSLCNEKDFVDGVKKKISEIEETWRNQITSKI KFWDFMKMKLREYIMKFSRERAKLRRVHIAK LEKEIKDLEYRLTINIDCHKSIQDEIDMKKMEL EKLFDYSRQGLKVRSRTEWSEEGEKNSKFFEQ LLKSNKRKSVIKEIYNENKETVRDGKGILKIIREF YENLYACKNGEITEDIENMFLNELPKLSKENKE YCEGKITLAECYQAVKEMKWNKSPGNDGFS TEFYFTFWPILGKTIVEAFNESFETKMLSNSQR QGVITLIEKEGKDPLYMKNYRPITLLNVDYKIL SKVLATRVKEVLEDIIHYDQVGYVKDRNIGEA VRLIDDTFFQSLYQPVGFLIAVDFEKAFDCISH KLLFKALEQFGFGKVLCSWVKTLYTNVSSCV MNGGHSTGYFGINRGVRQGDPLSPYLFVIVIE LLARYIRKDESIKGVGFGNKEIKQVLYADDITLF LKDMDSINKVKKVFKNFEKLSGLKINMDKTN FMWLGEESDRPGLPFFGHRVQHIKILGVYFA RHSKIKDDLNYKEILSKIKRLIGWWKQRDLTM FGKIKLLKIYVFSKLNYVSSLMTVPQNIFEEIEKI SFNFIWGGRDRIKRKILYQDYCAGGLRVTNFK IFIKTQRIMWIKRLLYGEKYSGWKVTFDHFFR SLGGRIIFLCNYDINKMDLKGIPSFYKDMLKA WQELDKCRHFGENKNNPIIFSNKCICLSNKTF FDVELFQKGILQASDIIVRGHLKPILHFLNLGFE TKSLLRIQKIFSAIRNDWIQSDFMEVDTENFEI ELMIFGRLQKLCDIKSRRLYDYFIEDLQKEYTL HVTDGNCQFDLGMTELKETFTRMKCAMLC	cd01 650	520	

						NKHREFQYKQLQGAVYTKRELFKFGFESDHF CSFCRLEVETYQHLFLDCLKVKRLWQEVIQKL HLIEIDIDDWKPIFMGISGNTARIFVINCIIFLVK YIIFISRKEGTLPSVQKTVHMLKNYRDEEYKIAF KKGKLGMHLQKWEQIDAFLL	
				CATACATGTACATC GCATTTCTCTTTCCT TTTCCATTTGGTGA GCAAAGTAATCCAA CTAATCAATAGCCA TGAACACGTGAGA GGCTTCCTCAGCTA TGATTGGTCCGCGA GAGCTGGAGGGCG AAAGTGTTTCTTTT GTTCTTTTCGTCAC CGGCTTCTCGTGGT CACCAACGCCGGA CATGCTGATTGGCT GAATGGTCTTTGCG GCAGCAGGAGGGC GGGACGTGTTTTG AGACAAGTCTCGTC TCCAGCTGGTCCTA GACATTTCAGTTGA TCGAGATAAATGG CATCTGAAATATAC AACATAATTGTCGA TTTGTTTCTGCTCTC CCCCCCCCCCCTTT AATAGTAAATATCC ATGTAATCTTACAA TCTTAGCAGGAGG ATTAACGTTAGTGC CTTGTGAAGCAAG CATAATTAGATCTC GTCACCGGCTGCTT GTGGTCACTGTCTA TGACGGCAGGTGG GCGGGACGTTGTT CTTTTATTTAGAAA AGTCTCGTCCCCAG CCAGTTTTAGACAC			


CAGCGAGGTCTTTA
AATCAACTATAAAT
TTTATATCTAGACT
CTGTTTTGAGAGAA
ATTAAAAGGGGAA
AATAAATAATGCCC
TTTTACGGTTACCT
TAAGCCGGTCGTTA
ACAGTGAAAAAAA
GAAAGAAAAAAAG
AAGAGGAATAGTT
CAAAAGTTAATTAA
TTCGTTTTCTTAAA
ATAAAAATAAAATA
AAAGATGGGAGGG
GAAGAAACTCCTTA
TATACTGTAAAGTT
TATAATGGTATCAA
AAGCCATTTGGTG
GGGTACAAGCTGT
TGAAGCTTAAAGTA
TGCACGGGGTTTAT
TTATTACGTTCGTTT
ATTACGTTTTTTTTC
TTTTAAAATATTAT
ATATAGCCGTTTGA
GCCTGGGGCTAAA
GTATCATTAGTTAC
ACATTATCTGAATC
CGTTAATTATGAGG
TCAAGTTATATCGT
CTGGAGTTATGATA
AATTGCTCTGTTTT
ATTTTCGGTGTTGG
TTTTTTTTTTTTTT
TTTTTGTCACAGTA
AGAAGACTGTATTA
ATACTTTGTGTAAA

			AGGCTGCCACGCC ATGATTTAGAATAA GTGATGGGAGGGG TTTCGAACATGCCA GCATATATTATGCT TATGTGTATGAATG TGGTTGATAGTGTT GAAGGGTGGGGTG ATGGTATGTGTGGT ATGTTTTCCGGTTC CGATGTCTGCCTGT ACTTGTTTAATTTG CAATCTAACTTGAA AGCGTTTGTTTTTC CTAGAGTTGGAAA TGCATCAAGCTTCC CATCTACTGTGAAA CTGTCACCACTATG TACATTTCTAAAGG TTCTTTGTTACAGG TATAGATTCATAGT TCTTG	TGACTGGAGTCTGGAAAA ACAAACTGCACCTTTGCTC ATCCTAATTGTTCGCTACT GTCATTGTCACTTTTGCAC ATTTGCACTTTTAGAACTG CCACTCGACATCTTGTGTT TTTGTGTGTATGTT	MQVNKALRRINTRKATGPDNIPGRALRACA NELTDILTSIFNLSLSQRVVPSCFKTTTIVPLPKK SPPTCLNDYRPVALTPIIMKCFERVMLTHIQSS IPDTLDPLQYAYRPNRSTSDAIAAALHTSLSHL ENKDSYIRMLFIDYSSAFNTVVPHKLTHKLFSL GIHPIICDWLLDFLTGRPQSVRIGNRTSASIITN TGTPQGCVLSPILYTLFTHDCVASHKDNIILKF ADDTAVIGRITGGDEAAYRREVASLVTWCED NNLTLNTDKTKEMIVDMRKERRNPHQPLLIR GLEVERVSSFRYLGVHLSQDLTWTLNTSQLVK KAHQRLYFLRRLRRFGMSPKILSNFYSCMVES ILTSCITVWYGSSTAADRKRLQRVVKTASKITR TPLPSLQSIYKRRVCRRAASIIKDPTHPQRGLF TLLPSGRRYRSVKCRTTRLKNSFFPSAIRLLNS	cd01 650	55	309
	Rex1-2_OL	GCA_004348 135.1	Oryzias latipes					
	RTE- 1_DaCa	GCF_0016252 15.1	Daucus carota	CATGAGGTGATAG TCATGAATCCAAAT	TGAGGGTTATGGTGTATG TTTAAAGGGAGGGTTCAG	MANTIESMTFWGCLDDLVRSITLDQFLFVGG DFNGHIGARADGYQGVHGGFGYGVRNDNG	cd01 650	413	662

			ATAATAGATAAATA	CTCAGTCATTACCAGATGC	STLLEFATAHDLVIVNSCFKRDDNLITFRSGG			397
			AGATATTTTTTTAA	ATTAGCGCCAATCAAAAAT	HATQIDYLLIRSRDFRFCSYCKVFPVEACASQH			
			GGTGCAGTTCTGCT	GTTTGGGTAAGTCCCATGT	RLLVMDLSLDGPPMEGLRVVTPRILWRNLHG			
			CCCTAACTAATAGT	GAAGAAGGCCATGGGCTA	SKVAEFKDMLEGFFRLEDDVDQMWIRMAD			
			AGAAATAAAGTAA	AGGTATTATATTCTTATTA	LIRGVARRTLGVASGNVKEQRESWWWNDD			
			TGGTTATAAGGGTT	TAGCACACACTATTTTATT	VQAKVQFKKGCFLEFMSCPEGPDRHIKRELFK			
			CTCTAAAACATGAC	TGTTGTGCACTAAGCCGG	LAKRMAKQAVAEAKTKAYQDMYRRLGTKEG			
			CCAAGTTAATAAAT	GGGTCTTCACGGAAACAG	VNEIFKLAKARNKRSQDIGAVRYIKDEGDRVL			
			GTCTCTAACCAAAA	CCTCTCTACTTTAAGGGTA	LYDGDIIARWGRYFSELFNAARGREIVFEQEN			
			AATGTCTACATTTT	GGGGCAAGGCTGCGTACA	VNVSHNDVGMSQDITMAEVRVALGMMGK			
			TAATGTGTTTTTTAT	TCTTACCCTCCTCAGACGC	GKTVGLDEIPIEVWQCLGEQGVRWLTALFNV			
			ACGCAATAAATTAT	TGCTCACTGAGTATGTTTG	IFRTSRMPSEWRLSVVVPIYKNKGDAQSCSN			
			AACTAATTGAACAT	GTTTGGTTTGGTT	YRGIKLLSHTMKLWERVIECRIRRIVTISVNQF			
			CACTAAAATTATAG		GFMPGRSTIEAIYLIRCLMEKYRECRKDLHMV			
			ACACAATTGGGATA		FIDLEKAYDSVPRVVLWRCLAARGVPEVYIRTI			
			AAAAGTTAGGCAC		QDMYSAVGTCVRTPVGDTQYFPVEVGLHQG			
			ATTTGAAAACTTTA		SVLSPLLFIIVLDVITRDIQAPVPWCMLFADDI			
			CAAATATAAGGAC		VLIAESRSEVNVKLEQWHMSLEGYGLRLSRSK			
			AAATCCGACAGTG		TEYLWANFSEEIHEADVAVCIAEGRVPQTNTF			
			GCTTAAAAATCCAA		KYLGSIIQSNGDISADVTHRISTGWLRWWAA			
			ATTTGGACCAATTC		TGVLCDKNVPLKLKGKFYRVAIRPSLLYGSEC			
			TAGCCTAAATCTTC		WPLRKAQERRLETAEMRMLRRICGNTMTD			
			TCAACGGGGGCCT		HIPNDTFRRLLGVESISKKIREGRLRWYGHVR			
			GAATTACGGTCCAA		RKCNSAPVRRVEHISVRGKRKRGRPRRTWAD			
			ATTTAGGCCACTTA		QLSLDMGALNLTGDMTLDKNDWRRRIRVVE			
			G		TRSRGSQRL			
	Penelope-2_LG	GCF_000327385.1	TACGATATATAGAC	TAATTATTCATATTCAAAC	MRSTMILSNEFLTKSVINLSQKNLTDEQINLLS	cd00304	316	
		Lottia gigantea	ACAGCTTTATGGAA	AATTTCCAACTTTAAGCAT	RGLKFSPVTKLKNHDDFIVNIERGLHQLAPGG			
			CAACAATCACTGGT	AGTTAGTTATATGAACAAT	HINYLRHQVTEILAKNKIQPNNLTKVQRDALT			
			TAAAGTTTTATGCA	GTAATTTCTATTAATAAAA	DLRQDNKIAILPADKGKATVIMDKKEYQETVT			
			ACGTTTCGACGTGT	GACTCCAATACCAGCAATA	ELLSDAKTYRRLRKNPTSIYKDRLDKKLNELESS			
			TTCCCCCGTCGTCC	GTTGTTTATTTAATCTTCAA	NQITNQLKFKLRSSRHPRSPYARATIKIHKNPI			
			TCAGGCACATAATG	TTAATCTTTTTTAATTCATT	KARLLICSRDSVFYNIAQHLSRILAPLGKTADSF			
			AGAAGAAGTGATG	TACTTTATTTCAATTATTTT	IKDSEEFVTKLKELQITGKMYSYDVVDLFTNVP			
			GGCAAACTATATAT	ACTTAATCAGCCTGTATGT	CKDVLSIIRDKISKLETPLDTTLTTDSIVDLIQM			
			ACATTAATGCTTAT	TATTTTAAGAATTATATAC	CMSSTYFTWGDEIYEQIHGLPMGSPLSPLITEI			
			TATTAAGAACACAC	CTCTTATTGTACCTGTTAA	YMTAFEEKTMETSPISPICWFRKVDDTFVML			
			AAGAGATAAATGG	AATTTAATGTAATGTATCA	RAEDDPSALLLNAQHKRINFTMEAEEDDSIPF			

			TTTATTGATACATT	ATAAACCATTTATCTCTTG	LDVRIKRLGGQLKTCVYRKPTHTDQYVNYHS
			ACATTAAATTTTAA	TGTGTTCTTAATAATAAGC	NTPQHVKRAAITALVRRAKRICDPEYFQQELQ
			CAGGTACAATAAG	ATTAATGTATATATAGTTT	HIKKVFCLHNNYPPAFVHQAFRSVLMAKPTN
			AGGTATATAATTCT	GCCCATCACTTCTTCTCATT	QQPNPNRPSTRLAIPFIGKASYQIKKILKKSVGI
			TAAAATAACATACA	ATGTGCCTGAGGACGACG	ETVFSGNSTLKDLLRASGKLNPSNNQKPKGC
			GGCTGATTAAGTA	GGGGAAACACGTCGAAAC	VYQVECSCGSSYIGETKRPLDVRLKDHQDKP
			AAATAATTGAAATA	GTTGCATAAAACTTTAACC	GNEGTSAVAEHLAMGRKTGHKVNFDNTNIL
			AAGTAAATGAATTA	AGTGATTGTTGTTCCATAA	TINKHIDKTRRLHEAIHIRRYQPDLNRDEGVRL
			AAAAAGATTAATTG	AGCTGTGTCTATATATCGT	SHAYDKLINL
			AAGATTAAATAAAC	A	
			AACTATTGCTGGTA		
			TTGGAGTCTTTTAT		
			TAATAGAAATTACA		
			TTGTTCATATAACT		
			AACTATGCTTAAAG		
			TTGGAAATTGTTTA		
			TTAAATACATTTAT		
			GAATATGATGATAA		
			GTTTATCATATGCA		
			TGCGATAAACGAA		
			CACCTTCATCTCTG		
			TTAAGGTCAGGCT		
			GATATCTTCTGATG		
			TGGATAGCTTCATG		
			AAGTCTCCTGGTTT		
			TATCTATGTGCTTA		
			TTGATTGTCAGTAT		
			GTTGGTGTTGTCAA		
			AGTTGACTTTATGT		
			CCTGTTTTCCTTCCC		
			ATAGCAAGATGTTC		
			AGCCACAGCAGAT		
			GTTCCTTCATTGCC		
			GGGCTTGTCTTGAT		
			GGTCTTTGAGTCGA		
			ACATCAAGGGGTC		
			TCTTTGTTTCCCCTA		
			TATAAGAACTTCCA		

				CAGGAGCACTCAA CTTGGTAGACACAT CCCTTAGGTTTTTG GTTGTTGGAGGGA TTCAATTTACCACT GGCTCTGAGCAAA TCTTTAAGCGTTGA GTTGCCAGAGAAC ACTGTTTCAATGCC GACTGACTTTTTCA GGATCTTCTTGATT TGGTATGATGCCTT TCCGATGAATGGG ATAGCTAAGCGGG TTGATGGTCTGTTG GGATTAGGCTGTT GGTTGGTTGGTTTT GCCATTAATACGGA TCGAAAAGCTTGAT GTACAAAGGCTGG AGGGTAGTTGTTGT GCAAGCAAAATACT TTCTTGATGTGTTG TAATTCTTGTTGAA AGTATTCTGGGTCA CAAATCCTTTTTGC TCTTCTGACCAGCG CTGTGATGGCTGCC CTTTTGACATGTTG AGGGGTGTTGGAA TGGTAGTTCACATA CTGATCTGTATGTG TAGGTTTCCTATAC ACGCATGTCTTCAA CTGACCGCCTAACC GCTTAATTCTCACA TCCAGAAAGGGTA TAGAGTCATCTTCT					

								775
							516	
						cd01650		
					MIPELCSTGLKICHININSLPHKMDELKILLGN DPFDIVAISETHCDSTITDLDISLDNFCVTRKD RSRHGGGIALYVKNNISYEVISQLDNELEVIFV KVKQKQTHPLYLGVVYRPPNSPNEFFDKLSEII HSISENAGEIIIIGDFNCDLLPEICTPFSRALLSL MDTNLLHQLISVPTRVTPHTKSLIDHVFSSIPE EHQHTGVVKTHISDHYLIFTVFGYARAKQKN DDNIIEYRCFRHFSNDNFVTMLQNLPFNENL NDKDPNHAWQCWVDLLISAVNKHAPVKSK RIRSKLCPWITGDLIEAMNRRDWLYTMATKP SNPDSSEFWSEYKSLRNEITCMIRNAKNDYIN ALLHENGGTPNAMWKTLKLISSSKKESNLKLE VDGEEVIDAEKIANHFNNYFVDAVDTMFNPL NFENDVNNSTIDRASEPSVHSDTVDTVPPVN SFDLPVITDEFVMKEIERIPIKKATGPDDVSIKL LKFACNSQNVIQSLTYILNSSLDQGIFFDEWKI ARVQPIYKAGSKLSVENYRPVSLLSIPSKILEKA		
				GTCTGGAACGCTCTTGTTC ATTGCCGTGCAATATGAAT GCTGCAATCTCCTTTAACA TTTTATTTTAATTAGCACAC GCCAATAATGAATCTTGCT TACTTATTTTATCAATATGC TAAACAACATTATCATTGA TGTTTTTTTTTTTAAACCTC AATCAAGGAATCTTTGAAA TTGTATTTCATGTACTTAA GCAATACCATGTTATGAAT TGATTATGATACTAACATA TGCATGTTATTATATTTAT GTGCATAATATGTCAACTA TTTGTTTTTATAATTAATAG TTTAAAGTATTAAATGGGA AAGTATTAATAGTATATAT	TCTGCTTCCATGGT GAAATTGATCCTCT TATGCTGAGCATTA AGATGCTGAAGTA GAGCTGATGGGTC ATCTTCAGCTCTAA GCATGACAAAGGT GTCATCAACTTTCC GGAACCAGCAGAT TGGTGAGATCGGA GAGGTTTCCATGGT CTTTTCCTCAAAGG CTGTCATATAGATT TCAGTAATCAATGG TGACAAGGGGCTT CCCATAGGTAGACC GTGGATTTGTTCAT AGATCTCATCTCCC CATGTGAAGTAGG TAGAGGACATACA CATCTGA	TCCGTTTCTGCTTA ATAAATGTGTTAAT ATATTTCCAGAGTT TTGTCTGGGGCTGT CGTATTATCAAGCA ATATGCTTACAACG ACAGCCCTTCTTCC ACATATAATTTTAT AATGTGTCATAATG TTTGATAATCCATG AATGCAGAAGAAA TCAAATCAAAATCA AATCAAAATTCATG TATATTGCATACTT TAACTTCTGGCCAA TCAGCGTAGTTGGC CTTTTATTGTCAAC	Strongylocentrotus purpuratus		GCF_000002235.5		Crack-23_SP	

		GATAACATCATCCC	GGTATAACGTAATGGTGA	VNTEFQEYLKVNKILTERQFGFRPNHSCETAL	
		TGGACTACATTTGG	CGGCAAACTTTTTTAACAT	LCMVDLWAQNVDKGNLNGVAYIDMRKAFD	
		CGGCAGATGGCGT	TGAAATATATTGCAATAAT	AVNHSTLLMKLKNVGCTDRAIKWFRSYLGGR	
		GCATCTTAACAACC	TTGTTTTGATATTTATACCA	SQFVSIKGKKSSTRTINYGVPQGSVLAPLLFSIF	
		ATGGAATACGCGT	TACTTTGTTTTAACTTTAAC	INDLPTSIHSADMFLFADDATMSVQGKTLSEI	
		AATGGCGCGAAAC	CTAATACTATGTGTTATTG	QTKLNLALDEVYSWTQKNKLLLNTTKTKVMV	
		CTCATCACTCGTCT	CGTACAAATGTCAACTACA	IGSKQKLQTLNENNLNVRIKTTPIECVTQYKCL	
		CCGAGATCAGTCAA	TGTATATATATATATATTTC	GVIVDSNLLFSKHVEKVALQMKQKLGILRRLK	
		GCGTCACTAGAAG	AATTGTGGGTTTTTGTTGT	GTFNKHYLSMIYWGYILPHAMYCCTVWTNR	
		TAGCCCGCCGTCTG	TGTTTGTTTACAATAGTTTT	SLHNYETINKLHKRAAYIISGCSWSTPSHEVLL	
		ACCCACTGCCCAAT	AATTATAAACTGGGAAAG	QLNWSTLEALYSRAMGIMAFKCVHHLAPALL	
		CAGCGCCCGCGAT	TATTTAATTGTATACATTTT	ANKFILHDDVAQRMTRNSHELKLESNRCNTE	
		CTCCGAATACAATG	ATAATTACTCAATGGTGAC	FFKRSFVNSAITLWNNLPIETRFSPNLNVFKSN	
		AGGACCGATCCATT	GGTAAACTTTTTTTAACAT	LSH	
		CTCGAATCGGCAA	TGAAATATTCGCAACAATT		
		GACTCCGAAAGCTC	TTATAAATATCTATAGCAT		
		CTCGCACGCCATGG	ACCTTATCTTAATTTGAAC		
		CTGCTGAGAATGAT	TTCGTATACTATGAAATGT		
		AACTCAGCTTTTCG	ACTATATCGCCTTTTTATTG		
		GAGACGCCAACAT	CATGTCCGAATTATCTCTC		
		CGCAAATTCCGCAA	ATATTATAATGTGCATTTA		
		AAAGCCGAAGGGG	TTGTTGTTTTGATTGACAT		
		AGAGTTTTCCCCCG	TATTATTGAACCATGTATG		
		AGACTGGCTCGACT	TTGTAAAGGGCCCTGATG		
		GTCTCCAGACGGCC	AAAACCAGCTCATGCTGAT		
		CGCACGCTCCTGAA	CGGGCTACCCTTTTTAAAG		
		TCGTCAAGCTCAGC	ATGTGAAATAAATAAATA		
		AATAGATGATAGC	AATAAATAAA		
		GATAATTTAAGTCT			
		TCACGCGGAGGAT			
		CTTAATGTAAACGC			
		TACAAATGTAATAA			
		ACATCAATGAGAA			
		ACAAACGCAAATTG			
		AAAAAACTACAGCT			
		TCCAAGGATGAAA			
		GTGAACTAAATGA			
		AGGACACAATGTG			

	HEROTn	GCA_000180735.1	Tetraodon nigroviridis	ACAAATACACAAG GATATTATGAACTT GTTGATGGC	TGATCACTCCCAGTCGGGT CGCCTGGGTGAGGGGGTC TGATGTTGAAAGACCCGA AACCCCCGATGACCCCAG GTACTATCACTGACGATGT GTCCAAGACATGCATCAAT AGGTGTATTTAGAAATC	MATTQASVKPTAVATCVCGKICKNPRGLKIH QTKMGCLASVQPEQRARFSLSESREVPARAE PYGPQQPHSPEALGETQEERGQESPHSAQN LRAQVAQAPDNPQHHRRVKWPPASKVSEW QQLDEDLEGILESTAKGGVDRKLQTMTTLVIS FATERYGTMEKRAAPEKYTKNRRAEKISQLR QELRVLKKQFKGASEDQKPGLAELRCTLRKKL LTLRRAEWHRRRAKERAKKRAAFLANPFGFT KQLLGQKRSGHLECAKEEVDSYLHDTFSDAER ENSLGECRVLISPPEPACSFNTKAPTWKEIQT VVRAARNNSAPGPNGVPYLVYKRCPKLLARL WKILRVIWRRGKVAHQWRWAEGVWVPKE EKSTLIEQFRTISLLNVEGKIFFSILSHRLSDFLLK NQYIDSSVQKGGIPGVPGCLEHCGVVTQLIRE AREGRGSLAVLWLDLANAYGSIPHKLVEMAL ARHHVPGPIKTLIMDYYDSFHLRVTSGSVTSE WHRLEKGIITGCTISVIIFALAMNMLAKSAEPE CRGPITKSGIRQPPIRAFMDDLTVTTTSVPGC RWILQGLERLMTWARMRFKPGKSRSLVLKA GKVTDRFRFYLGGTQIPSVSEKPVKSLGKMFD GSLKDAASIRETNDQLGHWLTLVDKSGLPGK FKAWVYQHGILPRILWPLLVYEFPISTVEGLER RVSSCLRRWLGLPRSLSSNALYGNNNKLTLPF SSLAEEFMVTRAREVLQYRESKDPKVALAGIE VRTGRRWRAQEAVDQAESRLHHKELVGAVA TGRAGLGTTPTTHLSRLKGKERRDQVQLEVR ASIEEQRASQWVGLRQQGAWTRWEEAMA RKISWPELWRAEPLRIRFLIQSVYDVLPSPSNL FLWGKVESPSCPLCQGRGTLEHILSSCPKALG EGRYRWRHDQVLKAIAESISSAMEYSKRLPLP GRGVRFVRAGEQPPPQPRAQPGLLATARD WQLRVDLGKQLKFPENIVETNLRPDIVLHSQS SKQVILLELTVPWEERMEEAYERKAGKYAELV EDCRRAGWRSRCLPIEVGGRGFAGKSLCKAF	cd01650	362	627

					SLLGITGMRRRKAICAASEAAERASRWLWIQ RDKPWTSASWTQAGN				
	CR1-65_AAe	GCF_002204515.2	Aedes aegypti	C	TGACTGTATGTTTAGCTTA AGTTATTTATTTATACATTC ATTAAGACAACTTCGTGTC AGATGGATTAAATCAGAA ATACAAATACAAAT	MNDHLSANMHRTLSHSGRTTSSTMEAPTPP STVESLQPALISRPDPVVGRGGGVFQPVLSGK YNKITDSLLPDISPPSSLRIYYQNIRGLRSKIDSF FLAISELDYDVIVLSETWLDDCILSSQLFGSNYS VFRTDRSALNSDKGRGGGVLIAVSSRISCSVD SAPVSNQLEQLWVKIVLPRNNVSIGVIYLPPD KKNDIDTIQRHIESIGAVHSHLHENDMAMLF GDYNLSDVHWTSDANDNVFVDLTRTRLNPA SSTLLDGFCFHGLSQVNTIVNSGGRTLDLVLL NDVVLPNHSIHEAVEALTTVDVHHPPIELVVS CPLPVVYETVSDNTGFDFKKANFESLNVALRS MDWDRLDSIEDVNEAVAFFTNSVMQAVAE HVPPRRPPSKPPWSNSRLRLLKRRRSAALRFY RSHRSQQSKLHFNRSSSEYRSYNAYLFAQYKS RTEQNLRSNPKQFWSFVNTKRNEDGLPTSM YLDELSADSASDKCELFAMQFKRAFNDYASP ASQIDRALQDTPQDVFSYDMFHVSEEEVIGAI GKLKFSYNPGPDGIPPALLKRCSSTLLAPLTKLL NHSLRQQVFPACWKKSFLFPIHKKGDKRCVS NYRGITSLCACSKVFEIIVNDSLFESCKRYISTD QHGFFPKRSVSSNLVEFSTLCIRAIDAGKQVD AVYLDLKAAFDRVDHRILLQKLRKCGVAINFV DWFRSYLTERSLCVRIGESESASFTNISGVPQG SNLGPLLFTLFINDVSLLLPTGNRIFYADDAKIY MIVESLDDCYRLQHLLNLFEAWCTRNCLTLSI EKCQVISYGRKRSPIRFPYELSGTTLERVKCVR DLGVTLDEKLTFNCHLNDVISRANKQLGFVLK VSSGFKDPLTLKALYCALVRPILEFAAIVWCPF QNYWISRIESVQRKFVRHALKNLPWRDPENL PPYHERCRLLGIDTLENRRHVAQTMFVAKVL RNEIDSPSLLADVNIYAPERTLRRRQFISLGSR NTRYGQHDPVRFMALKFNEAYHLFDFNVTIT TLRNHFIQYFRRN	TAAAATTAAAATAAATATA CTGTTAACTTAGCTAATTA TAAAAATAAATAAATAATT	MKINQANWDKFATHTNNNNNLYPPSTNINK ETAQINKIIMKSAHLAIPQTLPHRHTHTVPW WNKDLADLQEKKKQTWKALKRDTTQENIIKY	cd01650	586	838
	I-4_DAn	GCF_003285975.2	Drosophila ananassae	TT				cd01650	260	518

								1054
							788	
						cd01650		
				AAAATACGTAGTCGAAGG CCTCGTAGCTAGCACTACA CCCCAAACTGTTAGTCATA GTTTTGTAAACTGTACTAT CATATCCAATAAATAAATA AATAAATAT	RKTNALFKRAVKIRKKEATSKFTSTIQPNTPTT KIWNNIRRFCGLNPVKQIHCITNNNVTTNDK QKIANLLAQHWSQLSKDSNFSQEFTSHKYNT TTLNHIPSKSAMEIDKPITKLELTIALRSLKGITP GINRISYPMLKNSPLAIQNRILSHFNHIFENYIP QSYKTSLIIPILKAQKDKTDIASYRPISLNCCVG KTLDKIIANRLWWYLNSNKLLHPHQYGFKKG KSTSDNLLYADHIITKSLNTRKHTSLISLDFARA FDRVGIHSIVDQLIEWKCSHKIINYIINFTKNRK IAVRIGSSISNFYPLENGIPQGSPISVVLFLIAFN KLANILTIYKEIKFSAYADDFLLLVNLKNQKNTT INLTHIFRKIYDWGSYSGASLSQTKCQHLHICR KHRCNCTITTQNSNIKNVTSLKILGLTINNRYN WSSHINSLIPSLSNKHNIIKCLASHKFNCNTSTL LNITKSILISKIDYALHIYGYTPKNFLNKLKTTLN AALRTVLGAFRTTPTHNLLYEANIPTIEQQRD HLTTKLFTNLTRSKDTPLHKHIQPKTNNKKWT STLERTIENCQLHNIPFKPTKLYKTKPPWQLP QNIINTSLHNYKKTNTSSDTFQKLFFEIKETYK NHTFIYTDGSKINLNTAFAVTTETDVLRYGILP PSSSILTGEMIGILEAIKISTTKRGKSVICSDSLG AIEAIQNPLNLNKYPTYIRSLLQTKQPKIKILWI PGHSNIKGNDLADRTAREATKMPLIYTPNENI TDINKMLKNILIQEKETLINKCNTWYKNINTN NTHISALSKIQQEEISRIDQIKFIRLRLGHTSLTH KHLLDPNTQRNCYRCQSSTPTNMEHILDSCP AISQNKQNLLKNRPATEYLKNPTAENIKVIIKF LKETNLYNEI	ACATTTTAAATACTAACTTT TTTTATGTTTGTATGTTAGT GTGTTACAGCTGTTAAAGT TACAAACTCTACGTAGTTT GGACTCATCACTCTTTAAG TGTACAGCACTATTATATT ATATT	ATGGCGGCGTAAA GATGGCAGCCTAG TTGAGCACGTGTTT TGAAATAATTTTTA ATA				
	CR1-83_HM	GCF_0000040 95.1	Hydra vulgaris					

	CR1-22_SP	GCF_0000022 35.5	Strongylocentrotus purpuratus	TATTGTGAATGAAT TTAATGACTTTTTT GTAAATATTGGACC GAATTTAGCAGAT GAATTGCCGAATAC AGATGTAGCATAT GGTGATTTTTTGGG GCAGAATATAGAG CAATCTTTATTTTG	ATAAATATAATTCATAATT CATATTTTACATTATTAGTT TGTATATTGTAAGACATTG TTGAGTACTGATTCTTGTT TTTTTTAAAGTATCGATTG TAAAAGTAAATTCATTATT TTTTGTTTTGTTACGTTCTC TAAACTGATTTGTACCAAA ATACTTTGTTTTAGCTGTTT	MIPVQLIRDSAPYISHPLAYIFNLSLETGKVPDE IKIAKVTPLHKKGPREDPGNYRPISVLPVIGKV LEKLINIRLIQFLEKNDIFYYHQYGFRKKYSTKLS LINLINSTLKSIDQGKVTLGMFIDFKKAFDTIN HNILLGKLSSYGIRGLPLEWVKDYLSNRFQFV QCGQTLSERKKIRCGVPQGSVLGPTLFLIYIND LPTVTDYFNFRLFADDSNLFHTFSADQGEIDV NEVNNNIQKIHKWCTANKLTINSMKTNFMLI CSRRKKIGINGKLKIAGEEIKEVDVATFVGVHI	cd01 650	37	295

			GAAACCAATAACTG CAGTTGAAGTTAA GAGTCAAGTTCTGC TACTTGATACCAAA AAGGCTTGTGGTC ACGATAAC	AAGGGGTTAGGCTCGACT AGCACATGTGCTATTATCT GGCCCCTTTTACCAACCAA TAGTTTCTCTTCCTGTATAT GTAATTATTGTATAAAAATG TATGAAATGTATTGCTGTT GGTAAATAAAA	DRHLSWKSHIRFVNQCIRKKVGVLFRLRHFVP SRTLILLYKAFIQPHISYGIEVWGSTYKTSLRSIL YSQKMAVRAITFSEWRTPSDPLFKRLRIILNVY NLHDLSISTFMFDLVNGNLPHILIEYCDIIEHSY STRQKEDGQLRLPKLRTTQGQFSLSFVGSEF WNRLPVDVRLTRSRNRFRNSVTSHLLNKS			812
	Daphne-17_SP	GCF_0000022 35.5	Strongylocentrotus purpuratus	TCCATTTAAGTTCTT CTTTACATTTTGTAC ATAGTTGTTGTTGT TGTTTGCTTTTTACT ACTTTGTTTTAAGT GTAGAATTCATATT CCTTTGCTTTTTTGG TTTTCAA	TGATATTAGCTACAGATTT TAACGGTTGTTAAATTCAT TATATTTTACATGTAATTG TTGTTTGTTTCACTTTTTAT TACAGCTTAACTTTTTATAT TTGATTAATTGGTCAGCCG CATACAAGGACCACCTCCT TTCTGGCTGCTACCAGCAT TAGCAATTTCCTTTGTATT ATTCATGTTTGTAAATTGA TTCTGTGTATTTCTGTGTTT ATCTATGTTAGTAATTGTA ATTGTGTGAAACTTTGTAT ACTGCTTTTGTTTTACCCTG TATGTTTTGTATTGCTTGG CAAAATAAATTGAATTGAA TTGAATTGAA	MLCRARFVQCLNEKHVSVSTLREFLSPSDPQS TNNFTQSHYTPDEFNTFVTQKSLDKDVNNKL SFINICRSLQHNFDALSVLMMSLRMQVSIIG VSETWLRDENSQYSLCNLDNYVFVGNNRRL GRGGGVGIFIRKDINFKRRTDLDINESFLESIFI EIERFNSSILICVMYRPPSQPVDDFLMSLVNVF DVVNVEKKLFYLMGDFNIDLLNVGKVQSVN DFLNILMLHSMIPLINEPTRVTDRSSSLLDNIFT NDSCYIYSGILFTDISDHFPVYCICDTDIPKSSCE TDFYKRHLNENNIRMYFLQLLYSTEWYLSSDV NSSYNNFVFKFMKVYNLCFPVKLVRMKYKRK NKPWITNEIIKLSKKKCRLYKLLIKNPNVYRKEV YRKIRNELNNKKIKIEKKNYLKKSFKDSTGNMKK TWDIINKAMGKKCKMTNIKEINNNDTVITDK NDISNNFNEYFVDVGYNLNCKFDCFDDSECIK YIGMNDKSMFFKLVTSSEIEDIVKKIKNNTSA GYDDVDVKIVKKCIYLLSSHLCTIFNQCLEQGV FPDKMKIAKVIPVFKNGVSETMSNYRPISVLT VFSKILEKCMYNRLYAFLTKCNIINDNQYGFR AGHSTSCAIIDFIHKVASAIDNNEVMLGLFLDL SKAFDTLDHHILLVKLYQYGIRGIGLDLFKSYLL HRHQIVSIANVKSNYQEIKCGVPQGSILGPLLF ILYINDLPSVSRILKSILFADDTSIFLSHQNIQSL QDIFNQEIKLINTWFSVNKLSINISKTHYMIFT NKKCDTDSIIIRLCNTEIKQVSSLKFLGITIDNKL KWTDHINAVCNVLSKNIGVMYKLKSFPKCILK MIYNSLILPFLYYGILAWGNAHETHIRKVILLQ KRAIRTICHTSYLAHSHPLFKEMKILNFTDMY NYQCAIFMFQCYLQILPKSLLKLFILFNCNIHHY STRNAENFHLPLVHTTSYKRTVIFNGPCLWST IPRNIRELKFVNTFKRKYKKYLMDVY	cd01650	557	812

	R2NS-1_SMed	GCA_002600895.1	Schmidtea mediterranea	CCTTTTTACGGTAT AATTATAAACATTT ATCATAACATACTT TCTGTATCAACTAT ATATATGAGTATAA ATCTGGATCTTCTG AGAAGGACTCATAT ATATAAGATAATCT ATAAAACTGGAAG GTAATGACGATGA TTCAGAGGTCCAGT GACTATTCCTCGAG AGAGGAACTAGTT ACGCAGTTCTGAAT CGGCAATCACTAAC TCAACAGGCTGAT GTCAGGATATGGA AAGGGGTCTAATA GTGGTCTTAAAACT CTCTTTTGAAAGTG ACTTCTTATTGGAC ACCTTTTTATTGTAT AATTATAAACATTA ATTATAAATCATAT CATACTTTCTGTAT CAACTATATATATG AGTATAAATCTGGA TCTTCTGAGAAGGA CTCATATATATAGG ATAATCTATAAATC TGGCAGGTAGTGA CGATGATTCAAAG GTCCAGTGACTATT CCTCGAGAGAGGA ACTAGTTACGCAGT GCTGAATCGGAAA TCACTAACTCAACA GTCTGATGTCAGG	TGAGTAGTGGTATATTTAT AAATATAAGTAGCGAATTT ATAATATTTATCTATACTAT CTTTATATATTTCAAGTGG ATCTTTTATGACTAATTAA AAATACTTTTTTATATTTTT AACTTTTATTACTTTATTTT CTATTTTTATATTCTATGAT TCTTATAATGTAATTTTATT GGGAAAGAGAGTAAACG GGGGGCAATGGGGTTAGT CTCAGTATGGATCCGGACC CTCCGGAATCGGGATTCA ATGACTGGGGCAGTAGAT GCGGGATGGAAATTAAGA GTACTTTGGCATTAGGGTT TCTGGGACTGTGAGTGGT GAAACCGTAAAAAGGGAG TAAGAATCCGGGCTATTAC CTGTACGGCTAGAGAAAT CTAGTTCAGCCTTGATGCC GTATTTTTTTGGGAAGTTT ATTCGCACTGCCGTGGTTT GCGACCAAGTTTGATGGG AAGACACAAATAGATTTA GTTTTGGTTCTTTTGCCATT ACTGAAATCTACAACGAA GCCACCGGCTATAAAGAG TGGTACTCTTGTGGTATAC CTTTTAAGCACGGTAAGG AAGTGGAATGGATGGGTT GCTCATTCCGGACTGCGC GTTGGAGCAATTATATTGA TCGGGTGATACTAAAAAG CCGGGAAATATCAATTACT AGTTTTCTGGAATACTGCT GCTCACTTACAGTTTCTGG ATTACAAATTTACCCATAG	MAIITGEGDCHLLKKSLLKLKTMKNKTNKLNV NKGNDLDVTFPCDFCGLSFPTKIGRGVHISKA HRVERDAKLSGHASAKKAYYWSTEEETRMA KIHLQEESRGDSNHVLRAMHSACSTGRTLESI RLHMKAKRYVKIYTELRNEKQIQSSGVENILN NISDASPLTLPSETETLTSYLLNLPLKKGVSKW REQDLDFIIHSVSDSEEQKRLLDVYCKTVTEGY RKGKTGPLKKSTKVSPHHATTTKSKKFLKKKLK RELYGKVQQLWIKDRCEAAKLVLSGKVSTLIC EEDRSARQGEYWKKIFERESEPDIRPVPEVYA DPMLVRPISEDEIKKVVKDIKLKTSSGPDNIST EVLREIPVLVLVKLFNLIILLSHLPSEFKSSRTTLI PKVETPKDPGDYRPISVSSVFVRALNKILASRV SNGNKGAHAQKGFKALDGCLENTVLLNCILG DAKINNKSIALSFIDMKKAFDSVNHDSILRAAT IAGYPKLLLGYIKESYNGADTEVEKIKATFKRG VKQGDPLSPVLFNNIVELAIKKAYESGIGYRM NGTVFPIMAYADDIILLSESKEGLQDLLDIVLTE LRHNGLEANADKCASIIAKKDGKQKKWVISS DSVLAAGKTMPCMNANSMYKYLGIYFGPLGI TTERLTKIYSNQLYNIGIAPLKPQQRLEMINDF LLPGILHEAVLGNTNVGELKDLDKMTRVAIRK WLHLPKDTCSSFIYTGVKQGGLGVRSILAEVQ FGKISRLESLNKSKDTRIQALTTCEVVKKAFRQ WSKPIRVGDSIATSKVQFYEANRELLLSKCDG KGLREAGKIPFAHDWLRAKGKLMKGRSFINSI KVRGNLFGTPVRLNRLNKSGIPLIKCKMGCL MPCSLNHIAQTCPKVHGLRVERHNKVMNYI ADVARKGEWKTLVEPIIPYGATIRKPDLVVIKD KKAMVLDIQVVSDVGNLNERNTLKSEVYNDK IIADWVKNAGADSVEFGAITLNWRGLWASD SWETVSKIGVNKSDAMIISLKVLNWTARMYC AWNKG	cd01650	385	643

		ATATAGAAAGGGG TCTAATAGTAGTCT TAAACCTCTCTTTT GAGAGTGACTTCTT ATTAGACACCTTTT TATTGTATAATTAT AACCATTTATCATA TCATACTTTCTGTAT CAACTATATATATG AGTATAAATCTGGA TCTTCTGAGAAGGA CTCATATATATAGG ATAATCTATAAAAC TGGCAGGTAGTGA CGATGATTCAAAG GTCCAGTGATTATT CCTCGAGAGAGGA ACTAGTTACGCAGT GCTGAATCGGCAA TCACTAACTCAACA GTCTGATGTCAGG ATATGGAAAGGGG TTTAATAGTGGTCT TAAAACTCTCTTTT GAGAGTGACTTCTT ATTAAACACCTTTT TAATGTATAATTAT AATTAAATATAAGC AG	GTAATTGATATATAGCTGG CTGTAAGGGGTATGGGAC CCTTACAGGTATCATTATT AAAGGCTTGGATGATCAG ATGGGTATATAAGGGAGG GCTATCAAGTTACTGGTTT TCTGAATGATATTTGGATT ACATTTCTATCAAGGTAAC AGCGACGTCTATAACTTAT ATATCCATACGAAATGGA ACGATCAGATGGATAAAC AAGGGATGGATTATCAAG TTACTGGGTTTTCTGGATG ATATTCGGATTACATTTAT CTTAAAGGTAACAGCGAA GTCTTGAACTTAACTTTCC ATACGTCTAGGGAAATTC GCTAATAGAAATATTAGTG ATTAATAACCACACGGGCT AACAACCTCTAAGGTTGG GCAAGTGAGTTGGATTCA TCCAGCTTGCGCGTATTAT ATTTTCGATCCGATGTCAC CTGAGGTTACGGGGGAATT AAACCACAGCTTTTCTTAT ATTACTGAATCAGGATATA AGGATTGCATATTAAAAAC GTTGGTGGGTTAATGAACC TATAACTGAGGAGTAATC AAATAGTAATGATTTGAAT GCAAAAGGTTTCATTTGTT ACTGATTAACTTGCGGGTT TCATTTAAATATCTTATTTT ATACATGGGCAAGAATAA CAATGTGAAGTAATTCAGC TCTAGTAATCTAATGATAA AGCAATTACCAAGGTATA ACGCCCTATTTTGTAGGGT			

								742
							487	
						cd01650		
					MATNDNDDQFVAEAQGRTHGSLSGPALSIIS SNIEGLSGAKQELLADLCITTNCDILCLQETHR GPTRVRPRISGMTLIAERPHDQYGSAIFARNG LAVDKSYLSDINNIEVLSIDLMGITVSSVYKPP QEPFDLPGRATDGRMNVVIGDFNSHSISWG YREMNEDGELVETWADANQLSLIHDAKLPAS FNSARWRRGYNPDLIFTSTRIANQCEKVILDPI PHSQHRPIGLKVNAVITPRAVPFRRRFNLRKA NWEKFAKDFDSYIQDLPATPQNYDTFVELMK KSSRRNIPRGCRTSYIQGLTDHSKVMYEDYIE HFQMDPFSEETLESGENLTKEIGESRRRKWQ EFIQSTDMTHSSRKAWKTIRILGNDYTKAQPR		
				TAGTGACTCGATAAGAAG AAGAA				
				AGGCTCTGTATAAG CTTGGCTTGTTAGA GCTCTTGTGAGTTG TGTGCCGCGGGGT TTCATATCCTTAAA CTGCTGGCGAACA GAGCAACAAGACC AAGCTCTGTAATTG TTAGGTTTCTCCAA TGGTAGGTTTGTCC AATAATGCCGAGTT TGTGCCACGGAAC				
				TGGCCAAATATAACTTGAT AATAAATGCTGATTATTGA TCTACTAGGTATCAGGGT GCAATGCCTTACGAAGGT AAGGTTGGACAAATGTAA CTTGGTAACATGTATAACT ATAAGACCCCCGTTCAGG ATAGTGGTGATTGATAGT GGTGATGTAGTTAATACA GGCGTGTATGAATCTGGT GAAGGTGGGAAAAACTTC ATAACTCCGGTTATGATTT GAAGGTCAGGTTTGGATT GGGTCTGATCGGCTAATC AAGGTAGTGATGAACTCA AAGTAGGGGTCTATGGAG AGGGAGCACTTGATAAGA TAATTCACGATAAGATGGT TAGTAAAAACTCGCGATCC CTGCGTTGGAAAGTTTTGA AGTGCTGAAGGTTGGTTA GCAATAATCAACAATGCTC GTTAGCGAAGGGAGAATA ATTCCGGGCTATCATTCAA AACTGTATTTTATCAATGA ATTATT				
	Vingi-2_SPur	GCF_000002235.5	Strongylocentrotus purpuratus					

				CTAGTCGGACATTTCGATTATATTCTTTTTATGTCATACGTCTTTATCTTTTTTTTTTTAAATATTTTTTATAGTTTAGTCTACCTTGCTGATACCTCCTCAAAATATATACGATAGACTATCATTAATTCTGCTGCATTACCATATAGCTTACTTAGAGGTTACTGCAGATCAATCACCTGTACTGTATCTAGAGTATAGTTAAGTGTAGGAGAACAGTAAAAACAGTGGGGCCATTGAAGGAGCGTTCACGTCGCGGTCCGTCAGATGAAAATCTGCGAATCCTTCAACCACACTACTCAAGTTTCTTTTAGTAACTACCACAACCAACACC		SEVTADQVAHQLLLNGRGNPDHHPRRARLPKVPDALSSQDDPGFTRPFTMDDLLAAIKAMKNNKAAGLDDILCEQIKQLRPGALNWLLKMFNECLGTNNIPKLWRKSKVIALLKPGKDAASPKSYRPISLLCHTYKLFERLILNRIAPFVDEHLIPEQAGFRPGKSCTGQLLNLTQFIEDGYESKLITGAAFVDLSAAYDTVNHRILTRKIFELTKDVRLTGLIQTLLSNRRYFVVLNGNRSRWRRQKNGLPQGSVLAPLLFNVYTNDQPIHPNTRSFLYADDLCIATQNQSFVKLEESLSDALAGLIPYYATNHLRANPDKTQISAFHLKNRDANHQLRISWYGKRLKHTPNPVYLGVTLDRSLTYKNHIANTKAKVGARNSILKKLANTNWGTDARTIRTTALALCFSAAEYASPVWSRSAHAPKIDPALNSSCRAITGCLRPTKVEDIYLLCGIAPPHVRRLVLSQKEKDKQENNPLHPLFEQQPVGKRLKSRHSFLHSVEPLNTSPHNQRMTLWSNHLQTTPHKLSLSPKESLATGSSETWPTWLSLNRLRTGYGRCRVLMQRWGLGDEQTACECGQEQTMEHLLVCPLLPEPCTQDDLEAVNCRAKACVDHWTGLV			
	I-1_DPe	GCF_003286085.1	Drosophila persimilis	ACTCTGATATCGACTCAGAAATAGACAGTAAACACCCTATCAAATTCCCATAAAAACAGCACAACCTCTAACATTGTAAAAATTCTACCACTAACTACCTCTAAGCGCAGACTAGCAGCAATTAAGGCTGCGGACAAATTAAAAAGCTCAAGCAGCAAAAAAAACAAGACAGAAGAACCCCAATACTGACGCC	TAACTCTAATATAAAACCTATGGAATACATCTCACATATAATAAACATAATATAATAAGCATAACATAATAACAACATAGCCGAAGGTCGTAGTAACCAGTGCTATCATTAATAAGCTAGGTTAATAGTTTTAACTCTAACCTCGCTGTAAATAAATAAATAAA	MTLTIIQWNLRGYINNYNELQIIIKQYQPHIIALQETHLFNIHNIPIPINFSLYAKNYSPNFGGAAVLIHNSIESKEIHLQTDFDATAIEVESKKKFKIISAYISPSQNFNESNLHNVLSNNSSSLILGDFNSWSKLWGSPKNNARGNTVAKFLNNTNFILLNDKSPTHFTTHNTYTHIDLSFCSPDLFQYANWKIIDDLHGSDHFPILITLFNNNKQNKTQNNPIFKINSANWDAFKDKTDYFNTKNSPSNNINKESSCLNKILLQSAHLTIPQTSIKNHPRNVPWWNNGLQTLNNERKQAWNKLKKHITTTNIIRFRKAAAKFKKEIKSAKNKSLSSFTSEIQPQTKTSQTWSNIRRFCGLNNMKNIHTIIHPTTLQPTSDNPSIANSFAEYWSDQSKDSNFSSMFTNNKYSISTYQSSPPSQSAQKIESDITSIELSSALSSLKGHTPGIDRI	cd01650	491	749

			TTGAGTAGCAACAA	TGGCTGGCAAAAA	TAAGATTCTCCAGGACCAT	SYPMIANTSPAVKDRILKHYNNILNNYIPQAY		801
			CTGCTCAATGGAAA	TCCTCCAGCTGAAC	GTGAAAAACAACATCCGG	KIGLVIPIAKPGYIKTSITSYRPISLNSCFAKVLDK		
			TAGAATCCGACTAA	CCTCCACGAGGGA	ATAAATAAGGGCCATACT	IIAKRLWWFVTTNKHIHTNQVGFKRGKSVTD		
			AACTAACAGATCAA	GCCGGCTAGGCAA	GGCCTATGATTGGCCCCCT	NLAYLDHLITDALSHKQHLSIVSLDFNKAFDKI		
			GTAAAAACAACCCA	TGGGCACTGTATTA	CAGGAGGGCCGTCCGGTT	GLHAITKQLQEWNIGPKILNYIINFMSKRKIKC		
			AATAACCAGCACTC	CCGCGTACAATGGT	TAACCTAACCTAACCT	RVANNLSSTFPLHNGIPQGSPISVIIFLIAYNKL		
			AAGAGATAAAAAC	AATGCAGTGGCGC		ANLITIQRNLTFTAYADDFHIIVKFKKNQTQEN		
			TCAAAAAA	ATGCTCTGTCCGGG		FNLDDLFQKIDEWCSFSGSSLSFDKCKHLHICR		
				GTAATGCAGTGTCT		KHNCKCTITTRNLNIPSVDHLKILGLTFDSKYR		
				AGCTAAGGCTTCG		WTRHIEDLASSLSNKINIIKCLSSNKFNCNTHT		
				ATTCAATTTCTCCC		LISITKAIIVSKIDYGLVFYGNSSITWLKKIQTILN		
				GCCTAAACGCCATG		CAVRIALGAFRTTPITNLLMEANINTVADRRD		
				GTTGGAGGTGCCT		FLQAKLIKNLYFPNNSPLHKISKSLQSRIHFNN		
				GGCGGCAGGCATA		KIPSAITYSLQHCKKQNISLLIKPEIKTKLPFWKF		
						QPNTINTELEKLDKNNTHDSIYQQHFLTIKNQ		
						FNDHTFLFTDGSKTISTTSFAITTEKNTLKVGLL		
						QNFSSIYSAEIIAIFEAIQIVKSMRGKYMICSDS		
						LSSLQAIGNPDNFNYYESKIRSLLMQLHPKIKL		
						LWIPGHKNIKGNELADQKAKSASTMPLTYTH		
						NNNISDINKHLKTKMLKDKLIKLEQTSNWYQF		
						INPNQITADKFTKRLDQNTTSRIKQTKIIRLRLG		
						HTRITHQHYLDPTKINTCTHCLNNVPVTIEHLL		
						KDCPSSQLTDIRFKIFANNNPLTFLSNPTKTEI		
						DLIIKYLKETNLYPLI		
	LOA-2_DSer	GCA_008042835.1	Drosophila serrata			MSLRVLQLNLHKSKVASAELLIAMEQGSADV	SSF56672	436
						ALVQEPWIASGNAVAGLKSSNYNAFYTPSVS		
						RNRTAVLVRKGLQANLMSHYSSDDLTVVLLE		
						SKKKRLLVASCYMAHDRPAPPEELRSLVEPGP		
						KDQQLIVGADANAHHSVWGSPDINDRGESL		
						LDFILSTNLYIANVGEEHTFVGPTSSNVLDITLV		
						AGQSTMVSDWKVLERPSFSDHRYIQFQCEF		
						DHSPKPTVFRNPRNTNWGKFQDIISRKLATS		
						PGTVKSDEDIEKSLETLSKELLGAYHASCPISRS		
						KKRTRPPWWNKDLSLQRDDLREFFKLAKQA		
						KDDIINEEYKILLRNYKKEIRRAQRSSWRNFCS		
						NIETAPETARLRKLLSKQPTIQSQLKRDDGQW		
						TEGSEEALTALMHGHFPGCTKITNTHPAISFW		
						IGNLLSCRRVQSDWGTASMVKGAHRGTPQ		

				AAACGCAAGTGGG TGGTTCCCCGGAAA TAAACCAATTTTTC ACAGGATGGAGAC AGGAGAGGCAACC GGGGCTGGGATGT CAGCCCAAACGTCG GTGGAAGGCGTGA CTGCTACCACCGGC GGGCACGAAGGGG AGCAATCCTCTCTG CGTGTCACCAGCG GTCACACCTCTGAT TTGACGGGAGCAG GTCAAGTCAGTTGT AGCGCTAATGCCAA ATAACGACGAAAA GAAGGCTAAGCCG CACGGGTGCCGTT GTCTATACAGACTC GGACCTGGAGAGC GTGCAGCTTTCGAA AGAAGAGGAAGAA AGCCTTCTTCGCTC GCCAGAACCAGGC ACCAGCTCCCCGCG AATAGACGGGCCG AAGCGTTCAAGCG AGGGGAAGAAGGC TAAGGCGTAGTAT AAAGCTGCCCTCCG AATCCAGAGCCGG CTCCGGACAAAAG CTGCCCTAACCGAA GAGGAGAACGTAA AGGTCAACTGACA GGATCGATGAGAG CAGGCTGCACTTCG		GGVLSPLLWNLVVDDLIRRFERKAPKITAYAD DISIIITGVCPSTLSSIMESTLREICEWTDKVGLS INADKTDLILFTKRYKVPPWIPPKINTTRLTPKT QVKYLGTVLDSKLAWKPNVLERVKKATIALYA SKKMLSSTWGLSPALMHWVYISVVRPTLLYG VLVWWQAMEKRTYKKLMERTQRQALLCISG ALRSTPTKALETIIGIDPLDIHAQLTAGKAAQR LVALGSMSTKGYGHSSIGRTMSSVSDYMTPR TCPDVKTSTFMGPEDWKTVQDHARYLNIYT DGSKMEGGVGAGIYCTDPETRLSYKLPSHCSI FQAEVFAIRKAAELAQEIARPCEAVNLFVDSQ AAIRSMQSSAVSSKNVLASRDALDILSTTTSVR IYWVPSHQGIEGNEIADELAKEGVGLAIDRTE NVPVSLRTLQSELERRADVRAKSRWRGNSAT CKTSKIMCKERNDKLSHYVLHLSRKDCRLLVG LLTGHCLAAAHAVKLGITNNGNCRKCDESEA TETLEHLICKCPALTQARVKHLGAPVLASLAD ASKRSPGELLAFAKTTAILKDFNTA			


CCAAGCTTCCTAGC
TTGGCATCAACGGC
CGGGTTCGCCAAC
AAAGTGTAGGTGA
TGCTTGCCACCAAG
CGGCAACGCTCGG
CCGAGAGTGCCAT
AAAGGACATTCCG
GCGAGCAAGCGGG
GACTCGTGCGCAA
GGAAGCAGGCCCC
CCCTGGAAGGCGA
AAACCAAATCCAGA
TTAACAAGAAGGC
AGAGGACATCCTG
TATCCCAGACACGG
GAAGATGGCATGG
GTAGCGTCTACCTG
CGCCTCAAAAAGC
GCCATCCTGGGGA
TAAGGATGACCAC
ACCCTGAAGGCAG
GCGAGATGGAAAA
GGACCTAGGTCTG
GAAACCATTTTTGA
CGCCGCCCAGGAA
CTAACGCTGGATG
AGTCCGAAGAGGA
GGAAGACGGTGAC
CTGACGATAGTAGT
CAACCCGTCGAGT
GGAGTTGAGCCCA
GTACCC

Table 10. provides a listing of non-LTR retrotransposase systems with annotation of reverse transcriptase domain signatures.

A database of retrotransposase proteins and the associated retrotransposon 5'UTRs and 3'UTRs for use in novel Gene Writing systems. Reverse transcriptase domains in the proteins described here were identified using conserved RT signatures, and annotated to indicate the presence and location of RT domains within the polypeptide sequences. In some embodiments, a system or method described herein involves a polypeptide having an amino acid sequence according to Table 10, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a functional fragment thereof. In some embodiments, a system or method described herein involves a domain (e.g., a reverse transcriptase domain) having an amino acid sequence according to a domain (e.g., a reverse transcriptase domain) of Table 10, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or functional fragment thereof. In some embodiments, a system or method described herein involves a template RNA comprising a sequence according to one or both of a predicted 5' UTR and a predicted 3' UTR of Table 10, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or functional fragment thereof.

1. Family	2. Element	3. Accession	4. Organism	5. Predicted 5'UTR	6. Predicted 3'UTR	7. Predicted amino acid sequence	8. RT	9. Start	10. Stop
					TAATTTATT	MPNLKTLTKELRQEIVSYLYPLPSISSNLAFRLAIHIFTKT			
				TTTCATCT	CACAACTTC	RFSSHINFISRCLQHKVIPKGFRSNFHASIFSNVSNFNLKY			
				CTCCCAC	ATGTGACC	LHEIQRTQNTFSHNIMRSTIRAMCSKRNELNKQIIECRSQ			
				CGACGCA	ATATCTTCA	LSNVCPSVLLRSIHAKIQALNSKLFHHLEQTKKHKLENL			
				GCACCAC	CCAATGGC	TNSPKNRTEPLENQNHHTVVTIPENLPLSNAEKSVLSKG			
				AGTTTCTT	AAAGCTCCT	LNFVPISKKSDEFTTRQDVEKFLRRVQLKAFFHNKEDKS			
				TAGAAAC	CTACACTCT	DNTEKDAFETLTAKKSKWTPPEGQFASIDYFIKKCRHDV			
				TAGAAAT	CATATAAA	HKLKSNCNTKLSNLSKEEWTALINLKNRKDLVIKAADK			
				TCATTTA	CCACAACA	GGATVVWRTDLYQQEAIRQLSDPTFYTKVNKDLTPAN			
				CTACTTA	ACCCACAA	QKIVKDTIQELIRKQELPVTAQNLIITTPRTSCIYFKPKIH			
				ACCGCTG	TTCCTCTAT	KPNNPGRPIVSACSCPTELISSYLDKVMTPIVKSLPSYIKD			
				CGACCTC	TCGCTCTGA	SNHALETFGNFNFSGENKIIFTMDITSLYTVIPNNEGLQA			
				ATTCATT	CGAAGGGC	LKYFFNQRPIKKPSSETLLRLAELVLTLNCFSFGDNYYK			
	Penelope-2_ADi			CCTCGCT	TAACGCTCG	QINGVAMGTKMGPSYANLFVGFIENKFFSNYHGPKPDL			
				ACCGCGA	AAACGTCA	YKRYIDDCVGATSSTKEELNLFINSVNSFHPALKYTWEI			
Penelope		.	Acropora digitifera	CGCTGAA	GCTTTCTAA	SENSLAFLDIKLSINGNGLSTSVHYKPTDSHNYLLHSSSH	IPR000477	371	610
				ACTCACG	ATCTTTCAC	PQHVKNAIPFSQFLRLRRLCSDDTDFNNKCEEMCQFFKK			

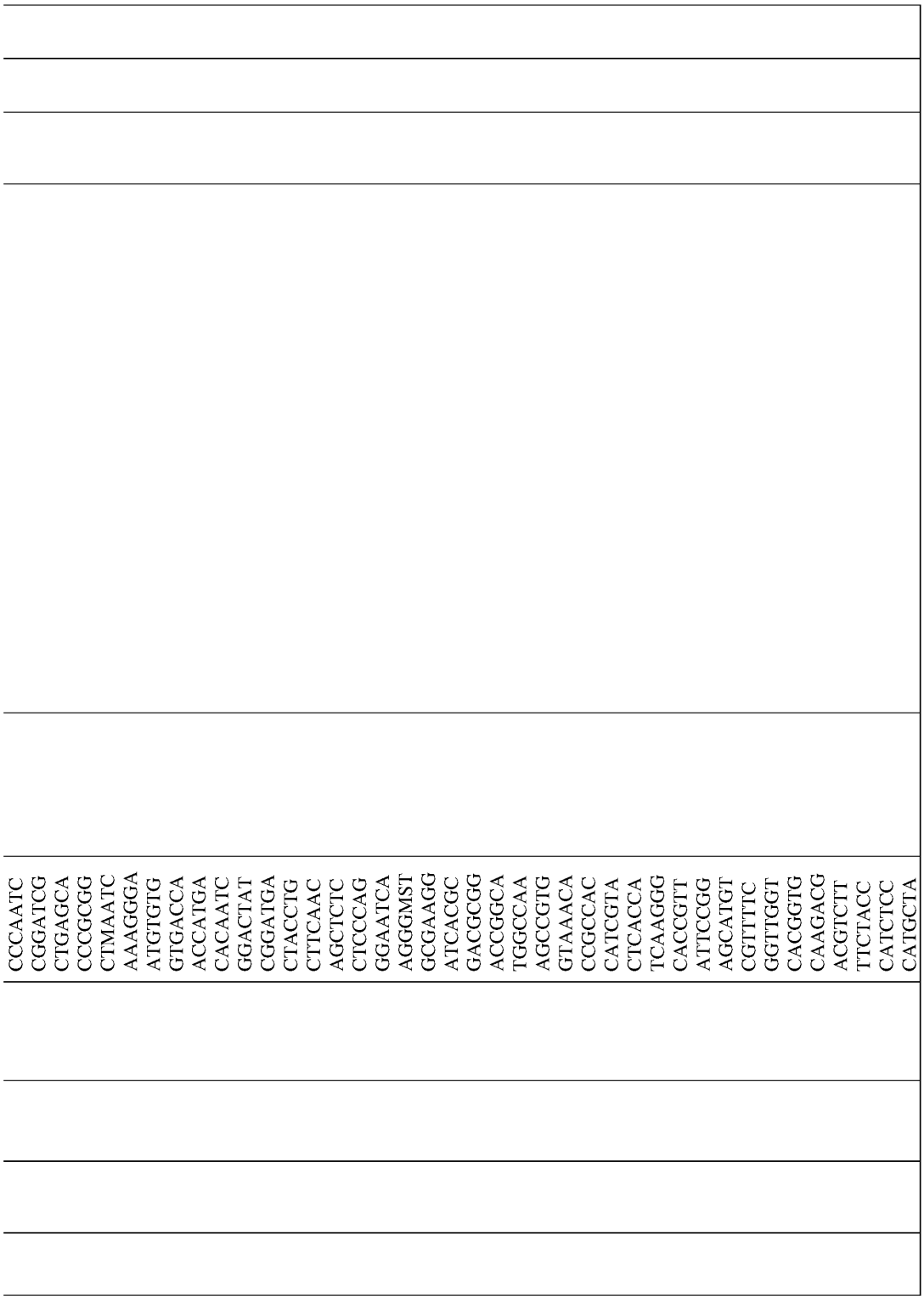
				GGTGGTAA TTCAACCTT TATCAACTC GTTTGATAA AACCAAAT TTTTGT	RGYPDSAVTTGKHRAQEIDRETALQTSQNEETDRIPFTL TYHPQNLAIKNVILKNFKILSNDPETKHIFSLPPLISFKRD KNLGNFLVRSAFKFNNKPGTIFTCKRTRCKTCPFISNTVN ISGPNRSVKVTDHFTCISSNVIYCITCTLCKKIYIGETGRR LADRFREHLRDAEQNNTDASKPVARHFNLPNHSHHNM TICGLYLHHGNTESRKNLEQKLIFQLGTLSPHGINERLSF H			
				GCCTCTT CGCTTGA ATCTCTG CTGGCCC ACAGACG TGCAAAA AAAGCCG CTCGTTC CCGCCGT ATTATTA ACCACAA TCTCACC GNCGATA ACCTCAA TCAACAC ATCAATA ACACTAA TGCCAGT ACTAATG ACATACT ATCTAAT GTTACAC ATAACGC CTCTACC TCAACTG TCACACG CAAACGC TCTCGCC GCGAAAC CAAACAT GCCAACA TTCATTTG GATCATT	TAATATTGT TACCTCTTA TCATGTTAT TTCATTATT TTTTCACCT TATACTTAT CTTTAGTTT ACACTGTCA ATTATTTTT ACTTATTTC CGGTTATTA CAAAATAT GCAACATA AAAATCAC GTACCAAT GTAAAATTC TCATTGTAC CTGAAGAA GGCTGGTTT GGCCAGCC GAAATATA GTACACCTA CAAAAAAG ATCCTTCTA CGTTGTGTC GACTTCTGC TT	MRLAEYFFDENSNNRTANEHDTPFHNKSTWNPPNDRER ALNTFLDAVKLDITTTKPKPTQDNLTVTERQAIGQLKQR QDIIIKPADKGSGTVVMDKTWYIDECNRQLTDTKFYQR LDEDITADIQKRVTFYVNRMHKDKLINDKTKQYLIQSD VKPGRFYILPKVHKPGNPGRPIVSSNSHPTERLSHFVDY HLQPLVHKLPSFVKDTNDFLNKLLTIGNLPANSLLVTLD VSSLYTNIPHNEGINACERFLRTSSHKTIPTSTLCDLIRMI LTMNNFSFNDNHYLQIHGTAMGTKMAPSYANLFLGFFE ANALKNAPFQPHTWLRYIDDIFMIWTEGLDNLKIFIDYL NNIHSTIKFTSSHSSTNIPFLDVNVSLTNDGNISTDLYTKP TDKHQHLLYSSCHPLHTKKAIPFSLALRLRRICSTDATFH TRTAQLATYLLKRGYNRNFVNKQIRRAADIPRQLTLQT KDINKPKRIPFITTFNPSLPHISHIIKKHFNLLLLSSNRCKSV FQHPPVVAFRRSPNLRDLLVTAKLSPNSTNPQLPCGSFR CGKNCATCPYISHGLTTYTFFSTGETRPIKFNFTCETKNL IYMIQCNRCNLQYIGETKRRLKDRFNEHRRTIDNPNNKS KPTTAAEHFLSSPNHTANDMILIPIEKIFSNRDSIRKAREA FLIQKGRTIDPDGLNIREETY	IPR0 00477	143	383

				CTTCTGTT ATCAACC TATCCAG TTGCTCTT TATCTAC CGACGAG ATATCTA TCCTTTCA CGTGGAC TCACTTTC TGCCCTA CTCCACG ACATATC AACTGGC CGGAAGT TTCAGCC GACATTT ACGACTT TTCTCGG CGC					
Pene lope	Penel ope- 4_AD i	.	Acropora digitifera	TTTATGA T	TGATTTACT CTCACAGTA CTGCCCAAC GTATTCTAC GATACACG TACTGACCT TAGACCTGT AGACGGAT CGAAACGC ACCAATCA CTGTTTAGT CTTTCCAGC CAATTACAT CTAGGCTCA ACTGACAG TCGACCAAT AACATCAC GAATAAGT TGACCAATT ACATCTACG	MTPGFKPFTILRSLGRQSERLANNYISYASNVRRIRQQL WFNHRCKDLRLVPSGLRLKSPLNTQEAIQIVKATCRRLI RARINDCHRRLNHYNNRLQQRLDKLKQLIPTNLLDVIR DIADQRAKKTTEQYRIKTELKLTRLQRTKNKKRHKTDE NWVRNISSRPLDKTETQVLSYGLKHSVTPKRIPTETIVSS VEAVLSRQRDLSEPTKDNIRSRIASTIQSASLTDNNLTKD ERQALKRLRNDNDILILPADKGRVTVVMDKTDYHDKM DELVNDKQTYEVLKRDPTPALQRKLNSKLLQLKKADAI DIRRYNRLRCPVPQPAKLYGLPKLHKPNVPMRPIVSFCG SPTYELSKYLTTILKPLTDESRHKLQSTETFIDAIKTAQVP DDHKLVSFDVKSLFTSIPLQLALDCTETAINNSTLQLPLL TNDLMDLLNLCLTSTYFQYNGQHYKQLHGTAMGSPVS VVVAEIVMQNIEEQALASYKRTIPLWLRYVDDTFTTLH KDEIDDFHEHLNRQNAHIQFTKEIEDNGKIPFLDCLVIRD NNRLQTTVYRKPTHTDRLLDESSYNPSSHKATTIRTLTR RALLVCNSHDSLADERKYLDNVFSKNNYNCDFVTRNT YRTEPNETNTNLTPTTTVTIPYIKGTSEIIARILQPYNIRVA HRPITTLRKLLTNVKDKDQPRDRQGAVYKIKCCDCQAT YIGETGRNLNTRLTEHRRATRNGDINNNIAEHHLQTNHR	cd003 04	459	541

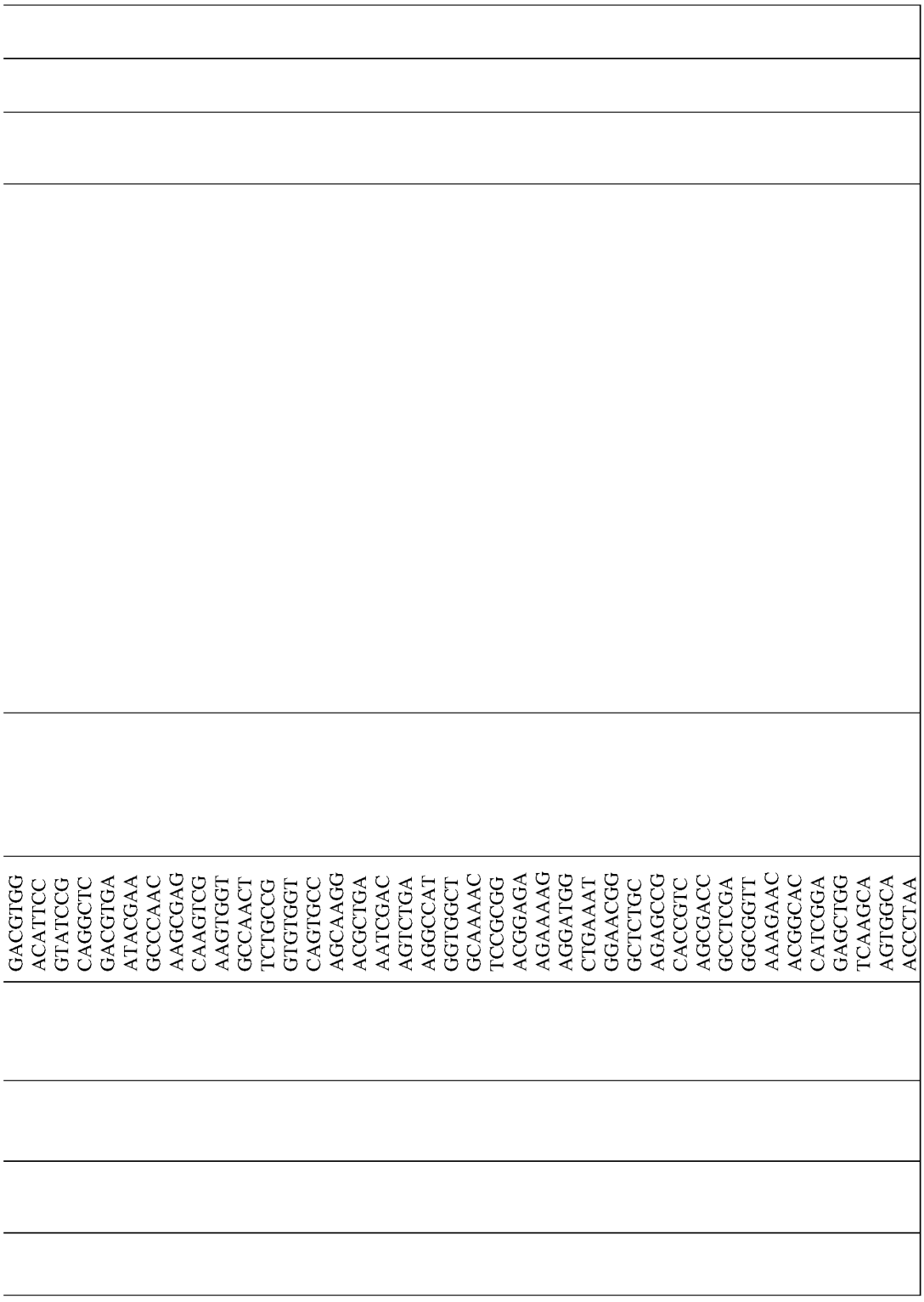
					ACCGGAGT TTTTATAGT ATCTACTGA CGTTACACA AATCACTTG ACTCTGAA GATGACTTC CGCTCAGGT TGTCGAAA CGTCAGTCA ATGTCATCT CAAACAGT CCTTCTCAG GACTACACT CACCCGGA CGATCGTAC TTTAC	IDWDSATCVTYNTNYYQRIVLESWFTNLEQTPINRCLQL PAPYKRLIDDINRQTTH				
					TGATCCTGC CCAATATG ACGACGAC ATTGATGCC AGCCCACG GATACGAA ACTGTTTTA CTAAATTGT ACTATGTAT AGATGTAA GCTTGAAAT GTTAGATTG ACAGTTATC TCCTTTTCA TCGGTTCAG TCTTCTGCT GACCGTATT TCTCCAGGG TCTCAGCCT TCTGCTGAG CCTTCTCCC GTGTTGAAC TAGCATAAT	MPGPALGGAPLESATRTLCPPNGFSQTYFSTAAPSPLPV VGSTPGPTLGGALSGSVSSTFLGSTNSKETSLSSFPVPAL PTSHNFILQWNLNGYRSRLCDLELIIQNNQPWILALQET NNISADEMGRTLGGQYAWTVKRLANQRHSAAIGILKGI PHKILDLPSDLPIVGIQLLGSPSVSVACAYLPGGNIPNLIH GQIQNCLQALPEPRIFLGDLNSHHPVWGSPKADKRGSTL INAFEEEDMLILNDGSPTFFNGHFSSEIDVTAVSLSFARNI QWSVSSDLHGSDHFPIRIALTTGASIPKNTRRPRWKYDK ADWAEYDLLIRSALRNSPPDNMADFISTIVESAKATIPRT SSTPGRKALRWWSDETKTAVKARRKALRQLRRLPLDH PDRQSTLENYRKLHLECRNIIGDAKKATWESFLESMNAS QTTSELWNRVNALNGKRKTTPLTLQVQGNAISDPEAVA SELGKYFASLSAIESYETHFXNSVKPSTSAINNFIIPPNTA DPEMEQPFSARELSFALSRSNGRSAGPDEVGYPLLKNLP SVGKLALLDLYNKIWANNDFPPAWKESLVVPIPKANIPT RDPTKFRPISLTCCAGKVMERMVNRRLRYHLEANGLLD FRQHAFRQGYGTSTYFAGLSDVLKEAYDKGNHAEVISL DISKAFNRTWTPLVLEQLAAWGIGGRTLHFVRNFLADR TFQVVIGNTKSSTFPEETGVPQGSVLAVTLFLIAMNGVF SRLPKNIYVFVYADDIVIVVCGSTPTMTRIRAQTAVKSV AKWASDNGFQLSASKSIRCHICPSGHRITGPDITIDGQPIP LRKTVRILGVIVDRALSFRQHFDTVKSACRSRLNLIKSIS RPHRSNNREIRFRVARAIVDSRLVYGLELTSIAMDRLVE	cd01650	575	833

				GGTGTGG TGAGAAG TGAGGTT AAATAGA CTCGGTA GCGTGGT TCGAGTA CTGTATA KTATTAT TTCGGGC GCGAGTG TATGTGT TAGTAAC TTGTGAG TGGTATA AATAAAA TTATCTTC AAATTAA AAAGCGG TTCAGTG GTTTTCA ACTGTAG TGTGACT GACCAAC CAAAACA GGTGAGT CTTCCAC GAGTGCC TCCTTCC GGCCCGC CGGCGCC ATGCTGG CGAACGG CCATCCC CTCCACC CAGGGGA CCCAGGG GGCCCCG GAGGAGG ATTTGGC	GTTAAAAA ACACGTTA ATAAAGAT GAAAAAAA AAA	VLSPIYNSYIRIISGLLPSTPSDSACVEAGLLPFRIFIRSSIC CKTAAFLSKTAGEDRVFLLDEGNRALSTAANLSLPPVTR VHWLGDRSWRSVPPKIDNKIKNSFSAGSNSAALRRSVA ELLQTSYSDFALRYSDGSLTSAGVGIGVAGDIPDVSMSL PSQCSVFSAEAAAAFIAATTPSDRSILVLTDSASVISALQS DSPTHPWIQAILKYALPDTVFTWIPGHCGVPGNEMADH LARSGLSGQRYTSEVPFMDLKRWIKSTFRQHWEDSWY RTRTLFLRKIKNSTTTWSDLPILKDQRILSRLRTGHTLISH NMGGGPFHKECESCHIPASVDHVLCACPMYEHLRQIHG LSDNIGEVLRDDATTIAALLSFLRDANLYSSI				

AAAGTAG
GCAACAG
AGTTCTC
GGCGACT
ACACCGG
TGCATTG
CTACCGG
AGTTTAT
GGACCGT
GAAGGTA
CTGCAGG
ACAACTG
CAGTATC
TAAGAAT
GGAAGCG
ACTGCTG
GAACCAT
GCCCCAG
GATCCGT
TTCTTCTT
CGGCTAT
CCGTCGA
AAAGGCC
ATTGGAG
GTCCCAT
TTTCGGG
GCCTTCA
AGGAAAA
TAAAGGC
ATGTCCT
ACGTGTT
GAAGGTG
CGCAGTC
AAGACCA
GGTMAAT
AAGCTCC
TCCGCAT
GAGCCAA
CTGCAAG
ATAAMAC


CCCAATC
CGGATCG
CTGAGCA
CCCGCGG
CTMAATC
AAAGGGA
ATGTGTG
GTGACCA
ACCATGA
CACAATC
GGACTAT
CGGATGA
CTACCTG
CTTCAAC
AGCTCTC
CTCCCAG
GGAATCA
AGGGMST
GCGAAGG
ATCACGC
GACGCGG
ACCGGCA
TGGCCAA
AGCCGTG
GTAAACA
CCGCCAC
CATCGTA
CTCACCA
TCAAGGG
CACCGTT
ATTCCGG
AGCATGT
CGTTTTC
GGTTGGT
CACGGTG
CAAGACG
ACGTCTT
TTCTACC
CATCTCC
CATGCTA

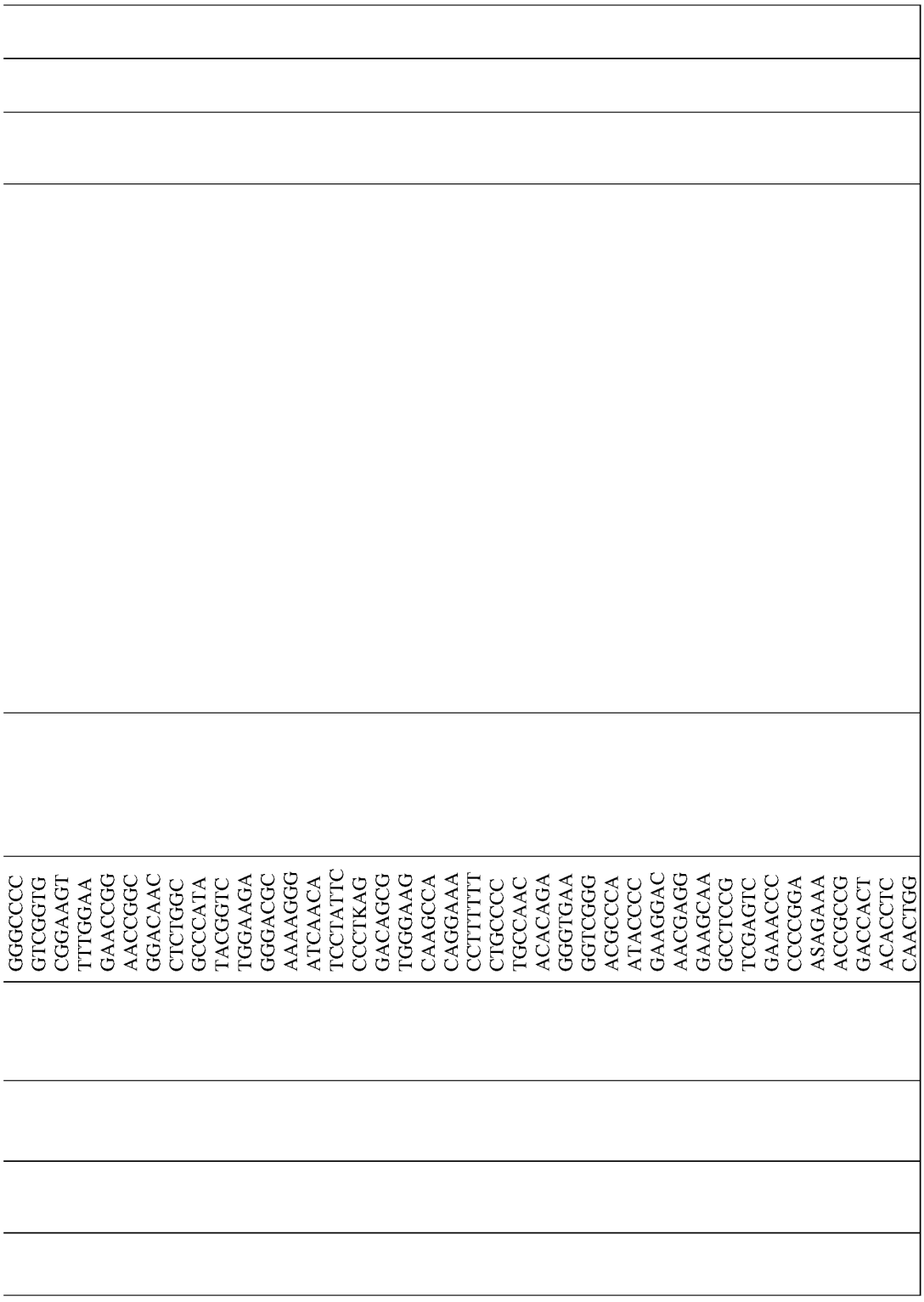
				TGCTTCC GCTGCTG GGAGTAT GGGCATA CCGGTAA GCGTTGC ACAAGCC CCCAGCA GCATCTG CGGACGC TGCTGCG AGGTACA CCCGGAA GAACAAA CCCCCGA TCCTACA AACACCC AGGAAGG AGCCGCC AGTGCCC ACGTTGC GTCGAAA ACGCGGT ACCAGTG CACCGCA GAGCCGT TGCTGCA AGACCAG TGCAAAA CAGGAGA CAGCACT CAGTCTC AAGCCGC AAGTGTC CGGTGTA CWWGAA GGAACAG GCCATAC AACACCT TCGCGTC					


GACGTGG
ACATTCC
GTATCCG
CAGGCTC
GACGTGA
ATACGAA
GCCCAAC
AAGCGAG
CAAGTCG
AAGTGGT
GCCAACT
TCTGCCG
GTGTGGT
CAGTGCC
AGCAAGG
ACGCTGA
AATCGAC
AGTCTGA
AGGCCAT
GGTGGCT
GCAAAAC
TCCGCGG
ACGGAGA
AGAAAAG
AGGATGG
CTGAAAT
GGAACGG
GCTCTGC
AGAGCCG
CACCGTC
AGCGACC
GCCTCGA
GGCGGTT
AAAGAAC
ACGGCAC
CATCGGA
GAGCTGG
TCAAGCA
AGTGGCA
ACCCTAA

CTGAAAC
GGTTCAG
AACCTCC
AAAATGC
TCTAGCA
GTAAGGA
CAAACAA
ATCCCGG
ATCAAAG
TCGCACC
ATCGCTG
AGCTTCT
AGGCAGG
CAATCAG
CTGCCGT
CCAACGA
CAAACCC
GCGACAA
GCTTTGG
CTTACCG
CGCAAGA
SGATCAA
ACGCAAA
CCGCCCC
AAGTCGC
CAGAGGT
TCCCGTC
ATGAAAC
CGCCCAG
TCTGAAC
CCCAAAG
TGGACGT
ACAGGTG
TCCAAAT
GGATTTC
AAATATA
ACACCGG
ACCTCAA
GAAACTC
TCACAGG

CGATGAA
ACCAACG
ATGGAAA
TGCGTCT
GATCATA
GCATGAT
GTCAGCC
ACATCCG
AAGCCAC
TGGAATC
TCCAACA
CGTCTCA
CGTTTCG
AACCCTA
CGAAAAG
AATGCAC
TCAACCT
CGGATCA
CAGTTGC
CCATCCG
GCGATTC
AAATGCC
TCATCCA
CCAACTC
GAAGCGG
AAAAACA
AGCACCG
CAAAAGA
GGTGGTA
GAACCAA
ACAGTAA
AACAGTA
ATTTACG
CTCAGTC
CTCTTCA
AGCACAT
GGTCTAA
ACAATTC
GACAATA
TTTTCTTC

				CCGTCAT CTTCACT CCGCAAA TTAGCGT TGAGACA GTCAAAT CACAAGA GTCGACT GAGCAAA TACCGAT GAACACA CTACAGC CTCCGAA CGAAGTA CTCACGG ACAACTC AGCGACA ACAGGGG TCACGGA CAGCCGG GGCCCCG TCAGTGC GGACGCT ACACCCC AACCGGA ACTGGCG GACAACC CCCGGCA TCCGGGC GTAACTG TTGTTATC GGGACGC AAAAGGG ACACAGT GCCTATC CCCCGAA AGTCCCA ACTAGCT GTATGGG TACGCCA					


GGGCCCC
GTCGGTG
CGGAAGT
TTTGGAA
GAACCGG
AACCGGC
GGACAAC
CTCTGGC
GCCCATA
TACGGTC
TGGAAGA
GGGACGC
AAAAGGG
ATCAACA
TCCTATTC
CCCTKAG
GACAGCG
TGGGAAG
CAAGCCA
CAGGAAA
CCTTTTTT
CTGCCCC
TGCCAAC
ACACAGA
GGGTGAA
GGTCGGG
ACGCCCA
ATACCCC
GAAGGAC
AACGAGG
GAAGCAA
GCCTCCG
TCGAGTC
GAAACCC
CCCCGGA
ASAGAAA
ACCGCCG
GACCACT
ACACCTC
CAACTGG

				CACAACC ACCCCGG TCAGCGA CAAGCTG AATCCCC CGACTTC CCGGGCA GCCTTGG GACCCGG AAGCAAG CGACGCC AGGTTGT ATGACCT GTTACAC AACGCTC GTTACAA TCGCTCA GCCCGCC GACTTCA GCCCGGG TATTCCG GAACCGT TTCCAGA CTTCAAC TCCAAAC GCGGAAG CATACGA AAACCGA TCGCCAT CCGCGAT AATKATA GCTCAAC CCGCTGC TGGCCCC TCAAGCC AAACAGC CTCCCCA GATCCCT ATCAGGA TCGCCCT					

				GGGCACCTCGGACTCAACAATCCGAGCGGTAAGGCCGACCCTTTCCCGCTGGGCCTCCACCCGGTGAATCTTCGTCCGCCACAACAACCTATGTCAACCCATTTTTCGACGCGGCTCCCAGTCTTCTTCGCTCTCAAACTCACAGCAGCACTCCTTGATAAGGTTTCCTAGCAGCCCTGTTGGAAACGACTGTCTTGCCGCAGCTCCCGAAGTTCCGGCAGTAGTAGGTCCC					
RTE	RTE_Ele3	.	Aedes aegypti	GGGCTGCAAAATGGTGAGTCGACGGAGTCCTCACAAGTTCCTACCTCAT	TAAGTAAGTAAGTAAGTA	MKLEILGLSEVRWPNFGEHRIPSGQILLYSGLRGEHAPRHRGVGFLLSAQAYAALMKWEPINERIIVARFRTRVRNLTIVQCYAPTDAAELQDKENFYSQLNAVVDKIPKGDIKIYMGDFNAKIGSDNSSYEPVMGRHGLGEMSENGELLAEFCGNNDMVIGGSLFPHRPVHKVTWVSRDGFTENQIDHICISRKWKRSLLDVRNKRSADIASDHHLLIGEVRLRIARIRRQEERVGRRFNTRRLEDATVKRSFVEELETXAANISEGGSV	cd01650	503	762

				GCTTCCA CGGGTCG ACGGATG ACAAATG ACCGCCA GCTAAGA GTTGCGT GCTTAGC TGGTAGT GCAGCCT GGGTCCT GTAGTCC TTCTGAC TTCAGCT AGATTGA GGAGGTA CGTCCAC CGAGCGT CTGTTCA CCAAGGA GGTGCGG CTCAAAC AGCGTCT GTCTGGC ATCCAGC GGCTGAA TAAGAAA TGCTCTTC CCCGGAA GCTACAC CTAAGAT GGCAGCC CCATCAC GGCGGAT AGGGAAC CTTAGGC TAACAAC CTACTGT CCCCGAA ACATCAA		EEQWTAIKNAFITTSENNLGELRTRRKQWITDETWRKIE ERREAKAAIERAKTRGAKALSRQRYSALEREVKRSCRR DKRAWADSLADEGEKAAATGDIRLLYDISRRLSGAKM NTTMPVKDTTGQLLTDPTDQLKRWFEHFEQLFQVSTRP PTPQHDPPRVRRISRVNTEAPSLHEIETAIQSMKSNKAPG VDRISAEMLKADPAISAQLLHQLFRNIWDTATFPADWM QGVLVKVPKKGDLTVCDNWRGIMLLCIVLKVLCKVILN RIQEKIDATLRRQQAGFRAGRSCVDHIVTLRIILEQVNEF QESLYMVFIDYEKAFDRLNHENMWGALRRKGVPEKIIG LIEAQYEAFTCRVLHNGVLSEPIRVVAGVRQGCILSPLLF LIVIDEILIGAIDREQNRGLLWQPITMEHLNDLDLADDIAI LAQRRSDMQSKLDDLAERSSAAGLNINVNKTKSLDVNT AGPSSFTVAGQTVENVENFQYLGSQLASDGGTKIDIGAR IKKARAAFASLRNIWKNSQISQRTKIRIFNSNVKSVLLYA SETWCVSAENTQRLQVFINRCLRYIIRAWWPHNWISNA DLHRRCHQQPIQTEIRQRKWGWVGHTLRRGGNEICKQA LEWNPAGHRSRGRPRGSWRRSLNKEIKEVDGNLTWPQ VKAIAGNRPGWKSFTSALCTRMGAQDP			

				AAAATGT TCAAGAA ACCGATA GAGAAAG ATTACGA ACTGATT TAACGGC AACGACT TTTAGCG CGAAACA ACGGACA CGAATTG GAACATG GAACGTA CTAACCC TAGCCCA GCAGGGT AAACTGG CACAACT CGCCAAC GAGGCGT GTCGC					
Pene lope	Penel ope_ Ele2	.	Aedes aegypti	AGTTAGG GATTTGT ATAATTA TTATTATC CCCCTTTT CAATTTTT GACCAAG ACATGAC AGGTATA CTCCTGT CGAAATT TTTTCCTA GGTAGTT TGATTGT GTAGGTA CTTATCT GGTGTTA GTATTTA	TGAATACAT AGGTAGAT AGTTAGGG ATTTGTATA ATTATTATT ATCCCCCTT TTCAATTTT TGACCAAG ACATGACA GGTATACTC CTGTCGAA AATTTTTTC CTAGGTAGT TTGATTGTG TAGGTACTT ATCTGGTGT TAGTATTTA AGGTAGAG	MASTLSESNTFIKLKKSIAGLKKDIAFLSECRKLNLTPKS HKIKVRGQVPQDIIKKMESEYLRSSIKKLHSKLNSQTLK CYSLHLKLAKQYPWEFPLFLTKVIQAEECEARRKKDIHS KKLAILKNKQHNKVVTALPKVHSIEGFVVNRSTQQFSEE QLTLLNKGLAYAVATKPDIEQTIIDVETAITHTIPIEFQES ARNDTEESLKGNKFQLRGQNSDEDKVLSELRDKSVYYV KADKGNAVVILDKDDYDNQMNEKITTGPYRHFRVDPL PTMVRHVEKTLKECKPLLGDNTSRLKVSNPVLPRIKGLP KIHKPGNEMREIISAEGSPTHKLAKWLVKEFQSMPKPFE SRSVNNTQEFASQLQASGYIQDDEIMVSFDVKALFPSVP VKEAINLLEEWLLNQHNESNWKNKVRYYIKLTRLCME ENYFSFRGNFYKQTKGAPMGNPLSPFLSELFMANFESIL KKKDLLPQRWWRYVDDVFSIIKKDYLPKILEAINSIHKD IKFTHEEEKDNKLPFLDLLITRESSHFEFEIYRKPTNTMR VIPSTSNHTYQHKMAAFHHMIHRMQTLPLSETGKSKEL DYIFETARLNGYRNSTIQAVVDKRARDKYRKSMTTLTA EKEDLKRVAVNFDTKITQPLAKKFRKYGVDLVFSSRSS QLKSRLGSTKDKINKLNRAGVYKISCPHCNKIYIGQTKR	cd003 04	445	525

						TLEIRFKEHIAEVKKADKELEKGLAYDFKSKVAEHVFQ				
						DGHQMTSDNIDIVRSVSSSWKLDVAESLEIYKQKPSLLL				
						NRDQGNGQSRLFKFVPKKYKRA				
			AGGTAGA	AGATGCGA						
			GAGATGC	CGATTTCTT						
			GACGATT	CAGTGGAT						
			TCTTCAG	ACGAGTAA						
			TGGATAC	CTGACACTG						
			GAGTAAC	AAGAAGAC						
			TGACACT	TGCAAGTG						
			GAAGAAG	GTAGTCGA						
			ACTGCAA	AATACGCG						
			GTGGTAG	TATCTGTCA						
			TCGAAAT	AAGATAAG						
			ACGCGTA	CATTTAGGA						
			TCTGTCA	GCGGAATT						
			AAGATAA	AAAAGGTA						
			GCATTTA	CACGGTACT						
			GGAGCGG	CCATCTACT						
			AATTAAA	TTAAAGTTT						
			AGGTACA	CTCTATTCG						
			CGGTACT	AGATTCTGC						
			CCATCTA	TCAGAGGA						
			CTTTAAA	TTCGAACAT						
			GTTTCTCT	TCATTAAAG						
			ATTTGAG	AG						
			ATTCTGC							
			TCAGAGG							
			ATTCGAA							
			CATTCAT							
			TAAAGAG							
			TTGTA							
					TAATGTATC	LDNFAFRGLKQRNLLRNHLDRTLDLIFTNDAITDCSSVD				
					CTGTTAAGT	EACQPIVDIDPHHPAFVLTVRSPTIVPFVEVSDEEVRDFG				
					GTTGTCCAA	KVDYETLHVLLSRVDWSTVYQCNDVDNAVSIFNGKLH				
					GAGTTGAT	ECFTHCVPLAMPRRKPAWSNARLLYLRRNQSKMQRRF				
					GCTTAGTTT	SKRKNDCNKLAFNRASRSYRYYNRYLYGKYVRRVQDN				
					AGTTTAGTT	LRRHPKQFWSFVNSKRNETGLPSFLSLNGHSASSAEGKC				
					TTAATTTTA	ELFASWFASVFDTSSQSQADPDAASQSLPTDIIDVDIFEV				
					TTCATTAAG	SHMMVEEALKKLKPSISPGPDGIPACILKHCKAVLIAPLK				
					ACCATGAT	AIFNLSLRCQQFPKAWKMSYMRPIFKKGSRTEVSNYRGI				
					GTCGGATG	TSLPAGSKCFEIIMNDVLLQACKTYISQNQHGFFPQRSV1				
CR1	CR1-61_Ae	.	Aedes aegypti	TT	AATCTTCAA	SNLCEYTSFCIKEMDKGGQVDSIYTDIKAAFDTVDHSIL	cd01650	332	579	

						LAKLNRLGASSRICSWLYSYLTDRQLLVKIGKAISAPFTP			
						HCGVPQGSNLGPLLFVLFFNDVSHVLSNCYLLFYADDL			
						KIYLIVRSINDCRQLQEQLDLFVSWCTLNHLSISVSKCCII			
						SFRRMKSPVLYNYAINGTILNRVDNVKDLGVLLDHQLT			
						FKMHYSETVDKANRQLGFIFKIASEFDDPLCLKSLYCSL			
						VRSILEFASVIWSPYDAVWIARIESIQRRFVRCALRSLPW			
						SDPLRLPPYEHRCALLGIETLEERRKMNQAVFGAKVLRS			
						EIDSPALLRQLNIYAPQRVLRSRQLLYQPPRRTMYGSNN			
						PIDTISRRFQDNYNLFDFNISTNTFKNRLRNRQV			
					TACAAATA CAAATACA AATACAAA TAC	TAGTTTGGG CCCCGTATC ACTCTGTAC ACATCAATC GACTTGAA CGAGTTCA AAAAAGTT TTATTAGGT ATGCACTCC AGAGACTA CCTTGGCGA AATAGGAT GCTGCTGCC GCCATACG AGCATCGCT GTGCACTCG TCGACTTAC CTACACTAC GCAACAGG AGAACTCTT CTAAAGCA GCTTTTCAT TTTTGATCT GCTCGAAA ATAACATC GAGTTTCAA GACTGCTCT GGTCTGTTA GGCAAATT AAGTTTCAA CACCCCTGT GAGGCTCCT	SQHQRGGGVLIAVKKALQSNVILINDANQLEQISVRITY GTMSLVVCTVYLPPNTEPALYQQHAASVNQLCKLLGTQ DEIVVLGDYNLPYLRWRFDDDIGAFLPINASSEQEISLVE DIVATGLMQMNHLLNSNERLLDLAFVSNASVCELLEPP LPLMKVDRHHTPFVLTLDVIQTPLENRKNETFFDFRQCN YDELNAEIAQINWRDILAFGSLDDAVTRFYSELNVIFER HVPRRSRACRPDNKQPWWNAQLRNLRNRLRKARTRYF RVRNRLNRDAVRDLESQFNELNVSCFQSYIDRLESNMK DDPKQFWSYLRSRTTSRGFPQLINYDGKTATTPDDAVK LFSSFFQSVLSITPPPSSASYLSALKTFDLNLQPTSFTQHE VYKKLCSIDGSKGPGSDGLQPSFIKACSASLAEPAALLF NRSLAESVFPAKWKEALITPIHKAGNVVHDVKNYRGISIL SCLPKVFESMLLDFLYPAVRHIITEDQHGFVRKRSTTTN LMSYVSTLIDKIEKRQQIDAVYVDFTKAFDRVPHQLAVE KFQRMGFPDWLTKWVLAYLINRSASVRLGTTLSDPFEIT SGVPQGSHLGPLLFILFVNDLCSELKSSKVMYADDLKIF RVINSPVDCCALQMDIDKILDWCSRNGMDVNIAKCNVI TFTRKRSPTVFEYSMHGATIERVRLVKDLGISLDCKLRF VEQISSVIAKAYAVLGIIRRNTKDFRDVYCLKALYIYHLF AAFWNMEL	cd01650	441	690
CR1	CR1-81_Ae	.	Aedes aegypti	T					

					ACGGAGGACTGCACTATTCCGACACGCTGCTCATCGAACGTTATATGGGCAAAATAATCCGCTTGATGTATGTTGTGAAATGTTTAATAATGTGTATCACTTGTACGATTTTAACATAGCTAAGGAAACATTTAGATTAAGGATTCTACGGAATTGAAGTCTGTGCGATGTATTTTTATTTAAATCGAAGACCAAATAAATCAAATCA				
Crack	Crack-13_AAe	.	Aedes aegypti	CA	TAATATGCATTTTTTCTAATGCCACTTCCGTATCTGTTCTTTGTCATTCCACGCCACCACCACCCTTCGCCACCACCCACCGCCAGCCACCAATCACCACCACCCACTGCTAGCCACCCCGCCGCCAGC	HGGLAVFVRNELQMDVCRNIVIEGFHHIHCRFKFEANAINLHAVYRPPSFDVRRFLHEIEGMISLSKNNEKCVLVGDLNIPLNITTNNIVGEYIRLLETYNMLPTNTVVSRQSSNNILDHVVCSTSMIEDVFNETIFTDVSDHCLICSSINMKCHTSESTLHKRIIDHSRLNELFERSILNIPQNLSVNERLNYVITHYNEILNDCSRLVTQRVKLKGSCPWMTLDLWKLIKIKDKLLKKKRNNPNDNHTTELFAHVSKILQDKKAKTKRDYYEGLLSGNNQKAAWKIIKNVMGNQKSNSKPTAIRNGAQLITDPQQVCSSFNDHFCAVGPKLSSTINTNRNINRFGTLGSLQSTIFLKPSTRNEIILLINRLDSKKSAGPDGIPVSFIKHHHSFFASLLMETFNEIIVSGEFPDCLKIARVIPVFKSGDPKDMNNYRPISTLSIIDKLLEKLLVSRILEFSTKHHLIYGQQYGFRQGSSTLTACHELVDEIYDALDNKQIVGALFIDLKKAFDTVDHDLLLRKLESYGIRGIAKDLLKSYLTDRHQYVSIGEHKSQLRLVTTGVPQGSNLGPILFLLFVNDVARLNLHGKL	cd01650	420	676

CR1	CR1-121_AAe	.	Aedes aegypti	AA					
					CACCCCGCC	RLFADDTAVFYRGTNCDTILNQIKMDLEVLLEYFGENV			
					GCCATCGCC	LSLNLNKTKYMLIHTPRRKIPAHGPISIGGHNLDKVYEH			
					ACCATTCAT	EFLGLTIDSVMSWAAHINKLKSKVSSLCGILRKISSFMPQ			
					CACTTCACC	TCLKQIYFALVHSRLQYVTAVWGSASKSHLRELQVLQN			
					CATCGCTTG	RCLKIVLRKPLLYPTVNLYSNRNDSLLPIKALYEYQVLV			
					CCACCAACT	QMQKIVKGTSTHHNTVLHMSQQSRSSRQANHIFLFRPN			
					TCAGTAAA	SEFGKKKITYFGSKLYNALPENCKSFQNMFHFKRFVRN			
					GATCAATG	ALKQKVEQFVL			
					CTCATAAAT				
					ATTTATAAT				
					GTGAATTG				
					AAAGTTTGT				
					GAAAATTCT				
					AAACTATA				
					GCGCTTCCT				
					TCAAAGAG				
					CTACGCTCA				
					TTGGAAGT				
					GCAGTTATC				
					GCTCAATCG				
					TTGAATTAT				
					ACTTATTTT				
					TTTGAAAA				
					GATGAGAA				
					GGTTTTATG				
					CCTATGGG				
					AGAAGTGG				
					CTTAAAGTA				
					GCTACACTC				
					TCACGGGCT				
					TTTCCCTTC				
					TCCAAGAG				
					AAAAAAAA				
					GCAGTAAA				
					TAAATAAA				
					TAAA				
					TGATTTTTA	VNDVVNENGRCLDLCFVSTQDVAPSIAIAPSALVKNVS			
					TTGTGTACT	HHPPLVLLLADSEPCEPLTISPVSYDFNRANQQGIVEFLA			
					TGTTTTTAT	SVNWLEVLDTNDVENAALTFTHILGHVIERHVPKKVHG	cd01650	316	566
					ATATCTTGA	TAGPPWQTQELRRLKTSKRAALRLYSKRGTQSHRNHY			

					CCTATTTTGTCGTAAATTTAAGTAACCACACATCATTGGGACTAATTGTCTGTTGGTGTAAACAAATAAACAAAAAAAA	VSVNNKYKSTSRRCFRRYQQHIQQNLKTKPKSFWRYVNQQRKESGLPSTVKLDDEVASDRNQACRLFAKKFSSVFCDEMLSEEYVALAASNVPLVMETLNAFSVDDGMIARAATQLKSSINPGPDGIPSSFVKSYIEQLLLPIRHVFALSLTSGTFPSVWKNAIIFPVHKKGDRKDVSNYRGISSLCAISKLFELVVMEPLIAHCKQFISNEQHGFIPGRSTTTNLLSLTSHVSDSFTERVQTDVIYTDLTAAFDKINHLITIAKLERLGISGLMLRWFKSYLTERQLTVGLEGHFSETFTATSGIPQGSHLGPVIFLLYFNDVNYVIKGPRLSYADDLKMYLRVRSICDAITLQNDLDAFAGWCSLNRMVVNPEKCNVISFSRRKEPVYFNYQLFGKILQRVDHVKDLGVILDAQLTFKHHMSYVIAKASRTLGFIFRIAKEFTDVYCLKSLYCSLVRATLEYCSTVWHPHYQNGVQRIESVQRRFIRFALRRLPWRDPHRLPSYHSRCQLIHLETLQSRRDINRAMIVADLLQGRIDCPAILGQVDISARHRNLRANSMLRPPLRRTNFGQNVAINGLQRVFNRVAPAFDFHLPRHTVRRNFAEILTRRIV			
Kiri	Kiri-9_AAe	.	Aedes aegypti	ATGAACTGTTCCACTCCTAAAAGTTATATTGGGAAGCTACTGTTTCCCCTTCTGCTGCTGTTGTTGCTGTGTGCTGTTGATGCTACTGCTGTTGCTGCTGGGAGGCGTTGCTGTCGAATAAGATTTTTCTGCTGCTCTCGATTGTGGTGAATTAGATCAAATTAGACTTTTT	TAGAACATTCAGGAAAATCGAATTCAGGAACATAATTCGAGTTAGTTAAATAGTTTTAATGGAATCACCACAGTGTAACATTTCAAGAAGGAGTTTCCTTACGTTACATAAATTCAAATAAAGAAATAAATAAAATAAATAAATA	NICHINVQSLIARNFTKFHELKLTFIDSKVDIICFTETWLNHTISNTMIGIEGYKLIRNDRNRHGGGICIYLRNSLLYKVVSMSQSSSNDFYGTEYLNIEIKVGFDKVLLGVIYNPPNIDCIDVLHDILENCTINYDCSFFVGDFNTDILKFSSRSRRFGEMLDAMSYECMNSEPTYFHQSGCSLLDLFITDSPDFVAKLDQISMPGVSNHDLVFCSLRLSLNRVEDIVQYRDYVNFDSSALLDGFYNIEWNNLFCMEDPNEILNFINYHILKLHDDFIPLRRKKKNRTPWFNSDISHAFINRNLAYDQWKRSRTEIDRNTYKRLRNFANDLVRQAKINHDKQNFNVDLPSKQLWKNIKKLGISNDSSFLDQDCLNANQINDYFISNYINDSSSVPIPFSSDGFKFNVFHEEDIVNSIFSIKSNAVGLDNIPIRFIKILLPIALPVYKFLFDSIIKTSVFPRAWKNSKVIPIKKKGNSASLSNLRPISILSSLSKVFEKLLKIQITKFITDNDLLHPLQSGFRQNHSTNSALIKVHDDIARVIDRRGIAVLLLIDFAKAFDRVSHSKLVKKLSSKFSFSHNAVKLIESYLNERKQAVFYNGVLSDFSFIESGVPQGSVLGPLLFSLFLNDLPASLEFCSIHMFADDVQVYLCASDDVDMNDFARKINHDLNNILRWSQNNLLAVNPSKTKAMLICKLKNRPLPPSIYFDGKRVQFYDKLENLGVIFTSNLCWDAFINGQCGKIYGTLKKLNLTTKHLDTAMKIKLFKSLILPHFIYCDFLFSNATASALNRLRVALNACVRYVFNLTRYSHVSHLQDVLLGCQFSKFYSFRACLQVYIIIKRKTPRYLYDKLQFMRNTRTMSLVIPNHSSAYYGQSLFVRGINNWNRLSPTLKSTANIYGFKGGLLRELERLQ	cd01650	467	726

				CGTTTTC GGCATAT TAGCAGA CATTATTT AGCTGTT TTTTCTTT GTAATGT GCTTTTCT CTAATGT CAAATGA ACTTCAG ACTAAAA GTTAATC ACAATGA ACATATT AGTAGTT AGTAGTT TTGCTTTT CCTGATT TCTACAC TGCATGT TGCTTGG GTTCGAT GGGTAGA CTGCTTTT GAAGCCT ATACTGT TCAACAT TTAGATA TGGATGA CTTTTCCC TTGATAA TGCTTCTC CGAACAA TGTAATT CCACGTG TTGTTCTT AATGCTG CATTAAA					

				ACATGAC TGTTTGA						
CR1	CR1_Ele32	.	Aedes aegypti	GG		LRTKIDALYCNAIDCNYDVIILIETGLDAAINSVQLFGDN				
						FNVFRCDRNAKNSTKSSFGGVLIAVHRRFTSSSISITTGE				
						SLEQVCVSVLVGNQRILLCAIYIPPDKSKDVNVMDEHIE				
						SVRELRNLSSGNDYLLVFGDYNQPELCWKRCGDEIRPT				
						NSGSISAASSSLIDGMDFCNLHQMNIERNHLNRTLDLVF				
						TSLELSPEITCSVAPLLPTDQHHPPLELDLMLPNGVNNVP				
						TVSNATGELDFTKIDFEALSLFLSETVWHVLLDGLTADN				
						MAETFCSVIANWLNTNVPTRRPAVSPVWGTPLLRNLKR				
						RRNALQRRLRRHRSLENIRSFHRACDEYRELNSSLYKSY				
						VLRVQSNLQRNPKQFWNYVNSKRKSSTIPMSVFLNDSV				
						SSSPLEKCELFAAHFSSVFDAKPASPSEVDSAVSDVPRDL				
						LDLDIFQITPAMVSQAAKKLKCTFSPGPDGIPSAVYCRC				
						MSALAEPLCRIFNQSFVDSTFPAIWKQSFLVPVHKKGEK				
						RNVINYRGITSLSAASKLFEMIVSDVILAQAKGYISIDQH				
						GFMPGRSVTTNLLDFSTTCFEQLENGAQVDVIYTDLKA				
						AFDTINHDILLAKISKLGASNRLSSWLLSYLTGRTLRVK				
					TAGGATTTA	VDSCMSSSSFTSMSGVPQGSNLGPLLFLLFFNDVMLLLGH				
					GTTTCAATT	GCRLAYADDLKLYFVVRTVEDCVRLQALLDMFVDWC				
					CATGTAGA	RRNRMTLSVPKCIVVSFYRITRPIMHNYVVVNDVVLKRA				
					CAAAAGTC	DSFCDLGVLLDHKLTFNLHRSNVIAKANRQLGFIAKISR				
					AGATGGAT	DFTDPHCLKALYCSLIRPILESAAVVWTPYQLSWSLRLE				
					AATATACA	SVQRKFMRLALRNLPWRDPLNLPPYPDRCRLLDLDTLD				
					AATATACA	RRRKTQQVAFIAKLMNGEVDCSRLLSMLNFRVVAPRVLR				
					AATATACA	GGALLQQRFHRTAFGSNEPFTSMIRLFSMFEDLYDFDGS				
					AATAT	SNQFVSRVRSSDLI	cd01650	497	748	
				CTCCCAG	TAATGTTGT	MTPKCNPVMSSNGQQPGRSVESLMEAPGPSDPVERNAV				
				CCTGCAC	TTTATGTTG	TVLHSRPGPAVGCGTGVFQSATSGEYSSIIDNAPSPSIVL				
				TTCCAGG	TTGTATTTT	NSRTCNARKMPISVNAKKSDDLLIYYQNAGGMNCDVD				
				CAAGTAT	TATCGTTGA	KYLLATSDDCYDIIVLTETWLDTRTISTQVFTPEYEVFRC				
				GCGTGTT	TTGTAATTT	DRNPRNSRKSTGGGVLIAVKKKLKANIIENDLWCSTEQL				
				ATACCAG	TGTTAGTAA	WVHVQFFGHSLFLCGIYIPPDRARDETLIETHLQSVTSVI				
				ACGTTCT	CTGCTTCAT	EIAKPMDEIVVLGDFNLPGISWISSGNGFLHPDMNHSHV				
				TTSCGTG	ATTTGTTAA	HSAASRLLDGYSTATLRQINHVTNENLRSLDLCFVSAQD				
				ATGAATT	ACTAGAAT	VAPPISNAPSSLVKQVNHHPPLVLKLIVQHYSTSESIAQV				
				GCCAAAT	AAGATATA	TYDFTKADHEGILEFLAAVEWTEVLDTRDVDNAVSTLA				
				TTCAGCT	CATAAGTTA	HILGHVIERHVPKKTLASKGLPWQTRELRTMKTAKRTA				
CR1	CR1_Ele41	.	Aedes aegypti	CTCTATC	TACATCATT	LRRYTKNPTPRMRDQYVRLNAAYKRTSRHCFNRYQQQ	cd01650	607	857	
				CGCCGCT	GGGACCTG	VQRDLTAKPKSFWRYVNEQRKVSGLPPSLFLNGDVASE				

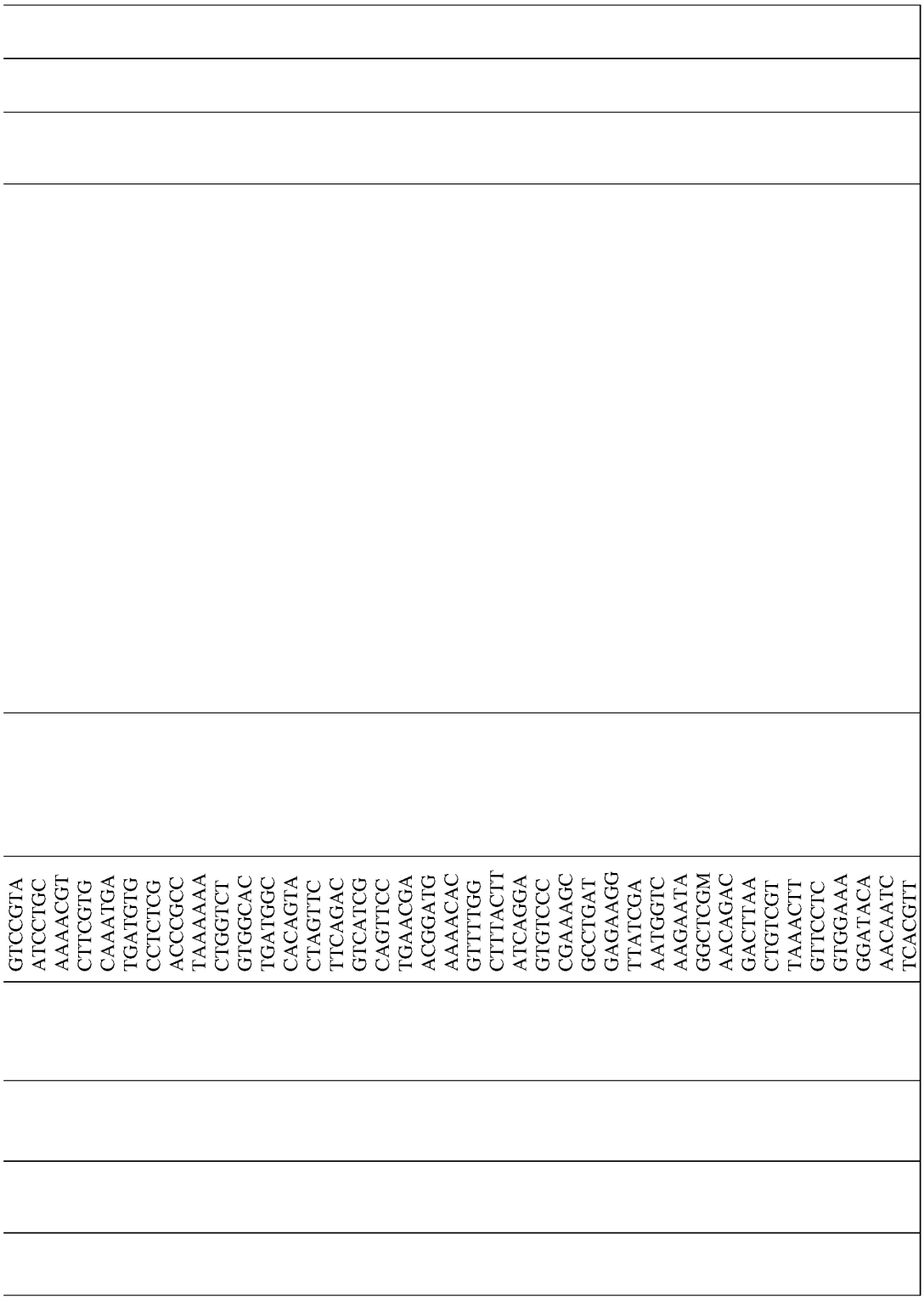
CR1	CR1-109_AAe	.	Aedes aegypti	GATGTAA CATCA	TTTTGTCTG TTGATGTCA TCCAATAA ATAAATAA A	PQSVCQLFAQKFSSTFCDEVIPEVLVTQAASNVPLVNDS WATLTVDDDMISRAASQLKSTMKPGPDGIPSAFLKRHID SLLVPIRHVFNLSLISGAFPSLWKNASVFPVHKKGERKD VNNYRGISSLCAISKLFELVVMDPLMTHCKNHLSDDQH GFISGRSTATNLIRLTSHIADSFAEGAQTDAIYTDLTAAF DKINHAIATAKLERFGISGNMLRWLHSYLIGRKLTVTIE GFQSNEFLATSGIPQGSHLGPLIFLIYFNDVHYTLKGPRL SFADDLKIYRRIHSDYDAMCLQQDLHAFASWCTLNRMT VNPGKCSVISFVRIRRPVVFRYELYNTVIQRVDSVKDLG VILDPQLTFKHHVAYVVAKASRTLGFIFRVAKDFTSVYC LKSLYCSLVRSTLEYCSVVWNPYYYQNGSDRIESIQHRFI RFALRQLPWRDPLRLPSYRSRCQLIDLETLQLRRDIARA MVVADTLQGRIDCPVLLQSLDLNVRPRVLRNNTMLRLP FRRTNFGQQSAYIGIKRVFNRVASVFDFHLSRQTIRRKF KTVFLSFLDE	AAPASSNILMYYQNVGGINSSLAEYNRAVSDSCYDVYA FSETWLNANTLSNQLFDSSYIVYRQDRSASNSNKRTGG GVLFAVRSNFKSRVLNSPNGSSIEQLWVAITTVDATLYL CVIYIPPDRVNDESLIEKHTIDSLDWVVSQLKPRDSIMIIG DFNLSGISWQYDXSGSLFPVASRSRTISLTSRKLLDAYSTA CLRQTVGIMNENSRVLDLCFVSEELRDASTVMQAPSPL VKTCRHHPPILLKMEICPQKRFHDVSESVFYDFRNADFN GMNDFFENVDWDDVLRDSDANLAASTISAILLYGIDQY VPVKVKREPSKPAWSNSELKNLKRVKKAALRRHSKFRT DSTRARYMQANTEYKQLNDCLYNAHQDHLQSILKANP RSFWKHVNEQRKESGLPPTMSDGLHEADSKEDIADMFR TQFSNVFTNEHVDPDIVNDAIRDVPRLSTFVQQIIITDDM VASASKELKSSTGSGPDGIPSLVLKRCAETIAAPLARVFN LSLASGVFPNCWKQSYVFPVFKKGSKQLVSNYRGIAAL SATSKLFEVIILQKLVQSYAHYISPDQHGFMPKRSTTTNL ACFTSFILRLMEAGQQVDAIYTDLSAAFDKMNHQIAVA KFDKLGMGNSMLLWLQSYLTGRNMSVKIGDHVSSPFR VWSGVPQGSHLGPFLFLLYMNDVNFKLKCLKLSYADD LKLYWTIKQPEDAVFLQHELEAFAEWCQINRMSLNVSK CSVISFSRKHAPFHFGYFLAGTQLERVSTVKDLGILLDSK MTFKDHVAYAVSKASSQLGFLFRYAKKFRDVYCLKAL YCSIVRPTLEYSSIIWSPYYYQNGIKRVEAIQRKFVRFALR HLRWRDPLNLPSYGSRCQLIKLESLVSRRNVAKACFVG DLLQGNIDCSTLLSCLDINVRRRNLRTHSFINIPSARTNY	cd01650	518	768

CR1	CR1-48_AAe	.	Aedes aegypti	AATTCTG GCACCAC TGCGTGT TTTCGTTT GTGTTGT TTATAAC GTCGCCC TAAAATC CTGTGAA TATTCGT GCAATCA ACTCTAT AAATCAG GGTTTAT TGCTAAA CTGCCTC GTTCCGC TTCTGTTA GCCGCAA GCATCAG GTTTTCG CATCACT TTTGGCA AGGTAGG ACAGTGA TTTGTGA TCTAAAT CCGAGCA GGCTAGA GAGTGGA AGCATCC GCCATCG CCCGATA ACATATT TCCCACC TCGCCGC TCTCCAC CTTCCAA	TAACAGTCT GGCGGATT AAGGTATT GATCCAAG ACGTAGCA ATTAAATA AATAAATA AATAAA	GLQEPMRSMSRVFNMCYHVFDFNVSRETNKSSYRRVLC	MRTKTNEFRTKMETCDYDIIVLTETWLRPDVVNAELAS NYCIFRCDRNDNTSTLQRGGGVLIAVKSTLRCSAVELTD YIHLEQVVVAVRLPDTTIYVCGIYIRPNSHFDVYDSHAS AVQEICNLSSTRDSIVVVGDYNLPHLTWHFDDDCNSYL PRNASSESEIMFTQSMIGSGLHQINSLRNMNGRILDLAFV SEXSFVELIDPPSSLLKVDSHHKPFVLRIEIETRTRAPVEE ADVDDFDFKRCDFAALNDMLVAVDWNETFGDAPLDRI TVLFYDKVFDIIRHHVPRRRRCSAVHSKLPWWTPELRH LRNIVRKARKRYFRNKSHDNRKNLKSLETRYSVCQENA FRSYTNRMESNLKRDPKSFWTYIKKLKNINRIPEELSYN GTTAETAEASANLFADFFSSVHNNHSPTFSSSARQNIQSF NVSLPMITISQHEVSSALKKLDVSKGAGTDCLPPTFLKE CADSLKTPISVIFNRSLREQVFPEIWKVASITPIHKSGSVH SVENYRGISILCCIAKVFEEIIHRSLYSAVRPIISDVQHGFV KKRSTVSNLMSFTNVLSNSVEKRRQVDAIYFDFAKAFD KVPHELVISKLKHIGLPDWITEWLRSYLTDRKAHVKVG RARSHSYDISSGVPQGSVLGPLIFILFINDLAVRLKSGKL MYADDLKIYREISSTLDCCALQSDVNELISWCTENGMEL NIDKCKTITFTRRQTYVTHDYAIGNNIIERVFSIRDLGVII DSKLKFTEHISTVTAKGFAVLGFIRRNSQSFRDVYTLKA LYCSLVRSVMEYAVCVWSPYHATHMLRIEKVQRCFIHY CLRQLPWNDPTNLPDYVSRCRLIDLETLTSRRTKIQRLF VYDLLSNNIDCSELTNELRFYAPTRQLRERQLLVIRPHRT TYGQNCPMSSCLRAFNDVGCYFDFNVSKCTFKHRIKYLN	cd01650	495	745

CCGATCG
AAAGCTG
CAACAAG
ATCAAAA
ATCAACC
CGCCCCC
ATTTGCC
CCANCGC
CCCACCA
ACGCCCT
ACTAYTG
TTTGCTG
CACCGAA
CGAAATC
AAAACAA
GAAATCG
TCCTGTA
ATACTCC
GCTTCAT
ATTGGAA
ATAAACT
ACTTACG
CTGCCTC
GTCTTTC
ATTCCAA
CGCTCCG
AAAGGGA
CCAACGA
ACGATCA
TCATACA
GCCTCAA
TCGCCGT
CCAAAGC
CAACCCG
TTGCTAC
CAAGGAT
TAATCGC
CAAACTC
TGCATCC
GTAGTAT

				CATCCAT TCACGTA AGGAAAT TCTGTTTT GGTTCTG CACTTTG TTTATGT ATTGAAC TGTGAAG CTCACCG TTGACAA GCCACCG TTTGTGA CAGAGAA CACACGC GCAGCTC GTGGTCA TTCCGAG AATCAAT AGTTCAC AGTATAT CACCCCT TCGTAAC TGGGAGC TTGTTTCT CGCAGCT AAAATGC AGGGTGT TTGTGAA AAATGCG CTTCCGA ACTATCC GCTGATT TTTGGAG CTGTGGT GGATTTT GTTCTGG TCGCTTTT GCTTCAG GTGTACC					

GGAATGA
GTCCTGA
TATCCAA
GCTGCCG
TTGCGAA
TAACGGG
TGCCTTC
ACTGGAT
GTGTAAC
GCCTGTA
GCGGAAT
AATGATT
AAAGCTC
GCTTTAC
TAAGGCA
ATCACAT
CAGTGAA
TGCTGCC
TATGAAG
GCATAAT
CGAATCA
ATGAAAA
TCCGAAA
TTCGAGA
CAACATT
TTGTCCG
AAATCCG
CTCGGAA
TTGCAGA
ATAACCT
TAAACGA
ATTCCGG
AGATTGT
TGCAAAC
ACACCGA
TTCGAAA
TATTCAA
CGTCGCA
CATTTGA
TTCGGCT


GTCCGTA
ATCCTGC
AAAACGT
CTTCGTG
CAAATGA
TGATGTG
CCTCTCG
ACCCGCC
TAAAAAA
CTGGTCT
GTGGCAC
TGATGGC
CACAGTA
CTAGTTC
TTCAGAC
GTCATCG
CAGTTCC
TGAACGA
ACGGATG
AAAACAC
GTTTTGG
CTTTACTT
ATCAGGA
GTGTCCC
CGAAAGC
GCCTGAT
GAGAAGG
TTATCGA
AATGGTC
AAGAATA
GGCTCGM
AACAGAC
GACTTAA
CTGTCGT
TAAACTT
GTTCCTC
GTGGAAA
GGATACA
AACAATC
TCACGTT

TGTATCC
TTCAAAA
TAGGAAT
GAAGCTT
GATCTTA
GGGCAAA
AGCAATG
ACTTCAT
CAACCTG
GCCAGTT
GGAATTC
GTTTTCG
CGAATTC
GAGAACC
ACAGTTC
CTCAAGG
GCGGGGT
TCTGGAA
TCCCTCG
GAGAACC
AGCCCGA
ACCACAG
ATCGTAA
CAGTATT
GAACTAG
TATCCCC
ATCGTGT
CCCATAG
ATCTTCA
TTCTCCTC
GTGATGA
GACCGTT
TTAAGCC
ATTTGAC
AATCTAC
TACCAAA
ACGTGCG
GGGT

CR1	CR1-64B_AAe	.	Aedes aegypti		TAGCTTAAGAAACAGTCTGTGGAATGTAACAATTCGAGACGGTGACAATAAATAAATAAATAAATAAATAAA	EQSIVCVKLQKSSIYICGIYLRPNSQPVLYSSHSAAIQQLSERISRFDSIVVVGDYNLPQLIWQTDDDINGLLPSNASSEQEVTLLETMVASGFQQINNIVNSNGRLLDLAFVNDNNDVELLEPPTPLLRMDNHHKPFVLRIDVNDSCTQVGDSNHADDFDFRYCNFDELKSAIATVEWATLFIGKDTDETVCTFYEVLHRILDEHVPRKRRRRTHPFKHPWWTSELQHLRNIVRKSRRRYFQSRTVENRDKLRLLEASYIDCQTASFRRYVARMETTAKEDPSAFWTFIRNRKRVNKFPAEMAFRDTVANSPEGIANLFADFLENVHCTNPPTFSPDSLRNCPTFDLNLEPFDFAQQDVLSALQKLDTSKGPGTDNLPPLLFKECADSFQIPLTIIFNKSLQSRTFPMLWKTASVIPIFKAGSTRGVENYRGISILCCLAKIFEELVHNVLYTASQPLISQFQHGFVKKRSTTTNLMTFTSFLSTEIENKHQVDAIYFDFSKAFDKVPHDLAISKLRHLGFPNWIADWLRSYLTNRKAFVNINGTYSRVISVTSGVPQGSVLGPLIFVLFINDLCFRLKSGKLFYADDLKIYRTIASHLDCCALQADVNELELWCQENGMELNIKKCKSIAFSRRQSRTEFDYKIGSEPLERVESIRDLGVIIDTKVRFNDHISVINAKAFAALGFVRRSTNDFNDIYALKSLYCSLVRSILEYAACVWAPHHTTQIVRLEKVQRSFIRYALRQLPWSDPVNLPDYPARCMLINLEMLDARRNNLQRLFVFDLIMGNIDCPALLEDVQFYAPLRQLRERDLLQIRRHRTSYGFNNPLSRCFRLFNCASALFDFNVSKYVFKNRLKDF	cd01650	413	664
CR1	CR1-8_AMi	.	Alligator mississippiensis	TA	TGATCTCAGCAGTATTCCTGCCCAGAGCAGGGGGGTTGGACTCGATGATCTTTCGAGGTCCCTTCCAGCCCTTAATTTCTATGATTCTATGATT	GIYYRPPHQEEELDLEFSRELMEAACSRDMVVMGDLNYPDICWEEHLARSDRSCRFLTCVEELYLVQEVNGPTRGKALLDLVLAKGDDLVGDLRIEGNLGDSDHEMIEFTIRRRAGKSASRAEVLDFRKADFNELRKIVGETLWEEGLEGKGVQEGWSFLKGTILRAQKRSIPVCRKGGKGAGKPSWLTRGVKECLKKKKEAYGQWKQGVVTKEEYTSIARICRERIRKARAVTEIRLVTEIKDNKKSFFRYIGSRRKTKGNVGPLQNGTGQLITDSQEKAELFNEYFASVFLHANGGRPPGRILDGCGRGDCQPMVSVELVREHLERLNVFKSAGPDGLHPRVLRELAGVIAGPLAKVFEDSWCSGEVPEDWKRANVVPIFKKGRREDPGNYRPVSLTSIPGKILEKIIKKSICERLAEGMMLKDSQHGFVAGRSCLTNLVSFYDQVTNRLDVGDAVDVIYLDFKKAFDMVSHDALIKKLGDCGLAESTVRWVGNWLRGRTQRVLVDGSVSSWREVASGVPQGSVLGPVLFNIFINDLDEDVESTLAKFADDTKLWGKAVTLGDRMRIQADLDKLERWAERNRMQFNKDKCRVLHLGRSNQRHEYRLGGDVLASTMAERDLGVMVDHRMNMSRQCDAVVNRANSILGCISRCVASRAKEVILPLYSALVRPQLEYCVQFWAPHF	cd01650	375	626

						KKDVENLERVQRRATRMIKGLEDKPYEERLRDLGLFSLGKRRLRGDLVAVYKYVRGEYRELGEKLFTRAPQGRTRSNGHKLLEDGFRLDVRKNFFTVRVTRIWNRLPKEVVQSPTLEVFKRRLDRHLAGII			
CR1	CR1-9D_AMi	.	Alligator mississippiensis	TCAGAACAAGACACATGCTCGGGTGGGATGGGCTTCACCTCTCCCCCAAAGGTAAGCATGTGTTCTCTTTCAGGTTGGCTGA	TGACCCCAGGGTTGGTGACTCTTTTCTTGCCTGCCAGGGGCAGGGGGTCAGACTAGATGGCCCACTGAGGTCCCTTCCAACTCTATGAATCTATGA	SPPAGFKLGSSGEGEDEYGGSHGTPNQETPTRSAQVCSKDPIRQQFKGGPGTATRGIRCIYTNARSMGNKQEDLELLLASNNFDLVGLTETWWDSSHDWVVGIEGYRLYRRDRVGKKGGGVALYVKEQYTSSLIKLGSEGGQAETLWVRIQGARGERDLVVGVYYRPPHQDDELDQEFSKQLEEAVRSRDMVVMGDLNYPDICWEEQSARSNRSRRFLTCIQDLHLTQEVRCPTRGNALLDLVLATGDDLVGDLRIQGNLGDSDHQMIKFTIQRRAGIIPSRARVLNFKKADFNRLRGLVSRVLGAKNLDALRVQDGWTFLKETILWAQSESIPWQRKGGKGARKPLWLTRGIQDTLRAKKEAYRKWKQGVASKEEYTSLARTCRESVRKAKATLELRLASGIKDNKKSFFRYVGSKKRVRGSIGPIQDGQGGLVTDSRDKAELFNEFFASVFLNGNLDEPSTYITDTNKGNISRPTVSADTVKRHLEGLDVFKSPGPEGLHPWVMKELAGVIADPLARIYERSWCSGQVPEDWKRANVVPIFKKGRKEDPGNYRPVSLTSIPGKILEKIIKEHICGSPLANVMRRGNQHGFLAGRSCLTNLVSFYDKATNCLDAGVEVDVIYLDFKKAFDKVSHTILINKLRGFALDDYTVDWVANWLKGRTQRVVVDGSVSTWKDVGSGVPQGSVLGPVLFNIFISDLDEGVGSTLSKFADDTKLWGAANTLERRNQIQADLDRLGKWAESNRMQFNKDKCKVLHLGRKNCHHTYRLGGDLLSSTEAEKDLGVTVDSRMNMSRQCDEAVGRANRTLSCIRRCFTNRSREVILPLYAALVRPQLEYCVQFWMPQFKRDADNLERVQRRATRLIRGLQGRPYEERLRDLNLFSLSKRRLRGDLVAAYKFIRGEQHEVGDDLFTRAPPGMTRNNGHKLLERRFRLDIRRKFFTVRVARIWNGLPREVVLAPSLGVFKRRLDRHLAGIT	cd01650	536	787
RTE	RTE-1_ATr	.	Amborella trichopoda	TAATTAGTGCTTATGCACCACAAGTAGGATTAGATGATAGAATTAAGAGACAATTTTGGGAGGACATAGA	TAATTGGGAAAAGGCTATAATGATGAT	MGETILDFAMAYDLAIANTYFKKRDEHLITYKSGSHKSQIDLFLVKKVDRPSCKDCKVIPGESLTTQHRLVVMDVCIKTWKRRGESNGCPRTRWWNLKNEKIDLFRDKLVEEGKWNSEHDVNVIWNEMADCIKKVAKEVLGESRGKRNFDKETWWWSDEVQKAISEKRSCYKTWQCNRNDQNLEEYKNAKKNAKQIVSKAKLKAYDNLYNKLGTREGEKDVFKLAKIREKKCRDLNHIKCIKSEDQRVLVKDDEVNERWRVYFQKLLNDEHNGNIVVEGTVGTSEGGGHRYIRRIRMSEVKEALKRMKIGKALGPDGIPIEVWKCMGEIGIDWLTKLFNKVLRTKKMPDEWRKSIVVPIYKNKGDIQNCANYRGIK	cd01650	353	603

				TGGAATA			LMSHTMKLWERVIEQRLRHETSVSENQFGFMPGRSTTE			
				GTCCAAG			AIFLLRQLIEKFRAKGKDLHMVFIDLEKAYDRVPRDLIW			
				ACATACC			RVLEKKGVSRHYIDLVRDMYDGALTSVRTTSGETSEFS			
				AAATGGA			VTIGLHQGSALSPYLFVLVMDDLTRHLQDEVPWCMLFA			
				GAGAGGA			DDIVLIDETRNGVNTKLELWRDALESKGFKISRTKTEYM			
				CCTTTTTA			ECKFSNNRSRDEEVVKIDGQEVQKRPHFRYLGSIIQENG			
				GGAGGGG			EIEEDVAHRIRAGWAKWRCATGILCDRRIPMKLKGKFY			
				ACTTAAA			RTAVRPALLYGAECWADKKQYTHRMSVAEMRMLRW			
				TGGTCAT			MSGKTRRDRJIRNESIRENLGVAPIGDKMRESRLRWAGH			
				GTAGGGA			VWRRPSTTPVRRCELVQVEGLKRARGRPKRTWLEVVR			
				AGGATAG			KDMGTYGLTEGMAFNRAEWRTRIHVADPN			
				TAGAGGA						
				TATGAAG						
				GGGTACA						
				TGGGGGA						
				TATGGGT						
				TTGGGGA						
				AAGAAAT						
				GAG						
				CAGTGTG	TAACCACC		MFINPIPLCALPRPKWSIHRRRNLSNMIYPPLSTPQEFTVT			
				CATCTGA	ACCATTCTG		GGLWNCQSAVKKADFITAFAKTKSLDFLALTETWITPD			
				TTGTGTC	TATATTTCA		NTATPAALSAGYSFSHTPRASGRGGGTGLLISLAWKFCV			
				GTCGCTT	CTCTTTAAT		LPSPNPTSSSFESHAITVTHPSKFHIIVLYRPPGPLGSFLDE			
				CTGCCGT	TTAAAAAA		LDTLLSSLPEDGTPAILLGDFNLSPETSQLSRVASLLQSF			
				CCCCCGC	AAAACAAA		GLSLSSSPPTHKAGNQLDLVFTRSCSIPQLTVTPLHVSDH			
				GATTCAG	AAAAAAAA		FFLSFPLSYLHPLNSSNSSPTVTFRRNLSTLSQSSLISTALS			
				ATAAGCG	AAAAAACC		TLPPPASFSNLPVDLATSTFNSCLSQSLDSLCPLVSKPAR			
				TATCTTA	CCCCCCCCC		SSPPCPWLTESLRSSRTSLRAAERKWRKSRSSDDLTAYH			
				ACTTGAT	TTTTTCATG		TLLAAFTSALTSAKASFFQSKISACVSNPRKLFSTFSSLLT			
				TTGTCTCT	TCACTTGTA		PPPPPPPSSLSADDFLAAFEGKVATIRNSFAHPVSPPPPRD			
				GCTGTTG	TTTGTCTTT		RPTATLTSFEMLEDSDVLQLITSHRATTCPLDPIPSKVLQ			
				CTAGTTA	GTCCTAATA		SISSEFLPYLSSIVNSSLSSGHVPSDFKMARVTPLLKKPTS			
				GAGAACA	CTGTAGCTT		DPSDVTNYRPVSLLPFLSKTLERAVLKQLTSFLHQNNLL			
				TAGTTGT	ACTCTTCTG		DPHQSGFRSGHSTETALLGVTEALATARASRLSSVLILL			
				GCGTAAT	CCTAGTTGG		DLSAAFDTVNHKILLSTLAGMGIAGTALSWFASYLADR			
				TTAGATA	CTTTGCACA		TYQVTWKGSASASHPLVTGVPQGSVLGPLLFSLYTRSL			
				ATCTTTTT	GGTTAGGTT		GSVINSHGFSYHCYADDTQLFFSFPPSVTQVNDKISSCLA			
				TAAACGT	AGAATAGT		DISTWMASHHLKLNLNKTELLFFPCKTSLLRELSITVDG			
	UnaL	.	Anguilla	GTCTTTA	GTTCACTGT		TTVTASHSAKSLGVVLDDQLDLKEHIKATSRSCRFLLYN	cd01650	513	
L2	2		japonica	CTGTTGC	GTGAACTGT		IRRIRPYLTTHSTQLLVQATVTSRLDYCNSLLASLPACAI			773

				TAGTCAG	GTTCTTAGC	QPLQMIQNAAARLIYNLPKFSHVTPLLRSLHWLPVAARI			
				CGAACTT	TAGAAATA	RFKTLTLAYTAANRTAPIYLQDMIRFYVPARPLRSAEAG			
				AGTTGTG	GCTGTACA	RLVTPPTRPKGSQSFSTLAPQWWNELPVPLRTSPSLPIFR			
				CGTTAGC	AAATAAGT	RGLKTHLFRLYLD			
				TGAGAAT	ATTATACCT				
				CTCTGTA	TTCTGAACT				
				GTTGACT	TGTGTTCAG				
				CACTGTT	CAGATGCCT				
				GTTAGTT	ACGACCAT				
				AGTTAAA	GATATGCA				
				CGCGTTA	CTTTTGTAC				
				GTGAAAC	GTCGCTTTG				
				TGTGTGT	GATAAAAG				
				GGGGGTT	CGTCTGCGA				
				GGTGTTT	AATAAATG				
				AACTGCC	TAATGTAAT				
				CGGTATT	GTAATGTA				
				GTTGAGC	ATG				
				TAATTTC					
				AAGTAGC					
				TTCACCT					
				GGTGCTT					
				ATCTGCC					
				TTAATGA					
				AGGTGAT					
				GCAAGCA					
				CGTAATT					
				GTCACCC					
				GGTATTT					
				ATAGCTC					
				CAGCGGA					
				GGCTGCC					
				ATAGGCA					
				GCCTCGT					
				CGTCAGT					
				TTGTG					
R4	R4-1B_AC	.	Anolis carolinensis	ATGMTGT	TAGACACTT	MVKQTLKKSILARSSPRIIRTTASATPGHLSTSGLQEPGPI	cd01650	476	742
				TGAGAGC	GGGAAGTG	AEQPGHGPVGQLEAKQLQKWQKFSMPQNRTVMNCYY			
				TGTGCTG	TCCGACGTG	RSEPQKRGYQKRMYQLWKLEYPDSTVTEQRLADQRRY			
				GCGGTAG	TGATCCAAT	ITKKKVLSEVELEEIKSNCETEEQTTVETAAEDLMTSEPT			

				TGTAAAC TACCAGC AGGTTCA ACCAAGC CAGAGTG GTCTCAA CTGAGAA GTAGAAC TAAGTGC ACCCAAC CCATTTA AAT	ACAACAGC CAGCAGAG TGTCTGCTG TGGACTCAT CTTGTTGTG TTTCAA	EHMEPEENAGIFQDDEHEEPVAIAYTTPGTLTMKQQEL KENFLTHAALKEGRQRLPTLKTVPRKHLAPLMNDVNA ALATVEITTMEETNQLAYSTAVIVTQELGIQQLKPARKM IGKPKWKIRLELKIRSLRSDASNLKNMKEQKLKNVKTK NYLIKKYWLNTRKIEEALEIVKQQITATARKIERYEARVI QYKQNQLFQSDQRRFYQSLNQQVAVKNEKPEKTATNA FWKDLWETEKGYNKKAEWIKDFENKFSENKMRTLEITT EMITERVKKVKNWTSPGNDQLHGFWLKYLSSLHSKMA EQFNEMLQTGNISKWITTGKTYLIQKDPAKGAIPGNYRP ITCLPTMFKLLTGIIADQIQDYLEENNILPVEQKGNKRRS RGTKDQLLIDKMILENCKSRKTNLYTTWIDYKKAFDSLP HCWIIKCLDVIGVCENIRKFTENAMKQWKTELFVGNES YGTVNIRRGIFQGDSLSPLLFIIAMVPLSVILQKTNLGYQ TARNSPKISHLLYMDDLKLYGKSETEIQSLTNTVRIFSTD ISMEFGLDKCATVALKKGKITHSEGIEMPNGQTIKCNQP EAYKYLGILQLDNIKHGHVKNVVSKEYIQRVRKVLKSK LNGGNTIKAINTWAIPVIRYTAGIVNWTQAELDDLDRKT RKLMTIHYSLHPRSDVDRLYLPRKSGGRGLLQVKQTVE EEKHALADYVKGSQEPTLREVNSSKLLQVQKTKSEYHK NTIQSRRENWRKKALHGQFLGKIESQIDKEKTWLWLTN GTLKKETEGLILAAQEQAIRTNAIKAKIEKSTTDPKCRLC KEADETIDHILSCCKKIAQTDYKQRHNTVAQMIHWNLC HKYHLPATKNWWDHKPEKVTENEHVKLLWDFRIQTDR VLEHNTPDLTIVLKDKVWIVDVAIPGDSRIEEKQLEKLT RYEDLKIELQRLWHKPVKVVPVVIGTLGAVPKDLGLHL NTIGADKITICQLQKATLLGSARIIRRYITQS			
L2	L2-19_ACar	.	Anolis carolinensis	ATGCATT ACAGAAA CCTGGAT ACGAGAA GGAGATA ATATTCC CTTGCAT GAACTAA CACCGTC AGGTTTC CAGATGT TTCAAAG ATCTCGA AAGGGGG	TAAATCAG ATCTAATTT GCCAAATA GTCACTGTT GAATGTTTT TAGTTGAAT GTTTTTATG TTGAATGTT TTTATATTG TGGAATGA TATTTTAAA TTGTAAGTC GCTCGGAG CACTCTGGT	MGIECIGLTLAIRESVTILLVYRAPGAPNENIPKLLDLVA DWALKFPSLIVLGDFNVHADRLLSSSIARDLVQSMETIG FSLMVTGPTHRAGHTLDLIFSIGIQVTLSSVLEIPWSDHYI LNTQIKSSPPQHPETEQIFARPRKLMDPIRLSSVLKELDY TADSLHERVENWNKNLSTALDIVAPKRAFRPRRNRSPW FTDALRQMKLERKRLERKWRKKPDETSKLALETHQKIY EAATREAKKEYNSSLIASARSRPAQIFKIVRNIINPTHPAK DMASQDITAETFQSYFINKINLLREGLPPTEDTLRELEAF WPSSDKLLELFVPLEMEDLEKTLAAARPTTCSLDPCPSR LVKECWEGLRDPLRDIINSSLEQGIFPECLKEAVVHPLLK KPSLDAAIPSNYRPVSNLPFLSKVIERTVAGQLQQFLDET AGLDPFQSGFRSGHGTETVLVAITDEIRRQLDRGGSALL VLLDLTAAFDTVDHDLLIHRLAMSGVRGLALXWFNSFL RDRRQRVEWSGQSSESSCLCCGVPQGAILSPLLFNIYVR	cd01650	388	648

				GTCGAGG GGGGTGG GAATAGT GGCTCGC GATCATT TCTCCATT AGAACTA TTCCTAC CCCTGAT ATT	GGAGAGCG ACTAATTAA GAAATAAA GTGAAGTG AAGTGAAG TGAAGTGA A	PLAGLVRSFGIECYQYADDTQLLLRLEPGTTTIPENFKLC LEAVMSWLRASRLKVNPAKTEILWPGQPPELIQSLPDFD GEVLVPSATVKSLGVVLDSSLTMEAQITAISKQAFFHLR QARKLASYLSTEALATVIHATVTTRLDYCNALYAGLPK TKTQKIRMVQNAAARLLAGSSIRSHITPILKQLHWLPIEH RISYKILILTFRARNGQGPLYLRDRLIPFFHQWSPRSSQE NLLYVPGPREVHLEATRRRAFSTYAPILWNSLPPYVRA MSELRPFVKALKTWLFGCAFK				
Pene lope	Posei don_ Ac	. Anolis carolinen sis		GGGGGAC ATGAGAG AAGCCTC CACGCAG GATCGTG ACACATC CGGGCGT CCCCTGG	TAGTATTAA TGCCAAGTT TTTGTTTTT TTTTCTTAA AAAAAATC TCTATGTTT GCAAATCC ATTACAACT	MPAIDVGETSGENASGTWPHSPKDIQQPCDPGHESLRQ HIEHLYGEKLFQDTRRLEKLRTRKAHLLCSLTFLLRCRD TDTIPQFLAAKRTFKTPQAHRIYNRLERNLLRERIHTIRK ELANTDKELLHLHINISQKMNTQDWDKIDSLTYKKMEK DMVLHTKRQKQKFDKCRKQQPKPELDKSRTIINLTDRQ LTEDQVSILEKGGNFAVTTTRIPVENIIANVESAIYQLPEE EAEEVRMETARILRNAKLPPSNITRKERQAIKDLNSDPEI IILPADKGNATVIMETKQYKEKIRQLLDPTIYKKLKQDPT NKITRKTNTLIKNSSINFDIRQQLCKSEALPPRLYGLPKIH KDSIPLRPIVSAIGSPTYNLAKFLATQLQTHIGLTAHYIKD STHFIEKISNLNLSTKDILISFDVVSLFTKVPVADTLTLIK QNFPEDITALFHHCLTTSYFQWDTGFYEQKDGVAMGSP LSPVVANFYMEYFEKQALETAPKKPTVWFRYVDDTFTI WSHGEEELSKFLDHLNSIHPNIQFTMEKEKEGKLPFLDV LVIRKPNQQLGHTVYRKPTHTDRYLHKNSNHHPSQKRS TIKALTDRAQRICEPHLLQGELNHLNWALQANGYSTTDI RRAARPRTSHESQDKDPPRGKVFLPYIKGTTDRIGKLMK KHNLQTIYRPTKKIQQMLRSAKDKRDPLSSAGVYRIPCS CGQVYIGTTKRSAQTRVKEHERHCRLIQPEKSAIAEHLM NQPGHRILFENTKMLDHSNNYHVRLHREAIEIHKHVDN FNRKEETMKMNKIWLPVLQNSKIKTVNKKQYSENRGFP DMNQPRAVNDSKQRMPQRQEEDSR	cd00304	470	554
Ving i	Vingi - 2_Ac ar	. Anolis carolinen sis				MTIISINIEGMSAAKEQLLYELCRDKKCDVLCVQETHRD KQQKRPKVPGMILVAERPHAQYGSAVFVKPGLQVSSVH NTEENNIEVITVELNNTITTSVYKPPNEEFCFSSPENFDSH QRTVVIGDFNSHSHVWGYAQEDKNGEAVLSWAELHKL SLIHDSKLPPSYNSGRWHRGYNPDLLFVSDSIVQQCSKS VGPPIPNTQHRPIVCQLFPTIRPLEVRFKKRRFNFRKADWS KFSDMLDDKIISVAPIPEEYERFIKIVKTCSRASIPRGCRT HYLQGITPDTAALLEDYYRLHNEDPFSEQTLQTAQSILSS	cd01650	459	714

CR1	CR1-9_AG	Anopheles gambiae	GCAACGT CTTTGTA GACGGCA GATTCTC TCACAAC CAGAAGC GACGTGC AACTTGC AACATGC AAGCAAA CAAACCA ACCAACG GACCTTT TCAGGAC CAACT	TGTACCCTC GGTTTCGCT TCTGACACG AGAAATAA ATAA	ISEAKKEQWIKLITEVDMGRSSQKAWRLLRRLSNDPSQT NTHANIKADQIAHQLLKNGKPKLSTKVGRKPIVRQPDN ENNHLHEPFTSTELDMALNKCKNGKAAGLDDLRMEQI KNFGPKARCWLLKMMNNCIASCQIPKIWRKARVIAILK PGKDRNDPKSYRPISLLCHLYKVLERLILNRIMEKIDPCL IPEQAGFRKGKSCTSQVLNLTQHIEDGFERQQITGAVFID LSAAYDTVNHRLLLRKTYNITKDYHLTRFIGNLLQNRSF FVEFQGQRSRWRKQKNGLPQGSVLAPSMFNIYTNDQPL PEGTESFIYADDRAITAQAGSFEMVEQKLSEALGALTAY YRENQLIPNPSKTQTCAFHLKNRQASRALRITWEGIPLE HCSTPKYLGVTLDRALTYKKHCFTIKQKVGARNNIVRK LTGTTWGSQPDTVKTSALALCYSAAEYACPVWNTSHH VKTVDVALNETCRIITGCLRPTPLEKLYCLAGIAPPDICR EVAANNERTKALTSPAHPLFGYQPARQRLKSRNSFLRST EILAGTPQQARVQKWQAKTRNLNQWLTPDERLPPGHTE DWATWKALNRLRSGTTRCRANLKKWGYKVESATCDC GEEQTTDHLLQCILSPATCTMEDLLTATPEALQVASYW AKDI					115
			TTGACAG TTGACAG CTGACGA CWGAGTG TGTTTTCT CACTGTG CTCCGCG TAAATCC GTGAATT ATTTTGTC GTTTCGG CCTWATT TTCGTCA AACTGCC GCTTAGC TGTGACT TAAGCTG TCCCTGG TAGTCCT GCCTTCA ATCGTTC GTCACAA	TAAATGTTC TGTTCGTTA TCTAATGTA ATGTATTGT AAACAAAT TTTGACTCG AGAGGGCT TCATAGTCC ATCGATTAA TAAACTAA ACTAAACT AAACTAAA	MEGVCCDCSVPLESLECGIMCSFCDAVYHSACTKSPFTV IEEVKRTASLHWSCVGCSNALGNPHSKAVKTTGMQVG FQAALTAAVEVMKAALVPPVMHEIREGFASFAAAHQA PSEHFNVPPPDALPSGKRRRLFRDVVASTEAIVVDNAPL YPVNTDTDNTHQRNRPSLPPIITGTDTTTTSIPTVSQLPRT NYMWLHLSRLAKTVTVDQVVSMVKSQLDTTDVIAFSL LKAGTTISSVSSLTFKVRIIAALRDKALTAGSWPVGLGV REFISLPQRPSHHRPDLTNTNTPIPVRHSPMHPVQLSPDM NCTLISPAPQKQLQPHQTPLKCTKLLFNTFFLSNNHTTVS SKGTVPLNLTSHLFTRNNNTSTTRSLTDDLAPTYTRIPLN VRSTSITVYYQNVRGLRSKADEFRLTVLEADYDVVVLT ETWLDPSLPTALLFGDAHCVYRCDRNAANSSLSRGGGV LIAVSSALRSHALPVLAQTLEFACICIQLPTYRLYIAAAY LPPNHSTDEGKINALLDTVNSICDTLGPNDRFVLVGDFN QPALSWSPAQSNHETPFVFFEPHTSSVRSAQFVDGLHQN ALYQLNNNTNSMGRILDLAYGNWSSAATSSLIRVSEHP LLPIDCYHPPLEFDLEITASSMNHATTVTEIKLNYARADL TKLKRLICSFNQSFECSNYTSIDHATNDFSEFMRAALREC APIKKPHRGPPWGDRTLHALKKAKRAAYDYLRANRSTP CRLYYNGAHALYRRYNRVCYRRYIRQTERSLRKYPRRF WSFADNMRKSTGLPGTIRYKDNCAHSPLDICELFATRFE DSFISNITPPEAVASALVNTPADAIHVTLPIVTETLVTQKI	cd01650	909	8

						ERLKSSYSEGPDGIPATILKRCAVSVAPVLTKIFNESLRT			
						GTFPSLWKVSWLTPIHKKGSKNEAVNYRGITSLAACAK			
						VFELIIYEPLMASAGKYISTNQHGFMPKRSTTTNLMQFV			
						SGCYKSIDDGMQVDAIYTDIKAAFDSVSHNILLAKLDRL			
						GLSPPLVKWMKSYLCGRSYAVRMGSHQSRSIGASSGVP			
						QGSNLGPLLFLLYISDLCTALPDNSCLLYADDAKIYREIR			
						EPEDHLQLQATLTEFVSWCKRNALSVCIDKCAIISFSRSR			
						APALFNYTLDGQNLERVNCVKDLGVLLDAKLSFEEQID			
						QVVASGNRLLGLVINMTRELRDPMCFKTLYCALIRPLLE			
						YASIVWWPASTRALARLESIQRKATRFALRDWPRRLDY			
						RTRCLLLGIPPLAERVEHTRLAFITGILNGTIDCPELLSRI			
						HLYVPARILRRRPMLAVAETRTTFGSRNPLLCMCRLLNS			
						ANDIYEPGMTITELTSLLSVRNAFNNNNI			
					TCCTTTATA	SPRIPRIFGLQCSPGFSSTTAAAIPSQAINRCDPRTRITSVY			
					ATCCTTTTTC	PYRFSAANFPPVDEPAATTAQSNSTSTPITPLHHAAERFE			
					CCTGATACT	TLPSVASPITVSAPLPLHASFPHPIDDLRFYYQNTRGLRT			
					CTTCCCTAT	KLDDLRLASADTEYHVLILTETWLSENIPSSLAFDNNFVI			
					CCTCTTTTT	YRCDRNSRNSSRSRGGGTLIAVSRKLAACEISPASDSLE			
					TCTCATCCT	QCWIRLKLPSCSLFIGSIYLPPTLSADASTLNTLDASIDEII			
					CTTTTTTCC	SRMKKNDRLLLLFGDFNIPSITWHASPDPAPPFLPLLVPAS			
					GACTATCAT	SAHSSSLFIDIVRYHSLSQLSSVSNFRQVQLDLIFTFPKNQ			
					ATCTAATAT	SNATSTCTYVSQAPVPILQIDPHHPPLELSLKITGSLTPSN			
					TATTCACTC	RSQHSSSRLTTAPSLDFRKTDFNLLNTSLSNIDWSQLYSC			
					TTCTTTTCT	KDVNLAVSTFTSLFLSTLPQCCPKRNPPRSPPWSDATLK			
					TACGTTTGT	SLKRDKSRYQRAYWSTRSVNALRTYKYAAAAYRLYNR			
					ATTTTATCC	TRYNGYILILQTRLRWYPRAFWKFMNARRKGSSLPSTM			
					TAATATCTG	HLESVYASSPEDICSLFRNHFSNIFSPPSQIPNTASQCLSY			
					TCTTTCATG	TPQDVIDVSCFGFDEHSVSEALSKLKPSFSAGPDGIPPSL			
					TGCTCAAA	LKKTAEVIAPLLSFIFNLSLRLCCFPSSWKSGWLCPIFKK			
					AGTTCGCTT	GNSADVSNYRGIVLLPACSKVFEIAVHSVISFQVKSWISV			
					TAGTATTAG	NQHGFMPSRSTVTNLMQFASHVIPSLDSGLQVDTIYTDF			
					GAATTTAG	KAAFDTVPHHLLLLKLKKLGFSNMLIDWLYSFLNSRSIR			
					ATAGGTTA	VKTDTTFSSSFCPSSGVPQGSVLSPLLFVLFVSDVHAILPS			
					AGTTAGGA	EGHLLYADDLKIYLTISTHSDCLRLQEFLNSFSLWCSYN			
					ATTAGTTAG	QLSLCPEKCSVISFCHNRNPIHYEYSLSNSVLSRTSLIRDL			
					ATTTAGGA	GVLWDDKLSLNEHLESVVASASRTLGLICRLTRDFRDPL			
					ATTGTGAA	CLKSLYCSLVRSTLEYASVVVGPVRSGWILRFEKVQKRF			
					ACCATCAG	TRFAFRRMAGPNAAMPSYESRCNLLGLESLEVRRSQAR			
					GCTGACAA	VLFIGRMLNGLIDAPTLLEKIGLYAPSRNLRARSALNMV			
CR1	CR1-10_AG	NT_078268.2	Anopheles gambiae		TTTGAATAA	SRQTTFGSNEPLLLCVRTFNTSTDPPHPRL	cd01650	634	882

					ACTTGAAATGAAATGAAA				
CR1	CR1-8_AG	.	Anopheles gambiae	TATATATCTTCAATCGAAAGGCATCCCCGAAAATTTTGGCGCTTCATGGACAGCAGACGACGCTCCAAGCACAATCATGGGCACGTCACCCGAAGATATTTGCAACATTTTCGCTGAACGATTTTCCG	TAATTTTATTTTCAAAATACATTTTCTATTCTTGTGCACCTGTTCTCCCCTATCGATTTATGTACCTTACATAGTCTATAAGCACTAATTTAAGTACATTCATGTAGGATCTTCAAGATCCCGATGGATTAATCAATAAACATATTAAACATATTAAACA	MPSPIVDQTVISRATSYTPLNMIDYTIPTFDIRTVEKCLKSLKPSTAPGPDGIPSYIISKCYQSLAPVLTNIYNTSLSSGIFPRLWKTSWMVPIFKKGDRSCASNYRGITSLCAVSKAFELLLYGPMLSATSNYIMDNQHGFVPKRSTLTNLTEFVSFCKRNLDSGGQVDAIYTDLKAAFDLISHDILIAKLQKLGFSEQIVKWLHSYLTERSYKIKVNDHTSREVLGTSGVPQGSNLGPLLFILYINDVGQLLAEHRFLLFADDVKLFAPINDTNDCINLQHAINLFSIWCRDNAMVLCIEKCRVLSFYRSRSCTRFNYQIDNISVERTDTFRDLGIILDTRLTFNDHLENIVSRGNQLLGWIIRSTQGFRNPMTIKTIYCAYVRSVLEYGSIVSDPCTEQWSNRIEAIQRKATRYAVRLLPWRQGDVLPSYHSRCLLLGIQSLKKRREIAKCLFISGLLNHHIDAPTLLATIEFNAPSRNLRTRQFISVPRYRTRFGQSDPMSAMCRTFINNFDLFDFNIPQNVFRDRLRTQLT	cd01650	92	340
RTE	RTE-1_AC	.	Aplysia californica	TTAGCCGTGCAGTCTCGGCATTGAGAAGAATAGGACTGACGTGGTGCTAGCGTGTCGTACGAGGCGACTAAATGGCGAACCAGGTAGTGACTCCTGAAACAGAGTCATCCTGCCTCCAGGTCGGGG	TGATGATGATGAT	MMQGGESRKEAAGQKPPPSSKSKSLRIGSWNVRTLYEAGKTAQVLQEMHRYQLHILGISETHLNQSGQKRMKEGELFVYSGNDSDGPHREGVGMIISKTAQKTFRGWEAHGPRIIMASFGTKKELKVRDGRKKVINMNIIQVYAPTNDAEVENKDDFYNRLQNVIDKLPGKDINIVMGDLNAKVGSDNSHNEEIMGRHGLGEANDNGERFQNFCTFNNLVIGGTIFPHKRIHKATWVSPDGTTENQIDHFCISRKFRRSLEDVRVLRGADVASDHHLILGKMKLKLRRYAQVAAGNRLKYEVSLLKDPVKKNEFKLELKNRFQALQDLDECEVNEHWDKLREAVNATCQSTLGPKDNTRNEWISQDSLDKVKRRRELKAQVNNSRTRTEKLAAQQEYTEVNREVKNSIKKDKNTYLNGLAEKAEKAATDGHLRIVHQTTRTLSGRFSKPTLPVKDKEGKTIFNPEGQLNRWREHFDQLLNRPPPENQPEILPARNDLPIDTEPPSREEIAAAIKTLKSHKAAGPDHIPPEALKAEKDTFVEILYALFEKVWQKEEIPSDWKEGHLIKLPKKGDLSNCNNYRGITLLSIPGKVFNRILLERIKLATDEKLRGNQAGFRKNRSCVDQIATLRIIVEQSLEWNSPLVVNFIDYEKAFDSVDRETLWKLMRHFGIPEKIVKLVKATYDGTSCQVFHE	cd01650	562	819

CR1	CR1-1_AC	.	Aplysia californica	GTTGAGT CACGAGG CTAACAA CCTCTCTC GTAAAAA GAACCTG TTACAGA AACACAT ACAAGCA AAAACCA TTATCTCT GCTTGGC TGATCGT GGATCTT CTCCAAA TGGAGAG TTT		GQLSEPFRVQTGVRQGCLLSPFLFILAIDWVMKEATRGK RNGIQWIPWQQLDDLDFADDLALLSHTHAQMQEKTKE LNSLSRSVGLRIHPGKSKVVKLGKVSNLAVTVEDTPLEE VDSFTYLGSVIDGKGGTEADVKARTAKARSAFKQLSKV WKASNISLRTKIRLFNSNVKSVLLYGCETWKTTKTVLN KVQTFINRCMRSILRIKWQDRIRNEELWERTGQKDIEVEI GLRKWRWMGHTLRKPSGNITRQSLQWNPQGARGRGRP RETWRRCVERDMREMKLTWNELGRMAQDRDKWKEL VCGLYPARGERH				
				AATTGAA CAAATTA TGGAGGA ACTGGGA GTGACTC GCAAACC CTTTGAT ATCAGGA GATTATT TGGTAGG AACAGAA CAGAGGA TAGCAGA GACAGCG GTAATGA CCGGGGA AGCGAGA AAAGGCC AAGACCT ACACTAC TGAAATT TGAAACT GAAGAAG	TGATTTGTC TAGCCTGCA CGCAATGC ATTATTGTC GTTCATTAC CAGATAGTT GAACAGGC TCTACTTAG AGCCTCAA ACAACTAA ATCGTCAAT GTCAATGTC AA	MSLKSMVEIYQPDIIAVTEVKPKYVRFKIQESELSLDGY DYELFHNLEEEGRGLVLYVRKALKPGLCDDLDNDFCEK RFVKCVLKDGVSLLIGLAIYRSQHSPTAEEDARKLNKLF ELVSLRKETHKIIMGDFNFRQIDWETETSNAGEKHIATK FLKATKDAFFIQHQKTPTRYRKGETPNILDLVFTNTEQL VTDIKVEAGIGKSDHLCLLMTLSLSAGTAPKVTRRNFRK TDVEKLREDLGAYDWESELANKNVEDTWSVIKSKIEGA IGASTPLMNISGRKNKRYMDKDTLEAVRKKHRAYRKW HRTGKTTDEIAYNKANNKARKACRKANRNYEKKLAIE AKKNPKVFYSYANSKIKSKTGIADLNKDDGTKTKSDTE KAELLNKFFQSIFTNEDTGPLPDFEEYDYNTELQDFNIET VTVKELLVSLDVNKATGPDGIPPRVLSQAADILALPLRIL FRRTLNDGVLPVEWKTAFVSPIFKKGNKAAVNNYRPVS LTCVICKVMEKLVRGHIMDHIVNNMLISKDQHGFVTGR SCITQLLEALDDWTSVMDEGGSVDVIYMDYQKAFDSVP HRRLIQKLSACGVKGRVLKWVRDFLTSRKQRVVVNGSR SEEAEVTSGIPQGSVLGPLLFVIYINDLPRGLKTTAKMFA DDTKLYVRSDTVNGPQDLQDDLDTLQEWSQRWLLRFH PQKCCVLKVGKDNDNEYFMDQAVDGSTSRHKLNVTKR EKDLGVVVDNKLTFKEHVAQMTSKANRIVGLIRRSFDY LSEDMFVLLYKSLVRPLLEYAHCVWQPFEKGLCSEIEQ VQRRATKMLSHLKEKPYSERLKKLKLPSLEHRRLRGDM IETYKYIKGYYITERPAFTKATTTQLRGNSEKLQKQRSR	cd01650	479	737

				CAAGTGC TAATACG ATGAGGA AATTTAT TGAGGCA AGAAAAG ACGCAGC AAAGGAC AAAAATG AAGGAGA AGAGGAC AAGTTAT ACTTGAC TATTGGG ATGCGAA GAGATAT KACAAAG CAGGAGA GAGAGGA GGACTAT GCCCTCT ACCAGGA GTTGAAG GCTAAGA AAGAGAG GTAAAAA AACTTGG GGGACAA TTTCGCTC ACTGGGT GAGGAGA GACGGCA AAATTCG GAACATT GGGAAAT ACCCAAC GGAACGT CAGAAAG AGAGGGA GACAATA		LCVRANFLENRVNTIWNNLPESVVSAPNVDTFKRRLDE YWKDLPTIYNPTCQN			

				CAAGAAT						
				AAAGCCA						
				AAGACAG						
				CAAAGAA						
				TGTAAAT						
				AGTTTTG						
				GCCATTG						
				TTCGCCTT						
				TTTCATAT						
				TTTAATT						
				ATTGCTC						
				ATCTCGT						
				TCCTGAC						
				GTWCAGA						
				ATAGAAG						
				TTTCACTT						
				CATTTTCT						
				TCTTTTTC						
				ACCTACA						
				TTTCATTC						
				CAATCGG						
				AATGCAC						
				ATACTTT						
				AGAGAAT						
				TGTACTT						
				GCTTGTA						
				TACAAAT						
				GCAGATC						
				AATTGTT						
				GAACAAA						
				AGAA						
Tx1	Tx1-1_AC	.	Aplysia californica	TCAAGCG	G	MSIRIASLNCRGLNKKLKRRAIFRKCTQYDIICLQETYVT	cd01650	507	772	
				ACAATCC		NTTAREWSLEWSGDFFYSEGTNKSKGQIILVNKHFKYK				
				GTGAGGC		TVELCIKTERILGLSFSXDDVSYVAINVYGPVVKNDKLD				
				TAACGCT		FIDKLYSVLKDEKADHTIICGDFNVVHNNTLDIVTGEKH				
				GACACAC		DDMSTQKFRDWIANCQIVDTWRGVHGEEKDFTWSRPT				
				AGGCGGA		PFTARRLDYIFCNTELFPFVIDSKHAFVYGTDHKAVITTV				
				GACACAG		RTEKFKRGPSYWKFNNSLLKDKEYVNSMEKCIDEYLER				
				GCAGAGA		TLDETDKTIRWELLKVMIKSHTIDYCIRKCRALKLRENLI				
				CACACAC		TEELLKLNDQVVKNPQDLTLLKQFESKKKELEIYQLHK				

				CCTGCGT GCCGTAA GGCAAAC AAATATG GACTCTT TTTTCCCG CAGCCTG ACCGCTC CCGCCCC CCAGTAC GCTCACC CCGAAGC CAGGACA CAAGGTC TCCAAGT GTCAAGA CGAGGAG ACATAAA AGACAGG TGAGCGA CGAGCCA GCTCGGT CACCGAG CGTGAAA NCTCGAA AACAGGA GGTGAGT GGCACGA GAAGAAG GCAAGAC ACGGTAG ATTATTTT GATTATG AGGAGCC AGTGACT ACCGAAA CGTGTGA GCACGAA GAGGTCG GTGCGTG		ARGAQIRSRQKWIEEGEKNTKYFLRMEQMRGGSNVITS LQDDKNNNLYGHVDIVDSIKNHFERLYKKDHNIETNRV DDFMQGCSFPVLDDVDKEICDSELSIVELDAIVNKLNIDS APGIDGITTPFYKFFWSKLRKPLFTNFQQCLETGELSPTQ RRGVISLLHKGKNLRRDNIKNWRPITLTNTDYKIFSKILA IRLQRVMPVLINENQTGFLKGRSISEHIRTLDDIIYVASEK NLPGMVVSLDFQRAFDTVEKGTILATLKKFNFGEGFIN MVGVLMANTECCVQNGGWLSEWLNTERGIKQGCCTSP LLFVLVAELMSIKLRNNSSISGMSTILSGEIINFCKVLQYA DDTTLILKDQEDLASALLDIDSFGSISGLKLNKDKSQGM WIGSDRTRTEKLGGISWVPNGNTIKILGIHFSATEEASRIE LNWIDKIENMKNIMKMWMKRNPSLYGKVIICKTFLLSQ ISYIIQSLSLPERVLNTIDTLFFKFLWQKRHSNKKAREKV KRTVMCRSIEEGGLGVISVKDQQKVFLLNWLKRMATV TEGRPNNIPNRISDLAFTEVGGCDYFFSATANCKKTGKL CHASKFWESVMDTWIEMASIKKEPSSTIEILMEPLFNNR NIKYKGKSLYLKRWMNSGIKYVYHLYDRRILKSKEELS HITGTYAAFDFDYYALMNALPTQWKQDVSDNNTEELL VNTREVINTNGKKTVLEILSKSNREIRQLLKYRRHAEICG RNFWMRKMNIDITCKYNIAHRSTKESRLRLLHFKILHNI FPSNILLNRLGIKNTELCEHCGVKDVIEHMFIHCTLLKNF WKYVFHTICLKTSSRFPMTDCNILFGFSERGQG			

				AGTGGTGTGTACACGTTTCGGTAGCCATTGGTAACGACTTGGACTGGTAAAAGGACTTTGAGTTTGGTTTTGCGTCTGACTGTTTTGG					
Hero	HERO-1_AC	.	Aplysia californica	GTAGAGAGGAAGAGCCCGAAACAGCGACGCATTCTCTCCTGATGATCTCCATGCTAGAGTCCGGGATACGGAGCGTTCGGCGGTGAGCAGTGCATACGGCTGTTTAGACTACGTACCTACCAGTAGCAGGTGTATGCTACGGACATCCTGGAAAGACAGCACGACAACCCGGAGTGACGGGTGGACATC	TAAAGTTGGCAGCAGAGGAAAACAGTCGAACTAACATGGAGACCTCACATCGGTGAGGGGGCTCGGTGCGTCAGAAAAGCTGTAGAGAGGAAGAGCCCGAAACAGCGACGCATTCTCTCCTGATGATCTCCATGCTAGAGTCCGGGATACGGAGCGTTCGACGGTGAGCAGTGCATACGGCTGTTTAGACTACGTAC	MDNDLSVKRTMQKENQENSFMCEQCYRKFDTWRGLRVHQGKVCLKKRRQCRSSDCKTRSKSSQDENHSGAISASVDSPFTLSQQEMEQQTEPGLRKPRILWPAANEKAKYKTLEEEVQKKLKEVKKSSNRGDAKEILSAFATIIYDSAIKEFGICEGKKKTEKKSGLSRRQRVLAQLRKQKRDLQKQKKYAPPEEIEGLRILFEDLKKKSRDLQRHERRCTRRKERKREREQFLKNPYEAAKKLFTEARSGKLECTKEELDAYVKQTYSDSRREEPLPYMRGLKRPAAPVVKYYLGPIREKEVDEFVKKARAKSAPGGDGVSYKVFKYCDTLRHTLFEMLRGLWKDKEIVKDWCRAEGVYLPKEQNAKEIGQFRPISIINVTCKIFMGILAKRTVAYLQSNGYVDETVQKAGVPGIPGCIEHAFTIWDTIQEAKKNKESLNVVWLDLANAYGSVPHKLLMKAMDFFYVPQEVQNIMREYYSNFQMRFSTANFTTEWHRLEIGIAAGCSISVIWFILVMEMLLRSAECCEEEAKVRSPKKAFMDDVTLLTRDGDTMQSVLNRLDKLITWARMKFKAKKSRSLTINKGKQKQLKFTIAGEQMPTVKEQPVKSLGRWYSGTLSDRSQGVEIMKQAENGLKAIDQTKLPGKHKIWCLQFALYPRLAWPLTMYEVALSRVEMIERKCNSYIRKWLGLPRTINTSSLYRHRGALQLPLTSIVEIYKAGKIRTVLMLRESRDKEISGKPPDVKTARKWKAEVAANEIISALEHGDIVGPAQPDRLGLGNGDFRPFRKMSQRDRRKAATRHMQKMEAERREIHLIQCSQQGQMTRWEEHVVERKVGWSDIWKWNTSRLSFLLRSTYDVLPSPVNLVRWKVREDNKCRCGKLGTMKHILSNCNLALNRYTWRHNEVLKVLTKMATQQVEEGKYAPKIQKHGPTKIAFVPQGSKVAKRKKANEAMKDQSGGTWEVAADLKGFERFFPIPTTKKPDIVLWEEKDKEVHLVELTVPHEDNIYAAHERKENRYEALVE	cd01650	355	614

					QCEEAGWKATHFPVEVGCRGFVATSARKWMRVAGLT AKKSNELTNSLQETAEKASHWIWLKREDSSWNES					
Ingi	Ingi-6_AC	.	Aplysia californica	GGGAAAG ACCGAGG ACGTCGG GAAAGAC CGACGGT GGCAGGG GCTAACT ACCCAGG TCACTGC AAATNTA AGTTACG TCAACCA NTAGAAA AGACAGC CCCGGGA TCCCCCG AGAGGGG GGAGTAG AGGATGG ATCCACA AGGACGG ATTCA	TAGAGACA CTTGAGAGT GGGACTGA TGTGAACA AAGATATA TTGCACTTG ATTTGGAGT CTACATAAT TTGATAAGT AGGCACCA CGTCAGAG TAGTCCTAC AATGGATTC CTGCCCACA CTGGGATA CCGGGAAA TGAAAGAG CTGATGCTC	MLFTNDRIPYELRRRLGNQGLPQVLVGNQAGLVDGYW KKKKIRKQRKWKNKKKTPPTTGNGKEKLIDFNILQANV CGLDKKKTQLGKILYERKVHIALLQETLHSSCDINTGY TPYPCKCEGCQGIITYIRNDIQGEVELLNWHPTDAHKVT LWYGNSKLILYNIYCPPSSTLNFLEQSTIFKKTVIAGDLN GHSPLWGYADQNNTGKRIEELCXSSNLILLQDEDSTKTL LHRRHGTLHRPDLTLVSADLHNYCSEEVLDCISSDHRPV LTKISILKHRQRKRRTRWNFKKANWDLFRKTAEDTVKL DDLDNKSIDALNSELTSTILDAAKKSIPQGCRTNYKPFW NERLEKATKTKNDARHEYEKDLKNMDKKIRYKKAVAE AKLVTKESKKEAWTQKCEGLNLHQGGREAWSLLNNLT GEKRKENPKPLHTETEVLTSELKKAEHFNKHFASINKSS KRNDLDRGLLSVLKGKERDLMTPSIFQDFLTKRELDCAI KKLKKRKSPGPDKVHNEMLQNLGEKAKLLLLLKLYNRT WTEGNLPKAWKLATINPILKKGKKANDPKNYRPISLTSC IGKLCERILNSRLYWWLESSGLISQCQAGFRSKGRTEDQ LFRLTQKIIDGFQRGEQTTAVFVDLQQAYDRVWRKGLL LKLQNMGIKGNMYSWIKSFLTDRLIQTQMDETLSSKAV	cd01650	550	806

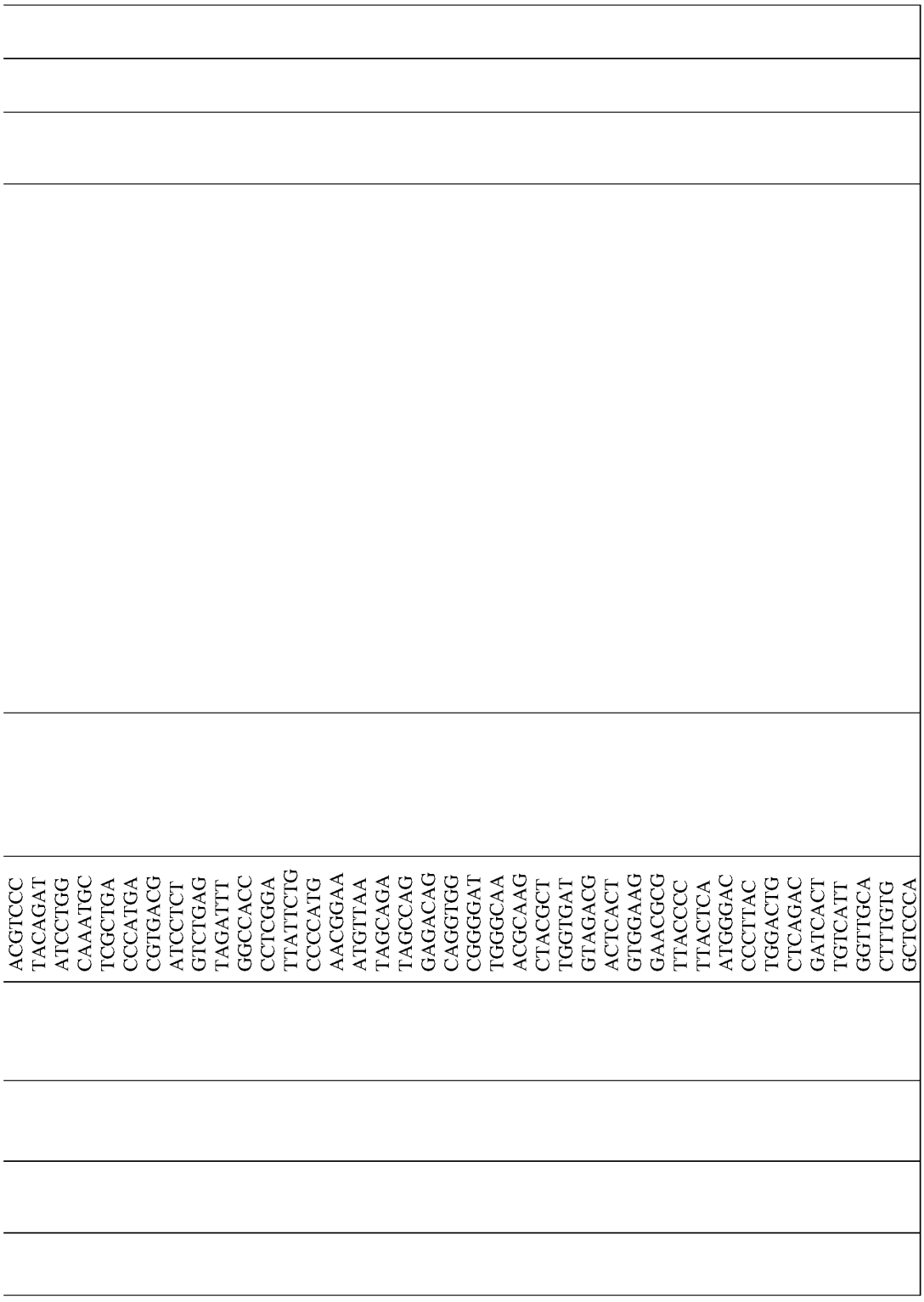
TTTCTTAC	TGGCAAAG	LEEGLPQGSSISCTLFLVFLNDVSDVLKTEKALFADDLV	
TCTTTCTA	AAAGGAGC	LWHTGNSTTISRRRLQDDLNSLGTYCRLWKMKVNCEK	
TTTATTGT	TAGCCTCCC	TVYSVFTRSYNLANAEIPLKIDGTNLKKENNPVYLGVQL	
TTTTATTG	TCAGCCTGA	DSKLNLKQHTNNLRKKAMKRLNLIKRLASTNWGSDKN	
TTTTACTT	AATTCCTAC	TLRGLYLGYTRAVFDYNLVLQNICSKTTKNSMDTIQNQ	
TATTACA	GAATTACA	ALRLISGGMRSSPTAACEIHTNIEPLETRRKRAALEVYER	
TTTATTA	CCACCTGCT	SKRLENNHPNRVLVDKWKPNQRLKNTKSVLDTVLNLQ	
ACTTTTCT	GCCAAATG	EKHHLPERREPLERVPSTLPPHLTIAKPEIRKTLLDKSNK	
TTTCTCCT	ATCAAGTC	NTYPVALKHSALETIDSYPPTWIHSYTDGSAFKATINAG	
CTCTCCA	AAACCTCA	YGSVINLPSGIKKKKSLTPAGPSAPISLPNNRLSPTQQTTS	
TCGACAA	AGGAAGAG	TTCLTLTHLLPQTSLSLQTAFQH	
CCTTTCTT	TGGATGAA		
TGACCTG	TGATTGGGC		
GAGTTGG	CACAGGTA		
GACGGGT	CGACAGGA		
ATTGGTG	AGGAGGAT		
GGGTGGA	GTATGGCC		
AAGCAGA	ACATGAAT		
GGTTCAT	AAACCACT		
TGTCAGC	GGTCAAAG		
ACTGTAC	ATCCAATCA		
TGGATGT	ATGAGTTG		
GACTTTA	AGGAGGGG		
CGTAGGG	AGATCAAT		
AATACCC	CCATAATCT		
AGCTACA	TCCAACTTA		
TACATGT	GAACACAA		
ATTAAGC	CATGTACCC		
ATGCACT	CTCAACCA		
TATCGGA	ACATCTCAA		
GAGCCCT	CCGAATTG		
ATGACTG	GAGTAAAG		
CGGAGAA	CAGTCTGCT		
GGTGCAG	GCCTGCCCA		
AGACCAT	CTATGTGAC		
TTGCGTC	TACCCCTCA		
TTGCTTGT	GAGACAGT		
TGACAAC	GGAACATC		
ATTAACC	ATCTGTTCT		
CTCTGGG	TCTGTAGAA		

				ATCTGCC ATGCCTT CCCCTGC CCCCATC TCTCGGA TGAGAGG ACTTGGT TGCTCTG GGAGGGG GGGGGGG GGAATCT AGCTTCC AATTGGT ATTATTA TTACTATT AGTATTA TTATTATT ATATTAT AATTTTT GTATTTC AGCAACA CACAGCA ACTATGC TCTTAAT AGCGTCT TTCTCTTC TTTATCTC CTCAAAC AACTTCT CAATATA ATACAAC AGACACC AACATAC CCATTCA TTTTTCTT ACAATCT CTCTGTG ATGGAGT TGGGTTT TTGTTTGT	AACTAACT GATTTAAG GGGATGTTT TCTTCCCAA GCTACCCCA CACATCTAA TTGCCTCTA CTCCTCAAG ACAACAAT TAGAACAA ACATGCAA CTATTACCG AATGGCAT CATGCAGA AGAGTTAA CGCTCAGAT GCTACTGG ATCGCTAAC AACAACAA CAA				

				GGCATTG					
				CACCTAC					
				AGCCACT					
				GCTTTAT					
				CTCCTAC					
				CCCGGGA					
				TGGAAGG					
				AAAATCA					
				CAGAACC					
				ACCAGTT					
				TTAG					
RTE	RTE-6_AC	.	Aplysia californica	CTCAGTA	GTAAAGC	TGATGATG AT	MDHITSINKDXDVKGSSERTASSGRVSHAQQVTPDRHQ	cd01650	549	806
				AAGAGAC	TGGGGCG		DTAHKRKNQTRNIATWNVRTXLEKGKLDNIKKEMKRM			
				TGTCAGG	ACCATCC		NVNIMGLCEVRWTGAGRILSDSYNIIYSGGDKHERGVA			
					CCGACCC		VIMDMPTSKTLKGYLTISDRVMLVRLKGSPFDTTIIQAY			
					ACCGCGT		APTSESTEEEIEKFYDEINQAKTHCKSQDIVIVMGDFNAK			
					CCACAGG		VGATRHEDIVGPYGLGXRNERGEKLIEWAILNDMIIGNT			
					TGGCGGA		WYEQPPRRLWTWKSPGDSARNQIDYIMIKKRYKNGML			
					TAGTGGA		NVKTRPGADCYSDHVPVVGKFRIRLKKTHMSKNKNMK			
					ACGGCCT		LNMALLQSNKEMQDKFRITVQNKFGPLEEIQDIEKQWE			
					ACATTAT		QLKTSINEAAKDVIPERTTKAKKKWMTEEILNLMEERR			
					GTAGGGT		TNKNNXEKYVNLHKEIRHKCNLAKETWLKEECEEMEN			
					AGCTGCG		HSNANSKDLHRKIKEVCHTRTRGSTGCLKSKDGNIIMD			
					AATAACC		KDKILERWTEYIGDLYNDKRDEQRRVENNNEGPPIMKE			
					AATAAGC		EVIKALRKMKNGKAAGPDNITTEMIKALEDFGITTLTKL			
					AGTCGCG		LNNIYATGHIPSDLSKSIFIALPKKPNANECETHRTISLMS			
					GACCAAC		HITKILLRIIMMRIRNKILPEIAQEQCGFIEGKGTVNAIYVI			
					TGGCGAC		RTTIERALEVQKDLHLCFIDYTKAFDKVRHDKIIDMLEV			
					AGTCTTG		LNIDGKDLRIIKNMYWDQTAAMRIENDISEYIQIKRGVR			
					GAGAGCG		QGCVLSPDLFSLYSENIMRHIEEMPGISIGGHNINNLRYA			
					AGGTGGG		DDTVLIAEDEEHLQNLLNVIVQESDNMGLSLNIKKTETM			
					ACAGGGT		VISRXTPIPTCNIKINDTTLKQVQHFKYLGTIISSDGRCHT			
					GGTGGCA		EIKSRICQAKATFHKMKQILTNINLSLETRKRVLKSYIEPI			
					ACCCTGT		LLYGCEAWTINKQLERTIVATELWFIRRMLRIPWTARKT			
					CTCTATA		NEEVLVEANYRKQLITNIRKRQANFFGHTMRKEALENI			
					AAATACC		VTTGKFDGKRGRGRQREKILDKMTTWLGAARTTDTIA			
					TCATGTG		AXKDRXKWRSMIAHASRQGTG			
					GGTGGCA					

				CGAGAGGGAGGCGCCGAAACCTGCCACCTGCCCCCACTGCCTTGTGGGTTAGTCCTTTTGGACAAAGGCTAGCGGAGTAAACCGTACAAAAATCCAGCAGAAGGCAAGGGGAAACCACTGCTGTATTCTTTCCCTAGAACATT					
Ingi	Ingi-7_AC	.	Aplysia californica	GGGCTGTAAGGGTTGAGGGTGAAACCCCAGACGCTCTCTGGCAAGCCAGCCTCAACCAAGTCAACTCCCCGTCCTTGTCGAGCAGCCTGGGGACCCCTGCTCTGTGCATAGGAGCAGGGGTTGCAGACGGAGC	TAAAGGGATTTGAAAACAACAACAA	MHTSLHNPXTNKMPLRQMKLYELGCRLGYQGPRSLPSPGGGSDKSATGTHTAGIGATSNTQTLNILQLNISSIQNKKMELAKTLSDLRIHVALLQETQLPKRIELHLTGYTMYHCKCTKCQGIATLIRNDITAEVSHPETDDETDMQEVXVQQSDGTTYTIYNVYCPPRXTLQLNFPNTDHKHTILAGDFNGHSPLWGYSDTDXTGTVLEELLRSSNLISLQNDKTPPTLLHRASGTLSRPDLTLVSADLHPHCSMTVLDDIGSDHRPILLQISSNRKREPTNPIRRSWNYKKANWELYQQKSESRLLDVCTTNPIDECYKDITTALKTSAKEAIPRGNPKKFRPYWDKELEQLTKARKKARRRAEKNPTAENRKHYNYLTAKVKLKTKSAKRQTWREHCGELDLRKDGRKAWNLLTNLSGTKRRKNPAPMKPTEKGETCRAEVMNKYFASINKKGNQNKLSHAIKQLTKRAERTPKANLEVFEKEFTLTELKAALKKCKKRKAPGPDHVTNEMILNLTDSSLSILLGFFNRTWNEGQLPTEWKKANITPILKKDKPASQPSSYRPISLTSCLGKVVERMVNDRLYWWMESTGTISTSQAGFRKGKRTMDHLIRFAQKTSDAFQDKDNVVAVFIDLRQAYDRVWRQGLFYKMQKLGIAGRMYSWIKNFLSCRTIQTSCDSKISSQRTLEEGLPQGSALSCTLFLLFMNDLPDVISCQKALYADDLLXW	cd01650	545	802

				TCTGGCC AGGAGGT ACCATAG CCCTGTG AGGTGCG ANGTGAG CAACGAC CCTTCGG CTCGTAG AAGGGTC CGTACAC ATCCACC ATTATGG TAGGTCT GGCGTCA CAGCAGG CCGACCT GTTCATT CAGGAAC TCACTGT CGACATC TTCTTCCC TCCCTCTC GTGGGAG CTCTACA GACGGGG CATGGTA CCCACAG ACTGCCA GATGACT CATCGTG AGATGTC TCCGCGG ACCCAGA AGGCCTT GTAGCTG TAAGGAA TAGGCTG CCTGGAA ACAGATG		SKGSDMRALNQAINADLTTISSYCQLWKMEINTDKTVF GLFSLKNKVTASQVNLKINGTTIRHEDYPKYLGVQLDR KLSLKEHLGELTKKATSRLNLLKHLSSLNWGADKATLR QLYIGYVRSVIDYSLPLQSIASQSSLAQLDRVQNQALRF VSGAMKTTPTSACEIHTNIEPLDLRRKSATLETYERFVRL EDTNENRKMADEWEPRTRIQKTSFMREATKLLEEFDPP ENREPITNISSEPPYWQTPEITIRTRLLDPSCDKTTPEEILR TSALETVDSYAETPIHAYTDGSATNATRNAGYGSTIKYR NNLGEKKQTQLFGPCGTYSSNYDAERIAIVKTLEQIAQD FFEGAVTPGDVVVFSDSMSVLQAVEGGGDYRGDGNILS LVKQMGTSYGVRVTLQWVPGHSGVPGNETADHLSKTG ASLPQTDTTASLQTCRGIIQAKSRDEWLNRWANAETGR SLFGYMDKPNPKDPSNLLCRADQTLIFRARTGHLHTNK HINRINPMWEPHCRHCDYHEETVEHLLLHCPNLAVLRR RLLHTPTDLQSLLYSGVDQMTRTCGYLREALRCQERNA RS			


ACGTCCC
TACAGAT
ATCCTGG
CAAATGC
TCGCTGA
CCCATGA
CGTGACG
ATCCTCT
GTCTGAG
TAGATTT
GGCCACC
CCTCGGA
TTATTCTG
CCCCATG
AACGGAA
ATGTTAA
TAGCAGA
TAGCCAG
GAGACAG
CAGGTGG
CGGGGAT
TGGGCAA
ACGCAAG
CTACGCT
TGGTGAT
GTAGACG
ACTCACT
GTGGAAG
GAACGCG
TTACCCC
TTACTCA
ATGGGAC
CCCTTAC
TGGACTG
CTCAGAC
GATCACT
TGTCATT
GGTTGCA
CTTTGTG
GCTCCCA

RTE	RTE-4_AC	.	Aplysia californica	AGAGGCG ACTAAAA GTGCATA	TCTCTCA GAACTAG AGTGTGC CTACACC CCCTATG GTTGTCT ACTGTCT AAGAAAT CCCCGTG AAGGTTG ATGGAAA ACAGGAC AGCGAAC ATTGCAC CTTTGTTC TCTTGTG GCAAGAT GTTTGGG CACAAGC CAGTGTC AGGACAA GACTACC ACCTACC TGAGATC CTACAGG G	CTCCCCT GCTGGTG GGGGGCC CCGTAGT TTTACTAT GGGGGAG CAAAGAA TACACCT GCCGTGA TACTGCG TGTCGTA	TGAGAAAG GCAATGAT GAT	MKDSGQSRKDSSDPKPSLSNKSLRIASWNVRTLYEAGK SSQVVQEMKRYRLHILGVSETHWIQSGQKQLSSGEQVL YSGREDKYHAEGVGILSKYAQKSLRGWEPHGERIIMAS FTTRSKSINMNIVQIYAPTNDASDEDKDAFYSQLQGVID KLPRKDVNIIMGDANAKIGQDNVGYERIMGQHGLGQM NENGERFANICAFNSFVIGGSVFPHKRIHKATWVSPDGV TENQIDHFCISRKFRRSLEDVKVQRGADVGSDHHLLLA KLKLRLKKHGTSTGIKRTKFQVSLLDGHKKEEFQLQLR NRFQPLADLGGEDTEVDTHWTKVKEVFTSTCQEVLGG KKWEHKEWISQHSLDLIRIRRQLKAEANASRTRGSRAD ALEQYRTTAKDVKKSIRKDKEEFMNRLAEKAEEAAVG GHMRILYQTTRTLAGKYGKPEVPVKDQEGKSIFGKEAQ SKRWVEHFNSLLNRPPPSNPPNILPARNELPICCDAPSKQ EILDAIKKLNKGKAAGPDQIPPEALKADANVTADILLPL	cd01650	551	808

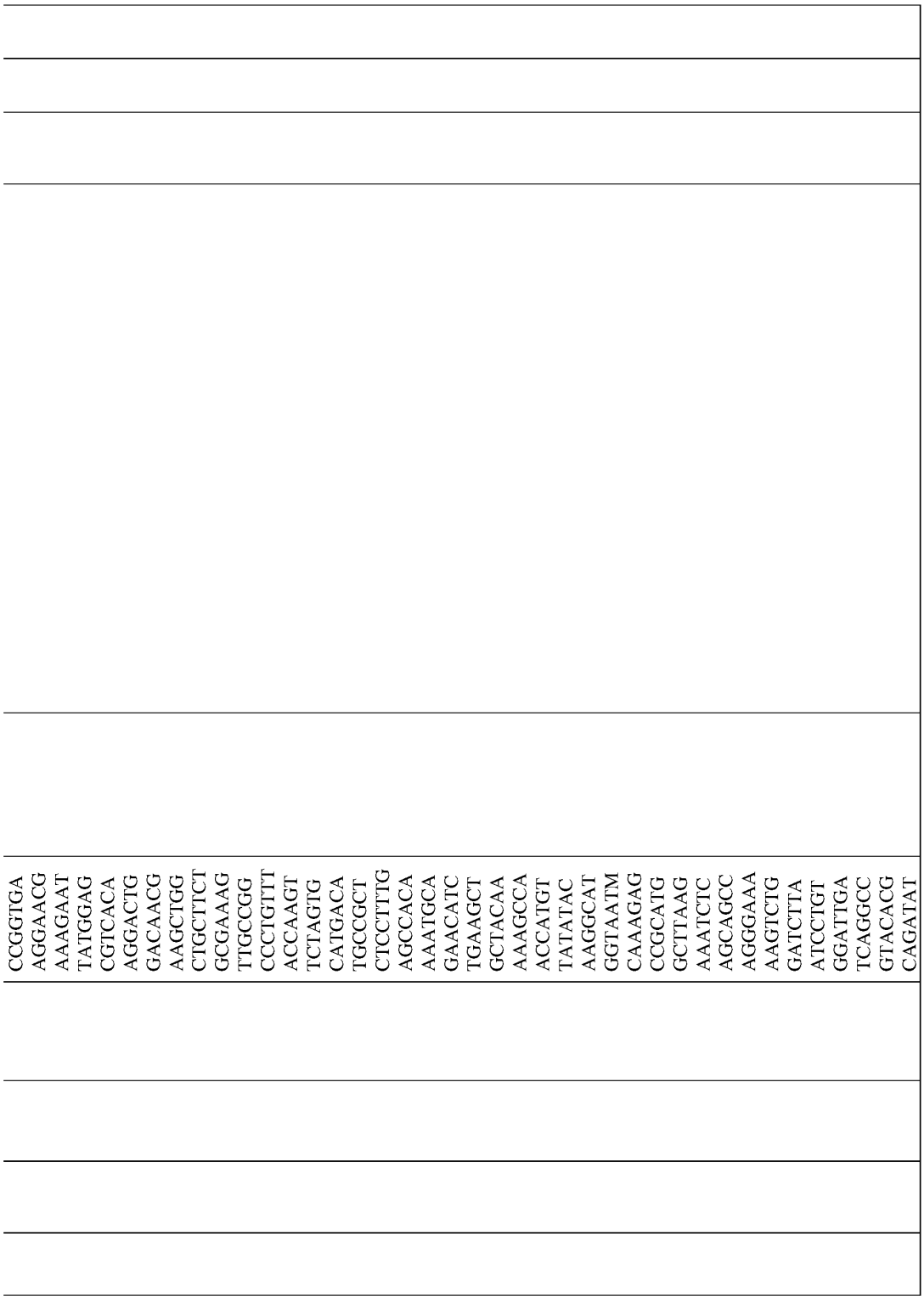
				GGTCTGGTCGCCCGAAAAGTGCCAGTTTGCGGAGGGCCAACTACCCAACCCAGTAAAAAACAGTAAGTTACAGAAACACGGACAAACACAGCTTTTTATTTTTCGATCTTGGGAGACGGAGGAGCTTCTGCATCAGCAGGAACA		FTRIWNESKFPKDWKEGHLVKLPKKGDLSNCSNYRGITLLSIPGKVFNRIVLERIKTGTDPRLRDEQAGFRKNRSTTDQIATLRIIVEQSLEWNSQLIVNFLDYEKAFDSIDRATLWKIMRHHGIPQKLVNLIKSMYEGTSCRVVHDGQLTDSFDIKTGVRQGCLLSPFLFILAIDWLMKETTRGHRNGIQWTPWTQLDDLDFADDLALLSHTQEQMQAKTSVLNSLSESVGLRIHPGKSKVLRTDAQRKEQVTIEGKALENVESFCYLGSIIDREGGTGAEIKSRIGKAQAAFISLDRIWKTREISLRTKLRIFNSNVKSVLLYGCEAWNASQTCIKRIQVFLNRCLRKLLRIRWTDRVTNVEVLRKTRQIPVRDEVGRRRWRWIGHTLRKSGSCITKKVLDWNPQGKRSRGRPRGTWRRITEIDVQKSGKTWNNIKRLAQDREKWGVFVRGLYPGAG			
RTE	RTE-3_AC	.	Aplysia californica	GTCAGGTCAGCCTTAGTGCTGCTACCATAATGACAGTAGATGGTACTACCTGCAGCTGTTCCTACTGCCTTGTAGGTGCATACGGGAGTATGAAAGGTTAAGGGAGTAAACCCTAAACAAATCCGGAGGGAGG	TAAGCTAAATTGACCGTCTGCATAGACGAAAAGAGCCATATATATA	MSPRVTFDDGSDYSVALISYRPSQTGQPPVSKVLIRHVCLLLWVYGARTKTERQTLNPAPGKKRVKLTPWKLGTWNVRTMTTGLDTTNLEKTEDIRKTAIIDRELTRLNVDITALQETRLANSGTIRERNYTFFWKGLGEDERRIHGVGFAVHNRLLNRISTPQGISERITVMEMETKTGKAHFISAYAPTLTTENEKKDQFYEELQDAINNVPKADQLFLMGDFNARVGTENKVWPTCLGPFGLGKINENGQRLLEFCTNNNLCITNTFYSGKDRHKVSWCHPRSGHWHQIDFIIVRKKDANMVKQTRSFYSADCNTDHLLVIAKAKIVTKKTPRAEKQKKPKINIRNTKDPALKAKFQEHVKEELQHLDSSSPDNMWNSMRSTIYNSASITFGIERPCKEDWVEASASTLAPLLEKKRKILIEHNTNPTESTKQQLRAAKSNLQRETRKCANQYWDKLCSEIQQASDTGNIRKMYENIRVAIGPTANKVAPLKSATGEQLTDKKKQMERWVEHYSDLYSEERHVDRQLHEAIPQLSEMSELDTTPTEEELSSAIDELACRKAAGNDNIPAEVLKENKETLLPYLYKLLTTCWLKAEIPHDMRDAKITTLYKNKGDKGDCNNYRGISLLSITGKAFARIILKRLQKLAERVLPESQCGFRSNRSTMDMIFSIRQLQEKCREQRLPLHTAFVDLTKAFDTVSRSGLYMVLGKIGCPPILLQLIISFHEDMKATIQ	cd01650	605	868

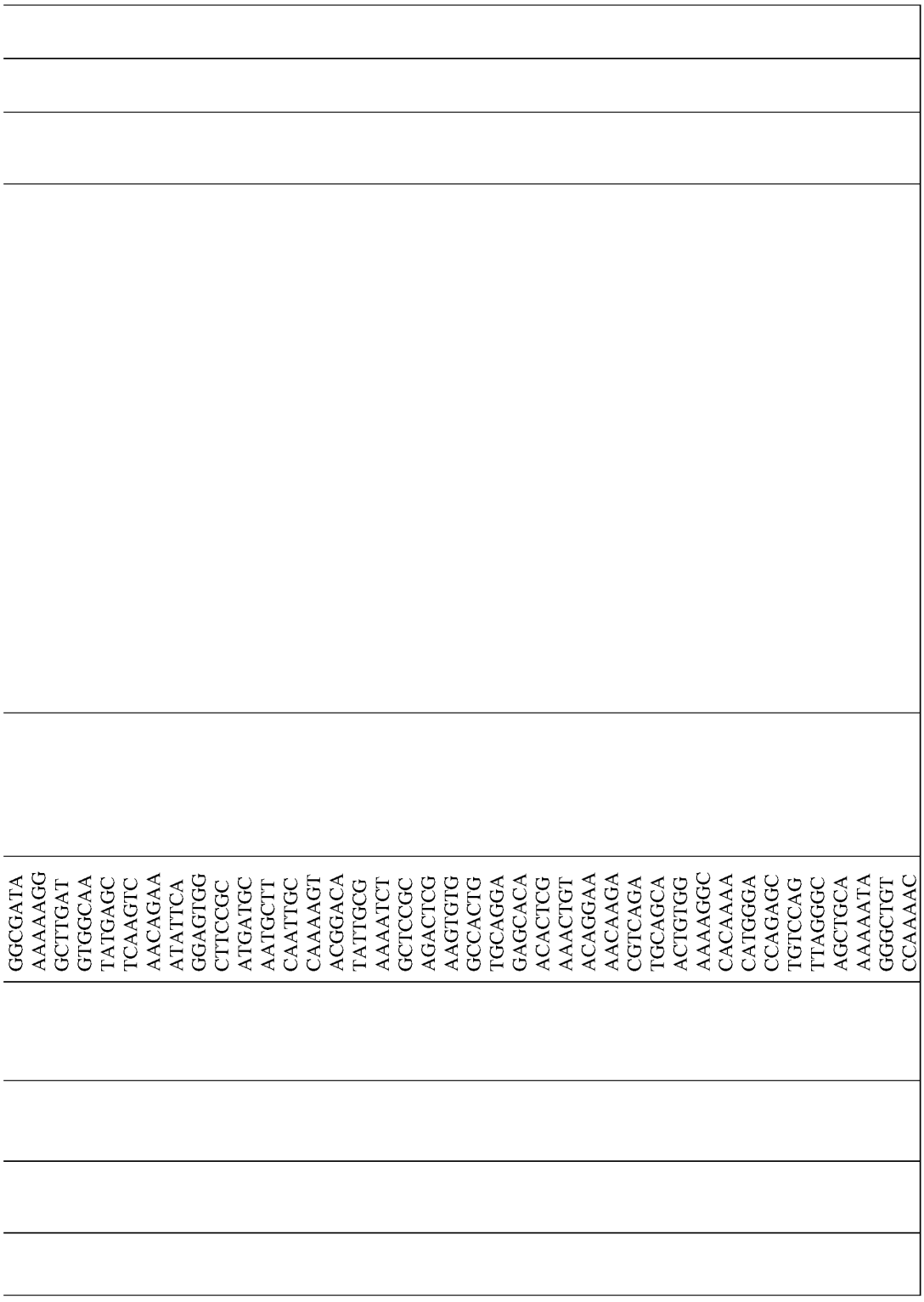
				GCCCCTA AGGCAGT CGACAGG GAGGAAC GATCCGA ACTATCT TCTGGCA GCCCCTG CAACTTT ACTGGTT CCAAATG TATTGAC CTCGTCT CTCCTTTG GACTAAA TCAGCTA AGTCGAG AGGGGGT TCGTGCA AATTGGG CAAGGCA CGTTCTC CATATAG TCAGGCC CAGGCAT TTAGCTT AGTCAAG ATTTCAA GATTTCT GGGAGGA CGCTGCA AAAGAGC AAATACA GCGGATC AGACCCA CAGCTCA ACCACAG CTACAAC AGCATTT ATATGAT		FDEGCVLAPTLFSTYFAVLLQHAFKESPGDVFLHWRTD GSLFNLARLKAKTKINKTTVRDLLFADDAALVAHSEST LQTMMDRLSTACKVFSLVISVKKTVILTQEGAEQSHIWL NDKELETVDKFCYLGSTISSTTSLDYEINIRIGKAATTFG RLTKRAWENKRLTIKTKIKIYEACVLSTLLYGSETWTLY SRQEKRLDIFHLRCCRKILGISWRDKVTNVEVLARANLQ SMTTILCLKRLRWLGHVKRMDDNRIPKQLLYGELAQG KRQRGRPKLRYKDTCKNSLSKLGVDVSIWEEIAEDXKT WRKXIKERTKMLENSLRNNQAEKRQRRKENSGNTELRS TTLTCKHCDRICASNIGRISHERSCMKTTLRPS			

				TATGGAG						
				AGGACAC						
				TTCATCCT						
				TGTTGTG						
				ACGAGCA						
				AGGGAAA						
				AAACACG						
				GGAGGTG						
				GCAGTTT						
				ACGGGTT						
				ATAAGAC						
				ATCCTTA						
				TTTGGCG						
				AAACAGG						
RTE	RTE-7_AC.		Aplysia californica	GGTGTAT	TAAAGTAA	MGDFNGRVGERRTPWTKHLGPHSDHRTPCNYNGNHVL				
Tad1	Tad1-23_BG.		Blumeria graminis	GCACCGT	TGAAGGTT	ELCAEHDLFTNTFFQHRASQIYTWYRWNNIMVSSQIDFI				
				GCAACGA	GACAGACT	LTRTRMRITTTDSRAIPNAGLDTDHRPIITTLTTKKERKPK				
				CTCCTAC	TCTCCCAGT	KHKRQQERLNMHKLGDDQVQAQIRSTLNEKLGSINSTI				
				TCAGACG	AGCAAGGA	LTVEEAWDTFKTTLLDTMKEVCGTKKTRGKHRKATAW				
				AAGAAAA		WNEEVKDAIKEKKRLYNVWVKSKNEEDYIKYRLARRH				
				GGACTCT		SKKVVITSKEKSWTQYGEKLSETCKTSPREFYKKVKTM				
				TTCTTCG		RTRDEPYVPTTVINDTNGEALHEEKEIIKRWEEYFQELL				
				AACAACT		NPSGVQAQNTQCPFTANYPEHSEPTILESEVREALKTSP				
				CTCGGAC		KNKAAGDDGITTEAILACGETGIHWLTIIFQKAWKERKV				
				TTAATTA		PADWQNAIVVPIWKKKGSKKDCNTYRGISLLSHAGKM				
				GCAACAT		YAKILERRTRAKTEHLLSDAQFGFRKGRGCTDAIFALRQ				
				CCCAGAC		LCERALEYDKDLHLVFVDQEKAFDRVNRNKLWKILEQ				
				GAGGAAG		YDIKGQLLDNIRAIYANSRSAVRTTSGTSDWFPVTSGVR				
				ACCTTTA		QGCNLSPLLFVIYMDQITKEANPDPEALNELMFADDLA				
				TGTA		MINNKTQLQEHINQLNASCEKYDMKISISKTEVMTISRR				
				ACATCCA		PGKVDININGSQLKQSREFKYLGSIFTENGKLDREIETRC				
				CATTCCG		QRANAVSYQLGPLLKHPNIPMSTKAKLINAIFLPTLTYQ				
				AAAAGTC		CQTWTLTKALERKLVTCEMKCLRRAVNKTRRDKIKNE				
				GACCCCA		VIRDTVGATPVLPHIEKQKVKWFGHLTRMPSNHPALRA				
						YNIKYSGWRARGRPRRRWSESVADTLKDHGMSLLQAT	cd01650	392	640	
						RLAADRHLCLPATPQGTSGRKK				
						MLGAHDLNLILPKGTITWKSRGSQSTLDLAFVSNALEST				
						VMACKPAFDLEASSDHIPICTQLHIEPPNREEQQPRPQW				
						KKADWEEVNRRLALKLSRLSTDHLELNSPESIDQRVSAI	cd01650	332	595	
						TQAIKETIPTARPSTFAKPYWTKECSETVKEARKARRLW				

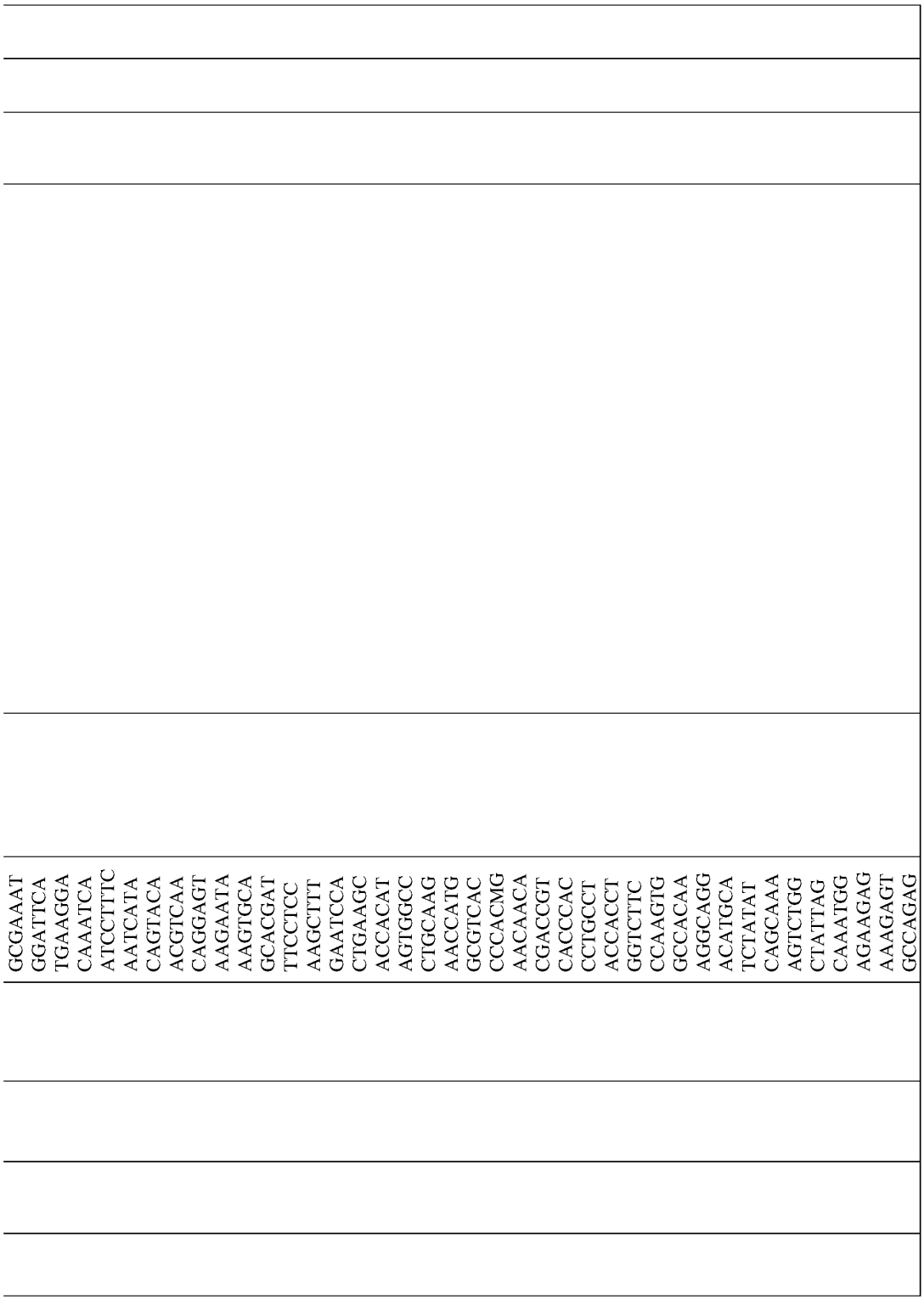
TAGATTT	AGTGGACG	TKLGSEESWVEYQKATNRKKSQIKKAKMIGWRAIVSEA				
TTATCCA	GTATCCGCA	TNDPAKLWKLAKWAKKSPEERSRLPQLPDIKDKDGNM				
TAAAGTT	CTGGACAT	FTEDTEKADILAQHFFPQPVQANLEDIDGTRYPEELTTVS				
GACAGCA	GACAGTGA	SIITASEIQDTMANLASEKAAGPDGISNWMLKNCRITLG				
AATCGTA	TAGTGCACT	KVLEKLFNACISIGYHPKGFKESVTIVLRKPQKESYDNP				
ATCGTGA	TCTTGAAGG	KAYRPIALLNTLGKVLEKLVANRLSKAAEAHHLLPDEQ				
TGACGAG	TAACACAA	MGARQLRSTISAVELITEQVHSIWGKNKRRVASLLSLDI				
TATTCCT	CATCAAA	SGAFDNVSHKRLIHNMRAKGIPRWTTQYIESFLEHRTTS				
ACGAGAG		MNLGSFKGEQFPTSTGIPQGSSLSPILFLFFVSTLLPELQS				
GAATTAA		PSSTAVGFVDDTNILTWSNSTEDNCRNLLERLHGICEKW				
TGGTGGA		ADRHGVKFAPEKYQLMHFSRARKRHNMQATVQIQGHL				
ACGCCTT		AKPQPSLRVLGIHLDPKLNWGAHVKSIQLKVEPQIQALG				
GCGCGGG		RLAQSTWGATFHKAKLVYNAVVRPAVTYGSPIWAEPG				
CTCTAGA		MAGKIPERVIKPLRSIQRRCLKIVTGAYNSTSSKVLEHET				
GGTGGTA		STLPMEIYLKQRRVQHAGLSKSLPVQDTISSACRKIKQP				
GCGTCGC		RAGRQDARAVTRIMDIADWGKIGANIXTKERQKAAGKI				
GTCAAAG		AAFKEWERSWTRLAARQPTNHHATADPKIWNAANIFT				
GCGCCAC		DDKTGRTRMDQRGTPSSLHKHLSRAQSSIAMQIRSEHIG				
AAGACAC		FNSYLHRRKVPGVDSPRCQCGYPSQNVKHMVMACTSW				
AGACGCG		AKGRGEILRKAENRSFEAMMNNPKDLARITRWIQQEGR				
TCCCTGA		LEQFRLTLEVEATMRAREEG				
AAAGATT						
CAAGGAC						
GCAACTA						
AAGGCGA						
AGAGCTC						
TCCAGTA						
GAGCGTA						
CAGGAGA						
ATGATTG						
AAGAGGC						
CATTCAG						
GTTATCG						
AAGGCCT						
TCAGAAA						
GCGAAAG						
TCACCCC						
CCCCATC						
CCCGCCC						
AGACAGG						

CCATAGT
GATGATA
ATTCTGC
AATGGCT
AAGATCC
TACAGGA
GATAAAG
GACATCA
AAGCCTC
CATAGCC
CAAAATA
GCCCACC
AATGCAA
GCCCCAG
GCCGCTC
CTGGGCA
AAAGTCG
CAAGCAA
GGGAGAG
TCGTACG
GAACTAC
TATCCGA
ATACAGG
ATGAACA
GGAAAAG
AGAGAAT
AGCCAAG
CTCTCGA
GCGAAGA
GTTAGTC
AAAAAAA
TCGGAGT
ACAGCAA
ATAGTAG
GTGCTAG
ACAAATG
ACAAATG
GTCAAGT
GAAGGTC
TACTTCG


CCGGTGA
AGGAACG
AAAGAAT
TATGGAG
CGTCACA
AGGACTG
GACAACG
AAGCTGG
CTGCTTCT
GCGAAAG
TTGCCGG
CCCTGTTT
ACCAAGT
TCTAGTG
CATGACA
TGCCGCT
CTCCTTTG
AGCCACA
AAATGCA
GAACATC
TGAAGCT
GCTACAA
AAAGCCA
ACCATGT
TATATAC
AAGGCAT
GGTAATM
CAAAGAG
CCGCATG
GCTTAAG
AAATCTC
AGCAGCC
AGGGAAA
AAGTCTG
GATCTTA
ATCCTGT
GGATTGA
TCAGGCC
GTACACG
CAGATAT


GGCGATA
AAAAAGG
GCTTGAT
GTGGCAA
TATGAGC
TCAAGTC
AACAGAA
ATATTCA
GGAGTGG
CTTCCGC
ATGATGC
AATGCTT
CAATTGC
CAAAAGT
ACGGACA
TATTGCG
AAAATCT
GCTCCGC
AGACTCG
AAGTGTG
GCCACTG
TGCAGGA
GAGCACA
ACACTCG
AAACTGT
ACAGGAA
AACAAGA
CGTCAGA
TGCAGCA
ACTGTGG
AAAAGGC
CACAAAA
CATGGGA
CCAGAGC
TGTCCAG
TTAGGGC
AGCTGCA
AAAAATA
GGGCTGT
CCAAAAC

				AGGACCC AAGACCC AGGAATG TTCTTGA CCGACGA AAAATCG ACAAGAA ATCAAGA TGGCGAA TGGCAGA TTGCAGG AAGCAGG AAAAGAA GAGCAGT AATGACT GCACCAC TTGTTGG GGCCGAT GGTAGTA TAGATGG TCGTAGG CGTCCAG GCAGACC ACCAAAG ATCCCAT CAGCTGT CCCCACA GTCGCAG GGCTCTT CGCAAAT CTCAGCG AGGTGCC ATCAAAA ATGACAC CCCGGCC TGTGACA CGGGGAA ATCAAGC AGATGAG ACTGAGA					


GCGAAAT
GGATTCA
TGAAGGA
CAAATCA
ATCCTTTC
AATCATA
CAGTACA
ACGTCAA
CAGGAGT
AAGAATA
AAGTGCA
GCACGAT
TTCCTCC
AAGCTTT
GAATCCA
CTGAAGC
ACCACAT
AGTGGCC
CTGCAAG
AACCATG
GCGTCAC
CCCACMG
AACAACA
CGACCGT
CACCCAC
CCTGCCT
ACCACCT
GGTCTTC
CCAAGTG
GCCACAA
AGGCAGG
ACATGCA
TCTATAT
CAGCAAA
AGTCTGG
CTATTAG
CAAATGG
AGAAGAG
AAAGAGT
GCCAGAG

				GAATCGC AAGGAGA CATCACA TCKGTCA GCCTCGA TACCGAC CTTGGTA AAATTTG GATACAT AGCGTGT ACAATCC ACCCCCA TTACACC ACAGCAG CCTAGAA GAAGGAA CACTCAA GTGGCTC CCGCAAA TCCTATCT CAAGAGG GGGAACA CATTCTA GTCGGAG ACTTTAA TGTGCAC CATCCGC ATTGGGG GGGACAG TCAGTAC TTAGGCA CCATAAG ATCGCAG ACAGCTT CATCGAA					
CR1	CR1-7_CBre	.	Caenorhabditis brenneri		TGATTTCTC TTGTCTTTC CCTTTCTTC ACTTGTCGC TTTTGTTTG	LLQKLINFGIPSSFVSWFAVFLADRTFSVKINDSIGDTSPT IPSGVPQGSVSGPLLFLIFINDLLLSLATIPLLHIAAFADDI KJFSNDPFVVQTGIDTVVGWAKSNSLPLAHSKTALLHLG SLNPFFPYTIAGNPIIPTSSVRDLGLLTEPNLKFSAHINRT VALALLRSKQLLKSFKSSSPQFYVYLFKCYVLPLIEYCS	SSF56672	45	118

RTE	RTE-1	AF054983	Caenorhabditis elegans	CTCTTTCT	ACCTCCCAT	VVYCPPSTSKLSLKLEYPLRFFTKKVFHRCNISYTSYPDR				
				GTACTAA	TCGCTCCAT	LAHLDLFSLRHRRLKAQLQLLYNFLSGSSYFPNLDSHVR				
				GCCTTTTC	TTTAGGCAT	FSSSPRRPMNLVNVNPRVKDFFFTTVPIWNAITCHCPTFL				
				GTGGCGA	ATCCCCCGT	RPSEFNSLLQCSIQRF				
				AAGGGAT	TTCTCTTTC					
				CAACAGA	TCCCCCGGT		IGLAGHYRVSRDTTFALNSNPSCLSPPHPASAKGKTVPG			
				TAACCAC	GAGGCCTC		ARRKRRSNLGVQTICQSDLEIKKPRDKTRRTDYRVCTFN			
				TTCAACA	GTGAGGGG		CRSLSSDDRLAELLEETRRIQFDVIGLCETKRAAEAHLIH			
				AACTCGA	CACTTGCCA		HDGTGVFLGKRNESSVSGGVGFIVRSTLLPKIVEVRFLS			
				TGATTTC	GGTTCTTCT		HRIGLLTFKVSRKFNCTVIQVYAPTADSDLEEICDFYDSV			
				GTCACAT	GATGTATCA		EDVFRECRSKYKLVIGDFNARMGCRQNNERYIGPHAME			
				CGAGAGT	TATTGATTT		PRNDTGELLATFCETNRLWHTNSMFKKPMHKRWTFVS			
				GGCTTGG	TTCCCTCTT		PDGNHRHEIDHILANGKFVTDTTVLPSFTNGSDHRLLRC			
				TTTGGAA	TTGGAAAT		NLHFNNCLAKLLEQVRRRKPPKRVLDPAIAYAISATVTV			
				CCAAGTG	AAATAAAT		QSDPDLDTDYDNLIQSLKELQDQAIVRPANHSSNRLSEE			
				CAAATTA	AAATAAA		TRKLLNKRRFMDRNDPQFKSISDKCREAVQKDHEAFAS			
				TCTGGCC			TRLLSAANQKKSLKRVARDINEYKSVIPCLKSTSTGERIT			
				AAGAAGT			SRVKMEQEIEKFYTELFKSAVSNSQTSSIPATATPPPFLPE			
				ACACCGT			EIRHVLRSFPNGKAAGQDKISADFLKSCHDNVIDLITDRF			
				CTCACGC	TGAAGAAC		NRYLHSRNVPKPWKTSKTTLIFKKGDRENLENYRPICLL			
				AGGCCAG	GGATCGTCT		PVLYKVFTKCLLNRMRRSLDEAQPVEQAGFRRSFSTIDH			
				AGGTAGT	AAGTATCTA		IHSLQRLLEVGREYQIPLTLVFIDFKKAFDSVEHQAIWKS			
				GTCGAAT	AGTATCTAA		LDEQGADGAYIDLLKECYKNCTTNFTPFHRPVAVPVTK			
				ACTCTCT	G		GVRQGDPISPNLFSACLEHVFRKLSWIELKGEAEDYDTIP			
							GMRVNGRNLTNLRFADDIVLIANHPNIASKMLQELVQK			
							CSEVGLEINTGKTKVLRNRFADPSKVYFGSPSPTTQLDD			
							VDEYIYLGRQINAQNNLMPEIHRRRRAAWAAFNGIKNT			
							TDSITDKKIRANLFDSIVLPALTYGSEAWTFTKALSERVR	cd01650	562	831
							ITHASLERRLVGITLTQQRERDLHREDIRTMSLVRDPLNF			

				GTCCGGG CGGCTCT AG			VKKRKLGWAGHVARRKDGRWTTLMTEWRPYGWKRP VGRPPMRWTDSLRKEITTRDADGEVITPWSTIAKDRKQ WLAVIRRNTTNS			
							NMKQPAKRTVMTCLDQRAAFDKVPIKKLCVDLHESLN EAKISKWLTNYLSKRIIKANFGNHSSKWCTLLGGVPQGS VLSPNLFSFYMRNMPIHPKTEITVYADDTTILAQDTNPF KAGELIQEHLDLLYDFFNERKMSISAEKSTTTLFTCDTK EFKTKIPVNWNGAEIPTTNQVRLLGVHLNTMCGSKDHV AETRKKMMVSTNILKALSGTKWGCSKELLISTAKALVK PIAIYGAPAWSQLLSETNLNKLEAAYRNTLRTCCGLTSD TPIEHIYAEAQMMPLKEEFELAKQQTYIATTKSGTHPNR NLESRRNTERRSNRPREPPLNLQEATKTHLSRLKGTKAE AQKANHTRFVREFLNSAPPNRILGATPPPLNSSEESSLPK ATRTELARLRCGHSLLIPRYKARIEGTPIGTCGCGDNEG GIDHYLSCGRSHPISKEKLWTSPMEAARALGLAYKPYD PGGGATTSV	SSF5 6672	8	150
Ving i	Vingi - 2_CE	.	Caenorha bditis elegans	AGTGACG GGTCTCA GTAATCA ATACTCT GTCTTCC GTACGGA TCGTGCT AAAGGTA AACGAGG TGGAGGT GTTTGTA ATGTGGT AAAAGTA GAAATCC CATGTGA AACAGTA GTTTCAG TTGAAAA CGCTTAA CGCTTAT GAAATTC	TGATCTAAT GTTACCTTA TCCTCACTT CTATCCATC ATTTATTCT TGCCTTTTA TACATCGAT ATTTGTGTT TTTTGCTTC TTCTCCAAT GTAGCCATT ATTTTCTTC TCATCTCTT TTTTATTTT TGTTCCGTT TTATTTATG TAATTTATT TATGTAACC ACTTGTGAT GGTTTGTAT CTATTTATG TGTGTTTTT TTTCTACCT CACGATATC	MPTRKSKILDVALSNSVISNLNIMAPIGSSDHNTIKFTVQ LLQNISVPKKMIPDFSKLDWPCCATYFENIEWGKVFDTH SHVNDIYECFLGHIRTMFSQAVPTMPMPETSKPVPQYLR SFEQMVKNQFSKTIMSKTMNDFIEYIYLARKYRKKLRK YLSNLEVKLFKKKGNSKFSKHAKTLLKPRNCSVPALIDN DENEIVSELDKANLLASYFEAQYTNQNNTNFALDDDFV SDQPIPWVTDAEVLKVIMASKKKSCSVTSDQVPFTFIKILS PLISSVLAQIFNLSMMRGQVPHAWTESIVIPLDKITNANN PTDFRPISLTSHLCRVYERCILAKIELFLIDNKFWSQNQH GFRPKKSTVTNMLESLNDWSKSIDLGEQIDIIYLDFSKAF DRLPHDLLLLDKLVSLRMNKNLLIWLNEFLSNRSFRVRIG QTLSSRKLATCGVPQGAVLSPLLFGIYVNDISSILPEKVA CKQFADDTKLYVSTPKTESENNLQSALDAVVDWTRGS KLTLNQSETVHVTVGKRRTDFKYHLDGYPIERKAITRDL GFLISEKLDFSEHWRKCINLAKFQLANMFNQYSTSNKKL MILLYKTFIRPRLEYGTVVSSPTKKSDEKAIESVQNAFTR RLYSRIKKRYINPTDKDYKTAMERNELFDLQTLSTRRKII DRIQVIRMNNGKVDLKTSEFFEQQETHTRSKKKYIWST GKSKLRRHFFVNRTLATMKQH	cd016 50	299	552

				ACGTAGT TCCCTTTA GAATCAT AAACATA TATCGTG CTCCTTCT TGCACTT ATCAAAA TACTGTT AGATTAA TTAGTAA AATTGAA GACCTTG TGGAACA TCAAGGC AAAGTTG TAGTTGT GGGAGAT CTTAATA TTGGAGG TATCGAC TGGGGCC TCAACGA CTCTAAA ACTAAAT TAGGTTC TTTATTA GTAGATC TGTGCAA AAACTAT AATCTGG AACAATA TGTCAAA	TTACTATCG TAATAAATT AAATTAAA TTAAATTAA AACTT				
CR1	CR1- 15_C Jap	.	Caenorha bditis japonica		TAACTTAAT GTAAAAAT TGACTTGTA TTAGTGGCT TTGCCCTTC AATCAATA AATAAAA	STKLDLXXXEDJXFNEXHLNKSSPKAFFNFINSRXKXXQ EIXXXRDXXXXQVTSDYMKAELLSNCFSNAFTEDKNTN TDFPYRTNHFINSLCFEPFIVESILSKLTPKCNTTPDGIPAI FLKRTCTSIALPLSIIFRNSYNQXVLPSFWKTALVKPLHK KGSRSDPNNYRPISLTSAVGKVMEKMVRKSIILFLDDFK LLSNSQYGFRSKMGTESQLLSYQNKLVTNYLSKQTTFS VYIDFRKAFDTVGTKKLLIKLERYGIRNHLLRWLENFLS	cd016 50	151	401

RTE	RTE-1_CJap		Caenorhabditis japonica	TGCTTGC TGGATAC CTCCTCG CGGTGCA AACTGGG CAATGCT GTGCTCC TACGACA GGGGAAA TTGGACA CCACTGT TAGGCAC AGTAGCC AAAATCC ACTCTTT GCCACGG TAGTGTC CGATACT CGTGAAG AAGGGGC TTCACGC CGGCCTA GTAAGCC TGCTGAA GTACAAA CCTAAAC TTACTAA AAAGCTG AATCAAA AGAATGT AAAGGAC	TGAGGTCG AACCTCAC AAGTATCA CAAGTAAG TAAGTAA	NRTQRVIINGVMSEEKSVLSGVPQGSVLGPLLFLLFVNDI GDNFNSDYLLYADDLKIFSTDKSSIQRDLDSLSTWCEN WQMSVAPNKCETIAFYHSKRGMDKDTKFTIGGATVPIT SRIRDLGMFFTSDLSFSNHIDTVLRKAHQRVNIFFNVLRH AGFDTFIKCFIIYVRPLIEYGSIIFSPTLKDLVKSVESVQKS FLYRCFKKFRIEYISYFDSLSYCKLESLEHRRLIIDLTFLY KLIVSKEIFIHNFTCTTRSPLPSLRRHPYIVRSTLRNSSKLS HQFFSNRIITCWNSLPSYLFPKLICSETFKSALKCLDLSKF LTLNMHNY	MFGRSTTDTEITEASYGPKNRRKGNDFHIGSLNVRSISD QSKAEAFDLLAVNSGCKVIALQETKRKEGEWVLSSGAE LICTRREAREGGLAFWIHKDWLDSLVETKRESDRISTAS FDFHGSHLHVINVYAPTSAANETQSDKFYKDLKKVYTR LSKWDSSIIICGDFNAVVGCSSDDLPFVGGHGTGQRSDN GDRLVNFAFSSRVSILNTIYPKRPNRKWTWISPNGKTKN EIDFILTNRREIAVDCEVYCDYLNCSDHRLLRARFRSTA KKEKSRKFAKTKRAGGERSWDPVRFKNTITEMLTENPP TCYEELTESIRSAASTATHATHQLSRISEATKKMMLSRF QMIHAGQTKNADLAKINNEVRRSLENDLDKWREVKSL EAAVKGRSLKKTKRALTLKSSPFTNLERPDGSTTYNRPE IRKLVLTHFSNLYAATTREPRSRYNHPDEPEILQHEVEH AIVTSKTNISPGPDQITMLQLKLGIESIAPHLTAALNQVL TNGRTPEDWKTVKISLLPKTTKPKKVKDYRPVALSSIVS KLFTKILTRRITAKSEDYLEESQAGFRRGRGCADNIQVV AQMWEKCTEFKIPLVAVFLDFTCAFDNVNWTKISEVLN NLQIGRNVIRALNNSNSSAIGELNVLNKKMRFKIKRGVR QGDSSSPLLFALALQAILDELDPAPCEDDKTGIDINGESI HRLEFADDVVLFANSVKEAEDRANRIAMGGPKYGLNIN PSKTQILMNKYVVTRSDIDILNTRIEISQKVKYLGRTFADD GSLTEEVARRIRAGWAAFNAIRRELLQSPQKIRIRLLTTT VIPAMTYGCETWTSKEADTKRLQRAVSNMFEIAGCDPP DMEKQVLRRKLKWAGHVSRMKYDRWAKIASEWDPRG SARPRGRPPKRWADDITDSIKIFIHNEALRGNLGHGRRRI GILEARRAWSTVGRDRVTWRSLVRSIQDGN	cd01650	531	773

			CCGTCCA GGAATTC ACCAGCT CGGCCAC GGTGCTC GCCGAGT TGTCTTG GTCGGAA GAAA	TGACCACTG ATGACCGA ACGTTTACG CMAAATCT ATCCTCTCC TCTCTCCCA AAATKTKTT WWTTTCAC GTGTACTCT AACCACAT ATCCCTGCA AATTCCTCT ATTCTTGCC ACTCCGTTT TCTTTGAWT CCTCTTSTA CACCTCTTC TGTATTTAT TACTTTTCA CGTACCGC AATGGGCG TATACAAAT AAATAAAT AAATAAAT AAA	MPTNHSQRTLMPRRCKYERTRQRIFSRIQKTNFAQHRXF NFFXHSPDLILSILSANXRSLMANKLEFQYIVNTTSPDILC IQETWLSSKFGNASFLDPNSNYHVFRADRCATSRGGGV CTLVKKEFPXTIITAETKNSAYELITVDIDVXVTPLRIINI YRPPSCSFENTITLLNKVSDLSNHVGQJVILGDMNNGIN WNSAVPTSQTRLGSLLCDFCNQNNLTQYIRKATRKENL LDVVFSNSRLSDVRILSPIGSSDHNSINFKLDILQRPPEKK ILFYDYKNIKWPKCLDYLGSIDWGRLLHPGNSISTMYQN FIDHMREAFELSVPKIYVEPTIRNTPDYLKSFYKLVQLKL ERVKASRGRXDFLDYIKTAKKFRKKLLKFRRNEEVKLL ESSGVSRFSKKAKYLLKPKSLRIPTLTNPLGVLSVTNAE KAEILADSFESQYVNRPLLNFELPSNFKSPQEIPWVTDAE MLSMMLKCKNSCSQTPDNVPYKFLKLIAHLISSPLAQLC NLSMMRXECPSLWKESIVIPLNKKEKPMAPTDFRPISLTS QLCRIYERCLLKKILQHLDTIKFWNNDQHGFRPKKSTVT CMLEAVNDWTESIDEGAQVDVIYLDYAKAFDRVPHDIL LGKLVDANLNPNLIKWIRSYLSNRXFSVKVNNESSTPKH AQCGVPQGAVLSPILFGIFVNEIPSILPSGIRCKQFADDIKI YTAIPKSDTIEDHETLMQSAIDAVCNWSENSKLALNDSK TLVISLGKNPRTTKYQINNNPIRRETSIRDLGFITPMLDF TGHWKKAICKAKFSVFQIFNHYCSNNPRLMTLLYKIFIR PILEYGTEISNPTRKLDCKAVESVQNAFTRRLYSRIKGRY LTPNDPDYLTAPQRNSLFNLSSLESRRQTSDRKLILKMIS GLVDIDTSPFFTYEKNTRTRAKNKFKWNHCKTKIRRNFF VNRTLSSFKL	cd016 50	528	785
				TAAATTAAT TCATTTCAA TTGTAGGAT TGAAACAT ATTTAATCT TTTTTTTTC	LINILNELLTELCPRIPERPXSKVCPLLRKLRRRLRRYKSSS LSHADIANRLKRYRALASRYRKWVLKHETKLINESNCT SIRKFVSKRLKSKQSIPALLDNGKAVTSDFVKCDLFSXSF MKNFACPPEEYSTPTPSHYGSDSLSSFSSPSSSRCLFLPYI VEKVLQHLPPKCGFSPHLANYLVLKKCATPLALPLSIIFS QSYCESEVPAAWKLAVVIPVHKKGNPACTDNYRPISLT	cd016 50	215	460
CR1	CR1-6_CJap	Caenorhabditis japonica	CAAGAAC T						
CR1	CR1-12_CJap	Caenorhabditis japonica	TT						

CR1	CR1-16_CJap	.	Caenorhabditis japonica	TGCCTCT CAGACTT TAGAACT TTTCAAA TCCCTTA CCGCACT TCTAG	TAAATTTTG TGCTTGTGA TATCATCCC ACCCGTCTT TCTCTCATC TATTTGACC GCGGATTTT TCCAATTGG	TATTTCTAT TTGTTTTTC TTCATACTT ATTGTACAT TTTTATTGA ATTTTTCCC CCCTTCATT TACAAAATT GTACTGTTT TATCTCTTT CTCTTTCAC TATAGCGC AATTGTAAT CTAGCTTAA ATTATCTCC AACATGAC TTTGTTTGA TTATTATCG AAGGATCC TCTTGCGGG GTGATGTCG AGTAATTTT ATTTATTCA TTGTGCCAT TCCCACCTC TTTCTCTGA TCTCGAAAT GTTTTTTTT TCTCAACCC GTGAATAA ATAAATAA ATAAATA	IFXTIIILLYWAISIFPQSTGDLLPDSPEFLVDXMTSFHLSQ LVTFPTRTSRLGKKNFLDLLIFSSNTSIIKNLKAGSPFILSD HCSVLFSLSLATSTIPKPPTEDPQHLHFRKCDFNKLNCHL FSYNWPKQLSYFATADSKLYHFLKIFNELIEQYTPISTKR TVHSKITGKKLYRKLRKRHPNSLSLSLKKRVKSRLKNIR KSFESQEKKVIASKNPRSLYSFIKSRIKTKSTLTHFEVDK KVIQKPSDMADLLVNIFSKNFNTPPPPFPALPSPKPLNTSI DTSPFGIYQVIAKIKPKIGYSKDNINNFIIKNCADSISIPLS	cd01650	343	590

					TCATCTTACTTATTTATTTTCTCTCCCCTTATTTTTATACCGTTGTCCTCTTTTCCATGTGCCATATACATATTTTTATGCGAGAAATGAATAAACATATTATTATTATTA	LIFGEFLLSRTFPDCWKISTVIPVFKKGNIHDPLNYRPIALTHPLARLFEKVVIGPIRRTFSPKLSKKQFGFLNNRSCTLAHLHASSIFHSVLSYPKSYLDVILIDFRKAFDSVPHELLLCKLSNFGFDTKLCEWFREFLTGRSSIVKIDNNYSSIKFKNSSGVLQGTVTGPFLFLVYINDLLESFPDDVHVVAFADDIKIFGNNRSSLQKSIKIIEEWCSIWQLKLAENKTQVLHLGVHNPRVEYHIGGITIDKCTSVRDLGIWVDEKLQFRKHINIIVNSCMYKVRQILNSFRSSNLNFYFKLFNTYIRPVLEYGCEVFNPCSVSMSTSLEKPLRFFSRAVFKRCRVNYLSYADRLLQANQQTLKHRRVLLILKTYHNIISGWYHFPNISQYIKLAHTPRFPYLLQPRGKWTPNNFLHKYLSLWNKIAKKFPKTVSGATFHLHISALPLDSYL			
Vingi	Vingi-1_CJ	.	Caenorhabditis japonica	TA	TGACCGAAGATCGTCAAGGAATGAGTAGAGTTGCACAACAACAACAACAA	RPAPPPPSQHTTKCRAKPAKTRILSWNVNGWSKRRSELATIIAKEKPDIVCLQETKLDKTNLGYTPTGYKVALASRNKFGGGVCTMVRDDWIFTISKTTSKKTAEILSITIDKLMVTNVYFPPRSSDSPTAEVREDVMEILDMASEKHIITGDWNAHAKTWHSHKTDERGAMIEDELDTRGVHVVLNEDSFTRAASGVLSSPDITISTSDLAACSSWRTCGVLNTDHLPVIIDTMAEIPKLVRPXRIMANFKKANWTDFSDSIETKVACYAGPRDAYSLENFLSATITRAASKHIPHGKFTPTREHTGDTGEIAELTRERERIQKRNPLDSNLKALNDKIEAARRELRRLXWHQKIEDMEDKRKIDCNALWRVIKGLKPKTKSTGSIRLPNGTTTTNPKTIANAMASSLVAVGKGRDCREERIAWRRTKRRLKKLRPQLPASSISHAEVKLAILKGKPSKALGPDGICHLHLKHVPDNCINLMAELFNASIDENKVPDSWKKANVFMLPKPGKDPSEIKSFRPVSLLSPVAKVLERIILERIRREIESPKDQHAFKKNHSTTTAITEATSCIVGGLNMEKPACRTIMTCLDMKAAFDTISIAKLLCGLEDAIADKKVCGWLGNYLHKRRINVQYDDVKSSWLTLRGGVPQGAVLSPNLFNFFIRDVPAIPDTKIITYADDMTLLVQDPLLSKAAEKSQDALDSLQKWFKDKHLDISAGKSTVTIFTADTKEFKYDPGLKWEGSVIPMQNRTRLLGVELDTMLRMTGHVEQVSARLAKANNILRALAGAKWGSSKETMLRTFKAIIKPLATYAGPAWHQLMSETQKTKLERQYVGGLKACCGLTKDTPRQLVYHETRMLPLSEELTLGSEQLAIAAIKTPGHPTSDLSRRRPKERTCGNRPRTPPLGCLEDRGDEWRRMRGDTRAVQKRNHTSFVKSYLSNHGIHPILGRQPPALSEEESTLPRNTRVELARLRAERSLLLEKYKAKVENRPVVCCIKCNDDV	cd01650	502	758

						GDLKHFLKCYPVKPLPMSKLWKDPVAAATALGLAVTP FDPGGDADS			
Ving i	Vingi-2_CJap	.	Caenorhabditis japonica	CACTAAA ATAAAGG ACCTGAA CGTGTAC ATTAAAA CGAAACA AACACAC ACTTGCA CGAAATG CGACCAA GTCACTA ACCGCCA CCGCCAC ACCWCTA GCAAGAC CAAAGAC CAGCAAW GGGACG WGCAGAC TCAGGCC AATCCCA CCACGWC GCGCTCA ATACACA ATCAAAC GTGTGAA GATTCTG ACATGGA ACGTAAA TGGCTGG GCCAGAA GAAACAG TGWGCTC GGTGCTA TCATTAG CAAGGAG AAGCCGG ACGTTKT	GAA	MPPGNEDRYRTRGIHSNWLQDRYGPEEQARRRSVXNG QARLGLHDHSIHDEESGLLITNVYFPPRSSDSPXAEVREE VLQILDMEDRKHILTGDWNAHSSAWHSHKTDERGSMIE DELDSRGNHVILNENTFTRAATGILSSPDITMATADIASR ATWHTCGVVNTDHLPIIAMDSEIPKREKPKRVIANYRK ADWKGFKTDVETRIRSYVGPRDAYSLEKLISKAILEAAG KHIPSGRFVPTRECQGDTKEMDDLLKKRDEIQKLNPLD GKIAEINDVIEKSRRELRKERWLNKVEDMEIKRKTDSNC LWSLLKGLKPKARSSGPVNLPNGTTTTNAKAIANAMAK NLVTVGEGRENKDEKMAWRRAKRNLVKSKTTLATTIT HAEVKLAITNGKSSRAMGPDGLCHIHLKNLPENGIHAL AELFNATISENKIPDSWKKANVFMLPKPGKDLTEIKSYR PVSLLSPVAKVLERVILERIKPEIQSPTSQHAFKKAHSTTT AIIEVTNRIVGGLNMPKPANRTIMTCLDLKAAFDTVSIA RLLCELDGAIQDKAITRWLGNYLYKRKIRTQFGEEKSK WRTLRGGVPQGAVLSPALFNFYIREVPTIPGAEIIGYADD TTILVQDPILQKAADTSQNALDXMKTWFDERHLDISAG KSSVTIFTADTKEFNFDPGLTWDGTPIPMQNRTRLLGVE FDTMMKMNGQVAKVSERLAKSNNILRALSGASWGSSK ETMLKTFKAIIKPLATYAGPAWHQLMSDTQMAKLERQ YIGGLKACCGLTKDTPKELVYYETKMMPLKEELELGSE QFAISAMRTPGHPAKDLHTRRPA	cd01650	442	698

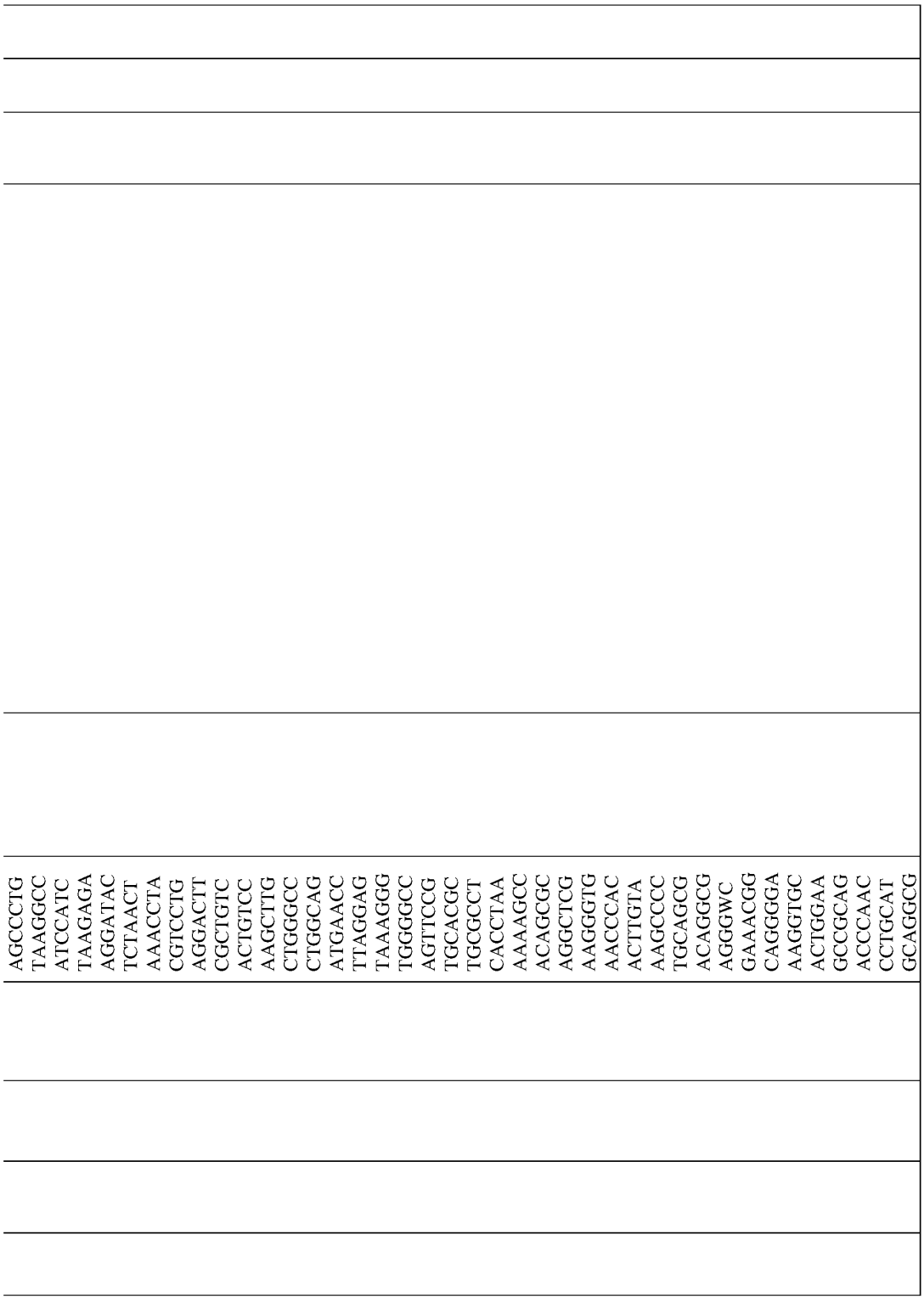
RTE	RTE-2_CJap	.	Caenorhabditis japonica	CTTCTAC ACGTAAT GAGCCTC TCCGTGG CGTAGAA GATCAAC ACCTGGA TCTTTTCT TCAACTC CGGCGGA TTTCGTC ACCCCGG CAAAGAT TCAAGTG GCTAATC ATTACAC ACTGCCT CGCTCAG GTCGGAT GTAGTGT CGAATAC TTCCTCCC TGCGCAC ATTCCTC GGTTTAG CAGGCCG TTGAAGA GTAGGCA TACCCGC TACGAAC AGCACCA ACAACTG CCTCAGT CCTCCGC ATCCGGC CTCTACT CTTGTTA CGGCAGA TTCAAGA GTATGTC	TGATCGAA GAACGGAT CGACTAAG TAACTAAGT AAGTAAGT AA	MMHRDGTGVFLGKRNTTSVSGGVGFIVNSTLLPKIQEIK FLNHRIAYLTFQVDRRLSCTVFQVYAPVADSDPEELADF YESLEEAYTACRSRYKLVVGDFNARIGPRQHTERFIGPH SMEPRNESGELLATLCESNRLWHMNSQFGKPSHRRWT HISPNKLHKHEIDHILANGKFVTDVSVVPSITTGSDHRLL RGNLHFNTKIARFEQVKRRKPPKRVLDPAVAHAITATA TITPSADIDLDYINLIAMLKELQDQAVTQPSNHSTNRLSG TTRQLLTKRRFMDRXDPNFKSLSNECREAVKKDHEEFA KDRLLQAAQNKKSMKKVARDIQEYKTFIPCLKSSTSGT RITSRTEMEKEIQQFYSKLFRSNRGPPVIASPIGMEIPPPFL PEEVRAAFRSFPNGKAAGDDKITADFLKSCHDNVILXLT DRFTKYLQSGKVPEKWKTSNTTLIFKKGDKEDLENYRPI CLLPVLYKAFTKCVLNRIRTTLEEAQPVEQAGFRRSFSTI DHIHSIQRILEVGREYQQPLTLVFIDFHKAFDSVEPQAIW ESLKSQGVQPAYIALLQQCYTNCTTTITPFHKKIEIPITRG VRQGDPISPNLFSACLESVFNKMSWQHLRGDEDDENDK ANLLPGIRINGQNLTHLRFADDIVLVASNPKAASLMIQE LVEMCGKVGLRMNTNKTKVMRNKFASQSPVNVTLNN VTTCIEEVNEYVYLGRLLNHNNELEPELHRRRRAAWAA FNNIKNTTDALSCPRIRAQLFDSIVLPALTYGSEVWTFTQ ALSERVRVTHAALERRLVGISLSEQRQRNLHREDIRAQS EVRDPLLVIKKKKLGWAGHIMRRNDGRWTRLVQEWYP IGEKRPVGRPRMRWSDSLKEQISLFDGNHLKTHWSTIAK DRMAWKAVIRDHIR	cd01650	447	720

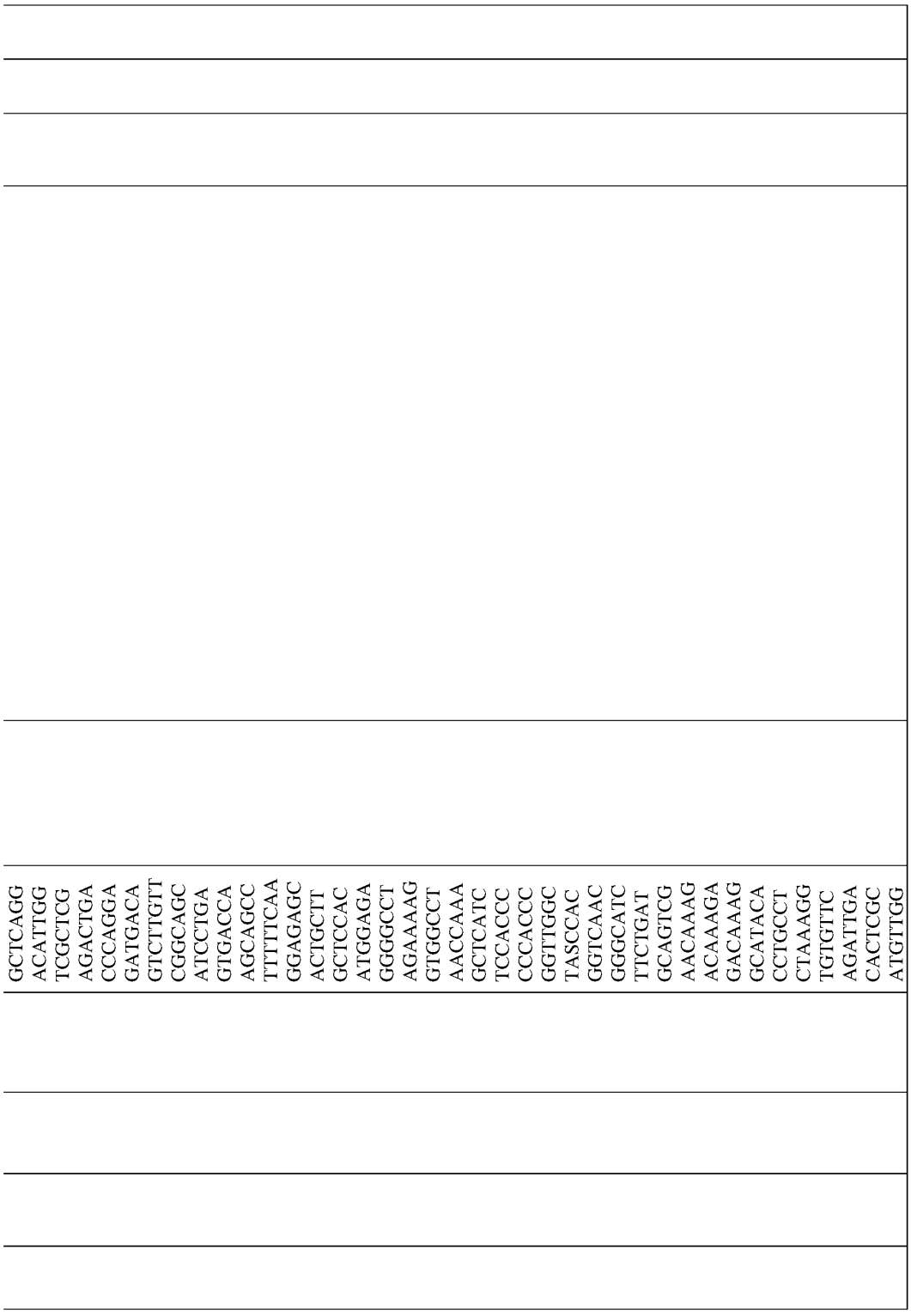
					GGAAGAG GCGATCT AATCACG GAGTATC GAACGTC TGCCACA CTGGACG GAACAAG AACACCC GGCCAAA AACAAGA AGAATCG ACTACAG AGTGTGT ACGTTCA ACTGCAG ATCGGTT GCCACCG ATGACCG ACTTGCA GAGCTGC TAGAGGA GACAAAA ACGATCA AGTACGA CATTATC GGACTGT CGGAAAC GAAGAGA GCAGCTG AAACACA C		TAACTACTG CTGAACGG ACGTCACC AAGATCCTT TATCACTTC TACCCCACC TTTTTTTTT CTGCTCTCA	VLVNRHEIQFLIDRAKPDILCLQETWLSEKTPSPTLVNPP SRYQIFRTDRAHEKRGGGVCTIVKSEIPAKVITSETKNM AYEILALDINISTTPIRLINVYRPPTCSYANTLTLLTKISD MTNXNGQIIILGDFNIKNIKWIPAGPMTKDRLGTILSDFC KNNNLKQFVNXPTRKNNILGIVFSNTKLENVEISXPIGSS DHRKITFTLDILEKPTKHLIEIWDFKGINWKVCAAELNLI DWPTIIDPVKTISEIYNEFLTIMRRICELAVPKKRINKELY NVPEYLVTFHKTIENKFDKLLKSRSWADYALYIELSRKF			
CR1	CR1-13_C Jap	.	Caenorhabditis japonica	G			cd01650	470	724

					TTCTTGATC	RKKLLKFRRNQELLTLKKSGNSHFTHIAKNILKPKFSMV					
					TGTTTCTAC	PTFVDQSGHHATSDHEKAEVLAVNFESQYVNKSNTSFT					
					CCAACTCCT	LPPDYKSDQEIPWVTDAEMLKFIQKCKNSCSDTPDKVPF					
					TATTGCTCT	KFLKLIAPFIASPLSQICNLSMMRGEVPEKWQQSIITPINK					
					CATTATTTC	IAKPQVPTDFRPISITSQICRIYERCILSKLLPFLDKIRFWN					
					CCTGCACAT	NNQHGFRPKRSTVSCMLESLNDWTNSIERGSQVDVIYF					
					ACCTTTAAT	DYAKAFDRVPHDLLLDKLVSIKLNGNLLKWLASYLSNR					
					GTATCTGTG	TFTVKVTKNQSSTKMAQCGVPQGAVLSPILFGIYVNEIP					
					TACACAAT	SKLPMGVHCKQFADDIKLYSEIPKDSDPQNNKLQQAID					
					ATCAACAC	AICQWSETKKLALNNSKTILISLGKTTKNCDYVINGSPIG					
					GTCACATTG	RKQLTRDLGFMISPKLDFSEHWHKATNQAKYALSLIFK					
					GCGTAAAT	KYCSDNSRLLTLLYKIFIRPILEYGTVISNPHKKSDIKSVE					
					AAATAAAT	SVQNAFTRRLYSRQKGRYIRADDPEYLSAPERNSLYGLS					
					AAATAAAT	SLENRRKYFDRKFLTKMIAGNVDIDTRPFFTFFQNTRTR					
					AAATAAAT	AKTKFIWARSKTKIRQEFFTNRTLTAITL					
					AAA						
					TAATTTTTT						
					GTCTACATA						
					GATTTTCCT						
					AACATCTCT						
					TTTTAACTT						
					CTTTCGTTT	KILVSGSSKNFFKFINQRMKESSDVGMIMDGKTXVYDN					
					GTTTCTAAA	YRKAELFAESFGKVYSTDDGQYTTCDLRTSAHLSEIIFEP					
					TCTTCCGTC	YAVELELSKLASKVNTSPDEIPAIFLKKICTSIALPLSIIFN					
					ATATCGTTT	TSMRTGFLPDLWKTAIVKPLFKKGSRSDANNYRPISLTS					
					CTTATACGT	TICKTMERMIRKEMMKHIYGNKLLSNAQYGFRPHRGTE					
					GTGTAATA	AQLLHYISCIQSNMEKKKDMSSVYVDFRKAFDTVSLPK					
					GTTTGTTAT	LQIKMESYGICGRLLNWLSSFLTDRHQSVVLNEVFSPRM					
					TTATATATG	VVKSGVPQGSVLGPVLFLLYVNDIGDSLECDTLLYADD					
					TGCTGATGA	LKIFSTDSAALQRDLDRISNWCRTWQMDVAPNKCEFITF					
					ATAAGTGG	NYSKKNSALSFPLMIGNFNLQRTTEIRDLGVIFSNNLSFK					
					CAGCTGCCC	SHISVIVRKAHQRINILFNIIRLSTIDVFIKCYVIYVRPILEY					
					TTCAATCAA	GSVVFSPISKELVKTIESLQKSFIYRCYKKFGIVYNSYFEA					
	CR1-17_CJap	.	Caenorhabditis japonica	AT	TAAACAAA	LEQNNLETLERRRLMADLVFIYKWLVSREINIGHDLFISS					
CR1					TACAACAA	ANLLPLRRHNYYIRCQLTGATRIGSQFLSNRIIRCWNSLP	cd01650	134	384		
					ATA	VKTFPVKFSSTAFKNRLQNLSLNNFLLLLHFHNF					
					TAGAGCTTA	VIPIPKKGSLSSPENXRPISLTDPFARLFERVICEYIKLHFA					
	CR1-8_CRem	.	Caenorhabditis remanei		TTCTTATTC	HKFSQNQHGFLAYRSCTSSLVHSISCYKSSLSSNNSLDVI					
CR1					TGTACTCCC	FFDFKKAFDKVNHKLLLQKLALFGIPHLFIEWFSNFLSG	cd01650	1	246		
					AAACCACC	RTFSIKIEDFTDTSITQIPSGVPQGSVSGPLLFLIFINDLLLD					

CR1	CR1-5_CR em	.	Caenorhabditis remanei		TTTCCCCCGC ATTACCACA TTGCTCAAT TCTTAATGC CTTCTTTTA TTCCCCTCA TACCGGTCA TTCTTTTTT TTTGTTGCC CCACATCTC TCACTTTAA CCCACCCTT TGAATGGTC TCGTGAGG GACTCATTC TCAGCCTCC CGCCCCATA TATTGATAA TTTTTTATA AGAATAAA ATACAAAT ACAAAATA CA	TAAATTATA ACTGTTTCT TTTTTTCTT ATTTGTATA AGTGGTGA AAGTCCTTC AAATAAAT AAATAAAT C	LALIPSLQVSAFADDIKIYSSNPVAVQKGIDLIETWASSN SLPLAHTKTSLLRLGSKNISFPYFIAGQPIETSKSVRDLGL ITDSTLKFKSHINKTIASALLRTKQLLKSFKSTSPQFYIFL FNCYVLPIIEYCSVVYSPPPASKLSLSLETPLRFFTRKIFQ RCNITYSSYSDRLAQLNLFSLRHRRLKAQLLLLLYKFLSG TSYFPHLDSYIRFSSSTRRPMNLICIKPKCSDFFSHTIPIWN AITSQSSYFLSPSEFNTLISSSITRY	VQKGPNSLAVDNSTKQVKFLSVGYANCFSVKNKLAQL ELVTITNNFDIVCLTETKLDNTFPDSLLSLSNNFSVVRKD RNKHGGGVAILISKSIRFLPIEIPSSLLSSEIAGVDIMAGG VSIRIVVGYHPSHHSKLDGMISCLEFLLSTRKNCVILGDF NMPHISWPSLTASDSHCKKFLSFVTKNGLTQHIASPTRL KPDNILDLCFTNTDILRDVRVGDLFSDHKLIHVTLSVKN RTKKLTKEVKLFRKADYASINYLLSNTDWVLRFSNLNA DGMYENLLSILHELIASYVPVKAINTLSKRHSAEILKLQK AKLNVWRKEGNSTNYKNISADLKVALIKEERRVNDEKL TNGSVKDFFKFINSRYKDNQEIGTLKNDSGAPINCDSEK VELFSDSFSKVFTEDNNVQPHFDKRTEELVSSPDFEPYII EHTLSKLTPKLNTTPDGIPALFLKNVCTAIALPLSIIFRESF RTSIVPTAWKTAIVKPLHKKGSRANPNNYRPISLTSSVG KVMEKLVRKQLTNYLNSNRLLSNCQYGFRSSMSTEAQL LSYQADILTNYICKKTTHSVYIDFKKAFDKVSLSKLKIKL RSYGIHDDFFNWLCAFLTNRTQRVCINNVFSSDRSVLSG VPQGSVLGPLLFLLFINDIGDAFESNYLLYADDLKLFSTN	cd01650	484	735

						ADCIKKDLVRLSVWCDNWQMGVAPEKCEVISFNHSKK		
						HSNSASLNFSINDAIIPQTKIIRDLGIIFNSDLNFSNHLDVT		
						LRKAHQRVNIFFNVLRHADFEIFIKCFKIYVRPLLEYGST		
						VFSPTSKEQVRLIESVQKTFIFRVFRKFNIPYISYFESIKHC		
						DILSLEHRRIIIDLVFMYKILVSKEVRIYNDVHISLPRFSN		
						NLRRHPYYIKSKLSNSTKNTCQFFTNRIISCWNKLPPHIF		
						PSFPNSDVFRCNISFNFVEPFLILKHSNF		
							cd01650	531
						MLDTADSGRPERRSALIAHELSRLNIDIAALSEVHLHEE		
						GSLKEHGASYTLYWSGKPKTERHLSGVGFMIKNSIASK		
						LENLXTGHSDRIISLRLPLHNKQHVVLFSIYAPTLQADH		
						AEKVKFYTDLRRLTQKVPAGDKIIILGDFNARVGKNYE		
						AWKGVLGKHGVGNCNDNGRLLLEFCAEQQLTITNTIFQ		
						QKDSLKTTWMHPRSKHWHLXDYVLVRQRNVGDVRHT		
						RAMPSAECHTDHRLVRCKLTLYFKPKPKRGGIPRRRLK		
						VNDLQKDTVRNSFQVNLQARLLDHPIDPSPEALWQHIK		
						NSILQSSEESLGFSSKKNKDWFDENNQEIQELLTKKRTA		
						HQAHLAQPSCHVRKAAFHLACSKLQQKLRDIQNKWWI		
						NLAEKTELCAVTGDHRGLYEVLKAVYGPTHQVQSPLLS		
						ADGQMLLTDKTSILNRWSEHFQTLFSASCVVQGSAIQHI		
						TQQPVKTELDIAPTLGETLKAIQQLKTGKAAGVDGIPSE		
						VWKHGGQALHAKFHELLVSCWEQGKLPPDLRDAVIITL		
						YKKKGEKSNCSNYQGITLLSIAGKILARILLNRLVPAIAE		
						EVLPESQCGFRANRSTTDMVFVLRQLQEKCREQNKGLY		
						VTFIDLTKAFDTVSRKGLWQILERLGCPPKFLKMIILLHE		
						DQQGQVRYGDALSESFPIINGVKQGCVLTPTLFTIFFSTM		
						LQRATADLDEENGIYIRYXTDGSLFNLRRLKAHTKTLN		
						HHVPELLFADDAALVAHTEAALQRITSCFAEAADLFGL		
						EVXLKKTEVLYQPAPQEDFHHPHITIGESELKSVQQFSY		
						LGCIISSDAKINKEIDNRLAKAYRAFGRLHKRVWSNKHL		
						KKSTKISVYRAIVLSTLLYGSESWVIYRHHLRLLERFHQ		
						CCLHTILNIHWTDCVTNVPVLEQAGVTSIEAMILRTQLC		
						WAGHVSRMEDHHLPKIVLXGELATGCRNRGAPKRRYK		
						DSLKQSLSFGHTDCHQXSTLASSRETWRHTIHNAAASFX		
	AviRTE_CAn	.	Calyptoanna	G		NTHRISLEEKRQHRKNRALAVPSQETFCCAFCXHTACPT		
						LALSAISVLAASVGRALLKSSFVKPSH	cd01650	531
								809


AGCCCTG
TAAGGCC
ATCCATC
TAAGAGA
AGGATAC
TCTAACT
AAACCTA
CGTCCTG
AGGACTT
CGCTGTC
ACTGTCC
AAGCTTG
CTGGGCC
CTGGCAG
ATGAACC
TTAGGAG
TAAAGGG
TGGGGCC
AGTTCCG
TGCACGC
TGCGCCT
CACCTAA
AAAAGCC
ACAGCGC
AGGCTCG
AAGGGTG
AACCCAC
ACTTGTA
AAGCCCC
TGCAGCG
ACAGGCG
AGGGWC
GAAACGG
CAGGGGA
AAGGTGC
ACTGGAA
GCCGCAG
ACCCAAC
CCTGCAT
GCAGGCG


GCTCAGG
ACATTGG
TCGCTCG
AGACTGA
CCCAGGA
GATGACA
GTCTTGTT
CGGCAGC
ATCCTGA
GTGACCA
AGCAGCC
TTTTTCAA
GGAGAGC
ACTGCTT
GCTCCAC
ATGGAGA
GGGGCCT
AGAAAAG
GTGGCCT
AACCAAA
GCTCATC
TCCACCC
CCCACCC
GGTTGGC
TASCCAC
GGTCAAC
GGGCATC
TTCTGAT
GCAGTCG
AACAAAG
ACAAAGA
GACAAAG
GCATACA
CCTGCCT
CTAAAGG
TGTGTTC
AGATTGA
CACTCGC
ATGTTGG

Daphne	Daphne-17_CTe	.	Capitella teleta	AACATCA GAACC	TAGGGCAA CTACAAGA TCTTATTCC TATGTAAA ACACCCTA GCTACTCTT ACACTTTTT CTTCTCTTT CCTATTTGA TTAGTCCTC CTAGTTCTT CTAGCTCTA TCCCCGCTC TTTATTCTT TAACCGTTT TGTGTCTCA CTTATATGT TTACCCTGT TTTCGGTAC TTTCTTTAA CAATTTCTA TTTCTTTTG TACTTTTAT AACTTTACT GTACTATGT TTGTATTAT TTTTTGTAT TGTATTACC ATCATGATA ATCATTCAA TTCTCACTC ATTTGGTGG CATACATAT GTTACGTTC AGTTATGA ACTTCGTAT GTCTCCTGA CCCTTGTCG	MTSLTTMAAIATTHARTDWMTNQPNMAERTDNDSLVI GTYTATVHLDSTDNDALLRSHYNLNQVLQTNQQEQDN TPINFSPYYDHEHLIQHLAQPTNNVKFKILSLNCQSLNAK FDQLMILIAQLNHNXVIFDAICLQETWIKESELPPVFNIK GYQAIHQARHCSGHGGLLIFLKDSYKFKIIPFEKSAVWE SQFINIDFENGNKDSLLVGNIYRPPHKVKDDLDLFSNQF TRNLLQCSKYKNIAIAGDFNINLLDVNSTTSPAAHDFPSS IISNDFLPKITLPTRISESSSTLIDNFFCKLKNDSTSSGVVIH KISDHQAYFTCLPIRSNKIHNIEKIKMRRVSQESIIAIGNDI RNATGRLDLSTETDPNVNYEALMETLHDSMSQHIPIKEV KFNRNHKHKGSSWITAGIVRSIRERDKLYCKLIKAKPNSN KRESIKLTLQLYTRILQNIIRKAKSSYYSMHFAQLKNDM KKTWKAINDILHRNKNSLNRNYPEQFKINDQLISDKNIA EEFNSFFTSVGKILASRIAHNSTNTATDYLSGNVTSRFVL QKVSESDIERIIAQLKSVSSCGYDMVTPKLLKALSPFLIM PLTIIVNQAISTGIFPDSLKRAKVVPIYKKRCPFIKKDDET VLTNYRPISLLPAISKIFEKVIHNQLQHYFTSNDLLYHSQ HGFRPKHSTELAALQLLDQITTDLDANKIPIAVYLDLSK AFDTINHNILLKKLNYYGITGTAYEMMKSYLTNRKQYV SFADCNSRDIMIDTGVPQGSILGPLLFIIYINDLHKSCSNF LPIVYADDTTLYKTINSPNSINSINQELEHVNLWLKIXKL SLNASKTRYMVFHKPRKRLPTLNISIDNENVERCISFNYL GIHLNQNLTWNDHIKRTSLKIARATGRINCLKHFLPKKIL LTLYNSLILPHLTYGILAWGAKSNDIAKLQKKSVRIITAS KYNAHTEPLFKAHHLLKIEDIYKVYQLNFYFKWKNGNL PKYFNSFSCNTNRDTHSYHTRTTSTVSARVNHEFAKHSL RHSLIILLNQTPNIITDKITTHSYHGFSSYVKNYYINTYTN VCYNSTCYICRNQQQLL	cd01650	615	877

					AATAAATA ATCTATCTA TCTATCT				
CR1	CR1-6_CTe	.	Capitella teleta	GAACAGG AGAAAGA TGCCATG AGAAAGA GAAATGT CATCATA CACGGAC TGCCAGA AAAGAAG ACCAACA ACAGAGA GGAGAGA AAGAAAG AAGATGA GCAACAC AATCGAG AGGATTA TGGACGC ATTGGAA CTGAACC TAGAAAT TGAAAAC TTCTACC GACTGGG AAAANTA CCCAATG AAGACGC CGACAGG GAACATA CTACTAG CAGACCC CTGAAGT TGTGCTT CAATAGT GAAGACA AGGTGAA GATGTTC	TAAAATGT ATCAAAAT AAAATAAA ATAAAAGG AGCTGATC GTAAGAGG CTATTTAAG CCTGAAGA TCAGAAGA GGACCTGT AAGTACCT GTAAGT	MTMYIKDHIDYTMLDIGEDTCEVQAARLLMEQGQATIF ASVYRSPNSSKENSDNINAFIRNFSNERAYTLIAGDFNYP TIGWKTMTGRTEADIEFIEALRDGFLIQHVDRPTRGRGS NNPSLLDLVLTKEHTATPTVEYHQPLGKSDHSILDIRIEM KQSETFIKTIPNYNRGNFDRMRRVLSDVDWEEGLGREE TTQKKWELFHETITNAVNRCVPKIKIKRKVLKNIPVDNK TLTKIRRKKRLWKQYLETGNQTIYHDYCRTRNQVRQLT RKNMRELEKNIAKEAKTNPKKFWAYANQKTKYRETIP QLLKTGNPNDGMLTENNQEKAQTLATFFNSVFSEKPND HWNLSPSSYPTIEQDLIITEASVLHHLSKLKKTKSPGPDG IHPRILIEIKEIIAKPLSIIYQHSLDQGQLPDTWKTANITAI HKKGDKRIAGNYRPVSLTSVPCKIMEKIIRKSIMDHLIRN NILSNKQYGFITGRSTLLQLLTVLDQWTEILDRGSEIDVI FLDFKKAFDRVPHQRLIDKLEYYGINGKILTWISAFLINR KQRVTIHEATSAWAEVKSGIPQGSVLGPTLFIIFINSLPDV TSSPTFLFADDAKIFREITKPTDSLLLQEDINNMYHWTLE SLLTFNPDKCSSMTISRSGTNHRQYQMNNSPLNKTIIERD LGVILDNRLTFDHHIQAKIKKSNSIMALIRRTYTYLDNHT FLLLYKSFVRPHLEYANQVWKPHLQKHITAIENVQRRA TRILPGMKDLNYEQRLRKLKLPTLQYRRIRGDLIETYKI ATEYYDTQITQNFFQFNSRTSRGHPLKLYHQPARTELRK NSFKIRIAQLWNQLPSEIVASPSIATFERRLDHHLKNHPK HNPFRPP	cd01650	423	675

				ATGGGGA GGCTGAA CAAATTA AGAGATG CAGAGGA AGATATC AAACGAA TTAGAGT TCAACAC GACATGA ATTTACA AGAAAGA GCAGAGA TTAGAGA ACTGGTA CAGGAAG CAAAAAA CTTAACA TTAACAG AGAGTGG AGACTTT GTACACG TCGTGAG AGGAAAA AAGATAA TCAGAGT GAAAGCC AGAAAGA ACGGTCA GAGCAGG AGAGAGA CCCAAGA GGACTCA ACATCTG GATCACG AACATAG ACACCCT TACAATA GCAAAAT TTAACGA					

				ATTAAAAACCATAATAAAGGAATCGCCACCGGATATCATAGCCCTACAAGAAGTTAAACCGAAAAACGCACGTACAAACTGGGAGGCAAAGAATTATATAATACCAGGATACAATATAGAAGAAAGGATAAGAGAAGATGGAAGAGGA					
L2	L2-6_CTe	.	Capitella teleta	T	TAAAAAGAAGTTAAAGACTTTTTTTATTCAATAATGCTTTTAATAACGCAGAACAGTTTTAAGTGGAAATGAACTTGGCTTGTTTTTATGTTTCTATGATCTTTTATCCAAATACATTTTCATATGTTTTATTTTGTGCTTTGAAACTTTTATGTATT	QLIKDKKAAYYKKLFSEANHKETFQLLKGTTKPPAKVLRQEQEDVTLAFEFQQFFNTKIRKIHDTLQTDENNSPHSHDKRTSASFTAFRTHTQSQVEKIIKSAPSKTCELDPMPTWLLKEQNVLDQVLPWVTGITNSSLELSHVPKDHKTANVKPRAKKDGAAYFDLSSYRPVSNLQFLSKVLERAVASQLTEYLEKETLMDPLQSAYRRGHSTETAVLKVKNDIDLAIDEGDGIILVLLDLSAAFDTLDHAILLKRLEKCLGIKGSALDWIESYLRNRPQRVKVGESYSSDSILQTGVPQGSVLGPLLFLLYVLPLQEVISRFRIDRHHYADDTQLYGRIRLNSIKTIDEDVREMEASIQDVRIWMKKNKLKLNEKKTEALIITSKEMKKHAEGIKLKIGGDTVTATSKVKNLGGWFDEHMKMQEQVNRTTKAGYFHLRTINRIRKNLDTPTCAKAINATVTSRLDYHNALISGARQCIMKPMQRLQNHSARVITMTDKREHIRPVLEDLHWLPVKERSDYKVLTFVHNALHDELAPAYMKEMYRVYHPNRALRSGEDRWTLTVGTAARHEGHRSPXIHGAIMWNTLPEALRSEASHPRF	cd01650	150	418

Nimb	Nimbe_CTe	.	Capitellateleta	ATAGACAAG	CAA	TTAAGCACT TGAATATA ATAACATTT ATTTATTTA TTTATTTAT TTATTTATT TATTTGTTT ATTTA	MTILQWNCRGFRANFQELKTLVAQKDATVLCLQEIKVK DSDSCSLNGFNTLKKCTLGIDGRAVGGAAVLIKKDTPH QHIPLTTNLQAVAVRVTLHRLLTVCSIYLPPGRSHGDLV ELDRLVQQLPPPLLLLGDFNAHSELWGSEEPRPSGRVVE DFIAGNDLSILNTGSQTYLHPASGSFTVIDLSLCSPSAHID FTWEVDTDQHGSDHFPIFISSECRPSLSHPKWVFKRADW TAFGSLCEIEITEDIFEDVDDPLDIFTTLLIKIAHETIPKSCS STKCRKPWYTDECKKAIRLRRAALQKFRHHPNDDNLLS YRKAQAVVRRVIKASKRASWKNYISKLTIRSSVKKTWD MVRKISGKYKGHQIHHLTTRSGDTATDKESIAEALADEF ALNSSTSHYTDKFQKYKDKAERKVLNFNSNNTESYNLP FTLLELLMSLKSSHDTSVGPDDIHYQLLKHLPINCKMILL NLFNRIWVGEPFPSAWRNAIVIAIPKPGKDISLPTNYRPIA LTSCLCKTMERMVNGRLVYFLEYNNLLSPSQSGFRSQR STADHLVSLETYIRDGFAHGDHVLSVFFDLEKAYDTTW KCGIISDLFNMGLRGHLPVFIRNFLSHRQFSVRLGSVDSS PHPQEMGVPQGCVLSVTLFSIKINSITNAVGADIHKCLY VDDFTISYRSRYLPTAERKIQLALNGLQKWADESGFRFS PTKTVSVCFSGRRGVFPDPLLTLNGQKIPIVEKTKFLGVI FDKKLNFKEHIQFLRGKCQKVLNLLRVISKMNWGADRE VLLRLFRSLLRSRLDYGAIVYGSARISYLERLKPIQNQAL RLCLGAFRTSPVTSLHTEANEMPMAIRQKLLSVQFALRV CSDTTNPAHQCIFNYNNDRFYLSKPNAIRPLALRIKEDL QTICPDIKAITPNQLFNTPYWLLRPPELDISLIHFGKKTAN PNYTLRNEFYNLMDKYPDHKVIFTDGSKSDSAVACSAT ADNLRIQIRLPDSASIFSAELLAIYQVLTLLECSANDQQQ FLIATDSLSSLQAIGNFNIKHPYVFFKILEKCTLLHKKGIDL VMAWCPSHVGVMGNERADLLAKEALSFTTCTIRIPSSD FKPITHEFYKEKWQEQWSSEQENKLYCIQPTLGKWAKS SREIRREEIVLARARIGHSHLTHGYLLRRREMPPVCIPCQSI LSVKHIFIECVDFDILRRRFFNVSSMK	cd01650	489	746

CR1	CR1-13_CTe	.	Capitella teleta		TGATTTTACCCTAATTTCCTTGACACATATTCTTTGTTGCCATTGTATTTTTTTAAGCTGGCAGTTCTTATAGGTTTTCCTGTTCTGCCATGCTGTTGCATGCATTGATTTTATTGTACATATTATTTTTGCATGCATGAAGCTTAATAAATAAAAAAATA	LDCPQIPDPHFTVENVLSVLASLDTNTAFSPQDIPPIILKKCRSLLAPSLCIIFNVSIRCKTVPLEWKKARVVPVFKKGDKSAVSNYRPISILCIVSKVMERCVFNHMYPLLSSFIHPLQHGFCKGRSCTSQLLDVYNRIGSILDKGGQIDVVFLDFSKAFDCVSHKLLLYKLKHFYGVDGNLLEWIRSYLSERSQSVIVEGKSSTWLGVPSGVPQGSILGPLLFLLFINDMPYKTLHCTTALFADDSKCFKSINTTHDCELLQRDLNAMHEWSLIWQMSFNPSKCKVLSITRSHAPIHFNYAMNGVILDHVNVFKDLGVFIDQTLSFSTHIESTILKCNKMCGIVKRTIGFNAPVNVKSSLFMSLCRSHLDFASPVWSPHLKSKVKQIESVQRSMTRFILNNADLSYIERCQFLEILPLSFRREITDLLFLFKYFNGAMNVDFTSIVAKRCFNTRLRSGPFIELLTPTPRTDSFINFYFNRVAHMWNILPLEIRSCLSVSTFKSKLLSFYSNKMCFYDVNNSCTWTSTCRCNGFYHT	cd01650	69	322
L2	L2-9_CTe	.	Capitella teleta	AGATCAGGTAGCACCCTCCGGTAGCACCTAAAATTTGCTCCGACCATTTTGCTGTTTTCCGATATCTTCCACTCCGCGGATTGCAGCGAATCTTGGTCAAGAC	TGATGTTTACATGTTCCCCCTTTTTATTTTCAAATTGTATTAGCGTTCTGAGCATTTTTATGGAAGCGACGCTTTATAAAATAAATTTTCATTTTCATTTTCATTTTCATTTTC	MWKVIFVICVLSQIKAKTGSGETRPGFTICLPQPRHASPTSMPSLCLEFIRALAVKAPARLPTDLYKRLNGLRISRRRLSKRGTKGYGRGVIPVQVSALRSADRYVNTARRDASLPTSLVTIEPMTIASSSSCVLLNCRSICNKASEIRDFISDRRMDFLCLTETWLTGTDEDKVVSAALLPDDYAMVHRPRDARGGGVAFIHRRSITLSQIRLGDYSSFEVLDALVCTARRLRMCVVYRPPPSRRNNLTVAQFMSEFADFLSDIVSLPGEPLIVGDFNFHLDDCGDADASTFSDLLTSFGLKQHVRGPTHEKGHTLDLVITKTDTALIWSSVSRDLCISDHYPVFTTLAIPKVLNPTREISYRRIKAIDKEQLDRLIEASSLSSPPENASLDENVTNYQRTLTGILDTIAPLRSRTVRIRVNADWYNDEIRAEKQRRRQAERRWRKSKLPADRELFKVAKNRTQSMIDEAKSLHFREKVSACSDDPKQLFRVVDTLLGNSTESPLPVHDSRVELSNRFSDFFVEKIATIQSAIPSVDNPPRPQAPLPQSQLATFQPLDVADTCALISGAAPKSCELDPLPSPLVKDLPSLRPIITRIINQSLATGVFPMEYKSAVVRPLLKKPSLDRDSLKNYRPVSNLSFVSKLIERAVASQLASHLRNNDLMEPRQSAYRPGHSTETALLCIHNEIVCAIGEGKAVLLVLLDLSAAFDTVHHPTMLSSLERLGVTGTPLKWFESYLSERSQAVQIGDAKSDPAAINCGVPQGSVLGPILFTIYTSSLGALLREQGVSYHLYADDSQLLLTCDPRDPAPAVARMEACIRSV	cd01650	620	878

						QAWMSCHSLKMNGDKTEILLITSKPMRKVLPDVQLCID GHQLQPSTTARNLGVTFNRHASMESHITSVCRLSYMQL RRLRRIKKLLPRDVLESLVHAFVTSRLDYCNSLYLGIPQ TQLDRLQRIQNSAARLVTDSGAREHITPVLQDLHWLPV KERVIFKVLLIVFKSVHNLTPQYLSEMITVHIPTRHLRSS NQNFLTVPFTRSSFIKNSAFSVAGPTIFNSLPINIRQAESV PIFKRLLLKTHLFHNYFLT			
Rex 1	Rex1-1_CTe	.	Capitella teleta	GGTGTAC AAACTTC GTCATCA AACACAC TTCGTGC ACAGCTG ATGTTGA GATCCTA TGTGTGC AGTGTCG ACCATTT TATCTAC CGCGTGA GATCAGC TGTGTCG TCTTCATC GTGATCT ACATCCC TCCCAGT GGCGATG CCAACAG AGCCACG GAAGCGA TAGCGTC TGTTGCTT TGGACAT ACAACAA ACAAAGC CAGAGGC AGCTATC GTCATCA CAGGAGA CTTTAAT	TAAATGTTAT TATCATCAT TAATGGCTT AATTGTCAT ATTCATCAT TCATTTTAC TCATTATTA TCTGTTTGG CAAAATCA TTCCTTCGG GCGAATAA AATCATCTT ATCTTATCT CATCTCATC TCA	MIYCLPTYIRKLERQKPMIKTIRQWTEDAAEQLGGCFAC TDWNIFEDTTTNINEYTNVITDYIKFCEETIIPTKEIKIYPN NRPWVTNDLIHAVRKRRRLWMSGHHEECKEAQKEVNT IIAQCKDNYKKKIEESFRSGNPKKTWSGLNTITGYKKKP MTRTSISNDEGWANELNVFYTRFEKEDEDLVFDQPAIAP IRFEEDVRRVFMGIKQDKAQGPDGLSPRLLKACAQDL APVFTGLFNRSLEESCVPTLWKCSTIVPVPKRPTAKELN DFRPVALTSIVMKSMERLVLRRLLHQVQLSMDSNQFAY KARRCVEDASALMVHLVSEHLETPGSYARVLFVDFSSA FNTMRPSILISKLEAQGVAPPLCQWILSYLKNRPQQVRV NQSLSTIRHTNIGAPQGCVLSPVLFTIYTDEHRATSQETR VIKYADDTALVALLSPGKEHLYTDAVSHFIDQCDNDDL KLNSTKTKEMCIDFRTHNTIITPIIIKGQPIEQVTTYKYLGI TIQSDLRWTSHISTQCKKASARLYHLRKLRQFHVDRHL QELFFTATIDSLLLFGITVWGGRCTARDRNNISRVQRIAG KIMNTTVQPWGVSHRQRTIMRAKKIIREPLHPLHYTFQR LPSDRRYRALKAKTNRLRQTFVPSAIRLINSQPI	cd01650	257	510

				GGAGCTTCGTTACATGACGCCCTCCCGACCTACGTACAATACGTAAATTTCAACACAAGAGGTCGCAGCTGTCTCGATCTACTGTACAGCAACATAAAAGATGCCTACAAGACCACAAGCCTGCCACCCATCGGCAGATCTGACCACACG					
L2	L2-3_CTe	.	Capitella teleta	GACTTCAATTTTCAACAACATGTGACTGATGCAACACATGTCAATGGTCACACAATCGACCTAATC	TAATGATTTTCCTTTTTACCTCTGTTTGTACAGCGCATTTGAATGTTCTTTGATATTGCTGAAAATGCGTTTTAAATGTAATAAATGAAATGAAATGAAAT	MSLTHEHLSIKVIDVDRSVTSDHFGVVFTINCPKPAREKQKLKVRKWKNVNISALDETIASHIPAIESAQDPFAAYNNCLTQVVDEFAPETTIEVTKRTDTPWYNAELRHLKTECRRLERKWRRTQLKIDQEIYRKKCRCLRRLRDRTKMTYISSKLQNATSTKDTYNILNQLLHKQSSSPLPRHDDEIELAETFSSFFSEKIERVRSAIGAHPDPDVTIEHHLSTEITPLTSFRPVSRDWISKTLSSTTPKSCNLDPIPTALLRKCPSAVPLLCFTINWSLANSTMPDSHKTAHVRPRLKKTSLDHDQLKNYRPVSNLSFISKMTEKAIVAQLKDHCFVNNIDVQMQSAYRAGHSTETAMVRVQNDLLDAVDRDGGAILVLLDLSAAFDTLDHPLILHALKYNIGVTDSALKWFESYLTNRKQLVTIGSQTSDTKNLVYGVPQGSVLGPILFTMYTRPLSNIISNCGMNHHLYADDTQLYIAVNVRSPTSVSSVKESIQECSNQVHRWMTANFLKVNEEKTEVLVITKPSLTEIAPSSFNIIGETITSSASVRNLGVMLDNKLSMADQVRNICQKSFFHLHMIRKIRNFITEDAARQLVHANVMSYIDYCNALLAGSPSYLIDRLQHIQNCAARLVKCADRHSPSLPLLKELHWLPIKFRIMFKINVITFKALRGLAPPYICSLINVYTPARSLRSSNFTLLS	cd01650	306	560

						VPKFHLQSYGYRTFAHQAPILWNRLPSNLQNQQSLDAF					
						KAQLKTYYFNIAFSSI					
						DDLNDLVSSFNRRCLDVTNAHAPLRQITLKGDGMKPW					
						YSDAIHSARVVVRRKYERQFRKSPLEVHRQIYVVDQCKEV					
						VQLIQNTKKEYFHQKFASASASAKEVFRLVDNLLHKEPNH					
						TLPTNVPLGDLPQTFNKFFYDKVHQIRAELDASPTLPSLT					
						TPQPLVAERGGAPALETFEICSAEDLRKIIQTSGSKTCMQ					
						DSIPTPLLKEGAVLEAALPAIVNIINRSITEGCVPAELKKA					
						VIAPILKKTGLDPNNLAHYRPVSNIGFIAKVMEKVIARQ					
					TGACATGG	LTDHLTRHNLHDQLQSAYRKGCSVETSLLYIKSDIELILD					
					CTTTGTGAA	AGDAVLMILLDLSAAFDTIDHALLLTRLSEIGLEGSALA					
					GCGCATAG	WVKSYLMERTQVVRIGDKISESSDLDIGVPQGSVLGPLL					
					AAATGCAT	FLIYMIPLGSIIDNRGVLRHSYADDTQLYCRVVNIKDPTKL					
					CATGCATCA	QKQIRDMEECITSVRSWMKSNKLKLNDAKTEVLLISRH					
					TGTTGCGCT	EPRQPVEIRVGNAVIKPAPLVKNLGAYLDCRLSMKDQV					
					ATATAAAA	NRTTSAVYYHLQRIAKVRPFLSDSACANAINATVTSRLD					
					TAATAAATT	FQNALLMGLPDCQLRKLQVAQNNAARLLTRTKKREHIS					
					CAATTCAAT	PVLTALHWLPIKHRIEFKVLSMIHKSFHSLSAPDYLKKL					
					TCAATTCAA	CRPYVPSRNLRSGDDPFKVVQLRTRNLYGARTLGGLGA	cd01650	233	497		
L2	L2-13_CTe	.	Capitella teleta	C	TTC	RLWNALPMSLRKPMTFSTFKKNLKTHFFRISQ					
					TAGTATCAT	KMFLDLMHSHAFFXLIDKPTRITATTSTLIDNIFSNVINX					
					CCTCGATCT	THLSGVLYTDISDHLPLFTVNQFICRNNSIVTKTIGYRSF					
					ACAACTAA	NETNKLAFQNAPLAIDXDXLFXCHNCGAAYTLFTSNFV					
					CATATGAA	DCYNESFPIKSHKINPKRNQPWLSPELKRSINRKNRLYR					
					ACGAACTA	RYHARPSVFNEINYKSYKSMITKEIRVAKRTYYHDLLEQ					
					TTTCACTCT	NRNKIXKTWDIIKEVIGIDNTKKPIASITINDSTVSEPIEIA					
					GAATATGC	QHFNKYFSSIGNNLASSLPINHLDPMRYMNPNTPRSLFL					
					GCTGCACC	SPVTSNELYTIIFNLKDASAGIDDIKPIIIKEFSQLIIKPLLHI					
					ACCCACTAT	FNLSFTQGSVPCAMKKANVTPIFKGGDPTSLGNYRPISV					
					TTTTTCATT	LPVFAKILEKLMYNRVYNYLTENRILFNKQFGFRRGYST					
					TTTTTTTCT	EMALLTALDELSSALDMKRNVIGVYLDLRKAFDTVNID					
					CTCCGCACC	ILLAKLFKYGIRGIAHQWFTSYLNDRSQRVKIGCSFSEM					
					AAAGCTGA	KSISCGVPQGSTLGPLLFLLYINDLPNVLTNIFPIIFADDTS					
					ATAAAATTT	LFMTGVDINDMTEAFNHQLSKINEWLIANKLSLNIQKTY					
					TAAATTGCA	SMLFTLSPHLYSSPIPITINNTEIQRVHSTRFLGVIIDDRLT					
					CAATTTGAA	WANHITHVSNKIAKSIGIINKVKHSLKQDTLMILYYSLIY					
					CAGGCGAC	PYLTYCFLLWGKASHSHINKLTILQKRVIRITNSGFRDH					
					TCACAAGT	TRPLFNTLRIIHMDDLFTYLCSVYIFKCIKIIFPRSFLDILN					
	Daphne-10_CTe	.	Capitella teleta	CT	AGCCATACT	YFTDYNSNAPRSSTRASLTRSIIVPLCRTTLRQKQLRFQSI	cd01650	337	594		
Daphne					TAGTCGCCT	KLYNNFLMPLEILDNSNSISHFRRFLKNVLS					

									425
								155	
							cd016 50		
						KRNQLTRATRRAKIKYYSDKINESASPKEIFQVTSFLLQS KREVVLPKATTMSSGSLADEFAEFFENKITAIRNGISAAP VQSSPLSDVVALHTFALTNAEEVSTLIRKSTSKSSAVDPI PTWLPKKCILTFSPILAAIVSASLVSGCVPDQLKAAVVTP LLKRPTLDPEVYKNYRPVSNLSYVSKLIERVVAERLKEH CALNNIQCKFQSAYKQNHSTETALIRVHNDILRAVDRD GGAILVLLDLSSAFDTIDHTLLLHHLEHQMAIGDTALKW FSSYLSERTQRVKVKESLSSPRRMLYGVPQGSVLGPVLF TLYTSPISAIIRAQQLQYHLYADDTQLYMSFTPTSAASTT SSGASMESAVSAIKRWMDDHYLKLNEDKTEMLVIQRP NPRSSQPTHRPTLTLDGNNINPSTSIRNLGVLFNSAFNLE DHVNALCRTMFFHVNRIWRIREYLTTSATRALMQAHVI SRLDYCNALFYGLPVVVRNRLQKAQNAAARVVVRAKR SDHITPILLQLHWLPVEFRVQYKICLMTYKALHGLAPAY IDELISVHRPTRRLRSSSDIRISPVQYKLKSVGGRCFSSAA PSLWNELPQSIRLSSHLLEYKSKLKTHFFRKAFNDIYLNK STSYHFIYILLYIAFEQFYFFLNFAIHKFEVINK			
						TAAATAAA TAAATATA AATTAATTA A			
						ATTGACATG ACCCTATTC TTTTTCGCT TCTGTATTT CCATTTCTC TAGTATGTA TTTCCATAT TTCATTGTC GCTTCCTGC TTTCTTTAC CAACCTCCC TTTTTCTAT TTTGGCATC TTATATTTG CTAATATTT CATTCTTTC TGTCGTGTC AATAAATG CAATTCAAT TCAATTCAA TTCAATTCA ATTCAA			
					Capitella teleta				
				L2-11_CTe	.				
			L2						

Proto2	Proto2-1_CTe	.	Capitella teleta	GCATCAAGGAAGGTGTGCAAGGGAGGACGAATATTGTTGATGGTCGAGTAGGGGAATCTGAGATCTGTGATGTGTTCCATGATGGCTATGAGCATCTGTATCAGTCTGTGCCTTCTGATCAGAGTGAACTCGAGGAACTCCAGCGCGAACTTTCTGCGGGAGTAGATCGCACATGTAAAGTAGGCTTATGTTACTGTGATCACGTGATCAGCCCACGTGATATTGAGTTTGCCGTTCGTAAACTGAAACCAGGTAAATCTGATGCCTGTCCGCAGTTA	TA	MSDHFIHGCPELYVHLSLFLNAMLIHLDVSSVMSRSVLVPIPKNRKKSMNSASNYRSIAISSIVGKCLDNVILVKHADVLSSCDLQFGFKRTHSTTQCTFVLNEVVDFYMRGGSSVYVTLLDATKAFDRVHYTKLFRLLVDRSMCNATAKLLLCMYVNQSLVVRWQSVTSDSFPVTNGVKQGGVLSPMLFCVYIETLLERLRRSGIGCYIGSTFAGAIGYADDVALLAPSLITAQKMLNVCEDFSREFFVQFNSSKSKAIVFNAPPNAVRSMRKLTLNGSPIDYVADALHLGHVVGDKSASLNMENACRDLVVRVNMLCSNFNFVHLDTILLLFNSYCTSYYGCSFWDFSPRTINGFVVCWRKCVRRLLDLPYRTRSKXLHLIVLKPEPRHQLFFRFQSFYVGCTTSTNHIVSLCTRLLPFSSSSVACNLRQLMSFVNINYDQLNNLIDNHSIRKRLSEFCEATTDQESLCCASSVIELMKIREGSLISPINPTEATDLLKHMCCSND	cd01650	35	295

L2	L2-15_CTe	.	Capitella teleta	AGCACATATCTGTTGCGACCCATCACTCCGGACACACTCTTGACCTGGTCATTACGCGAGAGAATGACCTTGTATTGAACTCAACTGTCAAGGATTCCCTGATCTCAGACCATAAGGTGATCCAGCTCAGCATTAACATTGCTCGTCCCCGAGCAGAACAGAAGAGAATGA	TGATTTTATTAATACGTCCTCATACTGTTTTTATCACTTTTGTGCTCTTAAAAGTCGCTCATTTTGTTTTTAGTCACTTTTAATTTTAGTTACTTTGTTCACTTTTTATATCCATTCAGTGTTCATCTTCTTTTACTCAGCGCTATTGAGCATTTTTGGATTTTAGCGCTATAGAAATAAATTATTCTTATTCTTATTCTTA	RTGSSRHSMLSLLLPASPMPWKSSETLTTLLTSTHTTKYLPQHLTDMHLRKPESSHYALTLLGTMMNSEHQNMSEKEGTQVENKWSRDRKETYKEQRDIYNRLLNQTKTQYYSKMVTDKARNPKDLFQLVDLLIHGKETNPVLPEHTSLDELSDRFATFFSEKIETIRLNLSNRQTDSQLLVHQEAEPELVSSWMVFSRVTXEEVTRVIKKSPSKTCSLDPIPTWLLKACLPALCETITAIINDSLSSGYVPDSMKVAAVSPILKKTNLDQNELKHYRPISNLTFVSKTLERIVAHQLNIYMAENHLHEQFQSAYRQHHSVETALVRVQNDILMNIDRGDAVLLILLDLSAAFDTVHHDTLIRRLEGIGIGGIVLNWFRSYLKDRHQSVTIDKHSSPRRQLGQGVPQGSVLGPILFSIYMIPLAQIIERHNMSGHYYADDTQIYISFRPSSPLADIEAKANACCSEIQQWMARNMLQLNEGKTDVILFGTKNILRTLDQKASISIGGSVVLSSASVRNLGCVQDSELKLDLHVNQICRSAYYHLRNINKIRRYLPLDVTEQLVHSFVTSRIDFCNSLLIGIKKANQQKLQRVQNAAARCILYIKSREHITPYLHALHWLPVEYRIQFKVLLLTFKSLNNLAPAYLTSLLHQNAVLASSRRNSSRLLSIPRSRLVGAGDRAFQHAAPTLWNASPPAMRTLSDVNVFKKGLKTFLFRRAYSDYC	cd01650	258	520
Daphne	Daphne-5_CTe	.	Capitella teleta	TG	TAATTATGTTCTTTTCCATTTATATATATATGCATATATATATTTGTTTATTTGTTTGTTTGTTTTTTTTTTTCTCGTTTTTAAATTTTTTTGTCGCTCTCTCTTTGTGCAAGACTCGTTAAAACTATTATTTTC	ISSNDPCDAYPRFLNKITTHYESSFPIIKKKPKKRDNNKWITNGLIISIKYKNKLYLDYKRHETIANELAYKKYKNRLSGIISQAKKNYYQGLIDENRHNMKNLWRTLKELIGSKKRSSLSNEFLINNKPCKDKDMISNTFNQYFVNVGPKLASIIPNVDSNPEDYLSGSYPNSLFLTPVTINEIKLALNKLKTTASAGQDKLKPNIIKLISDEISLPLTHIINLIFEHNKVPDEFKFANITPIYKAGNSMEVQNYRPISVLPAFSKLMERLLHDRIYVFLQQNEIISDYQYGFRKXFNSEMALAVTTDNIISSLDSQKHVMGLFLDLKKAFDTVDFNILMKKLVHYGIRGNALDLLKNYLSNRKQSVKIENSVSSELPVTCGVPQGSILGPLLFLIYINDLPNAXTSCNPIMYADDTNLFLSGKNIDEMENKFNTDLVALTKYLKCNRLSLNVSKTHSMLFTLNNSLRDSSLSLSIDGSIIDTVKCINFLGVKIDNKLNFSEHITHTCNKVSKSIGIIRKISKTVNRPTLLMLYNSLILPYLTYCNMIWGKSANIHINRLFILQKKALRLINKAPLLTHTAPLFHECSXLNIH	cd01650	239	495

					TCTCTTCCT TTTCTATGT ACTTTTCAT GTGGCGCA AAAGCGAT CACGCTTCT GCCACTAC GTTATTTTC GCACCTTTT ATGTTGTGA TTGAATAA AATTATTAT TATTATAAA AAA	DLYIYRSTIFIFKLSHHMLPTPVSKFFNLTPLVHGLNTRR QLNYKVIVPFCRTSLGQKSLNFNIAKLYNNILQPLSLLNL DSLEKFKKSIHCILVAKYL				
					TAAATTTTCA CATAATTTT ATATTTCAA TATGTTTTT CGGACTATT CTATGTGAA GCGCATAG CAGTGTTTT ACCTTTTGC GCTATATAA ATAAATAA ATTGATTGA TTGATTGAT TATAA	TLKIAQVKPRPKKKPGVNTMDPKLFRPISNLPTIEKLLEKC VIRDLHLYIQQHNIIDPLQSAYRPGHSTETAMLRIQSDIK TSLDEGNGVLLVLLDLSAAFDTLDHNILLSRLQNSAGIT GVAIDWMKSYLSSRQQFVQIEEKRSVTHDLDIGVPQGS CLGPLLFVMYTTPLGAIFRKHGLSYHQYADDTQVYCSL PSDDQNSLKNVLTRVEKCLDEVKCWMTQNKLKLNESK TECIIFSSKRNKQLFHQTTIRFGGSVITPVTSVRNLGAHL DCEMSMDKQINNTIKSAYYHLRRIAKIRRHLDRSSCTQA VVAFVSSRLDYHNGLLAGITQKNINRLQKVQNHAARLV SQTAHHCHITPVLADLHWLPIRARIDFKLMLNAHKAVH CNDAPKYMKELCMKYTPSRTLRSSSDPHLLAVKKSKRN IGASTFDVYSCKTWNELPRDVRDIEETICFKKTLKTHLY KRCF	cd016 50	12	272
	L2- 17_C Te	.	Capitella teleta	T	GCAACCT TTTTGTG GAAGCTA AGACCAC GGAAGGC CTCGACG AGCTTCG AAACAAG ATAGCCC CTCTCAC AAGCGAT CTCTTTG	TAAATATTC TGCAAATTT ATGACCTGT CAAGAGGA CCTCCTCAT GACATGTA CTGAATAA AGTTTTCTA TTGAAAAA AAAAA	MACEWDGKNSLLKLGSENQPSVSDVSISAMQAPPVAQS PVVSTARLSTLALPYAQVILNVPTAKASTRHPASPVHPT SKPKKENMDSHNIWITNIRGLVSNFDSFRLEIEHAHPYL VIVTETFLSPDVPDQALLIPGYSFLRRDRRTHGGGVALY YRSSLRLRRLTSLETQDHEVIWFQASTREHDIIGCAAYW PPGSDLDITQHLQHALDEVSIEFNTIVLLAGDLNAHHPD LSESISTNTAGRLLQRFATVNALTQVIQAPTRVTAHSSSC LDILLTNIPEEINFLGVYPGIGTSDHGVACVDLKTTSVITH TEDSATIRPRLHFNTHSTNWKTVNDHFSNINWHHRFAA CNVEETWTLFKDAYEEVLGRLSKKSRPTSRRTNHERHQ HNTLNQELIRLRATQQAAWRLHSTARTRETHQSLVTAR QEFTLACANHIRRTDEATVKRMTESSNAKQWHKYCKQ	cd016 50	587	851
CR1	CR1- 10_C Te	.	Capitella teleta								

Daphne	Daphne-4_CTe	.	Capitellateleta	GGGAAGC			LYRGSAVKEAIPIMTVGTESYITAKDKATILNQTFIAKSR			
				GGCTTCA			ASSRPRFPSLKKRTDSTLADILFNEYRVKKILQDLNINKA			
				TTGCGCC			SGPDGIQPSTLKNCSDSLHVPLAILFTKSFTEGVVPTEWK			
				TGATGGA			LAAVVPVYKNKGQRTDPNQYRPISLTSCVGKVMEKIIN			
				CGACAGC			DALLQHLMGNELIVDNQFGFLPHHSTTDQIALLLLHELLA			
				AAGACCA			ATTTKKSVVACFLDLAAAFDSVPHPAILQKLPAYGVRG			
				GGCCAAA			RLFRWIESYLSFRTQYVVVDQSSSPIAQVSSGVPQGSVIA			
				ACACGTC			PTLFLLFINDLCHSLSANTSLSHPDCHPEALVYADDTML			
				ATCATCA			YRTGKDPALICPAIEEDLRIAAKWADIWGMRFNASKTV			
				GAAATGT			AMYISRTTATPPPITFAGATLEYSHEHRHLGFILDSALSF			
				TCCTGCC			HAHVNALTRKGATEVFLLRRLSYKVKDRDLLLKIYKM			
				CACTATG			YVRPHLEYASPAWAALTVTQTGLLESLQRRAMRIILALP			
				ACGCGGA			RNHHITDADYAALNITSLRHRRNFALACFMFKLNSNRLP			
				CACCATC			RKLSKYRPIPHTNPYNTRNCVLRAPYVPYPCLRLLDRSP			
				CTGGAGG			FLFGLKILNFISHTFHSLPDISSFKSSLRESSYYMPLAQIY			
				AGGCCCG			Y			
				AGTCGAA						
				TACAGCA						
				GCGCCTC						
				ATCCGTC						
				ATTAACT						
				TGACCAA						
				GCCAGGG						
				AGCACCG						
				CCAGAAG						
				GAGGCCC						
				ATAAAGC						
				TGAGCCT						
				TGACGAC						
				AGCAATC						
				AATTGGA						
				GCGACTT						
				CTCAACC						
				CGATGTA	TAAAATCCT	MKNSIKMKNKLYCKYLKTKNPNDQKQYKSYKITLRSIL				
				CTCTTTGC	TCACAATCC	YKAERTHYDQLFCRNKNNLRKSWEVIKTVIHKQKNNN				
				GAATTCT	CCCCTGCTA	KQQYNMEVNGKKTSDLDIITKHFNEYFINVGPNLSKSIP				
				TTTCTGCC	AAATTATCT	YSNIDPISYITQTNRYSMFTKPTNETEIGNIIRALKNSAPG				
				AACAATC	ATCCCTAAC	PDGIPSSILKENTAHFIQPLTHILNLSLSQGIVPNEMKIAEI				
				AACAAGC	AGCTCATCT	KPLFKSGNKHLVNNYRPISLLPAFSKIFEKCMASRLIEFTT			cd01650	195
				CAACTAG	CCTAACGCT	KHNLLYQYQFGFREKHSTNMALHLLIDKITNAFDQNKV				449

				AATCACA GGGTCGT CTGCGAC TATTATT GATAACA TTTTCATC AATAATC TAACAAG AGCATCA ACTACAT CCGGAAT CCTTGCC AACTCAA TCTCTGA TCACTGT CCTGTAT TCTGCTTC ACACCAT TCCACAA TAATAAG AGCAACC AAAATAA CTTCATC ACAAAAG GCTATTC ACTGACC GAAATAA AAATAAA TTCCACG ACCTCAT AAGCTCC ACGCAAT GGGATCA AGTACTC AATGAAG CAAACTG CCGAACA GCATTCA CATCATT CTACACA	TTTGTGAAT ATCCTGTTT GTTAGTTTG TTTATTCGT TGTTTGTTT GTAAAATT GTAAATATC CATTGCGG GCACATAC CCATGTTCT GGGTCAGC CCCTCTCAC TCCACTTGT GTTTTTTTC TTTTCTTTT TGAATAAA CACATTTTT GAATTTGA ATTGA	LVGVALDFRKAFDTIDPSILISKLEKYGIRGCTLDWFRDY LHNRKQYVHLEGIKSSPLSIKCGVPQGSILGPLLFLLYIN DLPQSSQLLPIIFADDTNVFYSGDDSEECIGIINQEMINLI KWIRSNKLSLNVEKTQYIVFSKRKIAPDNNDVNINGERL NRVDSFKFLGVIIDDRLSWKQHIAYIKTKVSKSIGILRQA RPVLNKTTLLTVYYSFIYPYLNYCIDVWGHCSKQDLQSL FIVQKRAIRIINYAERNAHTDQLFKSLEILPLHHLYIFSVII FMHKFHHKRLPPVIDELFQTNASIHSINTRQRHQLHTPLL HTNISMKSIRYRGAFIWNQTFNEIDTNLTINRFKTCLKNK LFTDFPIRTSPIQ			

				CAATTCA CCACCAT TTTTGATC AAAGTTT TCCATTA ACCAAAA TAAAACT TGGATAT AAAAATC GAAAACC ATGGCTA ACAGATG CA				
				TCCTTGA	TAACGAGC	MVPDVNPSFSIRAEDCRYIAPXEIPNLCEHYTHPLRILHV			
				CACCAGC	AGACGACG	NIRSLRANYDELLNLIASYEESNRALHVLLLCETALSDA			
				AACAAGG	CCACAGCA	NTVLFGIPGFDAVYANRKSHQRGGVAIYVRKGLNFTQR			
				TCGTGTC	CCAGTAGC	PDLSIFSEGKFETIFIELSLCQNSQPIIIGEIYRVPSLNAPETI			
				TATCACC	GGCCAACT	AFFDSFTSLLHRSNASIIFGTDQNLDLGKIDEHALTRNVY			
				AGTGAGA	GAGTGTTAC	NLLESEGFLPTVNVPTRIEGVTKTIIDNTYIRGSLPSARTS			
				ACGACCT	ACAATGGT	VVEVKIADHLPIVLTVDFFHMHRKKNSPLVIMHRKMDET			
				AAGGGAC	ATGACGCT	IAQRIVDELGQCDWQSNENRDIDSAFEFFLSEMTRVLDR			
				TACTGAC	AATTATAAC	HAPAKQIKIPSKSFRREPWFTKGLQRSSRRLSKLYSTAA			
				ACAGATG	GCCCACAC	KCAKNSHQHTAYVKYRNVYNSAKRAAKQLYYDNLFSE			
				ACTGCCT	CCCACCTCT	SKNDIKRTWKILNDLTKKSGKDNSPVEKLQVDDVTITEP			
				AATATCA	CTTCCAAAA	KAIADAFARFFADVGPKQAGLIASHVQTDDQQNRAHLG			
				CGGACTC	AAGCAATG	VPPSHHSVFLHPVTQQEIIRIVDSLKNKKSLGLDLVSTHF			
				CGATCTC	TTTTACCTT	VKKLKHALSAPLSALVNKSFCEGAFPNLLKRAKTIPIYK			
				TCACAAT	TGCACAGC	KGERFNMDNYRPISLLLPIFGKIFEKAYCNRLMPFLEDRG			
				CATTCAA	CTGATCAAT	FLCDSQYGFRRKRNTTYAVSELYLNSINALLKNEITLAT			
				CCTCTTTC	ATCGAAATT	FIDFSKAFDTISHEILLNKMPSYGIRGRALDWIRSYLSNR			
				TGTTCTTG	GCCAAATTT	SMKVSIQGCLSEEVQLSQYGVPQGSILGPILFIIYANDLN			
				TTCTTGCT	ATGCACCTG	ACLTHTKAIFYADDTTLFLSGKSANVLAERMNDDLGRA			
				ATGTTAC	CACTCAAA	KRWCEANSLAVNIKKTKYMLIGPNALTRCTIDLSMANE			
				AACCCCC	AAATAGTTT	KIERVTTFKFLGITIDERLTWKAHIESLSKKLSSGLYALR			
				ATCCCTA	ACCACCTCG	NTKRLTSRHGLLSLYHSLIHSHLTYGCLLWGSASDNAL			
				CCTTTACT	ACTCTCTGC	KPLKIAQKKAIRLVYDAIV			
				TGCACAA	TATCCTAGA				
				AGCACCT	CTCACATGC				
				ATCTGCA	TCAACCTCA				
				CGTCTTC	TCAGTACGC				
Dap hne	Daph ne-15_CTe	.	Capitella teleta				cd01650	543	800

				CACTCTGGTGAGTTCCTATCCAACTTTCTTCATTCCTACCCATCATGCGAGTATCTCATACCTTCGCCATACCTTACCTACCTGTATCTGACCCTAGCTGCGCGAACAGAAACCTCACTACGAA	CACACGCTCCATCCTACCATACCGTCCTCCCAGAATAACCTCTCTAGCCACTGCAAAAAGTTTTTTATCTGAAGGTATCAGAATCTGGTCCTCCCTCAACCCTGAAACTCGCGCACTTCCCCCTCGCCGCTTCTCCAAGCAACTAAAACGAAAATATCTGAATGAACTCACCACCTCTCTCTAACTTACACATGTCCATCACCATTTTTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTACTCCTCCATCTACATTGTTTTTGTTTTTTTTTCTTTTTGCTTGACAAATTCCCCAAAAGTTTCAATCTGCTTTAT				

									1074
								786	
							cd01650		
Randl	RandI-4	.	Chlamydomonas reinhardtii			TTGGCCCCCC	KLGPTPAALAQDPPTNPTGGPPAGPNTTNSPSPPTSANP		
						TGGCCCCGT	GPAARSNTATRIVPKPAGPTKRPARRTPNPTPKLRKLAT		
						ACAGACGT	GTPSAPPPAKRLHQTTIATSHKRRHTATLQNTPTEPTNG		
						TTGTGTCTC	QAQASPSPPPDGRISILTLNTRELLSHHESILHTVTTHTPD		
						GCCGGGGGG	ILFLTETKIPTRLHYSMGSAKKGLSSHYQTYTSNAPYNP		
						GCCTCAGGT	REGTRPHAGVMVGVTHRFSRPSGTTDMNAHNTDSRLA		
						CACGTTTAA	GHVCHIRIAPPHSTPVHAIGIYAPSDDATRAAIYKYCTDQ		
						TGTCCAAAT	AAKTTAAKEHLLIGGDWNAVLFPSDRSTGQLNATDRA		
						TTATAATAC	HIAFTTSAGLHPLTPPTHEHPRPHTYLQDQATAPHMSRI		
						CCTGTAATA	DDLLVSHHTHATITTTGGGQEAVHECGDALDHRALLSS		
						TTCTGCCCT	VPLAALRLYPFPTAPAAPSDPDAWVNYHPPTKDQREKA		
						GAAATAAA	RIKMEETIGQEIADLLNTLHPAHLATINKLYPAGLTPAQ		
						CATTCATTC	AERNTPSPAEFRALPEVSAADTTSISNTLESILQKCLDVL		
						ATTCATTCA	KSYCDKKPPRASTTSPRHHTSKTVTRLLTATSKLKERLT		
						TTCA	AALTVTNAAGATPPTTSPHPAPTPPPNTTSPTPTPPTPAE		
					TAACATCCT	ADTPTAIRSQLLNARTLLRAAKKEARKLNTKRAARAFR			
					GGCGGGGGT	RLVATQPKRAHQKIFDETDQPRGTPALTDPTTKEVHTAP			
					GCAGGCCT	EKILSSIHQHFTAQLLPINGPRTGHYDIPHSEKQYPWSRP			
					GGCCTGCA	DAPDRMPALSKPTDAPATLLPGIADSTTFFGCVNHLSRR			
					CGCCCGCCT	KAPGPDGIPNELLQTIPTSLQTAIHQLMTLMWVKAHIPT			
					TCACATGCT	HWTQSDTVLLPKKGDPLLIQNKRPIALANTLYKLYTSLI			
					CGGTAGTG	TVNAALFSEQAHIFTESQEGFLRGRNTERQVQNLLHAIE			
					GGGGCGTA	DAATTGRDLFLLYIDFTSAFNTIDHDKLLIVMHDLGYPT			
					CAGGAGGC	DLIEVVRDLYGKAATSVRTEHGTTPPIPIQRGTVQGDVL			
					GGGCCTGG	SPLLFIIFIEPLLRWLHVGGRGYQYGCLTKAENDKHNLS			
					CCCACCTCC				
					TCATCTCCC				
					CCCACTCCC				
					TTAATCCAC				
					ACCACTTGC				
					CTACGGCA				
					GGGTGTCTC				
					TATTTCTAT				
					TTCTATTTC				

SGAFADDLVTATTQISNLRIQVEKIEKYAAWGGLKVNV						
SKCQLQGILHGRARADPTLGGPRSKPHQAALARMLEGQ						
IRIDGTAIPYKPPHESYKYLGIHITPTLDWSAQVQYIREKI						
KKRGQRIATSLATPDQKLRMIQLTIHPAVAYSLPFMPYS						
THDVTSLDSLIASTAKMCYGLKRGFPTRAILQPTNALGL						
GVGSIRELYVRRAGKALVQALNDGGRLGAVTRAMLHL						
QHSVAQKIPADQLRTQCRFYTSLRQLSLITDAGLELTQQ						
GSAYHTTLHGLSALLARTVPPGADAEQALPFRLIQPLLE						
LTSDFRELLTPDRTCVIDANTLKNTWPRAQRRHRLALN						
RLTLALHAETRTRAKTHNSTAPLTTAQRTLPDGHTFTGL						
HTAHTAPAAVDWGTLNRFLRQRAPATGQTANAGETAA						
ATTAAATTAAAEPEEDHHAGENHAPAQTRRPGTQQRK						
RACIKMVREPEGRGSPPADAPAAAFDWDPSLDPLTEAHC						
ADLIHTAPSLLGKSKPQNPAAHLRGLLLGPTPPATFTHR						
HKLSQEEQDAYLAATSDTWKHFSETYRTKALTTTNTLY						
GQQERVQQVIAHTTSKGAPFYRVQWEPTVTPGSLTLY						
ARNRYRPKPGGLTWLLTPGNEPRVLVEWEDLEEPAAIL						
DAQDNKQELLDALNARLAVPLGPRPPPPPATDTHLSHDQ						
RQGRWTDASNDTARGQRTLRHYVDINVEPCNPSRDIHP						
TQAYTVQLGTVLDPGSPACGTHLAHVYDPRGRCIGTISA						
EQLSLLWQRYQHAQTTWPKVHAKHDQGRGFAAAIARL						
LLRYKPGGGSKSHPVTEKGECMLCEGLTQTLHTAIGSTT						
ELFASPLNVSDAATTYYTLYEEDMLFGANYDAYSQPW						
DGVAVACPEPNTEDGAHAVRWALALAWAEHTETLTNTRN						
PPTTATLLILPTQRNAPYTRYLAHPRVTLLGSISATTTHA						
LLGQIHNDTGITLKHDLVLVSNTPGAQLAQPIRDALGTS						
LYLSGMPVDPQPQLTWAHTRWVTKQVRRMTRTRPAAP						
RVTPPPVPATPATLAPFAATQPLRLNPYEGFYTDGSVQK						
DALGQPSAWGAAAIDARTLRCYQIDVNPGQHGCNTITR						
CELSAIHQVLKLDEVQALPEVTVYTDSAASLFLIRKTINQ						
PQLLTFSKHKALLDNILALLKRRSHAGRRTNLLKVKAH						
SGIAHNELADQAAVKTAKGEQIPGAEWLVEESDNDPRS						
KYYWIRKIIKPTVEGVPREKNEGHYAHDLNRGIATLAAP						
HIAQGHTRETLYTGFVANAIPDLDIKSSTRTMNLQSGRH						
RKAKRCMDYLYGQLWNNKLAARFGRTVTGRRAPRAG						
PDTPAPCPLCRGPDSGGHILGGCHNPQLRAMYIKRHNIA						
VQTIARAISRGDCGGCYMVMDATSKADLPDYAADNRL						
PPWLLPDQEGKPAGRHLHYRPDILLLIPSISLAAALNPDFV						
VLPSEHDTIHIIEAGYTADTNHAAKQHEKAQQHQALAA						
DLREAGWKVQYTPQSAISLGFAGTIRKDLHPLLTSLGVT						

L1	L1-6_CR	.	Chlamydomonas reinhardtii	GGCGGCG	TAATTTGCT	SHKARQCCDTLHDHAVATLNHIVLTRRRLERGLPPGDP GGT			
				GCGGAGG	CGTGGGTG	QWGWWRRRWRRHAAGGGRPCRCAPRPRPHRGDLGC			
				CGGCGGC	GTGGGAGC	PAGGSAVVRRPRQQAAGPRPLRVGSHNVRGLRSRVQH			
				GGCGGCG	GGTGTGCCT	LGCPRLHALLRCWERQQLDVVCLQETHVRAEDIDLVTA			
				GCTGCCC	CTGGTGTGT	DLRAAALRLHVREWTAWWAPAPTPHQRVSGGTAILIRT			
				TGGCGGC	GCGCTGTG	CLLEQGAVTFPGGVAAPSVADGAWVGRGVSLPVQWG			
				GGCGGGT	GGTGTGGG	GHDFTLAAAYYPLASPEQRAFTQQWVASAVGGEGNHL			
				GCGGCGG	GCGTCGAC	LAADFNFVSDVALDTASGVPRPGDPAAATALAAACPGL			
				CGCCTGG	GTGGGCGT	IDVLRVRRPGQRVCTYIHSHGASRLDRILCSGSLVAQVL			
				TGGTGCG	CGCCGGAC	ACAVEAGVPSDHRLVTLTLAATPEALPPSPGLPKARLGF			
				GCACCCG	TACGTGCCC	LDFKDLTKDYRTWLTAALGARPGAPAELLAWWPQLKR			
				GTGCCGC	GGTGTACCT	AAAASAARLSRVAAVRRIAASEEEARALAAAAAAAAA			
				GGAGGCC	GCGTTGAC	VEVGAALPEAAAAAAVRARCAAATAAVQAAAGAARRT			
				CCCGCTG	GCCTCGCCG	RHEWLRGGEKPCPVLTRMLRPSGGPRTIARLKRPDGSA			
				CCCCGGC	GTTTGATTG	TSDPTVMGQVMATYWQDVCAAPPPAPAARAQVLAAL			
				CGCCGGG	CGCCTCGTG	QQHGRRCTTEEAAQLGQLAVTAAEVRAALAAAPAGRA			
				CCGGCGA	CGCTAGTGC	AGPDGIPVELYQYYESDFSPVLADVFTAVGEGAGVPGG			
				TGGCACC	GGTTTGCTG	FLDGVWTFFYKAGDPLEPSNYRPITLTDTDYRTLARVLC			
				CGGTGGC	TACGTTGGT	RRLQPVFGRVIDPEQTAFLSDRRIADNVLLLQLTPGLLK			
				GCCGCCG	GGGGCTTTG	AAGQNSGVAAFLDFYKAYDTVDRSFLLACLERMGVGA			
				CGGCCAC	GCGGGCTT	GFLTWVSRLLTDTRGAALVNGRVSGWVPLAAGVRQGC			
				CGCCCCG	GGGCTTAG	PLAPLLYLAIAQALLSWLRSKGHGVLAGLASALASQYA			
				GCAGTGG	CTTCGACTC	DDCTPFLAGMQAVPGFLADMDVFRCASGQRLNMSKVE			
				CGACCGA	GCTGTGCA	LLVLGTSAGLGVAPAAVAGPPLPPGWRVVNTAKSLGV			
				CGCGGCG	GCTGCCGA	HFGDWKTATPPVAFDGVVGVLGLIVRMPVSMFGRAAA			
				CCTGCGG	GACACGCTT	ASAYALSKVLYHLEFAGFPSPWAPVGKLLGQVAAVVD			
				CCGCTGC	TACGCGAA	RRLSPAKYAVAPHARPVGVSLDIMQVPPALGGVGLLPL			
				AGCCGCG	GACCTGTTG	QLHVQARQAGLAVRCVLGACGLLAFVPPWTKVAEAW			
				GTGACGC	CACGGTTGT	LRARHPAATPLRLLVRCDPSRVLDKLVTPECKPFLRLLG			
				TGCCGCC	GGAAGGTC	AVAHLPPITLVAPPAAGDWVYHVPLWANPALCDGAQT			
				CGCGTCC	GGTAGCCA	FDVAFADLFALPGLVYVGQLVAAHDQLEAVRAAHTRP			
				GCCCCTC	ACTGGGGG	GGWGTRFGEAMLDLYLTSVWRPVLGWGPRRHLPSPLL			
				TGGCTCT	GAGGGTGG	GPASPLAAAERIAAAVALLPGGWEAAARAARLARGRA			
				GGCGCTG	TTTCCTTTC	WVMPLPADAVRQAAAGALTQVVRGLGWVLDGGPPLL			
				GCGCCCG	GGGCCGGA	LLNYSVKAGTAMQLGPQVAALRSKHLLYVRDAGVSAD			
				GCGCTGG	GATGGTCCC	GALAAAGAFAGTLARLWALKWENYHKEALWRVAAN			
				TGCCGGT	TCTTGACCT	ACWSFPKNAAQRAAGVAMPLCRACGRDMCDGDRRHY			
				GCCGCCG	GGGTGCAC	FWDCVVARALRENMAMAMLGSPPDTADDCFSRADLW	cd01650	604	863
						LVQPPAGLAPPVWDVVCLAALSALDSGRQQVIMTGLR			

				CATCTGCGGCGGCCGCCACGGATGGCAGCGGCAGCGGAACTGGCAGCGGCAGCGGCAGCGGCCACGGCGGCGGCGGCGGAGTGGCGGCGGCGGCGACGGCAGCTGCGGCGGCCGCGGCCATGACAATGGGGATGGTGGGCGGCGGCGGCAGTGGCGGCGGCGGCGACGGCAGCTGCGGCGGCCGCGGCCATGA	GGGCACCCTTGCTTGGTCGGATTGTCCCTTCTTTGTAATATCGCAACCATAA	VRRALTSAEVMRISVSVIADFWGRLQSYVSLGLRPKAWATVPAHHPFLSRGAGDCVLLVLPDDVGSPPPSP			
RandI	RandI-2	.	Chlamydomonas reinhardtii		TGAGCCACGGCTCCCCCACACCTCCTCGGAGTGCGTGGTGAGGTGGCAGCGCGAGGGGGATGGGCCTCCTCGCCCTAGCCACCTTACCCCGATAATCCACG	MSSNPMTLQIGTQPGQQQPDLTSPSHPLRGGNNTERTTTPAAENGHIPVPLSWPRQIPSTYFREQQHLNFCTIHAINNMLGAPWLRAHAVLAFCQVVHTALADTPELQATWGHHFKHNGLFHEAAINHYLRHRPAPIDTTGTRLLLRLAAPNLDPGHSQDTILGLLPECSRNMGFMVHDGRHAYAIRKEEDSWFLVESFVHIGRKPMTPDLWQAVKGDIFCLHPYHLDMPLDPRTYPTIPNLPPPPQPPSPPPPNQTKPKRGPTQRALSTMFARQTTQLPTNHSAGDNAAPTPHPGPAETTVTRPLTQATLNITRREATEPARTQNERPATPNSAARPTRRPRHDNQLDPEPTSPTRANTTTRPRANANMHAPNTNPTQPTANNTTQPTWAQVVATHAEAAGNALRTVGRITTQAPHQRPETPPSAARPTQRPRADPPDSPTTHQTVNPNTPRRTPPGPTPDPGA	cd01650	1154	1442

					CCAGGCCC CATGGGCC GGCGTTTGA ATTATTATT ATTATT	GAQDNTETRPDNTTWLHIATHNVRGLGTNLLSTICHLR EWNADIVILTETKLGPRTRWLKDTLRQEKLSYRTFTSTK PGTEQYKRRSAGVLMAISKRYHSGSNVHTATIPPHLHG YIVHCTIRTPHSTPLHILGVYAPEDMPTRRRIYKYCEETI NTANKTDAHVIIGGDFNAVLNEGDRTGPPDEADKHHAR FVSKNKLKRLPEFGGTISMSFYQTRDGTPQPTSRIDDILIC QNIHSALPTASRTPATLQVQPCGGLFDHYSLHARVPTFPI RLWPSHTATQADTQPPETPGPWRTVVLPIPKSTLEAART KIEHTLANPIATLASAIQPATTAIQQHLAQHTAGATNASE LQHALQTDPAIQTPDLDALANQLLEILQDGLSILVDMCP KKPPTRGVFAPRKINKRIRESHNEIKQLRTQCDALLQQQ HPYANITTTTAPNTGIPLPDPTTEQDAENPPPQQPPPAHD ATTTAEPQRRPRPPPATPAIANLIDKIKRAGDTLRQILKK QTQHDIQKARDRFQHKLATRAKQAHKDIFSTQEQSRGIP ALIHPQNKAHCTDKASILDAIHAYFQGQGASITGPRSGK YKPEDRTQVYPFDPSLKNSAPDTFTLHSATTPPTHALLPS IADSTNFDSCIDHLSRNKSPGPDGIPNELLRALPQTLRTT LHNLLTLMWVRAHTPASWTASETVLLPKPGDPLLLKNY RPIALANTTYKLWTSLITVAISDISHDLDLFSETQEGFLR YRNTERQTLNLVHALEDAGLTKQDIYVMYADFSSAFNT ISHDRLLEIMWDLGMPHDLIRVVQNLYSNARTYIRSEHG LTAPIYIERGTVQGDTLSPVLFLVFIEPLIRWLHVGGRGY SYGCLPPHLNARYNCSSAAYADDLAVLTNTLSDLRIQC DKIHRYSEWAGLQVNHAKCRVTGILHRRAMQDKGLNG PTCNRTLRSMLENKIHIGTTSIPYLPATEPFKYLGVQVTM TLNWGPQFAYLCKNITDKSLSLQTCQASPEQCLRIIKSCI QSVAAYSFAVMPYTENDIRTLDNMIARLARTCYRLSRG FPTRAVLTPAEMSGLGVGSLLPIYVQKATRALTLALNDP GRLGMVTHALLELQTRIAGTTDMHKLRHESHYYTTLK QLSIMRAYNLTLWDHGAPYKGPPNTLLTAISRSPHPISH HLVLTLTQAGLDLRNLVGKERTPEGRACLIPSDILFQHL PRNQCQHKHRVALNRLSLCLSAECDGAREAPFNQAESY KCHLPLPQRYRALPNTPAFAALLPRNATPGELTTQSTMD AHLQHMTALHPNTEAQTPQHPNGLGPRQRQTTLAHML RTTGAPPTAPTQSPTATLTQTPTGTLPTPPLAQPMPPPLT QPTPPLTTPAGPDAPNTQTPPTVASLTTAHQQPKSKKQR PNAIKRLQQKIDADPRHDPLPPTDMRSHTEKSGFILNND PDENDSGPPNFVMWAPHYFANVQKDMQGTNRMPRLGI PRHHSRFPQNRTRPNPTRTIPAHLADPYFLGSQVDDTNP LNPVTAYISANPPTTPPTCAYDSRPRWADTGDFMIAPEP			

						PNWPNITGATPGDSHDPRQADHDPQGQMRAEYMQERERLRAHLDDHPPQPRPFKRWMGIRLQLLTGKLPWPTQAYHVAQAAKQLVYSLYHDIEEISHIQAGPITKAVTAPEKGASLHTERYYEVAWKDTTIVKGAAEAYTRMKYTGTAAAHMGYGQACHKYFLRMKWDPTSEPVDKCPTLQGWTPAISEHLRKLAEPPTPTPQPKPKDARLTPAQQQGRWTQADVAASHVRRTTKQNTRIDTNPCNPYLDVHPTGRYTIQIPDTYTDERTLAHVYNPTGRHIASVTLGTLLTLKARYEHALSTHHDATSPHRSQGGFPEDLAKLLLRTRVDSKGKTCKSAPARERSLPESLKLALYNCLASTHSDPDRPVGLVERFASPLNCSGRATKYYSQHAADSLFGASHDAYTAPFDGLSLAHPDPDPDACTKALAWATASIEHLEHQATPAITFMIIPQGKGAASRCLASPHVTILALIPPGVSASSMLDGLGAADEAASSKQGLQLLLIANTTGVQALNSNNNLTRLQTSLAELGAVPPPTSQPPGRRQPLEWPARVTLPRDRTAPPGRPYSHRAKARAKQLTRPQAPAARGLDTCPTPMPTLSALYTTRAAFPTRSSYLYDAATIVWTDGSCIKVAAGEDGNEMNKLGACAWSQQTEHIYVEPGGLNSTNTIQRAELSAILVALEYIRPRLCLLGMTQVTIASDSLVSLYLIRRALTAPQRLLLSKHKNLLENILHELQMYADAGVPVQLLKVRSHTGLHGNEEADKGAAKVAANDMEAMGQAIRLAPDEPWSNIWWPKRNSDNFYLSDLNRGVLNSLPPEAQQGFTNNTPVLEQWAEASANMCPTLSNRTLASSTRHPWLVKQILYARYGYLFNAQLKRRYRLGGNGNCPLCHTPDSGGHILGGCTHPKMKGAYINRHNWAVQRVAKCISKGTYGNGFMVMDAGPLADLPHYASAQRPPAWLAAERTSSPLITRLRPDILFIPNLPAHETHREGYRPPLNKARLPIYILEIGYSSDVRHLDKREEKSQQHQELKRLLEAAGWTVHYSTNECLALGVTGAIPCNLTKLLTSQLGVTHTAARVCGERLHTHAVNSAGGIIKMRRVTEQELKPRGG			
Daphne	Daphne-1_CSav	.	Ciona savignyi	TG	TAAATATTATACCGTAACTGTCCAATGCCTTAGCAATATTGTAACTTCATTAATTTAGATTTTCCATACCTCGATCTCTTTAACACATT	RFFEAEXFLSDLYXNFCNLLSHKPPEYSGDCLLQNFFSVFSEIVDKHAPLRSISRKKRRLLSKPWITKGLLQSIKNKNSMFSSIVKNKPNLQKYSVFKVYRNKLNHLLEISKQMYYKHEFSRSKNDIGRSWKLIRELIQIKSPTMTISKSVGITASDVTSDPINVANKFNDFFVDIGKNLASKIKPPTNHVPLKPFRKSSSFFISPSTPDEVFSLINCLDVKKGTRHEDIPTKFIKLANSIISPILSEIFNTCVKQGIYPDFLKNAEVIPIFKNGSDNLCTNYRPISLLSQFSKIFEKLLYNRLIKYLDKHKLLSNHQYGFRENCSTTLAVADITNMLLSNTDKGLYSCGIFIDLSKAFDTVDHSKLLAKLEHCYGIRGVALDLFKNYLTNRSQYTFVH	cd01650	266	523

R2	YUR E-2_Cis	.	Ciona savignyi	CCAAAAT	GTAAATCA	NVRSSIKPVLCGVPQGSTLGPLLFLMYINDLPLVSNFVT					
				TACTTCC	CTGTTATAT	KLFADDITFLCMSSKCPYELQAKVNLELIAIDDWLRSNK					
				AGCACCT	ATTCATTAT	LSLNYSKTSYVLFTGNLKPIGINDFHLNINNNKIIRSQSA					
				CCACAGC	GCATTTTGT	KYLGVLLDEKLNWNCHILNLSQKLARYCGIMFKLRKY					
				ACGAGCAC	TTGTTTTCT	VDIKTLLNIYYSLVYSQIQYAIVCWGASNKKTLYPLNVI					
				CTGCCCATT	TTTTTTAGG	HNKIIRAMSYSPARCHVTPLFSDLKILKLCDIYKLEMGK					
					GTTTTGTTA	LMYKFHNKTLPDSFNYIFQPLYCAHNHITRQVAHHDYF					
					AAGTTTTTT	VPYFHTLLGQNSIGYRGPVAWTAIDVSLKKLGWFSFKK					
					TGTCATTGT	QLKNQYLLTYIT					
					AACGTTTTT						
					GCACAAAT						
					TTTTCACAC						
					TGCTATTCT						
					GTTTATTTC						
					TTATTTATT						
					TATCTTTAT						
					ATTTTAATT						
					TACATTTAT						
					TATAAGGT						
					GGCACCTG						
					CTTCGATAA						
					CCGCATGGT						
					TTCTCGGTG						
					TCTCCTTGT						
					TAAACTTCA						
					GTTTAATTT						
					ATTTTCGTA						
					TTGTTTATT						
					ATTTTGTTT						
					TCTGTGCGT						
					TTAACAAAT						
					AAAATATG						
					TTTGATTGA						
					TTGATTGAT						
					TG						
				TAAGATCC		MAGHKITMSEGKLLEVAVRYGGVRNVSYECPVPDCTK					
				GCGGCTGT		TFSQANNLIRHLNNFGNTKHRAHNFTYFFTCEKCKIQIH					
				GGCGCCGA		SNTKHNISNHYKQCCATGGGPSCETGQYFCPACEQAGL				cd01650	710
				ACGAGCAC		GNLESALRHFQSSHPEFNLPPRSQFSKSHPNSYTLSLKPK					
				CTGCCCATT		DHLMKILYSGPLTPGQLVCPIKICLRSSAARLFHDVSKLR					

CGAAGAA	CTTCTTGTA	KHMLVDHNRTLVYETTCGKCLRPVDTSKNMRKTTSHF	
AGAAGAC	GGGACTTTT	EKCSGESFISSPSPIPQKTYKLDLPSTSTPPPRKSPKLQPY	
TTACGAA	TCACCCTCA	KPIRTFKNPLTKSSQSKSDNPPKPTPFFSPRTLERSASWP	
ATAAGAA	CTCCCCCCA	ALSEVVDPLPKLKEKHPSLPCALDKCPPSPRIKPSTLVPP	
TAAGCAG	ATAGTTTTT	CHTANNSPKPTSPESPSTLKPLPRPIRPSKPLEDWLTVRS	
TTTAAAG	TTTCGTTTT	VGPDREIVLNIGPRPRPGPAAGSRTTSPPSTAPAKRVAAN	
ACGAAGC	TTCGTTTTT	PIAAPLSGEPGATLDCGQTGRKVQPPKKRPTESAGSLPPP	
AGACGAA	TCACCCCCA	AEPATDLLTGREGLARLVEEYHLSGDFGAFCRDLERWT	
CACCACT	CCCCACTCG	ALSSTNRRPKPRRGRYNRGAAARATRNRGRDDRQDPQ	
CCACCAA	CCTCTGGGC	DRDDQGGPGPVTCGRPQRYKRAAALRSAFGRDMKATV	
CGACGCT	TGCACATCC	RRIIDGERGDARCEIDPKTIEGRFRDELSPPVREGPECSLP	
CTGCAGC	CACACGTA	PWMAEAQAGEHAPSNDSQPGDAYDGPITALEVEMVLS	
TACACCA	GGGACCTG	TLNVGSAPGSDGLSYGFWRALDPKGLVLSELFEVCRIER	
CCACCAT	TTTATATTA	RVPGPWKSSRVTLICKDAEGDLDDLGNWRPISICQTVY	
CTCGCCG	TTTGCCTTT	KIYAAVLARRLQSWALDGGVISRSQKGFMPFEGVYEHV	
CAACAAG	TATATGTAC	FLLDSVVADARATRRSLAVCWLDLRNAFGSVDHTTIVE	
AAGAATT	CACTTTTTA	ALSRFGAPAGLVEMISDIYTGGSCRIRTRAGFTPDIPVGR	
TTCCGCT	AATATATTT	GVRQGCPLSGIIFNLVMEVLLRGVEANNACGYRLSCAG	
GCTCTCC	TTGTACCCC	GASVRVLAYADDVALVGSSRAEXKIQLGVCERFAAWA	
ACTCTGC	ACAAGATG	GFSFNNKKCAAMVLKHQRGGRRLLDSAPLRLCGEEVAI	
TCCAACA	CTTTTCGCC	LGPDSFYKYLGAHTGYGRQTGGQLVDRVERQVVRLFT	
CCACCTC	AAAAAAAA	SFLTPTQKLSALKRIVLPAMSFHLRVRPCAEGHLRRLDN	
TCCTGCT	AAAATTTTT	TVRRCVKTALRLPKGSCRAFFHTSPDAGGLGITSVVAEC	
CTGTGGA	GTATCACAT	DILTVTQAFKMLSSPDHLVSLVAKGRLGMHAARMGRS	
CTGCTGC	TTTTATATT	ETASACAMADYLSGDSVMGHXSWKTGYRMPADLWTA	
CTTGCTG	TTGTAAAAC	TRAASRRLSLRFSPQPQGEFGLESGTFKIAPRERRSLTRR	
CTGGACC	ACAGATTTT	LHHRQNLWWRNQWAALPNQGKTVAAHSAYAASNNW	
AACCTCT	TATAAACTT	VKGPSSLAPQALFFGLKARLNQMPTRSVKACYSRAPNY	
ACCCGAA	TGCACTATT	DKSCRRCGAEVETLPHVLNHCPKSMKSILERHDSVLAE	
GGAACCC	TTTATATAA	VLAAIPRGTFASVDVDRTSREHFRRVGEALRPDIVARRH	
TTCGAAC	ACTTCGCAC	DGSVVVADVTCPFESCASALDTAAARKIEKYDQLCANL	
CCAGCAG	TTATTTAAA	RQLYRKPVESHALVVGSLGSWGRTNNTALAALGIRGA	
CAAGGTA	ATGAATCG	VRSRLAKQLVNLSVEGSHNIWLRWSGGIPKDLVR	
CGTGTCA	CATCTTTTT		
CCACCTC	TATATACAC		
TCCACAG	CAACACAA		
AGCCAAG	ACAGGATG		
GCCAGAG	TGCAGCTCA		
TGGATAG	GGGGAACC		
AGCAGCG	AATCCTGCG		

				GCCTCTA CAACCAA GTTCTCC ACTCCGA CGACAAA ACACCTG CCTCGTG GCCAGAG TCCTGCC GAAGAAG AATCTTC GACCCCA ACACCGT CTCGGCA GGGCTCC AAGCAAC GACATCA GCCACGA GTGCCCA CCGGAGT GGACGCG GCGAACC CGAGGAC CGCTGTC GCCAAGA ACAATAA CATCACC GCCTGCA GCAGCTA CCAAACA AAGGTTA GTCTCCT ACCTCAC CTACAAA CTCGTCT GAATAGA CGCCCCC GCGTGGG AGCTAAC CTTGGTA	TCCCTCCTA GCGGCGGG AGGGGCGC CCACCTACC CCCACGCTC CTCTTGAGC ACCAACAG GGACTCCCT TCCGGAGC CCCTGCACC CTCAACTTT TCTTATTTT TAAAAAAA AAAAATCA TATATATTG ATCTTGACG ACGGGGGC TACATTCAG CCCCCAAA AACCCACC CACCATCCC CAACGAGT GCCGGGGC ATTGAAGA GCTCCGGC ACAATTAG CACTTAGCT TATTTATTA TTTTTTGTC AACATTTTT GTTTTTTCA AATTTTTTC ACCCCTCAC CCCCACCCT AATAGGTC CCTCGGGCT TGGGCCCCT TTTTCGTGC TCGAGAAG CGTCACATC				

				GCTAGCT GCAGTGC CCTGCGA CAGCGGC CTCGAAG AGCTGCA GAGCGTC AGCCTGC TTCGACC TCGCTTG TTCTCATT TCAACCT ACTCCGT CCTGTGG AATCAAA AGAGCCC CACTAAC AATTACA TTCATAA AATCTAG CAAGAAC GAAGAGA AGCGACC AACTTTT AATCCAT AACTTTT AGATCTT TTTATTTA TTACTGTT TTTAAGC CCTAAGC ATATTGC CTTTTTTT AGATCTT ATAATTA TAAAAAT AGATTCA AAGTTAA CACCACC AGGCCGC	GCCCCACTG ACCACGAC CTTCCCCGA CATTGGAGT CCTTGGCGT CTCCCAGGT CGAAACAG TCCCAAGTG ATAGCACCT AATGCTCG ACTTGTTTC GGCCTGGG CCGCCGAG GATTCCCAG AACGACCA TTCTTCTAA ATAATATTT ATATTTCAG AATAAAAC TATATATAT ATCGTTGGC GGGACTTGT CCCGCCTCG ATACCGAG TGCTGCAG AGCGGCAA AATAAAGA AGAAACCG ACGTCGCTC TGCAGCCA AGGACCAC CCAAACTC AAGCCAGC ACCGTCGA CAACCAAC ATCCTCAAG TCGGCGGTT GCTGGAAC AACTCATA ACATCTTCA				

Pene lope	Penel ope-1_Co Fl	.	Copidoso ma floridanu m	TACAGAG CATTTTAT TTAATCA ATTTGTT ACCGACC TCCTGCT GCTTCTTT TTACTTTC TCCAGAC TACTACC CGGATAC AACCCTT GGAAACG AGAGGA	TCCAATG CTTTAAA ACGCTGG AAGGACC TGCACAT ACAACTA TCAACCA GT	ANATAAAT TATCACCCT GTGCAGCA GGAGGCCG TGCTTTTTAA AACTACTCT GTAGTGGCT CATGATAAT ATTTCGCTC CTTTTTTGC CCCGTGTAA ACTTAGTNG ATGCGAAT AAAATCAG TTGAATCA	TAAGCGCT GGCTTGTTG TTGTAATTT ATTGTCAGT CGATAGCA CGACAGGT GACGAGTA AGTCAACC AACTTAC GAATTCTCG CGTACCTGC GGTAACAA TTTACATCA AAATTTATT GTATTTTGT GTTTGAGCC TTCCAGATC ACTCATTGT GTGTTTTTA ATTTTAATT TTGTAGCCT GATGAGGC TGACGCAT ATTGTCAGC GAAACGTT	MAQRRFLLKCRYARVFPNHIKHLYNKCSSVNFRCKPIH ADFNKSTLHFCQKLLNFEISDINTHIRFLQNSISRLKDHIA SLKLDYNIRERHYKFVNIKQNTIYRRSNHTLHKKFKKIL PKTNTTTSNNDSIHNLTNIAIPKDVAEIASLGPNFNLPSTF RERDFLCIIKNLEVGLSHLDTNDRVKTEIRDKVSQYLVN ACKQKPAISRKDRQYAKNLKSTKNFLNKNPDIFFTSADK GNVTVCLKKSDYTERMHNLLSDKSTYKIISKNPLNSLQK RTSNILKNLNSNGFLSRTYHPNALSLTDTTLAKCYGLPKI HKVDAPLRPIVSLINSPTYFLAKVIYDELKKSVNVPISHIN NSFELKNKLKNIKPTTNYRLMSLDVSSLFTNIPLELVLRS IDRRIVTIMNKCRIPVDEIRQCTEFLFNNTFFTFDNIFYQQ IFGTPMGSPISPLFADLVMDDLENDCLSILKNNYNIIPSFY FRYVDDVLLFINVDHIDQILNIFNSYHPKLQFTYEMEIDK KLNFLDLTIIRNSNYFNFDWFEKPTSSGRILNYFSGHPVS YKRNIVFNLVDRAILLASTRFHGVNIARVKNLLIKNNYP FNFINKNIKVRLKRHKYSNHNRSTTNTYPLCKIGIPYIDN NFEKISTMLKPYNIHTYPLINNKLQHIVRLGKDTTNKHK QTNVVYKFYCKNCPASYVGETKRAAEVRISEHMKSKNS DSVVCAHQINFDHEFDWDNFKILDKESNYKKRIISEMLY IKLDKHNINRKEDVLSFNRAYLPLLRKFKCPDRVHRPLP STFAAPPI	SSF5 6672	363	534

Penelope	Penelope-7_Co Fl	.	Copidosoma floridanum	T	AACTGTATT	KRYIHLLLK	SSF56672	34	146
					CTAATTTGT	NIIDIEQNYYKRLLSEMIFIEKTNNFNRMLDCDNLNKVF			
					TGTAAGTCT	LSIRLNEHKSDIKHKREKSALYQHHLDTGHSFNFSDNEI			
					GAACAAAA	PVNNAIGSLVIPYCRELDYKLRVFCSKSFINVIFKSSNKLP			
					TTGGGTACT	FNRMYGKDILKKLNNPNVVYKIDCSDCSNCYIGQTLRK			
					GACGTTGTT	HSGSYLNFYSHCPISQKKAIVSHLVDKNYKLSSPEYFSN			
					TGAATCTTC	NFKYIYTVLLANNYPLDFIESCIKKRLYFLKSINNSQNNIS			
					GTGCCCTCG	ACLNLCPFNIPVYFRYVVDDILLCVPVDKVEYVLNVFNSY			
					GTCGACTAT	NNNYFQFNNNFYLQRRGSAMGSPISPVLADIVMRDLET			
					TTGAATATG	LFNNICFDAVKMAIEKFWHHIHNKINLSLTDFLNVIEFIM			
					TTTGATAAT				
					AAATTTGTA				
					ATGTATGTA				
					CGTCTATAT				
					TGTTGGTTT				
					GTAACACA				
					AATATATAT				
					TATAGATTG				
					TTTTCACTG				
					TATCGCGCG				
					ATTTGAATT				
					TTTATTGTA				
					TTTATATCC				
					ATTTGTGTT				
					TAAATTAAT				
					CAAATTCA				
					TTGAATTTC				
					CTGATCACT				
					ACGGATCA				
					TCGCCCCCA				
					TCTCACATT				
					TATTTATTT				
					ACTTGTATT				
					ATACACTTT				
					AAATTTAA				
					ACGCAAGA				
					ATAAATCA				
					GCGCATTA				

					TAAAATTGT TTATTATTG TTCTTAGTT TTTAAAATG TTATTAGAT GTGATAAC GTGGATGT ATTGTTACC TCAAGTTAA TTTATTTCG TGTATATTT TATAAAATT AGGACTTG AAAAGGCT CCTACCCAA AGAGGAGA GAAACGTC GTCAGGGG ATATAATA AAAAATTA TATACAAC AGCAATAG CCATTGGTT TTATTATTT GATGAATTT CGTCTTGCT CGCTGTTGT TGGAGTCAT AAACATCA TAA				
Pene lope	Penel ope-5_Co Fl	.	Copidoso ma floridanu m	TC	TAACAATTT TCATAATTT GATAATTCA TAACAATTT ATATCCATC AATCTCTTT TTATCGGTT AGTGGCCA CTTTCCTTG ATATTGAGT	SDINIYKNYLQKELDKIEKSLIFTLGIRFFELFERFFHYKS KIGIKTFDDILTNKFKILYFRKYKHTYEKKININNNSIRVS KTNSNIKSKWLHNLTTINLPKEVQFCINLGKNFNFFDFE NQNETKRVYLELFKSLNKPLHKNQSISNNIKFKILDLVK THIKNNNKSSLDILISNMLNNCNKFLKNNKNIMITKSDK GGAFVIIDKQDYINKVEESLNDKKLYEVVEHNPLPSLIN KTSNLLNELNGSDLFGDSGNKKVFNTIRDKCILAKCYGL VKVHKPNQPMRLIISNIGSPTYFLSKWVSEFLKNNLPPSP HNVKNSFELKNDIKNFLIPDGFCIVSFDVVALFNNIHFES VKVALESVWHILNLKINLEFFSFLNIIRFIMENNYFSFNN	SSF5 6672	321	494

TTGATTCTT	KNYRQKFGSAMGSPCSPSLADIVMRHIIKMLLFLGPFEIK					
TAAAGTAA	LIKVYVDDLLACIKISESDKVLNYLNSIDKHIQFTMETPN					
AAATAAAC	NNSINYLDMTLTWDNNGSIKTNWFQKPTHANLHLNYW					
TTAGTTTCA	SHSPLSQKIGVIYSLVDRAIKLSDTCYHNNNLKKVYEILK					
TAAATTACC	ENNYPPDFIKKHAERRKKIIYYNNKNDNNNNSNKKLND					
GCTGATTTT	DNKNYIVLPYIKNCDSNFKSLLNKVSYKTIFKSVNKQKF					
GTACTACAC	LNQLGKDSIKNKDKVCSVYCIKCKNCDKCYVGQSSRVV					
CAGGTTCAC	SKRINEHKRDVKSKKNESAIFNHVQDNGHSFDFDNFKVI					
CACTGTTGT	DNEPNFYKRLISEMVFIKKYNTVNKVTDTQSLSSVFEFYI					
AGTACTAC	NKYL					
ACACACAT						
GTTCTCACG						
TAGCTTTAA						
AAAGAATT						
CCCATATAT						
TGTACTATA						
TCGCAGTTG						
AGTAGAAT						
AAACATCTC						
ACATTGAC						
GACACTCTG						
CGATTATTG						
TTTAATATT						
GTAAGTAA						
AGATAAAA						
CTTTTATAC						
ATTGTTAGT						
GCGTTTTCT						
GTTGTGTAC						
ATCTTTATA						
TACATATTT						
ATATTACCA						
TTGCTGTAT						
ATTAATTTT						
TAGTGACTT						
GATAAGGA						
CTGCCTTGG						
AGTAGTCG						
AAACGTCG						
TTTTTTAAA						

			ATAAAGTC AATTTGAAC AGCATAGC CCTTAGTAC TTGTCATTA TTTATTGTA TTATTAGCT GTTCTTGGA ATCAAACA ATATC					
			TGACGAGT GTTGCTTGA TAGGTGTGC CACTTGATG GCCACGTTT CTGTTGGTC CGTGTTTCC TTGTACATA CATTGACCG TGGGTGCA CCCTTATGT GGTGACAA GACCTTAAC ACCATATTA CTATTTTTA ATATCTTCC AGTTACATG TGCTGAGTC AGGATGTG ACTTATATT GTAGATAA CTTAACTGT ACATACATT TTGACTGAT AAAGTTTAT AATATTAA GATTATTTA ATTATTAGT ATTCCATTG CCGTGGTTA	LLKPLEAALAVNKKIEVDTSKKIRQTYLQVIKKNIFNSK KISYFDKTINNHLNETKMFLKEHREILVTRADKGGATVV INKTDYINKVERALGDSMYYHRLDSQKRANHLTRIINST KKLVDVWDNRGVFDTVEIDRPDVNQLSGCNLAKAYGL IKVHKPGLPVRIIVSAIGSPTYTLDKWFSKFFNKFCKRSP SSLKNSYDLKSTLRDLIVPHNYLLVSFDVVSLYSNLPFSL IKKSVEEKIGRTLDSMHLDRKEFLDGLDFLINSNYFTFN NHTYKQTCGTPMGGCTSPWIAEIVLESLEKSVLPRCRDS CLMYRRYVDDCLLIIQSDSVDTVLGIFNDFNPFIQFTVER EDNGVIPFLDVTITRIADSVCEIDLYSKPSSSGRLLNFFSH NPVSQKIAIVFSLIDKCFKLADPKFRSRNIDRARRALTMN GYPTYFIEKYIKARISIINRFNKNKNHINSLTNCEMSPPV MEGSSRLSFIILPFYPRLEKTLRSFFSKHKLKTIFQNSNKF TNIIRLGKDQTPKLKRSGIVYRIECECGLFYVGQSKRRLA TRIDEHKRMVHKCDKNSVIAAHCSEFSHTFKFDEVKILD QEXNKFKREFSEMLYISYFNRNVVNRKLDTNKLHDHYKS FLYELRDRMHGWRCS			
							cd00304	283	365
Penelope	Penelope-10_CoFl	.	Copidosoma floridanum						

RTE X	RTE X-22_C Gi	.	Crassostrea gigas	GA						
					TGACTGACT					
					AGTCTTTAT					
					CAGTCATGC					
					TGCGTTTGG					
					CATTTTCAT					
					GTTCGCATC					
					TGCATAGGT					
					GACTTATAT					
					TTTCATTTA					
					ATGTAATTT					
					GTCATAATG					
					TTTTTTATA					
					TTATAAATA					
					TTGTAAATT					
					GTACAGTTC					
					CCTGAAGA					
					GGGAGTCA					
					GGATGACT					
					CCGAAATA					
					TTGGATCAC					
					ATTGTCCAA					
					CTAATATAT					
					GAAATAAA					
					TAGCCATGC					
					TCCAAGATT					
					CCAAGTTTT					
					TTCATTTAA					
					ATTTCCTTA					
					ATTATGGCA					
					AAGGAGCA					
					AGAAGACA					
					GTAGTTTTA					
					TTATCAAA					
					TAATTTATT	MDNPTYSPNETLNRNFLFDEINRLVFKTKSKSASGYDEI				
					CATTCSAGG	PYGVLKSQKVIETLTRLFQLILDTSIVPSVWRKAIICPILK				
					ATATACACT	DASTDPRIPMNYRGVSLLSCISKIYSAFINQRITSHLDAHE				
					TTGATGTAA	YLMSKTVSEKTGRVKTRIHLNSVIRIKITIRSFHDLRKCF				
					CATTGTGAT	DFIDREMMLYKLLVHDIDGKVYNSVVKSIYQNSESCVRW				
					TTAATATGT	QSVQLCEKYLPKLRIVCALNGQLTNSCNTGVKQGDNRS	cd01650	72	345	
					ACCATAAC	PTLFSIFINDLVQDINELGLGVDVGEKKLSLLLYADDIAM				

CR1	CR1-1_CGi	.	Crassostrea gigas	TACTACA	GAGGAGC	CCAAATAT	IAKSKAELQRMLDKLHQWCKRWRVLINTNKYKCVHFR				
				TTCATATT	AATCAAT	AAGTTAAT	KGKIKRSNCNFTIGSNALETVDHYKYLGVTFNYNSCFNS				
				TTTTATAT	CAATCCA	AATTGATAC	NAELLAKLLAKSGGRALGQIIAKIHNNKYFGFTSYEKLF				
				AACTCAT	GTCTTTTG	ACCTTGTAT	SSCVAPILDYGAGVWGYRKFQAIDNIQNRAIRYFLGVH				
				ATTTTTTT	TACATTT	AACGCGTG	RFTPILFLYGDVGWLPSQFRRWFCMLRYWNKLICFEDD				
				CATTTTCT	ATTTTCG	ATGAACTAT	RITKMAFNLDYERCRDNWCSEVKHIMNQIGIGQYYDTK				
				TGATTTTT	GTGTTAT	AATTGTTAA	CAIDLGLVESHLTNFYSEIWKDSLENVPKLRTYKTFKTD				
				AACTCAT	TGAGAAA	TATTACCAA	FGLENYISFNLSKAERSHMAQLRXGILHXRDPPLRIESGR				
				TTTTATAT	GTGTTAT	GGTGACTTA	YVGEPPNERLCTLCTQQVIENELHFVLDCPFYNQYRTEI				
				TTCATATT	TAGGCCCA	TAGGCCCA	LGEQLLSPALLNRDRQYVFTYLMSHKFRQLSKFVVKAF				
				TGTAATAT	ATGGGCCG	ATGGGCCG	RIRRQFLYR				
				TGTAATAT	GGTGCCCA	GGTGCCCA					
				TGTAATAT	ATATTCATG	ATATTCATG					
				TGTAATAT	TACAAGTA	TACAAGTA					
				TGTAATAT	GTTGTTATA	GTTGTTATA					
				TGTAATAT	AAATTTGTA	AAATTTGTA					
				TGTAATAT	TCATACGGC	TCATACGGC					
				TGTAATAT	ACAATTGTA	ACAATTGTA					
				TGTAATAT	CATAATATG	CATAATATG					
				TGTAATAT	CACATGTCA	CACATGTCA					
				TGTAATAT	CTATAATAA	CTATAATAA					
				TGTAATAT	ATAATTTAC	ATAATTTAC					
				TGTAATAT	TTACTTACT	TTACTTACT					
				ACATGTACT	TA	TA					
				ACATGTACT	TAAATTCCT	TAAATTCCT	MGNSLEQYRAAIGNFYFSCRCVCRSVLVNFFGFSFFSKIF				
				YRPVSLLSCVSKIMERIIFKHVYNFFHENNLFYKYQAGF	MFPSIGLFALAMKHLSRHIVVYNFYLQFFMFLLLLCGDI			cd01650	577		830
					EVNPGPIMTNVLDILHLNIRSIRNKVANLNTIVHDFDILC						
					FTETHLDRNVTNESLMLDGFNTIYRKDRNCFGGGIMIYL						
					SDLLRAIRRPEFETTCNESIWIEIESHTCSYFLCCVYRPPN						
					SDGNFWTSLSWSIEKVGEITDNIIIIGDLNVDLLSIPQTHV						
					IHDIMSNGNFVNNINEPTRVTNTSSTLIDLVLSTPAVKVH						
					ESGVLNIDSSVSDHKAIYISFSIDIELNRAYERNIWVYKD						
					ADFEKLNSLISNCDWDITINSAPNIDAAVAQFTSTFMAY						
					VRECVPEKTITIRPRDKPWYDSMLRKTMRIRDRLRNKA						
					LRTNNDADWTKFRHVRNKVNNMKKHAITNYYDNIESA						
					LHDSSRNGKIYWKLMKDLFKVKATTEIPPLQFFENGQPT						
					LAFSDIEKTETLNKYFSSISNIESANRSLPNLYYMCNNFF						
					SDIHIEVQEVFDIISILPVNKAIGPDNISHRMLKYTKFSISK						
					PLCMLFNKSLVENSFPNYWKIAHVLPLFKKDDPSVASN						
					YRPVSLLSCVSKIMERIIFKHVYNFFHENNLFYKYQAGF						

				TTCTAACGCAATACGTTTATGCGTTATTTAACTACCTTTGTATATTGTTATTTACCTTTTGTTTTGTCATTGTGCACATACATTTTTTATGCTATTACTTTCACCCCTACGTTTTGTTTATCTCTTTTGTTGAACAATATTTTGTCAATTGTCATATTGAGGTATCGTCATTTTCATTTGAATACCTTCATTCAGCTGTAAGTTTACTACTAGTGGTGTATCTTCCGGTTTTTGGTCTTATTCATTACCCATATACGACACATTATACAGATTGCAGCCATCATTTCC	ATGACAATTGTATATATTGTATGTATTATCATTAGGAGAGGAACTTGTAAGTTGTAAGAACTTGTTTCCGATCCTTTTGTGTTACTTACACAATAAAATATGTTCAAATCAAAA	LPGHSTVFQLLETYHSIVKSIEDGKMCCMIFCDLSKAFDRVWHEALIFKLQTYGISGHLLQWFKSYLSDRTQRVMYKNKLSSTLCINAGVPQGSVLGPLLFLIYINDVATNMLSFCRLYADDNSLQHCDLYTKNIEIVLNHDLKILSDWSDKWLLKFNPSKTKAVFFTKKHVEEFPKLFFQGCQLDYVSTHRHLGLLFSNDLSWSQYINNIVNSAYKKLGLLKKLKFTLGRTNLSKMYITFIRPVLEYASVVWDGCNMFDMERLEKVQLSAARIVTGLPILASTDSLYTETGWESLVSRRNKAKLITMFKIHTNQVPEYLSNIIPPVTKNVSIYHTRNSENYSVPISRLELYKKSFVPDTIRKWNSLKLEVREATSLNIFKKNITESIPQPPKYFSFGKRTINILHTKLRHNCILNYDLYRRNIIESPNCICGQKEDVYHFFFVCKNYVNARNNLFELLFNRLDETILINSHLLLWGSDNLSYNTNCFIFSAVQKFINESGRFNN			

			CCCCTAT					
			TGTTATC					
			AAAAAGT					
			TAATCAT					
			TGACA					
			GAGAGAG					
			AGTTGTT					
			AGACGAG					
			ATGAAGA					
			GAAACAT					
			CAAATCT					
			TTGCAAA					
			AAGAAAA			TYFCSRMLYTPESXVFYKKGKVDIFDCLEEXYCSFHRK		
			AAAGACA			GHIYXVGDFXARCGTLDDFISDDKLGKSVEXXLQSFXD		
			TTAGGCA			YDNNANVYCXHNADRGVNNFGKKLINLCRTTGLKIVN		
			ACTAAAA			GRNENDYNGNITFYNANGTSTIDYLLTDVYNFDNIAYFS		
			KATTCAA			TGIFNCFSDHAPVSFCIKTNYIDGGSKALCCANTSQGNG		
			TCGTTGA			SINQCTSVRWKEGADLEATWSISEHYDLFLSYVENQLY		
			TCTGCAA			TQEDVDNCVANFSKTLCKIMKPYCIITGGLANTVIDAER		
			TGTCGCT			CIDKPWFNDICKSKYRDYRGALHNFNVCKSDVNRIVLC		
			CTATGAA			AKKKVYKKNSPEVIELYFVQRKRFIKKTVQKFKHNYER		
			AWACAAT			YQGDRMEYLRKCNPKQFFRLFKNNKKISCDVSAKEFFE		
			CTAGTGT			HFKNLFENVKTTTYDNTSDDYEILPCYEELDVCISELEV		
			TTACGGG			EKAILKLKSNKSNAEDGVINXFFSKCKEVMIPILCRLFNN		
			GCTGGGT			IFDSGFYPDNWSKGCIVPIFKKGEMQDLNNYRGITLVSC		
			GGCGAAA			LAKLFTSIINTRLLEWDQRYNIITDAQFGFKPGVGTIDAIF		
			CAAAGGA			VLQNLISRTLKNKKRLYCCFVDYTKAFDLIDRSKLWYK		
			AGAGGAC		TGATTATGT	LQMEGFGGKLLRIIKSLYKNVRACVKHKGTFTETFRISA		
			ACAGGGG		ACACTTTGT	GVLQGEILSPLLFSMYLNDFERYFINSNCTSIELQMINMF		
			CAAGCTG		TTTTCTTAC	LIMYADDTVLLAESPEGLQDMLNTLHRYTDEWNLSVN		
			CGGGACT		TGTTTATTT	VQKTKIVIFRNGKCVKDNEKWEYNSKYYEVVNQFNYL		
			CTATATA		AATATGTAT	GMLFNYNGKFLSTQKHCAAQGNKAMFSISSKMKDLNF		
			TTTTGAA		ATGTTCATA	NVETQLNVFDTYVKSVLSYGAEIWGFHKGQDVEKVHIN		
			CTGGGAA		TGTATAACC	FCKKVLGVRKNTCNSAVYYEVGRFPLEIFRKQKIFKYW		
			TTGACTG		TATGAGAC	FKLRNSKNCILKACYEDMVKNNDIWISNIRKELSILGLA		
			GAAAGTG		ATTTGTCTC	DLYQSTMKESVILDIIFNRMCDVFKQKVVNDISNASKCI		
	RTE X-23_CGi	.	GAGTTCT	Crassostrea gigas	TTGGAATA	LYKHVVDHFCLQVYLKKPIHVKYRKLITRLRLSSHDLKI		
RTE X			GTAGGGT		AAGAACTT	ETGRYDNLTRDCRKCESCNLNVIEDEFHFLLICPSLCSLR		
			GCATCGG		GAACTTGA	DKYIKKYYSRKPSVYKVIQLLSTSNTKELCNLGKYLYY	cd01650	477
			TTCGGTC		ACTTGA	ANKQRTLHVNYPN		736

				GATTTCA GCASAAC AGGAGCA SACCTAC TGTAGCG AAATTCC TTTATGA AAAAGAC AGCTGCA GGTCTGT AGCTCCC GGTCTGA AAAAAGT TGAATGG ATCCCAG GTCGTCC ATCCGAA TTTKTTCA GACCGGK AGCTACA GACCTKC AGCTATG AAAAAGA CAGATTA ASTSTCCT TGAAASA GCCTATC AATTAAA WKGAMG TGGATTC GGTATCA ATGAACA ATTCCCG CACGCTT TTGAGGA AAGACGG AAACAGC TTTATCCT TAACAAA AATATGT					

			TAGACAT						
			GAACATA						
			TCACTTG						
			TAGACCA						
			ATTTTGT						
			AGTACTG						
			TTGTCTG						
			GATACGT						
			TTTGAAC						
			ATAAAAG						
			CAAACAG						
			GGTAA						
CR1	CR1-3_CGi	.	Crassostrea gigas	GATGATA	TAATCACA				
				ACTCTGT	AATCTGAGT				
				ATTACCT	ATTTTTGTT				
				CAAGAAA	GATAACATT				
				TATTTAT	TTGCTTATT				
				GTGTCAC	TTTATTTAT				
				AGTATGT	TTATCTCGT				
				TTGASAG	TCATCCTTG				
				TATGTGT	CATACATGT				
				ATTAATG	AACATTAA				
				AACAAGA	GTAGATAA				
				AATAATT	CATTAACTT				
				GATATAC	TCTTATTAC				
				TATCTATT	TACATTTAT				
				CTTCCTGT	CAAATATCT	SHPERIVFKHLYNYFHSNNMFYKYQAGFLPGHSTVFQL			
				AAACAAG	ACGTACTG	LETYDDIVRRIDEGKSCCMVFCDLSKAFDRVWHKGLLY			
				GCCATAG	GTGTAATTT	KLKMYGIDGKMYXWLKSYLANRTQRVMYKNLFSTKA			
				GTGCAGA	ATTACTGTA	QIHAGVPQGSVLGPLLFLIYVNDVAHNMSSFCRLFADD			
				CCTTATT	TAAGTACA	NCLQYSSKCVKTIENCLNHDLNVLENWSKKWLLKFNPS			
				AGTCACA	AACTGAGA	KTKVIFFNTRKNLEPPKLIFQDSQLEYVQTHKHLGVTFSS			
				AAATGTT	CGTATTTGC	NLSWSSHIEILLSNAYKKIGLLKKLKFKANRKTLSLMYT			
				GAAAGCT	AATAATATT	SFIRPQLEYASVVWGGCHAFDSEKLEKLQLXAARIVTGL			
				ACTAAAC	GTATTTGTT	TIFASRESLYFETGWEPLNARRKIQRLNIMYKIHNNLVPE			
				TTACCAT	GTTTAACAT	YLSDIVPDMRCNASAYYTRNSQNYNIPVCRLQIYKSSFV			
				TGTCAAA	GTAGGGAG	PTVVNEWNSLPLDIRDSSSLSLFKNKISKSLAVAPPYFSY			
				CCATTAT	AGGGCGAA	GKRRTNILHTRLRQNCALNIDLYRCNIISSPLCSCGKIED			
				TTCTATTG	TTAAGTTTT	AHHFFFSCPKYATVREEFFNKFLRFGKIRILDSRSLLWGD	cd01650	4	226
				TTTAACC	TATAACTTG	DSLSVDENKYLFSIVQIYIQGTGRFE			

				GCTCACTTAGAGAATCTGTTTTTCCTATTATGTGGAAACAAGCTAACGTTATTCCACTCTTCAAAAAAGATGACCCTTCTGTTTTGTCAAACTATAGACCAGTTTCGTTATTAAGTTGTGTTGGTAA	TGCCCAATCCCATTGCAATATATAATGCAATAAAATATGTTCAAATCAA				
Penelope	Penelope-2_CGi	.	Crassostrea gigas	TTATT	TAGACCATAATTAATCACCTAATTGTCTACGGACTTTTCCGGTTTTTTTCCCGATTACGACAAAGAGCACACGGCGGGTGTGACCGGTCAGCAGAGGATGCTCACTCCTCCTAGGCACCTGATCCTACCTCTATCTATTTGGAGGTCCGTGTTGCTCTGCTTTGAATTTGTATTTCGTTTTATGGATTT	MKGPKFREPRSFTWRRNFIHIMNSVEDYAKRWAKSEKEELDTLSEWVKSIRCILKSRIKRLRGQMKTIYPSVFNKAEVRKELDRLHDQYVLVPADKASNNIVFVCKAHYINCIFKELGFDSTHGNPTYTHSSLSKQEILQNHKSVMDIFNIPNTQNEFDLPYLYWIPKLHKSPYKQRYIAGSSKCSTKPLSLLLTKILTVVKEKLQEYCTTIYSRSGVNQMWILKNSKELLETLRSQNLTKINSIKTYDFSTLYTTIPHDKLKSRLFDIIDSCFFNKNGSRKNAYLVIGNLKNYFVKNHSDCTHKYSEVDIKKMLEFLIDNIYVVFGNQVFQQSVGIPMGTNCAPLLADLFLYSYEAEFIQKLVREKNKSLAVAFNSTFRYIDDVLSINNCYFHTYVDSIYPSELEIKDTTESASSVSYLDILLEMDIDGNLTTKLYDKRDDFNFSIVNFPYLCSNIPSSPAYGVFVSQLIRYARACSSYEQFLRRGKLLTNKLIKQDYQQSRLKSSFRKFYGRYNDLVSKYNLPLGRMLTDVFHTNC	IPR000477	147	467

				TTGAGATG GTTGACGGT TTGTTTATTG TCATTTTTTT CATT	RAILTLSCPFIKIKSTHCHRPLTLLWNDLXWLRMGIARF			
			TGCGCAT		MKXVSVGLYXIPWKTQISQLLQQIKFEPSEDILSVAKEL			
			GCGTGAA		VAHLDDSAQVVAAARLKNRKIDKPGLVKISFARVDQKI			
			AGTATCG		RVLRAKRKLGEISDYRSVFLRSSKSHTERLIELNARTILN			
			ACATGGC		EMPNGKDFRITANGRIVKKTDNEPNTRDESSSELPRPFET			
			GTAAAAT		IRIGHLNVCGWTNKNHDLRVEIIKYTDCDIVSINETHLLN			
			TCAGCTG		KENLDLDGYQWMGLNRSNVHVKAPKGSGGVGLFIKKH			
			CGAGAAT		LCKNYKVDIIDNSFDGILGVKFTNKFSDYVFVVFSCYLP			
			TTTCTGTT		PESSPWGRDASSFFSHLLTQIYIQYDTDSIFICGDFNARIS			
			TATTCCA		NIPDFDKDLDTIPPRLTIDNVHNQHGSSFLEFLNDCKFCV			
			AAATTTW		LNGRFGNKSNDYTFQSTRGHSVVDYICVPHDVYNSVLR			
			GGTWGTT		LSNNFLMEDLAINQMTIHFNRQEVIPWLTIYVFLMMFTI			
			GTTATGT		QCSDFQIISCSDIVENAGLVSMQGENCRIPDHGFLIFDFMI			
			CAACAAA		RDFENAHPSENTVTSSTRPIKKYKLRCIPPSFMSSVNVRG			
			AACTCCG		SLLNMIRIIETCRESRENIDRTYTQLCNIITKEMECSIPKFD			
			ACAAAAG		WSKRSRKKFRTHKPYWNNELTELWRVMREKEKTFLKH			
			TCCCGAC		KKGNSVLKQKSFSEYKRAQQELDKRMRFFRRKYNLDL			
			TAACGTG		AINLDNFSSNNPKKFWEELRSLGPKRKNKVPTECYTDD			
			CCAAATA		MYSSILNHRLVSYLEDNNLLVDEQNGFRRERSCADHVF			
			AATCTGG	TAATATGTG	NDGSISSDPTYVQNVWKSEFCNLYNPSTDVAQQPSNDA			
			ACCTGGA	TATATTTAT	FLQRVKESVNFMENNMLDPLREHNPYLNRPIELDEVKN			
			GCTGATG	GTTATAATT	VIMKSKNGKSSGVDEIPYEVLKYDCVIEVIHRLFMLCFE			
			TTGGGAA	GTATTTGCA	SHIIPSVWRQAIICPILKDSNSDPRVPLNYRGISLLSTIYK			
			GCGAAGG	GTGACATAT	MYSSILNHRLVSYLEDNNLLVDEQNGFRRERSCADHVF			
			CTCTCGT	AGGCCCAA	TLNSIIQNRKETFVAFIDLQKAFDTVDRDLLKFSLLNNGI			
			CGGACGG	TGGGCCGG	DGDFYNSIKSIYGNTVSCVRVSELCTDWFSTSAGVRQG			
			TCAAACC	GTGTTTCAT	DNLSPTLFALFINDLAKEIKELRKGITVGEINLGILLYAD			
			CCAAAGC	TATCTATTT	DIAILSDSKENLQKILDTVSYWCEKWRLHINMKKSGIIHF			
			CAACACC	TGTATAACA	RKKGPRSDFQFQINSNILQYNTDYKYLGVIFHEKMNFQN			
			CAAGCAA	TTTTGTAAT	NADTLSKAGGRALGGMISKIRSIFPKFGPYKEIGISTFEKL			
			GTCAAAA	TGTTATAAC	FLNCVAPVLDYCSEVWGLQNYHSIEMVQNRALRFYLG			
			TGGCGAC	ACATGTCAC	VHRFTPLPALHGEFGWKLCYHRHCINAVRLWNRLLTM			
			GAACAGC	TATAATAA	EDDRLTKRVFLWDFNQTGLSRWTSRIKSLFESMNLLHIY			
			TCACCGA	ATTATTTAC	DNLSICDVNEFENKLFLTYENEWLGKINQLPKLRLYKSI			
RTE X	RTE X- 26_C Gi	.	Crassostrea gigas	ACACACC	TTACTTACT	KTNFKTENYLEMNISKSQRSILAQFRCGILPIRIETGRYK	cd01650	108 8
				CGGCAAG	TAC	GEPLDERICNFCVLNEIEDESHFLLNCTIYSDFRRELFNNI	831	

RTE X	RTE X-21_CGi	.	Crassostrea gigas	TCCGGTA GGTCAAA CATTTCTC ATGCCGA TTATGAA ACATTTG TGGAAAA AATGACT ATTATTG CTGATGC TGTTGAT AAGTTAC AGAAAGG GCAATCG ACACTAC AATCAAT TGTTGAA AGCAAAT TAGACAA GTTCAAA AATGAGT TCGTTCA GGGCATC AATGAAA AATTCAA AGCCATG A	ATCACGT GATATGT ATGGAGG TGGACGT GTTTTTAC CGGAGTA CGACCTA CGGGAAT TTCATAC CATATAT AGCGAAG TTTAGTT AAAGGTT	TAACTTTAG ATGTTGTTC TTTTTTGAT TCTTTTCGC CTTTACTTA TGTCAATTT TTCATATTT TGTAAAAT ACCCACTTC TGTACTTTA	GYNPTLVMSNLDCIFFLSNFPRQTSKYLJESYLYRQSLL YKKRNA	MNLSSPSPSKPPKKKKQRTKSSTQGKMATTQQGACTNS VYPNDNLSQSVQPIQQVGNLTYHQLQPMQYIQSAHAPM QTHMQHQPQNVQSNTPQVTPQWVSDIVTQLKNMNQKL DTIIDTTLSSMKREIKGLSNRVVEVEKSQEFISKSLEDNEK NTSKALTELSDLKMALDQSLQENVRITESVLDMQCRSM RDNLLFFGIEEAQSGEQENCEDIITSLCASKLEIRSEILIER AHRLGSKKNNKPRPIVVKFNRFPHREMVRKCSYKLKDT HISIGEQFPKEIQERRKKLLPIYKRAKENNQKAVLARDK LFIDGKLYSDDGPVQMKATPTQPKASGRRGQKHGRGK RGQENQTGLRIIAWNIHGGLENKWSERNFKSFLCTYDII CLSECWIGKDFNISDQNFQSFVFPRKSGKSQGGGTVVLI RNVYTPFISFVESLHDTIVWLKIDKSFCNHTKDIYIAFTY FPPSNSTFLKERDIDLFMELENQLSYYMSIGDIFLFGDFN	cd01650	920	118 1

			GCGGTAA	CTATGTTTA	SRTKTLPDYIVNDQVHESVLDNNLQYSSDNQLTERVNP					
			ATGTGAT	ATTTGTAAA	DLGHNEYGQRLLSLCRSTGVRIVNGRHSGGHSNDYTFN					
			ACACCAT	ATCTGATGA	GARGLSTIDYLLTTYDMFDFISKFIVCNFTTFSDHAPLHI					
			TGATACG	GCTGCAAA	EFKCTTLLDPCDTSSHYRTRNVFRWDSSYVDSCNDFFNS					
			TGAGTAT	GCTTAAAG	IEGTCNFVDFEKVTSQSDLDECMLKITDMLSKDIGNFFY					
			TATACTC	AAATAAAG	VSSKNSNQSKPANSRSRYSTYLSTNDKPWFNENLKSLY					
			TGTTCCT	AATGAATG	RDYITALRTFNVCKNAQNHLNLVSKKRTYKKLEAKLKR					
					SYKQIEGNRLNILKKNNPRQFFAKFRKKPKSVGNVPIEIF					
					FDHFKKITNVDGSFDTNCEQANMEVLIPIVNKPTDECVF					
					YELDKEIDVDEIRKAISNLKRGKSHGEDGILNEYFINFQQ					
					HLLPLLYSLFNCILDTGFFPKSWSSSIIVPVYKKGDCTNP					
					DNYRGISLVSNMAKLFTSILNNRLLDWSKTNDVITDAQF					
					GFKPNCGTRDAIFALHSIVTNTLCKRKRLYCAFIDFKKA					
					FDSIDRSKLWLKLSKVGIQGKLLNVIKALYNDVKSCVV					
					VNGHLSDYFCNNVGLMQGEVLSPILFNLYVNDFEMNFL					
					KSGCIPHELLSLNLFLLMYADDMVLFSESVEGLQLLLNE					
					LSHYCKTWNLCINVEKSKIVVFRNKGKIKCNEKWCIGN					
					DVLEIVEQFTYLGIIIFQYNTKFTKAEKQLSDQGRKAMFA					
					FRKNIRGMILNHETLLSLFDCYVGGIINYASEIWGSHKG					
					NNVEKLHLDFCKQTLGVKRSSFNAAVYAELGRNTSYY					
					KKVLFNFKILEGCVMVGIPLITKRFCSILKFWKDVLCSN					
					NCIIQKCYEEMYINCELHGHKNWASDVKKILNEIGFNEV					
					WNKQTIDSLTLKVINQRIADLAKQDIFGKIENSQKCHFY					
					KYLVDGFYLQYYLRKSVPVKFQQCISKLRLSSHRLAIET					
					GRYNNTHRAQRNCFSCKNYVEDEFHFILVCPVYIEYRQ					
					KYIKKYFYEKPSMFKLLQLFNNENFKDMCNLGKYIYYS					
					MKLRNDICN					
			TATAATT	TAAAATGT	MDDKVQQSYLSILTLNCQGLGDINKRRDIFKVLKSKNY					
			GTATTTC	GTTTATAAT	NIYFLQDTHFVEKEESLIKGLWGYKAYFNSYKSNSRGV					
			ATGAACA	AGCTGCAA	AILINNNCDITILNEHKDSDGNFLLLDVTIESLPFLLINIYG					
			TTAATAT	GGAAATTG	PNSDSPAFYINVLNKIQDIYSNQHIVIGGDFNLVFDKDLD					
			ATCTCAC	TAAGCAAA	SMNYKNLNNPKARLEVLKIMETLTLKDVYREHNPTLKK					
			ATAATTT	TAATGACA	FTWRKKNPIKQARLDFFLISESLLNLSLYVKYENSYRSD					
			TGGTATG	AAAAAATA	HCPVVLYVRTNEFLKGKGFWKFNNSLLKDKEYVYLIKE					
			ATTTTTAT	TAAGTCTAC	KIKNVKQQYACPVYNLENIETIDDNIQFIISDQSFLDTLL					
			AATATGC	TTTTATGTA	MEIRGKSISYSTYKKKKTTEREKQLEQEILLLEQNFTVET					
			TGTCATA	TCCTCTTGC	KDELSNKQTELELIRKNRLQGQCIRSKVKWIEEGEKPSK					
			TTAACAA	ATATGTATG	YFISLESRNFSSKIIPKVEKEDGSTVFNQFEILNELRMFYD					
	Tx1-5_CGi	.	GCTCAAA	AGTTTTTGA	NLYQKRETIDSENIYEKLKLYDIPKLNKQESDLLEGPLN	cd01650	530			
Tx1		Crassostrea gigas	TCATCAG	TAGATAGT	KLEVYNFLKNMKNDKSPGPDGFTSEFFKFFWNDLGSFIT					791

AAAATCA	GTTCCCCCTC	RAINNTYELGHFTESNKLGVITCLPKAGKPKQFLKNWRP				
AGGTACT	TCTTTCTCC	ISLLNVLYKIASGCIANRIKTVLYKLINEDQTGFLKGRFIG				
TTTTTTTT	ACCCTCTTT	ENIRMIYDIMQYTEQNQIPGLLMLIDFEKAFDTISWKFIY				
AAACACA	CACCCCCCC	QTILDFFNFGISIKNWIQTFYSDIKSCVIQNGIASDYFFPQR				
TAGTATG	CCACCTCAT	GCRQGDPISPYLFLLCAEVLGILIRNNKEIRGIEIDGVEYK				
CGCAGCA	AACTAAAA	LSQYADDTSIIFDGSPQSMDGILRELDYFACISGLKINFSK				
TAGGGTT	ATACTCNA	TKMVWIGSKKISKEVFHHSRWKLDWNNPTFDLLGVKFS				
TAAGGTT	ACTTTCATG	VNLNEMVELNYPSKFLEVTRLIKQWKLRNLTTVGRLTV				
TTTAAAA	GGTGATAA	IKTLIIPKLNHLILTIPKPSVEYLQQFEKEIYIFLWGGKVH				
AAACAGA	AAGCAAGT	KVKKNIIIQDYRYGGLKMIDYMEFVCALKSTWIRRLIHS				
CAATAAG	TTTCGAGAT	NTKWVKLFESELGLKIETLWEKGSDFITNISNKISNKFW				
TTATACA	TGTTTGTGC	KEVLLDWVKIVMSYSQNDIKVLNEHIWFNPNIKINNSSIF				
TATAAAA	ATATATATG	IKNYFNCGYKFIKDLMTEEGNFVSLETLENSNVKTNFLE				
CTCCTCA	TCTGTCTGT	YRSLKGAVFERLKALNNNKVEYGPFIPNSLSIFYKQSKG				
CGACCGA	CGGTGTATA	CKDMYYVLIDNKKESIRSVKKWTESGFAYTEIEWNNIFE				
TTCTAATT	CACCTGTCT	LPFKTSKESKFQWLQFQILHRILPTNEYLFKLNIIDSPACS				
ATTTATA	GTACGTGTG	FCNQDIETIEHLFVDCFHVKEIWLYFEEWLRSKFNFLMT				
CTATTGT	CTTTTTTAT	FDKKTILFGKFKNRNMYRLQNILILTVKQFIFASKYKNV				
ACGATTT	GTGACTGTG	PNLSTLMLKRTVTDRIHIEKLLLLLKNCKFAEYEKYWQTI				
TTTTTTTT	TTTATATTA	CDLL				
AAGTTGA	TACTTATAT					
CAAAAAA	TCTAAATCT					
AAGGTTG	CAAATGAA					
TTTTTGGT	GTGAATAT					
CAACTTC	GTGATGTGT					
ATGCAGT	ATGCACTGT					
CAACTAT	AAACTAAA					
TTTTTAGT	CATATAAG					
GTTTGTA	AAGATGTA					
AAATTAG	AATAACTTG					
CAAGCGT	TACATGAA					
GATGTGC	AGTTTGAAT					
TTTCGCA	AAAAAAAA					
ATTCCAA	AAAAA					
ACTTCTG						
CCTTTTA						
ACTGTTTT						
TTCTCCCT						
TTTATTTG						
CCCCCTTT						

			CTATTCTT TTCGGGA ACCGTAA ACCTTAT AAAGCAT TTATATA GAC					
			AAAAATC GAGCGGC WATGACA ACATCAT CAACGAG ATGCTCA GTGCAGA TGTGTCG ACCCTTG AACCAGT GTTAAAT GTGTTAT TTAATTC GATATTT GATAGTG GTTATTA TCCTTTA AATTGGA AATTGGG TATAATT GTGCCGA TTTTCAA AAAAGGA GATAAAA TGGATCC ACAGAAC TACAGGG GCATAAC CCTACAC TCTAAGC TGCCTGG CGAAGCT TTTCTCCC	TAAGCGTA AATCAAGTT TATATATAA CTACTGCTT TGTTGATCT TTTTGATAT TTTTTCATC CACTTTATA CCTTTTATA TATATTTAT ATAACCTGC CTCCCACCT GTACCCAC AATTATTGT AAAGTACT GTTAAATTG TATGCTATG GCAATGTAT GTAATGTAT ACGCCAAT AAACTACT ACTACTAC	KQSRVLQGDILSPMLFSLFLEDLETHLHENXSAGIEIDQI NIFLLLFADDLAIISETPNGLQKHLNTLHTYCERWGLTV NIDKTKVMIIFQKGRTRKGYKFCYNNKNLEIVDSFAYLG TVLSPNGLFNKHVKITAEKTLQLYNTFMYKTQKLSLDC KTMMELFDVACSSKMNYGAEVWGYTNAKDLERVHLK FCKRLLGVKTSACTPAVYCELGRKPLLLTRTLRIIKYWF KVLRSQNCLMKSIYDYDLKNMLSSWIIFVKNTLLSNGF GDVWYYPESVNEKMFYVYFEQRLSDCFYQNLRVEIEN AHPLRLYKYLNLPVHQAEYLNHILYDKHRKALSRLRLS SHRLRIESGRYDKKCPINERKCIYCKSNDVEDEYHFVLV CDFYTELRKLYIKKYFYVRPSMAKFIDLLNSDSIFMLKK LGKYVYESFNKRDNFTIRRE	SSF5 6672	4	137
RTE X	RTE X- 5_CG i	.	Crassostrea gigas					

				GTATCCT AAACAAT AGATTAC AAAGATG GGCAGAA GAAAATA GTGTGAT ATCTGAT GCGCAAT ATGGGTT AAAGAAA GGTTGTA GTACTAT TGATGCA TTATATG TTTTAAA AAGTATC GTTGACT TTAAAAT TGCAAAT GGCAGTA AGCTCTT TTGTACA TTTGTAG ATTTTCG GAGAGCG TTTGACA CGGTTGA TAGAAAT GCATTGT ATTATAA AATGTTT AAGTTAG GTATAGA AGGTAAA ATGTTAC GCATAAT CAAATCT ATGTATG ATGATGT					

			GAAAAGT TGTGTAC GAGTCAA TGGTAAT CTTACAG ATTTTTTT TTGA						
RTEX	RTEX-2_CGi	.	GAATCAT GGAAAAG GATCTGG TGTCAAT AAAAGAA GCTAATG AAAGTCT ACAAAAC TCTGTTA CAGACCT AAAGGCT CGTTCAA TGCGCGA CAATCTG ATATTTTC TGGAATC CCGGAGG AGAGACA TGAAGAC ACCGAAG CTGTCCT ACAAGAA TTTATCC AAAAAAA GTATAAA CTGGATT ACGAAGT CCCATTC GAAAGGG TCCATAG AATGGGT AAATGGA	Crassostrea gigas	TAGTTGTTT TCCTATTAT GATCCTTTA TACTAAAAT CTAACATA AGTAGTGA TTCTATAGT CTGTTCTTT TTTATTTAT ATTTTTTCC CTGTATGAA AATGTACCT GTTTGTCTG TACCTCTGA CGATTTGTA TTTGTATGT ATTATATAT ATGTTCCTC TTGTACCAA ATTATTTGG TCTGAGTGA ATAAACTG AACTTGAA CTTGAA	MKNRHKLSNRRSGGLILAFKDTIVKYIKVLETDCKFIFLF KLSKTLLNISEDAVCGLVYIPPEGSRYSSPDCFSEIEQEIIN VESNCKYLCLFGDFNARVGKMQDYYTPDDFLAHVTHC EIQSEVATRSIIFDNYQIPLDRIVRDHTINNFGYKLVEFCK NNDIYIMNSRVGNDKNIGGTTCKDSSTVDYVLASANFF AIVESFNIKNFCDLFSDVHNPISFSFSICINDAKNNIPGKN QPSVKLWCADKADDFRNNIQMSAVCEIENELNVLSSCI DNICKDNISKIAKDICTVFVNTARETFGMSKHVDNFSNE KSSKPWFNRVCRTARRKYHLAKRLYNRNNSTDNKENL KVLSKQYKRTMDGCIKDYKKTITKKLVNLRSTNPKDY WKILHGSRRDEKCAVSIDDMFQFLKDMNEGGCENGESNI DNPDFIGTCMYDDILNGEISDEEIVAAVKKLKNSKAAGY DLIVNEHISSTLSIFLPVYKKKFFNIIFCSGFVPDEWLIGIIKP IYKNKGDPTQPENYRPISLLSCLGKLFTSILANRLEAYAT ETNLIYENQAGFRKNYSTLDHILTLQFLCDIFMKRKKKL FCAFIDFKQAFDTVWRTGLWSKLILNGINGKCFRYIKNM YMGIKSLVKLNGIASNFFECNVGVRQGENLSPFLFSLYI NDLENFLLEKNIVGLQSISNEIENELLLHLKLLILFYADD TVIMAESTEDLQNALNEFYSYCTQWKLNVNVNKTKIM VFSKGPTSKKLFYYKGEVIENVKEFKYLGIVLSRTGSFIK ARKHLCEQAQKAMYGVIRKIRNFNLPVNCQFDLFDKVV LPVLLYGCEIWGFENLDIIERVHLKFLKTVLNLRSTTPNF MIYGETGRYPLYINVYTRIITYWGKMLISPENKIVYTVY RYLFSQHFNSDCKSPWIDCIQRILNMCGLQYVVWQNHNF VNIEWLSATVNQRLKDQFMQIWSNDINCSSRGQVYKIL KPYFGFEKYLDILPVKLRKKFIKFRTSNHRFPVETGRWY NIPLNERLCLLCNKGLIGDEFHYILECSALEEIRKKYINTK YWKRPNFFKFSELMTNCNCKSLRKLCVFISKIFDAVCSS	cd01650	506	776

				CAGTTCCCATCCCAGAAATATTGTCGCCAAATTTACCTATTTTAAGGATAGAGAATTAATCCGTACACGTGCCCCCCAGAAGCTCCACAAATCCAACGTTTGGGTAAACGAACAATTTCCCCCTGAAATCGAGGAAAAGAGAAGGAAGCTATATCCCGTCCTCCGCCAGGCAAAGAAAGACCACAAGAAAGTAAAACTGGTCCGGGACGTGCTTTACATTGACGGACAGGAATATATCCCTCCTATAGGATCCGCTCCGAGAGCATTTCCAACGCCCACGC					

				CAGCTCA GACGCGA TCCGCCA GCAATCA AACAGCC GAGCGAC AATACAC CCCGCAA AACCCGA CCAAAGA CGACCGA CAGGCTT TTAAACG ACAACGT CAAGGAT CAACCCC CGACCGA ATTGGAT AGGGAGA CGCAGGA GAACAAC TCTCGCG AACAAAC AGCTACA CGTCTCT GAAATTT CTCTCGC TTAACGT ATGCGGT GTCAAAC AAAAACT AAATTAC CCCGAAT TTTGTGA TTTTATTG CAAACTA TGATATC ATTGGAC TATCTGA AACAAAA					

				ACTGATGATTCGGACAATATTCAAATACCCGGTTACGTCACTTAC					
CR1	CR1-6_CGi	.	Crassostrea gigas	CCAAATTCCAACAACAGCACAGACAGTATTATTAAAGATATTTGCAGAGAAAAACTGGGGATCCAAATAACCGAAAACGATATTTTTCGTTCCCATCCCATTGGAAACAGAAACAACCATGGGGGCATCCAAATTATTTGTCGTTTCCAAAATTGGAAAACAAAGAACGCAGTGTACAAAGCAAAGAGTAAACTAAAAAATTCCGATATCTTTGTGACCGAGGACCTCACCAGA	TGATCTTATAACTACTAGTACACCGTTGTTTACTATATACATGTCAATTGCATATATGTATACGTCCATTCTATCCACTGACCTTTATTTATGTTGTGTACACATATTTTGTATTATATCTTTGTATTGGGAGAGGGCGGTCTAAGTTATGGAACTTGTGCCCAATCCCAATTGTATTTTTGTACAATAAAAATATGTTCAAATCAA	MAGDTIEQYRAAIGLFYCCVRCVKKCFYLNLFQLRXLFKCLLQYLLWFGNLYKTLSMCLAQHSLVIDYFFQVIFMLLFLSGDIETNPGPCTGTQYTLDIFHLNIRSIRNKLDSLLSLVSDFDVLCFTESHLNSSIPDRNLFIDGFNTIFRKDRNSFGGGIIVYVSNSLRVKRRTDLEPTNIECIMGNQQTLNALWVEISDPTCNILLCCTYRPPNCDNSFWRNLSWSIEKAYDESDKIIVVGDLNVDFLNIPHTHLVKDLLSSHFLENTITEPTRVTQNTSTLIDPILVTRDVSFFDSGVLDVHHTISDHKATYIYVKSNLNLNSAYKRKVWLYKNADFDKLNFLITHTDWETLIMGADDVDQAAXNFSETFLSHIRKSVPEKTVTIRPKDKPWFDSALRKEIRQRNRLRKIAMNSRNALDMQKFKKSRNKVNNMKKYAIDNYYNNLEINLLDSSKNNSKLYWRLLKNVFDIKMSTEIPPLQFALKNGNQSIAYTNTEKAEVLNDYFSSISCLDDNDARLPEFNSECNNFICDVCISEQDVIDIISIIQVNKAVGPDNISHKMLKNTMYTIAKPLCSLFNRSLRDCTFPSSWKISNVLPLYKKGDSSQMSNYRPVSLLSCVSKIMERIIFRHVYNFFHDNNLFYKYQAGFLPGHSTVYQLIETYDHILRAIDQGQSCCMIFCDLSKAFDRVWHKGLLFKLHTYGITGNLFKWFNSYLVNRKQRVMYREVLSSIKSINAGVPQGSVLGPLLFLIYVNDIADNMLTFCRLFADDNSLQHAAINVHDMESYLNRDLLTLEVWSKKWLLNFNPSKTKAVYFSNKLTQTNTVTPTLKFQGCDLEFVPTHKHLGVIFSEDLSWSSYINSIIANAQKKLGLMKKLKFKLNRKSLSLLYTSFIRPQLEYASDVWGGCSINDSDKLEKIQLIAARIVTGLPIFASRESLYFETGWETLYSRRKISRLKTMYKIDRNLLPSYVMDIFPNKRANVSNYSTRNAHNYXIPKCRLELYKSSFVPTVVDEWNALSTDVRGSDSVRIFKEKIHVSVKNENXLTCRTRPNFFTYGDRFLNIIHTKLRHNCALNNDLFRCNIINSPLCSCGKJEDTYHYFFTCNKYKDARNVLFNDVFQIVNLNIVNTHVLLWGDSSISDNENKHLFTLVYNFIKNTERFN	cd01650	595	851

| | | |

				GTATCGT GTCTTCA AAAGTAG ATACATG TATTTGC AGATGTA TGAAAAT AATTTTT AGTTATA TGTACAC CAATCTA ATAGTTG TGGTTAA ATTACAT AGCTACA TTTTTGTT TGTTATA TAGTAGT ATTATAA CTGTGTA AACCTAT GAATTTA TATTTCCC ATTTTTTC TCAAATG CAATCAA TTTGGTTT TTACCAA TATTGTA CTTGATG CCATAAT GTTTTTTT TTTCTAC ATTTTCTT GTTTTTGT TCATTTGT TCCTTTCT TGTTTTTC ACTTTGC ATTCTTTA					

								195
							1	
						cd01650		
					CESALNLLLLKWKQCIEKGNIVLSVFVDLKRAFETIDRT KLIRVLKKSGIRGTVLKWFSSYLENRKQVTRYNDAVSA ETAVDLGVPQGSVLGPLLFILYINDIKQALKRVQVNLFA DDTVLFVVGDSFDECFDIMNEELGGFSEWLKWKKLQL NITKTKCMLVTTRQTKDCRRSVQMDGGIVERVETMKY LGVMLDEKLNFNEHIDYTIRKAARKFGVLCRINRYLTAE TKVQLYNSLIAPHFDYCSSILFLATQRQMKRMQVLQSKI MRLILRCDRLTPRRSMLECLQWMSVVKQRIEYNTLVFVF RVKERMAPQYLTEAMVRGRDIHEHDTRGADDLRLQIW RKACTQNSLFYKGFKLYNQLPEAAKMTSNINEFKRSCK EFVRQRPLE			
					TAGAAGTG ACCCACGA TGGTACTGT GACGAAGA GCACGTTAT GACGGTCG GCCATCTTC ATTATCGGT ACACATCG CATGGGAT CACTTTGGG CCGCATATG ATAAATTTA TTCAAAAGT AACGCGAA TCTGGGCGC			
				G	TTTGTTTG CATACAC TTTTTCAA GTGTTTG CTTACTG TGTTTTTA CCTTCAA CTTTTGTC TTGATGT TACGCAA CGAAGGT CATGGTC GTTCAGT ACATGTA TGAATTC ACACCTG TCTTTTCA ATCGCAG TGCTCTCT CATAGTT TCAATAA CGTTAAT CAACATG TCTTTCA			
			Culex quinquefasciatus					
		.						
		L2B-1B_CQ						
	L2B							

					GGTTTAACC CTATGGGCT CATATGTGC GTGGAATA GCAAATGG TTCCAATAC CCTAAATG GATAAAAA TACTGCAAT CTGATTGAT AAAATGAC TCGACACTG GATTTGTTA GAGCGCCTT GAGATAGG ACACGAGA GAAATGGA TGGGCATA CACGGAAG TAGAGTCA AGATTACA GGACACTCT CAAAACAC GAACGACG AATTATCGA AAGATATCT GCTCGTAA ACCTTCCAT GCTACAAA AACTGTGTA TGGGTAAG AGGTGGGC CATCCAAG GAAAAAAA AAA				
CR1	CR1-10_CQ	.	Culex quinquefasciatus	TC	TAAATTTAG TTTTAAGTK CATCATTGG GGCAAAAT AAGTGCCT	RRADFESLSRALRGIEWSSVLDPVDIDLAVDTYTXILGR LIEQFVPKIRKVVSHHLPWQTQELRXLKTQKRAAFKVS NGSVHLRDYYLRINHSYQRLSRRCFANYQRKMQKKLK TDPKKFWKFVNESRRESGLPTSMRLADXEGDDADKICQ LFAKKFAGVFCAEPITDEEVRAAAENVPRRDELLASIEID	cd01650	250	500

					GTTGAGAA GAAATAAA CAAC	EGTTQAAAAKLKHSCSPGPDGIPSTLLKRCSSSLSAPLLH LFRLSLSSGKFPVAWKQAYMFPVHKKGDRTNVENYRGI SALCATSKLLELVVIDPVFSHCRQDLADEQHGFFPKRST ATNLLCFTESIINSFDAHSQTDAIYTDLSAAFDKINHRIA VAKLERLGFRGSLLAWFKSYLTGRSLRVRIGDALSDIID ATSGIAQGSHLGPVVFLFYFNDVNYTVTGPRLVYADDL KLFARIDNSNDAEALQRDLDRFAGWCXVNRMILNPGK CQXITFSRRHFPVTFDYRINDTSVERVTHVKDLGVILDT QLTFKQHLSYIIAKASRQLGLVIRMTRKFTDIYCLKTLYC SLVRSTLEYCSTVWSPHYNNAVCRLESIQRRFVRYALR QLPWTNPLXLPPYEQRCRLIDLDTLQLRRDLARAMTSA DVLLGRIDCPDLRNQIQLRVPTRQLRHTPALEVPFRRTN YSANGAINGLKRAFNKVSSVFSLDVSRTTLRFKFVSLLR SLLRM						118 4
						RALRRIFLGADKRXQPDSAXSASSSCPTCTNKTVNAESH AKVNAETNTKVNVEVNPEIYAGHGADPKGYGTNFACE SQWFTYGSRDRLTNTLRSTPSLPQQSSVQHLPASSPGRA VESIMRAPRLPDATAPSPSYRTRSPASRSHQRRPGAGVG VGGGVSQTAPLGKYPSQLQVSLPDACPTYSSSLITSSSTR TANASPPAQAIQSSPGRTVESIMRAPRLPDATTSSFSYRT RSPAGCSHQSRPGAGVGVGGGVSQPVLFGKYPFQSQRV LPDASPTHRSCPSTSTSTRTANASPLPAQAIQSSPGRTVE SIMRAPRLPDATTSSFSYRTRSPAGCSHQSRPGAGVGVG GGVSQPVLFGKYQFSTPHTSPDARSDFSATVPPSTFVTL AAPGCVYHGTMEDLGSANSGMPAALTASDPASASSER GFLNVYYQNVRGLRTKTNSVLLALSSCDYDVIIFTETWL VPGIKNSEFAAGYDIFRCDRSAETSSLQRGGGVLVAVKT GRNCKDVELVGCGSLEQIAVRIVLPHQALFICCIYIRPAS NIDVYQLHTGAVQQILGMAGVRDLVLVVGDYNLPALT WFFDDDTHSFLPINASAEQEVLLLESMLSCGLLQINNLA SAHGNLLDLAFVSDATAAELIDPPCPLLKPDAHHKPILL KLDYDHGGNDDLDDADEIFDFNQCDYALLNERISAVD WDTLLDTSSVDESTANFYERLNSILHEVVPKKIRRMSHR YKLPWWNDELRTLRNRLRKATSRYMESKTAWDKVTV QNIEDEYNALNQESFRSYINNTQRSLKSDPSKFWSYVNS RKNANRTPTDVSYRDLNSTSPAIAANFFADFFKSVHSSD AIPPRDHEFEYVPTHYLPLPLPVLNDEETLKALSGVDGA KGPGPDGIPPSLIKACADSLTVPVKRIFNDSLRKGVYPA MWKVASITPIHKSGSTKKVENYRPISILNCLAKVLEKIVY DRLYAVVRPIISDDQHGFMQKRSTTTNLLTFVHSTVNAL	cd01650	933	

CR1	CR1-9_CQ	.	Culex quinquefasciatus	C						1157
					TAATCTGTT	ALEMFVDVPAPSATAEPPLNSQQSFNFRKANYAAIQEEL				
					TTTGTGTTT	SRIDWSFLETAPSIDAAVDAFQSILNNLFTEHVPLRRPPQ				
					CATGTTGTT	KPPWSNPTLRKLNRLKTAALQEFSNHRSPFFKSKFNLAS				
					AGCATTTGT	KRYRGYNRTLYRRYVIRTQRDLRRNPKNFWSFVNSKRK				
					CTTTAGTTA	ADGLPPTLYLDDVAAATSEEKCALFAAHFKDAFNSRFA				
					AGTTAAGTT	SEAEIIEATRNTPINAIECSLTEVANCQVSAAINKLKLSYS				
					TTATCATTA	PGPDGIPSSVIKRCSGALLSPLTCLFNLSLRQSTFPDKWK				
					AGACCACA	ASFMFPVHKKGDKSSVKNYRGITSLSACSKVFEILVNNV				
					AGTTGTCGG	LFDSCKNYISTDQHGFYPKRSVVTNLAQFVSHCLQEMD				
					ATGATTCAA	AGAQVDAVYTDLKAAFDRVDHGVLLKRLERLGVSASL				
					ATAAACAA	VRWLKSYLTNRILRVKIGNSESDMFISLSGVPQGSNLGP				
					ATAAACAA	LLFILFANELSMLLSEGCRLFYADDVKIFTVVRSVLNCS				
					ATAAACAA	ALQRQLNHFNEWCVRNFLTISIEKCSTISFHRKLKPVVFD				
					AAAAAAAAA	YAIAGNTLARVTQVRDLGVTLDSELCFTIHLHDIVSRAN				
						GESESRGREHGRVTRPPLAASPGRAVESHVRVPRLPDTT				
						TSSPSHRLIRPACCSLQSRPGAGVGAGGGVSQTATLGKY				
						XSXSQXPLPDARPTSSSCAISSTSPPVRTPNAVRAFWSSP				
						GRAVESHVRVPHLPDTTTSSPSHRLISPACCSLQSRPGAG				
						VGAGGGVSQTATLGKYHSSLQNPLPDARPTSSSCATSPP				
						VRTPNAVRAFRSSPGRAVESHVRVPRLPDTTTSSPSHRLI				
						SSACCSLQSRPGAGVGAGGGVSQTATLGKYHSSLQNPL				
						PDARPTSSSCATSPPVRTPNAVRAFWSSPGRAVESHVRV				
						PRLPDTTTSSPSHRPISPXCCSLQSRPGAGVGAGGGVSQT				
						ATLGKYPSGLQNPLSDARPTSSSAAFPSTINAKGTSTPCK				
						LRIYYQNVRGLRTKIDAFFLAVTECDYDVIVLTETWLDI				
						EILSPQLFGSTYRVFRNDRNERNSDKKSGGGVLIAISSRL				
						DCVLDPTPVCDTLEQLWVRLTVGASQRVCIGAFYLPPG				
						LNRSIDLIKDHLESIGNVLSHLNLNDSFIQFGDYNQPGLK				
						WIPSEEGGLEIDPVQSRVPLVSSYLIDGFNLHGLSQLNPC				
						VNVYNNCLDLVLVNEVALPNCAVLAAVDPLLPQDIAHP				
						ESGEQVDAIYIDFEKAFDKVPHALTVRKLRKLGLPAWIT				
						NWLLSYLTNRKAFVNLRIARSSTFDIPSGVPQGSHLGPLI				
						FILFVNEICALTNSNTLMYADDLKLFRSVKNNVDCCSLQ				
						MDVNALLHWCESNGMKVHVKKCKVISFSRSRNPLRFD				
						YTMGDGCPLDRVNSIRDLGVTVDRKMKFNEHIALTTGK				
						AFAMLGFLKRNTAHFDDPFALKVLYASLVRSVLEYAAV				
						VWAPYHVTLAARIERVQRAFVRFALRKLPWNNPLILPS				
						YEGRCLLFRLQSLAERRTLMQRLFVFDLLRGNIDCRSIR				
						NNVRFNEPARQGLRGERPLLWIPRHIFDYGYNNPLDVCF				
						RQFNEICSCFDYVVSKTVFKTRIS				

						RQLGFIFRIADDFRDVACLRSLYCALVRSILEFCSVVWCPYQSTWTAKIESVQRKFTRLALRRARNPAGWNTYEERCRVLQLETLERRRHVSQATFVAKVLAGEVDTPWIRAQIQIYTPARPLRQRQYLQLAHRDTNYGQHEPVRFMCSRFNVFSHLYDPNISTPQFQQRARLWYSHQL			
CR1	CR1-5B_CQ	.	Culex quinquefasciatus	ATACTACCAGAACGTCGGAGGC	TAATTTATGTATTAGTTTGTAAGGTAAAACACCATTGGGGCCCGAGCCTGTTGGTGTCCAAATAAATAAATA	MNTRALEYRLACSAISYDVIAVTETWLKNNTLSSQVFGNGYEVFRCDREQFINSQKKDGGGVAIAVRQGLNARIVSSDEWKCVEQVWVAIDLADHTVYICVVYIPPDQTNNASVFEINSSSVSTVSSMARPYDELVILGDFNLPRLIWRPSRDGFLYADPELSTQTLHVTEFLDNYSSALLQQINQFPNENNVLLDLCFVSCPDIAPPIAIAPAPLVKDVRHHPPLHLSLPVHLASEFTPIVCEVRYDYRKSDVPSMLELLRSIDWENTLDKDNLDDAVQTFTNIIGYVIDRHVPKKPVTNARVPWQTSELRRVKAAKRSALKKYSKHRTLPLHNYYVRLNSCYKRMSKQCHADHEVRTQKKLKSNPKGFWKYINEQRKEIGLPSSMFLGDKTASNTEDICQLFAAKFSSMFTQEDLSQQQITAAANNVLPNGLALNRITIDRSQILEAATKLKSSNTAGPDGVPSIVLKKCIDGLLDPLAHLFSLSLSTGSFPKLWKHAYMFPVHKKGDKRNADNYRGISALCATSKLFELAVMDPVFFFCKNLISDDQHGFMPSRSTTTNLLTFTTFVTDSFTAKSQTDAVYTDLSAAFDKLNHSIAIAKLERMGIGGVLLRWFQSYLTGRKISVKIGDWLSAVFEAFSGIAQGSHLGPLVFLLYYNDCIHSISVPRLSYADDMKIYCRIDNASDAQVLQQQLINFAEWCKINRMVVNPTKCSVISFSRKRNPIKFEYQICGTSIPRENCVKDLGVLLDTELTYKQHISYIVSKASRQLGFVFRTAKNFTDIYCLKALYCALVRSTLEYCCSVWSPYYENSVARIESVQRRFIRYALRSLPWRDPFRLPSYESRCELINLDTLAVRRNVCRSLLVADVLTSRVQCPAILRGLNIFAPIRTFRNSPFLRVPIRQTNYAMYSALTGLHRAFNFVASLFDFNLSRDVIKRKFLLFFRR	cd01650	499	750
RTE	RTE-1_CQ	.	Culex quinquefasciatus	TAGGGCTCCTTAGGCTAGCCACTT	TAAGTAAGTAAGT	MEPESGLRATETARRKTDWRNGTTLGMKTRTRIGTWNVLTLAQAGKLETLGREAVRLKLEILGLSEVRWPDSAEHKMPTGQVLLYSGLRSENAQRGFLLSPNARSALITWKPINERIIVARFRTRVRNLTFVQVYAPTDAADLQEKEEFYSQLSXVVNEIPKGDIRIYAGDFNAKIGSDNTDLERIMGPHGXGEMSENGELFTEFCGNQDMVIGGSLFPHRSVHKVTWVSRDGKTENQIDHICISRKWRRSLLDVRNMRSADIASDHHLVVGEIRLRVARVVRQEEKVGCRFNVQRLSNPEVXRAFVGELQARALEIPDSTVEEEWQFIKDAFAVTGENTLGMLRTQRKEWISDATWDKIEERKQAKAAIVSARTRATKARTRQ	cd01650	554	813

						TYAELEKAVKRSCRRDKRAWTDSLADEGEKAANNGDMRLLYDISRRLCGARTNTRIPVKDRSGQLITDPADQLKRWFEHFEQLLQISHPRPSPQYHPQNVRRINRVNSSPPTLREIEEAVKGMKSGKAPGIDRISAEMLKADAHLSARMLHRLFSKIWDTATFPVDWMQGILVKVPKKGDLTNCDNWRGITLLCIALKVLCKVILCRIKSKIDESLRRQQAGFRSGRSCVDHIVTLRVILEQVNEFQDSLHLVFVDYEKAFDRLDHENLWEALGRKGVPERIINLIKAQYEAFVCRVVHNGVLSDPIRVVAGVRQGCILSPLLFLIVIDEVLAGALDRVQNRGLLWQPIRMEHLNDLDFADDIVLFSQRGNDMQNKLDDLAACSSAAGLKVNVSKTKSMAVNTERAQGFTVAGENVEKVDTFQYLGSQIAPDGGTKLDISTRIKKARSAFAGLRTLWRSNQISLRTKLRIFNSNVKSVLLYGSETWCVSSQNTSKLQVFVNRCLRNILRAWWPRNWVSNAELHRRCQQRPIHVEIRERKWRWIGHTLRKDAHEICREALDWNPQGQRGRGRPRGSWRRSLHQEISAVDGTLSWRQVKAKAENRQQWKSFNAALCSTGEQXA			
Penelope	Penelope-4_CQ	.	Culex quinquefasciatus	GAATTGTGGTTTGGTTGAGCGTTTTAGCAGAATGCATTACTGTGAATACAAAGAGACGAGTTGTTCCTTTTAATTCCGCTATATATTTTAATATCTTGACAGATACGTATTTCGTCTACTACTTGCAGACTTCATCAGTGTTCTGTTCTCGACTGAAGGTTCA	TGATGAAGTCTGCAAGTAGTAGACGAAATACGTATCTGTCAAGATATTAAAAATATATAGCGGAATTAAAAGGAACAACTCGTCTCTTTGTATTCACA	MHSAKTLNQTTITMATKPSEHNKFIGLKKIIAGIHKDIAFLKKCVEHRVTPTSHKIRTKGHTDSKIIRRVELELVKESIKKLYGRLSIKTLECYSLHLKLAKEFPIEFEAFLTKVKVAEKCESERKRKILDNKFRQLISHIAPTMKPVVKPLEEFVVNRSSETFSEEQLTLLNKGLKYAVSSKPNLENIVIDIETVINANSIDEKQTKPPPLLPAQINSVRMTVSKLIREATTKHQDNAEVRVVKELKEKPVYYLKADKGNRVVIMDRDEYDKELLEKLGNREKYSLQRGYTLIDMVKKVENTIKECAGLLKSVGKTAKSLRVPNPTLPRIKALPKVHKPGNEMREIISAVDAPTSRIAKWLVEEFKSMPKPFPSRSIISSQVFTKELLESGPIAEDEIMVSFDVTALFPSIPVKNAIGTLRDWLQSQYEGEPWRQKTIQYITFVKLCMEQSYFQFRDCIYQQRMGASMGNPLSPFMSDVYMAALEHELQTKNLLPERWWRYVDDVFCIIKRDSLTTVLDTINSARRSIKFTYEMEVEGKLPFLDILILREPGCPSSFSFEIYRKPTNTRRTIPATSNHSFQHKMAAYHYFIHRMTTLPLSEAGKQKELEYIFETAEINGYSRTTIQAIIDKKERKLHRTSLTTLTPSAEPLRRAAVEFDYRITRPLRQKLAKFGIDLVFSSRNSQLESILGSTKDPIPTLGKPGIYEVSCGHCDMSYIGQTKRLLQTRLDEHIHTYVKIAMEEKRKGFVPHFKSAVTEHIIEEKHNITASDAVILRHITNPSKLAVAESLEIFKAGNKRLLNKDSGNGSTWLFKLLPIQAQKKNEEVVPTVTTNTDTPAMNLQSRTEH	cd00304	469	548

				TCGCTGG TGTATCC GTATTTG TAGTTAC TGTAGGT ACTACCT CCTCGTT CTTCTTTT GTGCCTG TATAGGT AACAGCT TAAACAG CCACGTT GAGCCGT TACCTGA ATCCTTG TTGAGTA AACGTTT GTTGCCT GCCTTGA AGATTTC CAAACTT TCGGCGA CAGCCAG CTTCGAT GGGTTAG TGATGTG TCTTAAG ATGACCG CGTCGCT AGCTGTG ATGTTAT GTTTTTCC TCGATGA TGTGTTC GGTTACT GCTGACT TGAAATG GGGGACA AATCCTT					

				TCCGTTTT TCCTCCA TGGCAAT TTTGACA TATGTGT GTATATG CTCATCC AACCTGG TTTGCAG CAATCTC TTAGTTT GTCCAAT GTAGCTC ATATCAC AGTGGCC GCAGGAT ACCTCGT AGATGCC GGGTTTG CCTAAAG TCGGTAT GGGGTCT TTCGTTG AACCTAA GATGGAT TCCAACT GGCTGTT TCTACTG CTGAACA CTAAATC GATGCCA AATTTAG CCAGTTT TTGTCGT AATGGGC GTGTTAT CCTGTAG TCGAATT CTACTGC TGCTCTTC					

				GCAGTGG CTCTGCT GATGGCG TAAGGGT TGTGAGT GAAGTTC TGTGCAG TTTTCTTT CCTTCTTA TCGATGA TTGCTTG GATAGTT GTTCTGC TGTAACC GTTGATT TCCGCTG TTTCAAA TATGTAC TCCAATT CTTTCTGT TTTCCTGC TTCGCTG AGGGGTA AAGTTGT CATCCTA TGTATAA AATAATG ATATGCT GCCATCT TATGTTG GAATGAA TGGTTTG ACGTGGC TGGGATT GTTCTTCT GGTGTTT GTTGGCT TCCTGTA GATCTCG AAAGAAA					

				ACGAAGA TGGGCAA CCTGGTT CTCTTAG GATCAGG ATGTCTA AGAAAGG TAACTTT CCTTCGA CTTCCATT TCGTAAG TGAACTT GATGCTC CTGCGTG CACTGTT TATGGTG TCCAGTA CCGTTGT TAGTGAG TCCCTTTT GATGATG CAGAAGA CGTCATC TACGTAT CTCCACC AACGTTC TGGTAGT AGGTTTT TGGTTTG TAGTTCG TGTTCCA AAGCTGC CATATAA ACATCGC TCATGAA CGGTGAC AGTGGGT TACCCAT CGATGCT CCCATCC					

TCTGTTTG
GTAGATG
CAATCCC
TGAACTG
GAAGTAG
CTCTGTTC
CATGCAT
AATTTGA
CGAATGT
TATATAC
TGGATCG
TTTTTTGT
CTCCATG
GTTCTCC
CTCGTAT
TGTGATT
GCAGCCA
ATCCCGT
AGAGTTC
CTATTGC
GTTTTTTA
CTGGTAT
GCTTGGG
AACAGTG
CTGTTAC
GTCGAAG
GAAACCA
TTATTTCA
TCCTCAG
CTATCGG
TCCCGAC
TCTAAGA
GCTCCTTT
GTGAACA
CCTGCGA
ACTGATG
ATCGAAC
GGCTCGG
GATCACT
AACCCAT

CGAAGCT
GGCTGTC
GCCGAAA
GTTTGGA
AATCTTC
AAGGCAG
GCAACAA
ACGTTTA
CTCAACA
AGGATTC
AGGTAAC
GGCTCAA
CGTGGCT
GTTTAAG
CTGTTAC
CTATACA
GGCACAA
AAGAAGA
ACGAGGA
GGTAGTA
CCTACAG
TAACTAC
AAATACG
GATACAC
CAGCGAT
GAACCTT
CAGTCGA
GAACAGA
ACACTGA
TGAAGTC
TGCAAGT
AGTAGAC
GAAATAC
GTATCTG
TCAAGAT
ATTAAAA
TATATAG
CGGAATT
AAAAGGA
ACAACTC

			GTCTCTTT GTATTCA CAGTA							
						ARNCFRRVYACNXDTANLQPAELQPKTPGPAVTPVPES				
						TESLTYDFTRCDFEAVNJALSQINWTGILAHCSLDEAVS				
						RFYDALXDVIQLHTPLRRLHPPSNRKPWWNATLRNLRN				
						RLRKARTRLHKRRCQPNIAEVRELEELYESQQNAAFSEY				
						IHRIETDVKQNPRSFWTFLKSRKKGSGIPTEISNGCVEAN				
			GGATACC			SVEEQASLFADFFSGVYSTSLPAPCNEVVDHIPTHDLHLP				
			ATCGTAG			VVHFSIEDVSRALDDVDVTKGPGPDSIHPSFLKNCAASL				
			TTGTAGG			ALPTTIIFNRSLAESTFPTAWKLASITPIHKAGNVHKVEN				
			CGATTAC			YRGISILSGFAKLFEHLVHGALYPVLKPIIANEQHGFVKR				
			AATCTGC			RSTTSNLLLFTSTLLTNIEKGQQIDAVYVDLSKAFDKVP				
			CTCACCT			HALLLEKLRRYGLPEWIVRWIHSYLADRKAFVKVRCTS				
			CAACTGG			SHLLAIPSGVPQGSVIGPLLFILFVNDLCSVLESDKLMYA				
			TCTTTCG			DDLKVFRSVSSPLDSCALQQDIDRLLRWCTANGMEVNV				
			ATTCGGA		TAAGTTGTA	QKCKIISFTRKRTPLLSSYRMGQNELERVNTIKDLGVTID				
			CATCAAC		GTCTGTGCG	SKACFNEHIALTTAKAFATLGFLRRNAADFVDLFALKTI				
			GGATACC		GCGAGTGC	YCSLVRSQLEYAVQVWAPYYAVQAERIERIQRAFTRFA				
			TGCCAAC		CGAAGATG	VRRLPWVRPQGLSSYDDNCERLKLPALASRRVLLQRIY				
			AAACGCA		AAATAAAT	VFDLLSCNIDCPTLRDKITLHIPARNLRNPAPFLEVPGHR				
	CR1-27_CQ	CR1	TCTGCTG A	Culex quinquefasciatus	AAATAAAT AAAA	TNYGYYSPLSACCRVFNDVADVFVFGMSKCVFKTRIKN RT	cd01650	297	547	
					TGACAAAG	ALKKLKSSKRAALKKFSKCPISVNRLNLNQISNDYRRLN				
					GTTTAGACC	KSLFLAHQSRVQNRLKQNPKGFWKFVNEQRKESGLPST				
					GGTTTATTT	MIRDGVEVSTSSEICDAFRVQFSSVFADNNLSDPDVAAA				
					TTAGTCAGT	ADSVPDSGLVCSDLQVSREDFNAACSKLKSSTSPGPDGV				
					TTTATTTAT	PAIVIKKCANSLSEPLLRIFNLSLSLGVFPNNWKESYVFP				
					TTATTTATT	VFKKGCKQDVKNYRGIAALCAVSKLFEVIVLEFLRHNL				
					TATTTTTTA	SHVVSDDQHGFFPKRSTNTNLVTYISKIQRAISDGYQVD				
					CATTGTGCT	AIYTDLSAAFDKINHRIAIAKLRRVGLHGSLLTWVNSYL				
					AGTTTAGTT	TGRTMLVKISDNLSLPFSVTSGVPQGSHLGPFLFLLYIND				
					TGWTAGCT	VNFLLRCPKLCYADDFKLYAIIKRPTDMEDLQAQVDLF				
					TAAGTTTGA	ADWCRLNQMVLNPSKCSTITFTRKRCPTFFSYKILGNPL				
					TTGTATATA	TRDSFVKDLGVLLDSKLSFKNHISYLCSKASKQLGMIFR				
					TCCTCTATG	MTKYFRDVNCLKILYLSLVRSTLEYGSVVWAPYYHNDI				
					TCATTTGGA	GRIEAIQRKFTRFALRHLGECPNYETRCSMLHLDPLSVR				
	CR1-23_CQ	CR1		Culex quinquefasciatus	TTGTAAAAT CTGTTGATA ATAAAAAA	REAAKAFFVADLFRSNINCPFLTSELNVNIRRRVLRSHSF LVIPGARTNYAFNEPVSSMCRVFERCYPFFDFHLSRPRL KHDIXNFLREAYVF	cd01650	190	441	

Kiri	Kiri-5_CQ	.	Culex quinquefasciatus	ATATTTTTT	GTATATTGT	RCKEHLKNSI	cd01650	519	778
				ATTTTTCT	GTATACAA	MNVDTERHLQVCFNSCIRFIYNLRRFDHVSQFQRSIFDLS			
				TACGCGT	TAAGTCAAT	YSTLRTLNCHRHTLSIDTKIKLVHTLLMPHFIYGNVIFSK			
				TCTTTTCC	CTTATCATT	PIQIDNENVEIVDECLNLGIKMNKHLTWXDHVSLIKKKV			
				CTACATA	AAAAGGTT	LSKFYQFQLLLFLFKLINFKEPQYLFESLVFSRSNRTNNIII			
				TGATTGC	ATAAGGTC	NSDLSAISRYCKDNGLVLNTSKTKAIXFSRQNTVLPHLP			
				ATCTTTTG	TTTAGTTTT	KSYLSDRSQLVQINGVRSQFRSTKRGVPQGSVFGPLAYT			
				CACTATT	TTGTTAGTT	YVTVLILLLDFTKAFDLIDHRIFIDKLKEQFQFSSDACDLM			
				AAAACAG	TTAATTTAG	QYVTFKKLITPFQSGFREHHSTSTIITKVCQDINASFDSN			
				GAAAACC	TTTTAGTTT	KLAIVKPIPKIKNPSNPNDLRPISILPSLSKGCEILLKDQIV			
				GTTTATCT	CTTATTTCG	SDAIGYDEIPLKFVKLFLDLLVPYITLLYNKIISTNTFPTIY			
				TTGTTATT	ACATTTTTA	LPSKTIWKNFKSLGFRDDPDNLDDKFKPEDVNNCFIQKI			
				GTTTATCT	TATCTTTAG	ENYPTGNRQPTIIPDIVEEDQFKFRAVTEDEVHKAILRIK			
				GAAAACC	TTTTAGTTT	KLAIVKPIPKIKNPSNPNDLRPISILPSLSKGCEILLKDQIV			
				GTTGTTA	TAATTTGTC	LVNKHAPLKTKTILPTSCPWYSREAYNACVDRDLAYRF			
				GATTTTTT	TAAACAAG	NRIDKEAVRNDFPNFNFDQLNGIVDCETLASNFNHKLIQ			
				AGTACAG	AATTCTACT	XSTLFSQVSVPAVSHHDLIYTVFNLPVPVASLKTISFRNF			
				GCCCAGC	TTATTACTG	SNNFTRLINSFSLSIVNKSFPTHFHNTGSSLLDYFICSENL			
				TAGATTT	GTCTTAAAT	KPPTLXDFSFFFEDLGNISSTYNNILMGDFNINILVSNPM			
				ATGATAG	GATAATCA	IEYMFCELKVVSCKLWLVLSINPQPIKSSQLQVVVGVIY			
				GATTTAG	ACACAGCT	NFEILRHDRNSTTRVXGGGVALFVRPEYKTKIIKTSPRDS			
				TTGTTTG	TAACTACAT	SIICHHDDVKNIFTRTGFDCIGFTETWLNSDIKDALIKFD			
				ACTCTTTA	MTNTVNNVQELRFEPINATLKNNLSKENGFKVGHLNTC				
				CAATGCCTT	TTTGGTTT				
				GTATTACAC	TTGATTAAG				
				AACA	TGAGTTTTG				
				AACATAA					
				AAACATAA					
				CCTCATATA					
				GGGCTTTTC					
				ACTCAAAC					
				AGGCGTTCC					
				CCCATCGGT					
				GAGAAGGA					
				TGCCTATTT					
				GAGGTTTTG					

						TGAAAAAT AAAATACA AATACAAA TACAAATA CAAT					
						TAGAAACA TCCAAAGTT AAATTATCT CTACCGAG CTTATTTGT CGTCTTATT AGCAAGTG GTGAAATG TAAATTCAA GAAAAATG TTGTAAAAT AAAGTGAA AAAAAAA	LYENIFGCLIANNFVKVVAYADDLNILVRNNHEFDTVL QLVSYFSIYSKIKLNFAKSQYLRFNGCLSGPHQIKEVESL NVLGVLICQNFNKLVEINYDSYLNKLKYTLSLHQKRILN LFQKAWLLNTYVLSKIWYICQIFPPNNKHLAAIKSICGQF IWQGFIFKVPRNELYLPVIKGGLALTDVESKAKSLFIKNI LYSNADGNAPLDNFMLGQINNSTLTRNMREWIVKADEL KLESHLNSSKKIYDFFIHSLNIKCKTMEDMPNLMWSNLL NNLDSNFLNSKCKTVLFCIFKDIVLNKKKLHLYGIRGTD SKFCENCGLVESVSHLVKYCQTSEEIWRWLKNILNKRL KILVKDPDELLSLSIGDRDFKYKAALWLTVQTIVYNLEK RGFGDLNEFKQGIRIVRFNNKAIFERHFKHFLNIF	IPR0 00477	11	83
Tx1	Tx1- 11_C Q	.	Culex quinquef asciatus	TG	A		IXRTQRRXFLGXFKCCREHSGKSATPPVPTLNDADDAEF AARVCTAIDSKDTGACFGVDWRSGRFTHGSSDRITCGTI VPTIPASFRSSSQHLPVRVLQSSPGRAVESLMRAPRLPDA TTSSPRQRLRTPASRSHQSRPGAGVGVGGGVSQPAPFGK YPSRSQHFLPDVRPTSSSCASTSATQTAIVSPPSFRAIQFS PGRAVESPMGAPRLPDATTSSPRQRLRTPASRSHQSRPG AGVGVGGGVSQPAPFGKYPSRSQHFLPDVRPTSSSCAST SATQTAIVSPPSFRAIQFSPGRAVESPMGAPRLPDATTSS PRQRLRTPASRSHQSRPGAGVGVGGGVSQPAPFGKYPS RSQHFLPDVRPTSSSCASTSATQTAIVSPPSFRAIQFSPGR AVESPMGAPRLPDATTSSPRQRLRTPASRSHQSRPGAGV GVGGGVSQPAPFGKYPPRSQHFLPDVRPTSSSCAGPSAT QIANASTLSATVPSQLANDVQPQRPGSVDRRDSSLLTIY YQNVRGLRTKTNQLHLSLSSCDYDVIAFSETWLNAKIV DSELSSDYTFYRADRSRETSVLERGGGVLIGVKKRFHPE PIHLRGGEKLEQIAVRIPLPGKTLFVCCVYIPPNSDATIYT QHSSCVQQLCDLAQGPDEVIVVGDYNVPNLTWHFDEDI NGYLPSGQSTEQEVALIEDVLGTGLQQMFDLPNANGRL LDLALVSDSSRFQLIEPPRAILKVDAHHRPFILVLETQST GTAAEQSGTYYDFANCDTMALNNSIASLDWANLLLVG SVDDIVERFYDKLNEIIREVVPLKHRRRPHPNSHQPWWNR ETRNLRNRLRKARKRFLKHPTEMLGRELADRELHYEMS	cd016 50	101 1	126 1

RTE	RTE-2_CQ	.	Culex quinquefasciatus	GACCCGC AATCGAC GGCTGGG	TGATTGGTA TGGTATGGT ATG	MARRLEIRNKRQKPIEKQTTDWRLGSWNCRSLNFLGFE FALAXELQPRNFDVVALQEVCLEEEQVLDWPGRQTNSR FXLSGGKDKKLGTGFIVRGKMQDRVFGFTAXSERMCK LRIRGRFFNYSIINVHCPHEEKTDDEKEAFYATLEEVYD GCPRQDVKIIIGDMNARFGREEMFRPTIGPESLHAVTND NGQRCIDFAASRGMVVRSTYFPRKDIHKATWTSPDQRT KTQIDHVLIDGRFFSDVTHVRTFRGADIDSDHYLVGVD MRSKLSTVFNQRRSRRAPPFNTACLQNGEVAHSYAQQL EANLPGEEELGAASLEDGWSRIRSAIGSAAEATLGSAIR VSRNDWYDDECQRITAEKKAAYDRKLHKATRGNVERY RQARNRQVAVFKLKKRQQEDRDCAEMEQLFRANETRK FYEKVNRSRKGFVPRADVCRDNGGNLIVNKSEVLDRW KQYFNEHLNGDEADGDGVGVNLGAPAADEQFPAPDLE TVKREIRKLKNNRAAGKDRLPGELFKYGGEKLARALHC VISKIWEEEKLPEEWMDGVVVCPIYKKGDKLDCGNYRGI TLINAAYKILSQILCRQLSPHARRFVGPYQAGFTGARAT TDQIFCLRQILEKCREYNVPTHHIFIDFKAAYDTVDREQL WQIMHENGFPDKLTRLIKATLDGVMCHVRVSGXLAEPF GSRRGLRQGDGLSCXLFNIVLEGIIRRAGIETSGTIMNKS YQLLAFADDIDIVARNLETVMDVYTRLKVQARRVGLG MNTTKTKYMRGRGSKDVGPPSLTPLTVDGDELEEVSEF VYLGSLVTADNDTSKEIRTRIFAGNRAYFGLRKTLTSDR VQRRTKLTMYRTLIRPVVLYGHETWTMRQEDERALGV FERKVLRTIYGGVQVSDGVWRRRMNHELHAFLEEPTIA	cd01650	547	807

				CAATAAAAGTGAAGTTATTTTTGCCGTCCATCGTCGTCGTCGCGGGAAAACAGGCAGAACAAAGCAAGAAGTGTGGTGTAGCAACAAGACCAGTGGCAGTGCCGTTGTGTCGGTTGCGTATACGCACACTAATCGAAACCGCGCGCGGTGTGCCGAGAGTGCGAATTTTTGAAAAAAGAAAACGGGCAAAAAGTTTTGAACTTTTTTGCAAAAAATAAAATAAATTTTATTATTGCGAAATTTGGCGCTAGTAGTGAAGCGCATTTACGCACGCAGCGGAGG		TQAKIGRLRWAGHVARMDESQPVRLLFDRAEPAGGTGRRAGKPRARWGDQVNGDIRKICNLGNWRVAAQDRERWKRLLATARVPGALC			

				ACGTCCT CTTCCGG ATATGAA GACCGTT AGGCTTC GAGATAA AAGCCCC GTTACGG TAKCGTG AGAAAAC CGGGGCT AGGTCTC GGTAGCC CGTAGGT GGTTCTC CACCCCG AGACTAG GTAGCGC AGCCCTG ATAAGGC TGCCTAC CGAAAAA GAGGATC GCGA					
Rex 1	Rex1-1_CCa	.	Cyprinus carpio	CAAG	TGAACCCCC CCACCCCCC CTCCCTCTA ACATACGA TGTCATGCA CCAGTCACT TTGTGCAGC AGTGGTCTG CTCACTACC TCATTCACC ATGGAACT GACGTCATT CTACCACCT CATCAGTCA GTTTAAATA AAATAACT	MAARPHAAASLCPDRTVFSCFLSCESQCFCTSSSRLSVF KRAYSREFLLDVGRAAFLELNSAHAEQLRDLGLLRRPT SSPTPTTASRPQRRRHKRRERKQKRGKRGGIRARLTANP HKPSIPTIVLANVRSLDNKLDYIRLLRSTQRTVRDCCVFV FTETWLSDSVPDCAIQLDQLTCYRADRARVAGGKTRGG GLCVYINDAWCRDAVVICKHCSPLVEFMIIKCRPFYLPR EYTAILLVSVYIPPNNNNSNRNDALNELYQHISEQQTAH PDAFLIIAGDFNHADLKSVFPKIHQHINFPTRGNNILDFV YTTQRGAYKALPLPHLGASDHITVMLMPAYRPLVKVA KPVQKQIQVWPEGSSDALQDCFDTTDWNMFKQAATYN NXTDLQEYSETVTAYITKCIEDVTVTKTITVRANQKPW MTGEVYRLLETRNAAFRAGDEGGLRTARANLSRGIREA KRQYSRRIAHQFSDSRDTRSLWQGIQTITDYKPPQRTCD SNISLLNELNTFFARFEAQNSTTAQKTPPPPSDQVMMLT PDSVRRSFSRINARKAPGPDNIPGRVLRDCAAELTDVFT DIFNISLSQAVVPTCFKATTIIPVPKKPSPSCFNDYRPVAL	cd01650	597	850

					GCTCTTGAG CCCTTTGTC ACTTTGTCA CTTTTAATC AGACTGAA TAAGCTATT TTTTTATGC ACTAAAAA TCTTTTTAT CTGCACTGT TTGTTCACT GGTTTGCAC TCTATCTGC CATGTGCCT TGCGCTGCT TTTATTTAA CCTTATTTT TATTTTATT TTTTCTATT ATGTCTTTT TTACATTCC CTTATTGTA TAGTTTTAT CTTATATTT TATATTTGA TCTAGATTT TAGGCTCTA CTGTTAGTG TTATCTGTA TGCACCGG GGGTCTGA GAGTAACG CAATTTCGA TTCTCTGTA TGTATGTAC TGTACATGT GGAAGAAT TGACAATA AAGCAGAC TTGACTTGA	TPILMKCFERLVMHHIKSALPPSLDPFQFAYRSNRSTDD AISTTLHSALTHLEKKDSYVRMLFIDFSSAFNTIIPQQLIH KLVELGLNTSLCNWLLDFLTGRPQAVRVGSNTSSTITLN TGAPQGCVLSPLLFTLLTHDCTPSHNSNLFIKFADDTTV VGLISNRDETNYRSEVSRLAGWCSDNNLSLNVEKTKEIV VDFRRAHTQHVPLTINGATVERVSSTKFLGVHITEDLSW TDNTAALAKKSQQRLYFLRKLRRARAPPPIMYTFYRGTI ESILTSCITVWYGACNASCRKTLQRIVRAAEKIIGVSLPS LQDIYGTRLTRKALCIAGDPTHPSHSFFSLLPSGRRLRSL QARTSRLKDSFIHQAVRKLNSLPNLPPRPSSASGTTEL			

RTE	RTE-2_DP1	.	Danaus plexippus	ATCCAGA GGGTGTC CTACTAA GTATTTTTT CCGATAT TTAGCCT TGGGCAT ACGGGTA CGCGACC CCGGATC CTCGAGC CACCACT TTCATAC CTCTATC CTCAGTC CATGAAG GTGGTGT CCTGTAC CTCCCCT CATTCCC CTTCCTA ACCTCTTT CCCCCCT TCTCTATC TTTTCCTT TCCTGGT TTTACCC GTAAAAC GGGGTTT TGGAGGT CAGACGG CAGTCGC	TAATGCTGG GATAATTGC CAGGAAGA AGAAGA	MVAKVYGLGTARASPINDAQAHRPTPAISGQGLPCQVR ARLKTLVPQRKLRFATWNIGSLTGRCRELADVLMRRRV QCAFLQETRWKGNKSRNIGQGYRLIYTGSPSGKAGVAV VLSEELRNGLLEVDRRCDRLMRVRVLIEGVITNLISAYT PQAGCSGSEKESFWEQFEEVLRAIPAAEVIIVGGDLNGH VGRAAETFDRVHGGFGYGRRNAEGENILRTCIASDLAV VNTFFQKTPQHLITYKSGSHSTQIDYLLTRRCHISKVTNC KVIPGESLTAQHRLLVMDYVVTPKKKVAEKRKPRIRW WLLNGTMQTSFRAEVESQNLSTNTETAQEVWDRAQSAI ITAGKRVLGLSKGGRVIDKETWWWNDEVQEVIREKKT AFKKWQQSNSPEDRLEYIAAKRASKRAVARARSDRLSP LYDTLETAEGQKLIYKLARARDKATQDIAKCLSVKDSQ GTLLCNHASVKERWRCYFKELLNTQHPCSLPTEIPPNLG LIAPITPDETRNCLRRMKNRKAVGPDDIPIEAWKSLGSL GVLILTDLFNRVLNTGTMPHQWRYSYITPIYKGRGSVQ DCGSYRGVKIMSHTMKLFERMIDLRLRRECTVSECQYG FQPGSGTLDAIFAIRTLMEAYREKRRALHVAFLDLQKAF DCVPRQCIWWALRFKGIPEAYIDIIRDMYRDSVSMVRTA VGDTKPFPISVGLHQGSALSPFLFNVVLDTVSANIQDQPP WLMMYADDIALIDESRLTLERRVNLWKGTLENGGLKL NVTKTEYMACGSPDSCTIHIGPEPAVKSEKFRYLGSILHE SGGIDHDVQARISAAWAKWRDVTGVVCDRRIPTKLKGI IYKSIIRPVLLYGSECWPTLSRHTQELHVTEMKMLRWM CGVTRADRIRNTFIRGSLGVRDVADKLQESRLRWYGHV ARRPENYVGKICLDMSVPGARPPGRPRKRWLDTVKQD MRANGLTTADAKDRAKWRSLSRKADPG	cd01650	561	
	Rex1-6_DR	.	Danio rerio	CCAAGAT GGCGGCG	TAAATTGA GGGTGTTTC	GHSENTGSPWFVIPGNNGRRREQKAETRKKSGCVDDSG KPHTSHRSPASSYECQIIANKMDELGFSWTMQGLLPLM	cd01650	546	799

				CGCATTG ACGCAGC GGCTCAG AGCTCTC CTTCTCA GTGCTTT GTTTGTTT GTTTTTAT ATGTCTG TTTAATTT TTTTAAC CTCAGCC AGTCAGG CGAAAAT ATTTTAT AGTCGAC AAGACCT CCTCTCA GTGGGAT TACAGAG CAAAGAG ACAGTTA CAACAAT GTTTATTC AGAACCA CAGCATA CCTGA	TGGCCAAA ATTATTTA ATCTTTTTA ACTTGCTGC TAGGATAT GCACTTGCC ACTTTATAC ACTTGGAC ATTAAATTG TATGCCTTC CTGGCTCAC ATTTTATTA ACTTTTAAA TTTGAAAA GCTGCTTGG AAATGCAC TTGCCACTT TATATGCAC ATCATACGC ACTTTTAAT ATTGTATTG TAACTCTTC ATTTGATTT TATATTTTT ATCTCTTTT TAATTATTG TTGTTGTTT TAGATAAA TAGGACTG CACTGAAG GGAAGGGC AAATGTAA ATTTCGTTG TATTTTGTA CAATGACA ATAAAGAT TCTGATTCT GATATT	LTESWLHPGIPDAAVELTDFTLHRHDRNHLSQKQRGGG LCIYISEKWAKDTKAVESLLPRFGVSDSNHCSPDLEYLT VKCRPFFLPREFSSAVFITAVYIAPDANANSALNLLHNAI SSQLNKYPEAAHIIAGDFNHVDLKIVFPKLHQHVNCATR GGNLLDKVYSNIKRGYKTIPAPHIGQSDHISLFMLPVYSS LRKRDPPTLKTVTVWPEGASLQLQDCFAHTDWSVFESQ DLEEYTTSVLSYIKHCTDTVTTEKHIRVYANQKPWMTG GLGFCALYRHCHHRETHQSVCQPKTLDDRGVRILLKER NIAFKSGDKERYSAARANLKTGIRGAKLKYKHKIEGYL HSNNSRQVWQGIQNITNYRPNPRSVGVDNTLAEELNCF FARFEQQSAEEATLHSLSNPSHVFTVEEDDVRKVLRTVN PRKAAGPDGITGRVLKDCADELTGVFTKIFNLSLSQAIV PLCLKASTIIPVPKKTTINSLNDYRPVALTPVVMKCFEKL VRSHILQRLPPAFDAHQFAYRENRSTEDAIAIALHTTLTH LEHKDSYARILFIDFSSAFNTIIPKRLVSKLNHLGFPFSICL WIQDFLSNRSQRVKVEHHVSSALSLSTGSPQGCVLSPLL YTLYTHDCTPFHSSNTILKFADDIAVVGLITDGDETAYR DEVNQLALWCTDNNLALNTLKTKEIIVDFRRNRPDTLPI HIHGDCVEIVSNFKYLGLHLDQNISWKTNTIALVKKAQ QRLHFLRVLKNNNLAQELLVSFYRCTIESVLSYCIPVWY ASCTAQEKAALQRVVNTAHKIIGCTLPSLEELFKARCYS KACSILKDSSHPAYHLFQMLPSGRRYRTIKAKTNRLKNS TYPRAVAIMNGH			

Rex1	REX1-1_DR	.	Danio rerio	TCGATCCAAG	TGACTCTTT TGCCCCCCC AATACACC CCCCACTCT CCTCTAACT TAAACTCCT CGCAATAA CTGCACTGT TTAACATTT GCACATTTA AAATTTGCA CATTCACTG CACCACATT GAACTGTTT ACTTATTGA ACTGTACAT ACCCACTGC ATATGGGC ATTTGTAAT TATGTACAC ACCCACTAT ACATATAC ACTTGTAAT CATGCTTAT TTATCTGCA TACTACTGA TTATTAATA GCAACCTGT ACATATATT CATATATTG TAACATATA CACTTGTAA TCATGTTTA TCTATCCCT GTACATATA TTCATATAT TGTAACATA TACACTTGT AATCATGTT TATCTATCT	MAARAQREAQGLSSFCIFAVLFLLISGLFVQNSGAFTSY TRQELLDIGLCIPDSFISNLRLIHEIARTPRAGRPARPGGS ARRPRRERKQRRGKRGGLRAKLKLTPHRLSLPSIFLANV RSLVNKMDEIRLRINHSKRLWNCNVMIFTETWLNSGIPD NAVSLTEHHTFRADRTADDSGKTRGGGLCIYINKAWCT NSVVVGRHCSANLEFLMVKCRPFYLPREFTSTIISAAYIP PDADAKLAMNELHAAISKQQTAHPEAAFIVAGDFNHSN LKTVLPKFHQNILCHTRGNKTLDHVYTNIAEAYAVTPLP HLGQSDHLSLFLTPKYSPLIDRVKPSVRTIKVWPAGVDS TLQDRFQHTDWGMFASQATHGSHTDIDSYTSSVLEYIN TTIDSVTTQKQITTYPNQKPWMNKEVRLLLKARNTAFR SGDAQAYSTSRANLKRGIKKAKHCYKLKLEEHFSNSDP RRMWQGIQAITNYKPSQSTPTATDVSFLNELNDFYARF DRDNKEPYTRITSSTDHSPITLTSSEVHTALSRINGRKAA GPDGIPGRVLKVCAEQLAGVFTDIFNLSLNLAAVPTCFK TTSIVPVPKHSSPTCLNDYRPVALTPIIMKCFERLVLAHL KDSLPSTLDPHQFAYRGNRSTDDAVSIALHSVLTHLDNK NTYARMLFVDFSSAFNTVIPSKLLIKLRDLDIDTSLCNWI MDFLTNRPQNVRSGHICSTTVTLNTGVPQGCVLSPFLYS LFTVDCRPVNRSNTIIKFADDTTVIGLISNNDETAYREEI QHLATWCTDNNLLLNTNKTKELIVDFRKGRTGSHDPIHI NGMAVEPVSSFKFLGTHISKDLSWTTNTSSLIKKAHQRL FFLRQLKKNQLSSAVLVNFYRCTIESILTNCVTVWYGSC SVAERKALQRVVVKTAQRITGTTLPAIEDIQKKHCLRRAR SILKDTFHPAHRQFSLLPSGRRFRLPRTKTSRLRNSFFPR AVSLLNSAPH	cd01650	584	838

CR1	CR1-10_DR	.	Danio rerio	ATTGTAT AGCTTTG AGGCGTA ATGAATA GTGCGAC AAATAGA GTTAAAC AACCTCT TTGCTTG CCATTTT	TAATATATT CAAAAAAC ATTTAAAA ATCTGGTTT ATAGAGAA TCAGACGT GTCATCATT CCGATTAA AATGAACC TAGCATGTT	GCACACTA CTGATTATT AATAGCAA CCTGCACAT ATATTCATA TATTGTAAA TCTGTTCAT AGCTTATCC AACCTGTAT ATAATGTTC ATAGTACAT CCATCTGTA AATATCACC ATAGTTTTC TATAACTGC ACTTTATAA CTTATTCCT GTATCCTGC ACTTGCTGC TATTGCACT GCTGGTTAG ACCTAAACT GCATTTCGT TGCATTGTA CTTGTACAT GTGTAATG ACAATAAA GTTGAATCT AATCTAATC TAATCT	MATGGTGVVITLLVYCYFTSLTAAIRVDPVRMFCSIYPL SYHNDNPHIYKDLGYKRKGNELAKTSTGYGTYALAVT MVYPSLYNKITKKKKAMTTTQSSRTGLIIAICLMLSGDI HPCPGPHHFKTPREANPQTEDKRATLTLQVCSSCTLLCS SAQQANPLLSLCPTPNTLKLSVQGVECVKGAAAVPGWS DGERQGEKTHPRLCVGMSDRSFGREFAGRVCAFRGNK RPDPETQQQQQTFLNDSASTMATNSVNRKRLNPGVVK NRKHAFFNTVVNNSKIVWDPKMKPKGLLGRHLNIRSILS KSEQITHLLTNSNLDYLCLSETWLYKYAPSAAFNISGYT MFRKDREGAKGGGVMIYVKDNLQCEEINWLCDYEFEC	cd01650	800	1056

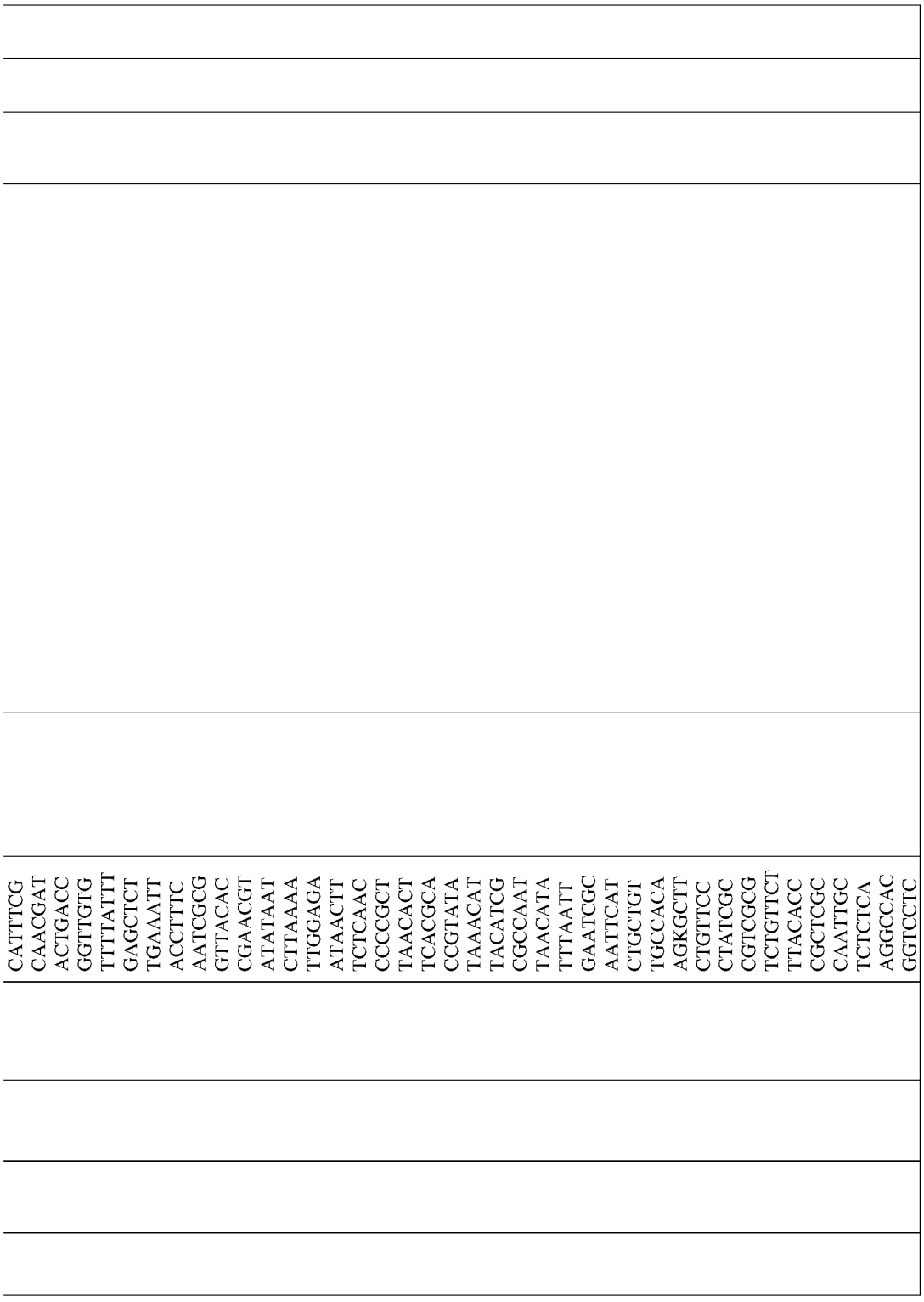
L2	L2-1_DL	.	Dicentrarchus labrax	T					
					AAGACTT	TCCAATTGT	LGLNVRLSPQMSFKIIVIYRPPSSDGTFYDRFEKLLKESN		
					GCTGTCT	GCCYACCA	LNKETVILGDFNINWEDTSCRNKLKRITDKFDLVQMING		
					TTTTTTAG	TCTTAATGA	PTKITTSSSTQIDLIFTNRPERIIKTYNMLAGISDHNLTLV		
					TCTCTGT	TCTGTTGTT	ARKLSKKRFKFSAREHEFIGIPRNRRDGFKNAVEQIDWD		
					GTGCAAA	TAAATGTGT	RLLLGTDLEKDRGSFVGTIEKTVKNSALNLNANTIKIYY		
					AACGTCT	TTCAACTAT	PGLTLTYHTKTNERERFCVKKVIKSKSASDRNMFTTLRN		
					TGCCTCC	TTTTATCTC	RVVKNIRKAKADYFLTLIEKAKGDSKTIWNQLKKVVRH		
					TATTCAG	TATTACAAA	DTKNKKMLEVMIDGHITNNACKIAEAFNNFLLTQFQIW		
					CGTTGGC	ATGTTGGTG	SHCVHQKFLTCALFFINSVSNMVPLCPSKISNMCSLNSS		
					GCTGTGG	TTTTAGYTT	QPIFSISNISESDALQAIKSLKSSRTKETFGIDSLMLKELSS		
					CTTAGTT	AATTTGTAA	TLLTPITKLLNRSISEAIFPCVWKSSAVIPIFKNGDHYSLS		
					GGTCAAA	ATATTCTCT	NYRPISIIPTVSKVAEKLIANQLIKHLNTTTFSLHPMQFGF		
					GTGCCTG	TTTAACATC	RSQHSTETANCFLIENIKHLLDKGGVVGAVILDLKKAFD		
					TCTGGTA	TAYATRACC	TVNHKILLSKLTKFNMSPESLKWFESYLSNRSQSVRINN		
					AGAAACT	ATGTCTATT	YHSSPLSLPTGVPQGSILGPLLFSLYINDLPSVCQNILIQM		
					TGGCTGA	TTAGTACTT	YADETVIXTHGKNTTQVAEKLTQAMAHVSEWLNQSCL		
					GGCCAAG	TTTAACTAT	KLNVSKTVAMYFSKTNIVNNLTDIYVDGDKLQIVSDFK		
						GTTTGATW	YLGVQIDSKLTFKSQIKKVCNRXKFNIRSHLSLQAAXIF		
						GATTGTATA	MHSMILSHITYCLTTWSQANKTTLKPIESLYKQSIKILXK		
						ATTCTTGTA	KPLRHHHCSILQKHKLLSWENVIKYSNICLLYKIIHGLSS		
						TTGTTTTTT	PPLSQFVNIRNSTHRATRGAARGELYRPTPEKYLWSGCL		
						TTAGGTTGT	FSYSCTGVEFCPWQHQRYNLL		
						TCAGGGAC			
						AACAGATA			
						AAAAATAG			
						CCTTCKGGC			
						TAATTCTGG			
						TGCATTTGC			
						AGAAATGC			
						TAATTAATG			
						TACACTGTC			
						CCTGCCAA			
						ATAAACAA			
						ATAAATAA			
						ATAAATAA			
						AT			
					TAGGGCTG	VRNHSINFHCYADDTQLYLSIKPDEXNQLSKLQTCLKDI			
					GATCAGGT	KSWMSCNFLMLNSEKNEILVLGPKNLRDSLSKDIVTLD			
					GAGTCCTG	GITLASSTTVRNLGVIFDQDLSFNSHIKQTSRTAFFHLRNI	IPR0		
					AACCATCCC	TKIRHILSQTDAEKLIHAFVTSRLDYCNSLLSGCPNKSIK	00477	5	93

						TLQLIQNAAARVLTGTRKRDHISPVLASLHWLPVKCRIE				
						FKILLLTYKALHGQAPSYLKELIVPYYPXRTLRSLNAGL				
						LVVPIVSKSRTGARAFSYQAPLLWNQLPVVVREADTLX				
						TFKSRLKTFLFDKAYS				
				TTAGTTATG						
				CTGCTATAG						
				GCCTAGACT						
				ATCGGGGG						
				ATTTCCCAT						
				GGTGCACT						
				GAGCTCCTC						
				TCTTCCTCT						
				CTCTTTCTG						
				CACTCATTC						
				ATGTCCCAT						
				TAATGCATG						
				TTACTAACT						
				TCACTTCTT						
				CCCCGGAG						
				TTCTTGTGC						
				TTTCTCGTC						
				TCGCAGGTT						
				CTCATGGAT						
				CCTGGTGG						
				AGCCTCCTG						
				CCATGGTCC						
				TGCTTGACT						
				GCCATT						
					AGGCAAT	TGACCAAA	MKAFMRQTRKALTEFRKQLTSSFEVFSIMETQFLDCKLL			
					CGATTTG	AAGACGAT	NESPKRKIANNNKNSLQIPSPCIAAVGIPSPISACLNSILA			
					AGGGGAA	TTAAGATTG	VNNSAGPGTSGLQSSFAFNNTSPGSEPGASGLQPNQNTG			
					AAAATCA	TGTTGATGG	TTAEGSSTPQQLAPGPGSLHINPDHATTIAPAALPSSQPA			
					CTCTACC	GTTTCCGCA	ANQPLSAIPLSGVAPRSNAADLPLRGIAPNRSVFVSRLNP			
					TCAAAGC	TTAGGATTA	EATADCVKSYLCQKLGAQPNSISVFKFNFKQPRNISSFKI			
					CACAGAA	TTAGGATTC	TLPESMFSNALDPTIWPEHTIMHEYIRKGASQDPNSVDA			
					AAATAAA	CCACAATGT	KNCKKLNRSISVAYQNVRSLLGKLKRVYSNSTLLDFDII			
					ACCATAC	TTTGGAGCA	AITETWLNTTVYDQEIFVNKFAIYRLDRPTRGGGVLIAV			
					ATGGAAC	ATTTCTTGA	LNSHSSELIPYDNMHSIEFIAVKVTTDKCNVFLTCSYIPP			
					TGCCAAA	TCTGCAGCG	GSDPAIYLHHLSAISAVSLAVGDADQLIVLGDFNLPHLS			
					ATTCGAC	AAGTGTTTG	WLPNSDSNILFPENQHEFIDGLTNLSLGQVNSIKNSRERI			
					GGTAMCT	CCAAATGCT	LDLIFVSDVSVATVSRCDPVSTPEDSYHPALYLSVHINFE			
					ACGTCGA	GTATCAGC	TTNQSASQRGNTIRDNYESICLTTREHYTGLNSHFDRLN			
					GTTCAAA	AATTTCAGC	WSTLYSFTDLDSAVLWFNESVCSALDTFVPSYRWKLIK	cd01650	775	1027
					ATTATTA	ATACGTTAA	ASLVLQAFVLHIPLSQVTVGNSLKPPWFSRHLSYIKNTK			
CR1	CR1-3_DBi	.	Drosophila biarmipes							

CTACTAT	CTGAGAGA	SLLKDIKNPRFFYAEYIAECRQNIRSDPKKFYSFVNSKRK	
TCAATGT	CATAATTAG	SNVFPPCLHLGNLSESSDINISNLFATFFQSTYSLPHINNI	
CAAAAAA	CGTCAAAT	HIPPFKISDVLNGMLTIDESFSAGPDKIPSCIIKQCAASLC	
ATCAATA	GTATTGACT	GPITFLFNLSLQLSFFPSVWKTSFLIPLHKKGSKTDVGNY	
ATTTTAT	TATGTATTC	RGIAKLSAIPKLFERLVTCQIQHMCSSIISTSQHGFVKKRS	
ATCCGCT	TTATGTTAC	PSTNLLEFSRLIHCSFLQKKQADVIFTDFSKAFDSVNHSL	
TTGAAAA	CTTGCAAAC	LIYKLSLIGFPPYLLDWISSYLSNRSQRVLFKNVVSSLVN	
GGAASAG	CATTTACTT	VTSGVPQATSHLGPLLFILFINDLPQVMSYCTTFMYSDD	
TAAAGGC	AATCTCGTT	VKICMSFSNLSSHLLIQSDLNKFQAWASTKLLSLNYSKC	
TTACCAA	TAAAAATT	KTMSFSRTTFHTTSYTLGNSLLERVVKINDLGVLFDQKL	
ATCAMGT	ATCTGATTT	CFNSHISVMVNKATGVLAFIKRWAKEFNDPYTTKSLYV	
CGACCAT	TGACAAAA	SLVRPIMEYCSEVWNPQYVYQQNLIESVQKQFLLFALR	
GAATAGG	TTAATGATA	NLNWDPDRHLPPHLSRLKLIDLNTLKHREFNCKIF	
GGGTATG	ATTGTTATA		
TAATCCG	AGCCGAGG		
TAATGCA	TTTGATAAT		
AGCGCCA	TAACCAAA		
TCCGCCG	TGCAATCCA		
CAGTGTT	GTTGGGCTT		
ACGATCT	TCCAGTTTC		
CTCTTCG	GAATCCGA		
ACGCCCT	ATGAATTTA		
TACTAGT	GTTGTGTTC		
TTTTACG	AGAAGAAT		
GATTTCTT	TATTGCACA		
CTCTCGG	ACTAAAGA		
CAACGAA	TGTTATGAA		
AAGTTGG	AAAATAGA		
AAAAAGT	TTCAT		
CATGTCC			
AGGTTAG			
ATCGAAA			
AACGCGG			
TCGCGAT			
GGGAAGA			
ATAAACC			
AGTTAAA			
ATTCTAG			
TCAGATA			
TGGAAAT			

				ATTCTTG GGAAGCG ATTAATA AACGTTA TCAACAC TTAGTGG CTGAGGA TTCCACG ACAGGAA GGAGAGC TTCCAAG CGGGAGC AGCCAAA CAAGCCA TGCTTTA GAAAGAA GGCATTC GTAGTTT CGTGACC TGMCGAA GACTCCC AATTGAA ATKTCAA TTTTGTA AGTCAAA GTTCTAT CAATTGC CAAACTG TTTACATT TCGCCCA AAGTCTT AAGAAAG GACGAAA TTTCAAC AGATGTA ATATATG CAAAACA TAGACCG CATTAAA CATATAT					

				TGTTTCG GTTTTCCC TTGCCCG GTTTCGC CTTTTGTT TCTAAGA ATAGATC GGGAGTT ACAAAAT TACGATA GTATTGA AATATTG CTTACAT AGTCATT ATTGCTTT TCGGCGA GCCGAGC AGCTCGA TTGCTCA AGAAATA TATAAAC TTAATTA TTTATTTA TAATGAC AATTATT ATTTATA TAAAAAT TATACAG TGTATTG GCCAGCC ATTCCAT TGAACAA ACGTATT GCCGTTG TGCTCTA TCCTTAA ATTACGA ACTATAG GCAAGTC AAACAGT					


CATTTCG
CAACGAT
ACTGACC
GGTTGTG
TTTTATTT
GAGCTCT
TGAAATT
ACCTTTC
AATCGCG
GTTACAC
CGAACGT
ATATAAT
CTTAAAA
TTGGAGA
ATAACTT
TCTCAAC
CCCCGCT
TAACACT
TCACGCA
CCGTATA
TAAACAT
TACATCG
CGCCAAT
TAACATA
TTTAATT
GAATCGC
AATTCAT
CTGCTGT
TGCCACA
AGKGCTT
CTGTTCC
CTATCGC
CGTCGCG
TCTGTTCT
TTACACC
CGCTCGC
CAATTGC
TCTCTCA
AGGCCAC
GGTCCTC

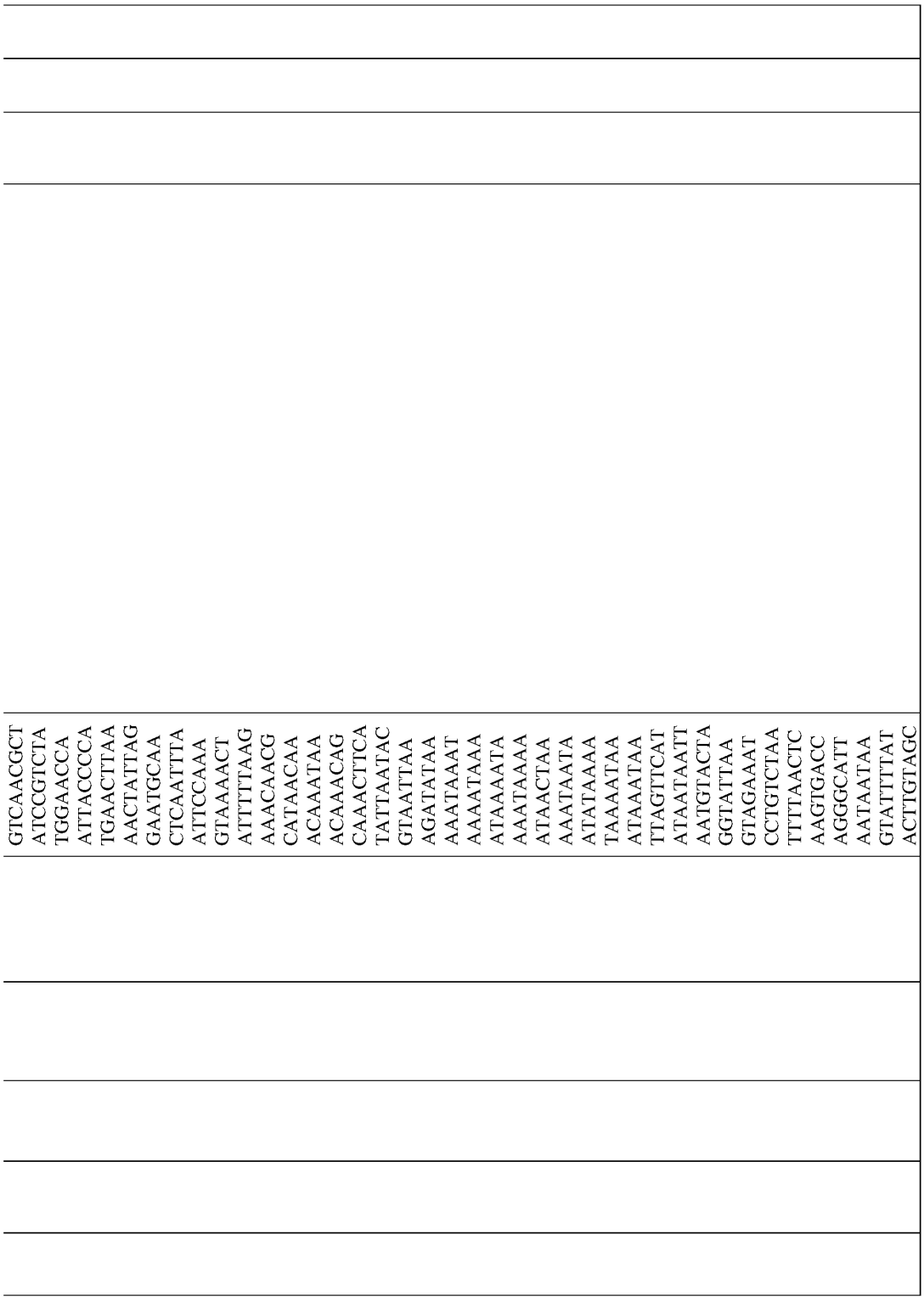
				TCTCCGT GCACTTC ACTTTGA GCGCTTC GCCTCTC TCATCGC CTGGATA TTCGAAT TGGCTCT CCGAATC GACGACT CTCCCAT CAAGTAT TCTAACG GTCCTCT CTCCGTTT ACTTCGA GTGGGTC GCCTCTC TCATCGC CTTGCTT GCTGCGG ATCTTCG AATTGGC TCTCCGA ATCGTCG CCTCTCC CATCGCC GTATCTG CCACTAC CTGGTGT TCTGCGT GTCTCTTT AATCTCC CACGGCC TGCTCTTC TGCAAGA TTCTCTTA ACTGCCT CTTAACT					

RTE	RTE-1_DF	.	Drosophila ficusphila	ATGTAGG GCCAGAC CAATAAC AATCGTG CAGTGTT ACGCCCC AACCGAA	TAAAGGAA AAAAAAAA	MGDFNAKIGPNFTRMRSATGGHGIGTRNNNGDRLVEV CQMFQLIIGGSMFPHRDIHKYTWTSPSGNTRNQIDHICIS RKWRKSLLDVRNKRGASIDSDHELIIGELSIKLDRNRPR NADDSRNHRPSPLNIHLLLGETDISARLASTLRTHLLPQQ GQTWEQTCDTLRTTAEGIIGTQQRRRNNWISSATWDTIN KRNSLKLLADRDSNRRDEHRALCKAVKRAARNDKRAL LDRLADNAERAANENNMRSLYQTIEQISGGYRRKTQPV	cd01650	381	637

						KDANGILLTSDDDQIHRWRQHFMEISNSDMPSRSDDYR			
						SVAPANGNRRISSAPPSVREIGNAIKKLKNNKAAGMDGI			
						PAEILQVDTQLMAEILHPHFSSAWESEHIPDSWKCGVIV			
						KIPKKGDLSDCNNWRGITLLNTSYKVLATILNERLTEKF			
						EPTLRDEQAGFRPHIRSCVDQANTLRIITEQAVEWRAPLY			
						LLFIDFKKAFDSINRAAIWGALARKGVPHKVIALIKALY			
						DDANLAVLHNGKTSAPFQTNTGVKQGCPLSPLLFNIVV			
						DEVMREVCKHKRGITWSFTRHLEDLDYADDICLISHKLS			
						DIQSKVDDLVNLASRVGLQINTAKTKAMRLNHASQGNI			
						IVDGNPIEYVESFPYLGTVISSNGGVTEDVTSRLAKARSA			
						FGRLYRIWRSRQIRRKTKIRIFNSCVKSVLLYGSETWLTT			
						KTIIQQLQAFTNKCLRIICGIFWPNRISNIDLWNMTGETAI			
						RLQIRKKKWRWIGHALRKDQXSIVRTALEWNPQGSRR			
						VGRPKATWRRTALQELSQYNITWDTAKQTAXNRVRWR			
						NLVEALCSQEEL			
					TAGGTTTAG				
					TTTATAACT				
					CCACGTCAC				
					TTGGCGCA				
					GACGTGTC				
					AAAATGTA				
					TGTCAATAT				
					TTAAATTGT				
					AACAAAAA				
					AGTGTATTT				
					TGATGTTGA				
					TTGTTTTTT				
					GTTTTCGAT				
					TAGTTGCCC	IVNNKWTILKEHTXTPSKVFLDIVKFCIQDNRYLKYETK			
					TGAAGACG	LXIXLKGMPMGSPASPVVADIVXETLLQNTTEKLRRPPR			
					GTTGCCGA	LLTKYVDDLFGIIKEDEIENTLTTLNAFNXNIKFXMELEK			
					GGAGCAAC	DNCIAYLDTVVIRRGNELKLKWYKKPIASGRIINFNSRH			
					CGAAATAT	PKMMIRNTASSCINRMLRISDAIFHEEIINEIKTILTNNDFP			
					ATTGGTCAA	KHMIKSLLDKYLKSDNNQRPPKKERTYISVTYVPGLSER			
					AAAACATA	LAQSDCYDKENILIAHKPDNTLKQIFNKTKSKIAPTDKK			
					AATTGTGTT	DVVYRIPCNGDGKNNCNSIYIGTTKSKLKTRLSQHKSDF			
					TTTAATATC	KLRHQTNTQKTALMSHCAKQGHSPNFEETTILAQEQHY			
					ACCCAGAC	RKRYMMEMLHIINTPTNIRMNYKTDTENCAHIYRHLLG			
					CTTGAGCCG	K			
					AC			cd00304	47	129
			Drosophila ficusphila	.					
		Penelope-1_DF							
Penelope									

					GCAAAACA CCAAATAT				
RTE	R1-1C_DK	RTE-2_DF	.	Drosophila ficusphila	GGGGGGA TGGGGGA GAGCATA TAAGGCA AAGTGTC TGACTGC ATTCAAC TTGGTGA ATGCCAG CATGATC CTTATCA CTGATCT CCTCAAG CCGACGC ACTACCT CCACGTG GTTCTCC CTGGATT GCCATTC AACATCC GTTGAGT GGCAAAC TAAAGCG CGCGACG AGAGTCC TAGCTCT TCAAACT AGGTTGT GGCGTAG GACTTGA AATCCAC CCATGGA AAAACCA AAATTCA	TGATGGCC AATTATTAT TATTAA	MKEAEEASERDRFNVDDHGKRKKDNDLRVCTWNVRT LYRPGAAQQLADTLKKCRADIAAIQEMRWIGQGCIKKK DCDIYYSCHPERHEFGCGFVVGARLRRRVSHFLPINERL ATIRITAKFFNISLICAHAPTEEKDDVTKDTFYAELDRAY GRCPSHDIKILLGDFNAKVGREDIFGATVGRFSLHETTSS NGLRLIGFAAAHNMVVRSTGFRHLDIHKATWLSPDRLT RNQIDHVVIDARHASSVLDVRSCRGPNIDSDHYLVAAKI RTRLCVAKVARRGLLRRLDIGKLQSQQTRDAFSTQVSG LLSRTTPTPEDISDQWALISHSLRTSAEEVIGFQRPPTRNP WYDQECRDATAAKDAAYRRTLQAGATRAVVEVYRLR RRDEKRLFRRKKREXEKRECESIESSRCRNEARNFYQRV KRLTQGCKTGGVACKDNDGNLVTDAEIVLKLWRDHFS SLLAGSGTGYEDYDPQTPIYGTDIDVPTPSHAEVKDAIQ RLKNNKSAGADGLPAELFKAAGDMLVGSMHQLISKIW RTESMPEDWNLSMICPVLKKGDATLRTNYRGISLLPVA YKVLSSVLCENLKPHAEALIGPYQCGFRPGKSTIDQIFTL RQILEKSYENQIDTHHLFVDYKAAFDSPRRDRLYAAMS ELGIPAKLIRLCRMTLSNTISSVKVGNDQSETFSTKCGLR QGDSLSCMLFNILMESIVRKAGVHRTGTILSNRCVQLLG YADDIIIGRTKRDVTAAFGAIEKESAKVGLAVNMDKT KFMVCSSRESRRLDSQLTADSYSFESVKEFIYLGTAITST NDVSLEIKRRVTLANRCYFGLSRHFNSRALSRQAKITLY KTLILPVLLYGAECWTVVQSDAAALGVFERKILRKIFGPI CVGDAYRIRWNHELYELYGDVDVVSRVKIHRLRWLGH IARMEEDAPARKVFDAVIVGKRRRGRPRMRWQDQVME ALSTSGAANWRRRAQDRDAWRQIVQQAVTRQGL	cd01650	551	807
	R1-1C_DK		.	Drosophila kikkawai	GGAACAA ACCAACC AA	ACAGCGCC CGCGAAGC AATGCCAG	MQLIQLNLNHCRAAQDLLTQTVRERGSEVAILSEPYRIG NSRDWATDRTGKAALWLCGVGAQQMCDIKAADGFVR ANVGGTWLYSCYLAPSLSLESFGRILDELSIDLRGHSNV VVGGDFNAWAIEWGSSRTNARGRAVLETFASLDVVLL	cd01650	455	717

Kiri	Kiri-5_DK	.	Drosophila kikkawai		GCGGCGGT ACCGCTGG GCGTCCGAT TCGTATTCT TTCTGTTAA TTGTTAAGC CATCCATGT TTATGTGTA GTGTATGTA TATATTGTA GAGTAATA AATATAAC GAATATAA AAAAAAAA	NEGSRQTFSRAGAGSVIDLTFVSSALARRARWTIGEVYT ASDHEAIECSLGMAGRPSGPPPLPRRAFRQDTFRPRAFA SALEGFAVDEADGANVMANKLADAVEDACDRSMLQR RPFRGNHRPVFWWSEEIAELRRICHRSRRLLQRARGSPL LVECSERYKAARKALKIAIRDSKRESFLKLCDAAEEDPW GGAYKMTVKKLNAGGNAPSDPEALEGIVRALFPDGQP VGNFPRAEPGEFRSVTEAEVLQAGRNIKPKKAPGPDGIP SRATKLALALLPEQVAGLFNKCLQEGTFPVRWKRQRLL LLAKPGKPPGEASSYRPICLLDIIGKVFERVIATRMTEAV EAAGGLSTEQYGFRKGRSTMDAIMKVVKTAREAIAGSR WKGGTKSYCLVVTLDIRNAFNTADWGRTLESLRMFNIP SYLLNVALSYFSNRVLTTDTCLGSRAYKVSAGVPQGSV LGPLLWNAMYDGVLRLPMPEGTSLVGFADDVAIVVVA KELATVEAHADAAVQAVESWLANAGLQLAAHKTEAV LVSSRKAVETASVRVGGTTIRSQRAIRYLGVLLDTRLCF KEHLEFVHKKASGTAGALSRMLLNTRGPKQATRKLLTT VVTSQMLYAAPVWAEAAKVKSYVRGVEATYRLCAVRI ACAFRTISEDAALVIAGQVPLTELVRERAEIYAAAQXRE ATQQTRAESKITARRRSLENWQNRWDSSQKGRWTHTLI PNLTSWVERSHGQVDFYLTQVISGHGCFRSYLKRFGHE TEDWCPVCGSGVVEDARHVLFECHRFDHERQELEAVA GSSIXAETLVPMMLADQRVWEATATFAADVMKSLREQ ERRRKEQTE	cd01650	72	326
					TAGAGCCCT TTCATTTTT CCACAAAA TAGTGACA ACCCATGA GCCTTCTTA CTTGTTTAA TAAATTAAT ACCGGCAG CCTCGCAA AGGACTAA AAATTTTAT ATGCCCTCG AGTCCAAA TGCTCAACT CCGAGAGG CAGTTCTTT	SVAPCDNIFNFVCVSQCDVLDCILKVKSNAEGLDNINPK FLKLLLPCLLPYVTYIFNSILTQSSFPTAWKLAKILPIPKQ NNEYRPISILPFLSKVFENVLAGQIQGFLDRNLLLHNKQS GFRRKRSCLTAITDITEDIRQQLDKNHVTILTLLDHSKAF DSVNHNILCTKLKNLFNFSDSSVRLIASYLTDRLQAVVA GAVTSAVKKLTRGIPQGSVLGPLLFAIYVNDLPNVLSDC EVHLYADDVQIYVSSPVREINVCISTCNRMLSLVQTWAE HNGLSINPKKSKSILISRRPIDTNDLTKPIINNIPISYVDHA VNLGLTFNNTLSWGNHISKATGRVYGILRQLANTRNFL PEHIRLLISKTILIPTLFYGIEIFGFCDTRDTRKLVTALNSI VRYVFNLNRFDHVSHLTTKLLDLNLNSWINL			


GTCAACGCT
ATCCGTCTA
TGGAACCA
ATTACCCCA
TGAACTTAA
AACTATTAG
GAATGCAA
CTCAATTTA
ATTCCAAA
GTAAAACT
ATTTTTAAG
AAACAACG
CATAACAA
ACAAATAA
ACAAACAG
CAAACTTCA
TATTAATAC
GTAATTAA
AGATATAA
AAATAAAT
AAAATAAA
ATAAAATA
AAATAAAA
ATAACTAA
AAATAATA
ATATAAAA
TAAAATAA
ATAAATAA
TTAGTTCAT
ATAATAATT
AATGTACTA
GGTATTAA
GTAGAAAT
CCTGTCTAA
TTTTAACTC
AAGTGACC
AGGGCATT
AATAATAA
GTATTTTAT
ACTTGTAGC

					ACTGGGCT AATTTTGTC AAAATAAA TAAATAAA	TIDPAEIANTLGDLWSKQAKDSNFSPEFRTNKMLITPHN TYAPTPTAIEIEKNINLIELNSAINSLKGHTPGLDRISYSMI KNTPISFKFRILQHFNDILDSHIPQSYKLSLVLPILKPNKP KTNPNSYRPISLNPCLAKVLDKVVANRLWWFVLSNKLI HNNQFGFKRGRSVLDSLLYVDHLATRSLSLKKHISIISLD FAKAFDKIGIHSLIHQLEKWKLGPRIIKYVWNFMTNRKII VKNGTSLSISFPLDNGIPQGSPISVILFLIAYNSLTNIISIPN XLXFTAYADDFHLILQHNRNNNPQVNLSTLFHQIGEWC SYSGSSLSIDKCQLLHICRKRSCRANIITGSINIPTTDHTKI LGLTINKRYSWNHHITNLKTSLSRPLNIIKHLSSRKYNCN TTSLIQIVNSILISKIDYCLPIYGKAPTSILNCLKSTQNSAIR AALGAFRTTPINNLLIEAKINTLENRCKLLTSKLGTKLILS HNSPIGKLSKSIAKTKRNYRYPSTLDRIIDSCKSFSLPFSP AKFYTSKKYNPGIILKATNTSLSLYNKNTTSPITYHQLFH SIKNSLHDHKFIYTDGSKTNDTVGYSVTNNTDIIKSGLLP QYSSVISSELIAIHEAIKICSGSPGKYAICSDSLSSIKSISNV NNYSYYPNTIRSLITKLFPKIRLIWIPSHIGIQGNERADTA AKETHKYPLIFTDNFNTKDITQLLKKHYTDKLNSLQAQT TQWYKQILLNTPANFSFKTTNLNRQQQIIINRLRLGHTQ LTNAHYMNRSLTSICPFCNTSAIDQHILIDCTALSHVRTI CFSNDNPIHLLANPTAANSNKIIYFLHKMNLIHKI	cd016 50	111	368
					TAATTATAA CATATTCCA CACACATAT CTTTGTTAA TACACTGTA GAGCGGAA GGCCTCGGT AGCCATTGC TCATTATCC TTTGTTAAC TTTTAGTTT ATAACTATT AGTTAGCAT TAAATAAA TAAATAAA TAAA	MHNICATGNFSYILEMEGSVDVSHSHTLLRILAKQRSGL KICHINAQSLPKKIEEFRYMFVNSGLDVICVSETWFVQD FRDSLISCDGYYVYRGDRVSHGGGVAIFVSKKLKTKFV CKHPDNSQIEYIFLEISNKFASSKILIGCIYRPRCDITYNEII SLLGDISLNYADIIVAGDLNSNMLRETHLKNEFESIGVHL VNSVFPTHFSRTCSSLLDVFFVSDVKKVYLFDQLIVPSFS KHDLIYLTYNFDLQYNDESIVFRDFKNLDYGLLMTEFD KCNWNNIFVSPTVDQQLAQFQNNISYLYENFVPVKTKVI KHSVKSWFNNDIKLAIKSRNDAYQKWKRYRTPEHYGR FKLLRKTVTKLIDSAKTNFYSSRFQTTASTSRTWKEIKSL GIGRKPAAPEQXVDIEELNTSLSKCNVPEICENNMYLRS VPRNLCDSRFRFVCVNQCEVLVNILKIKSDAVGLDDXN PKFIKVLLPKILKYITYIFNTAFTTSTFPDRWKIAKIIPVPK ANKEYRPISILPYLSKAFENIMASQINSFIHDNALLNNNQ SGFRKHRSCTSAILKIANDIREYADNSNVTLLVLLDYSK	cd016 50	505	758

				A		
I	I-3_DK	.	Drosophila kikkawai	GGCCATG CGATGCA AAAAGCA CAACCAG CTCACAG GTGTTTTT ACACGAC GTGGTGT GGTGTAC GTGAAGC TGAGCGA CGGAGGG	AG
Kiri	Kiri-1_DK	.	Drosophila kikkawai	GAACCTC GATATGT GCAAAGC		

					AFDTVNHNVLCAKLVNQFKFSKTAVKLMTSYLQNRRQ				
					SVVVSNVSSSFAFTKRGVPQGSVLGPLLFTLYVNDLPSM				
					LSSCDVHMYADDVQMYVSRPLNRINECVDICNNELVKI				
					SNWAKENGLGLNPSKSKCIIIYKKKLSLNIGPLLLNDSPI				
					EYVERATNLGFIFNNTLNWSDHINTAVGKTYGLLRSLY				
					LVRNFLPLNVRMLLAKSCLIPTLFYGIELFANCDSNSQH				
					KLKMAFNSITRFVFNLRRFDHVTRLSYNIFELHLADWIK				
					MRVLVYLQKIIFSKLPIYLYETLRFSTSRRSNIIIQIRH				
				TCACACG					
				AGCTTTC					
				TGAGCTG					
				ATCAACG					
				ATGCTGG					
				CGGCGCC					
				GATAAGA					
				GCACATT					
				TCGAGAT					
				TAAGAAG					
				CTTCACT					
				CAATTTT					
				GTATCTC					
				CTATGTTT					
				ACACATA					
				ACTATAA					
				TATTTATT					
				AGATTTA					
				AAATTAA					
				TGTAGGT					
				CGCTATG					
				TTTTCTTT					
				GTGTCCT					
				TTTCAGT					
				ATGTTTC					
				CAATGCC					
				TTAATTT					
				GCTCATT			SSGPRASDSFNQRSKKLNNSFSLHYQNVRSLLGKLSKLH				
				TTTGTCTC			VNSVSFDANVIAFTETWLNSSVSDCEIFAGSYVIYRCDR				
				TCGCCTT			SMRLGGGVLIAVKSDLVSQRIPIIGTDSILEFVAVRIKLH				
				AATCCTG			DYFVFISCSYIPPRSDISLYLQHLAQIQSIHSTLGYNDHLI				
				GTGCTTC			VMGDFNLPDISWMTSTDSIALVPTTAHEFVHGLIDLSLG				
				TGAAGAT		TGATGTGA	QLNSILNCSKKVLDLIFVSEFANAFVTRISPLSNPEDAYH				
				GACGTCA		GTCGGGTA	PTLEVTLEASAPLQSCSEIEPASFRCFRRAKFAELNRHLS				
				AAAACTA		CTATGTTTA	ELDWSFMDSNDDPLHIVEQFYGILYSAFDTFVPSFPSAN				
				CCTGAGC		TGTTGGCAG	APQKPPWFSRRLSFLKNSMSRLFKKFKRTGLQVDHSMY				
				AGCAAAC		ATATTTGTT	LIARSNFLVHNSVCYNGYLNRCRIEFQDNPKQFYAFVNS				
	CR1-4_DRh		Drosophila rhopaloa	TGAATGT		AACTTTAAT	KRKSQQFPSCLHLDTTSASSDADIANLFAKFFQSTYSSFS				
				TTCTCGA		GTAATAAA	YDNTSSYPYPFLTPTVIFTPLISEDDVLQNLKTIKMSFSSG	cd01650	514		
CR1				GCCGAAC		TAAATAAA	PDLVPSCVIKKCADALCKPLSKLFQISLLHSFFPEVWKES				765

						FIIPLHKKGSKSDIKNYRGIAKLSAIPKFFEQLITAQLQHQ				
						CSSVMSPTQHGFMKRRSTTTNLLEFTSVIRSAFKNNIQT				
						DVIFTDFSKAFDSVNHQLLLKKLTLLGFPIMLTKWISTYL				
						SNRTQRVAFKSTISNSVHVNSGVPQGSHLGPLLFGIFIND				
						LPQVVLSSYVMMYADDVKLCLPLSNPNSHDLLQADLN				
						NLHSWCIQNLLFLNNSKCKLMSFYRKRLLLLTSYSIGSND				
						LERINSVTDLGVLFNHNLNFSDHIGTIVSKAKGVFAFVK				
						RWSKEFEDPYTTKHLYISLVRPILEYCSCVWSPQYTNHQ				
						DLIESVQKQFLLFALRGLNWETHRHLPSYVDRLRLLNLP				
						SLKDRRTCHGVMFLHKLITGLVNSSYLLGQIRFSIPVRST				
						RYFRPIALDICKTNFEVHDPFRVLCSQYNELYFLIYSECS				
						VNLIVTNILTHLSLPPPT				
						MMRLIQLNLNHCTAAQDLLVQTVRERRVELALLSEPYR				
						TPASPDWAFDRTKKAAIWRCSRETQQITDVLSDIGFVRA				
						RVGRWWVYSVYLAPSLTLIAFSRALDRLAADARGRTQ				
						VLIAGDFNAWSESWGSSTTNARGRMVLEAFATLDLALL				
						NHGNRHTFRRAGLGSVVDLTFTSGSSYRLARWRLSEDY				
						TGSDHLAIICDLGSPPSSQAQTPAQARIKYKTDSLDTQVF				
						REQFLPTASRQGAELTAVELMRRLKAACDASMQSSRPH				
						SRQRAPVYWWNQEIETARRECLSARRRYQRARGAESFA				
						ERQSEYRARRKTLKLAIRESKRKCFLDLCDSADSDPWG				
						SAYKVVVKKAYTRTPKLLDPAMLRSVAEHLFPVMDGL				
						RPADPATGDHVDSDATVSSDEILELAKLLKDGKAPGPD				
						GIPNRALRLALHLQPDSFAKAFTKCLEEGVFPSCWKIQK				
						LLLLPKPGKPPEEPSSFRPICLIDGTGKLLEKLVCIRLERAI				
						ADAGDLSRSQFGFRKARSTVDAVNRVVEVAAQAIDGT				
						RWKGGSKEYCLMVTLDIRNAFNTARWDRILEALIGFSV				
						PAYLVRIIRSYFSDRVLLCESSEGVFRHQVTGGVPQGSV				
						LGPLLWNTMYDGILRLPLVGRSEIVGFADDVALLVVDK				
						YPGKAEESCNQNIRAIEHWLSSMGLELAPEKTEAVLISS			cd016 50	473	720
						RKAVETATVEVGCTKVSSSRAIKYLGVMIDTRLSFREHL				
R1	R1-1_DS i	.	Drosophila simulans	CAATCTG	A	RKAVETATVEVGCTKVSSSRAIKYLGVMIDTRLSFREHL				
				TCCCAGC	TAAAAAAA	YPGKAEESCNQNIRAIEHWLSSMGLELAPEKTEAVLISS				
				ATAGCAG	AACGAAT	LGPLLWNTMYDGILRLPLVGRSEIVGFADDVALLVVDK				
				TACGGAG	TATATAGTT	ADAGDLSRSQFGFRKARSTVDAVNRVVEVAAQAIDGT				
				GGAAGAC	TAGAATAA	RWKGGSKEYCLMVTLDIRNAFNTARWDRILEALIGFSV				
				TGCGGTG	GGGTTCAAT	PAYLVRIIRSYFSDRVLLCESSEGVFRHQVTGGVPQGSV				
				ATCGCAG	GTTGTATAG	LLLLPKPGKPPEEPSSFRPICLIDGTGKLLEKLVCIRLERAI				
				AAGAGCC	CCTCNTAAA	RPADPATGDHVDSDATVSSDEILELAKLLKDGKAPGPD				
				CAGTCGT	TGGATTTGC	SAYKVVVKKAYTRTPKLLDPAMLRSVAEHLFPVMDGL				
				GAGTGCC	TCCTCGCGA	NHGNRHTFRRAGLGSVVDLTFTSGSSYRLARWRLSEDY				
				CCTGCAG	ATACCGCG	SRQRAPVYWWNQEIETARRECLSARRRYQRARGAESFA				
				AGAGATC	TTCACGGCC	REQFLPTASRQGAELTAVELMRRLKAACDASMQSSRPH				
				GAGACAG	TGAGCAAA	RVGRWWVYSVYLAPSLTLIAFSRALDRLAADARGRTQ				
				GAAGACG	TCCTCGCGA	TPASPDWAFDRTKKAAIWRCSRETQQITDVLSDIGFVRA				
				CACCACG	AGCAATGC	MMRLIQLNLNHCTAAQDLLVQTVRERRVELALLSEPYR				
				CATGCCT	TTCACGGCC					
				ATGTAAA	AGCAATGC					
				CACGAGT	TCCTCGCGA					
				TTATTGTT	TCCTGATCG					
				AAACATG	TGAGCAAA					
				ATGGCCT						
				GTTTCTAT						
				TCTCGAT						
				ATAAGGC						
				CATCTTA						
				TGAGGAA						
				GACTTCC						
				AAAATTG						
				CTTCGTTT						
				GAAATAT						
				GAGACGT						
				TTAAGCA						
				TTCAAGT						
				CTTGAAG						
				TAACCGT						

				ACGGACC CGCTGAT CAAGCGT GGCAAGC TGAGGGT AGGATGG TCCCAAT GCTTGAT CAAGGAG CTGGAGC CACGCCA AAGATGC TTCAGAT GCCTGGA AGAAGGC CACATAG CGGCCCA CTGCAGA AGCACCG TGGACAG AAGCCAG TGCTGCT TCAGATG CGGGACC GCGGGAC ACAAAGC CGCCGAG TGCCCTA ACGAAGC TAAGTGC TTCTTGTG CGCAAGC AGAGGTA GCCAAGC GACCAAC CACCAAA TGGGCAC CCGGAAG TGCCCAT TGGCGGG		AYANSKAAGVNRALSNIMLNSRGPKQASRRLLTSVTRA TMLYAAPVWAKALGTESYAQGLNATHRLSALRVCCAF RTVSNAAALVIAGIPPLDLLAKENEVVFNQTHGRRLSPG VKKAIRAEARQHTLETWQHRWTNEPKGRWTQQLIPNL KPWLERKHGDITFHLTQLISGHGCFRSYLMRFGHEDAD DCPWCGGGRSETAEHVLFSCVKYARERSSLETVLGGRL NADNLVPFMLQGEAEWQAVNSFAAAITTELRRAERAR RVHE			

			TAAAGGA GCACCAA AGGCACT GCA						
			GAAACGT	TAACAAAM					
			CAAAAGG	GTCGACAC					
			ATAGATT	AAACCGCT					
			TTCACAA	CCCCAAAA					
			AATTAGG	ATGCAGCTC					
			ACAATW	ATTCTACAA					
			WTGGACG	AATAGGAA					
			CAAAAAG	TAAACAGG					
			CAACAAT	CTACATGG					
			AACGAAC	AAGATTCTT	LRRLPHLPRPKPQNNYQPNCPTHLSKSHRNRLYTKCSHS				
			GGCAGCC	AAAATCTA	PNPLKWHPCKYDNCCPIYCPPVYKWTHENFHKLFSHFE				
			CCCCCCT	CACTAAAG	TLLDGKSNFILSGDCNSKHRCWGNYSSCARGRSLMQAI				
			TTCAACA	CCATGGAT	MANKLLDVVATGQATHFPFDRRKQPSALDFAICKGLRR				
			AGACCCC	GCGGGTTAT	RITRTFSSDDLSSDHLPIQILDLEEANLQLKKVSNSILKR				
			CCACCTA	TTCACTGAA	NTNMEKFKRILEDKMLLNTEIRTGQDIDDCIDIFLNNIKT				
			CCATATA	ACAATAAT	AANEATPKNRPFTNPRENSRRRENNLSLDEATTRLLDEK				
			GTTCCGA	ACTTAAGC	KELNRILRAAGTEEAKKWFKSAQNKLAKEIRKLRTSQIN				
			AATGGCC	GCCAGCAA	NELQLIDTTDRYRMQKLWKWTNNVKKQPQPIWPLKKD				
			AAATCTC	AGTGTCTTC	DDCISMRTASPWTRTLDERAEAFASHLESRFKANLRNSI				
			GGAAGGA	ATCTCTATC	EDREIIKKELDKFKSLKPRIETGTSGLKPVTLAELDLLIKS				
			AGCCGCC	CTGATGACC	LESKKAPGADNISNKVISNLPKKARIYLILIYNKILNTGHF				
			AGGAAAA	TGCAAAAC	PKKWKHACITMILKPGKSPSLLNSYRPISLLSGLSKIFERI				
			CTACATC	CCGGACGA	ILNRLYNIESFSKAIPNHQFGFRKGHGAEHQLARVTQFIL				
			AAACCCG	AGTATCCCT	KAYEDKEVCSATFLDITEAFDRVWHDGLLLKLTKLIPK				
			TTCCGGC	CCGACGAC	QLFNVLESYLTGRTFQVRINNETCSRIGTINAGVPQGSIL				
			AAAATGC	GCATCCCTG	GPVLYSIYTSDMPFPIIKGYIRNISLPDYHPKNIILATYAD				
			GAAGGAC	CTGACCCCG	DTIIISRSNNTRSAININQTYINKFCRWAKKWDIAINGKK				
			GATAGGC	ACGAAGCG	TGHVLFSLKNISENVHTTPIVDGQRAIKLCKHGYLGVML				
			CAAACCA	CAAGTGAA	DRRLTFGAHLILLKGKIIAAYKRLEWLLGKNSHLPKRTK				
			AACGGGA	GGATGTCAT	LLLWKQIISPIWHYAIAIWGSLTSATQTKKIQTLENKLIR				
			TTCTCTTC	CCACTAGG	RIIKANRYTKQSAIREKYKIKSFEEVFDQASKRYANSIAD				
			ACATGAA	ACTGGCAA	HENPLIKELLTKSFKPIRLAASRTRYVVNQYTKYLLPLLQ				
			GCTGCCA	AGAAGTAT	NIXKPPDEPTYKSLYSYTDKEELEIVSKIRYDYRQTLPTL				
			TCCAAAA	CTAGAAGC	LQLEDEEMEARDCKAEEERKEREKEERKKALEKGPPDR				
	Jock ey	Jocke y- 3_DY a	ACGGCTG	AGAGGGAC	WCELLINKFSKMYKKGMITRERVIELMLGQPPTVIKLVI	cd016 50	479		749
		.	AATAAAT	ACGTGAAG	PDYEP				

				GGCGCCG ACACGCG GACAAAG CGACCAC AAAAAAC AGGACCA ACTCGAA GGACAAC CCCAAGC CGCGACC CACAAAA ACGACAA AAAACCC GGCACAA AGACACC TTGAGGC CTTCAAC GACAAAC TCCGCAA GGAAAGG AGGGATG CTATAGA ACACACC CGCGACG AAGCTAT GACAAGC CCCAGGA GGGGTGA TGCTGGT AGCCCGC CGACAAC AAGTAGA GCCGCCA AAACTAA TTTCAAG ACAAGGA CCTAAAA ATTAACC GCCGATG GGACCGA	GATCAACG TGCGGCAA AAACATGA TGAATCTAA GTAGACTC ATTGCAGC AGCGCATG CATAATGTC ACTTACCTG AACCTTTTG CTGCAAATC CCTTGAAAT GTCCTGAA ATAACAGC TGCAATCTC CCCGAAAA CCGTACTCA ATGACAAA AGACGACC CGACTTGTC GCAATAGA TAAACTAA CTTTTTGGC ACTTAGAA CACAAAAA ACATACAA GGTTTTGAT CCAGTCTCT TTATTGGGA TTCCCCCAC GTCCTCACA GACCCTAG GCGTGTTGC TGCCGACCC GGCAAAAG ATCGAAAT CGTAACCCC TGAAATTG ACCTTTTTT TTCCCTTAG				

				GGACGTG CATAAGC ACTTGCA AAGAGGA CCCCCTG ACGACCA CAGCCTA AACCTGG CGCAACG AGTACTG GAAATGG AAAAGAA AATTGAC ACACTTA TGGAACT CTTCTTCA AATGCTT AGAAAAT AACAAAA ACCTGAC CACAGCA CACCCCA TGTCCTC CTAAACT TCTAAGT TCCTCGC AAAAATA AGTAGCA CCATGAC ATTAAAC ATAAACA ACACACA TACTATC AATACCC TTAAAAT AGGGTAT TGGAACT CATGTGG AGTTACC AATAAAA	AATTAATAT CAACGAAA TCTTTGTTA AAGATCGA AATAATTTA ATCAAAAC ATAAAACA GAGTATTAT ATAATAAT ATATAAWA AAAAATTC ACAATGTTT TGAAGCCG TCTCAACGT CAAGAAGC CCACTCGAT TTTATTGAC CCTTGGCGC GGCMCTGC CGACCCGA CAAAAGGT CAAAATCG TGACCCTTG AATTTCGAC TTTTTCCTG GTAGCAAC GATTATACT GACAATAT CCCTGAAA AAGAATAA CAGATCAA TAAAAGCTT AAAACTAC GCAACCTA CTACTCCCC TACTAGGC GCAACATA AAACGAAG ATTTGGCAC CTAAAACA				

				GAGACGA ACTAGAA GCTTACA TAAAAAA AGAGGGT GTCCAGA TCATGCT GGTAASA GAAACCA AGCTAGA ACGAAAT TCTACTA TGCTTAA	GTAAAAAC CTAAAATGT TTTGATGTC GTTTCCATG WTGGGATG CCCCCTCGT CCTTCTGAA CCCTAGGTG CGCCGCTGC TGACCCGG CAAAGGGT CAATCTCGC GACCCTTGA ATTCGACTT TTTCCTGGC AGCAACGG TAAACAGA CAAAATCC CTGAAAAA TTATAAAAC AAATTATA AAACAAAT TATAAAAG CATAAAAC CACGACTC GGCGAAAT GAAWAACG ACGATTTTG GCACCTAA AACATTAA AAACCTAA AATGTTTTK ATGCCGTTT CCATGCTGG GATGCCCCC CCGCCCTT MTGAATTC CAAGCGCG GCGCTGCTG ACCCGGCA				

										376
										120
									cd016 50	
R4	EhRL E2	AB097 128	Entamoeba histolytica	AGTAATA CAAAGGA AGATAAA TAAACTA AAAGACA TGCTAAT AAGAGGA AACTATA AAAAATT AAGAAGG ATATTAA GAATAAG AAGAATG ATAGAAG AAAAGAT AAGGGGA GAAGTAA ATGAAGA AATAAAA CAATTAG AATATTT AAAAGAA ATGAGGA AAAAGAA GTAGACT TAAACAA AAGTGAA AGAACAA GAATCTA	AAAGGCCA AAATCGTG ACCCTTGAA TTCGACTTT TCCTGGTAG CAACGGCA TACTGACA ATATCCCTG AAAAAGAA CAATAAATT AA	TGAACAAT AAGTTGAA AGGTATTA AACCTTAAT TTAATATTT ATTCTTTAT TCTT	MSTTGKKEIRKEIGREEQPSEEDTKEFLNKIYNRGKRIKK ENERIKEIIEESNRRNLKRIIITKENIENVITRIPSWKAPGIE NMHGYYWKRLESVRDRLVIVFNEWLDTPENIPIEMLTG RTILIHKGGDSKKAENYRPITCLNVIMKIFTSIINRKIEEQ LANNEEKYQINKSQYGNKKRSLAAKEVNINSLLERNER EVQKKRYLESFYDIRKAYDTVNDEWLENVLRYFKIPFQI TELIKSMSSRWRIRIGYRYKEEIREVKIKNGILQGDSISPL LFILQMNIISDVIDRTNEKRLKVKHVLYMDDIKVTSETK KGMEMVHKNIIETIEVIGMEVNVNKSGVMKKGNIGIQD NMKDIPIVDSEHQYKYLGVYQFTRNDDVGCIERISKSIEE GIKENNKGNDSSMNYINRINTEIIPKFTYSASVVKLPIMK LIEIDKMIRIEVKEGKIIGAGMCTSRLYVVNRNKMGDGLK SIRDEYAIKLIRSILYYTWKRGNRIGDMIGINENKNRMLN RLYKSMLKKRINKEEFRKIVEEERGKRDKNRIIIKYVREK FQEDYINKWRNKKVNSEILKGFEDKTMEQKKTIKVWK GINVIRKAYIQIKKMEEKARNVGVRKALMKHDMRYAIC PICNKIATINHILLNCIVTKKTQMNKHDRIGEYIWRSLKI KYQLYENKEKMKMREEENTKKLIIRKEELGKNERRIEK DKILIVWNHEVVNRSEDKFHKRIDIYVRDLNKKEAMIID MTIVRDKNISKAFTDKINMYSKLHEHIRKIERLVRVIVIP VVITVSGLINKETVILLNAEGIEIPREIMTREIVISNMKDL MRYCGDHSGEYSNVEIPEEEDSQESEGREEILT		

				CAGAAAC							
				AAAATC							
							MVQNNPMYKRGRGGVTLPGGSPGPPSGARPRGRARRR				
							APGGRVCHGARSGIARKSNVAPPSPSYGPTTHGRNRWG				
							RVRCHMGGSEGQGPRRTRPGRQGLALGTWNVTSLWG				
							KEPELVREVERYQLDLVGLTSTHSLGSGTVLLLDRGWTL				
							FFSGVAQGVRRRAGVGILTSPRLSAAMLEFTPVDERVAS				
							LRLRVVGGGISDCCLCICPEQQFGVFGLLGDPEWSPVW				
				CCTCTCC			GSSRGLHNSAGGLQRTRGQFCDRVIGSETACFGHSGKE				
				GGGAGCC			RGGAVNRSPSGGELGQGVGEDSGQTGKTQTGSAGELG				
				ATCACCT			TSGGGPRPKDLHSHLRRSFSGIPVEVGGIEPEWSMFKASI				
				TATCGTG			AEAAAGSCGLKVLGASRGGNPRTPWWTPVVREAVRLK				
				GTGGAGA			KEAFRDMLSRRTPETVAVYRQTRRAATSAVKEAKQRV				
				GGTTTGT			WEEFGEAMEKDFRSAPRCFWKTVRHLRRGKRGTIQAV				
				GTGTTCC			YSKDGTLLTSTEEVIGRWKEHFEELLNPTNTPSIVEAELE				
				TATGAAC			ADGEASSISMAEVTEVVKQLHSGKAPGIDEIRPEMLKAL				
				CTGAGGG			GVEGMSWMTRLFNIAWESGKVPKEWRTGVVVPLFKKG				
				CTGTGTT			DQRVCANYRGITLLSLPGKVYSKVLERRVRQIVEPQIEE				
				GTCCGGG			EQCGFRPGRGTTDQLFTLARILEGAWEYAHPVYMCFVD				
				CCTTAGT			LEKAYDRVPREILWEVLREYGVRRSLLRAIQSLYVQSES				
				GCTCCTG			CVRVLGSKSDSFQVGVGLRQGCALSPILFVTFMDRISRR				
				GTAGGGT			SRGGEGLQFGGLGISSLLFADDVVLMASSVCDLQHSLD				
				TTCCCAT			RFAAECEAVGMRISTSKSEAMVLSRKPMECLLRVGNEA				
				GGCAAAG			LPQVKEFKYLGVLFASEGTMEREIGRRIGAAGAVLHSLY				
				TGGTCTC			RTVVTKRELSRKAKLSIYRSIFLPTLTYGHEGWVMTERT				
				AGGGGAG			RSRVQAAEMGFLRRVAGFSLRDRVRSSAIREELGVEPLL				
	RTE-2_EL	.	Esox lucius	GAGCCAG	GC		LCIERSQLRWFGHLVRMPPGRLPREVFQARPAGRRPRG	cd01650	558	813	
RTE				ACTAAGA			RPRTRWRDYISTLAWERLGIPQSELANV				
							MMTPTSTTIAANLSTHDDQHCPISTCPGASNTDQWILCD				
							HCQKWYHTACVAVVIIPEEEWLCPTCVISVTSSQSQVTP				
							SSLNTTTTTKKKHIIATTQYWLKSNLVEDPLGLIPKSEIV				
					TAAAGAAT		ARYAHHIQTSGEGRVVDATVGKYILQLFPSTKFSRPPTR				
					TTCATTTTAT		GAANAYIGIRWKDQPPLLPHAPRNWGAELSQLKNTVPT				
				AATTTTTTC	ATATCTATTT		LRRVPKGARMEVAKALAKIIEEVTTTNSEESWFKLFSFP				
				AAAAAAA	GTAGTTATG		YAVVPVPTKSDKVKNLTTWVKSRVLQWSITQIVTPSRP				
				AAAATAA	CTAATATTA		RPQFAKPSRPQETIAKKVEAKLADGDVRGAIRLLTSDDT				
				AACGCCC	TGGATTACA		IAPNNQETLDALITKHPPHPEPTYFPNAPDKAEVLAPVDP				
	CRE-3_Fo Ca	.	Folsomia candida	AAAAATA	ATTATGTAT		KEVLIAISSFAPGSAGGLDSLRPQILKDLVISENGEPATNL				
CRE				TAGAGAA	TATGATAA		LNAMLKFLDLVLRGDVPKSICPIFFGASMTALRKKTGGI	IPR000	404	656	
				TGAA	AGTA		RPVAVGNTWRRLISKIAVARMSSTLVEIFSPHQLGVGIK	00477			

						GGAEAAAHAARIYWNYSHDSPKAFLKLDFRNAFNEIRR				
						DTLLQIIREKFPNFYRFVSQAYSSPSEIFFAGTPISSLCGV				
						QQGDPLGPALYALVIQPIVDLIDTELNMWYLDDSTIADS				
						PEKVLMAMQTITRMGTEVGLHLNTTKCEVGILGVVDD				
						QTRDEIFELFRSTAPGIQKISPESAMLLGVPLTDEAIPTIL				
						GAKTEEMEKISTRLLKLSSHSAFFLLRASISTPRLIYFLRC				
						APTWRRFDALTLYDAKLKNAMEGIINCSLSPQSWLQSS				
						LPVNKGGLGVRHAIDVAIPCFLASIYSSLPLAERLLPPHL				
						HGIDTAIQEGEERWKSLGSSPLPPLEIRGIQAIWEIPQQEG				
						VINHLTTNATTLEDKARYHAMSQPNVGAWLNAIPSPQL				
						GTHLPNETFRIAAALRLGCDICQPHKCSCGEQVSTKGYH				
						GLSCKNSAGRHSRHAAAANDVIARALRSAGVPIIKEPAGC				
						AREDGKRPDGLTLIPWARGKSVVWDFTCSDTFAPSYLP				
						QTSIRPGVAAEKAEDRKKKRYEQLMDSFLFVPIAIETTGI				
						WGKDGLSFIKEIGSRITNTTEEKRATSFLLQRMSITVQRG				
						NVASILGSLPKGHKLEEIFML				
									cd01650	472
			CTTCCGT	TGACTATAA		MRRDRDFSTSSVLCFTETWLCGSIPDSALQLAGFQLLRA				
			GGAATTA	CACTCCACC		DRDTELSGKTKGGGICFYFNNSWCNDVTVILQHCSPDL				
			CTGGACA	GGTCACTCC		ETFFINCKPFYSPREFASFILVGVYMPPAGNVHEAQRTL				
			TTCTAGT	CCCTCATAC		ADQIRRVERTNPDSLVIVLGDFNKGNLTHELPKYRQFIK				
			CAAAGGT	TGCACATGC		CPTREENTLDHCYTTVSRAYHAVPRAALGHSDHVMVH				
			GCGCTCA	CACTTTAAC		LIPAYRQKLKLCKPAVREFKKWTSEALEDLRACFDCTD				
			CCTTTCGT	TGCAATTCA		WDVFRTATDSLDEFTEAVTSYIGFCEDSCVPSCTRVSYN				
			CACGCGG	TCACTTTGT		NDKPWFTAELRTLRLQKDQAFRSGDKDLYTEAKYKFS				
			TGAGACG	CGTCACTCG		KAVRDAKRLYSEKLQQQFSASDSASVWRGLKHLTNYK				
			CCGTAGG	TCACTTTGT		PKTPHSMNDLRLADELNEFYCRFERQCPDPIPHSSTNLL				
			AGAGGAA	CTCTTGTCT		QSPSPPSPSPSGAHASSSISPSPPPPATTLSILERDVNRLFR				
			AACGGGC	GTTACTTGT		RLNPRKAAGPDSVSPHSLKHCADQLSPVFTDIFNTSLET				
			CGGCTCG	TCGTTAGTG		CHVPACFKASTIIPVPKKPRITGLNDYRPVALTSVVMKSF				
			CTGGTGA	CACTTTATG		ERLVLTHLKSLTDPLLDPLQFAYRANRSVDDAVNNMALH				
			GACTCCG	CTTAATATT		YILQHLDSPGTYARILFVDFSSAFNTIIPSLLQDKLSQLHV				
			CAGGCGT	TTTCTTTTT		PDSTCKWITDFLSDRKQHVKLGKHVSASRTISTGSPQGC				
			GGTTCTC	AACTTTTTA		VLSPLLFSLYTNSCTSSHQSVKLLKFADDTTLIGLISGGD				
			GCACTCC	ATATTTAAA		ESAYRWESDHLVSWCSQNNLELNALKTVEMVVDFRRN				
			GCTGCCA	ATTTTTTAA		RAPPSPITLCDSPVTIVDSFRFLGSIITQDLKWELNISSITK				
			GGCATCT	CTTTATTCC		KAQQRLFFLRQLKKFNLPKTMMVHFYTAIIESILCSSITV				
			TTCTCTCC	CTTGTTTTA		WYAAATAKDKGRLQRVIRSAERVIGCNLPSLQDLFASR				
			AACGTGC	TACTAACCC		SLKRAKKIAADPSHPGQKLFVPLPSGRRLRSIRTKTSRHT				
			GCTCACT	ATAGCCTTA		NSFFPSAVGLINRARSPTA				
			GTGCAAC	GCCTTATTC						
	REX1		Gasterost							
Rex	-	.	eus							
1	1_GA		aculeatus							

				AAACTGG ACGAAAT TCAGCTG CTG	TACTAACCC ATTGCATTA GCATTTCAT TCCACTTTA TTTTATTAC TTGTGCACT GTTGTCTTG TCTGTCTAC TGTCGCGCA CTAACCGCC AAGACAAA TTCCTTGTA TGTTTGACA TATTTTGGC TAATAAAT GTTTCCTGA TTCCTGATT CCTGA				
RTE	RTE- 1_GA	.	Gasterost eus aculeatus	CCTCTCT GGGAACC GTCACCT TATCGTG GTGGAGA GGTTTGT GTGTCCC TATGAAC CTGAGGG CTGTGTT GTCGGGA GCCTTGT GCTCCTG GTAGGGT TACCCTT GGCAAAG TGGTCTC AGGCGAG GGGCCAG ACTAAGA ATGGTTC	TAAGCGGTT GAAGATGG ATGGATGG ATGGATGG ATGGAT	MKKRKREGVTRPGGSPGPPFGARPRWRARQRASGGRV YRGARPGTARRSDAAAQHSTPHGLTTCGGNRWGRVRC YKGGSDSGGSRRNRPGRQMLALGTWNVTSLGGKEPEL VREVERYQLDLVGLTSTHSLGSGTKLLDRGWTLFYSGV AHGVRRQAGVGILISPRLSAATLEFTPVDKRVASLRLRV VGGKTLTVVCAYAPNSSSEYSAFLETLNGVLHGAPVGD SIVLLGDFNAHVGDDGDNWRGVIGRNGLPDLNPSGCLL LDFCASHGLSITNTMFEHKDVHKCTWYQSTLGQRSMID FVIVSSDLRPHVLDTRVKRGAELSTDHHLVVSWVRGW GKTRDRPGKPKRVVRVNWERLEEAPVQKIFNSHLRRSF SHIPVEVGDIEPEWSMFKTSIAEAAAVSCGLKVLGASRG GNPRTPWWTPVVREAVRLKKEAFRVMISGGTPEAVAV YRQARRAAASAVMEAKQRVWEEFGVTMEKDFRSAPK CFWKTIRHLRRGKRGTIQAVYSKDGTLLTATEEVIGRW KEHFEELLNPTTTPSLVEAELEAEEGSSSISLVEVTEVVK QLRSGKAPGIDEIRPEMLKAMGVGGLSWMTRLFNIAWK SGTVPKEWQIGVVVPLFKKGDQRVCANYRGITLLSLPG KVYSKVLERRVRPIVEPRIEEEQCGFRPGRGTTDQLFTLS RIIEGAWEYAQPVYMCFVDLEKAYDRVPREILWEVLRE YGVRGSLLGAIQSLYAQSESCVRVLGSKSKAFPVGVGL RQGCALSPILFVVFMDRISRRSRGEEGLQFGGLRISSLLF ADDVVLMASSVCDLQLSLERFAVECEAVGMRISTSKSE	cd016 50	616	870

			AAGACCC CC			AMVLSRKPMDCLLQVGNVSLPQVKEFKYLGVLFTSEG KMECEFGRRIGAAGAVLHSLYRTVVTKRELSRKAKLSI YRSIFVPTLTYGHEGWVVMTERTRSRVQAAEMGFLRRVA GVSLRDRVRSSAIREELGVEPLLLCVERSQLRWFGHLVR MPPGRLPREVFQARPAGKRPRGRPRTRWRDYISALAWE RLGIPQSELVDVARERFVWGPLLELLPPRPDPG					
CR1	CR1-1_Gav	.	Gavialis gangeticus	CC	TGACCCCA GTGCTCTTC CTGCCATGG CAGGGGGT CGGACTAG ATGATCTGC TCAGGTCCC TTCCGACCC	LVRGLRVLDPLGDSDHCLLEFTIQRKVAKASSQVVALD FRRANFNELRRIVGEALKTQRGGEPGVQGEWSLLKETIL QAQRKVIPRRVKGGRRAYKPPWLTKNIRECLLAKKEAY SQWKGGVITKEDYSAVARGCRGMVRKAKAEMEQGLA SRIKDNKKSFFKYIGSRKKAPGNVGPLQDKTGNLVVTP DDKANIFNNFFASVFLSKDQNCHLPAGPPLGPGGGSPGP RVSEDLVRELLEGLDVFKSAGPDDLHPRVLKELAGVIAS PLARLYEHSWCSGEVPEDWKRANVVPIFKKGRKEDPGN	cd01650	521	770
CR1	CR1-7_Gav	.	Gavialis gangeticus	CC	AAGCTTAG TTGCCTG TACACTA ACKCCAK AAGCATG SGAAACA AACGTAC TGACCAA CAAGCAA CACGACC AGCTCCA CCTMTGA	TGATCCCAG CAGTCTTTC CTGCCCACG TGCAGGGG GGCTGGAC CTGATGATC TGTAAGGTC CCTTCCAGC CCCTAACA ACTATGAA A			cd01650	288	539

				TAGAAACT ATGAAACT ATGAAA	YRPVSLTSILGKLFEKIILAHIQEGPAGNGMLRGNQHGFI SGRSCQTNLVAFYDQVTKSLDAGVAVDIVFLDFRKAFD TVSHPILIRKLGDCGVDTHTVQWVANWLESRTQRVVV DGSFSTWKDVGSGVPQGSVLGPALFNIFISDLDEGVQST LFKFADDTKMWGDVGTIEGSNRLQLDLDRLQGWAEEN RMGFNTDKCKVMHLGRNNRHHTYRLGNSPLVSTEVEK DLGVIIDSKMNMSRQCGEAARKANCTLSCIHRCISSRSK EVILPLYAALVRPQLEYCVQFWAPHFKRDVDSIERVQR RATRMIKGQQGRSYEERLRDLNLFSLQKRRLRGDLVAC YKLVRGDQQALGESLFPRAPQGVTRNNGHKLAEGRFK LDIRRRYFTVRAARVWNQLPSEVVQAPTLGVFKRRLDD HLTGVI			
CR1	CR1-17_Gav	.	Gavialis gangeticus	ATACAGG ACCTCCA CCTAACT CAGGAGG TGCACAG TCCCACC AGGGGAA ATGCCCT GCTGGAC CTGGTCC TTGCCAC AGGCGAT GACCTGG TGAGGGA ATTGCGG GTTCTCG GCCACCT GGGCGAT AGCGATC ATCGCCT GCTGGAA TTCACTA TCCAGCG CAGGGTG TCAAGGG CTAGCAG TAAGGCG GAAGTCC	TGACCCCA GCATCCATT CCTGCCCAA GGCAGGGG GTCGGACTT GATGATCTG TTTAGGTCC CTTCCGACC CTAGCTACT ATGATACTA TGA	MGVQDGWSFLKETILRAQRESVPMHTKGGKSAKKPPW LSKGIQECLRAKKEAYIRWKQGAITKEDHTSVARGCRA AVRKAKAEMELGLVTRIKDNKKSFFKYIGSKKAAPGNV GPLQDALGNLVIAANEKADLFNKFFASIFLSRDRDISPTG STEGLRGGTARPKVRDDVVRELLEGLDVFKSAGPDALH PRVLRELAGVIAGPLARLYEHSWSSGQVPDDWKKANV VPIYKKGRKEDPGNYRPVSLTSVLGKIFEKIIKEHICEGP AGEIMLRGNQHGFIKGRSCQTNLISFYDQVTKSLDEGVG VDVIFLDFKKAFDTVSHPILIKKLGDCGIDAYTVRWVAN WLDGRTQRVVVDGSFSTWRDVGSGVPQGSVLGPVLFNI FINDLDEGVESTLSKFADDTKMWGEVGTPEGRDRIQLD LDRLQRWADENRMEFNADKCRVLHLGRRNQQHTYRL GNTLLISTEAERDLGVIVDSRLNMSRQCEEAVSKANRTL SCIHRCITSRSREVILPLYAALVRPQLEYCVQFWAPHFKR DVASIERVQRRATRMIRGQQGRPYEERLEGLNLFSLHKR RLRGDLVAVYKFTRGDQQGIGDALFPRAPHGVTRNNG HKLIESRFRLDIRRHYFMVRVARTWNQLPREVVLAPTL GVFKRRLDRHLTGVI	cd01650	225	476

				TTGACTT					
				CAGGAGG					
				GCTGACT					
				TCAATGA					
				GTTAAGA					
				AGACTAG					
				TATGAGA					
				GGCATTG					
				GGATCCC					
				GAAGTTT					
				TGGGGAG					
L1	L1-5_GAr	.	Gossypium arboreum	GACTTCG			MKFMCWNCRGLGNPAKLRELKQLIAANNPDLIFLSETK		
				TTTGTGT			MSANDFRRVQNKCRMQNGLAVNSEGRSGGLALMWRD		
				GTCTGCC			GMNVSIQNYSKHHIDSMVKLENNDNIRVTGFYGHANPN		
				TATTATTC			LRSSSWDILRRVGESVREEWIVGGDFNAILNDAEKEGGR		
				TTTGAAT			RGVRAQMNEFKEVMDDLALVDIKPDSGWFTWVNNRR		
				TTTATGG			GGVTWGGLIKERIDRFISSVSVAEKFPFISTKVVRQTQSD		
				CAGTCTT			HDAIILDLWGRKPKAYPTDKRLCFKFDVCWAGDVEAK		
				GTGGTGT			NVIERAWNRAATDFGEKIERVRLALGPWQRKKYGNLK		
				TGAGTGT			RDMQKLEKHIDWIIDSVSSEESGRTLKETRRRLDFLYAR		
				TTCTTTGT		TAGACCGA	EESYWAQRSRSIWLREGDRNTRYFHAKATGRLKKNVIE		
				TGTTTTTG		ACGGTCTTG	KLQNAEGQWVTNSEDINKVAKDYFVKLFRSNGQNANT		
				GTATTAT		TAATAATGT	QEMGFIKDCVTRETNEWLCMDYTEKEVLQAIKQMDPN		
				TTTATTTT		AGCCCATTT	KAPGIDGLSGSFFKHHWDIVGKDTISYCLEVLNGNKNIS		
				TTTTTGTG		AATTTGTAA	SVNETMIILIPKIKNPCELTNFRPISLCRFVYKIISKVLANR		
				TCTTTCTT		GACTATGTG	LKAALPSCISQNQSAFVPGRMIHDNVLIAHELLHYLQSS		
				CGGTGGG		TAGTGAGG	KNGPNKGLVVKLDMSKAYDRVEWNFIENVMKKMGFV		
				TTTGCGG		CTAAAAATT	DKWIAKIMECVRSVTYTVKCNNTLSDIFVPERGLRQGD		
				TGCCGGG		GTTTTTTTG	PLSPYLFLFCMEAFSRMLIHAQENNTLRGIRASRNGPRIN		
				GGACCGA		ATATGGATT	HLFFADDALLFVRNKRSDIECLVTMLDIFSNISGQEINFD		
				CGTGACA		ATCCTTCGG	KSMILFSPNTPRAQRILFSDLLGMTVVEKLNNYLGLPIPI		
				TGTTGCT		AAATCCAT	GKKKTEAFKEITNRMSCRINSWTKRLLSFEGKEVFIKAV		
				GTTGAAG		GATGCTGTT	LQSIPTYAMSIFLAPKGIIDDIQAMLSRTWWSGKDKGGF		
				TCTGGCG		ATTAGTGAT	WTMIPWKTLCKPKGMGGLGFRDVRLFNLALLGRQVW		
				TTAGGTG		GTTACTTAA	RLLNNTDSLCFKVLSSKYFPDGNIFHAKKVDKASFTWSS		
				GGGATTT		TAATATGTG	IAKAVEAFQDGFGWQVGNGEKINIWADNWGMEGLNG		
				TCAGAGG		TTGACCCTT	AAVISNTLNHSEKFVKDLWHLDRRSWDENKVKQAYGQ		
				GTTTCTTG		GTGGTCGTT	DWGDKICDIPIGDEVQEDKIIWFHNTHGCFTSKSAYSWL		
				GCTCTGT		TTTGCTCAA	LLKEMGCGPHRFFWRALWKLDTLPKVRVFFTWRVGHEI		
				GACGTTC		AAAAAA	LPTNGKIASIRPGFNKGCPRCGAEVETLLHALRDCPTSR		

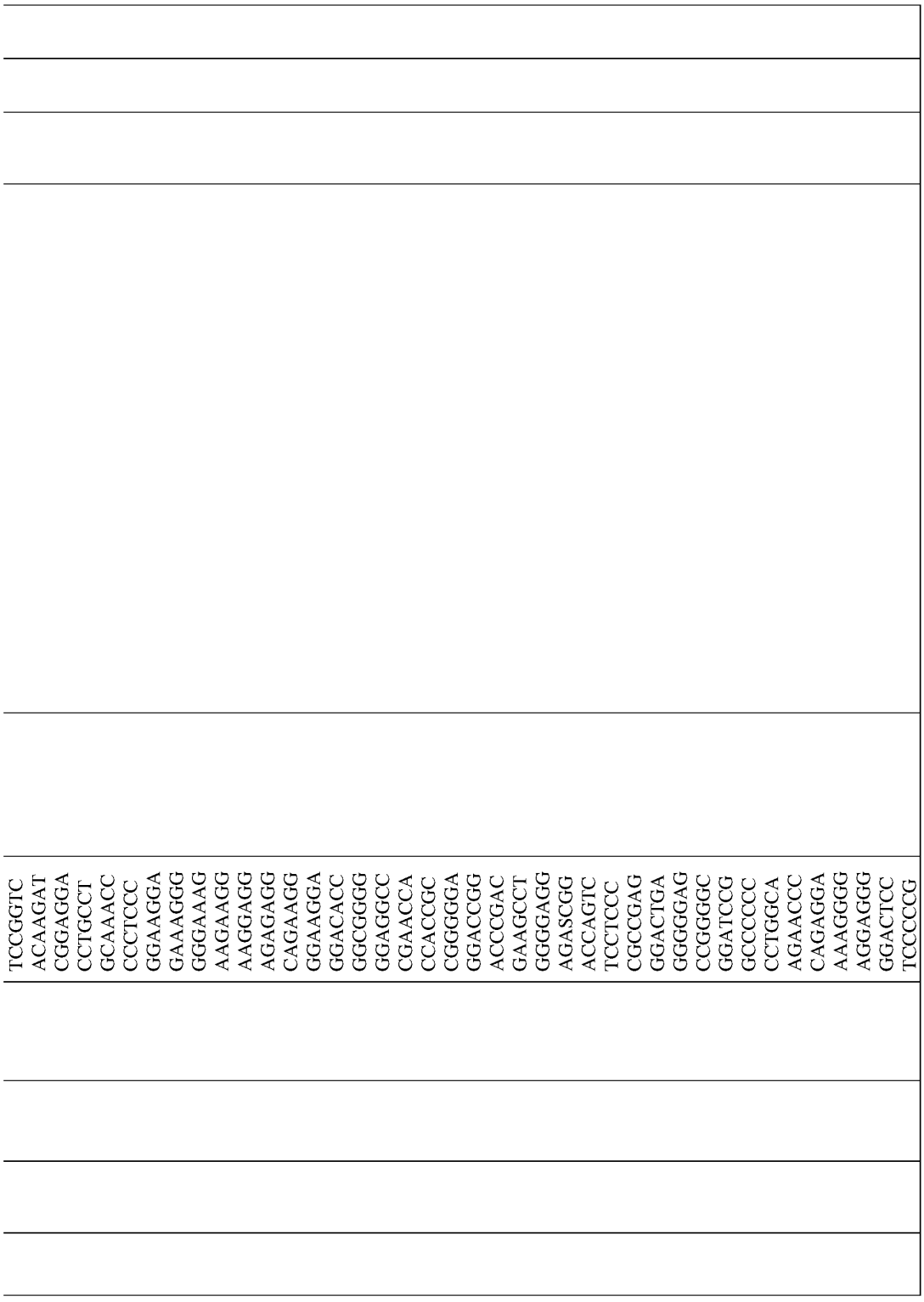
				GTTGCTTT GCATGGC CAAGGAG TTCTTGC GACTTTA TGCGCTT TAGCGGT GGCGGGC CATTCGA GGTGGGA CGGGCGA TACATAC GTGGCTC GTCACGA TCTTTTCT AGCATCG TCAGACC CTGTTTTG ATGGGAT TGTTTCAT TGGCAGC AATAAGG TTTTGTCG TTTCTTTT GTCTTGTT CAGCTAT TTTATTG ATTTTCTT TACGTTC AGTAGGT TGCATGG GTCTTGC ATGTATC TTGCATG GATTGTT TCGTCTT GCACGTA TTCCACT GGTGGTA TTTCTTAT		AVLSIGGWSRSFISKSYDHCIDWLEDLMRVLDKRAMADLMTILWNCWNNRNNFIFRGKEEEAQLIWERASNLSKEFRICNMLHAPLISQNIAEKRWVKPPKGFLKINFDATVGETGIGYGTVIRDEEGFVLGGGGGFKEGRRSVEEAECVAFEQSIDVACKLNIKDHVLFETDHAGLVNRINNLANDVTIIGARLKDCTVAFTFFKSASLI			

				TTTCTTTG ACTTTTC AACGTTA TTCAATC TGAGGGT TTTTTTGG TTTAACA GGCATTA TATGAAT GCGGAGG ATAAGTT TTAATGA TTTTCCCT ATTTGTTC GATTTAG TTGTTTTT GTTACAG TAGCTAT TAGGTTT CTGGGTT TTCTTGG ACTGCTG GTGTTAG GTTGCAG TTTTTTGG TACTTTCT TGCATGT TTTCCTTT CCTTGAT GTTCGGG GATGGGG GTTTGAT GGGGGGT GTTTGTTC GTAGCTT GAGTCGG GGTGTTC CTGGCTG TTGTTGTT TTGCTGT					

				GTTTTTGCAGGTTCTTGGACTGTTTCTTTGCTCTGCTCCCGGTGGCGGTTTTCGGGGTTCTGTTTTTCTCTATTCTTGGGGTGTTTCTTGTCGAGGTTGGGGGAGGCGACAAGAGAAAAGTGTTGATCAGTCTCCCTGGCACTCGTTTAAGGTAGCAGTCTCGTGAGTTGTCGTTTGAAA					
R1	R1-1_HSal	.	Harpegnathos saltator	CGCCGCGCTCKGCGGCCGTCACCATAACGTGCCCCGACGGAAAATACCATGAGTTCATGGTCGAGGCCAGGCGCCGAATCAGACCGGCGGACCTSGGGATAAACT	TAGCAGGGAGGGGGGTAAGGAGGGAGGGTGGATGTCCCCTCCCTCCTTGACAGGAGACACCAGAAGGCCAACCGGCGGGGCAACCCGCCGGGGACCGCGCCCTGGATTCCCTTCGTGGAACAGG	MQRMSELGAGLGVVAEPYWVPVRSVNWFGSPDGSVAMVANPAAGTPPCVFRENGSSYVAVDWGPITVVGVYLPHHRNKAGLGELAPVKRVLDRIGELVRRCHPRPVVIAGDFNAHSEGWGCSPRQRDPRGEAVIGWAAGLGLLLMNRGSTSTCVRPGGQSVIDLTWATPSAAGLFQEWRVEAEGESLSDHRHIAWALRLPKPARVLARCPGEGNFPLPKRWALNRLDTDRFRAAVLAATWHLDVGPGEWDPSPAEGAGEIVGIVASACDAAMPQSHPRRRRATYWWTEELENLRDSSVQARRRYYRARRRGGDVAGTKEGYLRAREALKAAIVEAKRAAWEQMVSSLDDDPWGRPYKRVMGKIRPWAPPPTESLDRQALEGLLDALFPRAEGGPPEIPTPPLGEGDWDEGLAVSPEELTAARRRLGAKGKAPGPDGIPGRAWALALGEGNLSAATCRVLSGCLREGVFPPEWRRAKLVLLPKEGKAPGSASAYRPICLLDEAGKMLERIVADRLVQHLSSKGPNLHDRQYGFRPGRSTLDAIQCFRDLTRAVVEEEGGVLLAVS	cd01650	474	707

CTTCCCCT	GCGCGGTA	LDITNAFNTLPWPEIGRALEHHGVPVYLRRILSAYFGGR				
GAGGGTC	CACAGGTG	DLAFPVEGGVQGRRAMERGVPQGSVLGPLLWDIAFDR				
CGCCGGG	GCGAGGGG	VLRTPLPPGCHVVCYADDTLVVAGGRAGGKPASERRR				
CCCTGAA	CTCCCGGA	HWHRWWEPSETWVSRWRPKRPRPCSFTAGVGGRRRRP				
CGGGGCA	AGGGAGCC	GSSPGGELGGSPPPSGDGIGIGGGNHQRPGSQGGAPKDP				
ATGCTGA	CTGGGGCC	GRVLSRQGSGGAAGDPGPRRWRPRPRRADDEISRPDAG				
TAGAGGT	GGGGGGTT	QPMELRTPLQGTGPQGAQGRTGCSQPDAHPRGPRLEGS				
GCCTGGC	GGGGAGGG	PPVRHGGALDRPVRGAHMGATAGGLPRRKARAAAGPE				
CTCAACG	AATACGGG	ARANQGDSGVPYRVVHGGDHPGRRPPCGAPRRGEAPP				
TCGGGGC	AGTAGTTCC	LLEGEGAPRGGEPVGQGPEGSEAPGRGPDPGEVEGLAV				
GTCCGCC	CCTGACTCC	RPPGARASDGRGCPPLPVRVAGRRGRGLSFRLAQVLTG				
GACCGGC	CCTCCTCAC	HGCFGRYLRRIGKELTTRCHHCPALEDTAHHTLAACPA				
TTGCGGA	GGTCCGAC	WAGERRVLAGAVGCREGEJSLPRMVKAMCGSEEAWA				
GGAGCTC	AGGTAGAC	AAVSFCEAVMSQKETAEWGRRPHRARGAAPGGRGTRQ				
CGCAAGC	ATCCCGCA	RARGGAGRGGLPDPRGPGAPSAARARHRSQGGSRGRR				
TGGCGGC	GGGGTTTTA	AAFPAFPPPSDRTSRDRGERCPGWGWARSPTNPPLPD				
GGAAAAA	GTCGGTAA					
GGCCCCG	GAGTCCGA					
GTTATCG	CATAACCCC					
CATCCAG	CTCCCCCCC					
CGGCCGG	AAGGGGGG					
CAAAGTT	GGGTATCC					
CGCAGCA	ATGATGGA					
CTGCGAA	TTTCCCCTG					
TTGCTGG	CGTGACCA					
GCTGGAT	AAAAAAA					
GCGACGA						
CCGGGGC						
CGAAGAG						
GTGGCGG						
CCGCGGT						
GGCCCTC						
GCTGGGG						
GATGCTC						
GCCTGAG						
GAAGTTT						
CCTCGGG						
CGTGATG						
AGCGTCT						
CCCAGCG						

				GGGAGTC GGCGCCA TGGTGGT GCGGTGT CCGCTGG CGGCGGC CGCTAAG GTCGCCG CCGCGGA GCGGGTT CCAGTGG GCTGGAC CCGCGCC GGGGTGA CGGCGCT GCCCGCT CGCGCAA CGCTGTG CTTCCGC TGCTTAG AGCGGGG GCACGTG CGCGAGC GGTGCAC CAGCGCC GTCGACC GCAGCGG TATGTGC TACCGCT GCGGTAA CCCCGGC CACCGCG CCAAGGG GTGCCAG GAGGCGG TGCGCTG CCCGGTG TGCGCCG AGGCGGG CAGGCCC					


TCCGGTC
ACAAGAT
CGGAGGA
CCTGCCT
GCAAACC
CCCTCCC
GGAAGGA
GAAAGGG
GGGAAAG
AAGAAGG
AAGGAGG
AGAGAGG
CAGAAGG
GGAAGGA
GGACACC
GGCGGGG
GGAGGCC
CGAACCA
CCACCGC
CGGGGGA
GGACCGG
ACCCGAC
GAAGCCT
GGGGAGG
AGASCGG
ACCAGTC
TCCTCCC
CGCCGAG
GGACTGA
GGGGGAG
CCGGGGC
GGATCCG
GCCCCCC
CCTGGCA
AGAACCC
CAGAGGA
AAAGGGG
AGGAGGG
GGACTCC
TCCCCCG

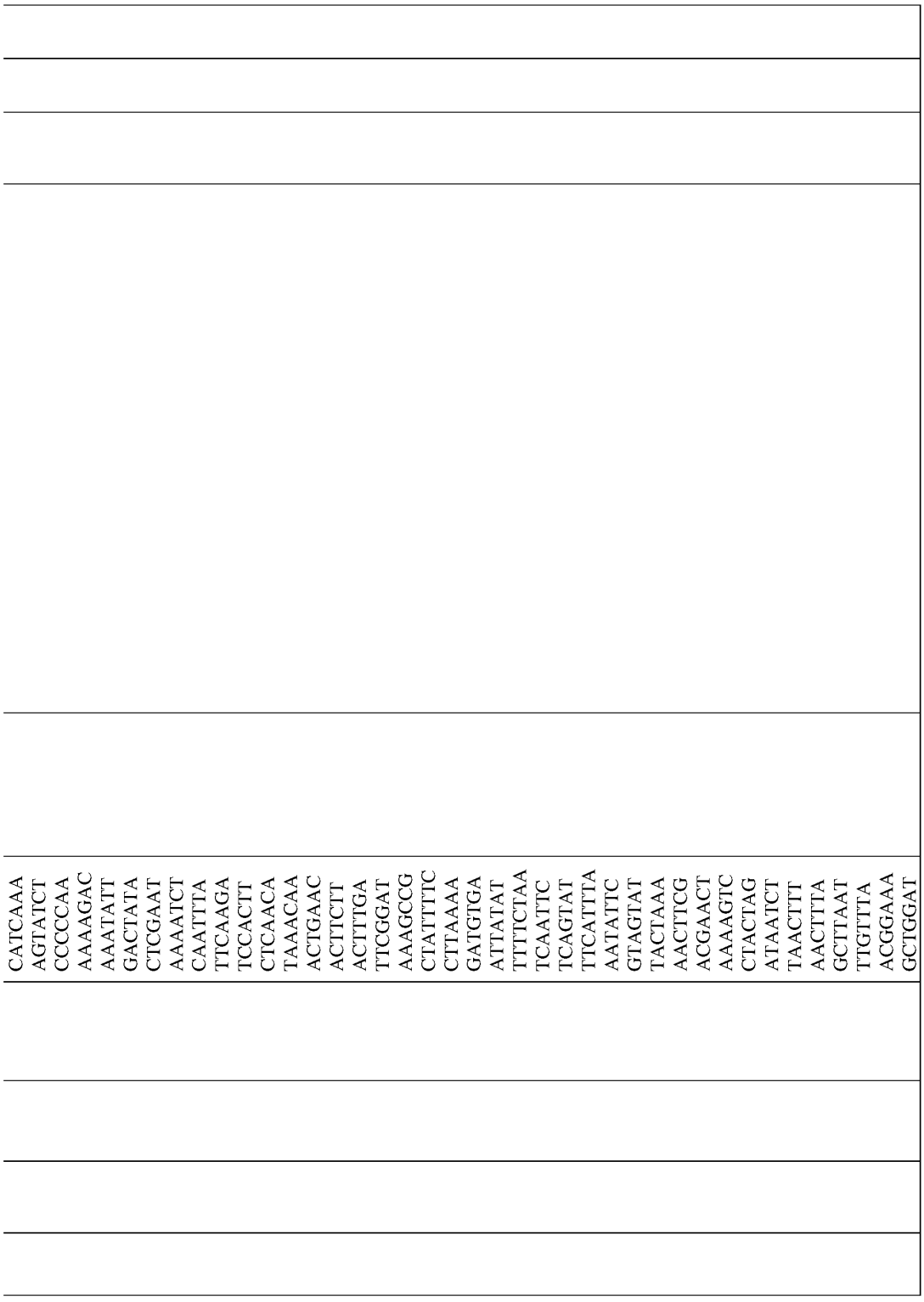
ACCTAAT
AGACTGG
GACTGGG
AGCCGGA
GAGCCCC
GATCCAG
GAGCCGG
AGGAACC
TCTGCCC
CGCAGGA
AGAGGAG
AGGGGAA
GAAGAGG
CCAACAG
GGACGAG
GACGACC
CGGAGGA
GGASGCC
CAGGCCG
CAAAGGT
GAGGAGG
GTGGAGC
CCTCGGG
CACACCC
CCAACAG
GTGTAGA
GGAAGGG
GAGTCGG
TGGGAGC
GCTGCCC
CCGACTC
CCCTGCA
GGTTGAC
TCCCCCT
CGTACCC
CCGAGAG
GAGGAGG
TCACGCA
GGTAGGG
ACCCCCC

Daphne	Daphne-14_HMa	.	Hydra vulgaris	GTGTCTC CGCGGAG ACACCCG CGTTGGA GGCCACC ATGGCCG AACCCCC CTCCGGG GAGACCG AGACGGC TTGCCGT CGGAAGT CTCCCAG GAGGAGG GGGTGGA CGTAGGC AGCCCGG AGGAGGC GTGTAAG GCTTAGC CTTCACG ACACTCC GGGCCCT ACAGGCC AACCTCG GCCACGG CGCCCGG GCTCAGG ACCTCTG C	GTGTATA TAAAAAC CTTCAAA TATTTTCA AAATAAC AAAAACA TACAAA	TATTTCAAG TTATTTGAT ATTTTTGTC GACTGTGTG ATGAATATT ATTTAAACA ATATATGCT GAATACTTT ATATTATAA TATTGTTCA	MDSTYRNDFEKISNLFQTNTFNLDNESDPDINYFNEINGS KFECSYFYPDKIKSIFKNNDNRDYFNIIHINIRSLKKNFDN FLNMICQTENFFNVICLTETWCTEEDFKSNSNLHLTGFN TVFLERNVNKRGGGVLFYIKENLAYNIRSDMSVSDDNIE ILTVEIINKKSKNILLSCCYRPPTGRTEYLSNYLVHNIIEK TSQEKKKNYILGDFNLDCFQYNEKKYITKFYNHLFETGT IPIINKPTRITKSSSSLIDNVLTNDYFNITLKKGIIKTDVSD HFPIFFCLGADHKTNTNGKITLKKRIYSNNNLNLFKEQL LHQDWSYINFNEDINIIYKSFFNIFYKIYEANFPIREIVLN AKDISCPWITKGIKKSSKIKQKLYIKYLKNKSDKHKNNY	cd01650	570	829

CR1	CR1-53_HM	.	Hydra vulgaris	CGCGCGC CGTTAAA GATGGCA GACGTGA CTTTGTG AAAAAAT AAGAATA ACAAAAC ATAAACA TTTTGTA ATAATAA CTAAATA ATAAAAG	ATATTGAG AATGTTATA TATCCATAT TATAAAAT ATTATGTTA AAACTTGCT TAACATCTG AAGAATAT TCTAACGCT TATGTAATT TTTAATATT TGTAAATTT TCTACCTTT TAATTCGTT ATATAGATT GTAAAGGG CTTCATGAT AAGATCAT GTGATCTTC TGGAAGTC CTACCAGTA ATATATTCA ATGTAAATT TACCTTGTT CATATTTTT CTGGAAAA AAAAA	TCTCTATGA CTATTATGA AACTGTTAA TAGCAATG GCTCATAAT AGATTCAA CATGGCGTT GAATCTCA AGGAGCAT AAATAAAT AAATAAAT TATTT	TVYKNLFEKLRKKAKQIYYSDLLIKHKHNSRRVWQIMK EITSKHKTNTNTMPNAIKFENKLITDSNEIAAEFNNFFSSI GHNLSNKIPYVDNNSVDEFISSPNSTINFSNLTYNEFEIAF KSLKRNKAIGPDDINGNIVIDSFNEIKDILFKVFKASITQG SFPNSLKIAKVTPIFKTGDHTNITNYRPISVLPVFSKILERI MYNRIYTYLIDNKLLYKNQFGFQRNSSTEHAILQVTRSI ADSFKNSQFTLGIFIDLSKAFDTVNHHILIKKLESYGIKD KTLNWFKSYLSNRKQFVYYNKESSSQIIKNITCGVPQGSIL GPLLFLLYINDLYKASTLTTVNFADDTNLFQSHENIEILF NNMNNELKRISYWFRKNKLSLNTDKTKFILFHPTSKKK RIQNILPDLFIDNTIIKREKFTKFLGVIIDENLSWSQQIDSIS MKVAKNIGVLYKARYILNKKQLTQLYYSFIHCHINYANI IWGSAHKTKLKSLYRHQKHAARIINFGNRFSSSTPLLIEM KALNIYELNIFNILCFMFKCKENLSPSVFINLYNLKPKNK YELRDSKNIQEPFCKTKIDQFCISYRGPFLWNKIVLPNFD FSFKWSYSSFKKKTKEIIFSTQNIFIYF	MVRKEDFEALVSRVLKLEAELQIKDKEINDLKIVVNDM KNQSKLVSNSNDVSNTWSDVVKGRKKNTSVQLDLLNV VVKESHDREIRQKNLVVFGFKTSTNKDESVAKEDDKQE LQNLFKQINVKVTIQRVFKLNTKTDKPPPVIVVLTNKEE RNAVLKTAKNLRNSPETNNIFINADMTEAERFKMKALR EERAKLNEANQDLKYYYGIRNERVVKLLKKLYESTQPY QQISPKKVLYYXNLNNITKNPNDRHLLNSNKRLIVYYTNS TSLNNKKNDFRALIDIEKPDVFAVTETWYTPLSDVNIEG YVIYRNDRLNGKKGGGVCIYVNSSIKSYEISNIELKSDLI EQIWTWIEYGQEKVLLGCIYRPPDSSCVTNKNIINSIIAA KRAQDKKQFTSIIITGDFNYNTLKWKDSVEESHSQQAED FVACLNDCFFEQCVTEATFQSNENELTNILDLVIVDDSN RIFGLSHLPPLGLSSKGHHILKWFYKMKDRIDDFSSSTK	cd01650	757	1007

				ATAAACT TATCTTGT CAAATTT GACA		KFAYRRGNYEKINVYIKSINWADEFQNKCIDKCYDKFL YHYNQSCKNYIPIVKLKNKASEVWMSNSLRKEIKYKNN YWTTNKKMNWSIAGSKEKYNKMKRNILKQINNSVKNY ERNLVNQVKNEPKLLYSYIRKKQKVKNHIRALNDNIGS TMYDKHSIVDIFNKYFHSVYIKEENINDCNFEVQTNSRF GKIFVDQALVKSYLSKLDINKSSGYDEVNPYVLKVCSES FSIPLSIIYQQSLTNSDIPDLWKMANVSPIFKNGSKLQAIN YRPVSLTSIPCKVLEKIISDNLKNYLESNNLFCKQQHGFL KGKNCITNLLESLDFLTYQLMIGNSIDILYTDFQKAFDRV PHKRLISKLCGYGIEKEMLNWIETYLSNRKQRVIIGDTKS DWLEVLSGVPQGSVLGPLLFLLYINDLPSKIKNKCELYA DDNKIIAVVNNQVDSKSLQDDINNLTEWSDKWMINFNI KKCKIMHFGKHNKNFEYFIKDSKLDSTTNEKDIGLLISN TMKWNQHVNMVVNKANSRLGLLSKSFTYKDKNTMKI LYCTFVRPILEYASPIWNPYNKNDIHKLEIVQQRATKLIP ELRHLPYEERLKKLQLTSLKIRRLRYDLIQYYKIFHNLD KVCWYQQPKTAISISANGPASNIRGHKARVDIELIKRNR PRENFFTNRVAKWWNKLPEEIISAKSVNIFKDKLDKHM VIMKL			
CR1	CR1-7_HM	.	Hydra vulgaris	AAACGTT ATCATCT ATGGACT AGAAGCA ACCCAAT GTTCTTCT CAGACGG AAAAACT AGAAAAT GATAAAT CGCAAGT AAATAAT TTACTTA GTGCTAT CGAATGC GATAGAG CAAAAGT GAGATCT ATTCGCC GTCTTAA ATCTAAA	AACATTAA ACCGATAC AACACGCG ATCTACCGG AAGATTTA AAAAACCT TTCCGTTCA TCAAAACTT ACGAGCTTT GCCATTTCC TATCGTGGC CCAATGCTA TGGAACCA TTTTACAAG TCTGAGTGG TAAACTTAA AGAAGCCA ATACCTCTC AAAAATTT AAAAAGAA CATAATCG	MMPIINKSTRITQHSCSAIDNIFTNTLFHSTFKSGIIRSDIT DHFPVFAIFKKPKDLQKKDWYFKKRIFTTRAKNMFKLE LSTVNWNFVFLETCANNSYDLFLDKFFNLYDKHFPNKII RIKTKGLLNPWFTKSLLKSSKRKQKLYIRFINSKKERDK HNYLEHKRLYERTKKIARTTYFKNKLLELKGNCKKTW DAIKEILGKQKIHNKALPSVITVDGTQISDVKSVATKFNE FFVNIGPNLASKIKHTPAHFLSYLSKSDKSLAFKKLTSKE LDDSIKKLNKNKATGYDAISASVIKDCYELIKDVLLDIFN KSLKTGVFPDKMKVAIITPIFKSGDESLVTNYRPISVLPV FSKILERIMYNRIFAFLYDNNLLYYNQYGFRSNYSTEHA LMQLVEVISEASSHNKFTLGIFLDLSKAFDTVNHQILLEK LKYYGISHEYHKWFTSYLTDRKQFISLNDGFQTNKLKIS CGVPQGSILGPLLFLIYVNDLYKASNTMSTIMFADDTTL FYSHSNIISLFQVANNELEQISNWLIANSLSLNADKTKFS LFHSKRKFQSIPKSLPMLYINNINIERSFTNKLLGILIDENL SWRDHINHISNKISRSIGILYKTRQILDKSTLTQLYFAFIN SYLSYGNIVWGSTNKSKLDCLLKQQNHAARIIHFQNRFT NAKPLLKSMRALDIYELNIFQNYTIYVSTSVKKNAEGFR KYFP	cd01650	330	590

				GGAACTA TCGATCC TTTGCTTA TCGAGCT TGAAGAC GTTAATA ACAAAAA TATGGTT CTTAAAG CATCAAT TAAATTA AGAAAAA TCAAAAA ATATGAA AATGTGT ATATAAA TCGCGAT CTAACAG CAAGTCA AAGATTC ATTCTAA AAAATAA ATTAATA GAACGAA ACGAAAA AAATGCG TTAAGAA GTGAACA AGAAAAA ATTAAAT TTCATTA CGGTATG ACAATCT TACTCGA ATTTTTCA CAAATAG CCACTAG CAACCAC TAATTTA GCAACGT	AACTTATTG CCAATCAC GATTCTTTA CATGATTTT TTCTAACAA TACACGATT ATTTCCAAT TTTTATTAT CAAAGGTT ACTGATGA AAAGACTTT ATCTTCTTC AGGTTCTCC GGTCTTGTA ACTATTATA AATAATGA AATGTACTT GTATTTAAT ATTTCTAAA TATTTTACG GAAATTATC TTTAATTGT AAATGGTT GCTTGACA AAAATTGA AATAAAAA AAAAAA				


CATCAAA
AGTATCT
CCCCCAA
AAAAGAC
AAATATT
GACTATA
CTCGAAT
AAAATCT
CAATTTA
TTCAAGA
TCCACTT
CTCAACA
TAAACAA
ACTGAAC
ACTTCTT
ACTTTGA
TTCGGAT
AAAGCCG
CTATTTTC
CTTAAAA
GATGTGA
ATTATAT
TTTTCTAA
TCAATTC
TCAGTAT
TTCATTTA
AATATTC
GTAGTAT
TACTAAA
AACTTCG
ACGAACT
AAAAGTC
CTACTAG
ATAATCT
TAACTTT
AACTTTA
GCTTAAT
TTGTTTA
ACGGAAA
GCTGGAT

				TTTAGAT ACAACAT TTAATCA AAATTCT AAATTTA TATTGCC CAACTAC AAAGCCA TTAGTTTT GAACGTA AGTCTTC AAAAACA GGCGGTG GTATATG TGTCTAT GTGCTTG ATGAATT TACTTAT CAAATTA GATATGA CCTGTCG TTCTCTG AGCCGGA TTTTGAA TCTTTCAC TTTAGAA ATCTACA ATAAGTC ATCTAAA AATATTA TTGTGAC AACTACA TACCGAC CACCAAA CGGATGT TTAAATA ACTTTCA AAATCAC TTAAGTG ATTATTTT					

				GAAACAGCATGCAAAAAAAATAAAGACCACTCATTTTTATTTCTTGGAGATTATAACGTCGACTATTTAAAATATGGTGAAAATAATAAAGTACAACAATTTTACGATAGCTTTAATGAGTTTGGA					
CR1	CR1-5_HM	.	Hydra vulgaris	AGAAGCTGGAATGTTTGCGTATATTTCATACGACAAGTTGGTGGTTCGCGATGGTCTCGAAAACCCACGTAAATATTTCTTTTAACGTTACTTTATTTTTATTTTTTATTTATCAATAGTTTTTCAAATTTTAATTTAACTTTTAGAA	AAACAATTTAAAGGTTTAAACTAAATCACTTAAAAATAAATTAAATCTTTAACACTTTCTTTAGCTTTTATTAAAATGGTTTCAAACCATTAGTATAAAAGATAACGAAAATAAAATTGTCTTTTTCGTTCATATTTTTGATGATATTAAATAATCAGCCCAAATACTTATGTTTACATATATTGCTTTTT	MENNLSYNINAKNIVFSDNNSDNHTETLLQSEYYSIPESTQYLCENKNKFSILHINIRSMNKNFENFKCLLGELNHDFKIICFTETWLKSNETNSNFELRNYTSVHQMRDICVGGGISIYVHNSIDFIQRKDLNVNNTDCEALCVEIINKLAKNFIINAIYRTPAGSLKTFKTYLRTFLTTKNILQKHVYVVGDINIDLLNHALNSEAKTFIDILLEYNLIPTINKATRVTKKSSTLLDNVITNNFHNSRFKTGIIKTDLTDHFPIFFITESVTLNNATHKSTVFMRQINESSICQFKNLLNNYVDWNLVLQSHDVNNAYDLFLNQFSKMYDKAFPLKVKLINSKSVXSPWMTKGLLKSSRKKQKLYDKYLKNKTYKNETNYKNYKNLFEKTKKQSKVNYYAKLLEKNKGNPQKTWSVIRDLIGKNKIEKNNLPQKLIIEGKMIYHKEVIIEKLNNFFLDVGPNLAAKIPIGQKKFDSYLATTDLIMEEPILTKSELHTAFNSLKKNKSAGIDQINVNVIKSVFDIIEPSLFHIFNLSLKSGHIPDKLKIAKITPIFKSGDETNISNYRPISVLPCFSKLIERIMYNRLYKYLSENKILYNNQYGFKKNHSTDHAIXELVKHISNGFNSDCYTIGVFIDLSKAFDTVNHEILLKKLENYGVKNQSLLWFKSYLTNRKQFILKESKDSKNNLITCGVPQGSILGPLLFLLYINDLYLSSKVLNTILFADDTNLFYSHKDITVLFKIFNEELDKINEWFISNRLSLNVEKTKFILFHKPSKAENIPLKLPNLLINNKIIKRESTTNFLGVLLDEKLSWKFHIKYIEGKISKNIAMLYRTKPFLNNESLKNLYFSFIHSYLSYCNISWGSTNHAK	cd01650	551	810

TTTTTTTTT AATTAAAC AACACATTT TTTTTGTGT GTTTTTATA ACTTTCATG ACTTTTTGT CGTTTTTAC TTTTTTATT TTTAAAATT ATTCTGCTT GAAGATTG TTAGCAAC AAATTATTT TATTGCTGA AGTTTTTGA TTTGTTTTT GTTTTTGTA TTTATACTG TCTTTTTGT CTTTTATCA TTTATTGTG CATGTAAAT TTTTGTGTG ACGGGGCT CGGTGATA AGACAACC TGTCTTCTT CTTGCTCCG GCCAGGAT TTTCTTTTTA TTGTAATTT CTTTTTTGT AATATTTAT CAACGGCA AATTAAAT AAATAATT AAAAAAAA A	LKKIYSKQKHACRIVFGANRKTQCEPLLRQLGALNIYKL NIHQVMIIFMFKAKLGFSPVTFQSYFNEISHKYPTKFSVN NFVVPINFLKLSSYQIQYRGPLVWKEFLPIFSKKQNFVLN CTLQSFKDDSKRHLLNMEFDILDFKYLF					

CR1	CR1-42_HM	.	Hydra vulgaris	TGCGCGATGAAGGAAAGATGGCGGTCTWAATTCTGAGTTGAAGTTTGGTATACTTTTATTTTAWCGAGTCGAAATTTATATTTATCCATCTATTATTAAATCAATGAAAAAGATAAAAAATTAATAAAGTAATTAGTATCAAAATAATTTATTCGTGATCAATATTAATACAAAGTAACCCTAAAGTAGCAAACAAAAAAACGAGCCATATTTCATTATCCATCAAACAAGAAGAAAAGAAGCAAGAAGAAAAGTATAACATAAATAAAATGAGGAGAAGAA	ATATTTCTACCAACTAATATTGACATTCTTTACTATATATGTCAATTCGTAATCCTTCAGTTTTTAATTTGATATTTTTAACATTTTGAAATTCTCTACTTATCTATCTACATCTAATTTCTCTGGTTTTCTATCTGCCATTTTTAAACTTTATTATTAAACTATCTATACTAATATTATTAATTTTAGTACCAATGGGTTGGCATTCTTTATTTATCTATCTATTTATCTCTTTTCATTCAAAAATTCTCTATTTCTCAACTATTCTACTTCGTACTAAATCTTTGTTGTTGCCATCACTCTTTACTAAATTTTTGTTGCTGCTGTTATTGTCTTTATTTATGAAATTCTTCTCTTAT	MKNKTLCIFCFKTVTDNHKAIYCDCCSSWVHAKCNQIDLLTYNLLSEDSSEWYCFNCISKNMPFGLLTDNELLTTLSCKNPQTASSNFLVSPIHSRNILKELNTISGSAINCKFYDVVDFNKAIKPNSDIYLHLNIASLPYHIDDLCSLISTLNIIPMAIGITESNLYLNDTNITDITINGFNIEHCPTEAKKGGALLYLQSNLNYVVRHDLKIYLPKLLESVFVEIIKPHKTNVLIGCIYRHPSMNQNEFNTTFLTPLLEKLSKEQKLIFLMGDFNVDLLSYSESNPTSNYLDTLCSYSFNPSIILPTRITTKSQTLIDNIFMNFYSSDHLSGNLTISISDHTAQFVCVPGTSNQPNKVKISRCLKRCLKKFDPNKFIIEISKINWELIIKKDDNINESINNFLKSFNRVLDQCAPYKTLTKNQIKLKSKPWISKGLLKSISVKNKLYKKFVRCKNINTKNELFTKFKFYRNKISNLLKSSKKSYYITYFNNNINNIKNTWKGIKDIINIKGSLHKTPTQLNLNENIITDHATVSNIFNNYFVSIREKLLNNIMPSEYTYSDFLNFPNANSFFIDPVSEDEVSSLISNTLKNDKSYGPNSIPTFFLKLTSHIISKPLSIMINNSFKNGKFPDVFKVAQVIPIYKAGSLLDFSNYRPISLLSNLSKLFEKAMHNRLYKFLDKFKCLYQHQYGFRSKHSTTLALIEITEKIRRALDNKHFACGVFVDLQKAFDTVDHSILLTKLEYYGIRGIPLKWFTSYLKNRTQFVSINSTKSDLQNCSNGVPQGSVLGPLLFLLFINDLNTSFKFATAYHFADDTNMLLVDKSLKKINKHINHDLGNLVQWLRSNKLSLNSNKTELIIFKSNLAPIKKQLNFRLSGQKIYPLDSIKYLGIKIDSNLSFQGHLNDLAMKLSRSNGMLAKVRHYVNLETLLNIYHAIFGSHLRYACQVWGQSHKQIILRLSHLQNKALKIIYFQYNNSKSNVLYLISKILKISDLVVFLNCLFVWNYHHKNLPLSFNNFFSITNCHYPLRSISSLNLVIPKHQSAKYGNKSIKYQCISSWNKLPPDLKSLNSISEFKIKLFNYLLKKYN	cd01650	640	894

				CCGAAAAAAAAAATAGCACGAATCTATTGGAAAGCATAAGATTTAAAGATACTTAAACCTCCTTATTTTTTTCTTTTGATGATTTTTGGTTAGCCACATTTGTTTATTATTTTTCTGGACTATTTATCAATATTTGTAGGTTATTTATTTTCATTACTTTATAGAATTATAATTATATCTAAATAGCTAATATATCTGTAATATAATTTATTCATATATATAAAATATTATTATATACCTATATAACTATCTATATATCTATATAGCTTTTATTAGTGATTATATAA	TATAATTGTTATTATTTCGTTACTATTATTATTATTGTTATAATTATTATTTTTGTTGTTGTTGTTGCTATTATTATTCTTATCACAATCATTGTTATTATTGTTATTATTGTTATTATTATTTAYTGTTATTATTGTTATTACTATTATTATTGTTATTATTGTTATTACTATTATTGTTATTATTGCTATTATTAGTATTATTAATAACAGTTATTATTATTA				

				ATGTCAT ATAACTT AATCTTT CAGTTTA TGTTAAT TATATAT TTACATA ACTATCA TTATTTAC ATAACTA TCACTAT TTACATA ACTATCA TTATTTAT CATTTGA ACTTTATT TATTATTT GTACATT TCTCTTCA TTTATTCT TTTCTTTG TGCTTGTT TTAACAT TTTATTTG TATATTT ATTTTTAT TTATTCTT TTCTTCAT GATTATT TTAACAT TCTTCCC AAGTATA TTATTTTA TTATTAT ATTGGCT TGTTAGT GTTTTTGT TAGTTGT CAACATC TATTTAA					

				TCTCTTCAACATCATCTACTTGATTTTGTATAACCTACCTGTACTGTGTTATATATACTGTTCTTATGCAAATCCATTTTGCCTCACAATTACTCTCTTATTATATTATTATATTATTATAAGTTTATCTTATTTATTATATCCTTAGTCACTCCTTTTTTTTTCATCACCTTCTGTGTTTATCAATAATAAAACATAATAACATTCTGCCTATCAAGTT					
CR1	CR1-20_HM	.	Hydra vulgaris	AACATACAAAATGGATAACGATAAAGTAATTACAATGGGAATGCTTCGAGAACTCATGTCAAT	CATATATTTTATAGTTTACAGTATTTGTTTTCAGTATTTTTTTATTTTTTTATTTCTTCTTCTTCTTATTTTTCTTGTTTT	MSENSLYNLVSYKMISQPRVNSKGGGVAVYILDKFEFKKNNTFSKTNKHIETLCIEVLNTSGKSFLISSLYRPPCGKQVNFLNILKXFMSQMSKQNKHVFFAGDTNMDALSYEKFANIKNFFNLLFEFNIIPTIFKPTRITNQSSSVIDNILTNCSLNTKFESGLFIADFSDHFPVFHIAREVLSNIDNKKCFVTYRNLSIKKLKKNLNNFKNKLQVEQWDKVYSSTDPNIAFDHFSSTFQNIFDNVCPKITTEIKNKEYKNPWMTITLIKASKRKQKLYVKFLKSKNKEDEKIYKNYKKFFQHSMKEEKIRYFSNQLDKHKFDIKRTWAVINELTGREKKKHTCTCLPQKLI	cd01650	484	751

				ACAAGGA GATACAA TTTTCAA ATGCTTT AAGGAAA GTTCGGT CAGTGTC CTGGAAA AAGTTGA GAACTTA GCCAAAG RCATTGA AGAAATA AAACGTG GGCTTAC TTTTGTTG GTGATCA AATTGAG ACAAAGT AAGCAAG GTTGATA GTAAAAT GGCAGAA ATTAAAG AAACTAT TGTGTAT TAAATCA TTCCACG GAAAATT AAGCATT GAAAATA AAAKAAA AACTGTC GATCTTG AAACCGA CGTAACA ATTTAAG AATTGTT GGAATCA AAGAAGA	TAATATTTT TTGTTTTTG AATTTGTTA CACTAAAA AAATATGA GCTTTGTTT TTATTGAAT TTTTATTTT ATTTTTAAA AATTAAATT GTTTTTATA ATGAGTGA CTATAGGG GCTCTATGA AAAGGTTG TGATGTTAA ACACATCAT CCTTACCTT CTTTGAGTC CCTGACTGA TTAGAACA GAATATAC ATAATATA AAATAAAA AATTATGA AAAAACGT TTTTTATCA TGTTTATCT TTACGACGT GTATTTTAT GTATTTTTT ACATGATTA TATGAAAA TTGTTATTG TAACTGTTT AATCAGCA AAATAAAA GTWTATAT AATAATAA TAA	SDKKNYNKYNKNMSRIFPIKKIITSTTKICQEFNKYFVNV GPSLASNIVQPNKTFDNFLGEKPTTSINNEKITKLEFNKAI IELKRNKSCGYDGISSNVAIYIMDSISKPLFYLISSSFENSI FPNKLKLAKINPILKKGDCNNVSNYRPISLLPVLSKVFER IIYNRVNNYFTLNNLLYKNQFGFQSKCSTEHAIIKLADKI FKSFDSDELVLGVFVDLSKAFDTVDHSILLSKLKYYGIIG STFKWIQSYLSNRKQFVLLEKSGALDITCGVPQGSILGPL LFLIYINDMHKASQKISTIMYADDTNLFFNHPBAKTLFET MNIELEKFNLWFIANKLSLNCQKTSFTLFHKKKNNQLIS LKKKQSINLPLKLPKLLINKNEIKREKCTKFLGVLFDENL SWRNHIGLLETKMSSVIGVLYKSRPFLDSKSRNMLYFSL VHSQLLYANIVWASTHKTKLLKLQSLQNHACKVINYLN RLVNPAPVMKSMMVLDIEKLNTLQILIFMYKYKNNLLP SVFSDIFLISFSEKYNLRSNTRYNYHLGKIKSHQSNFLITN RGPSIWNHFQNKNIKDLKSISSFKRETKMHLLNS			

				AWATAACGAATCTTTTAATGATGTTGTARAAAAAGTAAACCTACTAATTAAAGAAGATCTGGAAATTGCGGGTCYAGTTATTATAGAAAGAGCRCATCGTGTCGGTAAATTTAGTGGAAAGCCTAGAACTATTGTCTGCAAAATATTAAACTGGCAGGACAAGGAAAACATCTTATCCAACGCTCGTAAACTCAAGGGGAAAAATATATATATTAACGAAGACTATAGTGATGAAACAATGCGAATTCGTAAAGAACTTTTTGTACAAGCYAAAATACAT					

				AGTGGTA ACGGAAA ATACGCG AAAGTCA TTTACAG TCGATTA GTGGTAA GAGAAAT GATAAAC GAGAACG AAGGTGT TGAAAGY AAGTAAT TTATTTAT TTACGTTT TAATAGT TTTATTA GATATTA AATTTTA WGGTTTT TATTTAA GGATAAT ATTAACA ATGAATA TRTNTAA AACAATT CATAAYA ATGAAGC AAATTTA ACAAAAT TAATATC TCCAATT TTTGAAA TTTTGAA ATATAAC AAAAATG ATTAYAA TTTTTTGA CTGAAAT AGATATT					

				AAWTGCTTAAATAATGAATACTATAAACCGGAAGAATTTAAGAGAAAAAAAGCTTAACAAAGATTTAAACATTTTAAGTATAAATATACGGTCAATGAACATTTTTATTATTCTTTAAATTATCTTTTTGACGTAATATGTATATCAGAAACTTGGGAAGATGCAAACTTGCCC					
CR1	CR1-3_HM	.	Hydra vulgaris	AGGTCATCGCAAATGGCGGACGTATTTTTCTAACGGTAAAAAATAAATAAAAGAAAAAGGATAAGATAAACAGTTAATTTTTCATTTTCAATTCGCTAAAAAGTTTT	ATCTTATACCTTTTTACAACTTATTATTTATAACCCTAACGCAGTGCTTCTATTCATCTTTTAATTTTTCTCATTTCTATCGTGATATATTACTAATTTTAATATCAATTATTAAATTTATTATTATAAACTG	MKICHHCKFCNKKIKDNHRAIQCDCCSFWIHAKCIPVDKHAYNMLSNDSSEWFCPNCISNNMPFGVLSDLELKSTLSCKNPSNKIKFHEPPDYLKNLFKTMNNVFNPTTKCKYLDVNELNKSINPGTEIYLHLNISSLPFHINDLSSFVGTINTPPMVIGITESNLYANDSNITDITIQGYNIEHCPTESKKGGALLYLNSNLNYIVRSDLQIYATKFLESIFVEVIYPLKSNTIFGCIYRHPSLNITEFLSIHFNPLLEKLSHEAKNIVLMGDFNIDLLNYRESQVISNYFESLCSHSLFPTIILPTRVTAKTKTLIDNIFMNSFPTDIVSGNLTISISDHMAQFVCIPNNPPIKKKVKMFKRSFKKFDSDSFIQEISDINWELLIKDDDNINESINFFLKTFEKILDRYAPYKELTKKQIKLQSKPWITSGILKSISNKNILYKKFTRSKNVNTKNILFTKFKLYRNKISNLLRYSKKLYYASFFNNNINNVKNTWKGIKEIINIKPSTSTKSFNLKINENVISDHTSVANIFNNYFTTIQDELLKNTVPSKCAFGNFLKIPSINSFFIEPVTENEVANLIKDTLKNNKSLGPNSLPTF	cd01650	634	890

				TTTAAAT	ATATGTTTC	LLKLVSHIISKPLCTMMNNSFKNGIFPEAFKVAKVIPIHK			
				TGTTAAA	TATAGTATT	MGSYLDYSNYRPISLLSNLSKLFEKAMFQRLEHFLEKHK			
				AATCTGA	CTCTTATCA	FIYKHQYGFRNKHSTTHSLIEITEKIRQAIDNKHFACGVFI			
				TTTAAAT	CTGTTGTTG	DIRKAFDTVEHTILLEKLKHYGIRGIPFLWFSSYLCNRTQ			
				CAAACGC	TTATTATTA	FVSINGINSGLAKSFNGVPQGSVLGPLLFLIFINDLNVSLK			
				TTAAAAA	CTATATTTT	FSTAYHFADDTNLLLINKSLKKLNKNMNHDLANVVQW			
				TAAAAAT	ATATCGTAA	LRSNKLSLNSKKTEIIIFKSAKTKINKQLNFRLSGQKINPV			
				CAAATAA	TGAATATTA	NSIKYLGIKIDSNLSFASHLQDLALKLSRSNGILAKIRHFV			
				AAAAAAG	TTGTAATTG	NHETLLNLYHAIFHSHLRYACQVWGQSKQLAFLRLTYL			
				AACTTCG	TTATTATTG	QNKALKLIYFQHTNSNCSILYFLSKVLKLCDLIQLSNCLF			
				AATTATT	TTGTTGTTA	VWNQNHNNLPLTFINFFSYRENCKYILRSALNFKLSVPK			
				AAAAAAA	TTATTTTTG	YRTVHYGYESIQHKSIQTWNNLPSQLKSLKSFSKFKTAL			
				AAGAACA	TTGTTATTA	FNHFLEKYSL			
				CTTTAAA	TTGCTATTA				
				CAAAGAA	TTATTGYTA				
				AAGAACG	TTATTATTA				
				AAGAAAA	TTATTATTA				
				AAAAAAA	TTATTATTA				
				AAAAAGA	TTATTATTA				
				AAGAAAG	TTATTATTA				
				AAAGAGA	TTATTATTA				
				AAGTAAA	TAGTATATT				
				CAAAGAA	CTTTTTATA				
				GAAGAAA	ATCGCTACT				
				AAAAGAA	ATTATTTTT				
				GAAATAA	TTTATCACT				
				GTAAACA	ATAATCACT				
				AACAAGA	ACTATTACT				
				AAGAGAA	TTTATTATT				
				GAACAAA	CCTACTTGT				
				ACCAATT	ATTATAATT				
				TTAAAGA	ATTATTATT				
				GAATAAA	ACTATCTCC				
				GAAGATT	ATCTTATTT				
				GGTTATT	TTATTTATT				
				CATTTATT	TTGACAGGT				
				TTCGCGA	GTTTACTTT				
				TTTAATTT	ATATTAGTA				
				GTAATTA	CAAATTACT				
				TTATTTTT	ATTTGTAAA				

				TAATATA ATTATTT AACTTAT TTGTTATT TTATTATT ATTATTA TTATTTTT TTTTTTTT TCACTTTT TCACATA TTTTTTAA TTATTTAT TTATCATT AATTATA TTCATTA GTTCTAG TTAATAA GGTTTTC CCGTTAT CTATCTA AGTTGTA TTATCAC ATATTAT TTATTTAT TTATCATT AGATTAT CTATTTA GGAAGTT TTGTTAT ATATTTT ATTATAT TTAAATT ATTTTACT ATATATA ATTTTATT ATATTTTT TGATCAT TATTATA ACTTTAT CATTATT	TATCCTTGA TGTAACAA CTTTATTCA GAATATAT WTGATTTG ATTT				

CR1	CR1-12_HM	.	Hydra vulgaris	ATATATT CTTATAC ATTTATTT ATTATGG ATTTATA TTATTTTA ATTTTATC TATTGTTT ATTATTTT ATCTTAG TTATTAA TTAATAT TATAAAT CTATACT CTTTATTT ATAATTT CGATAAT AATTAAC TACGTAC TACTACT GATTACT GTTGCTA TTTATCC ATTTACT GTTTATT GTTTGAT TATTTTCT CTTTTAAT TTTCTATA TTTTCTGT TCTATAA ACGGCTT TACCATC TATATC	AAAGGTC TATAATG AAGCCTA TTTTCGG AAAAAAA GTTGCAT	TCAAATAC AGGAAACA CAGAATAA AAAATCCTT TACAAACG AAGCGTTTG	MEGNIESKTFENISFCAYNCQSFKSNSIYISELVKSFDIIFL SEHWLLNIELFVLKNIALPTHKVFFHAAEKKAFGRPFGG NAFIIKKDVLLSPHIIYEDDNILAIKGKNKCHNLVFIGIYL TSCRNNEESLSKYEQQLNLITSIVKNYEDVSECIIAGDFQ SYPETLYDSEIRNNKKRNKFSNHLTNFIKSNKLSFIDIISG AGPTYTYQHKTLSNSSYIDHIAALKESSLSFINTSVIPPSP	cd01650	542	799

				TTAACAG AAACTCT ATTGCAG GCGAACG TACGGTG CGAGAAT CTTATAT CTTTGTA GGAGGGG TCTGTAA CTCAATG TCGGAAG AAGATCT ATCTAAC CACATCA AAAATGA AATAGGT ATTGAGC CTATTAA GGTAGAG CTTAACC GGGAAAA TAAATAC AATCGTT CATTCAA AGTTATC ATAAAAA GTACTGA AAAAGAC ATAATCC TAAACCC TGATATT TGGGATA CTAATAT TATCGTT AAACCAT TCAGGCT TCATCGT GTTACCT CAAGAAA	ATGCATGC AACACACTT GATTTAAAT TTCATACAG TGCATGTTA AATGGCAA AAAAATTA AATTGACG CAACCCGTC GTAGAGCT ACAAACTA ATGTTATGC TAATATTGA AAGAAACG TTCCAGTTT TGGAATTTG AAAGAAAA GAGAGAGG CGTTTAAAA TCTTGATGG AGGAGAAT GTTCCTAGA ACTCTGTCC CAATAATAT GTGGAAAA CAAACCGG CTGTTTATA AGCATTTAT TGCTAACA ACTACAAG AGAGCCCG TGTTGTTTT AGCTGATTT TTAATTTAT TGTTTGACT ATTTAAATA TTATTTTAC CTGTAATTT ATTACCGG GTGATAAA	FNFSDHLPIATNIVVTHTNLSTVINSIMHKYNYIPKYAW NNEKFINIYNHNLSVAFNYYTFNEEEIEQELKQTCEKIQN AALLAVKQCFKQQICKYSKSWWTTDLKKCKENLSFNY KQWQKSNYNKTQECIEYKRYIFSRKNFRKAVKAAHNK KIHEKTINIENLRTTNPQRFWDNIRKIKTKATTRLFTINKS QNIKDIVQEFRQHFQTLLNTPLITNNNHNPSQIPPLSKEP NSLIITSADIINCISQLNSNKSPDSCGLSAEHLKNSHNNNL LLWLAKFYNSILSNGKVPNDLSTSTIIPLVKSYKKSLNNP DNYRGISILPIFTKLLEYLILLICPSITDSHSHQFGFKKNSS TLHAEFLLSETVKHYKNNNTPLYMCSLDANKAFDTCN WDLLFEKLYFQKKLPLSVVHTISSLYHSGTANVSYQGV TSNQFSLSQGVRQGSILSPHLYNIYTESILNEIVSECKVGS TINGIYTGVIAYADDVILLSPTISGLQELINRFQTYGVANF IKLNTEKTEFLISGKSYIXNNVIKINGVSIYLQNKLKHLGF QWDNDTSKNKIATLQKTNLNERITQFQSVAKALIDSGIR YCQPSTIVHLYNTLTVPKLIYGLELCKYDTNFLSKIDAV GRAVLKSFFCISKYSKNYLHSFFKIEDISTILLRNKVNLFI RLLKNQTCLXTDN			

			CGAAACA CCACAGC AGAGTGC AGTTAAT ACGCTAC ACTCTGT ACAACCA ACTCAAC AAATATT TTCTACA TCTAAAG AACAACC ATCCTTT GCACAGC AACATTC TTACTCA ACTCGAA TACTAAA TTCTAATT CAAA	GAGAAAAT ATAATATA ATAATAA							
			GGGAAAT TTTTCAA GAAATGT TYCAAAA ACAACAA AANGATA TTTTAAM ATTAATA AGTGGAA ATTTAAA ACTGACT AATGATC GAATCGA TGGATTG CTAAAAG AAATTAC TACATTA AAAGAA WCTTGTG AAWCTTT	AAACTCAA AAATTATG GTATAAAA AATACAAA TTTAGCTTG GTTTAAAA GTTACTTGA CTAATAGA AAACAATA TATTTCATA TGATGGMG GAAAAACT GAATATAT GACAATTA CTTGCGGTG TTCCCCAAG GATCAATTC TAGGACCA CTTTTATTT CTAATTTAT	MIFLRFYIIFLNFNIALSYFLFKMADETVLTNLNFNSFESK KSILLNNYSDPDINFYNNDEVIKNINPLYYNPNSKNIANG MDDNSFSMLHINIRSIQKNFESLKQFLYKIKINFQIICLSE TWCHDKNIENNSNFQLPNYKVIHQVRASNKEGGGLCIFI KNSLLYKVKPNLSSTTNDYESLCIEIINETSKNIIIHALYR PPSGSIKVFENHIKKIIKNKSSINKTVYFVGDINLNILDYD KNKNIKNFFNVIFQNSYIPLINKPTRITKESATSIDQIITNE FINKKIKTGIFKTDISDHFPIFIJSQKCIYNAQNKRKKIITTR LINDNSIKHFHNLLLSCVNWDTLKLNLNABKAYDIFLTEF LKLYNKAFPEITKFIKTKTLLNPWITKGILKSSKKKQRLY NKFLKKKTLKNETDYKNYKRLFESILKRSKKHYYSEQLI KHKNDSQKTWQIIKEVIGKKNLYRNLLPKNLKFNNKYI VNKSLVAETLNQFFVNIGPTLASKIATTQLNFKSYLTSN NINVMXNYKLTEKELLDAVSLLKPNESLGFDDISSNVIIK SIKYITIPLLHIFNLSLKQGVFPEKLKIAKVIPIFKSGDPSN VANYRPISILPCFSKILERIMYNRLYSFLNINNILYNKQFG FKSGHSTDHAIIHLVHDIFKAFDEKKFTLGVFIDLSKAFD TVDHKILL	cd01650	592	694

				AAAAAAA	ATAAACGA				
				GAAAACG	TTTAAATAA				
				AAAACMG	ATTTTCTAA				
				AAAWGCC	CATTTTAAA				
				GATCAAA	CTCAATTTT				
				ATAAAAA	ATTTGCWG				
				CAAACAA	ATGATACTA				
				GCAGTTA	ACTTATTTT				
				ATAAAGT	ATTCTAATA				
				ATGAAAA	AAGATATT				
				AATACAA	AATATTTTA				
				AAAGACA	TTTAAAACA				
				TTGAAGA	GTAAACAA				
				AAGTCTA	AGAACTTAT				
				ACGTTCT	AAAAACTA				
				ATCAAGA	ACCGAATG				
				TAGCCAW	GTTTAAATC				
				GATAAAA	AAATAAAT				
				AAGTTAG	TGTCCTTAA				
				CGAYTTA	ATATAAAT				
				GAAAAAA	AAAACAAA				
				AAAAGTC	ATATACTTT				
				ATCGAAT	ATTTCATCG				
				GCAAATA	TCTTCATAA				
				TAGGTGA	AAAAGAAA				
				GAATGAA	ATATTCCAT				
				AAAAATA	TAAAACTTC				
				AATTTAG	CAAATCTTT				
				ACARCTG	TTATTGATA				
				GAAGATC	ACTCATTAA				
				GACAAAG	TAAAAAGA				
				AAGAAAT	GAGCAATC				
				AATCTTC	AATAAAAT				
				GAATTGA	TTTTAGGCG				
				AGGAGTT	TKATTCTTG				
				AAAGAAA	ATGAAAAT				
				ATGATAA	WTAACATG				
				CGAAAGT	GAGGGAMC				
				TGGGATG	ACATAAGT				
				ATACGGA	ATAATTGA				

				AATAAAA ATAATAA ACATTTTT GAAAATA ATCTCAA TGTAAAT GGAGTAA TCATAGA AAGAGCG CATAGAA CTGGAAT TATTGAA AAAAAAA AACCAAG AACAATT GTAATTA AGTTACT AAATTAT AAAGACA AAGTAAA AATTYTA AAAAATG CCAACAA ACTAAAA GGWTCTG GAATTTA TATCGAC GAGGATT TTTCGCT AGAAACT ACAATAA TAAGAAA AAAACTT CTTGAAG AAAGCAA AATGCAT AGGAAAA ATGGTAA GTATTCT GTTGTTA	AAATAAAA TTTCAAAAA ATATWGGT ATATTGTAT AAAGCAAA ACAGTTTTT AAATCAAA CTTGTTTAA AAAACTTAT ATTTTTCTT TTATTCATT GYTATCTAA ATTATGCAA ACATTGCTT GGTGCAGC ACTAATGC AACAAAAA TAAAAAAA CTGTTTAGT AAACAAAA ACATGCTAT AAGGATWA TTACAAATG TAGATCGTT TCTCTCATA CTAAAACA CTATTTAAT AAACTTAAT ATTCTAAAT GTATATCAA CTAAACCTT TATCATATT CTTATTTTT ATGTTTAAA CTTGATAAT AAAATATC ACCAATTTT ATTTAATTC ATTTTTTAA AAAAATAA				

	ATCATATAT ACCCTACA AGATTTTCA AAAAACAA TTATATTCA ATCCAAAA CACATTATT CAGCAACT AAATTCTCA ATTGCAAAT CGAGGTCCT AAATTATG GAATACAA TATTAAATA ATGAGCTTA AAACTAATT CTTCTTTAA ACCAATTTA AAACAAAA CTTAARCAA AAACTTTTG ATAAATGA AAATGAAT TAAATTTCT TCTAAAAA GCTTTTTAA ATTAGCAA ACAAAAAT AAAAATTA ATCTATTAA TTACTCTTT AATATTAAT TTAATATTT TAAACAATT GTCTTATTG AAATCTATT TATTATCTT ATAAGTTTT ATTTACGAC GATTTTTTA	TATACGA TAARCTT ATTGTTA AAGAATT TATAAAA AAAAAAT AATGATA A			

					AGTTTTATTATTTTTTTTKAATTTCTTTTTATTATTTTCTTTTTTATTARTACTTACTTATTTTACTAATCAATTATTATTATAAATTTTTCGAATTACWAAATAGTTATATATTTTATTTTTTATTTTGTTAACGTTTTACTTATGAATGCATTTTGGGGCTTAGTGATAAGACGAATTTTGTCTTCTTCTAGCCCCAGTCATGTAAATATTTGTTTTTATAAATTACATTGTAAATTATTTTCAACGGCAAATAAATAATAAAAAAAAA				
Daphne	Daphne-16_HMa	.	Hydra vulgaris	GAAAGTCTTAACTTCCAAGAAACAACTATTATGGAAAAAATAAATAAGATCACGAAGCGTTAC	AAATTTATGTAAAATGTATTTTTCAAATCTTCAGCATATTGAGTATTAAAATGTTAGTATATTGTATATTTGCATTATG	MALISDFETLVFNFFETKFFSLHENSDPDINYFHDTKIECSYYYPNELKGILFKNVNNDQIRILHVNIRSINCNFEKFQNLLEETKYFFNIICLTETWATSNDIKKNFNFYLSHFEMISLERQINKRGGGVLIYIHKDIKHCCRNDLSVSDGDGEILTIEIINERTKNILVSCCYRPPDGVSENLSMFLQQNVIAKGNNEKKKNFLVGDFNMNCFVYNDDNKVKNFYDTIFESGAIPLINRPTRITKNSATLLDNIITTDLFNNDIKLGILKTDISDHFPIFLTINTDQLPKQKLNKIVKKRFYTSSNIEAFKNQLSLLHWKHLNFNENANNLYEKFYETFISVYESNFPIVTITLKTK	cd01650	569	828

GAAAAGG	TTATATTGT	HFDNPWITKGFKKSSKTKQRLYIKYLKTKTPDSEKIYKD	
AGATAAA	ATTAGTATA	YKNLFEKIRKKLKKNYYSKLINKAKNDSKRTWQILKEIT	
CAACTTA	TTGCATATT	GKQKTCSSSLPQMLKVDNNSLHEPQIIAHEFNKYFTEIGS	
AATAAGA	TGCATTATG	ILSSKIQKTQTSFYDFLLPFDKNISSEELSAELSFDEFEKA	
AGACACT	TTAAATCTT	FKSLKKNKAPGADEINGNIVIDCYEQLKNILFKIFRASIH	
AGATTTG	TGCATTATG	QGIFPERLKLARVTPLHKEGDRSNINNYRPISVLSIFSKIL	
GAGAATC	TCTTTTTTA	ERIIFNRVYNYFNYNNLLYNNQYGFRKGSSTEHAIIQFIH	
GATCTCG	AGGTTCCG	NISKSFEKSQYTLGVFIDLSKAFDTVDHHILIQKIKYYGL	
AAGAAAC	GTGATAAG	NNKILKWFESYLTNRKQLVYSNDGYQSEPLSITCGVPQG	
AACCTAA	ATCCTTCTG	SILGPLLFLIYVNDLNKASKLKSIMFADDTNLFLSHTDIY	
GATTAGA	ATCTTCTTT	ELFSTTNKELKHISNWFKANKLTLNINKTKWIIFHSCAK	
TGGAATA	CGGAAACC	KRFLTNNMPQIYIDETIIKRDTVIKFLGVYLDENITWRKH	
TTTGAAA	TAGTCTTAT	IDHISTKVSKNIGILYKARNYLNKKNLTQLYYSFIHNYIN	
AGCCAAA	ATTGTTAGT	YAIIAWGSTEKSKLQHLYRRQKHAIRVVNYADRFSRSK	
TGAAAAT	ACGCTGAA	YFFDYMNVLDVYKLNIFNVLCFTFMWKNDLSLFVFNDL	
TGGAATG	TTGTATATA	FSLKPINKYTLRSNSFLNEPFCKTKFNQFCIAYRAPHLW	
AGTGCGA	CATATATAT	NKIVLPNFDPSTTLPVFKTKLKNFILCMDNILKFY	
AAATGCA	ATATATATA		
GTAAAGG	TATATATAT		
AAATGTT	ATATATATA		
TAAAAAA	TATATATAT		
CAACTAA	ATATATATA		
AAATTAG	CTCATTTTT		
TAATGAA	TTTTTCTTT		
ATTATCA	TCTTGTAGA		
TAGAAAG	AGTCTTAAA		
AGCTCAT	TATTTTTAT		
AGAATAG	GTATCAGA		
GCCAACC	CTTATATTA		
TAAAGAA	TTATATATC		
GATAAGA	ACGACTTAT		
AACCAAG	ATTATTATA		
AACGATT	TATGTATCA		
GTCCTAA	GACTTATAT		
AGTTATT	TATTATATA		
AAATTTT	TCACGACTT		
CAAGACA	ATATTATTT		
AAACAAA	TGTATCACG		
AATACTT	ACTAACATT		
AACGCTA	GTAAACTA		

CR1	CR1-11_HM	.	Hydra vulgaris	CAAAAAA CTTGCGA GGAACTG GGATATA TGTTAAT GAGGATT TCGCTAA AGAAACG ATGGAAA GCCGAAG GATGTTG TGGGAAG AGGTCAA AAAGTTG CGACTTG AAGGTAA GTATGCA GTTATAA AGTATGA TAAAATT GTTTCTA GAGAATT TCGAAAG TAAATGC GCTCATT CTTAATA AGAACTC CTTTTTAA TTTATTTG TGATAAT TAATC	AAAAAAAA AAA	GGCTAAA GGATTAT CATTTAC TCAAATT AAAGAAT TACTGGA AATTCAT GAAAACA CAATATT	TTTTTTGTT TCTTTTTTT AACTTTATA TTTAACACT GCATTAATA AATTTTATA ATTTAACAA TTTATAAAC AGTTGTTTA	MDSNEPVNFESTLFNFFQTNDFLLDSDSDPDLNYFSEAG ALQNKCSYFYNHEIKEFLDRDNFNVIHINIRSLKKNFENF RNXIEETFNIFSIICLTETWCDSDDVNFNSNFHLPGFNLIS LARKTNKRGGGVJIYVKEXLXFFTRLDMSISDGDKEVLT IEILTKKTKNIIISCCYRPPSGVIESFNSFLRNDIIKKSNHEK KFNYLIGDLNLNCFEYHVNNNIKRFYNDIFENGAIPLINK PTRITQSTSSLIDNIITTDVFNESLKKGIIKSDISDHFPIFFSI NTNNEXVTKTHSLCLKTIKFNDKFLYDKKEIANELNKYF VSVGPNLAKKIPKAINPINNLFIPNTSQFNSFELSFEEFES	cd01650	418	677

				AAAAATMTTTARCGATAAATTTGATAAAATGGAAAACAAAATGTAAGTACTTAAAAGMGAAAATAAAATTTTAAAAAATGAAATAGCAGATATTAAAAGTGCAATGACTTTTCATAAATGAAAAGTATGAAAAAGAGACTAAAGAATTATACGATTTTAAAGAACATCTTAAACACAACATGAAAGTAAATGAAAAAGCTTTAAACAATGATGAATTTAAAGACGAACTTGCKGAACTTGAAGACAGAAGTCGAAGAAATAATTTAAGATTTAATGGAAT	ACACTTTTGTTAAAGGTATCTTAAAACTGTATAACAAAACCTTTAACAAAAAGTTTAACAATTGTGCTAAAGATATCTTAAAACTGTATAACGCATCTATGTATTTATAACGGATTTCTTTTTCTTTTTTTCTTTAAATTTGTTTATGTTTTCAATTCTTAATTCTTAATWATTTGTTATCTTCATGTATGTTAAGTATATATTCCTATTGAAAGTTTATTTATTTTGAGAATATTTTATTAATACGAAATATTAACACGAAAATTTTTTAGCGGTTCTCGACGACAAGACCTTACGGTCTTCTACGAGTWCCCGCGTTCTTTTATTTCTTAATATCATTTATATTA	AFKMLKPDKAVGPDDINGNIVINSYDIIKDILFKIFSCSIKQGIFPDQLKIAKVIPIFKGGDLTNVSNYRPISILSVFSKVLERISYNKIYNHLSVNNLLYSNQYGFKKNNSTEHAILQFTRSITDSFENSQFTLGVFIDLAKAFDTIDHKILFKKLKYYGITGNVLNWLKSYLNNRKQFVYVDVVSPVNLFDITCGVPQGSILGPLLFLIYVNDLHKASNLMTIMFADDTNLFLSHNNITTLFQNMNIELVKISDWFKLNKLSLNIDKTKWIFFHPNRKKHLLPSEMPLLFIDNILIKRVTVTNFLGIYIDENLSWKNHIENLCSKISKSIGILYKARSVLNKHTLIQLYYSFIHCHINYANIAWGSTNKSKLXPLYRQQKHVARLINFKDRFTHAKPLLFEMKVFNIYELNIFNVLCFMFKCKTNISPVSFHNLYXKKEKNKYILRNDNFIRQPVFQTNFGKFCISFRGAFLWNKIVLKNFKDFSQECNYRSFKQKLKKVIFSIDDISIYF			

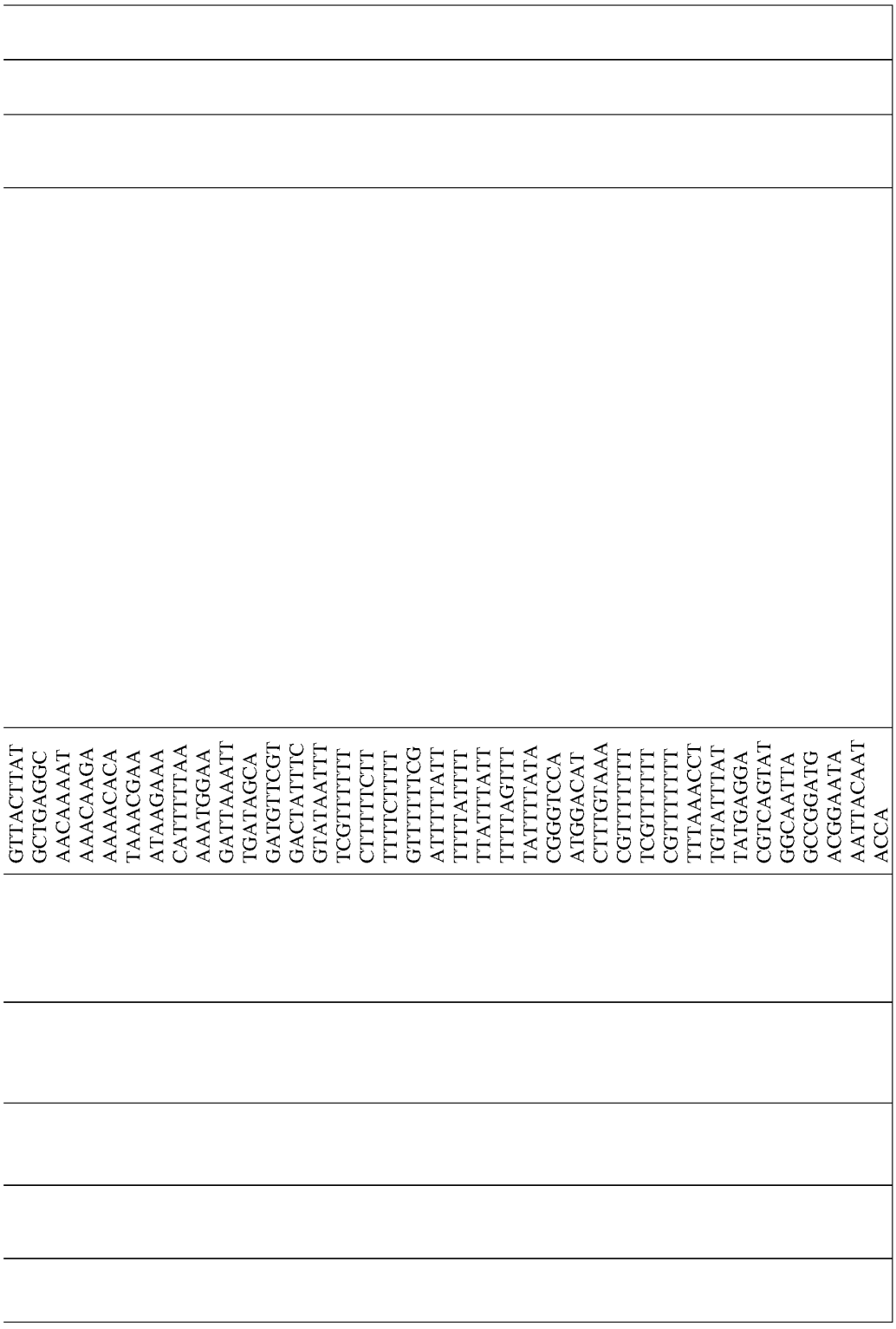
				TGAAGAA AAGGAAA ATGAAAC GTGGGAA GAAAGCG AAAAAAA AATTAAA GAANTCT TAAAAKA AAAACTT GGTATTA CGAACGA AATATAA ATTGAAA GRGCTCA TAGAACA GGAAAAA AGAATGA AGTTGAA ATGAAAA AACGAAG AACTATT ATTGTAA AATTTTT AAACTAT AAAGACA AAGAAAC CGTTTTA RAAAAAT ACCGATT ACTCAAA TTATGGA ACGAAAA ACTATAC ATAAATG AAGATTT TAGCGAA AGAACAA TGGAAAT ACGAAAA	TTATATCAT TGTTAACGA ATATTTATT AAATACAA AGTTGTAAT AAAGAACA AAAAAAA				

									1046
								791	
							cd01650		
CR1	CR1-72_HM	.	Hydra vulgaris	AATYTWT TTAAAGA AGCAAAA GATTTAA GAGCCAA AGGTAAG TATGCTA AAGTTGT ATATAAT AAATTAG TTACACG CGATTTTT AAAAGAA TTAATTTT TTTATTTT TGGATTT CGAATAA CTAAA	AGTCCCT GTGTAAT AAGCAAA ACTAATA ATAATAA CATTACA ATTTCAC TGACAGG AGGTAAT CCACTGT TATTTAG GGTGGTA GCCCTCC TATCCAG TAATCCA	CGTTTTTTA TTATTTAGT GACTGAAC TAACTTGTT CATTAATTA ATTTATACC GGCTGTCAT AGATATTGC CTGGCGCTT TATCTCAAC AGCATTAAT TAATAA	MVNLQLGDFNTYKCEQEAIIVDLKKLIETLRTDNANLTK RVLDLENSKSKSSSAVTGDWASKVKKTPDQMQIINTIA NETKERDKRENNLVIFGLEESTLSDKVKAIEEDKQNIETI MKKLKVNVKIKNVIKLKSKNGYKAPYIVVLNDKLERNS VLKNAKELKKDKTYESVFINPDLTESERYKSKLLREECK KLNNDNINKANFYFGIRNDKVVKIKKKQLNAKKDLNHIR CFNYETNNKINLFSGGGQNLNKLLSNKRQLDIEIVNFHE RKKCDLSINKEHDTQKCNELMIWYTNATSLNNKIDELR LQCEIYKPDVICVTETWFKAESDVNIISYNIFRSDRKTHG GGVCIYVKQCINQFETNNLQLNNNFIEQIWTCICFKNNK MLIGCIYRPPDISHEYNNYVLLSINAAKMSIYQKEYSDLII TGDFNYNTIDWSDNNCPFSATDNESQTNKFIECLEDCFL FQCVCEPTFQIANYKSTNILDLIITNDKNKIVSITHSAPLG NVKQGHHLLLKCCYLMNDYKFSTEHKKSIRLYQKGNYE LFSCHFNNVDWENGFKSKSINEMYDLFLYHYNYYATNLF IPRINYQNRRNRNKWITPELKHKIRTKNNIWLANKRNG WNEESTKIQYNKLKIRIKNDIKNNIKQFESTLASKFKSNP KLLYKYINEKKSVKTHIRALEIDNGEIINDQFAIATELNK YFHSVYVNDNCTNIIEFHNIPSNLLLNTVLITQNDVQTRL QALDKSKSVGFDFVHPYVLKECSSSISYPLYLIFKKSMET GTLPDMWKKANVMPLFKKGSKLKASNYRPVSLTSIPCK VLERIIADCIMQHLIKNNLLSKKQHGFMKSKSCTTNLLE			
									1046
								791	
							cd01650		

						YLDILTDAFHNRTPVDVLYTDFKKAFDSVSHKKLLSKLL			
						TFGISGKLLTWIVCFLANRKQRVVLGKSVTDWVDVLSG			
						VPQGSVLGPLLFLVFINDLPENFINECRLYADDNKIIAPIS			
						SQNDSQLFQNDINKLDEWSTKWKLGLNFEKCKIMHFGS			
						NNNEYSYFMNNNNTLMEIEKSKLEKDIGVYISSDLKWA			
						KQVKYAASKANSILSLLNNTFKYKDKDLIKTLYCTYVR			
						PNLEFSIQAWSPYYTKDINELEKVQRRATKMIPELRHLE			
						YKKRLEILGLTTLETRRLRCDLIQQFKIFKGIEIIDIKQNQ			
						TMNSICSSGPAANIRGAKHRMKPELVRNCLKRQYFFSN			
						RVVDGWNSLPAEVVQSTTVKSFKLNLQSVDLEAITNKI			
						KTKKAYY			
						RWQPKFCIKYYNFNFLPKLKIFIQLTLIKMVRKEDFEAL			
						VSRVLKLEAELLTKDKEITQLKILVNDLKNQPSNVLNSS			
						ACAWSDVVKGKKSNKAAQLDLVNVVVKESQDRDLRQ			
						KNLIVFGFNTSKNKNDDVAREEDILGIQNIFNQINAKATI			
						QRVFKLSIKSDKPPPVIVVLTNKEERNAVLKTAKTLRNL			
						PETNNIFNADLTVAERLKMKALREERAKLNETNKDLQ			
						YYYGIRNERVVKLIKRQFISSQQDQLLFSKNLLFKNSNK			
						FIEKNNKRQLINLKNQELFNHPKNNYCGKWALTNIRSL			
						GNKIADFQHFVHLHSLDICCVIETWLNKNLPDSLLCPRF			
						YTVLRCDRSSKGGGIAIFFRSFIQFTKVDIPPEFHELEIVC			
						ADLNLNGQSFRVIGYYRSGGFGKVAIDYMQSSVKCFSIL			
					TACAATTTT	CSTKKNVVILGDFNLPDIDWHYYHGPDSVIHNSFLHFIN			
					GTTGTAACT	RYSLTQHVDQPTRENSILDLILSSSSSFISNLCIEPPIGTSD			
					TGTCCAGCA	HNVIIFTPNLSCLYNKTEVPVQQFFCWKNADYQVINKYL			
					CTTGATACA	KSVDWSIIFQTCFNIEACWHTFSTILNNIITQHXPKVRVC			
					ATTTAACAT	EEHKLSHKIHYPYYVKRLINNKAMAWKRSKVTGKSND			
					TTCAGTAAA	KAAYKSIAAQCSDSIRKYHAAKELKMIRKDNLGSFFNF			
					AAAAAAAA	VNGKLKTCSKLNDIRRSDNTLCHDKEEISCIFNNFFASVF			
					AATGTTTTT	TVDNGLSPKFTDRVNVHSTCLSSINFSPSTVEKALRALK			
					GTTAATGTA	PTTSTGLDGIPNVFLKNCANSLSIPLSHLFDTSFKDGAIPL			
					TACTCAGTG	DWKVASVIPIHKKGPTVDPGNYRPISLTSTCCRVMERIIN			
					TTAGTGGCC	NKLINYLLLNNLITKEQHGFIRKRSTCTNILESLHDWTLN			
					TTCGGGTCC	LQRRIPTDIIYFDFKKAFDSVSHPKLLMKLPAYGIVGNLL			
					TTCTGTTGT	NWLTNFLHMRSQVVNLNNVTSHAISVTSGVPQGSVLGP			
					TTTCAAATA	TLFLLFINDVSHINKQLDVKLKLFADDIKLYSVYDVRGL			
					AATAAATA	QSDLHTAVNXLYEWSCTWQLQIATEKSFVCTIANQSQN			
	CR1-55_HM	.	Hydra vulgaris	TTAA	AATAAATA	IAQRVYTINNHELAHVNCIRDLGVTIDRNLKFNQHINLIV			
CR1					AATAAATA	HKAMSRAFLILKSFYSRDRLLMVKAYCTYVRPMLEYCS	cd01650	787	1043
					TAT	SVWSPYTICLINKVEKVQRFFTKRIAGLWSVSYDKRLAI			

					LKLNTLEYRRVFKDLVLCYKILNGKLDTDLSNILILNSY LNNRGHNFKLYKHHCSLNSTKCFFSNRIVNVWNSLSED VVSASSVLMFKKRLAVLLCSS			
L2	L2-1_Hydra_vulgaris_Ma	.	AGGAGCA	TAAAGGAA				
			TGTTTAT	TGTCAAAGT				
			GAAAATG	TATGATTTT				
			TTAAGGA	AAAAAGTC				
			TTGGTGG	TTCATTTTA				
			GAGGCTG	TTTAGCCAA				
			GTAAGGA	AAGGAAAA				
			GAGAGTA	CAAAATAC				
			AAAATAA	GAATATAA				
			TCTCGAA	AAATCTAA				
			AAAATTT	CTGTCTTAA				
			AGCAAAA	CTCTTTACA				
			GATTAAA	AATTATTAA				
			ACAAGAA	ACAAACCA				
			AAGAACA	ACAAAAGT				
			AAAAAGA	TTAGACTAT				
			ATTTAAA	AAATTTTAT				
			TTAATCA	AAATGGAA				
			AATTATT	AGATTGTTT				
			AAAAAAC	TGAAAGCA				
			GCATCGA	TGAAAAAT				
			AAAAAAT	ATAAATGA				
			GCATTTA	AAAAGATT	MLKKEISESEIKQALGQMKNNKSPGSDGLTIEFYKHFFX			
			AATTGCA	GGAAACGT	LIKKDLIESLNNVLFTGEMTNSQKEAJITCIFKKGNKNDI			
			ATATGAA	TTCTTTTATA	GNWRPISLTNVDYKILTKVLANRLKDVLPNIIHSNQTACI			
			AAAACTT	AATCTTCAA	KNRTIYYNLSYTRDIISIAKKSKLDACILSIDQVKAFDRV			
			TACGACG	AAAAGTAA	DRAFLFDSLKHFGFGKIFISFIKTIYNNINAKVKINGLLSE			
			ATTTAAA	TACTAACA	KIDILRGVRQGCPLSMILYIIQAEIFSCYVRQNKNIKGIHV			
			AGTTAAA	ACAAAGCG	KESETKJQQYADDTNFFLIGDDSIREVGKALKLNQKATG			
			CTTGAAA	TGCGAAATT	AKINIAKCQGIWLGKNIFKDKSEYLNFNWDXNTFKSLGI			
			ATATAGA	CTATTTAAA	FFSNNANFQYETHWKERFNKMENSMRKWEKYKLSIKG			
			AAAAGAA	ACAGCACA	RKTIINSLIIPQALHLAFIAPLPSNGILNKIEAKINNFYWGG			
			AAAGCCG	CTTTTTGTT	KICRLNKTKTQQKINEGGAGVNSLTNKLKAIQLQWISKL			
			TAGGCGC	GCCAACCA	YNESICGPWKDAMVYILNLYRNANQSKHTFETIPTQMK			
			GGCAATA	ACGATAAA	LLPPFYSQLLVNWHSLKQKHVLPITKIEEILNQPLFFNSM			
			AGAGCAA	ATGTTTGAT	IKTMPNKFLTPSETHIKNDIILVADITKTFQPGFLHHKSIG	cd01650	64	321
			AAATTCA	TTTAAAATT	LTEKELLEFLTKALPRKWYKKF			

			AGAAAAGA	GTGGAGG		
			CAAACTTTG	TTAGAAG		
			CCCACAGT	GGGAAAA		
			GCAAACTT	ATGCTCC		
			AAAAATGA	CATTTTTT		
			AAATACTG	TTTCGGA		
			CCCACATG	TTGGAAA		
			ATGTATGA	AACAAAA		
			ATGTGAAA	ATTAGCT		
			AAGTACAG	AAACAAT		
			CCACTAATA	ATGTCTC		
			TTTTATGTT	AGAAATA		
			CTAGATATT	AACGGCA		
			TTAGACCA	GAGACGG		
			GCTCTACCC	GCAGGTT		
			TACTGAATA	AAAATCG		
			CCCTCATAA	ACTCAAA		
			TAACAATTT	AAAATTT		
			TAAATTTAT	TATTAGA		
			TTTATTTGG	ATTTGAA		
			ATACGAAA	CATTTTTA		
			AAAACGCA	TAGAAAC		
			GCCCAGAG	TTATACA		
			GCTTTTTGG	AATTCAA		
			CAGTTTATT	AGAAAAC		
			ATTAGAAA	GACAAAC		
			GTTTTATTT	CTTCTCA		
			GTGAAATA	AGAATTT		
			TATTTTAAC	TTGTTCTT		
			AGAATGAA	AAATGCT		
			AAGTCAATT	AAAGTTA		
			CGAAAAGA	AWAAAGT		
			AAACATTC	TACTAAG		
			ACTAAGAA	GAAGAAA		
			ATATTTGCT	ATAAA		
			AGAAAAGT			
			TTGATAGA			
			AAAGTCAA			
			AAGATTATT			
			TTTAAATGA			


GTTACTTAT
GCTGAGGC
AACAAAAT
AAACAAGA
AAAACACA
TAAACGAA
ATAAGAAA
CATTTTTAA
AAATGGAA
GATTAAATT
TGATAGCA
GATGTTCGT
GACTATTTC
GTATAATTT
TCGTTTTTT
CTTTTTCTT
TTTTCTTTT
GTTTTTTCG
ATTTTTATT
TTTTATTTT
TTATTTATT
TTTTAGTTT
TATTTTATA
CGGGTCCA
ATGGACAT
CTTTGTAAA
CGTTTTTTT
TCGTTTTTT
CGTTTTTTT
TTTAAACCT
TGTATTTAT
TATGAGGA
CGTCAGTAT
GGCAATTA
GCCGGATG
ACGGAATA
AATTACAAT
ACCA

CR1	CR1-6_HM	.	Hydra vulgaris	GGGAATT GGAAGAT TAACAAT TTAAGAG TAGATGG AATTGAT GAAAATG AAGGTGA AAGCTGG GAAGAAA GCGAA	AGTAAACT AMAATAAA AAAAAATA AAAAAAAA CAGATTCA ATTATATCA ATGATATAT GATAATAT GAATAAGA AAAGAAAG AAAAAGGA AAAAAAAA AAAAAAAA AAAAAAAA AAAGAATA AGAAATCC AAGTGATG ATGTCACA AGTGATGA TGTCATAAAG TGATGATGT TACATATTC ATTTAGTGT TAATTATGA GATTTTATT TTTTCAATT TCTTTTTATA TTCAATTTT TTTTTTTTTT TTTTTTTAT ATTAACTTA TTTCTTTAAA CATTTTAGC AACTGTTAA GCATCTTAA CGCTTCTTA ATTAATTCT TATTTTTTA TTTTTTTTTA TTATATTAG	MKVKKIFEDILSVKNVQIERAHRTGNKKLKMPRTIVIKL LDFRDKVAILSKSSKLKGKNIYINEDFCFETTQIRKNLRE KMKIERAAGKFAFISYDKLIIRDWVAKKSKNLSSTVLK MEENVIFKSSNENDIIGDNDTDSNYFNHKMLESQYYSVS ESSQYLGQTKNTFSILNINIRSLNKNFENLKILLDEIKHDF KIICLTETWCKCGETNYEFELVNYTSIHQPREVNAGGGV SIFIHNSVNYILRNDLCVNETDYESLCVELVNNKEKNIIIN STIFRRQINENSLKQFKNCLKNCVDWNLILQSNXANNA YDLFLXQFCEQYEIAFPEIKKVINTKSLXNPWMSKGLLK LINYASNNNAKNFINTLLQYNFIPTINKSTRITNRSSTLLD NITTNNFHNNPLYTGIIQTDLSDHFPTFLATNNILJNNISTK SSKKKEKLYXKYLKKKTYKNEITYKNYKNIFEKTKKRS KKLFYDQLLXKTNGNTKKTWNLIKDIIGKNKRKKNNLP KNLLIDGELVYDKSIIAEELNNXFLNIGPNLAAKIPTNSX PFDSYLKTYDKLMDETNLKPSELRTAFSSLKSKKSEGFD KISVDVVKSVFDIIEPSLFHVFDLSLKSGIVPDKLKIARITP IFKSGDKSNILNYRPISVLPCFSKLLLERIMYNRLYNYLTE NNMLYCKQFQFKKKQFNLRTICCIVNNLVLKKNNSTXH AVVELFNQISNAFNNDCFTLGVFVDLSKAFDTVNHDILI KKLEXYGVKNKNLLWFEDYPTNRKQFLYYNQNNTIKH SITCGVPQGSILGPLLFILYINDLYLASNLLBLILFADDSN LFYSHRDIKALFKIVNEELCKVNEWFISNKLSLNVEKTK FMLFHKPNKSENIPLKLPDLIINKTYIKRESNINFLGVILD ENLSWNSHINSIEKKISKNIXMMYKAKPFLNIXSLKKLY FSFIHSYLSYCNIAWASTNYTKLKKLYSKQKHACRIVFG ANRNVPCEPLLSELGALNVYKLNIHQVLLFMFKTKNEL SPKIFQSYFNEIXHKYPTKFSDNNFVVPKYNLKLTSYSIQ YRGPYLWKKFPXIVNKKKTLRDISLEQFKNESKRLLLFM DLNKFDFKCFF	cd01650	670	948

CR1	CR1-19_HM	.	Hydra vulgaris	ATTTTTCT TTTTACTT TCTTTCTT ATTCTTTA TTTTTTTT TTTTTTTT TTTTTCTT TTAAATT AGGAATT AGTTTTC GAGCGTA TTTTGTTT GACGTGT GTAAGTAA CTTTGAG CTTAAGC	AATAAA ATAATAAT CAACTATA TTACACTT TCCCCTGCT TGGGGTGTT	TTATTTTCT ATTTCATTT CTAATTGTT TCTTTTATA TTTATATTT GTATGTTAC GGTTATGTA GAATTTTTA ATGGGGCT CGGCGATA AGGCGAAA TGTGCCTTC TTCTTGCTC CAGCCATC AATTTATAT AAATGTAT ATGCAATGT AAACATTTT TTAACGGC GAAATAAA TTACACTAA AAAAAAAA AAA	NRSVKTEIFKAKKAHEKEIISKSQNNPKIVYKYINKQLKI DPWVTPHIKNLIREKKSAWHINCSTKWENSELKIKYNRL WEFLFLHKTADQMFNILISNYRTACTYFIPKINKTKKKY FPTFQTDYGSSNNMLDLLFTYNSNSILNLSPSTILGNTTK FNFPNIKWSNDGIESFNPNNPTEYDFWNIFQNSLFTQHIT LIGCIYRPNDASLIKYVAESFKLAHSLIKKPYDDCLIVGD KYKSKNLHNKKTNYSVNCYYTNATSLDNKMADLISISR YQLRMKRNELNSNINLKTSPFRYGIRSSGVVKLKNIQTT SDVSERNPLLTAAKKLKNSDNYKNIFISPDLTEAERYLN HIDDKETVQNILVTISSDAKPKSVKRIKSTSSKPGPIIVEL PIEQLKIINTFAIEQQQKKKREKNVIIYGIKESAKLSVADK QQQLIIDKLIQNPNELTNTTTSSIESSWVTVASKNIKKKQ MTNLTPAQLKQVEDAIKKLIHAETAKWSAKISELEQVIK	791 cd01650	1046

				AACTTTTT CTAACCA CTACTTCT TTTTCTAC TTATTGCT TTATCTA ATTAGCT TATTAAT TACTTTTC TACTTAA AATTATA GTAATAT ATTAACT ATTCGTG TCTATAA TACAGTT TTTTGTTG TACTTCG CATACCT AAATTAA ATATAAT TCTTATAT TTTATAC CTATACT TATATTTT ATTATAT ATATCTTT ATTTATA TATTACT ATTATAA AAATAGT TATAACT GAGATTT TTTTTTAT GCTTCAG TTCGCTT AATTTTTT TTTTTTTT TTTTTTAC AGAGTTA		KDRIRSLKDTSGRIIEALPDIVDLLNSQFKSMFTLESSENF PLLNKSGILSECIYNENLDFSISDIIQELADLKPYKSFGSD GLHPLILKICAKSFAIPLTKIFHKSFNSGVVPTQWKLANIS PIYKKKGDKLLAENYRPISLTSVACKIFEKLVKKKILKHI NDNKMLSANQHGFLYHKSCVTNLIETFDIISSSIASHQFI DVIYLDFAKAFDSVPHNRLLLKLRSFGINGKLLQWIKSF LVGRKQRVTINNTSSAWVDVLSGVPQGSVLGPLLFILYT NDLAKLTETSEIFADDTKLFSKTLSQSNFVDLQDRLRLV YEWSNTWLLKLNTSKCCVIHYGRNNPHYNYFLQNNQG INQSLRSSTKERDLGVIFTSNLKWSAHIQTVVSKSLQILG MLKKTFTFIDEKMALKLYKVFIRPHLEYGVSIWAPFLKA DMALVESVQRKATKWAYSLRHLQYCERLKRLNLPSLE FRRLRCDLIQIFKLIHRFDIVNDKYLPKYINSVSRGHHSKI RREYVLNCLPRYNHLYNRSAKYWNNLPEQITLYSTVLS FKNRLDSWLWKKFGTNIYKTGCL			


TAAAAAG
TTTTTTTT
TTTTTTTT
TTTTTTTT
TTTTCTTT
TTCTTCGC
CAAAAGT
TGCGTTT
GCACCTG
TTCTTTGT
TTTTAGTT
TTCTGAA
CAATAAG
CTAAATT
TTTGCGC
TCTGACA
CAAACTA
TATTAGT
AAATTAT
ATATCCA
TATATTA
TATATAA
AATACAT
ATAAAAA
AGTTATA
TATAAAC
ATATACA
AATAATA
TAAAGAA
ATTAAAT
ATATACA
AATAAAT
ATGATAA
CACGGAA
AAAAACA
CATACTG
AATCTTTT
TAGTCCC
TGTGTAA
TAAGTTA

									1046
								789	
							cd01650		
CR1	CR1-69_HM	.	Hydra vulgaris	TACAAGT TACAATT TGTGTTT GTGTTTA CAAGTTC GAGTAAA ATAA	TTTGTCAT AATTGTA ACAAATA AATTTGC AATAATT GTAACAA ATAAATA AACATAA AAACATC ACTGATA TAAAAAA AAAAAAA ATTCATC TGTG	CTGGTTTTT GTTTTATAA CTATGGTCT TATATATAA TTGTAATAT ATATTTTAA ATTAACTAG CATTTGTAT TATGTGATT GTAGGGCA CGTTGTCAG AGTCCTTCA TATTTTATG GACCTGCGT GTCCTTTAG AACATGTAT CACTGTTCT AATAAATTT TAATCTAAT CTAATCTAA TCTAATCTA A	NLIKILIANMVNVQSTEFNVYKADQDKLIKSLIKTIDELK NCNLLLTERLNVLEKAKVVPSTSISWASVVSGNMPPKR STEQLHLMNAVVAEAKEREKRESNVVVFGIEASKKADL LSAREEEKNSILHMMNSINAKIEIKQIIKLKSKSDKESPFV IVLNNRNDRNSFLKAAKSLKNSTEFNKVFINPDLTVAER YKAKLLREECKQKNLENSEPSLFYYGIRNEKVIKISKQSF YNCKNQKYFSHGSLTSLVIPKSLQKNHASSISKQRITAV NNKSIKKVICSKRSNFINSYNHLTTNSFNNFTVYLFNARS LANKLNDFFLYVESNKPALICVCETFFNDKLPNSMVCPK NYSIYRKDRVRIGGGVAIYCRNDVKSAIVSITQTDKDIDI ICVDLIFGSSKLRLITCYRPPSYSLDDYIYLESMISIISNLC ESVPQFVIVGDFNLPKIDWVNYVSPIEKCHNLFLTLVNN LGMYQYVLEPTRENNILDLVLSNIISLLSNISVECPFSTSD HNSVFFYINTFNYYQPETNSESFYDFKNTDIIAFNSYLSKI SWDFDFSFVFTTEEYWNIFSNHLRKGIDRFVPIRTIYHNK KSRSYPKYIYKMLIRKSFLWKRWSITKTTLHKKAFIQYA SKCKKAISTHHKNVELKLVKENNLSKFFSYVNKNLNNP KKIYPLKTNSNFTTTNVEIAEVFNKHFGSVFTIDDGILPKI NTFVSEHIKMNFVDFSAKIVYQKLKNLKPSTSFGPDGIP NILLKQLAHIICIPLSYIFESSFTSSSLPKQWLQATVSPIFK KGSTSDANNYRPISLTCTSCRVMESIVSSSIADYLNINNLI TPNQHGFLSKRSTCTNLLESTNDWNKALDSNLITDVVYI DFQKAFDSVPHPKLLKKLAMYGIIDNLLHWISAFLSNRY QRVLVGNSLSNPTCILSGVPQGSVLGPTLFLLFINDLPSIV KDLNCSLMLYADDIKLYSSFKDADYSHDLAVAIYKVYL WSKKWQLKIANNKCFTHRIAPTSFCKDINYNYQLGDHN	789	1046

						LTWSTNPKDLGVTMDSYLNYSNHISNIVRASNIRGYLIL KCFKSRDPRVIVKAYTTYIRPILEYCSPVWSPNRTKMNN TVERVQRRFTKRISNLNNVSYSSRLSILGLELLEIRRLKQ DMITCYKILNNLVCLNKSSFFLFNNYIYTRGHNFKLRKL KCKLDVRKHSFSLRVVNIWNNLPSDVVNAKNIKSFKIKI NSIDFEKYCSG			
CR1	CR1-81_HM	.	Hydra vulgaris	TTTTTTGG ACTTTTAT AGCGATC GGACATG TTATAAG AGCTTTT CATTAAA AAAAAAA TTTACGG AAAAATT CAAATAA AGTTAAC TATACTTT GTAACTT TTTGTGA AAAAAGT ATATTTG CATAGCG TACCAAC T	CTTTTCTTA CCGAGTTGT TGACATGTG GAACAATA TGACCTCGG AGGCTGTA AATGCAAA TAACATTTT AATGTTCAA AAATAAAA TTAAATCTT TCGACTTTA ACAAATAT CTATTGTAA AAGGAACA ACTATTTGT TTCTTTATA GTTTTGTTA TTTTATTTT TTTTTTGTA GCGATATTT TTAAATTTC TTTGGGCAC GTTGTTAGT GTCCATCTT TGTATGGGC CTTCGTGTC CTTTAGAAC ATGTATTTA TACTTTGTT CTAATAAAT ATCAATCA ATCAATCAT CAT	MVNVQSTEFNVFKAEQEKLIKALNRTIEELKNLNLLLSE RVKMLENAESVKSSPLLSWASVVSGKTPVKKTVEQLNL INVVTAEKKEREKRENNVIIFGINPSKAIDKSSAIKEDKN SIQEVLNSINLKVEVKQIIKLKSKIDNQPPFVIVLNNKTDR NSILKAAKLLKNSSEFKNVYINPDLTEAERYKAKLLRDE CKKNNLENSESLLYYYGIRNDRVVKIAKQLLFNNRKLS HMSHCATELANSSRISCEYLISKSSKPDKIICTKKCRNKPI NLNLINGNSNNGCNSNNGCSSNPYSKNSVYLLNARSLA NKLHEFFMFTEANKPAIICVCETFFTSKISDSIVCPKDYSI YRKDRIKIGGGVAIYCRNDIKSEQVQITQTDTDIDIVCID LIFSSHNLRLITCYRPPFYSIADVAYIELMTAFIYNLCHSV SQFVIVGDFNFPKMDWLNYVVSPNEKCHILFLNLVNNLG LHQFVQEPTRENNILDLVFSNNISFLSLISVECPFSTSDHN TVHFSINDLDQSNKNIESFYDFVNADIDGFQSYLLNIN WDFEFSFLFLQLKIIGTFSPIIFAMVFINLFQLERVILIKKL GLRVFLFVFTVEDYWNVFSNHFCNGIHQFVPIRKSNLNK KIKHYPSFIGKMLNRKAFLWKRWSNTKNITHKKAYNK YASKCKKSLLTYQMRVELKLVKENNISKFFNYVNRKLN NSKSIFPLKDKNNNNKITCSNEEIADVFNKHFGSVFTND NNYSPFINSYVDDSIGIEDCTISTRSSVLVVTKKSLTCLTN TLAVSLQMITITLHLLIVMLMIQLVLKTVLFPPEAVYKSL KEIKASTSYGPDGIPNILLKKLAHLLCIPLSFIFDASFKSSS LPKQWRQAFVSPVFKKGATADPNNYRPISLTCTCCRVM ERMINVYVIKYLGHHNLISLNQHGFLNKRSTSTNLLESA NDWNTALDNHQITDVVYIDFQKAFDSVSHTKLLVKLAS YNIKGNLLNWISAFLTDRYQQVKVGNSISNSIRIVSGVPQ GSVLGPTLFLLYVNDLTAIXNNLDCCVKLYADDIKLYSS YCDMDLSQDLVVALNRLLQWAETWQLTIACSKCFALRI APPSLCKGIKVNYKLGEYSLAWSSNPKDLGVTMDCNLN YKKHIFNVIHQANSRAYLILKCFKTRDPCILVRAFTTYIR PILEYCSPVWSPHQIGLIKSVESVQQKFTKKIRSLRNLSY HNRLQLLGIESLEVRRLKNDLVTCFKIFRKLVCTDQSMF FAIDNNSHTRGHNYKISKQSCRLDIKKI	cd01650	881	113 8

Dap hne	Daph ne- 78_H Ma	.	Hydra vulgaris	TGGTAAA GATGGCG GAGTTAT AGATAGA ATTGTAA ATTTATC AAATAAA GTAATAT ATCTTTA CATGACA GCAATTA AGAGGAA AAATATT TATAGAA AAAAAGA CTAAAAA GTAAAAA AAAATAT ATCGGAA AATTAAT TAGAAAA CCAAAGT ACTTAAA GTTTAGA GATGTGT AAGAACT TCATAGA CTCAAAT AAATATA ATTAAAG AACAAGA TACATTA CAAAATT AATAAAC TAAGAGA AAGTAAA GAAAGAA TCAAGTT AATGAAA GRAAAAG	AGACTCTG AAATATGTT GTTATGCTT TTATTATTA CTACTYTCT TGATACCTA CTAGTTTTA TTATTAACT TTTTGTTCT CATATTTTT ATTGTTATT ATTAGTATT GATATTAAT AATATTCTT CTATCTTTA TTAGTAAAT TGTTATTTT TATTTATAA CATAATAA GTATTATTA TTGCTTTTT TAATTTTTT CTTAAGTTT ATTTACTAC TTATACTAT GTAATTGTT ATTACCACT ATTATTGCT ATTATCTAA TATTGATTG TATAATATC TTTATTATT ATTGTTAGT ATTATTACT AATAATATT CCTTTTCAT CGTCATCTT ATTATTATT GTTAATATT ATTTCTATA	MLIRNLCKFCNKTVVDSHKGILCNYCSSWLHPKCNYIN KHTYSLIEKDSTDWYCSDCINKNLPFCSLSDFELSLTLSS QTIPTNNLKSLKSPPHFKKLFNNINKISNKSINCKYYDILE LNKVLNNNCNLYLHLNIASLPYHIDGLRSLISSLTTLPIVI GISESRLNINDSSKTDVNINGFTIEHSPTEAKKGGALLYL RSNLNYFIRSDLIVYSPKYLESIFVEIVNPLKQNIIVGSIYR HPSMNSNEFITDHFNPLLERLSLENKQVVLMGDFNMDL LNYNESKVISSYLDTLCSYSFFPSIIQPTRITATSKTLIDNI FLNFQTPDLISGNLTISISDHMAQFICIPCLSPKKTKEVIFR RTFQNFDNKKFIKDISNADWQIQITNNDNVNESLHYFFK TFNKILDQHAPYKKLTAKQIKLKSKPWITNGILKSISIKN KLYKQYIKAKNHNIKIHIFNKFKWYRNNICNLLKLSKKS YYINFFNNNLXNVKNTWKGIKEIINIKPSSFSSKSINLKFN GNTISDSKAVSNIFNNYFVNTQQTLLTKIVSPKCNFSDYL NNPNINSFFIDPVTENEVSNLIKSLQSKKSLGINSIPTFLLK LVSHIISKPLCTMINSSFENGLFPDVFKIAKVIPLYKKGSR TDHTNYRPISLLSNISKIFEKAMHNRLYNFLDKYQCLYK HQYGFRSKHSTTHALIEITEVIRKALDDKLFACGVFVDL QKAFDTVDHSILLSKLEYYGIRGISLQWFTSYLYNRFQF VSINGNTSSLVKCTTGVPQGSVLGPLLFLVFINDLNLSLK YTTSYHFADDTNLLLINKSLKKINKYINHDLSNLIQWLR SNKLSLNTSKTEIIIFKSKQSKISKHLNFRLSGQKINPVMY LKYLGIKIDSNLTFSQHLKDLSIKLCRSNGMLAKIRHYV NLETLINIYHAIFGSHLRYACQVWGQIRTQEILRLSNLQN KALKIIYFQHYQSNSDHLYFLSKILKIDDLVQLLNCHFVF NYKQKCLPSTFTNFFTVRESCRYHLRSANSYNLHVPNHS SVKYGLKSIKYQSVKSWNSLPSPLKSIDSITNFKNCLFNY FLARYYT	cd016 50	635	891

				ATCAAGA GTTTTTTA CATTTGTT TTGATTA TTTTTTGC TTTAAAT ATTATTT GTAAATT TAAGTAT TTTCCAG GATTCGT TTTTATTA TTGACTTT TTGATTA TTATTATT AGCAGTT ATCATTG TTATTATT ATTATAA TTAAATA TTTATCAT TATTATT AAATAAT ATAGCAT TATTTATC CTATATTT ATCCATT ATTTACC TCTTATTA CAATTTA CATCATA ATTTATCT TATATAA ATGTTAT TTACCTA ATATTAC TACTTAT CTCCTTCT TTACTTAT TATTACTT	ATTATTATC AATACTATT ATTATTGTT ATTATAATC ACTTTTATT GTAATGTTG CAGGTTTTT ACTTTAGAT TAGTACTTG TACTATTTG TAAAAGTC CTTGAAATT TTACAACTT ATTTCAGAA TATATATTG AATTGAATT GAATATAT ATA				

ACCTTAT
TATCATC
AAATTAT
TTAAATA
ACATATA
TTATTTAT
ATTATAC
TATGTCT
AAAGCTA
CTAGAGG
TATTAAT
ATATCTA
AATTTAT
ATTCTAA
ATCAAGT
TAATAAT
ATTCTAA
TTTTATCT
AGCTGTA
ATACTTA
TTCTCCTG
TGTAAAG
TATTTAA
ATCTCTT
ATATCTT
ACTATCA
TTTCTTTA
TAACTGA
TCATTCCT
CAATATT
TAAAATA
GCAACCA
TTTCTAG
ATTAACA
TTTCTTAG
TTTTAATC
TCTCAGT
TTCTTTAT
AAATATA
TAGTAGA

			TATCTTTA TTTAATT ATTTTTTC TCTATAC TTTGTTTA AGCATAA TAAGTC		TAAAGCCC GAACAAAG GAGAAGCA GTGGACAG CGGCACTCT CCAACATA GCCGACTG GCCTAGACT CGAAGCTG TTGCAGAAT TTAGACTGC GTACCGGA CACGATTGC CTGGCGAA ACATCTCCA CAGAATAG GAGTATAC GCCCAACC AACATGCC CCCTATGTG ACCTTCAGG AGGAAATG GAGAAGAC CCATCTGAT ACGGTGTCC AGCCCTAC AGACAGCG ACAGAAAC CCAGAGAT ACTGGGAA GCCAGAAG TCAACTAA AGGGCTAA	MTSLKRNSNRRNRPCQGEPQLLQILQWNAGGMSQDKKI QIQKILDTYNIDIFTIMESNLTDDKLTYYQFPGYTLHLLP KDRQVASGILTGVKEGLTSHCEIIKSMGSMQDICEIIRLN VWKSQNHFKMYAVYNPPQNRPNFDLLNISHKTIVLGDF NAHSTRWGYKNRNTAGKEIEDILNSSPLELIYSDEDPAT YLHYNGTRTTPDLLLVSSDISELTQRKIIDDPGSGHKPVI ASITINSKSMTPKMPTKVLWKFKKADWPKFTNLLETEL NASPINYNQHPDKLCXNTNIIIKCAKKTIPRGKVKHYRV FWSKNLEELKRKREVLRNTAEQTGKTEDVQAWRRQSA VLRQAILQAKRTTFNSFISNINYQNDSQRTFKFLGNLLN NKKKPKKEPIRLNNKLLTTDSEIANSFARFYSXKQKKNP FVRNSSRDIRKQVTKFDEIKEDTAIPHKVLTEPFSSHELN AAIKQLKCKKSPGEDCIHPEFLIRMGPKAKEILLTLFNKI WETSLVPTQWKVAIVIPILKKGKDPSNFDNYRPISLTSM LAKLMERMVNTRLTWFLETNNILGNEQAGFRPQRSTNQ QVATLSQHIKDALDARNTLTAVFIDFKSAYDLVWKENL ILKLAKIGIRHNMLNWIKAFIGQRTCKVRYGNDLSKSNL LQTGLPQGAVTSCSLFNVFINDIAELVQTVTGVKCLLYA DDLVLWYSAPKKNAQKRTEIALNYALKLLANWCDNNG MVINTSKTAYQTFSLAHHSINPHLMYKDTPLEQTNESTY LGVTFDTKLTWKNHIAKIAERASNRLNVLKRLAGSLWG SARSTLNTTYKMFVQPIMLYCCETLVTASEAILKPLEKT HNQALRLITGGIKSTPIDAMLLVTGNSTIRSSVEEKALTL YEKLLRIPSDTFWSSYENRPRHLKTQVGLIQKAIELKKR LQTTDKPESLSLPQNPLADNEVQFCTQLLGHSMKADTPP DQMRLLALETINANYPADQWLQVFTDGSYVENQANVG AGVYSELFTFFAAAGQNRSAFEGEIEAIKIALGQLCCRD TKFTNAVILSDSQSAIQSIGNREPPKTAEIHECKKLYQLL KEKNKTIVLQWIPGHCGIIGNERADTLAKKGTTILQANIF LHNEDLNQERIQDSKKQRT	cd01650	521	782
Ingi	Ingi-1_HMi	.	Hymenolepis microstoma	ACTACT						

					TGATAACTA TTATAAAGC CTGCCATTT GATATAAT AATAATAA	MECTLYYQNVRGLRTKSSEFLSGVIGNSYTIISLTETWLE			
						DTIPSSQYFPPCYSVFRKDRDYAATRQRYGGGVLTAVD			
						TSLCARRRYDLELYSECVWIEVTCRDGLNFLIGNYYFPP			
						HLSSDVFARHFRDLEGKLDVTKFRVHIYGDFNLPGIDW			
						LANYNLITSDVTSLKASSLFDFIYFNGFEQRNTIQNSAGN			
						VLDLVLVSTPINYISPITDPLTRVDHFHPPFAVSFCFPLRK			
						QALCSYEYLLYSSGDYLNMYKYFQCYDWKPILNNKCA			
						DSAAEALTEVVSSAINDFVPKRVSRLNKYPAWFSKELRS			
						CLRRKLHYHRLFKRSGLDKWYAKFSECRALAKWLFKK			
						DESAHLHGLESDMYAKPQNFWKYVPTKKKRNSSVAPM			
						RNSSGYTCDSIEIANMFAEHFKSCFSSASVPDNNATCHG			
				CATTTGG	TAGTCACG	TSDFLSFIPIDEPEIRESIRNMKSKLSAGSDGIPSFIVKGCA			
				CCGGCAG	ACCTTGTTT	ELLTPVLLHIFNLSIHSGIFPSIWKRSVVVPIHKAGDPSTV			
				GGTGCGT	TGTTGCCTG	NNYRPVSLLCSFSKVFEMIMHKRLYRHFRHKMNPEQHG			
				TTTTCGCC	CATTTATCC	FLKGKSVETNLCSFLDYAVPLVCNREQVDVIYFDMTKA			
				CTTTTTTC	TTGATGACG	FDKVNHSTLLTKLDSYGLCLTYCQWFKSYLTCRTNYVR			
				GGCATTC	AGCGGTGT	VSGSMSDYFSSPSGVPQGSNLGPLLFLMFVNDISLCAKN			
				TTAAAGA	ACCACTATT	SKVLLYADDIKLFRRVQSHSDCVLLQEDALQISQWCAK			
				ACTCTCG	TAGGCCTTG	FLHKVVNGAVDSPLLLSTVCFHAPLRSNRFQSPFYPSAL	cd01650	493	745
				GACACGG	TGCTGTTTT	NYLPLNEQKTRTMSLTRKRDSLAFSYTIGGSVITRVSVIR			
				ACACCTC	TGGACACC	DLGIFIDSSLNFNHHVLTICNASFRNLGAISRFTRKFSSPV			
				GGTCGAT	CAACAAAA	CLVRLFCCLVRTKLEYASSVWNCINLTASCCIERVQKRF			
			Ixodes scapularis	GTAAACG	ATAAAGAT	IGIIYDRYFDRTCYYSYEQLLKKLGLCTLLDRRRLRDLM			
	CR1-1_IS	.		CATAATC	TATTATTA	FLHKVVNGAVDSPLLLSTVCFHAPLRSNRFQSPFYPSAL			
CR1				AGGTACC		NVASPLTRMQLEFNKLDQSAVDIFSDLIVFRHRLESIICT			
				AGGATGT	TGACTGTGT				
				CGGATGT	TAATTTATT				
				GCCTTTA	CTTTGTGTG				
				ATAGCCA	CTGTATGCA				
				GGAGATC	TCTTATATA	RRGIIFFISYAAPLGGSSADTLRTTALGLVYSTAEYCAPV			
				GTTCTGA	CCTTATTTT	WMNSAHTRKIDVQLNGTMRIVSGTIRSTPTSWLPILSHIP			
				TTGTCTG	ATGTCTCTT	PADLRRSNALLREYRKITNNPRLPIHHDTPDLERNRLRS			
				GCTGCCA	TTGTATTCT	RHPTMKTAKELHNTDFNLIESWKERWRQACPEQCQALP			
	Vingi-35_LMi	.	Locusta migratoria	GTTAGCG	TGTATGTTT	CITERPPGYELPRKAWSSLNRIRTNHGRCADSLYRWGKI			
Vingi				CTAGTTC	TACACATTT	PSPNCDCGADRQTIRHIIEDCEMRAYDGNVSDFLTAKDA			
				ATTTAGT	TTGTATTTT	AIDYIRNLDVCI	NA	NA	NA

				TCACGTTTTACAACGCTATCCGGGGTGCACCACGGGGAGGCGAACTAGCATTACACCCTTACAAACGCCCGACTTATAGATAGCACGCTTGCCGCAGAACATTCCAGGTGACAGTGGGGGACCAAATATGCCCAAAGAACCCTAAACAACGGATTACCACAGGGCTCCGTCCTAGCCCCATTGCTATTCAGCATGTACATCGCTGACATGCCCCAAACAACTCCAGAAAATTTGAATATGCAGACGACTGGGCTTAGCGTCAACATAAGAGCTTTGAGAC	TTGTATGTATAGTTGTATAAGACCTGCTAAATGATGTGCACTGCCATACGATAAATAAATAAATAATA			

				AATGGAGTGACCTCGTCGTCCAACTTGGCGACTGGCGCATATTTCCGGGGCCGAGCCTACAAAACGGGACGTGCTTCCATATCTGTAATAAATCAGCAGGCAGACGACGACGTCTACTTTGAGGGAACTCGCTCCGGCATAACGGTATCCAAAGTACCTCGGAGTCACATTAGACAGAACCCTCACCTATAAGCGACATCTGGAGAACACTGCAGCCAAACTAG					
Ving i	Vingi - 24_L Mi	.	Locusta migratori a	CGACTGTTCGCGTCCGTAGCAATCCTCAAGCTACACGTGTCCCCTGGGTGGTGCTA	TGACTGCAACATCAGGTCTACCCAGGCCAAAAACAGACCTGTATAAAGTGACAGATGATGCAATTTC	MIMKKKQQTPESVLTRELGSTTRICQLNVEGLSSSKSQFLSKFLMENDVDVAVLQETHIENXEQQRSRGKIHGYDLLGAVYHRHYGLATYVRSSIENAALVSLSADCFQQIVTKIGELTVVNIYKPPQTEWPQNVIETHPHPTIYAGDFNSHHTDWKYRTTDINGEKLSNWAEENHLHLVFDAKDKNTFKSAIHNTETNPDLSFTTRNSQSHPIPTRRKIXESFPHSQHRPILLEIGINIPLIRSLPRPRWNLQKANWDKFSNDIDQTLRWIPARWYNYDRFVNIIIAAAKRYIPRGYRKEYIPGWNEETEEL	cd016 50	481	707

AAGAACA	ATGGCTGG	HQRFLQENDAEIADKLLRNLDVARRKKWEETTEALNFT			
GTCACTT	ATGGTTTAG	KSSRKAWNLLRKLGDASAPEREVPGVSPNQIASRIVSLS			
CGAGGCA	ACATTAATA	RAPRDKKRTTFVKRHLKSFRQSVEAPNKISRPFTVEEVN			
CGAAGTC	TATAAGGA	TAISKIKKGKAPGFDGVLPEFLIHSGRRTREWLSQFFSHI			
AGAGCCA	CCATGTTCG	MQTGLVPPXFKKAKIISLNNIFVKPGKPKDKPESYRPISL			
GAAGTCA	CTGCTTATC	LSTTYKLLERLILGRIEPILEEHLPVEQXGFRHNRGCEDQ			
CGGCAGC	TTTCTTAAT	VLSLTSHIEAGFQQXLKTTVAFVDLTVAYDTVWKDGLL			
CAGACCA	TCAGTTAAC	YKLYRTVPCQKLLRLFNNMLRDRTFQIIMGDKASRVRK			
CTTGTGG	ATAAATTGT	INNGLPQGSVLAPALFNLYISDMPETKGRKFGYADDLAI			
ACTGTCC	GTAAATTTA	AVQAKXMEDADTLLSQDLEVVGKFYADWRLIVSTSKT			
TGGCTAT	CTGCTCATT	ETTSFHLTIGKQQNLXKYFSTIKCSPIISSQSILVSHWIAPS			
CAAACAC	TCATTGTTG	PSGNTWKQRQPKSTLAITSCTNCVAAGGEHMLPLFSLN			
TTGGGCT	TTTTTCTAT	NREATKSPKVLLNNKVLPHNQFPKYLGVTLDRTLSFRK			
GATTAAT	TACTTATTT	HLEATAAKVNTRNNILHKLCGSRWGAHASTLRGSALAL			
GCA	TAGTTTCCT	VYSAAEYCSSSWLNSSHVSKLDARLHDTMRTISGCIKST			
	AACTTAATG	PVQWLPVLSHIAPPPLRRKGCLVRQLKRIQSDPDLPVHK			
	TTCCTTTAT	DIPHLSLRRLKSRKPPLVSAQQLQDADFNITEQWRTDWS			
	TAGTGTTCA	TVNFTALWPDPTERPTGSNFPADNGKRLTGSEHLTESVR			
	TTCAATTGT	ILCTSGAYRKSSVRLWCFASDNRPYCV			
	TCATCTTAC				
	TGTACGATG				
	TATATATCT				
	TTCATAAAT				
	TTGTATTAA				
	TATTTTATA				
	AATATTTTT				
	TTCCAAATT				
	CCACTGCAT				
	CACACATGT				
	AATGGACTT				
	TTTACTGCA				
	TACCTGTAT				
	ACTTCTCTG				
	TAATTAATG				
	CACTTAAAT				
	TTTGATATA				
	AGCTGTTTG				
	TTGTCTCTG				
	CTTCCCACC				
	TGGAGAAG				

				TGTCATTGT					
				GTAATGTAT					
				CTTCATGTC					
				CAAGTGTTT					
				ATGTCATAC					
				GATAAATA					
				AATAAAT					
Nimb					MRCRSSHPSILQWNCGGFFHHLAELQQLMCHSLLCIAY				
					RKRGFLFRGPLPYVATGAISKIAPSVTGSLEVCASMSLPL				
					SPVRCAPHHESQGLWRCVPLCPCHCPQCAVPLTTNLEA				
					VAVRIRVRQELTICSLYLPPDEVVCLADLVPWSLNFPRR				
					FSFWGILMPTTLCGVGRRPRLVVGMPRALVAQLPPPFLL				
					LGDFNAHHPMWGGTQTSARGRDVEMLLTQRNLCLLNT				
					GAPTHFSVARGTYSAIDLSICSPSLLPHLNWNVHDDLCG				
					SDHFPIFLSLPETQAANRLPHWSLHRADWPAYASAAVV				
					TTPLPAGIEVAVQQVTKAIVSAAEKTVPTASASLRRKSV				
					PWWTQEIAEAIREERRRALQRTGDTLLGHPGTSARASAH				
					RRHPSVENLIAFKRLRARARRLIRQSKKDSWERFVASIG				
					SRASPSRVWTQVRRLFGYRPPTGIPSLTIQGALCTDPQAI				
					AEYMAEHYARCSATETYSADFRTFKERVERKQLSFRTR				
					RQELYNAPFSEWEFETALAACPDTSPGPDRIHSQFLRHL				
					SAEFQHQLLSLFNRIWADGVFPSQWRESTVIPVLKPGAD				
					PLVADSYRPISLTNVLCKLLERMVSRRLMWILEDRGLLA				
					PQQGGFRQGRSAIDNLVSLESAIRTAFTRREHLVAVFFD				
					LRKAYDTTWRHHILATLHGWGLRGSLPIFIQNFLSPRSF				
					QVRVGTCLSSPHIQENGVPQGSVLSVPLFLIAINGLAAA				
					VGSSVSSSLFADDFCIYFSSATIGVAERRLQGAIRKAQSW				
					ASAHGFHFSGSKTRVMHFCRRRALHLDPELYLGNELLR				
					VEETHRFLGMVFDARLTWQPHLRQLRQKCWKHLNILR				
					CLGHTAWGASRPTLLQLYKALIQSRLDYGSIVYDSAAPS				
					ALQVLDPVLHSGVRLATGAFRTSPVTSLLAEAGVPPLRF				
					RRQRLLASYAAHVHALPHHPNYRLLFPDVALPSPRRRP				
					RSGLPIAVRVRSLLSSLESLPIPPTFHYPSRAPPWSVPVRS				
					FGIPSYSPHFSLPFPCSSLVRTRPQLRLDLSRDSKASVAPL				
					NLRRQFQAVLGSFHGAEIVYTDGSLVDGRCGYAYAHD				
					EHIEQHSLPCGSSIFTAELIAIRRALEHLRACPGESYVICS				
					DSLSGLQALDQCFPKHPLVRSIQESVTFLFSLAALWSSFG				
					PRDTLEFRGTNAWIGWLSRQLLQAP				

				CTCAAGGTGTGCTACTGCCGAGGAGATGCATGGCTGTTGAGGTTGAACAGTCGCTAGCGAGCAACCTCTGAGGAACCTGCCGCACCTCAGTTGTATAAGGCTTACCCAGGCAAGCGGGGCTCTGCCTGGGTGGACTTAGTTCCCTAGCTGCTCGTGGGACGAAATGAAACGAGCCAAATAGCGTGGTGGACGTAAACCACCGTCTCCAAGCGTGGAAGAGTTGCTTTAATGCAGCGACGTATGATCCCCGAACGTTCCCTTCCCTGGCCACCCCTGGGAG	GTTGGACTGTCCCCTTTTAGCTGCCTTACGACAGTCTTTTGACCTGCCAGCTGCGCTTCCTTTTATGCTTGGTGACGACGCTGCTACAGCTGATCGTGTTTTACGTTTTATCCGTGCAGCTGGGTTTTATCATTTGATCTAAAGGTGGGTACCTCAGCCTGATCTCTCAGGCACCCATCTCTCACAGATTTTTACTCTTCTTATTTTGTGTTGTCTGACCTGCTTTTACTCGTATGTTTTAGAATGTGCAGAGTGGCTGGCTCATCCTTTATTATCCTCGTGATCAGCCAGCCACAATATTCTGCTGTAACTTTTTAGTCCAGTTATGTTTATTTCCTCTTTGTGTTTTCTT				

				GAACGGC GGGCTAA AGACAGG GATGAGC TTACTCC CCAAGTA TCTTGTTT GCAGTCG AACAGAT GGGGAAA CTTTCAT GTCGACG AAGCCAC TTTTCTTC GTGAAAA CCTTGAA GATAAGT TTGGTGA AGTCTCG GATGTCA AGATGAG GAGCGGT TCGGTAC TGATCGA AACAGCG TCATCCG CCCAATC GAAAGCA TTGCTTG CTTGCGA TAAGCTG GGTGACG TTCCGGT CAGTATA ACACCCC ACAAATC GCTAACA TGATCCA GGGTCTG ATCTATC	TTTGGTGTT CCGAGCTA GTGTTCCTT ATATTGTGT ACCCATTGT GTGTTTCTC TTCTGTGTG ATTCAGCGC TCGTTTTTT CAGCTTGCA TTACACTGT ATTGTCTTT GCTCGCAA CAGGGGAC TGATGACC AGCAGTCT AGTCCCCTC TAAACAAC AACAACAA C				

ACAGGGA
CCTTCTTT
TACAGTC
GATGAGA
GCTACGC
GCCAACC
TAGAGCG
ACGGGGT
GTCCACT
TCGTCCG
GCGCGTG
CATAGGG
GTCCGAA
GGATAAT
AGGGTAG
CCACCGG
TGCTTTC
ATCTTGG
CCTTTGA
GGCGATA
CATTGCC
CGAAAAG
GTCAAGG
TCATGAT
CTACCGG
TGTGACG
TAAAGCC
TTATATC
CCTCCCC
CTATGCG
GTGTTTC
AAATGCT
GGAAATT
TGGACAT
ATGTCCT
CCAGATG
TAGCTCC
AGCCCCA
CATGTAG
GGAATGC

				GGGCGTCCATCACACCCGGACGCTCCGTGTACTCCTCCACCTATCTGTGTAAACTGTGGGCAGCAACATCCACCCTGTTCTCCAGACTGTACCGTGCTCCGCAAAGAGCGAAAGATTATGGAATCAAGACCCTGGATCGACTGACATATACGGAGGCACGTAAAAAATATGAACGGCTTCACCCAGTTTCCATGACGACGTCTTTCGCTGCAGTTACATCGTCGCCACCCTTCTGATAACACCTACCGTGCAACAACCTTCTGTGGGCCCAAAGGGCCGT					

				CAAAGCT CATCTGC CTCCTCA AGGTCGG GGGCACT CCTCTCA CTGCCCC CCACTAC CACCTGG GGGGCAA CTCCACC CCCACCA TCGGGGA CACCAGT CCCCACC TCCCAGC CGGAGAA GGAGCAG CCTCCTC CGGCTCC CCTCGCT CGGAAGG GGTCCCT CGGGGCA CTCCCTTC TGAGGCT TCTCCTA GTGCTCC AGCGATC CAGCCAA TGGCTGA AGGATCC AAAACCT GCTGGTA GAAGGGC TTCCCGC TCTTCTTC TACCCCT GGCGCCG CTACAGA					

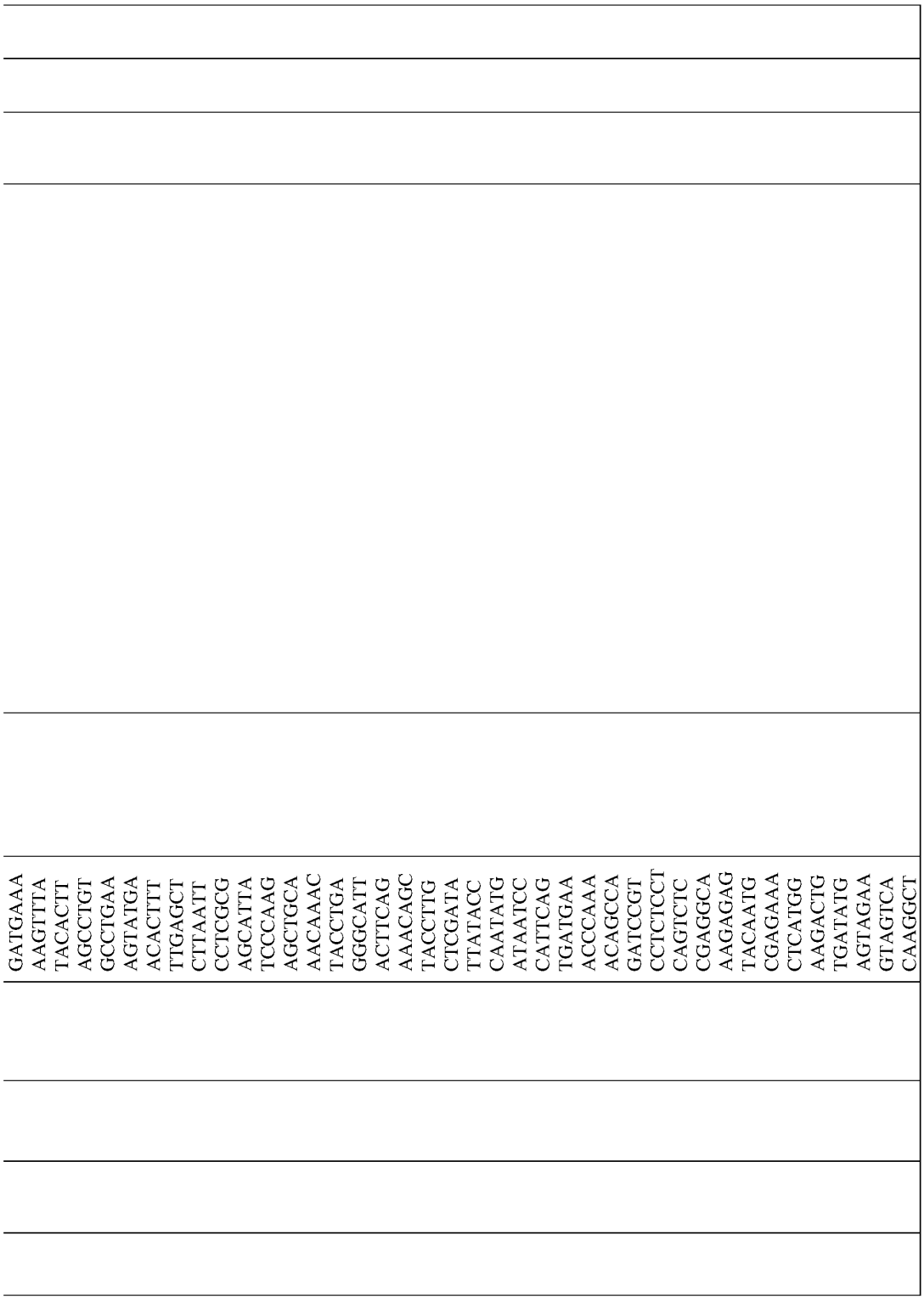
			GAAGCCC							
			TCCCGGC							
			CCTGCAA							
			ACCTAAG							
			GGCCAGC							
			GAGAGGG							
			ACGGAAG							
			AAGAAGA							
			CAAAAGA							
			CTCTTCG							
			GTGGCTC							
			TAAGCCA							
			CCGCTAC							
			TCATCTCT							
			ACATCGG							
			ACG							
					TAGTTTGTT					
					ACACGTGT					
					AATATTATA					
					TGTATCTTT					
					GTTATTAAA					
				GAGCGAA	ACATCAAA					
				AATTATA	TAACTGTAA					
				GCCTTAC	ATTTTGTGA					
		Vingi-46_LMi		TMAACCA	TGTAGCCAT					
				GTAGCAA	ACGCTAAA					
	Ving i		Locusta migratoria	ACSAAGT	TAAATAAG					
				GATAA	AAATA					
				ATGTACA	TAAAATAT	TKTWCMMGGFPLTLLFICSQTSNRTVIXXEQAGFRPNR				
				TAGTCAG	GTAGAACG	NCIDQVLSLTTFIEAGFQNKEKTSVVFIDLTAAYDTVWR				
				AAGTGCT	CCTGGTTAG	PGVILKLLQAIPSIKINHLINEMLSNRCFQVIMGNNISSPM				
				CAACTGA	GTGTAAGA	KLNNGLPQGSVLAPLLFSLYISDMPETKSRKFGYADDW				
				ACAGCAC	GAGGGCCT	ALATRHKDLETTESILTNDLSTIGEYFRKWRLQPSATKT				
				ATGTCCT	GAAGGCCC	EVSTFHLSNKLANKELTVHFEGKVLRHNRNPKYLGVTL				
				CTTCTCA	TAATCTTGC	DRTLSYKQHLAKTAAKIKTRNNILQKLCGTSWGSSAST				
				ATCGGAA	CAGGTTAA	LRTSALGLVYPVAEYCSPAWLNSSHTKLVDAQLNHTM				
				TTACAGT	ATAAATGA	RLITGTIRPTPTFWLPILSHIQPPSLRRKNALVSEYKKLQK				
				TACTCAC	GTAAATAA	NPHLPVHEDIPSIRKARLRSRHPPLETARTLVTNSFNTDD				
				TATCAAT	ATAAA	CWRQTWQINATPEQLKMPCISTKPPGFDLPRKDWSALN				
						RIRTNTGRCADLLYKWKKIASPSCDCGAEKQTVYHVVQ	cd01650	26	233	
						ECPLRAFAGDQTEFLMVTEGSLDYIHGLDISL				
						MMNSRDAAHRRFKASTKPEHYEEYRKLRNKVKQCIRN				
						AKIRYARSFVCSNTNPATLWKNLRSLGVGKATPESVFQI				
						SADELNEFFSAPINCQAQINNRPEESPSHITNDDTFHLKH				
						VTINTTRKAIMRISSEAIGNDGISITMIKNVADSLLPVLTD				
						IYLISLSFSANQHIACVSKALEYIVHDQITEHLCKFSLFDK				
						YQSGFRKNHSTNTALIKVTDDLKFAMNNREATILTLLDF				
						SKAFDTVNFDILLNKMRQLKFSNSALKWFESYLKNRQQ				
						SVVCGNEKSSWRNVVSGVPQGSVLGPLLFSLYVNDISS				
						VLTFCKYHFYADDLQLYQSTRPEDINTAIAQMNEDLSSV				
						VTWAKNLGLKLNAKKSQVILVAHQKLISSELRDQLPPIQ				
						LDGIPIPYQKTVKNLGVIIDEHLNWADNTVTVCRKTSAC	cd01650	173	410	

				ATGACAC GTCATCG TGATGGG AGACCTT AACATGA CTTGTTA AGAGACA CTCCATC CACAATA AACTAAG AGATTGC TCCAACA ACATGAA CATTCTG CCATTAG AACCTAC TCACCAT ACGGAGC ACACTCT TATAGAT GTGATCG CAACGAA AAAGATC GACAAAG TACAGAT CTAAGTC AAACGTC AGCCCGC CTCTCAG CACACGA TGCAGCG CCAAAAG TTAAATC ACGTTAT ATTACTC GTAGAAA CGTGAAA CAAATAG ACCGTGA GGCTCTA		LYALKKFRNIFPQDVKCKLVQSLVLPNLHYCDMIQHGT SSENTRRLELTMNACVRYICNIRRYDHISASYTQLGWLR PDKLRDYHTLCLLHRLLSEQHHSTSPQNLTIFRAIIILTQD PTYLVPLCPFIKRKHLQTPSLSLLFASGTSCPALFARFSLL LRLRKSNTSLCHPLKTLAKTHVYSQVSLCVIVKPMLFSP LFLFLLFLSPNQCSSLRYFSSSSFFHHNLLFRYISLCEYYY THFSPPFPPLPPCPLLLAAIIKTFPYCNTCHVSYLSNTNIN CVYVLLLLVLIENFNFCRFSITSVILEM			

				ACAGCCAACTGCTCAGAAATCTCATGGCATCAAGTAACCAGAGAACCCACGATTGATGGAAAAATTTTTGAACTTAGAATAAACTTATTTCCTATATGACAAGCATGCCCCCTACAATCCGTGTAAGAAAACCCCTGCTCCATGGCTGACTAAAGAATTACGTCAA					
Vingi	Vingi-16_LMi	.	Locusta migratoria	TCGTTTCAGAACCTGGTCTAACAGTTACTTCAGTAAATATAGAAGGCATGTCACAATCCAAGCAACAGCTGCTTCAAGAACTGTGCCATGAACAGAAGTTTGATGTACTGTG	TAAATAATTGTCTTGTAAATATGATGTATTCACATAACATGTAAATATTGTACATTAGTTTATACATTGTAAGTGCTTCTGACACGAATAAATAAATAAATAA	PHQNLLRPFTQAEVDLALSKCKNGKAAGLDDIRVEQIKNFGPITKAWLMELMNNCVQSCKIPKLWRKAKVIAIRKPGKDQNDPKSYRPISLLCHLYKLLERLILHRLMEKLEEYTYHTGWLTGEELHISSDETYTAHRRWVSKERDHRSCIYRSDSSLRYCKQKIFLSNFRDKEADGDRKRMDYPRESVLAPALFNIYTNDQPLPEGTQSFIYADDCAITAQAHCFETVEENLSKALREMSAYYRENQLRPNPGKTQTCAFHLKNRQAARTLKITWENTTLEHCTYPKYLGVTLDRALTYKRHCLNTKQKVAARNNVIRKLTGTTWGAHPNTVRTSSLALSYSAGEYASPVWFKSSHAKIVDVALNDTCRIITGCLKPTPLNKLYSLAGIAPPDIRREVAAKSERGKSETSEAHPLYGHQPANQRLKSRKSFLRSTEALVGTSQQCRTELWRARYTGTAGWFTPCEKLPPGYTEKWSTWRSLNRLRSGVTKCKSNLVKWGFLVESPLCECGATQTTSHLLQCTQCPATCTPEDLLHATPNAIQVANFWASKV	SSF56672	19	256

				CATACAA GAAACTC ACAGAGA TATCCAT CAAATGC GACCGAC AATACCT GGTTTGC AGCTAGC TGCAGAA GTAACCC ATGCTAA ACGTGGA GGTGCTG TTTTTGTC AAACCAG GAATACG AGTCTTG AGTGCAT ATAACTA AATGACA ATAACAT CGAGATC ATCACCT TGGAACT TAGTAAC AGTGTGA TCACTTC GGTGATA AGCGGTC CTCCTTCC GAAAATT CTCCTTG CTCCTCT AAAAACT TCATTCA GGAAAAA TATCTTG GTAATAT GGGATAT					

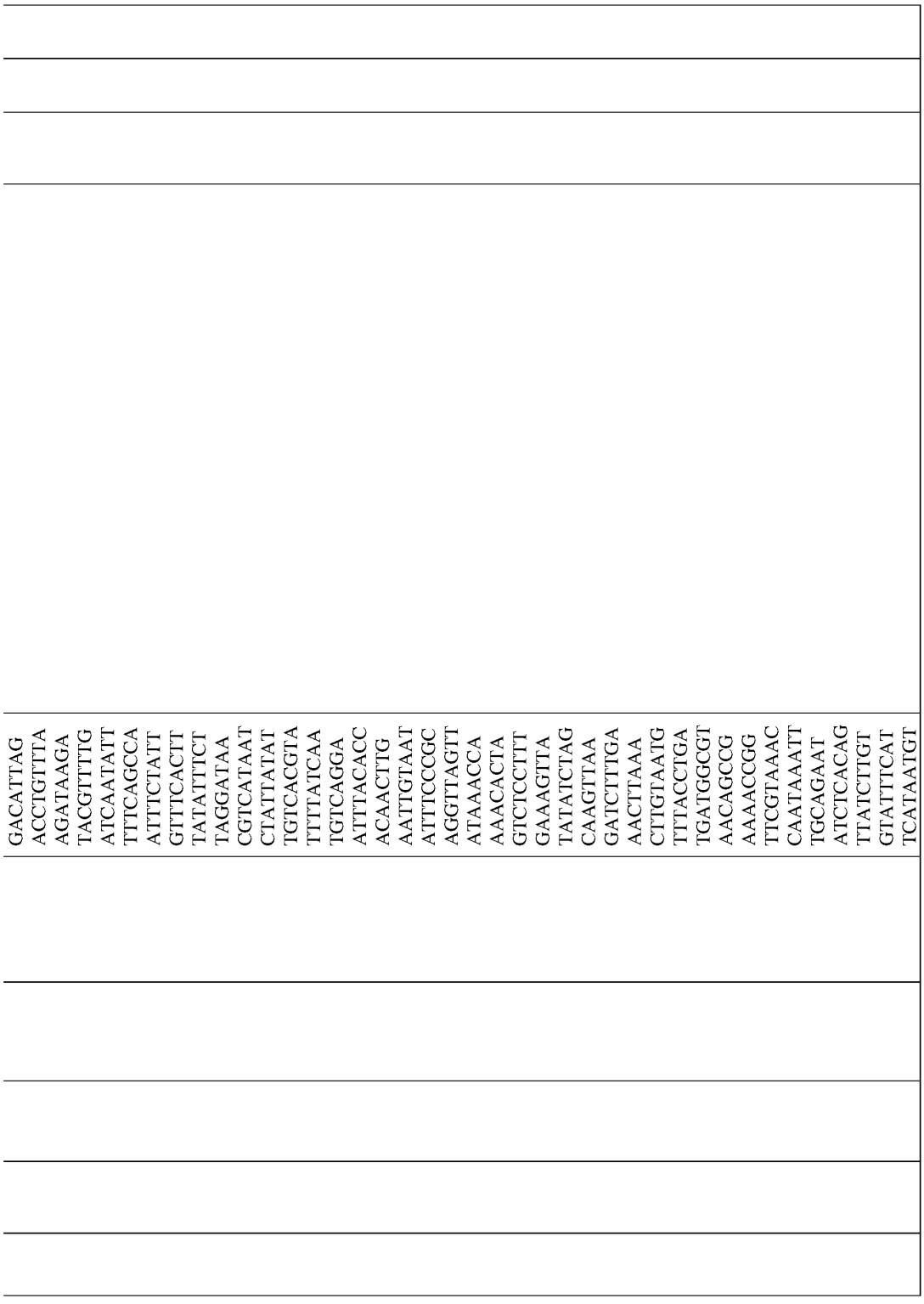
				GAAAATG TGAACCC GTTATGT CTCGAGC TGAAACC AACACCC WCACTCT CACTCAC GACAGCA AACTACC TCATCCTT CAAAGTG AAGGTGG CAACGTG ACCTCTA TTTGTGA GTGAAAC TAGCCAA CAGTGTT ACAATCA ATGTGTA ATTTCAT ACCCACC ACACAAC ACGGGTA CCAATAA TGTGTCG AATTCTT CCCATCA TAAGACC CCAAGAA ACTGGTT TTAGACG AAGGGAT AGTTTCA AAAAGGC AGATTGG CCGAAAT TCTCCAA AATGCTG					


GATGAAA
AAGTTTA
TACACTT
AGCCTGT
GCCTGAA
AGTATGA
ACACTTT
TTGAGCT
CTTAATT
CCTCGCG
AGCATTA
TCCCAAG
AGCTGCA
AACAAAC
TACCTGA
GGGCATT
ACTTCAG
AAACAGC
TACCTTG
CTCGATA
TTATACC
CAATATG
ATAATCC
CATTCAG
TGATGAA
ACCCAAA
ACAGCCA
GATCCGT
CCTCTCCT
CAGTCTC
CGAGGCA
AAGAGAG
TACAATG
CGAGAAA
CTCATGG
AAGACTG
TGATATG
AGTAGAA
GTAGTCA
CAAGGCT

				GGAGGCTTCTAAAGGCTTAATAATGACCCTTCACAGGAATATGCAGGAATATTCGATCTGATAAAATTCATACATGAAAACACCTCATTCAACCAGGTCTGGTAAGAAGACATGTGAGAGATAA					
Penelope	Penelope-2_LMi	.	Locusta migratoria	AAGTTGTTTGAACATTCATATGCGGGAAAAGTTATTCACATTGTTTTTACCTCTACGGGAGGAAATGTATCGGTGTTTACGGAAGTTGGATAAACTCCGCAGCAGTAGAGTAAGGTTATGTGCTTGTCGTGTTGAAGTGTCCATATTGTTCCACG	TAACAGTCACTATCGATATTTTTGATGTTTGGTCATCTTTGGTTGTGTTGACTCTGCAGTGACTTGTACTGTGTGTGTTATCTTCCTTGTTTACCCTTTGCAAACCGAGGTATTAAATACGCTTGCACATTGCTTCCTTGAGTGCAGTTGTGCCTTGAGAATGGCAGGGTGTCCTCCTGCCGAAATATCGG	MIAANVSEFSWDWLDGASWTQSDWEFQSVTARHSAKFERFLGSESDSNLISRRSVINLTEKTYDDAALSVLGKGLNFAPTPKRLPVVDFISSIEQAVSKLPSDSAEEARREACRVLTGARPPKCNITAAEASALRSLRMDPDIVILPADKGNATVLLNKDDYIQKMQCLLSDSAYRRIGADPTKSVERKTNNLLKKSPFSQETIKSLNSYCAVPPRLYGLPKVHKENIPLRPIVSNIGAPTYRVAKHLASLLSPMVGRCEHHIKNSADFLRRLQGLRLNDSDILVSFDVVSLFTRVPLSDSLQLIEVKFGVELTNLFRHVLTSTYFLFNGQYYEQTDGVAMGSPLSPIVANLFMEDFEERALVSATLKPACFFRYVDDTFVVWPHGSENLNGFLEHLNSIHPNIQFTMEVEKNGCLPFLDVLVRRKVDGTLGHSVYRKPTHTDLYLRADSCHHPAQREGVLRTLVHRAHVISDPESLSAELSHLEVTFRQNGYSERQIKRALRYQPSVQRVSDDSNEVAPKSMAFLPYAGSISNKIGRILRKYDVKCVFRPPSKIKALLGSVKDDLGLRKAGVYRIPCSCGMSYIGQTIRTVEDRCIEHKRYTRLQQPNKSAVAEHCLGTGHPMEYNNTEILACTSTYWDSVIKEAVEIKLASDLINRDGGFCLNSAWNPALSLVKKQRDKINAASPVD	cd00304	339	423

			TTTGCTA AGGTTGT GCACCAT ATTAATT CTCCTGC CGCCACG TATCAAG CAACGTG CTGCTTA GCTCTTG TTCGTGA AAGGATC CACTTTA TCGTCGT CAATTAG ATTTTGTT TCCAAGG AGTTATT TCGGCTT CATTTA	CGGTTGTCG ACGACGTC ACCCGGCA GCATACCC GTAAGTTGT TTGAACATT CATTAC					
Vingi-26_LMi		Locusta migratoria	ACGAGTA CTACTAA AAGGATT GCAAAGA ATACTCA CGGCCAT CCCAGTA GAAAAGC CGGCTTT CGTSCCA GCCGAAG CGTGCAG ACCAAGT CCTGCTC TTACTAC ACATATT GAAGCAG CTTCCAA AAGAAAT TAAAAAC	TAGATGTTT GTTGTGAAT ACATGTAAT GTAATATAT GTAATGAA TGTACCATA ATGTAATCA GACACTGA AATGTACTT AAATGCCA TACGCTAA ATAAATAA	TYENNSLSKSDKLIGNMIGNRGFQVIMGNISSQKFLQNG LPQGSVLAPLLFSLYIADMPETACTKFGYADDWXIATR HTNFDTMZETLTSDLSALAAYFRKWRLRPNATKTEATC FHLNNRGADRTLEVYFDDTRLNYNKEPSIQHTSVNGGL GLMPRRQKRLASISITEERTELWKYTLTTHASTTTRSQV SRVTLDRTLSFKTHVTKTAAKVRTRNNLIHKLCGTTWG STATTLRTSALSLVYSAAEYCAPVWLNSSHVKKIDVEL NNTMRIISGTIRSTPNIWLPVLSHILPPKMRRETALLREY NKIEANPELPIHSDIPDLRIKRLRSRHPPLSLDSLTGADIN SWRREWLQNAPQLATHQLLEKRMAPKRPTTCHLMPCIT EAVPGFDQPRKIWSTLNRLRTGHGRCADSLFKWGLLSS PACDCGAPKQTIQHIREECSARAFSGDFEEFLMATQKLL NIYKD	SSF5 6672	34	112

				TCTGTGGCCTTTTTGACTTGTCAGCAGCCTTCGATACCGCCTGGGACAAGGCCTGCTCTATAA					
Penelope	Penelope-69_LMi	.	Locusta migratoria	GTATTCTGAACTCAATTGATTTAGTGAACAAGTTAAAAGACTTCCCTATCCCCCCTAGTGCCAAGCTAATTTCTTTTGATGCCAGCAGCTTATTCACTAATATACCAGTTGGTGAGTGCCTAAACATCCTCACAGACTTACTGGATAAATCAGGAGTTAATCCACTTGTCACAGATGA	TAAATTGTAATGCCTTATTTTACTAGTTATAAAATCCTTGTGTAATCACAACTGTTTTGGCTGCTTATGTGATATAAACTTAATGTCTTATTTTACTAGTAACCATAGTTTAATTTAACCAGTTAAAATTTTAACATATATTTACAGATGTTTCATTCAACTTTGTCACACGCTTCCATGTTCCAGACATTGGACCTGTTTAAGATAATTTTTTGGTCAAGTTTAATTTAACCAGTTAAAATTTTAACATATATTCACAACGCTTCCATGTTTCTA	HKVSHTYLPITKLFSFQWQGIQTVRWSSYGFPLSPILAEIFMDNFENQFLESNPLSSKVGYWFRYVDDVLCLWTGTTRQLQTFLKQLNSAHKSINFTMEFGGNCINYLDLTIDISNHCHTFKIYRKPSTTDTVIPAQSQHPVTHKHAAFHSMIHRLLSIPMSKESFQIELNTIRQIASNNGYPLPWWLPPTLVDSIMHKKQKQHVSSLLYPVINTPTSQDHSKWCTIPYLGKVSVNVAKKLKSRNFKTSFYTHNTIGHFLPNGKDNPPALEKSGVYKITCNCGKFYIGQSGRAICTRLKEHHRSWKLRQGNSTFADHLLKEGHSYNVTHEVLHHVQKGRKMNFFEIMEINRHMYTSPNLVLNDQTQFPYSPLLSAEPNHNPVQTMPCADNVNYHLLPTCLIPLLLFHSHVIKALCNHNCLAAYQI	SSF56672	12	120


GACATTAG
ACCTGTTTA
AGATAAGA
TACGTTTTG
ATCAATATT
TTTCAGCCA
ATTTCTATT
GTTTCACTT
TATATTTCT
TAGGATAA
CGTCATAAT
CTATTATAT
TGTCACGTA
TTTTATCAA
TGTCAGGA
ATTTACACC
ACAACTTG
AATTGTAAT
ATTTCCCGC
AGGTTAGTT
ATAAACCA
AAACACTA
GTCTCCTTT
GAAAGTTA
TATATCTAG
CAAGTTAA
GATCTTTGA
AACTTAAA
CTTGTAATG
TTTACCTGA
TGATGGCGT
AACAGCCG
AAAACCGG
TTCGTAAAC
CAATAAATT
TGCAGAAT
ATCTCACAG
TTATCTTGT
GTATTTCAT
TCATAATGT

Dap hne				ATCAAATAT TCTGTCAAG AACCGACG GAACAGTC AGTGAATA T				
	Daph ne-60_L Mi		Locusta migratoria	TCAAAAT AAAAATA GTGTAAA GATAACA GACAATA AGGAACT CAAGGCG AAAACTG AAGCAGT GATGACG AAAGAAG AAATTTT CATATTG TCAACAA GTCATGG TAAGGGT TAGCAAA AATAATG TCGATAT CCAATCA GAATATA GTCAGTG GAATCTC AAAACCT GGAGCTC GTTGCGA GAGGTAC TGAAAAA CAGTGAA ACTACTG TAAAAAG TGGAGTG AGTAATT	TGATATAG ACGTACAC AGAATTTTA AATTGTAAC ATTACCATT GTGATATAT TGAGTCACT GTTCTATAT TGTACTGTA CAACAATG AGATGTAT ACAGAATTT TAATTGTAT TTTTTCTAA CAAATTTAC TATTGTTAC TACTTTTCA TTGTACTGT ATTGTATTA GGAAAATT TAAAACTA AGTATATG ATAATTGTA CCTATGATA AACACTGC AAATTACTA GCTGTTTAG ATATGTATT TAAATGTAT ATGACGAC ACCAATTTAC	MDSTAPCKKVLAKEILQAIKTASVDTFEKLALPAIGLLP MANEKENVPTVHKEETTMATPPEETCPQPTPTSASEALS PVETPEPSVQEESGGARASQAIPGVSQETVTASTPKGGK RSTADEAVSSVVSKRKTVPPLRLNDFLVECRKIKKPTKI CIPLLHKKKYVEMTSRISYPDESEADSLVPPCKEKKLIFL HTNVQSLKNKLFELQYWLEKLNCSILMINEHWLSCGEL ELSVPNGYELATGYCRXNKAHGGVSIYINKMLQLNYET LDVNEFCIESHFEVAAVLLTELEIIAVSVYYRPPNHSEDLFI ENLQKLIVSLLLKKKKKSQSSNFWRHLSHSSRKKNHRVLIF GDINIDIRYRSSTVTNLLDTLRSLNFYCINDKPTRKNACL DNIICNFHIGRDQFNTIKLAISDHDGLWLRLPLSHKTDTS QQRIIRPINEQKLNTLKSKLTLVNWNELLTRDMGANEST KRFLEVVTKIFEETCPAYYKKHSKNRSQSKPGTLKQWY TAELNNLRILMLMYYDRSQHSNIDRLRYKNIKKYYRSAI KEAKCKSNDDYILNSVNKCKAAWEVIKTEINHQKKDQ NIPISPDTLNKYFANPLMENNTEEDVNKAEELLPQLHNK AMGHFEWNHVTCRQVSVAISKLSNSRTEDHYGLSNFIIK QISQEILTPITNIINKILTEGVYPDSLKISVIKPVHKKGSTL SPDNYRPISLVPIISKITEYCMHQQLNNYFEHNNLICSEQF GFRPTLSTVKAVDNIITKVYDCFEKKTLSCATLLDLRKA FDMVSHNILIRKLQYYGIGENALRLLQLVNVNNKKSDP LPVKKGVPQGSVLSPFLFIVYINDLPSILPCEAVLYADDT TLIAQNSDMDELLLNVAVAIENCNNWFQANXLHINSSK TENIIFDLNAHKHENKSVKLLGFHIDQSLSWETHTDKLC SKLARVIYLLHRLRGCVSKTLLLYAYFAFFSSHLTYGILL WGNSPGSKAVFKWQKKAIRCVAGLAPQESCRHFFKEM KIMTVPSLYIYSCLVYVKENLHKYDLRMSVHNHNTRHR HAIDKPHVRLAKVHSSYKYLGPTFFNKLPNEAHSVPLN KFKNIISKWLKSKAFYTIKEFEESTIDDLEF	cd01650	693	926

Nimb	Nimb-24_Lmi		Locusta migratoria	GGTGGTG GTAATGA TGTGTAT AGAAATG AGAGCAG CCTTGCC GTGAAAG CACGAGA GAATCAC TAGGAAA ATTGTCA CAAGCGA AGGTATT TTTGGTA AGCATTC CTCATAG ACATGAC CGGAGAC GATTCAT GTGTAAA TAAAGAA ATCGTCA CAACAAA CAGGAAA TTCAGGA AGATCTG CAGCAGT TTTGTGC ACACTAC CTTTATTG AGGCTAC TAGCTCT ATCAGAG AACATTT TAGAGA	CCCCCCA GGGGGCC CACAGCT CTTTTGTG GGTACGT	ATGTAAAAC ATGTTTAAA GGTGAATA AAATATTCT ATTCTATTC	TGAGTCCTC TGGGTGGTC TTTTAGATC ATTATCTAT GTTCCACTT	MSKLRDGSVLIQAETPAQSRSLLTCDHLGDIPVSITPHKS LNRVQGVIYHRDLLLQTDEELRDNLQRRGVHFVRRVFR GPKDNRVATGAFVLAFEGDNLPEKVKVMMYRCDVQP YIPPPMRCFKCWRFGHMSSRCNAAATCRDCGRPSHPDT ACAPPPYCVNCGGHHSPCSQDCPALHKERKIVEYKTLD	cd01650	901	1159

GTGTGGC	GACGCGTCT	RLTYSEARRKYEKLHPVSLLTSYAAATTMAPPPVATSRP
GCGCACG	TTCGACTAC	ALPPSVGPMGRPSTSASQTSGGTPSSAGPVLPPTSGVVTP
GGTCCCC	AGCTTTTCT	APQPEKGSPSLASSSGKGTCSPLPTMSPQKSDTSQRLKD
GAGCCAT	TGTGTACGC	PRPESRLAPQSSHTPQAPPEKPPSSGRPKERREDKKEKKE
TGCGGCC	TATGTGGTT	RRKAQGDAGGLGAPSPCTGHPSRDARRFGLCRSRGCHG
TTCCTTCC	GGCTTTTCC	HGGSDFTARDECSCHPSSIFIVFMHRQSGDSLILQWNCG
TGTTTCCC	CTTTTTATA	GFFCHLAELRQLLSSFPTLCVALQETWFPGDRTPTLRGY
AGGCCGC	CCCATGGTC	RGYFNNRTVYNRVSGGVCLYVHTSVYSSPVPLQTALEA
TTTTCCTC	GGTCAACC	VAVRVRASRDTTICSLYLPPDADVAPADLAALIAQLPSPF
ACCCCGG	ACATTCATT	LLLGDFNAHHPMWGSNVTTGRGRDFELLFNQLDLCILN
TTCCCTCC	GGCTCTGTC	TGAPTHFSVAHGTFSAIDLSVCSPGLRPLLNWWVHDDL
CCTCCCT	ATTTTTATC	CGSDHFPIFLSLPHHQRHARLPRWSFHRADWAAYTSAA
AATCCTT	TCGTATTTG	MLDTPLPADATMAVDQIISVVVSAAESAIPSSSALPRRRP
CTTCCCA	TTACTGTGT	VPWWSPEIAAAIRNRRRALQQYRRHPTVDHLVAFKRLR
CCCTACC	GTCTGGCGT	AQTRRLIRQRKQACWEEYVSTIGPRVSSTQIWTKIRRISG
CTTCTCC	CTTTTGTAT	YHPPAGVAGLSIRGTIHTDPTIIAEYFAAHFALNSASTAY
AACCCCG	CGTCTCTTG	SQGFCTLKERAERKQLVFRTNRQEPYNAPFTEWEFQSIL
TCTCTAG	TTGACTCTG	ASCPNTSPGPDRIHCQMLKHLSVNSQRHLLDFFNRLWIE
GTGTTCT	TCAATATTT	GAYPSQWQESIVIPLLKPGADPLNVDSYRPISLTNVLCKL
CGCGATG	CATTGACG	FERLVSRRLLWVLEYRGLLAPQQGAFRQGRSAIDNLLSL
ACGGCCC	GTGCGTTCT	ESAIRTAFARRHHLVAIFFDLRKAYDTAWRHHILATLHG
GGCTATT	TTTATTCTG	WGLRGSLPLFIQNFLLSRRFRVRVGTSFSSPHIQENGVPQ
TCCACGG	CGCTATTTT	GSVLSVPLFLVAINGLVAAVGPSVSCSLYADDFCLYYST
TACTGCT	AAATCTAG	ASTSVAERRLQAAIHRAQNWASAHGFTFSPAKTRVMHF
GGTTTCG	CTCAGGCG	CRQRTVHPHPELYLDGELLQVVDTHRFLGLLFDSRLTW
CAATTGT	CATTTATCG	LPHLRQLKQRCSRHLDVLRCLSHTVWGAARPILLRLYN
GTCTTTGT	ACGATTATG	ALVQSRLDYGSQVYGSASPSALRVLDPVHHTGIRLATG
TTGCCGT	TTTTTCTGC	AFRTSPVTSLLLAEAGVAPLRYRRNTYSPSMLPESTVPLGI
TTTCACC	GTGTCGCCT	QITDSFSRGRSFVCRRGDPGRDSRRSGAVTSGLREPGLW
GATTTCG	AGCTATGA	LSLPICTESSRPSPPYRHPSGHWSLSYEPCHESSRRGWSR
GCCCAGT	AGTGGGGG	PIAIPAQHLLAQYAARVHCTPRHPNNRLLLFPGEVIRLPQR
TAGGTCC	ACTGATGG	RPRSGLPIAARVQSLLSTLDSFPLNRPFRSTFHAPPWTIPR
CCCTCGG	CCATAGTA	PRIRLDLSREPKTSTSPLVLRHQFYALLREFRNSEVIYTD
GGTTTGA	GTCTAGTCC	GSSDGDHTGYAYMHGAYVVEQRSLPPGSSVFTAELVAIY
CCTCCAT	CCCTAAAA	RALKHIRTCPGTSFLICSDSLSGLQALDQCYPRHTLVRAI
CAAAATT	CCCCCCATA	QESILALDDSGRSVTFIWTPGHVGIPGNESVDQLAKQAT
TCAGAGT	CCAACCAA	TCRPLAIGLPEADFRPLLRNKIRELWNAEWSALDVPNKL
AAGTGTG	CCAACCAA	RTIKDTTSVWKSSQLGSRRDSVALCRLRIGHTRLTHGHL
AGCCATT	C	LRREDPPVCQCCGAPLTVVHLLLLHCPILAPLRHSFNLPNT
TGGGGAA		LSLVLGDDASVANMVLRFLRAGGIYQLL

				GAACTCC CAACCTA GCCCCTA CGGCGCG GATTCCT ACCCTTT CCTTCCA AATCCCT CTTCCTC GCCGTAG CTCCCTC CCCCGCT AGGTCAC CAGCACG GTAGCCA GTCCGTG TGGTGGG GCTGTTA TGTACCC TCATGGT TGAGCCC CCTGAAT TGCACAG GGATCAC ACTGCCG ATGCCTG AGCTGTG ACCTCCT CATGAAT GCCAAGG AGTCGAT GCTTGTC TCTGCGG GGCATCA GGACTCC CAGCAAC GGTCATC CTGCCAG GTGGCGT TGCTGTG					

				GCTGGGT GGCGCCC TTGAGGA GAGCCCC TGTTTCG GAGTGGG TGGCATC GGGGCGG ACATTCC GCAATGA AGAGGAA ACAAACC TCTGAAA ATCCTGG CTCTAAG TCAGGTG TCTCAAG TGTGGTA AGTTTAA TTTCAAT GCTGCAA GATACGA CCCCAAA TCGTTTCC TTCCCTA ACTCCAT CATGGGG GGAACGT GGGCCCA CAAAGCA GACTGAA CCATACT CACCACG ATACCTA GTTGTTA CCAGACA GGATGGG GAATCTT TCCAGGC AACTAAG					

				CCTTCCTT CTTCGTC GAACACC TCGAAGA AAGGTTT GGAGAAG TTGCCGG T					
Dap hne	Daph ne-23_L Mi	.	Locusta migratoria	C	TAAAATAG AATTTTGTA ATGAAGAA CTGTGACA AACATATAT ATGTAAATT TGTATTACT GTGTTAAA GTGTTCTAA TAATATATG TGAAATTGC ATTAATCCT ATAACATTT AATACTTAT CTGTAAAAT TAATTACCT GTGTTTCAT GGTATTCAT TTATTAAGT GTGTTTCAC GGTGTATAT TCATTTACT AAGTGTAC CAGTACTTA ATTTTGAGT CAATCATAT TGTGTAAA ACCTTTCAC CATTATATT CATTTATTA CATATACCA GAACATAA	LYLKQPQQLHKVGLPIIHSLVKCLKPQEQHQKIALSLQC KTVSTPTTSPGRHPRRSYASALNTAATTPSSPVNTEETAT VAVPSEEGGGKSASFSSKFFSRKQSKEXHNITRXKTLKE VSGSSEVTKSISCNKQNKSFHIMYHNVQSLTNKLQELEII LSSDLNADVICITEHWLEKHQIEALKIPGFHLATSFCRTTI NMVGQLFLQNRITWWDSYFCKTEFKAMQSYSDLIKEK VFEMSAIELTDKHLVIVTVYRSPDGDIEEFLCKFEMMLN IVEPTKKKIICGDFNVDFLKSSNYKEELLYITKNFSLIPTV NTPTRITKHSKTTPDQIFIQGTFSSYNSKVVKMGYSDHE AVLLTLELYSEKTPPIKMIECRDFSDGNIANFCNQLSQET WYEIFYETDVNKMFNTFHQTYKYYFDIAFPLQQRKLKP AKNKCWITNGIKISSTMKRYLHSYKKQNHLSLEMQHYI KEYNKIYRKVIKEAKKMGNDEYIRKSKNKIQATWNIVR YETDTKPVAKSISLKLENRQVSEATTVANIFNDFFTKIAD DLIQSNFQTPNLHHTVNYKSPKGSMFLKQVTGSELMLA MSSLKKLSCGFDDIPDYIIKQTGYHIVEPLMHIINHSFSTG TFPNLLKIAKVTPLHKKGATDDVNNYRPVAQLSGFSKIS LKSSFLIDCQLQACSSAQWLLKNIFEKLFLNRLLEFVNK NSLLSKHQHGFRKSRSTTTAIYEYLNSILKAVDEKEMAT GIFLDLSKAFDVIDHGILLHKLERKGIRGISNEWIKSYLT NRQQRVVLKQDTTIGNISRRSTYTSEKTTIKYGVPQGST LGPILFLLYIDDLNDTVLAKQKTLFADDTSILITGIDQKH LQQSTDTSLLELTHWFKINKLIINTEKTVAMNFHLSHQK QQNNHTLKINNRSLKLVDDVKFLGIYLQSDLKWNTHIK ALTSKLSSICYMLRILSTSCNKEIVIQAYHTHFQSAIRYG VIFWGNAPCSHSVFKMQKRALRIICNMKKQESCRPLFPS LKILNLPCLYMFETIVFVKEYLGSTSAYQLNSSVHSYRT RNDSNIHVSYARTSAYQKSVIHIGTQLYNHLPGNIKSIKE DSIFKNKLKSFLMDHSFYSVDEYLSD	cd01650	637	930

Ving i	Vingi-19_LMi.	Locusta migratoria	GGGTACG	TAGTACTGT	GTCATAAAT	MATLVKNEILRNVDTITPPTEHVIGIKLGELTLYNVYKPP					
			ACGTGTA	AAGAGCCT	TGAACCAT	SQKWPVDVLPKANPQQFMLAISTPNIPDGVTRDPQLKA					
			GACAACC	AVLKPGKTGDNPKNYRPISLLCTSYKLFERILLTRLTPYI	ATTTTTATGT	LDPPAVYAGDFNSQHTRWGYPRSSAEGIGLNDWADNQ					
			TCCCCGA	EANLPTYQAGFRAGRNCCDQVLSLTTYTEKGFQNKMK	AGGCCTAA	DLHLLFDAKDKATFKSKAWGTTSTPDLTFVTRGATGLP					
			TGGTGTA	TGLVLIDLTSAYDTVWIEGLLLKLTRIIKCKQTYTLLKN	CTACGCTTT	VQGTRRVVEAFPRSQHHPVIVEIGLSIPVIKSPPIPRWNLR					
			CCCTGGG	MLTGRKIKVSLHGTNSKSVVLNNGLPQGSALAPALFNL	TTTGTTTTTG	KADWNKYRCFVEKTINRIPPKPENYKRFTNLLHKAALN					
			CGCTGTA	YIADLPETKSRKFAYADDLAIAYQSKNFNDLNETLNAD	ATTTAATGT	AIPRGSRKEYTPCWTKECETLMEEFEKCGDTETADRLIN					
			ACATGGG	LKVLNTYYSKWHLHPNPNKTTSTIMHLNNREAHTKLQL	CAATCATAT	AMNEERRSRWKEAMENLDFTHSSRKSWGLLRKLGGAT					
			ACAGCGG	TMNGGQIRHENKPKYLGVKLDRTLSYNAHLEDTKQKL	TGTATAAA	IPQRPSLGTTASEISAVLKHTSKIQLKTAEKKEVSRDLSN					
			CGAAAAG	KSRNAILSKLAGTTWGCGASTLRTSGLALVYSVAEYCS	ACCTTTTGT	LLVHNPTNMDIVKEVETEEITQALKLMKKGKAAGPDEI					
			TCTACAC	PVWMRSAHVAKVDVQLRETMRIISGTLRATPCAWLPVL	AAATACCTC	LPEFLKELGTLGRRWLAKLCTECISTGALPKIWKEAKVI					
			AACTATG	ANIEPPEIRRSQLAIRTHQKILGNLDLPIHEDLRATDRLKS	ATGACAAC	AVLKPGKTGDNPKNYRPISLLCTSYKLFERILLTRLTPYI					
			GCGGCCG	RTAFWKTSEELQEFNSKTAWKELWATSAPKNSNLVDD	ATCGATTGC	EANLPTYQAGFRAGRNCCDQVLSLTTYTEKGFQNKMK					
			GACGCAC	PNKKIEGFSLPRKLWTTLNRVRTGVGRCKQLMKKWGL	ATGGACAT	TGLVLIDLTSAYDTVWIEGLLLKLTRIIKCKQTYTLLKN					
			AAATGCC		GCTTAAAG	MLTGRKIKVSLHGTNSKSVVLNNGLPQGSALAPALFNL					
			TCCCGGT		ATGAATAA	YIADLPETKSRKFAYADDLAIAYQSKNFNDLNETLNAD					
			CCACTCA		ATAAATA	LKVLNTYYSKWHLHPNPNKTTSTIMHLNNREAHTKLQL					
			AAGGCAA			TMNGGQIRHENKPKYLGVKLDRTLSYNAHLEDTKQKL					
			ATGGAGC			KSRNAILSKLAGTTWGCGASTLRTSGLALVYSVAEYCS					
			AAGCAAG			PVWMRSAHVAKVDVQLRETMRIISGTLRATPCAWLPVL					
			GAACTTT			ANIEPPEIRRSQLAIRTHQKILGNLDLPIHEDLRATDRLKS					
			AACATGA			RTAFWKTSEELQEFNSKTAWKELWATSAPKNSNLVDD	cd01650	421	675		
			AGAGACT			PNKKIEGFSLPRKLWTTLNRVRTGVGRCKQLMKKWGL					

Ving i	Vingi-22_LMi		Locusta migratoria	GGACTTC CCTCATA GCCAACA CCGCCCA GTCTTCA TTGA	TAATTAAAC AATGTATA AAGTGTGT AAAGATAG CGATGTATG CCATACGAT AAATAAAT AAA	GAATGTA TGTCATTT CAACATA GAAGACT ACACAAG GACGAAG GGAGAAA TCCTTGA AAAAATT TTAACAG AGAACAA GTAAGTG TCGTACT TCTACAA GAGACCC ACCTCAC AGATGAA ACTATCG ACCGACT CAAAATA CCGGGAT TCACTAA TGTGAGC TACATTG GCCACGA CAAGCGT GGT	TDSSECECGETQTMDHLLACEVYGYNGDLKDIHRLSDS ALKWLEKICNVV	NWNHIPLITTTQRPRWNFRKAKWDIHPEKLDKCLGWIH PTANNDRFMGAVKSAAKAAVPRGFRKQYVRGWDKEC ENKYQRFLESGDPEIADDLLHSLDNARREKWIETVESLD FKTSSRKAWSLLRKLGGANKLHHNSTDISANRISAHIQA TSRAPRDKKHTRTIRRELSFLKKNSTPLDQYSSKFTTTEV REALKSVKAGKAPGLDGVHPEFLLTVGNMPKNGLLNFI QMSQRSAEECQGRKGTWIGWSTPRIPSYCGEYAKKWL AEFYSDVMATGNIPQKMKNSKIIALLKPGKPSNQPKSYR PIALLSVLYKLLLERLLYNRLSPKILESIPDEQSGFRPNRSC TDQVLSLTTHIEAGYQKSKKHQLHLSSTFPDNAYRGRLS KKQKTSIAFIDLSAAYDTVWIQGLLYKLIKIIPCSSTIRLIN NMLTGRAFQVITGERTSSTKVLNNGLPQGSVLAPLFFNL YISDIPNTQSRKFGYADDWAIAAQHESFEATEDILCSDL	cd01650	287	563

						KLLCGYFRKWRLQPNPTKTEVSSFHLCNHSARRELNVIF				
						EGKRLSHNNYPKYLGVTLDRTLSYKEHLTNTANKMRT				
						RNNIIHKLCGTTWGSTANTLRTSAVGLVYSVAEYCAPV				
						WINSPHTNRIDVVLNDTMRTISGTIKSTPTVWLPTLSHIP				
						PPEMRRKNALLREYNKITNNPQLPIHSDVADLERRRLKS				
						RRPALLSAKEMMNTNFEVEDLWKQHWQEHCPPKCQSF				
						MTGEKPPGFNQPRKVWTTLNRIRTGHGRSADAFYKWG				
						KTPSPNCDCGAEKQTVRHIVEECPFRAYKGTFNDFLAAT				
						EEAINYIKCLDVNL				
					TAGTTCCTG					
					CCTCTACCA					
					CCCTTCCCT					
					TCCTCTCCC					
					TGCCACTCT					
					CTATCCTTA					
					CATTCGTCT					
					CCCATCCCG					
					CCACACATT	PPPVQISSIHSTHQPNPTKPSKPGPNSLAAPHLHRKILLLC				
					TCATTCCCT	NPKYLHPIPQTDVLALQVLEQHSHTTSKKLSNLFTSYSR				
					TCCCTTCCT	LGLPLSTPSTTSRHTPQPLIADKPCLADLLHLPHPQKLPH				
					TTTCCGCAC	TTTQNPQPKQTVMNLSSKSLSPIEVSVLSKGLTFCPTPKF				
					TGTCCGTCA	NHAGLVKDLLSFSQSLQWKHHFATNPTIQTHTNTNIEPC				
					GACAGCAC	LTQFTPPSSQDPPTLPPNHPLLTFQNFLTSNLASPSFPKSL				
					TGGCGTATT	KMDTNLTPSERNAIYHLKTDPDLMILPADKGSTTVVLD				
					CCCCCACCC	RKDYVAEGLRQLSDSSTYTPCHNDPIPKTQQDLQSLLKS				
					ATGCCACA	LGPSQDLSPESISLLTPPTPRTPTFYMLPKVHKPNHPGRPI				
					CATCCCCCC	VAGYGAPTERISALVDQHLQPITHNLPSYIKDTNHFLHR				
					CCCCTCCTC	LSTVPVPLPPGTMLVTVDATSLYTNIPNAHGLAAVEHYL				
					TCCTGCTTC	SQCPTDCKPTTSFLVTLTNYILTRNYFSFEGITYKQICGT				
					CCTCTCCTT	AMGTRMAPSYANLFMGHLEESFLTTQNPKPLIWFRFID				
					GCCACCTCC	DIFIIWTEGEDTLSTFLQNLNTFSSIHFTWSYSTQRATFLD				
					TCACCCCCA	VDLFLKDGLISTSVHIKPTNHQQYLHFDSCHPFHTKKSL				
					CCACTGGTT	PYSLATRGRRICSDKQSLSKYAKRLTEAFTDRNYPPNLV				
				AACCATT	GCTCTCGCC	QKQISRALSVQSPSSSPTPTVRPQRSTPLVTQYHPGLEQL				
				AAATTAC	TAGGATGC	NQVLRQGFDYLSSCPEMRNTLPNILPTPSYSGIPPPTKPT				
				CCATTTC	CATGCTGG	EIPYQISSPPPPTVVFRRPPNLRNILVHPYSIPAPNPLPHGS				
				AGGCTGT	GTAGCGCA	YPCNRPRCKTCPIHHPTTTYSSPVTGISYPIKGRATCESSH				
				AATCCCT	GATCGGCC	VIYKLSCNHCAAFYVGMTSTQLSIRMNGHRQTVARRQL				
				CCCTCCA	ACAGCGGC	DHPVAEHAALHSVTHFNDCYTACAIWILPTNTSFSELRR	IPR0 00477	326		566
				CAATGA	CGGAGACG	WELSLQYILHSRNPPGLNLR				

RTE	RTE-47_LMi	.	Locusta migratoria	GCCATAG	TAATAATA	MKIIGTAELAEEATLQLNVQVVEDDAPLMAANMVRWY				
				CCGGTCC		RAAAGTSIELMHRANDASDGSRSKTHGRNQSCFRWIKI				
				CAAGCCC		EDSWAKPIIVWFFGHDTAMWKGNGKTTTLLYFLNVGR				
				GGTTGAG		RRIMMAFLGVKYSGDKLLPRSDLRVGSISGGVIKRDEHF				
				AAAGGAG		NLRIGTWNVRTLQRAGKLENLKREMDKCEVNIMGLSE				
				GAGGGGG		VRWTDSGEIVSGNYTMFYSGGKSFERGVAIVMGNELVQ				
				AGGAGTA		SLTKVVCHSDRLMFVKISAKPVDILIVQVYMPTSDHEDE				
				ACACCTC		EVEKMYDEIEDIIQSEGTGKVNTIIMGDWNSVVGQGNE				
				GTAAAAA		ANIVGQYGSKHSWTIWLGKRNDRGGMLVEFCQRNNLV				
				ACCTTGG		VMNTWFKKRKSKLYTWKSPGDMNRYQIDYILVKQRFR				
				TTCACCT		GSVKDAKTMPAADIDSDHNLLVADISTRMKKIQQHGRR				
				GGCCCGG		KQRWDLEKLRENNVAVRESLEKKFNEDEIRASAEEQW				
				GTGATAG		QNVKMILLNTLKSEVGRETKRARKPWITSEMLEKMEER				
				TACCTGT		RKCKNVNERNTEDRNRGTTKKRKVRPHVPENEGTGEG				
				GAGGGCC		ENKSVRTFGIEDSRGQVVTEQVEVLKTWEEYIRELYDTE				
				CCTTGCC		HRPHDIDIEPEEAIDEDEKGPSILTSEVEKAIKDMKSRKA				
				AGGGATA		TGDDDIPVDLLKDIGNDSLKALTQVINKIYETGEWPKDF				
				CGGTGAA		LDVTMVTLEKKKQAKKCSDYRTISLISHVAKIIARLLNR				
				GACCTCA		RLEHKIEEVMGEDQFGFRKGKGTRDAIGMMRILSERVL				
				ACGGCAG	ATAATAA	AVNEEICACFIDWQKAFDRVNWNILMNILKEIGINWRD	cd01650	655		900

Penelope	Penelope-38_LMi	.	Locusta migratoria	TGAAGGC GGAG	TGATGCGA CGGGTAGC ATACCTCGC AGGCGCAC ACCGAATTC GAAGCCAG GAGAGTAG CGTGGCCA CGATAACA TTCTGTAGT GGTCACCA AMAGCAGA GGAGATCG CCATGAAA GTGCTAGG AACGCGCA TGGCGGGC GCGGACCA CCTATGGA ACGCCTGG GACGGCCG CCACTAGA GAGCCGCA GGGCGCGG ACCGACGA GGGATCAC CAGCGTCA GATACCAC CGCCAGAC TGGAGCAC ATTGGGCG	CACTCCT AGGGAGC TACCAGT TTCCAGT TTCATCA GCTCTGT GGAGCAA GTAGTCA ACACTCT ACCATCA GTGTTGC AGAAGAG ATCGGAG AGAGACG CAGGGCG CTACCAG AGCAGGC CCCAAAA CAAACAT CTCAAGA GACGAGA GGTGGCG CTAG	RRLIRNLYLGQRVKVRLKCGETNSVEIGRGVRQGCCLSP SLFNLYGEMLAREALKGCGNFKVGGRLINTIKYADDLV VLAKNQRQLQHMMNRLVEMGRKYGMEINIEKSKVMRI SKREKHLKIIVDNRELENVVNQFKYLGCFITKDAHCTNEI RARIAMAKAAFNKKRTLLTSKLSLALRKKLVKCYIWSI ALYGAETWTMRKKEKKYLESFEMWCWRRIEKIKWTDR VTNDDVLRRVGEKKSILRTILQRKANWIGHILRHEGLIH DVIEGKINGTAGRGRRRIQHLDDLKDGRSYIRLKEEAED REHWHHKFSVRRHGPATRQNTT		cd00304	203	287

					CGGACCGCATAGAAAGCACCAGACGCCGCCTAGGCATAAATATGGGACGCGGTCGGCTTGTCGATCAGTCTCAGGGAACACCTGATGAAGACGCCGAGTTACGGTGTCGAAATATCGTGTTTGGAAGACGCGGATCACCGGCAGTACACCCGATTCCACGAGA				
Penelope	Penelope-32_LMi	.	Locusta migratoria	TG	TGACGTCACAGCCCGCAGTGCAGCTATATAAGGACGACAGCAGCAATCACTCGACAGTCAACCACTTGACAATGGCCGAGGAGATCTCGGCCGAAACGTCGTGGACAGTAAACCACATGACGCGGCTGGAAGCCCGAGAAAATTTTATCAACATAACTCGCGCGA	WNKHFARCLVNYRQIRRETSHSIQATKTEHDKSRTQWPSELRNDPLIVVLPADKGNAAVIMERATMTRRFERCWRRRHMQKLSRDPTSKILRKTSELLKRSSLEDSVIKNLLPQAPRPPTLYGLPKVHKMNTPLRPIVSTIGAPTYRLAKHLANLLTPHVGHCEHHIKNTADFISRIKCLRLGEEDMMVSFDVVSLFTNVPVKDCIDLLSQLFDDTVVDLFKHTLTSSYFLYGGEYFNQTDGVAMGSPLAPVMANLFMEHFEDIALDTAPVRPSCFYRYVDDTFVVWPMDVKNSVNSSTTSIASMTRSSSRWRWRKRVPCRSLTSSSAVKQMDVSAMVFTENPPTQTGTCTPQAIIIHRRNSLFHVRCLAHGREKLGEFFDHINSIHDKIKFTMEVEKEGALPFLDVLVSRKANGCLSHGVYRKPTHTDRYLHATSHHHPSQKQSVLRTLVHRAEVVSDTENLQKELSHLRKVFRDNGYSNSQITKAIFPKTQKQKTPEEDSKKLAFLPFCGSVTGKISRLLKRHKIDSVFRAPSKIRQLMRPVKDDVGLRTPGVYKIPCGCGQFYIGQTVRTIEQRCVEHERCYRLRYPEKSAIVEHALENRHRIAFEETSIITRTNSFWDSVIKEAIEIKIADNTLNRDGGLQLSTAWNPAIGRLKRARQARDQNIAFYGHVAEHQ	IPR000477	105	351

RTE	RTE-33_LMi	.	Locusta migratoria	GCCCTAG CTAAGTC TGGCATT GCTTCCA CTACTTG TGTCAGA CTTCTGCT TTCTTCGT TCCTATC GGCCTCT CTTGGTC AACTCTG CTGCTTC GACCCCG ACGGTAT TAGGTTG CGAGGCC GAGGGTG CTTTCATT TTCTTTTC TAACACA TGGCTGT TAACCGA ATGGCTC CCGGTAT AAGGAGA CCAACCT TACCAAG TGCCAAC GGCCAGT AAGGGGG GATGAGC GGTAAAG AAGGGGC ACCTCCT TGTCCAG GGATGAG AAATTGT	AACAATGC ATTCA	TGATGATG A	MRIGTWNVRTLLKAGKLEEVKMLMEKSQLDILGICETR WGGNGDFMSGDFRIIHSGSARSGRNGVAIILRGKWKNN VLNTYHCSDRIIMIKIKAQPMDLCVIAVYLPTTNSKDEE VEVVYEQIDDLMKLVKEKDNLIVIGDFNASIGEDKKGL WNGKFGLGKMNSRGERLLDFCEQYEMIVTNTCFEMPK RRRYTWKSPVDGERYQIYILVKKKFINQIKSSHSYPGYDI DSDHVLVIADYYNSKFKKYKRTNQHRWCVEKLKEKQVA ADMKELTDNLIGNREGWEGFKETVEVAAKELVGLRKL EPRKPWMTKEILDLIEERNRLRKKDFDKYKKRVKNAITQ KCRLEKEKWMHDKHKEIQQDLSGGQTDKAYSKIKTMQ TKTRTKTNMVKNKNGDLLFELEDVCERWKEYIEELYEG EDIANEGEILEEENSVDIDMIGPPISKEEFNSALNDLSNGK ATGTDNIPAEILKEIGDQQKSTYINFTFKEYYEDGKEPPT DLLNSKTVIIPKKEMLWTVKQEKTNLVSDLEKAKAILAL RMVLERRIGVNRKTYLTFIDLQKAFDMVVNLEGQRLIWN LYKRQEAEIEVNGVKKKAKIRRGVRQGCPLSPYLFNLFI EKAIEKLKRITKGIKINGERIHCIRFADDIVVLADSVRKW SLCCLSKEMEFMLLKFAKILREFNLKINKKKTKTMVVG KTKEKGKVNIKLEDDVLEQVENFCYLGSTITDDNKCTID IKRRIAWQSKLSKIKGIYMYHRYKEKNSLAKQAFQNKR NLLTDNHISLESRKKFAKTFVWSVLTYGCEAWTLGKAE KKRLEAMEMWIWRRMTKTSWVDRKSNEQVLREVNEN RTLLNYIGRRKSKMIGHIMRHNQFLKNVFEGKVLGKKP RGRPRATYFNNIKEDMGIKSYEELKRMTMERGTWLNR QGIAFSL	cd01650	533	722

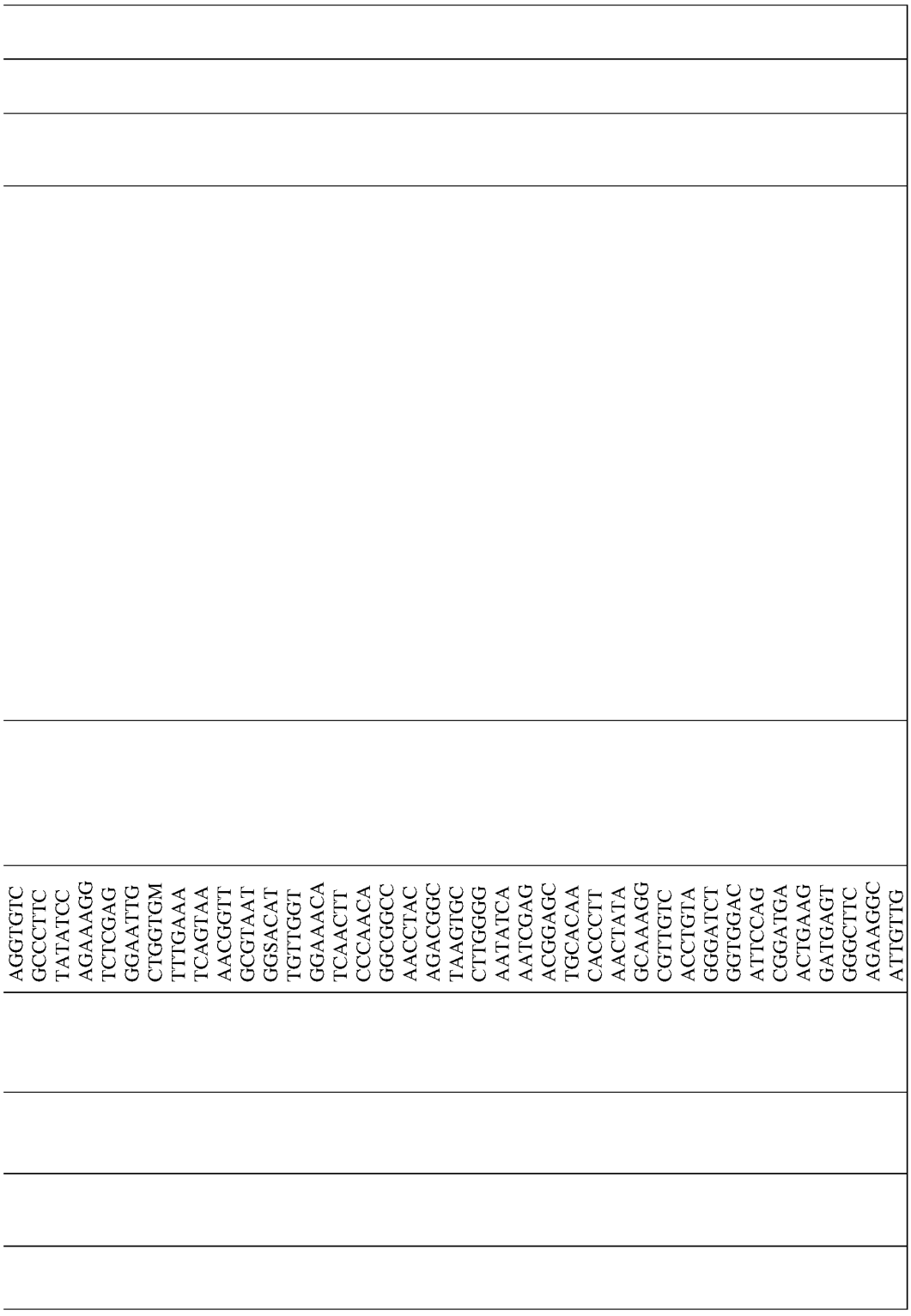
				CCCTAAA						
				GGCGGAG						
				GAACCAT						
				TAACATG						
				GTCAACG						
				GCAGGAG						
				GATGCAG						
				AAGGCAA						
				CGGGAAC						
				CACTGCA						
				TTAACGA						
				TCCCATA						
				CAGGATT						
				CCCAAAT						
				AGAAATA						
				TGGCTAC						
				GTCCAAT						
				TCAAAAT						
				TTTCCGG						
				AGTTTCA						
				ATTCCCC						
				CATTCGG						
				ATCTCCG						
				AGGGAAA						
				TGTCTTG						
				AGCAAGG						
				CTGTAAG						
				GAGTAAA						
				GGAATAA						
				AAGAACA						
				G						
				TTTAACA	TAGCCTTGT	ASVSRIPEDQADEIRQEVAYKLRRAPPPRSNIPSSERAAIR				
				GATGCCG	TTCGACCCA	SLREDKDLVILPADKGNATVLLSRGDYVSKMNLLLDDS				
				GAATGAA	TCGTTTTGC	AYRRLRNDPTERIKRKTFHFSHGTNKTEDLSLLKKSSLS				
				GGTATTG	CGATTTTCT	EDVVKKLHPGAAVPPRLYGLPKVHKHGIPLRPIVSNLGA				
				TTCCTAA	CCACCGGC	PTYNLAKYLASVLAPHVGKCEHHILNSSHFIRRLKSVCL				
				GTTTGCT	GCGCTACTA	GPSDLLVSFDVVSLFTRVPLEESLLLISDKLDDETVKLCR				
				AGAATAG	GTTCTATTG	HVLTSTYFLFNDAFFEQTDGVAMGSPLSPIVANLFMEDF				
				TTCACTA	CGCCTAGA	EDMALRTSFLKPTCFWRYVDDIFMVWPHGMDALNHFE				
				ATTGATT	GGGCCTCTA	DGKLPFLDVLVQKNSDGTLGHGVYRKPTHTDLYLHATS				
Pene lope		Penel ope-49_L Mi	.	Locusta migratori a			IPR0 00477	119		362

				CTAATGC ACAAGAT TAAGAAG CTGCCAG TCTTGCG CTAGTAC GTGAACG CTTCGTTT TACTCGT AGACAAC TGGATTT TGTTTCC AAAGACT TATATTTT TTGCATC TTGAGAT TTCATCA CGGCTTT CGGCCGC GTCATGG GAGTGGG TTGATAG CGCTTCTT GGGCCTT TTCCGAC TCGGCAC AGAGGGG TGCTGCC TGCAAAC AGTTATC AAATTTG AGCGACT TTGCACT CAACAGA CAATTCT GGTGACG ATTTCCG TCGGCGG GTTGTGT CAATCTT	CTTTGGTTT CGCGCACG CGCGGTGG GTCGTCAGC GGTGTATA AAGGGGCC GCTGTTTCT CGTTCTGCC TCAGTTTCC GGCGAGTA AGTGTATCA GCGCACTC ACCTGAAG ATGGCAGC CAGTTGGTC TGCCGAAA TATCGTGTC TAGTTGACG ACGGTATCC GGCTGCAC ACCCGAGA TGAATATTA TCA	CHPPHQRMGVLRTLVKRAYAISDAEHLGEELEHLKVVF KQNGYTAQQICRAFQFGPKGEPLEDNYKSIAFLPFVGNI SSKISRILRRFDIKSVFMPPPKTRALLGSVKDDLGLRKPG VYRIPCYCGKTYIGQTIRTVQDRCVEHHRHTRLLQPEKS AVAEHCLNEGHKMLYDETKLIAPASRYWECVLKESMEI RMADNIINRDKGFPLSKTWNPILSDIKQQRSSFRPAVSFN			

			ACAGATA AGGAACT GGACAAC GCTGCAT TATCGGT TTTGGAA AAAGGTT TAAATTT TGCACCC ACACCTA AGGTTGT CCCTATT ACGGAAT TTATTTGT GCTGTTG A		TAGTGGTTT TGTAATTTG TTTATATAT AATCTCTTT TGTTTTGTA TTATAGTTG TACTCTTGT CTTTGACCT TTGTATGC AACGCTTTT GACACGAA TAAATAAA T	MNKELTSKLHEYLRLFDEDPFSTDTIRSGMTLSQELAKA KQKRWSEMVESMDMSKSSRKAWQLLKRLSNDPTTPKE RFNVTPNEVAHQVLLNGKTEGRKEKIRITRHNLETNDLS AKFTLEELKAAIKMCKSGKAPGLDDILVEQIKHFGPVTM GWLLRFYNNCLTRRHIPKIWRRTRIIAILKPGKPASDAK NFRPISLLCHLYKVLERMLLNRLSPVIEPQLIPQQAGFRP KKNCTAQILNLTQNIEDGFERGLITGAVFVDLSAAYDTV QHKIMLHKLYQLTRDYDFTSFIGTLLQNRKFYVELDGK KSRWRIQKSGLPQGSVLAPTLFNIYSNDQPQPPNTKSFIY ADDLALVSQAKDFETVERHLTYGLKTLSAYYEKNHLKP NPAKTQVCAFHLRNKDANKKLKVIWNGTELEHCHNPK YLGVTLDRALTFRKHCLNIKQKVSARNNIIRKLTNSGW GAQPTVLRTTAIALCYSAAEYACPVWYTSVHARHVDV ALNETCRITTGCLRATPTHKLYNLAGIAPPDIRREVAANI ERXKVEQNAAHPLHGHQTPPSRLKSRKSFMKTTVNCST SPSTARQNLWTTRNADIATWIPPLEALPPGENLPWKTW KSLNRLRSGVGRTKDNMARWNYLGDESTLCECGSEQT TKHLYSCPQCPVSCTLEDLLQAKENAVELANFWSDCI MKKRVDGELVKSDSFILTFNSPRLRTMLRQASCVSQTPD HVKAGFLRLNVRPYVPNPMRCFKCQRFGHTTLGCKGR LPVANVVRQPMMGLTVLHLRXVSTVQGTTQSGVGIVQ CTXRRRKYRKLRPQGVSPMQRPKRSSRPCNLLYFPPPSL QYCKNQLLKLMPPPRLRLLVWAPTPALASARVRQQKL QLRSPTDNADKATGTAFAEPWFHNCRACHHVTATATTP	cd016 50	176	431
Ving i	Vingi-2_LMi	Locusta migratoria	AGAGGCT GCCGTAC CTCATAC ATTCCGG GC	CCCTCAG GGGCTCA GCATTAA TTTGCTG AGTACGG GCTTGGC	TGATGCTTG TGGCGTGC AAATCACT ATACGCCA CATTTTAAT TGAGTGTAT					
Nimb	Nimb-37_LMi	Locusta migratoria							cd016 50	865

GAACCCG	TTTATATCG	PTPPPHHSCNHNLCLHHQATLQSLPSRHRHRHRHNPTNATT	
GGGTCCT	TGCTGAGA	SPQLQPQPLSPPPSHLNPRRLPANTVTGKTVIXLSDFVEV	
GAGCTGG	GGGCAGCG	IPPEPSSDPAPELMELDLPQGQXSRPNTEPPPRSVSPSRK	
GGACTGG	GCTAATCTT	KERVKIQPPRSNLRDRWPQYCSGTCRVSGRMWRNSVIL	
TAAGCGC	GGTCGTGCC	TQGRPICICLQETNFKSTDNPELRGYVLHKKDDLSGERA	
CGCCAGT	CTACCCACT	RGGVAIFTNTAYHSEPLSLTTHLQAVAVSVMCVTVAPL	
CCTCTAG	GTTTTAGCT	SHHTFTSRCSICYVRHRLTVCSLYLPPHDTFDVAALEDL	
ACACCGT	GACGACGA	LTQLPQPFILCGDFNAHNVLWGSASTCPRGRATERLLLS	
AACCTCT	GGCGAGTG	SSAHLLNTGQPTHFCTATGSFSAIDLSLCSPALAHTTQW	
GGGCACG	TGACGAAC	EVVDDLHGSDHFPIWLHLPDRVEPECRPPQWVISKADW	
CTTCAGC	GTTTTACAA	TLFSQLAELETRDNVQEWVDCITQTIHSAAEASMPLSLG	
GACCACC	TTTTGTGAA	QPKRQPVPWWTAECRSAIRARRAALRRFKSRPTADNLA	
GTGTGGC	CGTTCCCGA	AFRMARAKCRQIIRESRKRSWQQFLSTVTRSTKSSTVWE	
GCGGCGG	CTCTGGCCT	TIRRISGKGGRCPVAAVMGNGVLHTSPRDIAQTMAIYFA	
TGGAACG	AAGCTCTTA	NVTATAHQDPAFQRHRVVAERGNLDFRSTGHQDYNCP	
TTGTGTG	GGCTGGAG	FTLWELDSALSVARDTSPGNDRIHYSMLRHLTQRNKEIL	
TACCGGG	GTTTTAGTG	LNLFNAIRPGHFPDSWREAILIPLLKPGKDRTCPSSYRSV	
GAATGGG	TGTTACAGG	ALTSCVGKTLERMVNRRLVWMLESRQLLSRFQCGFRR	
GATCTTG	AGGGCTCG	YRSTFDNLASLEAAIQQAFLRRHHLLGVFFDIEKAYDTT	
GCTTGAC	CTGATGCCT	WRHLILRQLHEWGFRGCLPIFIRSFLSPRYYFRYRVGGVLS	
CGCCCGG	CTAACATCA	DRFEQENGVPQGSVLSVTLFAIAINSITSVVKSPVQCSLF	
GTCGCGA	GCGATCAG	VDDFSVFCSSSSLATTTRQLQLTLRRLDEWAQKTGFKFS	
GGATGGT	CGTCCCTCT	TEKSVCVLFNRSRSIFNLPELRMRDTVLNFKDTVRFLGL	
ACAAACC	ATACTCCGA	TFDCKLTWLPHLKDLKRRCLKALSILKCLSHTSWGADR	
TCTATAA	CATTTTCTT	TRLLQFYRAFVRSRLDYGCTVYGSARPSYLKMLDVVH	
AAAACCC	TTAAGGACT	HEGIRLATGAFRTSPIPSLCAEAAEPPLSIRRQLLMVRQA	
TTCAATC	TGTAAGGTT	YAALSTPYTPAYNTVARPPMERLFRNRLRATRPFGIRAK	
CCAAGGT	GGATACTTC	DCFLDMGVSDFTVLRRGWSRYPPWLLQRPSLLLDLTSF	
GTGCTAC	AGTCTGCCA	KKDCTPDITFQSEFFNILNKHRDCTVVYTDGSKQGDSVG	
GTGCTGA	CCTTACATC	CSAVFPGHIVGIRLPAAYSVFSAELHAILQALEHIRRARG	
TGGGATG	GCTGATTTG	TKFLLCSDSLSALQSLLVMYPAETITQLIHDQLYLLQRR	
CATGGCT	CAATATGTG	GKEVSFLWVPGHVGIWGNEAADVAAKAACSDPNVTEC	
GTTGAAG	TTCGTGTTG	AIPLRDVYSVLRQRSLQLWEETWLAVTGNKLRLVKPTV	
GGGAACA	TTGCCAAG	RPWRTSCQPFRRFEVTLTRLRIGHSPLTHSFFIPSGGSTDL	
GTCGCTA	ACCTACATT		
GCGAGCA	CTATTGCCT		
ACCTCTG	GTTCTTATT		
GGGATAC	TATTTACTG		
TGCCGCA	TTGTCTAAT		
CCTCAGT	TCTTTTGTC		

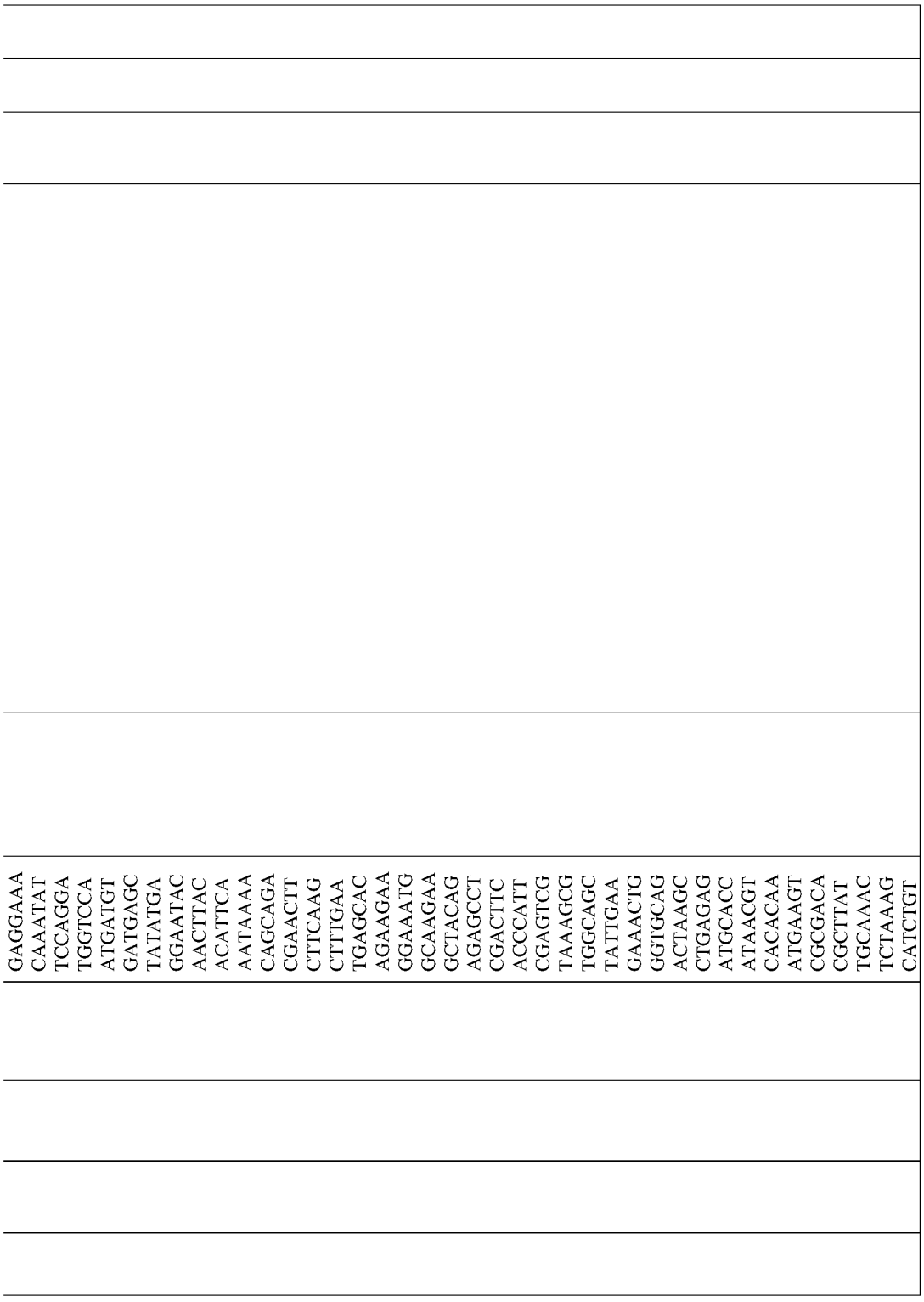
				TGTATAA GGCTTAC CTAGGCA TGCGGGG CTCTGTCT AGGTGGA CACTTAC TTCCCTA GCTACTC GTGGGAC CGAGATG GAAAAAT TCTAATTT ACCCATG AACCTCC TAGTGGG ATGGGTG TACCACT GGGGGGT ACCAACA CCCAACC AAMCAAG AGGGCTM GGGTAGC CAGTCCT CCGGATT CAGGTTC TCAAAAA AAAGTTT AGTTTAT AGTAACA GAACACA TGCTGCT MMACAG AATGTGT TTATTGTC GTGAAGA GAAAAGA GGGCTCG TTTGAGA	AGGGCGCT AAAAACCT TGCTGTTGT GCGCCCAA AACATCCTA TCCTATCCT A				


AGGTGTC
GCCCTTC
TATATCC
AGAAAGG
TCTCGAG
GGAATTG
CTGGTGM
TTTGAAA
TCAGTAA
AACGGTT
GCGTAAT
GGSACAT
TGTTGGT
GGAAACA
TCAACTT
CCCAACA
GGCGGCC
AACCTAC
AGACGGC
TAAGTGC
CTTGGGG
AATATCA
AATCGAG
ACGGAGC
TGCACAA
CACCCTT
AACTATA
GCAAAGG
CGTTGTC
ACCTGTA
GGGATCT
GGTGGAC
ATTCCAG
CGGATGA
ACTGAAG
GATGAGT
GGGCTTC
AGAAGGC
ATTGTTG

				ATGTGCAAAAC					
Penelope	Penelope-46_LMi	.	Locusta migratoria	GTTCGTAAGTTTCCCAGGAACAGAATGGAAACGCCCATTCTTGTAAAATCGAACCTCCAGCCAAACTGACAGCTGAACTCCGTGGCTTCTCTCCGATCCTCTCACAGTGTTCTGCCAGCGATAAGGGATGCTACCTGACGCCGATTATGATGGAAAACCGACAGATGTTGA	TGACGTCACTGGCGGCGGCGCGGCTATATAAGCACGGCAGCGGCCGCGAAATGACAGTCATCCACTTGACAATGGCAGAGGAGTTCTCTGTCGAAAGCTCGTGGGCAGTAAACCACATGACGCGGCTTGTAACCCGAGAAGATTTCATTAATGGATATCGCCGCGAAAACATGCACTCA	RNTNSELRKSELKDPGRYNYLGIVTIGLLQDLPNTCTLLSMHVGHCEHHIKNTADFIDRIKGRRLREEDIMVSFDVVSLFTCVPITDSIDLLAQLFDDTIVDLFKHTLTSSYFLYGGDYFNQTDGVAMGSPLAPVVANLFMEHFEEIALDTAPAKPSCFYRYVDDTFVIWPHGQKKLLLSLRRRYLRHLASWTEKIEEFLEHINSIHGHIKFTMEVEKDGSLPFLDVLVRRKPDGSLGHSVYRKPTHTDRYLHATSYHHPAQKQSVLKTLVHRAETVSDAXNLSNELSHLRKVFKDNGYSKRQISQAVSLKARNQKTSEEEVKKTAFLPFCGSTTGKISRLLKRHKIDTIFRPPAKIRQLMKPVKDDVGLRTPGIYKIPCECGQFYVGQTVRTIEQRRIEHERCFRLRYPEKSAVVEHALENGHRIAFDETSIIKRTNSFWDSVINEAVEIKISENTLNKDGGLQLSTAWDPVIAELKRARRTRDQNIALYGQGTGHQ	cd00304	126	229
Vingi	Vingi-45_LMi	.	Locusta migratoria	GACGAACAACATGTTACGGAATAGCAATGTATGTCAGGATGAACATAGCTGATTATAGAGTGGTTTCTAAGGACTCTACAGACAACATTTA	TAGTATTTTCTCTGGTTTGTTTGTTTGCTTGCTTTTATTATGCATATTTGTACATAGTTTTTGTATTTGTAATTCTTCTATATATGAAAATTTTCAAATAAGACTGTAGCATGAC	SPIQSHGFKPRAGGSPGKCAYKQSGCFDGIYPEFLKNAGPKTKRWLLQLYNIILTTGKLPTTFRAAKVIAVRKPGKEGSEASHYRPISLLSVVYKLLERMLLNRIQPLIDKATPLHQAGFRKNRSCCDQVLALTSFIEAGYQRKLKTAVAFVDLTAAFDTVWRHGLLLKFAEAIPCQKLVNLMDSMLSNRRFQVFLGDKGSRWRRLNDGLPQGSVLSPTLFNLYTSDMPNTSAKKIQFADDLALAYQARNLEECEGTLTTDLHIMGDYFKSWRLKPNPTKTEVNAYHLNNREANRKLNVVFENVPLEHNPYPKYLGVILDRALTFKNHLEYVAKKVNTRIDIVRKLAGSGWGANADTLRTATLALVYSTAEYCAPVWKNSAHVRKVDVKLHQAMRIISGTTKSTPLPWLPVLCNIPPPSIRREAAAIREFHKSSTHKNSLLYNQLRNIPTSRLKSRKPPWPMQDSPLVYDVCQEWKDVWKKFPPANSRLITDPTKRVN	cd01650	65	319

				TATCTTA GCAGTGA CGATATG TGGAACT ACTATTG TCAACAT CTATAAA CCCCCAT TAGTCAT CTGGCCT ACGCCCC CATTCAA ACTTTTTA CTGACCC TGCCATC TATGTTG GAGACTT TAATAGC CATCACC ATGCTTG GGGATAC AAGAATA ATGATAT AAATGGT AACATTT TACATGA CTGGCTG TCCTTGA ATAATTT GTATCTG GTCTTTA GTGCAAA GGACAAA GGAACGT TTCGATC AGCAAGA TGGCAGA AAGACAC ATCACCA GACCTAT	CATGTAAG CCATACGA ATAAATAA ATAAATAA	GFGLPRKLWTSLNRFRTGHGRCSACLYEWGYKDAPAC DCGAPVQTMQHLIEECPPRKFPGGLSELHLVTTRAVEW LRDLDVQL			

GTTTTAC
ACTACCA
GTCAAGC
TAATGGA
AATCAAC
TGGCAAA
CAGGAAA
GTTATAG
ATTCTTTT
CCACATA
GTCAACA
TACGCCC
ATTATCA
TTGACTA
TGGCATC
CAGATAC
CATTAGT
AAATTCA
ACTGCCA
AAGCTCG
TTGGAAT
TTTAAAA
AAGTGAA
CTGGCAA
CAGTATG
CTGACAA
ACTTGAT
GTGACTG
TTGACTC
ATCCCTG
TGCAATT
TAATAAC
TATGACA
GATTTGC
TAAAGCT
GTCAGAG
AGGTAGC
AAAAAGA
ACGATAC
CTAGAGG


GAGGAAA
CAAATAT
TCCAGGA
TGGTCCA
ATGATGT
GATGAGC
TATATGA
GGAATAC
AACTTAC
ACATTCA
AATAAAA
CAGCAGA
CGAACTT
CTTCAAG
CTTTGAA
TGAGCAC
AGAAGAA
GGAAATG
GCAAGAA
GCTACAG
AGAGCCT
CGACTTC
ACCCATT
CGAGTCG
TAAAGCG
TGGCAGC
TATTGAA
GAAACTG
GGTGCAG
ACTAAGC
CTGAGAG
ATGCACC
ATAACGT
CACACAA
ATGAAGT
CGCGACA
CGCTTAT
TGCAAAC
TCTAAAG
CATCTGT

				GAACAAG AACCGCA CTAGGCT GGTGAAG AACAGTT ACGATCC AAACGCT CTCAGCT AG				
				AAAACGG CCTTGAC TATGCAA AAATGAA GTTCGCA AACCAGA ATTGCTTT GACGTAG CTAACAT AGTTAGT AATAATT CATCAAA TGCTGTG TACATGA TAGTTAA TCTTAAA TAATAAG CTAAAGA ACCGATG CTTTAAT AACCAAA TCTGACA AAGGTTG TTCCTTG GTTATTG CTAGCAA TTCTGAA TATGTCT CTAAAAC CTTAGAA TTTTTTAA	TGATGTTGT ACTTCGTGC AGTGTCCG AAATATTTC TTCCGACTT GATCCAAA GCATTTGTA TTCCAAAA GCCAATTTT TAACTGTTT TGCTTATTA AAACTTTGT TACTCGTTT ATTCTCTCT ACTAATATT GTATTGCCA AAATGACG CCTCAGCCA TCTCTCATG TTAAACAA AGCATTTTG TAGCAATTG TTTTTTCAA TACTGTTTC TCTCTGTAC AGTAGCCT AGTTGTAAT TGCGGCTAC TAGATGGC ACTACTGA GGTAACGT	MNPQPPKLRSQFKVHKLNHPIRPISNSLNSAYHDLARFL HETLKKSFVFQNNYSVPNSHVLVHKVKDLDCGPDTKLF SLDIKNLYTNVPVVETLKIVEKNLCQFKKLSSEQITDFM NLINVVVKYNYYFEFNGKLYQQDDGLAMGNPLAGILADI FINSLEEKFFHKFSTTTLGIVSYSRYVDDILVIYEGPTTGI DHIFTLFNGLHENTIFTCELENEARQLNFLDLTLTVEENK ISFNIYRKETYTDQIVPASSMHPQCHKNAFFHSAIHRAISI PLSKEKFEEEINLIKTIAVNNEYDPAIVDNILKKKTAKRV TTLGTPIESDQKKKYFSIPFLGPISYQIQRLLRNKYNCNV AFSTNNNLKKNFIHNLKSTCSPLDNSGVYKIVCDTCSSY YIGQTGRAFKIRFKEHLLRKNGTNPQNSSFADHLLITGH APKAIHDLNILHTEKKGHRLDVLEELQIFKHFSRNDGLIL NEQLQLRNRSFFNGIKPLLGL	SSF5 6672	34	231

RTE	RTE-34_LMi.	Locusta migratoria	TGAAAAT AATATTT CGGAACT GCACGAG GACCCGA CCGCAGA ATTACAA AAAGAAG TCAGGTC GGCCGTC AGTGATG CAAAATA TTTGCTTA AGCCATT TCATAAA AAAACTC TCGTTAA T	TAGAGAA ACAGTAG GAAATGA AAAAGCC TGCATGA GGTGTCA	CTTCCTGTT CACAACTCC ACTTAGCAC ACGCCAGT CTGATGCCT CCGTGCAG GCGTGCAG CCCATAGTG GCTAGCTTT TTCTAAAAA TTTCTTTTT AAATTTTGT GACGTAAG CTTTACAGC TGAGTTGTA TATTTTATT ATGTGACTT GTTTTATTT TATAATTTC TGATGAAG GCAATCAT AGTCATTGC CGAAACCA GGTTAAAT AATTTTTAA TTTGCGACC GAGGGCTG CWTTTTTAT CCTTTGACG TTGTAACGG TCGCTGACA ATAGCAGC TATGTTCAA AATTTTA	TAGTGACTG AAATTAAG TGTAACAG ATTGAGAT GGCTGGGC CATGTAGA	MGNAVVVKVHGSRETQESTRCSPGTTFNQTDIGNHRG MHRLIAAGRLQGGDRSLYGKMEYTRDPEEGKALKTQC FQIEGKGAKPEASGGSTIKEALQETAEEVLGLEERVKRA DGTMKIVKRLWKKEMRQGRKGKEENDKQDNKIYYRK FMKKRKEADRICRRKKRELERDGLWSRGMDYGAREEY NDEEVRKFYKKVRELKKGFQPRTLQRRKASNHVLCRD	cd01650	345	632

RTE	RTE-44_LMi	.	Locusta migratoria	AAGACTA TAACTTT GCC	GCGAATGT CGGAGGAT CGAGCTGTT AAAAAGGT CTACAAGG AAATCCAG GTGAAAGA AGCAAAGG GAGGCCTC GGAAACGA TGGCTGGA GGATGTGG AAAACGAT TTGAGAAA GATGGGAG TAGGAGAT GGAGGAAA CGAGCAGA AAATCGTC AGGATTGG GCTGCGATC ATCAGAGA GGCCAAGG CCCTACATG GGCTGTAG CGCCAAGG AGTAAGTA AGTA	KEGNFIGGEEQVLDQWAKYFSELLNVTTEREDKMSLQK EKTNETEGDQPWQNVSQDERPVPIPTIEEVRKAIKSLRN NKAPGNDNISAEMIKGRGEHLTRLVHQLIVEIWEKEEMP DQWNMAIICPIHKKKDKANCENYRGIALVSVVYKVLAK VIAMRFAPITEEILRRLPVRLSAGQINNRPNFHYKANNG KVLDLHQLQKKYSGDYQCGFRQGRSTTDQIFIIRQIMEK CYEYNVDVHQLFVDFKQAYDSVKRKKLYQTLMEMEFP SKLIRLVQMTLTSTKCQVKIEGKLSKEFEVNQGLRQGD VLSTLLFNIVLERVMRRVEIDNPGGTLFNRMNQNLAYV DDVCMLSRRLKELEEGFDRLEKEAEQLGVNNTKTQYLF ASRKNAVVKEKYIKLNRGNYERCEKFQSWGAVTEDSR QIREEIKSRIAAGNRAYWAMIKILQSRSLSRKSKVTVYR TVIRPVVMYGSETWTMTVAEEELLRRWERKILRKIFGA VIEVGQWRIRRNKELEELYRNLT	MTIKCKNKMYTLVNAHAPTNIDNKKNPTKVQTFWEEL EDTIQKIPKENIILLGDFNAQLGKERKYKNIIGEYPAHKR TNNNGIRLVELCQVFNLKIMTTHFKKHPKKKMKTWKAP NQLLGEFQIDHIAISRREFKTIQNIKVQKGAKIDSDHYLV KSKIKFIPNSVKNNKKPTRKNFETNNIDADTKNIFKSKLQ GETWEEMKSQIMKAAEESLPTKIKRKNIWWNDKCNEA LDNRQTLWQKFKSNPTQINLERFREGRKIASYTIRNEKR LFEKLSLQQIQVHFTQHKSRDYYQSFKRHLNGYEPPSLN FRDKSGKLGTNDQENCNILAHYFKDLLNCKAPPEKPEF KTTEEIPAESHPPSLDEVRQAIKSLKNNKSSGEDGITAEL LKLGEEKVTLHLGTIIKEIWQTEKIPEDWRTALIHPLHKK GDRSDVNNYRGISLLSVPYKVLSKILLNRIEPVLDAQIGE	cd01650	419	678

				TTGCCAG CGGACTT GAGCGAA ATAAAAT AGCTCTC GCGGACC AAACACA AACCCTC TGGATTA ACACTCC AGTTGTA TCTCGAG CAATCCT CGGAAAG CTCCACC ACTCCAA CCTTACA AATKTTA TGGATTA TCAGTGG AAAAGAA AACAAAT TACCCCA GGGAGTA AAATCTC TCGACCT CCGGTCG CAAGCAG GCATTTC GGATTCT GGGGAGC CTGCACT TTCTGCTC TTGTTTTA CCAAGAC AACCTGA AAAACTT AAATTCA AGAMGA AATACAG		YQGGFRKGRSCSEQIINLKNIIRTRAGQPHVTTFIDFRKA YDSIDRESLIRTLEEFGIDRKTREIVKQTLTDTKSKVKFR GKISDPFKIQTGLRQGDGLSPILFNITLEKIIRETLAEYTQE KVKGISFGGSVKGLTIHALAFADDIALLSTNIXDAIKQIEI LAKHAAKFGLQISQEKTKFMSNKFKEVNKLHTIIGCIER VKHFKYLGETITDTGTEKEAVKLRVAKLRRAQILTMKL YNKKCLSRNAKLL			

				TATCGGAACACTGAACATTAACTCGCTCATCAAAACAGGAAAACATAAAGAACTGACTCAAGCACTTGAGCATCATGATATCCAAATTCTGGCACTACAAGAAACCCGCTACACTGACGAAAACTCTTTTGAAACAGGCAACTACAGAATTTATAAAGGAAAACCAGCAGAGTATATCTTTGGAAAATGCCCCTTATTAGGTACAGCTTTTATTGTTAGAAAAAATATAACGGAGTCCGTAATAAACTTCAATTCTACGTCCAACAGGCTCTCAACA					

CR1	CR1-2_LMi	.	Locusta migratoria	AGTCGCC		MNRDCDCCVRMQSELVTLRSQLQAVLTSVAQLEAAIEG	cd01650	852	1140
				GAGTGGA		RRCCWPVVGTERTPVVSTDRPAPATCPDTASSEADTPPV			
				CGTCGTC		VAWEVAPVRGRRRKTFRGTDREASPVRLANRFEALSVQ			
				CAGTGCA		DNVPESDGVSHPEVTTQPARSGLSQRWVCLRDGCQGGE			
				CACCTGT		EGQRALRVHTGGVIPDVERVLPDAMKGTGCSQLQVVA			
				CCAGCGC		HVGTNDVCRFGSEEILSGFERLSEVVKTASLACKMKAE			
				CAGCACA		LTICSIVDRTDCCGPLVQSRVEGLNQRLRRFCDRVGCRFL			
				TTTTCAGT		DLRHRVVGFRVPLNRSGVHYTQEAATRVAGAVWRGLG			
				TGTAAGT		GFLGVRSPLYAGSGYTGSGGCVAWTGRFFRIKSFGEIEN			
				GTAACCG		GQSQRVQVDHRRNVDPETIGITVVNCRSCVGKVPELQA			
				GCTTCTCT		LIESTETHIIIGSESWLKPEISSEEVFSKDLTVFRKDRLNKI			
				CTAGAAG		GGGVFVAVKSSISCSEVEVDSSCELLWVEVVPENRKKLI			
				CTACATA		IGSFYRPPDSDDTIAERFKENLSLITKTYPTHTIIVGGDFN			
				TTTTGACT		LPSICWQNYIFKSGGRHKTSSENALNAFSENYFEQLVHE			
				TTTCAGT		PTRIVNGCENTLDLLATNNPELIMSIKTDTGISEHRAVVA			
				GCGCTTT		KLNIESPKVTKNKRKVYLFRKADRNKIHCFLKDSFHSFL			
				CCCTTCA		TLNVGVDQMWLKFKEVVSTAIESFIPNKLIKEGGDPPW			
				GTTCATT		YTKQVRTLLQKQRKRHTKFKRTRNPQDWRSYTEARNL			
				AAAGTTA		ARTSMRDAYNTYHNETLSRNLTENPKRFWSYVKYTRG			
				TTGTGTG		KTQTIPSLRDSNGDPADDSATKAELLNIAFRNSFTKEDET			
				CTCGCGT		NIPDFELRTEARMSNVEVDIIGVVKQLKSLNKNKSSGPD			
				TTTTATTT		CIPIRFLSDYAEAIAPYLTAIYNRSLKEGSVPKDWKVAQ			
				GATTCAT		VTPIFKKGSRSNPLNYRPISLTSICSRVLEHILYSNIMNYL			
				ATTCACG		EQNDLLTQGQHGFRKYRSCETQLALYSHEVMSAIDKGS			
				TTAAATA		QIDSVFLDFQKAFDTVPHKRLLIKLRGYGISPQICDWIGD			
				TCTGGTT	TAAATAGT	FLSERTQFVVIDGKASTKTEVISGVPQGSVIGPLLFLIYIN			
				AGTGTAT	ATAAGGTG	DLGEICYWSSAVPYLHKRFRGNLSSSLRLFADDAVIYRL			
				TATTCCG	GTTCGATGA	VKSSEDQNQLQNDLNKISLWCKNWQLTLNNEKCEVVH			
				CCGTGTG	ACCCTCTGC	MSTKRNPLNFVYKINQSKSRPHEYEKKSVKLCLQDKSIK			
				AATCGCG	CAGGCACTT	VKAVNSTKYLGITITSDLNWKEHVDNVVGKANQRLRFI			
				TCGCGAC	AAATGTGA	GRTLKRCNRTTKETAYTTLVRPFLEYCCPVWDPYQVGL			
				AGTCGTC	TTTGCAGAG	TEDIERVQRRAARFVLSRYRGDSVTEMIQDLGWESLQQ			
				TGTCTTTG	TATCCCATG	RRLSLRRGLLTKLQSPTFSSECENILLTPTYIGRNDHQNK			
				TTCTGAT	TAGATGTA	IREIRARTERFRCSFFPRAIREWNGR			
				CGTCTGC	GATGTAGA				
				GCGTTCT					

				CGTTTCAGTTTGTAAATTTATACTGTGCTCTCGTGCATATTAGATTCAAATCCGCGTTTTAACAAACAGTGTAGTCTTCCTTTGAGCGCATCGCGACTGTAGTTTTTGTTTTCTCGGCACTGTAATC					
RTE	RTE-14_LMi	.	Locusta migratoria	CTCTGAAAGTACTGGAGATTCTAGGAAACGCCAATACGGTCCTAGAATCAATGAAAGGAAAATCCTGACATTCACCTCCACGGGGTAGAGAGGGGGCAACGTTTGCATAGCAAGCGGCTAACGAATGCAGGGACCGAATATCCCCCGGGATAAG	TGAGAACGAGAGAAGAAGA	MMTGFVTKQSGGKIVPNSDLRQGTVSKCVIKDDRRQVRIGTWNIRSLSICGKLENVKKEMEEHDIDILGISEIRWPQKGDFWSGNFRIIYSGKEGSYAGVGIILRKHVGTRVKSYVQYSDRIIMLRIENDPIDTVIIQVYMPTSSHADEEVEQVYEQLEEIMEMVKDKDNLVILGDWNASVGEGEDGNTVGKYGLGRRNDRGQQLVDFCKEKELIIANTWFQQPKRRRYTWIAPGDRARYQIDYIMVRRRYRNQAKQCKSYPGADVATDHNLVMLKSNLRLRRKGQNTKGYCKYDVKKLRQDSVKERYQKQAQQKVKEIPNGTVNEKWIGLKHSILQAADREIGNIKKQATQPWISDHTLELMEERKKFKNSKSEEGIKSYKRLRNMVNREAKKDKEEYIMRRCKEVEEKLEKGNIETAYELIKKQFKKVNGRSNNILDENGKRLFENKEVGERWKRYIEKLYEDPDEISNLENEETVTKDQLGPSLTKEEYLKAVRELKNNKAPGCDALTGEMIQGLDNETHNYIFKLIQDMYETGEIAEDFKKTIIIPIPKKPGTIKCEEHRTLSLISHMSKILTKIIYNRIISKIEQNLDEDQFGFRRNRGTREAILMLRQIMDKMLDLKRPLFIAFVDLEKAFDNVNWNLLMEIVTNVGLDFRDRRIVHELYKNQRGVINIHGEIQEAHIKKGVRQGCNLSPVLFNLYIEEGIKECKESMEGIKINGVQIKMIRFADDIAIIEESEENLQKALSRMNEILPAKYNMKINTKKTEIMVCSHEQITTDIKINDVPLKQVNVFTYLGSKITPDGKSSADIKSRIAQAKLAFNNKRMLLTSRNIKEETKKAFIKCYVWSVALYG	cd01650	545	801

Dap hne	Daph ne-8_LM i	.	Locusta migratoria	GGGATCC CCTATAC CAAGTGC CAAAGGC CTCCATG AGGGCGG ATGAGCG GGTATAG GATAGGG CACTCTC TTGTCCTT GGAGTGT GAAATCG CTCCTAA AGGCGGA AGAACCA GCAAGGT CAACGGC ATAGAAT GGAGAAG GCAAGAG GGAAACC ACTCCAT TAAAGAT CCTTTTG GATATCC CATGTCA AGCAGCA	TAAGCTAAT TCATTTTTT TCTTTATTG ATTCTTTTTC ACTTTGTCT TAAGATTGT CGTGTGCAT AGGACTAC TTAAAAATT TTCTATTAC CATTCTTGT GTAAAAAT	SETWTLGVCDKRRLEAFELWCWRRMQKIPWTAKMTN EEVLRRIGEKRALVKTIMKRRDKLIGHLYRHNEFIINVM EGMVEGKRGRGRPRLQYLKQIQRDIGVESYQEMKRLSE NREVWRRRIAANQSSD	MIEEQERLICALQVQLKNAEEELGRLRRGRGAGEWELA VGKKVARRRYSDSFTLQVANRFDQLSELSGDEPPVAVG VGSTQQTSAFGKPKSVERSLRKKKVLLLGSSHGRGVGQ QLQEVLGSEYQVTSIVKPSAGLDQVTCNIGDLCRNFSKE DQVVIVGGAGNSLDRDGEYDIGGDLERIATQTGGTNVH FVQLFQRHDRPCLNTAVSRVNMGLEKALMTEGMDHIA VVPVETIQRSDFTRHGLHLNRYGKGRLAKLIGDSVVGG GGXTHGKIPVVVGVRAAPFLDGCTFFRLKTADRYSCLK EVPLRXEPPLTKSKNISNKGITVFHQNIQGIRDKVNEXLI DTDSDIIGISEHFLSGVTIQRLPLPGYRLAAFYSRSSLQCG GVAMYVKKRFFIGVNRFFKNDFSSESIDFSKHCTEQIFEC CAGVVNVNGTKLVIVVIYRSPDSDFKMFLPKLELVLGSL	cd01650	851	111 2

ATTGTTA	TAATGTATT	YRKYKNLIVCGDFNINCVGDSSRKSQLLDLFNSYNMMQ	
TTTCGTG	GACTTGTTC	TVFFPTRVQGNSSSTIDNIFVHSSLLEGHSVSRIVNGLSD	
ATAATAT	CATGACCAT	HDAQLLILKSICSQTNVKHKYQLLRKGKVFVLKQMLNI	
TCAATAT	GGAGTCTC	NINYSEKVNPVAIETFLSLIKEQEWQDVYSADNADDKY	
AAAGGTG	ACTCCTCAA	NAFLNTFIMLFESCFPLERSKQGTSSNRQPGWLTKGIRVS	
TTCTAGA	TTGGGTCCT	CRTKRELYQNVRSSQNEDAVAHYKQYCKVLKNVIKNA	
AAAGGAT	ATGGAACA	KSMWYAKRIANSRDKIKTIWSVVKEVSGQQQQVEDITS	
CGTCAGT	GAAAAAAC	VRSRNVSVTDKLDICAIFNNHFLNIVGELNKTLVSTGNHI	
AATTTTA	AATAAATA	TFLENAFPRLLPEMVLHDTSIGEIESIIKSLKTKDSCGYDG	
GCGTTTG	AATAAATA	VSSRILKYCSPYVSPVLSHICNFSFQNGQFPDRLKYSVVK	
TAGTGCC		PLYKKGERDNVDNYRPISMPSVFSKVIEKAVYSRLIDHF	
GTATTGC		SSHNLLSNVQFGFRSGLTTENAIFSFLCEALDGLNKRLRT	
TTTAAGT		LGVFFDLTKAFDCVDHKILLQKLEHYGIRGVAHNWFTS	
GGTGTTA		YFNNRQQKVILRSVENGYDVQSQWGTIKCGVPQGSVLG	
TTTGTGTT		PLLFLIYINDMPSSITGDCKIFLFADDTSLVVKDVARNID	
CGCGTGT		TVSNNAVQEISSWLVENKLMLNHSKTQFLQFLTHNLTK	
GCGGCAC		ADILITHNGHMISETEQFKFLGVHIDSKLSWKAHVQDLV	
CGACGTC		PKLNSAIFTIRTISEISDFSTRKVVYYAYFHSLMTYGIVF	
TTACTGC		WGNSPDAKRIFLAQKRAVRAISGVSSRTSCRPLFSSLGIL	
TTTGTTTA		TLASQYVFSLMSFVVNNISLFPRVSSFHSVNTRQKSNLH	
CATCTGG		LHHTTLTLVQKGVQYSAASIFNKLPQELKNLSTNPHAFK	
TGGCGAG		SKLKSYLMAHSFYSVEELLEKLNN	
TGGTTCC			
AGCCGTA			
GCGGTCT			
CCGAAAG			
CTAAGTT			
CCAATAA			
GTGTTTG			
TGTTATT			
GTAGTTA			
AATTATT			
AAAGTGA			
CTAGATT			
CAGGAGG			
TGTTATC			
ACGCTAG			
TTGTTTAT			
TAAATTT			
AGTACTT			


TTCAACA
GGTGACT
TGCTGTTT
TTAACGT
AGTTGAG
GTTGGGA
AACTTTA
AGTTACG
GTATTTG
ATAACTC
TTKTACT
CTCACTG
TGTTTAG
TTAGTTT
ATTAGTG
ACGCTTG
AATTTGA
ACTTGTTT
ATACAAG
CAATAGT
GTAAAGT
TTCATGT
GGTAGTG
TTATTTGT
TCTTTTCT
GATACTA
TTAGTTG
TTGTTACT
TGATTTT
GTAGAGA
CTGAGGT
TTGTTTGA
TTGCTGT
AGTTAAG
AGGTACT
TCCTAGT
TCTTTTGT
AGAAACG
TAACTTA
ATAACTT

Jockey	Jockey-7_LMi	.	Locusta migratoria	AGACGTC TAGTTTTT TAAATAT TGTGAAA TTTTACC ATGAGTG AAAAATG TGGGCTC TGTCGTA GGGTCGT GCGTAGT GGGCTGC AGTGTGG GATTTGT AGTAAGT ATTTTCAT GGGGGA GTGCAGT GGGGAAG ACAGTGG GCATTGT GGTGAGG TTCTCTCC TGGAACT GTAGGAT TTGCAGC AGGAATA AG	TGACCTCTC TCTGTAGAT TGCCGCCGC CTCCTCCCT CCCCCTCCC GTCCCCCGC TTGGCTCAC CCCCTCCTT TCTCCACCT CCCCGGGA GCTCCTTCC TTCCGGTTT	PPSVVPIMAGSPHRILYHNIRSLPANKPLFMHTLSQHCV DAFLLNETFLQPRHSVSTTPYTLHRTDGPPPLARGGVAI GHLRHHPVRPQPPVNNPAEHLILTLFLPSLTVTLATIYIR ARVPIPYDFISHIDNSFTTYVIAADLNLHSRSAAELQRWH QFLPTLRGDLVPLPSPTRPESNTTPDAVLVSPNLWDACR RKSLTSLAVTTPPYSSPSTPAAVPAAPPVLCPLLGPLGRL QAEVLDLIGSDHAPVLLTFNPRGRPGRATRPVPPPRLVH DYSRADWDAYRRSINTSVASHPLSLHTPDDIVHASSFLH QTVADAVATHIPTKAIHPHRPPLPPQAVLLLRESRRLYR SFLRTRDRDTLLRHRQLQRHIRNLLNAKKRRDWRQTCT RLNATLPINSSKYWKAFHRLTGSHSVPAYPLLHNNLPLP DNASKAQLFADHLSNIFDLPDNPRFDYSRFPDVHERTDT	cd01650	551	810

					CCCCCATTC TCCCCTCCC CCACCTCCC TTCCCCTTC CCCCGCGCC TTCTCCATC CTCCCCCCT CCCCCTGCA GGGCCAGT ACAGTGCT AGTGTTTTT AAGTGTTCA GTGCTTCGT GTGTATAGT GTTTCGTCG CCCAGTGTT TTTTAACGT TGCGTACAC GCCTGTGG GAAACCAG TGTTCAGTG TTTTTTACC TCAATGTCG CCGTCATGT TATGTCTCG TCGTCTGCA GTGAGTGTT TTTTAGCCA CCGTGGCA ATGTCTTGT TTTAACAAC CATGTATGT TCTGTCCAC TGTCATGTT GTCTTTTAA CATTCTCCC ATATTTTCA TCTTGTAAT TGTAAAAG CGCCATTTT	SVPPLAPGFQYLDHIPPSDLNTPITAQDIVHTLRTKRNTA PGHDRITYRHLRACPPSFHSALASLYNVILFTGFYPDLW KTSRILMFLKPDKPPSEASSYRPICLTSVFSKVLESILSRRI HEHLRLHQLLPLSQCGFRPSFSVDDQLLRLVHLLSVQVN SRRAAILVSLDIEKAYDRVWHSGLLFKLQTYALPINYVR LIASFLSKRPSYVTVNNQNSRLFYPSAGVPQGSVLSPLLY LLYSADMPRPQPPVHLLQYADDTAFLALAPSLARSQHL LQSHLSEFSSWCNQWLLTVNPTKSQAIIIGRTTPSFRHRD LYLTLHGRPIPQTKTLRYLGLTLDRDLSWNPHLKTVRQ KALSRLNLLKCLSGRTWGLNPSTILYTYRSLMRPVLCCF HPAWISAPPSYYESLKILERHALRLAYRIRLPSPTRILYDL IPFPHLVQFLQHIRLKYIHRKIESHNPLVSPLLLDPGPLPR LYRHIPPSLHLHTLYVLSEGGFRHLPLPDDVLHNRPDIYP SFQL			

Penelope	Penelope-9_LMi	.	Locusta migratoria	CACACAT	TAGTGATGT	MIGERKALKELNADNDIVVLPADKGNATIVLNKDDYES	cd00304	206	298
				GTAGTGA	TTTATCTCT	KILDLLTSGQYRKLKKDPTASVLKKTNTLIKSSTSIPKED			
				TGACTGA	GACAAGGA	KKALLKSEAMCPRLYGLPKVHKINIPLRPIVSAINSPTYEI			
				CACTGTT	CTATCTCTG	ARYLATCLQPYIGKTDSYIKDSSHFIGKLEGLTLRPEDM			
				ATCAATA	CTGATAAC	LVSFDVVSLFTMIPVNEVMTFITDIFPEDLTTLFKHCLTSS			
				AGTCAAA	GTGACAGC	QFQWGGDFYEQVDGVAMGNPLGPAIANFYMEKFEQSA			
				AAGAGTA	TTGGCCACG	LGIAKKKPVLWYRYVVDDTFVVWPHGKEALDEFFSHLN			
				TTGCTCA	CCCACTTTT	NINKKIQFTMEMEKDSAISFLDVLVMREVDGKLKHKVF			
				AGACGAG	CACAGTAT	RKATHTDRYLHKNSNHHPQQKRGVIKSLVDRAKKICTP			
				ATTTCCG	ATATGGCG	EHLDAELKHLKQAFKKNGYSDKEINRVLRPNNNSLRTR			
				TCCTTAC	CTCTCAGAC	DQTKRWKNTVSLPFIGKVTDQIGKILRKHNIRPVFRPTK			
				AAAAGGT	ACCCGACTC	KICQVLRSVKDKRPTLSACGVYKIPCTCGKVYIGTTKRSI			
				GGAAACT	ATCAGTCG	ETRLKEHKSLCRLGKTEKSAVAEHALQSGSHIVRFSETS			
				ATGCTAT	GCAAGACT	VLSTMTNYYPRQYREAIEIYKHKENFNRKEEAMKLSDI			
				AACGCCT	CAACGCAA	WTVALKDRQVLSLTDCNR			
				GTGGAAG	GTAAGAAG				
				TACAGTT	ACCTCCGA				
				GAGGATA	AGATGTCC				
				TTATTGC	GGCGCAGC				
				GAATGTG	TCTGGACG				
				AAGCAGG	AAACGTTA				
				AATAAGA	GGAGAAAC				
				ACTCTCC	TGAAAAGT				
				CGAAAGA	TTTATGGAC				
				AACTGCA	CACGGCCTT				
				AATGCAA	ATAACCCG				
					ATCACCTTT				
					AATTTTATA				
					TTCTGTTTT				
					GTCCATTGT				
					TTACCATTC				
					GCTCGGCTG				
					TAGAGCAG				
					CATCATCGG				
					CTGCTGCCA				
					GCCTGCCTG				
					TATTAGGGC				
					ATTAAAATT				
					CAATAAAG				
					GAAAAAAA				

RTE	RTE-25_LMi	.	Locusta migratoria	TTCGGAT TGAAACC ACCAGAG TACTATG TCGCAGT AACCTCT GAAAAGT AATTTG	GAAGTTCTA CCAGCAAC TATGATCCG GTCGTGAA AGCCTTCAT GCA					
				CCCGTGT		MPRKNWNCGRRTGDEKRKMMFGCWNVQGISTKIDLLP	cd01650	536	788
				GGAGTTT		AELDMFNIDVVVLSETKRKGKGEEELDNYVHIWSGVSK			
				GCTGGTC		AVRAKAGVSIMIQKKWKKRITNWTFINERIITVEMTLFA			
				TTCCATC		REVVIIGVYAPTNDTKDKEKDAFWDTLRETIEKIPRRKE			
				GGGCACC		LIIMGDMNGRVGIRESCKIVGKHGEAEYNDNGERLIDIC			
				TCCCCAG		AQFDLKITNTFFKHKDIHKYTWQQNTKELRSIIDYIIIRQT			
				GTGGCGG		SSFKAADVRSYRGAQCGSDHYLVKMKSFWPWKNATN			
				ATAGGGG		DTSNINKMNCSEKVQNVHFNIDSLQDESIRTFFKARMER			
				AATGCTC		TLDESFEGSTEEIYEYIKTKVKNVASEVLGIKENNPKRA			
				ACCAGAT		AEWWSEEIETSVKEKRNAFVQWLNDKSEGTRSKYKEK			
				ATGGTGG		KNEVEKKIRLAKNEAWERTCANVNSKLGFGRAKEAWS			
				GTACCGG		VLKALRQDTKGKSNLQLVTQKEWEEYFKKLLNEDRDE			
				GGAAATA		YLEEGTVEENEHHDDEIXISESEVLQVLRTGKNGKSPGP			
				AAATACC	TAGTGTAA	GNINMEFLKYGGDKIVKLILQLFNKMIHGDSVPKEMKL			
				CGGGGTG	AACCTTATG	GYISTIFKKGDRKICSNYRGICVTNTLMRIFGKIIKNKLEK			
				GACCAAA	TACTAGGTG	NFRTQQEQCGFTAGRSCVDHIFTLRQILEKHREKSKNVG			
				ACCAGCA	TATTCATTT	LIFIDLEKAYDTVPRKLLWRALHRANINTSLIKIIEQMYK			
				ACTGCTG	CTGGGCGT	DNICQVKIGNTLSQKFRTSKGLLQGCPMSPTLFKIYIDIC			
				CCTTGTA	ATTAGTATG	LRTWSQKCNSMGLEIRDGVYLHHLLFADDQVVIAQDG			
				GTATGAT	TTGGAGGT	EDANYMCNQLAIAYKNWGLKINYQKTEYLTNDPHELRI			
				ATTGGCT	ATAATAAA	EGKKIKKVNTFCYLGSILETEGKSDSEINKRISSGRKVIG			
				TATCAAA	AAACCTCTG	MLNSVLWSRNVMNRTKKIIYKSIFESAVLYGAETWTIN			
				GGCTAAA	TAATGAGG	QKHTKKLQALEMDFWRRSARISRKEKRRNTEVIKRMEII			
				GGAAGAA	ACAATCTCA	ERIDEVMDRKKLRWYGHVRRMEDTRIPKLVLEWQPEG			
				AACCTTG	ATAATAAA	RRRRGRPVTTWIKNVQLTMNRLGAEEEDTQDRHTWRN			
				AATACAA	ATAAAATA	IVNN			
				ATTACCT	AA				
				GGTCCTC					
				CAGGTTG					
				GGGGTTG					
				TGCAGTG					
				GGCCAGC					

			TCCTCAC TCACATA AAAATAT AAAATGC TAAAAAA CCTAATA AT		TAAACAAT CCCACTGCC AATCATAG ATAATAGC GCAACAAA ATTATCAAA GATAAAAT ATCCCAAC ATATTTCGT TCAATTTTA AATAAAAG CTTTGCCTC				
					ATAGATAT				
					ATATTCTCT				
					GTGCAAAT				
					AATGTAAG				
					CATATTACA				
					TCTCTGTAG				
					AGCTGTCA				
					AGTAAATG	RRHLVLINKQDYIQKVFEFFDKNGIIKIPCDPTNKYIAQT			
					TCAAAATCC	RKIIKSCSSILSDKDKRKYIPMNPKTPQLRGHPKTHKELS			
					GTAACTTGA	PIRPVVDSINCPTYLISKKLNTLLQKYYTYKKKYTIKNSR			
					TAAAGATCT	TLAANINKIDIPINGKFVSFDIINLFTNIPVKETINIIELNLL			
					TGCATGAC	SXNKLPINSTYEIINLLHLVLEQNYFKFEDNFYTQPDGVA			
					AGCACAAT	MGCNLASTIADIFLNTLEERFFLHQETNKIIYYYRYVDDI			
					ATTATAAGT	IVLIDGNETDINNIHGLINKLHPKIKFTMELETNKTINHLD			
					TACGTGATA	LTIQNVQNKHRFEIYRKPTTTTNIIHAHSCHPSSHKLAFF			
					TATAATACA	HSMINRLINVPLDDAAKQKEMTILQYIAQQNEYDPNIIL			
					ACATAACA	KLYNNKTSPRPPTTRTEITLQRETSKLNKPIYISTPYLGKI			
					GAATGTGG	SNKVQQIFRKTNTTFAFQNRHTLKHKFVTHTTKPSKFLN			
					TATGTAGCA	SGIYKISCNSCNKYYIGQTGRNFTIRYKEHTRESEKNPST			
					TTTCTTTTA	FYQHLQSENHTVDTIDNALEILHTVQKGYEMNILEEME	IPR0		
Pene lope	Penel ope- 55_L Mi	.	Locusta migratori a	AATAAAG	CATCTGTAT	VCLHAIKNTNNLLNDQTDFKHKNYFDNFIHILVHDNG	00477	124	263

RTE	RTE-5_LMi	.	Locusta migratoria	CCCTCGG AATACCA CCTTAAC GTTGGTG GGGAGGT TTGTGTG TGTCAGT GATCCAC TGAGCTA TGCTGGC GGGAGCT GTGCTCC CGGTAGG GTCACCC ATGCCAG ACAGGTC AGTTGGT GAGATTC CAGACAA AGAGTAT TCCCCCT	AAAACAAG TAGGCAGC ATAGTATA AATATTATG AAAACTTAT GTATTGCAG GCATAACA CCTGAAGA TGACTATTA AGTCGAAA TAAGCTTAT MGTGTAAC TAATACGA ATTAAAAT ATACAGTGT TGAAATTAT ATAATTGTT ATTATTAAA A TAGGAGAT ACAGGAAA TAAGTCAA GTCAAG	MNRSRMDKKDLTFSMKSDTRIGMWNVQTMWQAGRLA EISREMKRYNIQGDEKILGLSETRWTGFGEIKTQEGDTL LYSGREEEHRGGVGLLITQKTKKSLIEWYPVSERIVVAR FASRVRPITVVQCYAPTEAGSEEEKEQFYGQLHRTLLKA KKKDIMVVMGDLNAKVGNDNQGVESIMGKHGIGNRN DNGERLIEFCSNHRLVIGGTIFPHKNCHKVTWVSPDHRT ENQIDHFTISKTWRTSLLDVRNRRGADVGSDHHLVIATL RLKIASTRNTQQKRRIKYNWDKLKNKEVQQCFSLELSN RFQQLREDEEQDIERNWKSIESTYIDTCEKVLGKRERGR EEWTTSETWEAVNQRKECKKRLNTARTQEEKVQARTEF NELNKMAKRRFRRDKRAFINNLAAQAEMAAQKNDSKT LYDITKRLTGKCFRSEKPVKAKDGKLLTSQKEQLERWR EHFLEILNRQTVEEEVTEAATEEEQEDVDISIEPPSRIEIQ RAIQQLKNGKAAGMDNITPEALKVDVSTSVEILFPLFKK IWEEEKIPTDWKKGLLIKLPKKGDLSECDNWRGITLLSV PSKVLSRVILERIKKRVVNSRLRKVQAGFREGYSCTDQIN TLRIILEQSTELQSSLYLLFVDFEKAFDSVMRKKMWKA MQEFGIPKKIIKITQEMYENYTCQVLHQGMLSEPIPVDT GVKQGCLLSPILFLMVLDLVMQKAMNRRRGIQWGLTG RLEDLDFADDLCLLASSYKDMEAMLNDLKTEASEVGL KINGKKTKEMRCNNRNEALFFFKPRDRASQAVLLPGKF	cd01650	550	782

				AGGGGGC TCCTGCC TTGGGAG TATGTGG GTGGCAA CCACATG GGCCCTT AGCTAAG TCTGGCA TTACTTCC ACTTACT TGTGTCA GGCTTCC TCTGCAT TCTTTCCT ATCGGCC TCCCTTG GCCAACT CTTGATTT TTCCGAC CCCGACG GTATTAG GTTTCGA GGCCCGA GGGAGTC TTTCATTT TCCTTCTT CTACTCTT ATTGTCA ACTGCTG CCTTGTA GTTAGAA GTGGGAA CTTCAAA GGCTAAG GGAAGAA ACCCCGA CAGAAAA TCAGTAG GCCTGGC		SHPGGRTIEDIGNRINKAKGAFAMLRPIWKSRELNSYTK LRIFNTNVKSVLLYGCESWKTNKQLISRLQIFVNRSLRNI LRIWWPQVISNEELWRRTHQTPIDIEIRRRKWKWIGHIL RRGKDNVARQALEWNPQGSRRRGRPRNTWRRSVVAE GTEKGGKAWTEIKTLAQNRVRWRKFVDALCST			

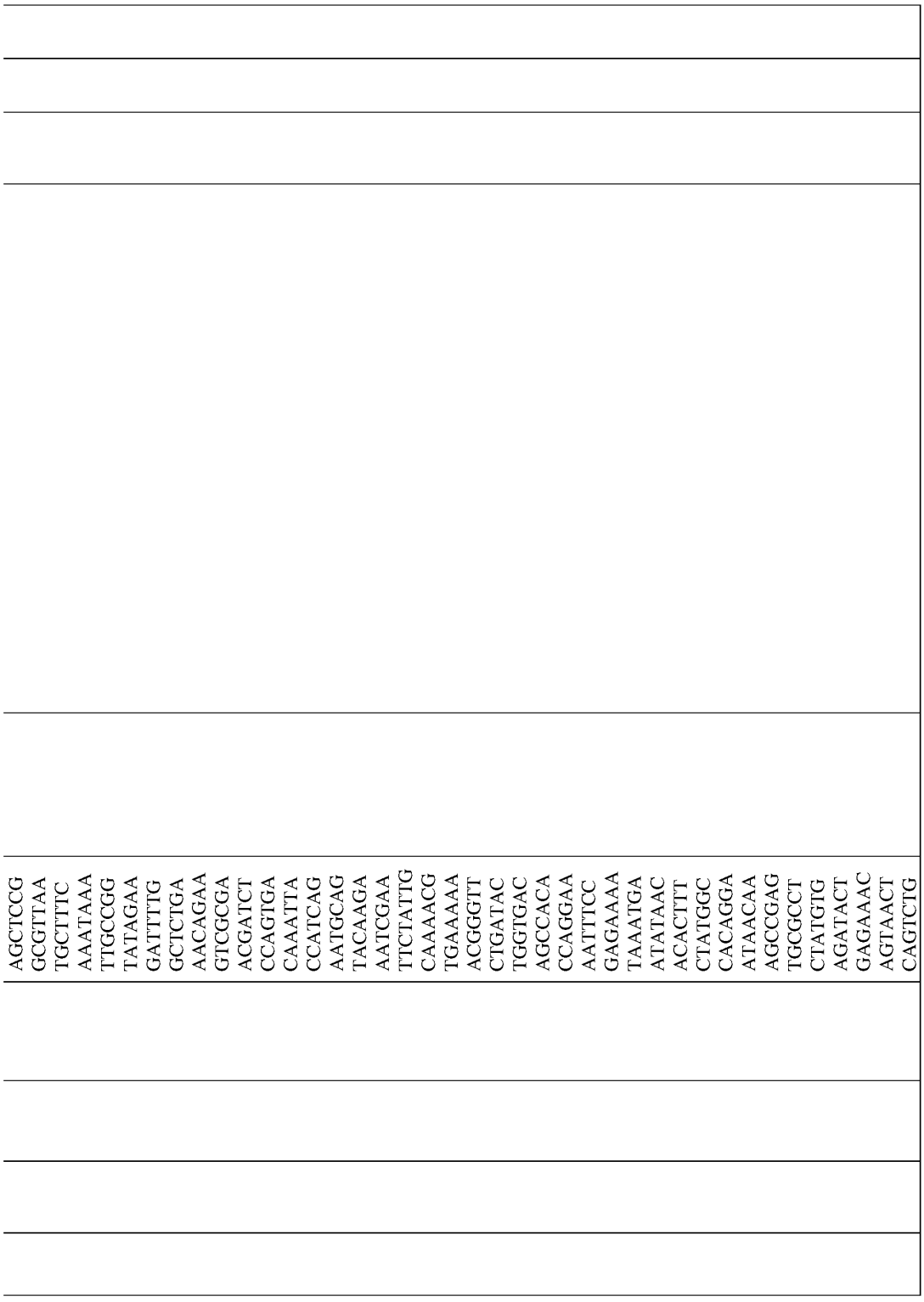
				TACTTAG AAGGCAG CCCCTTC ACCAGGC TATGGGT GCTGAAG GCTAATA ACCTACA CACCCGA GATTAAA CTATTAC AGAAACA TCA	TAAACCAG TGGTGAAT GCCCGAAA CGAGCCTTT GGGGGATC AATAAATG ACATCCACC ATGCATCAC CTGAAGGG CTCAACTGG ATCACCTGT TGGACTTAG AATCTAAG AACTTGTGT AATGTGTAC ATAATTTAA TATAACACT GTACATATG ATCCTGTAC ATATATTGT CCTTGTCAT ACATATCTT ACAATGTAT ACAATTAG AACTATGTA CTTGAACCA AGCTCTGTA	MRLSIQIIEKPAVDVIALQETHTADDQQIKSRGKIPGYKL IASNNHSKCGLAIYVKHTINSLKQPFKMWSSNLRQAYIK NYQVVSVSSENNIFTISAKIDSATIVNVYKPPRVNXPTPP LPSQQHPTLFLGDFNSHHNLWGYCENDSNGEALCEWM ETENLKVIFDAKDPPTFHSGCWNRGYTPDLCFTTQLHSS HTQIRRSALRISLTVSTDLSFSRDWNMCTPCAVTPETKV EFQESRKSNVDKATDLLQSLDENRKKIWHETVESFDHT HSSRKAWSIIRKLDATNQTPVRPKTAIKLNDFAKHLISNS KSPKDKQTRREIKTQLNTKKRAAPNSSKFSNPFTEEEVR DAIRNMKVGKALDSMGSILNFSHTVDQQLSNGFLTSSQ TFSPLEIFHHYSRKPSRGSKRCHQKHESWKGAGFDGIYP EFLTHCGPATIKWLSNFFSDILSTGNLPPLFKKTKLJAILK PNKDPTKVESYRPIALLSVTYKLLERLIYNRIHXTINKXIP IEQAGFRPGRSCNDQVFALTTHIESGYEKKFVSIAAFLDL SAAYDTVWREGLLLKFVSIIQCRKLTNLLSNMLSNRLFQ IYSDNDRSKVFKLNDGLPQGSVLAPLLFNLYIYDLPATS SRKFIYADDICLVTQSPSFTTAESTLTADMEILDTYFRKW CLRLNLQKTVTSAFHLRNRHSNYKPEVKLRGQTLQYCS TPKYLGVTLDRSLTFRQHLSNVAAKVKTRNNIIQKLAGT SWGCSAQTLRSTALALSYSVAEYCCPTWMNSVHSRKID VQLNQTMRLITGCIHSTQTAGYQFHPTSSSTRSDALLKI WRNIQKNPQLPIHSDTSTAEHQRLKSRKPSWKTAHHLD VTNFDINEAWKSHWEESSVSTVIWLKTLQCEFQVSHSP DASGESSTASEQDKGTADIYSTNGAGQHLPSVTVELPYK Q	cd01650	461	715
Vingi	Vingi-20_LMi	Locusta migratoria	.	GGTGGTG CTAAATG AGTAATA GCAACAA GATGAGA CGGAACT GCCCGAA AAGGCCT CCCGGTC AAAGAGT GCAGAGC AAACAAC AAGAAGT CCTTTTCA GGATGCA CAAGTTC GTAGAAT TGGACCA TCTATTC GCATACT CCAACTK AACATAG AAGGAAT AAGCAGG TCGAA						

Nimb	Nimb-22_LMi	Locusta migratoria	CCCCTCA GGGGCTC AGCATTA ATTTGCT GAGTACG GGCTTGG CGAACCC GGGGTCC TGAGCCG TGGGCTG TGTCGCG CCAACAG TCCGCTA CACCATA AGCTCTG TGCGTGC TTCAACG ACCATCG TTCGGCG CGACGGT GGAACAT TGTACGT TTCAGAG CCCGGGG ATCTTGG CTTAATC GCCCGGG TCGCGAG GACGGTA CAAACTC TATAAAA AAAACCT CAATCTC AAGGTGT GCTACGT GCCGAGG AGATGCA	TCATCATTC GATAAATA ATAA TGAAGATT GGTCTCCAA GACAGGGA TTTCAGACT GCAACTGA CCCGTAAG GTTCTGGGA TTTGGACCT CGGAATGG GCTGCCTTG AACGAACC GAATAAGC TGCGCGTA GTCAAGCC CCCACAGC CGTGTGGA GATCCTCCC TGCGAGCCT CTCGTAGG GAGTCTGTA GTCCTCTGC AGACTACG TATTGGMC ATACGTGGT TGACACAC GGCCACCTT CTACGCGA GAGGACCC ACCAAATT GCCGTTGTG GTGCACTAT TGACGTGGT CCACATCCT GGTGGACT GCCCACTTT TAGCCGACT TGAGGCAG	GQVWRSGIFFENASGAVLIQTASPAQSRALLACDKLGD VPVSITPHKSLNMVQGLIYHRDLLSQTDEELRANLERRG VHFVRRVYRGPKDQRVATGAFILAFEGDTLPEKVKVMI YRCDVKPYVPPPMRCFRCWKFGHMSSRCTSSPNCRECG RPSHPDAPCTPPPVCVNCGQHHPPCSPNCGVLHRERKIM EIKTLDRLTIQRHVKSMSGSIRAAPSGKDDNIFCRSHSGT SFTAINGQCSLLCGPYGPTTSTAPSMGGGTSSSVAPTTPP VGAITPSPSGTSVPAPSRRRNSLLRLLSLGRGPWDTPFRG FSQCTGDPSQRLKGPRPAGRRASRSSSNPDATTEKPSQA SKPKGQREGRQKKSKKKETLVATKPQAPSSSASEDEVEI LASPEDLAIPGSPESRIADTGDSEQLAADVMDEPVLLYG FSARRPVLLQWNCGGFFNRLAELQQLLCRFPVLCLALQ ETWFPAQRTPALRGYRGYYKNRAAYDRASGGVCIYVL ATVPSAAVPLHTTLEAVAVRIRAMQEFTICSIYLPPDEAV SIADIAALIAQLPPPFLLLGDFNAHHPLWGGTQTTGRGR DVEALLSQRDVCLLNTGAPTHFCVARGTYSAIDLSVCSP SLLPHFRWKVHDDLCGSDHFPILVSLLHFQPAGHQPRW SLRRADWQRYASAAVVPVPLPADVDMAVHQVTSSIVS AAEATVPTTSGSPRRKSVPWWTPEIAEAIRDRRRALQRH RRHPTMENLIAFKRLRARARQLIRQSKRDCWEQYVANI GSRASSSQVWSQVRRLFGHRPPAAIPSINIQGTVCTDPTA IAEHMADHYACNSATAAYSAAFRALKVRVERRHLSFHT RRDETYNAPFSQWEFQTALAACPDTSPGPDRIHSQFLRH LSAESQHQLLALFNHLWAEGAFPSQWRESIVIPRNVQCA IQSVGIPNCPCCMPRYLPGTRPDPQPVPSPPFSRVPTPASR PVQPSMGRGGFPITMAGKHRYPVMKPGADPQVADSYR PISLTNVLCKLLERMVSRRLMWILEHRGLLSPRQGAFRQ GRSTVDNLVALESAIRTAFSRREHLVAVFFDLRKAYDTT WRHHILATLHGWGLRGPLPLFIQNFLSLRSFRVRVGTSL SSPRIQENGVPQGSVLSVPLFLIAIDGLAAAVGSSVSSSL YADDFCLYFSSTGMGVAERRLQGAITKAQSWAMTHGF QFSASKTRVMHFCRRRAVHPDPELYLGXDLLRVEQTHR FLGMQFDARLTWVPHLQQLRKKCWRHLNVLRCLSHTA WGAARTSLLQLYKALIQSRLDYGSLVYDSAAPSALRVL DPITIVASVSWGSPPSESAGSSYWLLVARIHASPDHPNHR LLFPAAVLPSPRRRPRSGLPIAVRVRSLLSSLASLPIPPTF HYLSRTPPWFVPRLQLRLDLSSDSKASVSPMVLCRQFQ	cd01650	993	1246

				TGGCTGT	TCTTTTACC	ALIDSFHGAEVVYTDGSLVDGRAGYAYVHGDHVEQHS			
				TGAGGTT	CTGCCTTCT	LPCGTSVFTAELVAISRALTHFRDCPGESIVICSDSLSGLL			
				GAACAGC	GAGCTTCCA	ALDQFMPKHHLVRSIQDAVFTLGESGRSVVFVWTPGH			
				CGCTAGC	TTGATTTTA	MAFRETNESITWPSRPPHHGP			
				GAGCAAC	TCTGACGAT				
				CTCTGAG	GCCTCTATG				
				GAACCTG	GTTGACCGT				
				CCGCACC	GTTTTAGGT				
				TCAGTTG	TTTATTCGC				
				TATAAGG	GCAGCTGG				
				CTTACCC	CTTTTATTG				
				AGGCAAG	GTCTATTTA				
				CGGGGCT	GTGGTGGG				
				CTGCCTG	GCCTATTTG				
				GGTGGAC	CCCGCCTTC				
				CTTTAGTT	TCGGGCTCC				
				CCCTAGC	CACCCCTCC				
				TGCTCGT	AGAACAAT				
				GGGACCA	AGGCCGTT				
				ATGAAAC	GTCCTTTTA				
				GAGCAAA	CGGTTTATG				
				TAAAGCG	ACTTGTTTT				
				TGGCCGG	TCAATTAAG				
				CGCAAAA	GATTTTAGT				
				CGGCGTC	GTGTGCAG				
				TCTAAGC	AGTGGCTG				
				GTGGCAA	GCTCATCCT				
				AACGAGT	GTATTATCC				
				TCAATGC	TTGTGATCA				
				GTCACTA	GCCAGCCA				
				TGACCCT	CACTCTTCT				
				AAGACGT	GCTGTCCTA				
				TTCCTTCC	TTTTAGTTT				
				CTGCCAC	TCTGTCTAT				
				CCCGTGG	TTGGTCAAC				
				GAGGAAC	TCCTGTCTG				
				GAAGGGC	TCCTCTGTT				
				CCTTAAT	ATATGTCGT				
				AGATCAG	CATTTTACC				
				AGCCATA	ATTGACTGT				

			CTCGCCA CAATACC TTGTATG TGCGCGT TCAGATG GTGAATC CTTCATG ACCACTA AGCCCAG TTCTTTGT TGAACAG CTGA	CTCTCTGTA CCCGGAGT ATACTGGA CGTGGAGT ACTATGCG ACAAAAGG GAAACTTTT AAAAGTGA ACTCTTAGC TGGGAACA GGGGACTG ATAACCCA GCAGTTTAG TCCCCTTTA AACACCCA ACAACAAC AACAAC	MLIKLLQKDDHTTRQPIAIQSVSTKDTQKCQNNPKRRKL CQLSLKIRLHHGWKKNLHQYQPQQIPLKKTRPNRRNER RQGPTEETSSEVSDKIMTPFLKILKKKKRFKAVLFSVPSK HPVHKNKVQQLEVELQTLNSSLVCITEHWCKGHELTHV ALKSYVLASHYSRTTLQCGGSCIYVKEGIVFKVRDDIIA LSEDKHIEISAIEISKLNVLKRLTVLCIYHAPSGDMDTFIA KLNTSIGHGLMYSKDSLYYVFTMLPVVTWILLLQSLTQ ALDMVSKPKGRILICGDLNINTLNADGTCITFLNTLHSY NISPLINTATRITKNSASCIDHMLTDIEKDMCNVVVDHL GISDHSSQFLKVNINVETNTKLLKYIRVFSESKRSVFASQ LASESWAEVYRESTVNGKFSKFLSIFKFKSEEVFPKVLR VIKAQNKNNWVTNGIRQSSKTLRHLSTIKNYNNTIQNFY NTYKKVYRKVLLAAKRYQNDKLIEKANNKIKAAWDVT NRVTNRKKAGQINIQFQLNGTSVNDPKKLANFANKYFS NVATKLQQKFVNNTQHPHTLNNVPHTIMFLQPSTAEEVS KVVQNLKSKSSAGIDEIPVCLLKDCTQSISHPLAHIINESF TKGCFPETLKHAKVLPIHKKGDPESIESIENYRPISLLSSF SKVMETIMKNRLMAYLTKYNLLCKEQSGFRSGKSTSSA IAQFTNAVLEALDKGDHVTGIFLDLTKAFDTVDHTILLN KLEALGVRGVANMWFHSYLQNRVQQVEITKTNSNHTI KHLSEPEYINIGVPQGSVLGPILFLIYINDFPQSVKHGKSV LFADDSNILITDNSQIKLNQKAEDTLVHVSRWVENNKLT LNVKKTTASTSTQINPQCKKNNSINFHIKRKPVTNFLKIK				
								cd01650	634	915
Daphne	Daphne-12_LMi	.	Locusta migratoria							

				CATGACT GGCATGC CGGCAGT GCATTGT GAATGCC ACGAAAA GTGAACT AGTGAAT ACACTAA AGTCAAA AGAAGAA ATAATAC GTGTGCT GACAGAC GATTTAG TGGCTAT AAATGCA AAACACC AAGAACT TCTGAAT AAATATC AAGAACT AGTAACG AAACACA AAGAAGT GGAGCAA AATCATT GCCAAGA ACAAGAA AATATTA TTGTTAA TGGACGT TCTGTGC GACTAAG GTGGCAT GTACCAA ATAAAAA MTGCAAA CAAATGA TAAGTGC		DENIECVSTTKFLGMHIDSQLKWTEHITVLGKKIASACY ALRIINSVCSSSCTRTAYFANIHSLISYGIIFWGTNEQNRQ AIFKLQKRAVRIMTNSSSRAHCTELFKYMGILTVPCEYIL QQSSTLEGPLYRIIQIYGNTNCSMRIHFTAVIYIKENINQY PPNSSIHEHETRSSRHLHLNRKNTTKTQNSILYNGIKLYN RLPQDIKNKDKIQDFRNSLKKFLEKCYYTVQDYLN			


AGCTCCG
GCGTTAA
TGCTTTC
AAATAAA
TTGCCGG
TATAGAA
GATTTTG
GCTCTGA
AACAGAA
GTCGCGA
ACGATCT
CCAGTGA
CAAATTA
CCATCAG
AATGCAG
TACAAGA
AATCGAA
TTCTATTG
CAAAACG
TGAAAAA
ACGGGTT
CTGATAC
TGGTGAC
AGCCACA
CCAGGAA
AATTTCC
GAGAAAA
TAAATGA
ATATAAC
ACACTTT
CTATGGC
CACAGGA
ATAACAA
AGCCGAG
TGCGCCT
CTATGTG
AGATACT
GAGAAAC
AGTAACT
CAGTCTG


TGACTCA
CTAATAA
TGATTTT
GTCATTA
TCATGGG
AGGCAGA
ATGATGT
CTATCGA
AATGAGA
GTAAAAA
TKCCACA
CGGTGTT
TAAAATC
TACACTG
CGTGATC
TTAAAGT
GCCAAAA
TCATAGT
TACTAGC
ATGCCTC
ACAGATA
TGACCTG
CAAGAAA
ACTCAAT
TGTAAAT
ATGGAAC
TTGCAAA
AATAAAC
AGACAAT
TCCACAA
AATGTGC
AAGTCCT
TCCAAAA
CKTAACG
TTCTGTG
AGCTTCC
AAATTCA
AGAGATC
TATTCAC
AAAACAT

			GGACAAC ACTASAA CACGCTG GGAAGTC ATACATT GGTAAA						
Nimb	Nimb-26_LMi		CCCCTCA GGGGCTC AGCATTA ATTTGCT GAGTACG GGCTTGG CGAACCC GGGGTCC TGAGCTG GGGACTG GTAAGCG CCACCAG TCCTCTA GACACCG TAACCTC TGGGCAC GCTTCAG CGACCAC CGTGTGG CGCGGCG GTGGAAC GTTGTGT GTACCAG GGAATGG GGATCTT GGCTTGA CCACCCG GGTCGCG AGGATGG ATAAACC TCTATAA AAAAAAC CTTCAAT CCTAAGG	Locusta migratoria	TGAAGTGA CCTGACGC GGCTCCGA ATCGGGCA CTGCCCATT AACGCACA AGTTTTTAT TGTCAAGA AGATCCTCC TTTTTGTGA TGCCTGTGG CGTGCAACT TACATAGCC ACATTTTAA TGAGTGCAT TTATATCGT GCTGCGAG GGCAGAAG CCAATCTTA GTGGTGATC TGCCCTCCA TTTTAGCTG ATGACGAG GCGAGTGT GAAAAGGT TTTAAGATT TTGTGAAGC GTCTGCACT CTGGCCTAA GCTTTTAGG TTGGAGATT TTAGTTTGA TGCGAGTG GCTGGCTG	MGVPLGGTNTHSKKRARLGSPPDTGLCNMSMLKSNRA HSAVQNVFLVIKRKEGSFEKISPFYIQKGLEGIAGDLKSV KRLRNGTLLVETSSSQQAVNLRNAKCLGEYEIETEMHS TLNYSKGVVSCRDLVDIPNEELKEEWASEGIVDVQNITK RVDGGLVKSESFILTFNSPKLPEYVKAGFLRLKVRPYVP NPMRCFKCQRFGHTTLACKGEATCGKCGKAAHEGAEC TPPAKCVNCSGEHPVWSRDCTVYLEEKKVQEIKTTKRIS YGEAKKIFKAMQPPSFVTSFAAVVQKPVPKVDVSTQTD VASVGTNTCTCECTCKAAVVXEPELPHAHQKKLLXSTL QSFCHPPRQRPLARPWHLQSQRLFRRSSLTFPLLLSXTTA TRTSEKATGVNITELLSPSPAEAIGQALAPSKPEAIPKVIT DLSTPAVTNIKVAKPKKAPGKSQEQAKAGNKPSDYIEVI IPDLPSDSAPELMELDVPQGQTSRPKTEPPARSVSPTKRK ERVKIQPPRWLPYCSGTAGFRTHVEELRLLTRDRPMCV CLQETCFKETDSPELRGYFLYRKDDLTGVRAKGGVAIF VNATYHSSPVSLTTGLQAVAVSVYLRHRLTICSLYLPPH DALDEDALRDLLTQLPQPFILCGDFNAHNVLWGSANTC PRGRAIERILLSSSAHLLNTGQSTHFCAATGSFSAIDLSLS SPALAHSAQWEVAMTYMAVTTSPSGFICQTQWYPPEPH HDGCFVRQTGPCSANWLNLNNVIASRNGWTILHXSLAL FSSPCPLCSMGSGDDLHGSDHFPLWIHLPDTVVPARTPP RWVLRQADWTLFSQLAEFEQRDSVQEWVDHITQXIHR AAEASIPQSTGQLQRHPVPWWTEECRSAIRARRAALRK FKCRPTAENLAAFRLARAKCRKVIRESKKRSWQHFLNS VNRSTPSTKVWETIRRISGRGRCPVAAVMANGTLHTD PRDIAQTMAEYFATVTSTAQQDPAFQRHRMGAEKGSLD FRSTTAEEYNCPFSMWELDSACLQLVIRALDMTGYITVC YSTLQNGTEKSSSHFLMSFGRQVIFPTHGVKQFEGQFGL QVHYCRRIQLPFLHVGARFCLSAARDTCPGHDRIHYSM LQHLTERNREILLSLFNVVWASGHFPDSWREAVLVPILK PGKDRTSPSSYRSIALTSCVGKTLERMVNRRLVWMLES RQLLSRFQCGFRKYRSTFDNLAFLEAAIQQAFLRRQHLL GIFFDIERAYDTTWRHLILRQLHEWGFRGAFPFLLCPFFL HAILGIESVTCSRIVLSKRTVFLRVAFMGFSWSLPLFIVSF	cd01650	1150	1451

				TGTGCTACGTGCTGATGGGATGCATGGCTGTTGAGGTGGAACAGTCGCTAGCGAGCAACCTCTGGGGATACTGCCGCACCTCAGTTGTATAAGGCTTACCTAGGCACGTGGGGCTCTGTCTAGGTGGACATTTACTTCCCTAGCTGCTCGTGGGACCAGCATGAATATTTCTTTGTTTTTCCCCTCCCAGTGGG	AGTTCCTCCTTGTGTGTCAGCCAGCCGTGCTAATCTACCTAGTCCTCTCTTGGTAACTGTGATATGAATGTTTTTATGGCCTCATTTAACTAGCTTGTCTCTTATCTTGTTTGACACTCGTCTCACTCTGTTTTCGTCCGGGTGCTAAAGACCTCGACGTCTGCGCCCTGATCCATAATCACCACCACACCACCA	LSPRYFRYRVGDVLSDRFEQENGVPQGSVLSVTLFAIAINSLTSVVKSPVQCSLFVDDFSIFCSSSSLATTRQLQLTLRRLDDWAKKSGFKFSTEKSVCVLFNRSRSILNLPELRMRDTVLHFKETVRFLGLIFDSKLTWLPHLKDLKKRCLKALSILKCLSHNSWGADRTCLLRFYRAFVRSRLDYGCTVYGSARPSYLKMLDVVHHEGIRLATGAFRTSPIPSLCAEAAEPPLSIRQLLLGRQGYKMLSTPHAPAYHSVARPPMAGCFILGRGQHDPLGFARIVSHRWVYLILMFTSGLEQISSLAPSEAQHYFRLGSFKKGCTPDFTYRLNFSKFIGITILRWSTPMVQAGDSLGCSIVFPTMLPGFASHSLFCLHSRASRDLSIRNAACSEHTVSYLLRFTRTAITDSLYPDEKIHSSYRTNCTPSSGGVSRCHSAGSWHVGIRGNEVADGLQRRPAATTTYTICHSPAGCPLCGSAEVLSAVGRHVAAVTGNKLRSVKPTVRRGNLLPVTDA			
Vingi	Vingi-14_LMi	.	Locusta migratoria	GAGAAGCCTCCCCGCGGAACATATTAATACGTTTTGGCGTCTCCCGGGACGGCGGCAAATTCCTCAAACCAGAACATCAACACA	TAACATTACAATTACCTATTGTACATGACATTTATCCTACATTTTAGCTTATAAATTTTATTACTCTCTGGACATGGCCAGCTTTAGCCGAAAAGTCTTAGCCCTA	MLTIISVNIEGLSSSKENLLHEMCQSNKCDVLCLQETHRDNNHRRPNIPGMKLVAERPHAQYGSAVFTKPDLQVKSTCHTEENNIETITVELKDYTVTSVYKPPSEIFAFTPPANFVKEKISFVIGDFNSHSHVWGYASEDKSGEAVLAWAESSQLSLIHDNKLPPSFYSRRWHRGYNPDLLFVSENIAIQSCKSVCPPIPNTQHRPIMCQLFPAIRPQNVPFRRRYNFKKANWDRFSGLLDEKIQTINPAPEEYDLFVDTIRTCSKASIPRGYRSNYVKGLDSEAATNLDRYYTLFERDPFGEETLAQANLVTSSISEAKKKEWVKQMEELDMGRSSNKAWRLLKRLATDQLNLKHANVKSNQIAHQLLKNGKPVRRNRNVQNSRQIERDISQETSEISKPFSMTELDKAISKSRNGKAAGLDDICVEQIKNFGPAAKKWLLAXMNNCMSACKIPKIWRKTRVIA	cd01650	461	717

				GTCATTAAAATGCCTCATAAAGAACCCAAAAGCTCTTTTCAGAACCA	CAAAATACATGTTCATGTAATATCTATGAAGAGCCTAACAAACCTATCTATCCGATGTCTAATGTACGTATAAAGTTTGATGCTTCTGACACGATTAAATAAAATAAA	IPSETWEEPDDXXSYRPISLLCHLYKTLERMILNRLTEVVEPLLIPEQAGFRTSKSCTAQILKLVQHIEDGFEERKITGAVFIDLTAAYDTVNHRLLLQKFYTSCKDYGMTQMIGNLLQNRRFFVDFQGQRSRWRTQKNGLPQGSVLAPLLFNIYTNDQPLPEGTQSFIYADDRAITAQGNTFETVESKLSEALEQLAIYYKENQLTPNPRKTQTCAFHLRNRQATKKLNVSWDGIPLVHCATPKYLGVTLDRTLTYRDHCMNTKQKVAARNNIVRKLAGTTWGAEPSTLRTTTLALCYSAAEYACPVWYNSSHAKQVDIPLNESCRIITGCLKPTPTDKLHCLAGIAPPSIRRKVAASKEKLKSVTSAAHPLFGQQPARRRLISRRSFLHTTECLTDTPQNTRTEMWRNKNVHLHDWLVPTEELPPGHGEKRNVWRSLNRLRSGVARTKSNMHRWGLSDESVLCDCGQEQTTSHLMQCELCPSTCTVEELMKATPNALDVASFWSKIS			
Vingi	Vingi-4_LMi	.	Locusta migratoria	GGGGGTCACAGCAGAAGCGTTCCCGTTCGCGACCCCTGTCGAGTCTGATACCAGTCTGGTATCAGTCCGCGAATTCTGCTACAAGACGAGCTACATCGGATCAATATGCCGAAACAAAAAGATAGCCAAAAGAGTGGACTCGTTTTAGAGCCCACCTGACAATTATTTCGCTAACATCGAAGG	TGATTCAATTGTACTATATGTTGTATTTGTCTATCTCTGTTTTTATAGTTTTAAATCTGTACTTTATATTTTATCTCTGTATTTGACTACTGCTTTCTGTAAAAGCACTTACTTTGTAGTTTAAATTATGTGCTGGCTTCTTTCAGTTACATTTGTGATGTAGTTCGTACACGAATAA	MRLVNEIAHEKHGSAIFVRQGLQVSGSRTIQHNNIELIGVELDNTEIISIYKPPAQTFNLNIPLLSHTRDNMHTVVLGDFNCHSTQWGYSETNEDGHLLEQWAENNQMCLIHDPKQPPSFNSGRWRKGYNPDLIFVSKDLSALSYKQVLNAIPRSQHRPIKLEIRPVVQPVEIPFRRRFNFKRAKWEDFGNKLDESLKLIEPKPEHYDKFVQAIKDASFKTIPRGCRTEYIPGFTDKNKDAYERYCDLYDKSPLAEETIECGEELLEALSDQRRQTWMKTVVGMDMTHSSRKAWRLIKKLSNDPTIRTPHVNITPNQIASQLLLNGKVNGKNNAARSSLKVKKSKESNVSDVDFSVEELEEAISSLKVGKAAGLDDIRNEIIKNLSENXKKWLLKFFNCIWRTAKIPRIWRHAKVIAVLKPGKDPNDPKSYRPISLLCSLLKLFERLIMNRIGDYIDSCLIDEQAGFRPGKSCTAQALNLTQFIEDGFERGLPTGVVFIDLTAAYDTVNLRTLISKVHRMTNSDRLAEVLRALLFNRRFFVDFQQKRSRWRNQKNGLPQGSVLSPLLFNIYTNDQPIGKKTKSFIYADDRAIACQAKTSNEIEAILTEDLTTLTEYYRRNHLKPNPSKTQVCLFHLNNREAGKVLDIIWNDVRLNNCSHPTYLGITLDRTLTYKKHIEKTKAKVQTRNGILNKLTSAKWGANAKVLRSTALATCFSVAEYGCPVWERSHHTKKLDVVLNNTCRKITGCLKPTRVDYLYSLSGIAPPNIRRSFLSSIERNKQTNDHRHLLYKNQPVPKRLQSRRSFMYSVEPTVASIKSLWQEQYPEALITEEALPQGSHLDWQHWKTLNRLRVGVGRSKQLMKRWGYTDSDECECGAIQTMEHLLKCPLVDASRECTQEDLLKYTGTAAAYTDFWTSKI	cd01650	412	667

				AATTTCA						
				GTATCCA						
				AACAAGA						
				TTTAATT						
				GCAGAAA						
				TCTGCAG						
				AACCAAC						
				AGCTGCG						
				ACGTCCT						
				GTGCCTA						
				CAGGAGA						
				CCCACAG						
				GGGCCTC						
				TCAAAAT						
				AAGCCAA						
				ACATTGT						
				AGGT						
CR1	CR1-3B_LMi	.	Locusta migratoria	TTCTCGG		MEPLREMAAKEGKKANVHSVCIPGGVIPDVERVLPDA				
				TATTGTA		MKGTGCSQLQVVAHVGTNDVCRFGSEEILSGFERLSEV				
				ATCATGA		VKTASLACKMKAELTICSIVDRTDCGPLVQSRVEGLNQ				
				ATAGGGA		RLRRFCDRVGCRFLDLRHRVVGFQVPLNRSGVHYTQEA				
				CTGTGAC		ATRVAGAVWRGLGGFFFFRELCGVDWAVFFFFRLEGL				
				TGCTGTG		GNTQKGFQSQRVQVKDRTNTDLKTIGLTVVNCRSCVG				
				TGCGGAT		KVPELQALIESTESQIIIGTESWLKPEISSAEIFTKGQMVF				
				GCAGTCT		RKDRLNKIGGGVFIAVTGGLSCSEIEVDCSCELVWVEVI				
				GAGCTGG		PDNQNKLIIGSFYRPPDSDDSIAEMFKESLDQISKTYPKH				
				TGACACT		TILIGGDFNLPSICWQKYMFKTGGKHKTSSEIVLNAFSEN				
				TCGCTCC		YLEQLVHEPTRIVNGCENILDLLATNNPELIKSIKTDTGIS				
				CAGCTTC		EHRVVVARLNIGLSKSTKIKRKYIYSKKQIKIRLTLSKKI				
				AAGCTGG		YLFKKADKNSINAFLSENLQSFETNNVSVDQMWLTFKEI				
				AGGCTGC		ISTTIDRFIPNKLTKRGLTRRGTQNRLDHCYKSKEKSMP				
				GGTCGAG	TAAAAGTG	NLNAQKIPKTGNLLQKLEIQIHTKQINKEGADPPWYTKQ				
				GGGCGTC	GTTCGATGA	VRSLLQKQRKKHAKFKRTKNPQDWQSFTETRNLARTS				
				GTTGTTG	ACCCTCTGC	MREAYNSYHNETLSRNLVENPKKFWSYYVKYTSGKTQSI				
				TTGGCCG	CATGCACTT	PSLHDRNGDATDDSSAKAELLNTAFRNSFTREDEANTPE				
				GTCGTGG	CAATGTGA	FVSRTAASMRNLEVDTPGVLKQLKSLNKSKSPSPDCIPIR				
				GACAACG	CTTGCAGA	FLSEYADAIAPYLTIIYNRSLSERSVPKDWKVAQVTPIFK				
				GATCCGT	GTATCTATG	KGNRSDPQNYRPISLTSVCSRILEHILSSNIMNYLEENDL				102
				TCCACCG	TAGATGTA	LSHSQHGFRKYRSCETQLALYSHDVMSAIDKGCQIDSVF			cd01650	772
				ATCGGCC	GATGTAG	LDFRKAFDTVPHKRLLIKLRAYGISPQLCDWIRDFLSERT				2

				TGCGGCA GCGACAT GCCCTGA TACTGCC AGCAATG AGGCAGA CACCCCA CCGTGGT TGCGTGG GAGGTGT CCAGTGC GTGGCAG GCGGCGA AGACATT CCGGGGG ACCGATC GTGAGGC CTCCCCG GTTCGTC TTGCGAA CAGGTTT GAGGCGC TATCTGTT CGGACAA TGTTCCT GAGCCGG ATGGTGT CGCTCAT CCTGAAG TCACCAC GCAGCCT GCAGATC CGGGCTT TCACAGA GGGTGGG TTTGCTTG TAGTTGG GAGCTCC AACGTTA		QYVVIDGKSSSRTEVISGVPQGSVIGPLLFLIFINDLGDNL SSSLRLFADDAVIYRLVKSPEDQKQLQNDLDKISAWCK NWQLTLNNDKCEVIHMTTKRNPLNFSYTINQSNLQAVN STKYLGVTITNNLNWKEHIENIVGKANQRLRFIGRTLRR CNSSTKETAYTTLVRPILEYCCALWDPHQVGLTDDLEKI QRRAARFVLSRYRGESVTEMMHELGWKSLKERRYLLR RDLFAKFLSPTFSSECEHILLAPTYIGRHDHHNKIREIRSK TERFRCSFFPRAIREWNYREIS			

Pene lope									
	Penel ope-59_L Mi								
		.							
			Locusta migratoria						
				GGCGGGT A					
					TAAAGTAC				
					AATCCATCC				
					TACCGTAAT				
					ATACTTCCA				
					AATATGCA				
					ACAAAACG				
					CGATAAATT				
				CTTGATG	ATAGTAGC				
				AAGAAAG	ACCGCAAT				
				ACGGATA	ATTATCCCT				
				TATAAAT	ATATTACAA				
				AAAACGC	TATAGACA				
				TTGACTT	CAAACAAT				
				CTTTAAM	TTATTACAC				
				TGCAAAT	TTAAAATA				
				AACATTA	GTTCGGAAT				
				CACGTAT	GAAATTAA				
				AGACAAA	AATTTATTG				
				GACCCTA	TACGATATT				
				CAAACCG	TGTATGGTA				
				ATTTCAA	CTGTTTATG				
				GCTTTGA	TCACACTGC				
				TAAAAAG	ATATAACC				
				AGCTATA	AACGATATT				
				AAAATCT	GTTACAAA				
				GAGTTTC	CAAAATTA				
				TTTTAAC	AGCAAAGA	NKCSDSTNCESDKQPSNLIEKKLNKIICENYTYSRQYTV			
				ACAAAAA	ATATTTAAT	KSSTDLIAKIQDIKIPDNSRFLSLDVTNLYTNVPIDETIAII			
				GAACAAA	AAAAACAT	KENLLKYRKIPKQSLNNSTITIILSHNYFSFNGHVYTQND			
				ACAGATA	CTTTGGATC	GLAMGSCLSGTLANIFLDHLENKFFEEQNQWKTKVVYY			
				TATAGTT	AACAGCTCT	FRYIDDTIILFNGCIEEIEDFHQRFNNMHKKITFTIEHEINN			
				GTGAATC	GTAAATTAC	SINFLDLTISKTNSVHTTFNIFRKPTTTDIIIHNSSCHPQSHK			
				CACAATG	ACCAAAGT	EAYFRSAIDRITKYPLCNKNINSELNTLKDIAISNGYNPE			
				CCCCAAG	AATTACAA	LVNNLYKSKLNSKQKSIKDEITLTKTETRKKDKFVILPY			
				CTCAGAG	CAAACAGC	WGQISDRIKYALKKTNLKIGFYTNNTLGTNIIHKIEKIDR			
				CACAACC	GTGACCTA	YSQSGVYKIECSKCDKAYIGQTGRPFKIRFREHINTSENN			
				AAAGGTC	AACAGTAA	VSAFGTHLRESKHSTGSIEENLKILHIVCKGQLMDVLEE	cd00304	123	210
				CATAA	CCAAATAT	LQIYTHMTKYPEQMLNEQTEMRNKKFIENFHNTLASI			

Pene lope		Penel ope-6_LMi	.	Locusta migratoria		AAATGTAT ATATGAAT GTAAGTAA TTACAACTT GTACAGTC AATATAAC ATTGCCATA TTAATGTAT TTTCATGTA AAACACAC TTTGTTCTC TGTAATAG GTGGCTTAG AAAATGGG CAAAGCCC GAAACTAG TCGTCAGCA TTTCAAATA AAAGCAGC TTTGGTGCA TATATAAA AATGTCTTT TCTACAATA TTTAAGAGC TGCAAAAC GCGACATA TCAAAAAC ATTT	TGAGAATG TCCTCCGCA GTCGGGGA CGAAACGT TGGGTTTCA ATAATAAA TCCACAAG ACCACGGC ATAATAGC CCGGAAAT TTTTAGTAA TTTTGACCA	GKQNCKLLSVHITISNGVCSSDWDRIDRLLHDTINKRVL TTSSRQRKKFDSLRPQPTISQLDSSRTVINLSGHLLSDDE RSVLARGGNFAVSPRVVPTEEIIANIEAGIRNLDKVTAEE IRIEAVRVLARAKPPKSNLTTGEKNAIKSLNEDDSIIILPA DKGNCTVVLNVTDYHCKIKELLDPQTYRKLSKDPTDKV LRTVTQLIKKSSIEQCVQKSLFNSVALPPRLYGLPKVHK EKENVPPLRPIVSAIASPTYLIAKFLTTLLRPHVGRSDSYI KNSRHFISKLNDILVNPEDILVSFDVVSLFTMVPVNEVLQ QLNLLFPPDIADLFKYCLTTTYFKWNDDFYEQTDGVAM GSPLSPVIANFYMEKFEEQALKTANKKPAIWFRYVDDTF VVWRHGREELDRFLQHLNGINSRIQFTMEKEADGKLNF LDVSVFKKGNGRLGHTVYRKPTHTDRYLHRDSNHHPQ	cd00304	354	438

					TTCCGGCCG TGAAAGTTT ACATTTTAC A	QKRGIIKTLADRARNICEPELLDAELKHLSNALLKNGYS STEIRRALRPTRRVENDERVPVKSKVFLPFIKDVTDRIGK ILTKRNVAVIYKPTRKIQDHLKSAKDARKPLEKAGIYRIP CSCGDVYVGTTKRTVKKRLDEHKGNCRRGEIDRSAVA EHVFQPGDHLIQFDKTQVLAATNGYHERLYREAIEIVKH PNNFNRMEEGVKLNEIWIPVLKKKCNSLPPWDDINGRR QSATTNCARMSKTSDVTPLRASPRKRSLASVSSEPGCAS DIQKCHDPP				
Ving i	Vingi-27_LMi	.	Locusta migratoria	GGGGGCG AAGTTGT TCGCGTC CGGAAGC AATCCTC AAAGCTC CACGTGC CCCCTGG GTGGTTC TAAACGG ACAACAA TCTCTGG TAGCTGG TCCTGTC ACCACGG TGATCTC CCAGCT	TGACAGAG ATATGTTTT ATCTTATTA ATTAGACA CTAGTATAT CTTACTAAA GTATTATAT GTAGCTGA CATTATACC ATA	MSATTTIPNGSCQGQQNCQQNDLNRAIGPVLNVCHLNI EGISRSKCDFLSKNLLSNNTDVLALQXTHAEDLKNCAR GNIHGYRLVGEIGHRTYGVATYARQDLDDVSIIHESCDD DIFVLAVKVSGTTIVNVYKPPNIQWQDTVLPVFEHPSLY VGDFNSHSSHWGYSSDNENGIALAEWAENNNLSLVFD AKDSSTFFSARWKRGYSPDLCFCSCDQTLRSLSVTRKVL GTFPRSQHRPVLISVGIKIPLIRSFPRPRWNFNKADWEAF TSKLEDNIRFIPPTTENYDRFAGLLKATAKACIPRGYRKT YIPGWSQECEDLYNEFISTEDSNTADLLLNSLDNARREK WVKKVQSLDFTHSSRKAWSLLRKLSTSVSEKKLPPTHP DPSDIAKRIVALSKSKRQQSDRLISYKLNEVKKCQQAPL EVSKPFCLDELHVAISDVKTGKAAGIDDIPPEFIKHAGLK TRLWLLNLFNNILKDSSLPKAFKRSKVIAVLKPGKPADR ADSYRPIALLSVCYKLLERLIYNRIVPILTASIPVEQAGFK QGRNCCDQVLALTSFIETGFEKKCTTSVAFLDLSAAYDT VWKKGLLVKLYNLIPCPILLRLLNSMLSGRLIKIYLGNTE SGFRQLHDGLPQGSVLAPILFNVYTADIPATQSRKFIYAD DIAIASQHRDIHVTETTLTSDLTLLRNYFAMWKLKPNPN KTEVACFHLNNRMADYKLNISFDGVQLEHSEHPKYLGV ILDRSLTFKYHLHKTADKVKTRNNILQHLAGSSWGASA KVLRTTALSLVYSAAEYCSPVWLNSAHVRDIDVQLNRA MRLISGTLMATPLQWLPILCNIVPPDLRRKEALMREYYK ILSTPELPILQDLPPQVSRLKSRKPPLKTAVQLHESGFSPK SNWARTWDSFNGXNSIWIDDPSKELQGMDLPRREWSLT NRFRTGVGRCNYWKFKWGQADDQSCDCGADIQTMTHI LTNCPLRSFSGGMTELHJVSQEAVQWLKELDLAL	cd01650	490	744
Dap hne	Daphne-6_LMi	.	Locusta migratoria	AATATGA TAAACAA ATTAGAA AACACAA ATGTAAC AATAGTA	TAGACATG CTAAAAAC TACTTAATA AATGTAAT ATAGTAAT GACTCATTT	KEQKISKGAQWCPSAINRPIQDQDRSSGTRSKKSNATPH PARPQQKGERKTSKPQILEPSADLVVFHQNVQSLTNKID EFQIMLENELKKVSVLCLTEHWLQDDTLNMVHFPNFIL GSHFCRKKRYASPQHRKQSFVVADLHKSIHLHPSFPEAF QRKIPWKYQDLSYCMKTYLNXECVLHKLSKKNRHIIC GDFNIDFSVGLETKPKTDILNLLLLTFNMTPTIQTPTCVTK	cd01650	538	802

AATCAAC	TTGTTGTGA	TSATTIDQIFASVPMEYKARTLCTGFGDHTAQEITIKMD				
CTCGCGA	TGTGCCATA	LPLHCKTSLLTITTRSFNLQNTNTFTQYLTNESWEEVYSC				
CATGACG	TATACTATG	RDTNEKFEKFLSIFQHYFELCFPLKKRKIKKSNIKSSWITP				
CGTAAAT	TGTATATAG	GILTSLREKRVLHTISKQTKTTPEFDTYFKRYKQILNKVI				
AAATTAA	AAATCTGTA	KKSKKLTNFSYIEKPQIKCNVDYCERNRQNTGHTHSNIN				
TAGAAAA	TACTTATGT	LLENGEEITDPHRVADLFNVYFTEIASNLIDKQQKPASEN				
AGTAAAC	AAAAAACC	KNQQKLFIRKNNFPIPNRRKGSTEYYKRIKTKIFIRHKCA				
AAAGAAA	TGTAAGCA	TVVTKPLTHLCNSSLSNGVFPEKLKLAKVKPLFKKGDK				
TTGATGA	ATATCCTAT	RNEKNYRPISLLPGFSKILEKVFYRRLISFIEENNFLSNTQ				
CTATGCA	ATCAATAAT	HGFRKQKSTETAIFDFVNEALSTIDQKEHVSALFLDLTK				
AAAAATT	TTGATGAA	AFDMIDHNLLLKKLEYVGIRGLAHKWVTSYLKNRAQIV				
TCGGTAT	GTGGATGCT	EIAHQEGSNIYTYQSERKDIKYGVPQGSILGPVLFLFIND				
GTAAACT	TAAATAAA	LETSSSSQKYIKFADDTNILIKGKTAAELQHEIQNCTSHIE				
TGATTGA	TAAATAAA	EWFTTNNLVINIEKTVTMHLHTIQNKHPLTPTIKLFNKK				
TGTAAGC	TAAAA	LETVNSTKFLGISIQDDMKWDTHIKYISKKLNTICYGVRI				
AAGTTAG		LARCTTKETLKNVYYAQAESLLRYGLIFXGNSPESRYCF				
AGAGATC		VAQKRIIRSIENATPRKSCKPLFKALQILPLPCLYILEIVTY				
TATGCAT		IKKISVAQNNLVYTNKSVHTHDTIRQTLDLHSEHLRTTLR				
ACCAGTC		QNGPLQMGEKIVQQTSSPYQNNQRTPEVQIGCKKLFTR				
ATGGTCT		KIIYKINVITLSVNICQLNLMSCFLTYNYCNYCHLLYDV				
TCATCTC		DCIII				
AATTACA						
GTGGGAA						
ACAGTTC						
TTAGCTG						
AAGAAAC						
TGTAATT						
CAGTAAA						
ATTTAAC						
CACAAGA						
GGTTTAA						
TCTGGAA						
GCTGTAC						
AAAGTAC						
TTTGGAC						
AAATGGG						
TAGTAAG						
GAAGGAC						
CAGGTCT						
CCCAGTT						

				GTCAAAT					
				ACGTGGC					
				CCCAGTG					
				ACTGAAG					
				AAACTAA					
				TGCTGCT					
				GCAAGAC					
				CTGAAGT					
				CCTGAAG					
				ACGCCAG					
				CCCAACA					
				CCAAAAT					
				GGTTTTCT					
				AGACAAA					
				ATAA					
Nimb	Nimb-36_LMi	.	Locusta migratoria	GGGGCCC		MGDNFAMTSSSSSSTLPFTQKRKLDESKSKKVLPNLPKF	cd01650	987	1244
				ACAGCTC		LVVSRTDGHDFSSVNPFVIQRGVDAIAGPVKTCSRQRNG			
				TTTTGTG		TLLLETEGAFQVQKLLRTTLLHTFPVKVEPHRTLNSSRG			
				GGTACGT		VVFSRSLDGLSDEELQEYLSDQGVTAVHRVMKRVNNN			
				GTGTGGC		LTPTQTNCVLDLHRPKQTVFLTFDKVDLPSKIKAGYEIIS			
				GCACGGG		VRPYVPNPTRCYQCQRFNHTRQSCPNAAKCVTCGRDA			
				GCCCCGA		HEGDCPPPSPRCINCMGDHAANSRECPVYKDERAIQEIR			
				GCCGAAG		VKEKVSTSAARKLLASRKPAVLPAGTYSNVLSSPRPTKE			
				CAGCCTT		TGTQTCDFTFSDTVVRSASPKIARSTSPMPSAHSKSQSTS			
				CCTTCCC	TAGCTCTTC	VSQKSRATKSEARTSKKDPVREDFLKTSSSQPSAAHSSN			
				GCCCCAG	CTTTTGTTG	RTASKKAHKKRSSSPSPPRRDSSVTPSSSRRPRPSSVSPG			
				GCTGCTT	TATTTTACC	RSTACRSAGISPAEEAAPSDQSRKEAEQMSEPMEHDCPS			
				CAGTCCC	CTATGTTTT	VDCSSSAASQPGPQRPSRSPLAFSLFFFSAFFMALLQWNI			
				CCCCCCC	TAGCTGTTT	RGIRSNRDELRLLLLRLHCPLIVGLQETKLRPTDHINFPQY			
				GGTTTAT	TCATATAGG	TAVRFDLPPVEGIPAHGGAMLLIREDVYYHPIPLQTPLQ			
				GCCTTCG	CTTATTATC	AVAVRVTISTFTFSLCNIYTPSSSAVLRADLLHLVAQLPT			
				CCCGGAG	CTTCGGTAT	PFLLIGDFNAHHPLWGSPASCQRGSLLADLFNHLNLVCL			
				ACGGGTG	ATCTCAGTT	NTGEPTFLSDTTHTYSHLDLAICTTPLAHRFEWYVLSDR			
				TCCCGTG	TTGATTCGG	YSSDHFPCIISLLHHTTSPCSTSWNISKADWGLYSSMAVF			
				TGAGTTT	GCGCTAAT	HDQNFPSCDSQVNYLTDVILTAAEHSIPRTSTSPHRVPVP			
				AGGTGGT	GACCTCAG	WWNAACSDAIRARRRALRLFQRRPTLENCINYKRLRAK			
				CTCATGT	CAGTTGTGC	CRRVIKESKKACWTAFTSSFNRSTPSSVVWGNLRRLSGT			
				ACCAGCT	GCCCTAAA	KVHSPIPGLNIANQVLVEPVDICNALGRSFAEVSSSANY			
				TTTTTCTT	ACCACAAA	DPAFLPRKQAEEARTPSFTSPNRESYNATFTMQELERAL			
				TTCTCGG	AAAAAA	ALSRSSAPGPDAIHIQMLKNLSPAGKCFLLRTYNRVWTE			

				TTTCTCTT CTCCTAA CCTTTGCT TCCGCCT TTCCCTCC TTACTGT ACTCTTCT AGCCTTC CTCTCCTC CTTTAGT CTCCGCC TCGGCGT TTGAGAC TCAACCT TTCCCAC TTCCTTCT TCTTTCCT TCTTTCCT CTTTCTTC CTCCCTG TGCGTTC CTGAAGG CCGCCCA CGCACTC GCTGCGT AACAGGT GACTGGG TAATGCG TAAGTCC CAGCCCC GGGTTGG CATGTAG GGCCCGC ACGTACC CCCTGGT AAAGGCC AGGCCCA GGGAGGG GTGATTA CCCGAGC		GEVPTRWREAIVVPIPKPGKDKHLPSSYRPISLTSCVCKV MERMVNARLVWILESRHLLTNVQCGFRRHRSSIDHLVT LSTFVMNNFLRKRQTVAVFFDLEKAYDTCWRAGILRT MHTWGLRGRLPLFICAFLKDRMFRVRVGSVLSDTFRQE NGVPQGSVLSVALFAIAINPIMDCLPADVSGSLFVDDFTI YCSAQRTCLLERRLQHCLDSLYSWSVANGFRFSHEKTV CINFWRYREFLPPSLHLGPVALPFVETTTFLGLTLDRKLS WSPHVSSLAARCTRSLNVLRVLSGTSWGADRTVLLHLY WSTVRSKLDYGSFVYSSARPSILRRLNSIHHRGLRLAIGA FYTSPVDSLYAEAAELPLTYRRDILLCRYACRLLSMPHH PSYPSFFNDSLDSHYGLSAIALLPPGVRFRRLLRQLDFTL PSTFQIGESPVPPWLQAQVRVHLELTSLPKESTSDAVYR SRFLELRAWLENSVFIYTDGSKTNEGVGCAFVIGDDGFQ YRLNDHCSNFTAELFALYQAVLYIRRHRSPSYVICSDSL SAIQSLRDPFSAHPVVHRINRSLQLLAADGPRVHLLWVP GHAGVPGNEAADAAAKAATLLPRTATYCVPSNDIRGAL SRRLISSWHNVWSSLTNNKLRTIKPLPVAWRTTSRRSRR EEVILARLRIGHCRFSHRHLLTGDPAPLCSCGHPLTVRH VLLSCPNFTPVRIALNLPDTLEEILADDARAAARVLNFIN LTNLSKDI			

				TGATACC TTCCGAC CGTGCCG ATTGGTC CCTCCGT CCGTTTCT CGGGAGG TGTGACC TGAGGTG TGGACAA TCACCTA AGGCGGG TGAGCCC TCAGTGA GGGTTCC CCATTAG GAAGGAG CGCGCCA TCGGAGA CGCCGGT AATC					
Dap hne	Daph ne-57_L Mi	.	Locusta migratori a	CCAAACC ATCACTG GAACAAC AACCCCA CTAGAAG CAACTCC ATAATTG AAGTACA CAGTCAA CAAAAAC TGAACTA CCAGCAT CAGAGGC AACTGCT ATGCCTC AGCACCA CTGACAG CAGCAAC AGGACCA	TGATGTTCC TTATATAGC TAATCAAA ATTGAGAA TATAAGTTT TGTTGCAAA CATTGTACT ATATACCTC ATAAGTATT CCCCCTATA AATATTGTG CTATGTATC CCTTGTTTT GTTGTAAAC ATTGTACTA TATACCATA CCTACCTCA TAGTTCATA AGTATTCCC	KTTYIVTTAEHRRAKDQEDLLCVRIFYGFSTRRKAVTNQ PGDSSNTQVGAKSSVTVQESQCSFRIIHQNFGSLGNKHN DLDVLAFVNKPDVLVITEHWYPKDKISVMHPLMMNLC SSYSRMNKSGGGVAIFAKSGSNFIVSFYCVIDVSSLSVEG VIELSAINYRWERQHIIIVGVYRPPSANIDIFLTNLNQLLV EIDNKYSGKNGCINIIITGDINIDTLSNNSTTKRLQLLLQE FNLNLLIKEPTREVGGTRTCLDNIITNMDKFTYKTAVLK LAISDHHAVEFQCVPNINLHLVKHKPPPGKMKRYETKR TINDKNIRNLTHVLQSATWAETNCYSYENFASDFFYYFD MTCPTKSVIVNNNEKCRKVPGRKMQKSTWIDAEVTQA RNMLTLFQCKMLENRENTKLKEAFNKYHEYYKNLIGD KRSKHVAHILNNGNNTAACIWNVINSFRAGKDNVINDL VISYNGSVIECQQDIANIFNEYFSSVGKSINDRLHNNKYA FKGKASSYSIYFTPTNSNEIRNIVKTLRNSKSAGYDGLNI NIVKSCIDSISGPLSTAINNILTSGSFPDSLKIAQVTPIFKK GDRLKIENYRPISILPVLSKIIEKVIYNRIVTFLNKHNIIFE NQNGFIKNKSTVTAAAQYIQLLHSIFDEVITAIENKEYVT GIFLDLEKAFDCVNIPILLDKLWNIGVRGIAHDLIKSYLT NRKQFVVIRTSSGTYRSTVTDMKYGVPQGSVMGPLLFLI	cd016 50	581	845

				ACTGCAACACCAACCCAACATTATCACCAGTAACAGCAGCTAA	CTTATAAATATTATGCTATGTATCCCTTGTAACCAACTGACGTATTCAGAAAAATGATACTTGTATCATTTGATGTTGAATAAAGTTTATTTATTTATTTA	YVNDIKYCTQNSKVILYADDTSIVCKHKELDQLEILSNCLTNEIVQYFNESHLSVSTAKTAVMDFNPSRQTDFEIQLCIGKDIASKENCVKFLGITIQDTLKWDVHVDTLASKLAKNIFAINMLSKHCNDTQVLRTAYHALMSSNISYGIEVWGATTQANLNKILKLQKKAVRIICGAKPKDSCRSLFPKAGILTVVNMYIYKSIMLIRKLNLNLNSDTHNYNTRNKDKFHLLSHRTALYERNPLYAGQRLANTIPSCITSLPTEKFKVTLKRILITNPHYSLDEYFCFAQNKENFIPDSLTSV			
Jockey	Jockey-9_LMi	.	Locusta migratoria	CAAAACGGCGGACGACTATAGGGCCATCACAGCGAAGTTTGGGGCGCTTGATAAGCCCTTCTATACATTTCCTTCTGCCGCCGACAAAACATTGAAGATTGTTTTCCGTGGCCTGCCCAGTAATTACAGTCTGGAACACATAAAACAACAATTGTCGGACATGGGGTATGGTTTTCGACTCCGCTCCGTGACTCGGGT	TAGGAGCCGTACGCACAACTAGGATCACCAAGACGACGCAAACCGCAAGGACAATCAACAAGCCAGGACACAGAATACCAGCACCCAGTTATTTTGGAGGCGTAGCTCACCAGCTTGGACTCCACCGACCATCGGAAACTCATATGTTTCCGTCATTGCACGAGCGAGCGAGCGA	MPPLRQMQRLTRDQNQQMRNPRNHTARRQRNGGENSTTQPLTRSERTQEQRDRTHQAREEDVATAALSSSSTVPATPRENAPRFEDMVTLLQKLRDLIKEIPRQITAAVTAAIQAISTGPVMARSQQGLTIAAWNANGIRTQEGEFRQFLREEAVDVCLVNETFLKPGVHVRAANYVCYRTDRTTHGGGTAVYVRASLAHHRVQLPPLQAIQESHSGSGRHIHRQDHIRRSLQTTKDCGTTCRRRNFTVTPGKVFAAGDFNAKHHIWNSRVINNHGWHILREAQHNDAMAIGPNDPTHFPENGGQPDVIMLRSDPFPREWGTAGCYYVALIKGVPYVISASTRNVLSSDHVPVVLDVDVDAGPAKRRGLPTNSTNWDRYREHLGNSLQDAPMVEDAGVEDNVRTFTKNAIQAVKRATPRLQRDPPDRSRQLPPDILAMITAKNRMFREWQLTRNPATKRAVNRMQREIKAAVKAHRHQEWSNKLATLRVDDGTGWKATQMFLRRRLKIPPLQVGPTIVSAPEDKANAIADVFEANFTPVQDPIDPEHVRLVQERLPTFLNAQKSVQKRCPGFYAASINEKLEAPTSSLTKCFDRCRQKPQRTLQQRHCIMRTGHXPQAWKHAEVVPVPKPGKDHRLPSNYRPISLLPAJSKVFERVYLKRLMDHIRRENILPDEQCGFRPGHATTHQLLRLVEEALDXTERREYFGAVFLDVSRAFXSVWHDGLLFKLLDXRPAPAARPINSVVSPRQDVPGAHWKRPSXRPNVGNTSGQYSLTFRERSXPSGMTVYSSSCWTSGLPLPHVRLIRSYLQDRTFQVRTEELSTRRPIRAGVPQGSVLGPILYILYTADMPRIDRVHLALYADDTALFSRSMNARTLGTRLQRASDELSRWATKWRLSFNGAKTQALIIARKRVPVDLPRLSVRGTAVPWGRTAKYLGVTLDRRLTWQQHVDDVIGRARGRLRDLYPFLNPTSTLPPQLGVAIYTTLIRPIFEYAAVVWGNAAQTHIKRLQGLQNRILRLALHLPRDFPTEDLLRV	cd01650	621	939

				GAATTCT GGACGAA CGAGGAA AGAGACA CCACTGG TCTTGGTT GTCACCT GTGACAT ACCTCAT AACAGGG GTCTCTG GCAAGTT ACCAGGA TCGCTGC GACGCCT GTCACCA CGGAAAA ACTCAGA TCAAAGC GGGGTGA ACCACAG TGCTACA ACTGCCA AGCTGTG GGACACG TCGCCCG ATTTTGCT AATTACC TCCGCGC TGTGTGA AGTGCGC GGGAGAA CACGCCA GCCGAGA TTGCACA CGGAAGA AAGAAGA CGGCCCA CCCAAGT GCGTTGG		TEIPPIQARFRQAAKLFYQKAARSPNALIRALGHRVQRL NATRWPHLLRSD			

				TTGTGGTGGGGACCACGTCGCCTCCTACCGTGGATGTTCTGTATTCCTCCGGGCTGCTGTAAAGAGCCGCGGACAACGTCCTGGGGCCACTAGGAAACCAGCGACAGTACGCCCAGGAGTAAGTT					
Vingi	Vingi-21_LMi	.	Locusta migratoria	GGGGGATGGTGTTATTCGCGTCCGCAGCATTCCTCAAGCTGCACGTATCCCCCGGATGGTGCCAAACGAATAACAATTCTGGTAGCTGGACTGTCACACCAGTGATCTCCCAGCTATGTCAGCTACCCACCAAAATCCAAATGGACCTTCTCAG	TAAATGAAGCAGGACTTAAATGGCTTAAATCACTGGACCTGGCTCTGTGACGGGAGATCCTATTCACACATAATTTTTAATCCCCTGAACGCAAGTACTATGTACTGTAGTAACAATAGTTTAATATGGTGGCTGACAAGAACCATACGATAATAA	GLHQXNVDILALQETHIEEPLNNARGNIPGYRLVAAIGHRAYGVSTYVRQGLDDVTVKHNSSNDDTFILAVKVSEITIVNVYKPPNTQWTDTVLPVFEHPALYVGDFNSHHSQWGYPADDENGRRLTSWAENNNFSLIFDAKDNSTFYSARWRRGYSPDLCFCSCERSSKPMTARRRVLGAFPHSQHRPVIISVGTQIPIIRSFPRPRWNFSKADWETFTSKVEQSIRFIPPIITTTVSRIFSWLLLNSVSPGASGKNTCQDGVENATIFTRSFKTPKALTQPTICCSLLTLPGKRSGQSRLGDLYFKGRAEHQVYPSHNNYDRFANLLLAIAKQCIPRGFRKEYVPGWSRECDDLYKEFQDTESPDTADHLLQSLDTARKEKWIKTVENLNFTHSSQRAWSLLKKLSSGKPADNCTVTHPNPCDIAKRIVSSSKSARLPSDKRVAYKLNKILRSQFTPLEVSRPFTEEELEAAITTLQTGKAAGLDKIFPEFLKHIGPKTKFWLLTVYNAILATGNLPKAFKRSKIVAILKPGKSPEQAESYRPIALLSVCYKLLERLIYNRISQILDASIPVEQAGFRPGRNCCDQVLSLTSFIEKGYEMKKKTSAAFIDLSGAYDTVWRKGLLLKFYNIIPCPLLLRFLNSTLSNRLIKVYLGHKESAFRQLNDGLPQGSVLAPLLFNVYIADIPATQSKKFIYADDIALATQHTDLPTTEATLTNDLATLQEYFRTWKLKPNPNKTEVSCFHLNNKLANQKLNIYLDGVKLTHNNHPKYLGVILDRSLTFKSHLQKTAAKVKTRTNILQQIAGSSWGANAQVLRTTALSLVYSSAEYCAPVWLNSAHVREVDVQLNRAMRIIS	cd01650	524	778

						GTLLATPLPWLPVLCHIPPPEIRRQEALIREYKKILSTPQL PITQYLPQNLSRLKSRKPPLRLASQLIAEDFSPNTTWAAS WDXFEGRNKHLIPSPSTKVGGMNLPRREWTLLNRFRTG VGRCNYWKCKWGMTLDQSCDCGADTQTMDHIAIDCP CVLSLRGWLVFTI				
				TGACTGAC		MTTPAVKAFIYDQTPLRRGYVNIGSAVREHAKKSQQLC				
				AGCAGCGC						
				CCTCTGTTT		SLTFLLRCRDTSTVPKFLKVKRTFRSAQAHRIYSRMESA				
				TGCCTATAT						
				ATACGACG		MLRERIHQTRRALAETNSKLFQLHLLISNTMQCDDWNK				
				CTTACTCGC		IDNLTFRTMEASAEIAGNRQKRKFDSLQKNVLVGTMED				
				GCTCTCGTT		NSRTVINLSDKVLTADEVLVLAKGGNFAVTPRKVPTEDI				
				GCAGTACG		VANIEAGIRQLSQQSADVIRMETARILRTCKPPTSNLSQA				
				CCGCTTTAC		QRKALRDINADENIVLSADKGNATVVMKSEDYHEKIM				
				CTCCGAGG		DILDPSNYKQLRKDPTSKILKMTNQLVKASSFSSDDKKH				
				ATGCCTTCC		LCKTEAYSPRLYGLPKVHKQLVPLRPIVSAIGSPTHEVA				
				GCAGACGA		GHLASLLQPYVGGTDSHIKNSCHFIDIIKELSVDPDDILV				
				GGGCGAAA		SFDVVSLFTMIPVNEAITCIADIFPSDIVALFKHCLTTTYF				
				CGTCAGGG		QYKNVFYEQIDGVAMGSPLSPAVANLFMENFEQRALQT				
			GGGCGAA	GAAAGTTTC		ARKRPAKWYRYVDDTFVIWTHGEEELDAFLLHLNSINP				
			ACGTCAG	ATGTATGG		RIQFTMEKETNGQLNFLDVTVIKREDGSLGHKVYRKDT				
			GGGAAAG	ACCACGGC		HTDRYLQKDSNHHPRQKKRRDRLSNERTDHWVIRSIGK				
			TTTCATGT	ATCTCAGCC		THTLIVIFKKILTTTHDKKRGVIKTLVDRAHKICEPAHLD				
			ATGGACC	CGGAAGCT		DEIEHLRVTFKKNGYSDREIDRALRPRRTTRTDEERQQP				
			ACGGCAT	TTAAGTAAT		KGKVFLPFINKVTDRIGKILSKYDVETVFRPTTKIKEFLR				
			CTCAGCC	GACGACAC		SAKDKRHPLATPGVYKIPCSCGKVYIGTTKRSVNTRLGE				
		Locusta migratoria	CGGAAGC	CTGCCGTGA		HKRNCRLGHTDKSAVAEHVYQEGDHEIKFNETSVIAKT				
			TTTAAGT	AAGCCTTCA		SHYYARLYREAIEIFKHRNNFNKKEEGLKLDKIWMSTL	cd00304	435		526
			A	TTCTATGA		QQQHDNRLISIEEVGNSRDHSTADTT				
	Penelope-80_Mi	.								
Penelope										

Ving i	Vingi-15_LMi	.	Locusta migratoria	ACCAAGG							
				TGCTGCT							
				GCACAGC							
				AACAAGT							
				ATGCAAG							
				TAAACCT							
				ACAATGA							
				ATAGTCC							
				TTTTCAG							
				GACCAAC							
				CAAGCTC							
				AGAACAC							
				CCAGCTC							
				ATGGTAA							
				CAACTTA							
				AGAATAT							
				GTCACTT							
				GAACATG				MLVILSHSHVWGYQEEDRNGAILNEWMDMYNLQLVH			
				AGTCGAT				NAKDPGTFRSGRWQKDYSPDLCIISKSGNNNTTQRKVIK			
				TTCATCT				SFPHSQHRPVIIQFGTEIPLVKSVPLPRWNFQLANWEQFR			
				GCTAAGA				KSLDNSIQWIPPLPSNYERFLGAIKVNAKRYIPRGFRKEY			
				CTCAGAT				IPGWSEECDELYAEYQDSHSDSTADDLLRALNKNRREK			
				CTTTCAC				WHKTTAGLNFTHSSRKAWSLMRKLGTDPRLSNRVAHV			
				GCCTGAT				NANSVAARLMNISKSKVEKAYTRTVKKELHIKRSQLEP			
				GCTGAAT				HSDYSREFSIEEVETALKMVKVNKAPGFDGVYPEFLKA			
				CGGTGTG				AGPKTKLWLTDFFNAILRTGKLPKLFKQAKIIAILKPGK			
				CAATTCA				DGTDASHYRPVSLLSMTYKLLERLILNRIQETIDQIIPAE			
				GGAACCA				QGGFRKHRSCCDQVLTLTTLVETGFQRGLKSAAVFVDL			
				ACCAATC				TAAYGTVWREGLMLKLIQIIPCQKLAFLVNNMLSDRRF			
				TACAGAA				KVFLNGQSSRWRTLNNGLPQGSVLAPILFNLYIHDMPDT			
				AAGAAAT				NSTKIQYADDLALVFQSNDLKHCEQVLTSDLIKLSDYFH			
				CTTTTGG				KWRLQPNPNKTEVTAFHLNNKEAQRKLHVTFDGIEVPH			
				ATATGTT				NYNPVYLGITLDRSLTFKEQCLKLSKKLASRVNLLRKIA			
				CTGGCTG		TAATTGTAT		GSNWGADANTLRSAALALVYSAGEYCAPVWENSAHTT			
				GGCTATA		GTATATATT		KIDVQLNEAMRIITGTVRSTPIQWLPVLANIAPQDLRRK			
				AATGTGA		CTTGTATAC		AAKVKLIKNVVKAHKNSPLYDALQDPPACRLKSRKPPWC			
				CAGTAGG		TTGTGAAA		DTESTTSFDVTSKWQERWNENPPHNGHLVQDPSKKVE			
				CATTGCA		ACCATACG		GFHLPRQTWSRLNRIRTGQGRCAYNMKKWGFLPSSDC			
				ACATATG		ATAAATAA		DCGAEEQTVGHIVKQCPLRAFNGDLESLHQVTESALDW	cd01650	335	588
				TCAGATC		ATAAATAA		LSNLDISI			

				CATTTCTC						
				TACAAAA						
				TCATCAT						
				AAGACAA						
				ACATCAG						
				AGTCTGT						
				CTTGGCT						
				ATAGAGA						
				TGCAGGC						
				ACACTAT						
				TGCGAAT						
				ATCTACA						
				AACCTCC						
				GCTTATTT						
				GGCCTAA						
				TCCACTG						
				TTAGGTC						
				CTTCATC						
				CCCCAGT						
				TGTGT						
				CTTCCCA				MFYSCDDNRHEFGTGFLVHKYLKHLVMNFTPINPRLSVI		
				GGGGACA				RIKGRFFNYSIINGHAPTEESIEDLKDDFYDALEKAYDKC		
				CAGCCTT				PRNDIKIVLGDFNAQLGREDMFQPVIGAHSKHGSTNDN		
				GTAGTAG				GIRLILFAASRNMIVGSTSFPHKEIHKGTWKSPDGRTINQ		
				GCACCTT				IDHVLIDQRHKSNLLDVRTFRGPNVDSDHFLVMAKIRA		
				CAGTGCC				RISNAQKERNHRTKMAAIKRAIDISAKETLGIERESIRPT		
				GGGCGGG				WFDEECKRVTEEKNTAYQKTLQANCTRSLLEDYREKR		
				AGGGCTT				RIEKMTHRRKKREWTKAEIECMELSKKKKMKQGILKTIC		
				GCGTGCT				SITSLIRDEEGNIISEVNRLWKDGEDTLMTCESRDEQAED		
				TCCGTGA				ISNDCEQGLDDEQPLDGPGLEIKKAKSLRKLKNNKAPGT		
				AGTTGAG				DSIPAELLKYGGATLEQSLFELITLIWEQERIPNEWNEGIL		
				AGCTGTG				CPIHKKGDRMECRNHRGITLLNLGYKVLSNIIFDRLLPQ		
				CCGGCAG				VDKEIGMYQSGFRRGRSTIDQIFNLRQILEKTVEYRIGTH		
				TAGCGTA				HLFIDFKAAYDSINRNQLFHALQEMGIHGKLLRMVKMI		
				GCTACTG				VVQGNYSDPLEIKNGLRQGDALACLLFNVALEKVIREA		
				GCAGGGT				GLQTRGTIIFYKSVQILAYADDIIIARTVPALKEAFTALS		
				CACCCAA				QASKKMGLMINEDKTKYMACGKAYQPNMPSTFSVNG		
	RTE-56_LMi	.	Locusta migratoria	GCCGTAT	TAGCGCCA			YTFERVDTFKYLGSTVTCFNDTSTEIKIRLISANKAYFAL		
RTE				TGGTCTC	CAAAAGAA			SNLFQSRLLSRPTKFTIYKTLVRPVLTYASETWPLTETDA	cd01650	427
				AACTGAG	GAAGAAG			RLLDGFERKVLRRIIGPVYDGEQWRKGTMQNCTLYMP		677

				AAGCCAGACGAAGAGTGTCCCACAACATAGGGCAGGGAGGGCTCTAGAGGCGTAGTGAAGCCAGCTTCCCTAGGGGAGGAAACCTTATACAAATCCCGGTCCTCCAGGTTGGGGGTTGGGCTTGAGGCTAACAACCTCTTCCTGTAAAAAAAGTCAGGTTATGAAAACCCAAAGAGCAGAAGCGGACAGATCAATAGACGACGACATGGCACGGATATTGGACAACGAATGGTTAAAAAGCTGAAAACTGATCTTGAATTGGTACCTGGAATGTTCGATC		MLLWRTMEERYNAELYTIYADAPVTNIIRSGRIRWAGHVARMEESEVTKKILFGNPGGQRGRGRPRARWLDGVEEDLRKLGCRNWRRVAYDRDGWRKLVGKAKVHPGL			

Ving i	Vingi-31_LMi	.	Locusta migratoria	TTTATAT AGCCCAG GCGCTCT AAAGAAG TTATTAG ACCAGCA GGGGAAG CTCAAGG AGAGCAT TATTGCC ATCCAAG AAATTCG ATGGACA GGGACGG GGTATTG GAAAAGA AGGCACT	GGGTGCA ATGAAGG AGCCTCC CCAAAGA TTGCGAC AGCAAGA TAAAAAA AAAACCG CAATCGG GCGCACC CCGGGCA ACGTTGA TCATCAG CGGCTAA TCCTTCA ACAACCG GAACGGA CTTCTGA GAA	TAGTTCCGG ACACGGAA AGTAAGTA AGTA	MKTNKTPPKRSFPRPLVTISINIEGISRDKEVLLSDLCKEY QCDFLLIQETHRGITRPSPKIDGMKMILERPHEQYGSAIF ARPGINVTSASLTDRDDVEILTVRTQQYSVTSVYKPPDS TFVFTEPANFSYQDINFILGDFNCQSTAWGYKETNSSGA ELEKWAEVHDLLLIHDPKLPSSFNSGRWKKGYNPDNIF VSKKLAAQSEKLIGEPIPHTQHRPIICTVSAVVKPEQVPF KRRFNFKKANWKKFREIIDLEVTKILHNLRCMTLLSVFR MCLGRQYARMQNTVYRWTSEDAKPLLERYQSSSTGNE LMSVIAENRKAKWRDLVENLDMKVNSRGLGNCRTSTV MPQSQTINLLMSLQTSRHHTLLNGKPRQKIPQPLTRVLA EKPHLKAPFTMSELTAAISSLKTNRAPGIDELRVSKLKTL AVIPYSGCTESQQIKNLGSNTLQWLLDLMNACVERLHIP KMWRKARVVALLKPGKDPNDPKNFRPVSLLCHCFKILE RLTLSRIENVLEQQIIPQQAGFRQGKSCCGQVLNLTQHIE DGFERKEITGVAFIDLTAAYDTVNHRRLLRKVYDVTKD YHLTSLIATFLQNRRFFVCLQGKKSRWRNQRNGLPQGS VLAPALYNIYTNDQPVPNTTRQFMYADDTAVAAQGTN FLDVEEKLSSTLESLSQYYEANHLRPNPSKTQVCAFHLK NAEANRELNVTWRGERLAHCPTPKYLGVTLDRTLTYK HHCNATREKVSARNNIIRKLTNGSWGANPKVLRTSALA LSVSAAEYAAPVWSSSTHARKVDIAVNEMARIVTGCLK PTPVHKMYLIMGIAPPNIRRDIAAEAEKTKQEKDMRHPL YGHQPADRRLRSRKSFLSRTRALEGSAAENRIARWEMD	cd01650	473	728

						LRSPHKAAKEELAPGGDLPYTVWRTLNRMRVEVPKCK TNMQRWGFLKDNENTLCLCGAVQDPQHLLICPYLEKPC TMEDLWEANDKAISAANFWKLEV			
Daphne	Daphne-7_LMi	.	Locusta migratoria	GGCGGCT AGCGTGA AAATACG GAATGTG ACTCCGC GAAAAAA CGTGAAT TTTAGTG TTAAAAA TAAACAG TGTGTAA ATTGTAA AGCCGAA ATTCAAA GCTAAAT GAACGTG TGTCCAG TTTACAG CTTATTAT CAGCGCA CTGAAGA TGGATAT TACGAAA CTTAAAG ATGAAAT CAGTGAA CTAAACA CATCAAG ATCGCGC CATGATA TTTCGAC TCTGTCG GAAAAGT TGGCGGA AGTAAAG TTAAAGG ATGCAAA	TAATTTAAA GGTGAACT GATGTAGT GCTGGAAA CAGAGAAG TGGACTGTG TCTTTTCAC TAAATTTCT AATACTTAA ATACCACAT GTTATATTT ATAAATGA TAAAATTTGT AAAAATTTTC TTTCTTTCT TATGTTATA TAAAATTA ACAACATG ACTTGTGTC TAATATCAT TGTAAATAT GATCTACA GACAAATA AATTATTAT TATTATT	MESKIDVFGDSHAKGISQELQVISKSSTVTGMVKPGAGF KEVTKNIVKYKYGKQDSVVICAGAKDIYENNAMKMYK QLLSLILVLSNTNIILVNFXLRHDLPEWSCVNKEIHRINE KLKSCCRLFSNVKLIDSSTLHRECFTKHGLHLNNNGKAR LGNLIRDAIKCNSKWTVTELGWYKCPTAETPAQTATAP RSTIQKTVATPTEQPTTEPKLTATSEPTSEVTVKTSKFNQ PESSAFTTTVAEQTSVAVAPKQITSQPESLAYTATTAVA VVVAAVVTPEEGSHLQQSSHQKKAATFEKPGFLLGPGS EQQQIIGKHLNDCNQKNLQQGQLVQEADSNQGRESGNI NRTAVSKNSNLKIMHQNIQYASKKSEELEVLLAAEVPNI FIMTEHGLKDSELKVFNLEGYKNVAYYCRNEHKSGGV AIFVDSKVGTTPLDISQIAQISTLKQQQYKSSWGGKCLII GIYRSPSSSIENFLLELEVLLQKVMKKYTFVILAGDLNID MLRDNDCTRRFQDILGSYNLQLHVKSPTRCTETSSTLID HILTNIPTERQKVQVINTVIADHSGQTIELKDYAKAEDK KMIKYRQTNIANLNSLEILVKAEEWDDVYNEHSVEKSW ESFCDILRHHINVACPVICKTVSAKQNKKWITKEIKKAR EELIKLYEDYQKDKEQASQEAYRSKKKLYTRMIREAKA SQYKHEIKTSTNISKTVWNLINDNRKESQTSKDTNIILEM EGVKVTDPLQVANSLNQHYVAIVEKLLNQXTPVTKSIK SPKWVAESLYLLPTNELEVSKIIGALKNKNTEGLDGISA KIVKHCAKYLATPLTFLINLTMKEGVFPNCLKRSKVVPV FKGGDKEDPGNYRPITITTVLSKIVERVINDRIVNFIDKH NILNEAQHGFRKGRSTKTAVANFITYITEALEQRRKVSGI YLDLSKAFDCVCHKILLKKLNCYGIRGNAADIIKTFAEN RQQCVELKHKCRNKEEKVYSEFQQVKYGVAQGSILGPL LFILYVNDMPTEINEKVIMYADDTLLVCGDTVKDLEK RATENFQMAERYFTENNLLINRKKTKCTKFETKITKSSE VFTLKIGSADLAEVDCNRFLGIEVDKHLSWENHIDKVCS KISTNIFLMRKMCYTWTQKPCGLFENKHKYILNEKNVL YVDTETLLKMYYGLIFPHLSYGIPLWGGAADVHIQRLL KLQKKAIRCCTHTETFKTSKESHQMLAKLSPRESCKNKF KEFKILTVPSLYIYMRQDTNCTKSVYIYETIVFLMVNCT APLNSDIHNHNTRTKEQYHFFSKRLNMTDRDPINAGCRF FNKLPLSIRCIKERSAFKRKLKEHLISGCWYKVKEYLTE NL	cd01650	847	111 1

RTE	RTE-7_LMi	.	Locusta migratoria	CAGGGAC AAGCATC TTGCGTG ACGTGGT GGCGAGC ATGGGGA CCCAAGC TGGCCTG GCATTGC TTCCCCCTT ACCTGTG	TGACAAGA AGAGGATC AACCCATG ACAAGACC TGCCGCAA GGCAGAAT ACAATTGAT GATGATGA T	MDRNKLDVLGMCEVRWGGCGELESGDSDCTTXVKIK MECMELEYLRLYYSXEDKNGMYGVGILIGKRLKDKVM KVEYINGRIMMIRLKGKKEDLVIIQVYMPTSNHSEEEVE ECYEKIENLAEREKRNACIVVMGDWNAVVGEGSEGKA LGKFGLGERNERGQRLLEFCNENLMVVGNTLFDNHKR RRYTWISRFDNKRHQIDYIMIQQRYRNCLKSAKSYPGA DIYSDHILVVAEVLVKLKRTWKGQKREFINIEKLKDKGI CEKLENAFSEIIMPGQEMECIERRWDHYKNGLKRQLRK RKWVKKAAKETLGVKERKRVKKEWITPEMISKMDDRR KWKTVNTEEGKRNYRRLNNELRRETRKARVKWLTEKC DKIENLZKEGKYDLMYKEAMDLVFREKNSNSGIKEIEA	cd01650	523	779

				CCAGGTC CCTCCTG TTATCTTT CCTATCG ACCACCC TCGGTCA ACTCTTG TTCATTTC CGACCCC GACGGTA TTAAGCT ATGGAGT CTTCATTT TCCTCTCT CAGTGCA AGGTAAC AGGAAAC CACCACT GACATTA TTTTCTCT AGACGCT CATGGGG TGACCTC TTTGTAA AAGATTT GGAGGTA AAGCTTC CCCTCAA TCGGATC TCCGGGA GGGGAAT ACAACGA GGTTTCC AGACATA AGGAGAG ATGGAAC AGGAAGG AAGGAGA AGGAAAA AGAAGAA		ADGKMVNETQEVMRRWEEYIEWLYDGDTRPSEDELGL EEENTVAEDDKGNAILTSEVEASVKEMKNGKAEGIDEIP VELLKSLQREGKKELIGLYNQIYMTGNWPKDFCESILIPI EKKKNSKKCEEHRTVSLISHTAKVLLRTLNRRLYGKLN AVMGEEQFGFRRGVGTRDAIGLLRVIGERYMEKKRKV YAIFIDLEKAFDRVRWDKLMAILRKHGVDWRDRKLITN LYLNQTARVRIGGVISEEIELGRGVRQGCCLSPTLFNIYL EDIIKESFNGSRGVCIGGRRVECIRFADDMVVMAESAKE MEQMLDDLNTKLEEYGMKINKKKTKSMVIGGRNTKCR LKIGGEEIEQVHSFKYLGSMIMDDMYCTAEIKKRIAMA KEGFKRKQRLLCGPLKKDLRKRLAKCYIWSVVLYGAET WTLRKEDERRLEALEMWIWRRMEQVKWEDRIRNEEVL RRVGEERNMLEVIRKRKRNWIGHCLRRDCLLKEGIEGM VKGKRGRGKPRYQMLDNIQGNKYEEMKRLALDREKW KRINP			

			TTATTGA GGATAGG AACGTGG AATGTTT TGACCCT TTTGCAG GAAGGGA AATTAGA AAACGCT AAACATG AA			MRMSVNKRSFSEPALKMTSVNIEGLSANKELLLSDICSK CKCDVLILQETHRGPTSHRPNIKGMILTLERPHEQYGSAI FTKPGLMVRSAAMTHVDDIEILTVELKACTVTSLYKPPG ADFTFREPSNFQTQPLKFITGDFNSHSTAWGYQETDSNG EKVERWAESSGLKLVHDPKLPASFNSGRWRRGYNPDNI FVTEQIAQQSFKSVEDALPRTQHRPISLTVKAVVCPVTV PFKRRFNFKKARWRLFTEELDSKIDEILPAACSYDKFIQL IRRVSRKYIPRGCRNEYIAGLDDNIEILLKKYEELYNTDP FSEDTLETGEDLLRALTTTRREKWCDLVTSLDFKQNSR QAWSLLRNLSNDPTSSKTLHTSVTPNQIAHQLLLNGKTT HRSPRPSIVRDMENEHHSLEAPFTLEELQAAIDKMKNNK AAGEDDLRVEQIKHFGPKTLRWLTHLMKQLRQHHADS QSVEKSKSGSPIKTRKGANRCQKLSPSILAVPSVQGPRE NGANNCVNTMQIPRVWRKARVVALLKPGKEPTDVKNF RPVSLLCHLFKALERMVLNRISNHVDEVLSPEQAGFRPG RSCTGQILNLTQHIEDGYERKQVTGVAFVDLSAAYDTIN HSRLVHKVYQATKDFGLSRFVQCILQNRRFYVTLQGKS SRWRIQKMVLLRAVYGLWTFKICAVHTTEQAFLCHTAR KKQQMENPKNGLAQGSVLAPVLFNIYTNDQPMHPDVR QFIYADDTAVAAQGSTFEEVEEKLTRALELLGNYYHQN HLKPNPAKTQVCSFHLRNKEAKTKLHVSWKGQELEHC YAPKYLGVKLDRTLSFKQHCLDTKKKVSTRNSIIRKLTN TNWGADPQVLRTTAMALSLSAAEYAAPVVWKNSAHVK HIDVAVNESARIISGCLRPTPTEKIYPLIGIAPPSIRRQVAA EVERKKQLTDNRHPLHGHQQPRRLRSRRSFLWTSEPID VTVEERRLQLWKEATPNPILALKEQLAAGGNKSFTTWK ALNRLRTGVSNCKFNLSRWGYLDGEPLCDCGNVQTDD HLLTCPNLSTCCTPLDLQLANDKAICVADYWARNI	cd016 50	524	819
Ving i	Vingi - 37_L Mi	.	Locusta migratori a	AACACCA GAACACC AG	TAAATTTTA AAATTTATT ACATGGAC TCAAACTA GATGTAAA GGATGGTG TTCTGGACA CGAAAAAT AAAATAAA T				

RTE	RTE-21_LMi	.	Locusta migratoria	CCTGGTCAACGGTCTCAAAGGCGGAAGAGGAATCTTAGTTCCCAAGGAGAAGAGCGGAAGAACTAATGGAGAAAACCATGAAAAATAATCATTTCGGACTAACAATCCGATCTTATCTTGGAATTTACTTCCTCCCTGTACA	TGAACAGCAACAGGTTAACCAAACAGATTTTTAAATATCTCTGGGACAAGAAGTCAACAATAGCTGGATTCAAGAAGTTAGGAAAGATTTGGAAAGAAACAATATAAGTGAAAAAGATGCAACAGAAAAGAGAAACTTTTAAGAAGAAAGTGTTAAAAATGGAAGGATTCCAAGGCAGGAGGGAGAAGAAGACAGGAGTAAAGTGGTCTGAGGAGAGAAAAAGGATGCATAGTGAAAAGATGAAAGAATACTGGAGGAAAAAGAAAGAAGAACAAAGGAGGAAACATTGAAATTGGTGCGTGGTCCTTAGATGGCCGAA	MYNFNCRMESPLKVSTTPGGNSNEESTIALCNASGPRVLGSPNICIKGESEHPWYPSKLKFNRKFFAGTLNVNSLLKIGKQKELEILLDKHEIQILAVQETRFLDENITETKNYRIFKGKPAIKIMKGMPILGTAFCVHKNLVDKITEFKSSSERLSLLTIKCKNKQYTFVNCHAPINDDNKKNLSKVEAFWELLEDEISKIPKSNVVILLGDFNAQIGREKTFRKIVGEYPAHKRTNKNGTRLIGLCKTYXLKLMSTFFKKLPRKAKTWISPNPMLGEFQLDHVAISKHNIKEIMNVKVTRNREFDSDHYLTKIKMKLLPXKIQKRQQRLIKYNTDRLTITKDAIKHFQEEIKISKSGKWEELSKEINRVAKNVFGATKIMKKAWWNETCEEAVAERTKKWKKWKCSGKTEHLDEFKIQRKETARIIKSTKNFYQMFKNHLKGYQAPSICFRDENGKVGLSNXHNCEILRKYFEQLLNCEEPNCKLEFPDLHENTETDPPPTAEEIKKAINALKNNKASGEDSITAELIKWSQPKIISNLQLMFEEIWETEKIPEDWKVALIHPLHKKGDKQDVNNYRGISLLPVGYKIFSTIILNKVEATLDSHLGEYQSGFRKGRSCSEQIFNLKSIIRHRLTNSKDIVVTFIDFKKAFDSVDRETLDKVIREFGIKSKLANLIRETLTDTKSKVKLQGEISKAFDIKTGVRQGDGLSPLLFNCVLEKIVREWKQKLKDHKLSPIRLGTKHKGIEIHCLAFADDFAILSENLTNAEIQINLLEEIANKAGLRISVEKTKFMTNIKNAPEILVTDIGXIKRVNKFKYLGEMIQENGLEKFAVEERVHKMERAYGITRNIYNKRSLSKNLKIQHYNTVVKPECLYASECLVLNYKLHKLEVLERKIIRKILGPLRNTEVWKLRSNAEVYRNIENITETLRKRRLLFFGHLYE	cd01650	561	821

					ACGAAGAA AGAAGAAA				
Ving i	Vingi-42_LMi	.	Locusta migratoria	GCTAATA					
				AGCATAA					
				ACATTGA					
				GCGCTTG					
				GCCAACG					
				ATAAGCA					
				GGAGCTG					
				CTCTCGG					
				AATGTGC					
				AGAAATA					
				CAAATGT					
				CACATTC					
				TGTGCAT					
				CAGAGAC					
				CCAGACA					
				CCAATTA					
				ACCTCGT					
				CCAACAT	TGACATGG	RNGQMLIILYLFMTQNCHHHFTVKYGNKALIRISASAVL			
				ACCAGGA	ATTCGGAC	IWRQXASRQFITHSQRHNIDQLATSWPECLRRHKTADD			
				ATGAAGC	ACGAAAAA	TSRRRFNFRKANWTDFTSTLDNCIDELEPIPGNYERFVE			
				TTGTAAT	GTAATGT	KVRYISRITIPRGCRQQYIPGLSGESKGLLDQYNEMYEK			
				TGAGGTA		DPFSNETEEVGNLLLKELSRTRQEKWVETIERLDMTXSS			
				TCACACA		KMAWNMMKKLSADPIAPKQASNTTADQIAHQLLISGK			
				AAAAAAT		TGTKPRSRKIHRNLESEVSLIDTPFGIKDLDNSIQSLKTGK			
				ATGGTAG		AAGVDDICTEQIKHFGKNKTVDPTTNKVLLNKQIPKSW			
				TACAATC		RKSKVIALLKPSKDPNLASSYRPISLLCHHSKFMSGNSEL			
				TATGTAG		CGRKTNSRTRRFPPRKIMYRAGIGADRTHRERLRGENIT			
				AACCGGA		GVVFVDLSAAYDTINQRTLLAKVYNITKDYELTKVIESL			
				CCAAATT		LQNRRFYVTLEGKKSRWRTQKNGLPKAVFCPQCCSTFT			
				ATAACAT		QTTNLLLLVIQSTSFMLMTWPLGHSSSGQKLPAVEERLQ			
				CAGTGGG		NALSQMSVYYKENSLKPNPAKTQACAFHLKSKEARRK			
				CTCAACA		LQIKWDGFSIEHTAQPKYLGVILDRSLTYKSHCQKTRQK			
				AACATGA		TATRNQLLKKLIGSKWGASPHVIRTTAEAVCFSTAEYAC			
				CAATTTA		PVWNRSAHAKEVNIALNETCRLMTGCLKATPLGQLYK			
				GAAATTC		AAGCEAPATRRLAVEHTERFRQAFDDRHPLFEAPEPLRR			
				TAACAGT		LKSRRSFLRSVPMEPPDKYPLTQKPPPGYHLDWRSWKS			
				GGAAACT		LNRLRTGVAPVKTNLVKWGIIDESDVRCNCGERQNME	IPR0 00477	299	551
				GACAATG		HLLTCKNCPRTCSLDDLWNARDNALDVARYWADVL			

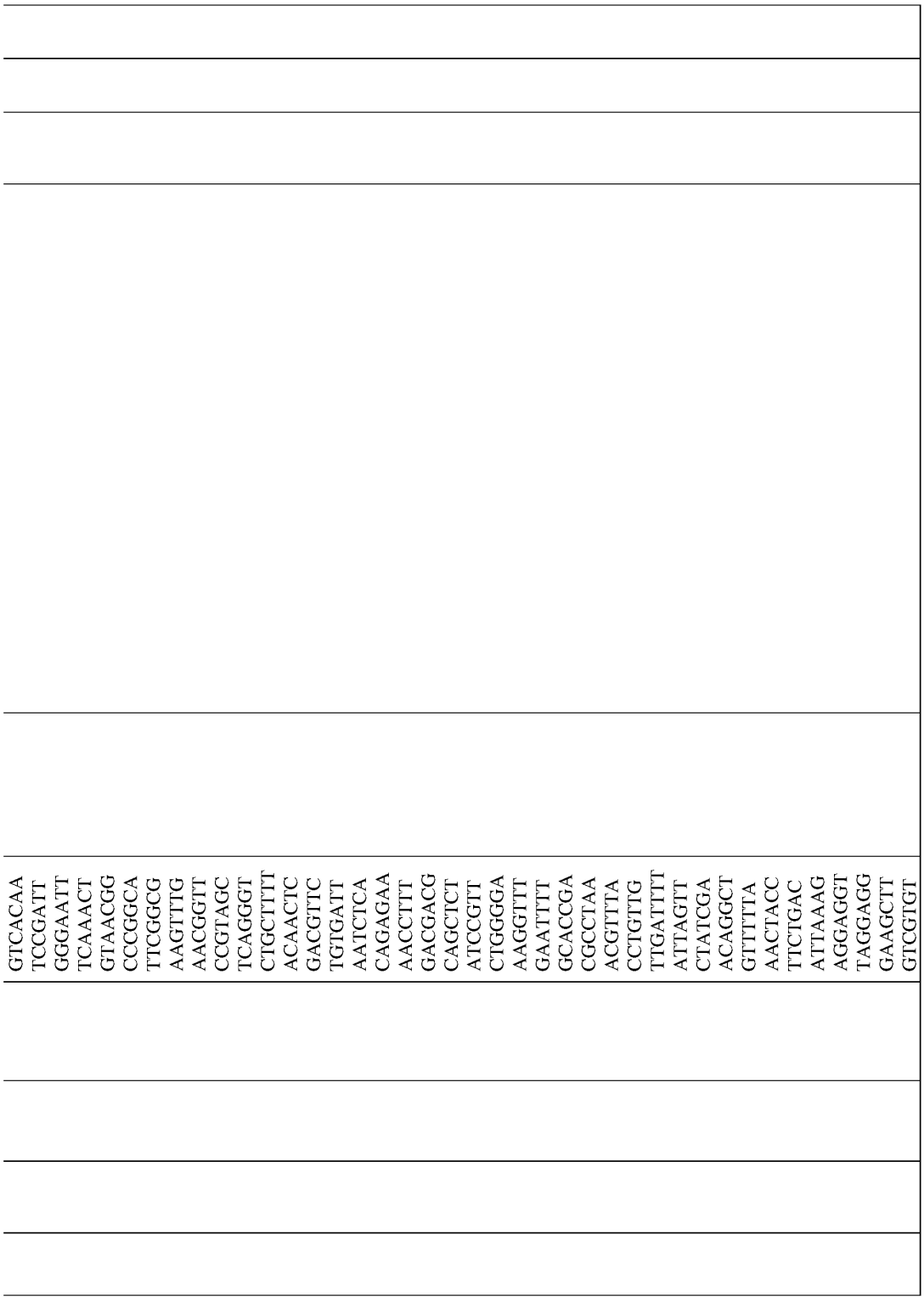
				CACAATAACAAGTGCGTACAAGCCCCTAGCCAAAAATTCATATTAATGAACACATAATTTTAGAACAAAATAACAACTGCTTCATCCTGGGTGATTTTAACAGCCATAGCACTACATGGGGATATGAAGAACTAACACCAATGGAAGTCTTGTAG					
Penelope	Penelope-30_LMi	.	Locusta migratoria	TCTTAAACGAAAATGGCTAAAAACTACGGCCAACATAAATTTCAACAAGGCTTGTGTAAAGAATAAATGTATCCCGAACTATATCAAGTTTAAGATATGTAATCCAAAGTCCAATGTAG	TAAATGAACAAGTTCAAGCAATACATACAGCTTTTTTAAAAACTTTGTGGACCTATTATAACAAAAGTCGTTTATTTTAAATTTCAACCTCTGTCTTACTGTGCAAATGATTTCGTGACACCTGAATGTACCCACTAGCTTGTGGATG	MIWDTVKSELDEQLQVEFLNLKSRQDQKLKKLMRQIDCRVNNAQNVDNFQFYPRVVNETDISFNDEELRLLNKGLKFNVQPSLNDKTVKELIVATKMAVDVAKLNENEKGHVGYKLANAIKAETNKSCNSMHNNSDHKTVTSINDKLISNKAIVTKADKGNTTVIVSEEQYINKTLQFFSENGVTEIAKDPTKKVQAELKQLLKRTSFLLTEKEKFQCYVMNPTAPRLRSQLKVHKNGHPIRAIVNSMKGPLCKITKKLDKILKEFYVFGDNFSVSNNTTLINEIKDKPIADTSHLMSLDIKNLYANVPVDEAVELIKANLLKHKRLSQQEIVEFIDLLQFILKYNYFAFNGKFYIQKDGLAMGSSISGTIADIFINNLENRLFTNMQHPVDKILYYKRYVDDTLLLIDGGQTEITEIVNKFNSLHEKIEFTVEQEKDGTLQFLDLNITKQDGRHLFKVYRKPTTTDILIHNSSCHPNGQKLSALRCLVDRAFKVPMNEKDRTTEFNIIRQLAKVNGYQENVVNNLIRKKQRAKSNKNGTXNIKRVMVPYFGNLSQKVANILKPFKVEAAFKTNNTLHHKLRNNLNQETDKYDKAGVYKIQCSSCNANYIGQTGRSFKV	cd00304	353	448

				GTAGGGC AGTGGTT GATAAAG CGCGGAA ATTTTGG TTGAATT CCGAAAT TAAATTC CTCCATA AGAAAAA AGATTGT ATTAGTG GGCAATT ACTTAAG GCGCATT TAGAAGC AGCTAGT GCGGTAC ACA	ATCTTTGGA ATTTAGAG GTCTTTGTT TTCATGTGC ATACGTGC GCCCCACTC GTGTTGTTG ACTCCGTGC ATTTCGCTA GTGTTATGC GCCGCCAG TGCTGGGGT CTGTTTGAA TCCTGACGA AGCACAGT ATGCTGAG GACCTTTCT GCGTAGAC ACGGTTAC AAGGGAAG TAACCACGT TTTCTTTTA TTACGTTTT AGATTTTTA TTGTCTCAG GTTTTAAGA TTCTTGTAT AGATTATCT GATGATGA GGGCTTGA CCCTTGAAA CGCGTCATA TTTTAGTGT TTTTACACG ATCTCTGTG GTTGAGCCT ACTTGACA AAATAAT	RYKEHMDAFRLNNFDKSAVAMHIRETGHPLKRIEDDLK ILHKQDKGRKLDLLEQLEITIHNAKATFH			

Nimb	Nimb-19_LMi	.	Locusta migratoria	CCCCTCAGGGGCTCAGCATTAATTTGCTGAGTACGGGCTTGGCGAACCCGGGGTCCTGAGCCGTGGGCTGTGTCGCGCCAACAGTCCGCTACACCGTAAGCTCTGTGCGTGCTTCAACGACCATCGTTCGGCGCGACGGTGGAACGTTGTACGTTTCAGAGCCCGGGGATCTTGGCTTAATCGCCCGGGTCGCGAGGACGGTACAAACCTCTATAAAAAAACCTCAATCTCAAGGTGTGCTACGTGCCGAGGAGATGCATGGCTGTTGAGGTTGAACAGT	TAATGGGAAGCCTGATATTTCAGGCACCCACCATTCAAACGGTTTTAATCTTTTCGTCCTGTCAGTCCACTTTGACTGGCAGTTTAGTGTGTGCAGAATGGCTGGCTCATCCTATATTATCTTTGCGATCAGCCAGCCACATTCTTCTGCTGTGATATTTTATGTTGTTGTCTGTTTGGTCAACTACTTTGTGTTTTCCTTTAAATGTTGTCTCTTAGCACGCATTATCACTCTGACCGTCTCTGCATTAGACGTGGAGTACTTTTTACCGCTAATTAAAAGATGTGCTCTCTTGGCTAGGAACAAGGGACTGATAACCGTTGCTGTTTAGCCCCCTTTCAAAAACCCAAC	MALPGAKPASLSVXGPNLMLRHMTPRRSLPWPRHGRNAGREIEVRLMPPXTWYVPGLTANRLPPRSPCFLWKTLRISSGRWMACQDEIWXCXHXDVVPSPVTGTACLVRGGGWHVKMRSGXVLITTSSPAQSRALLACDKLGDVPVSITPHKSLNMVQGIIFHKDLLLQXDDELRKNLERRGVHFIRRVHRGPKDNRVATGAFILAFEGDSLPEKVKVMIYRCDVKPYVPPPMRCFKCWRFGHMSSRCXGNPTCRECGRAXHPDAPCAPPPVCVNCGQHHSPCSPXCTVLHKERKIMEYKTLDRLSYSEAKRKYXKLHPVSMKTSMPQLHHTASAISATALHCLVFHSGPPGXTAHJCPHGLWGHLNLHRPQGPYFGSSNHHSTCGXISPHPPPLHLWGHQSPPLSQRHNSXLWLLSRGRGPSDPFLLRLLLLKDTSRWLKEPPTAGRRASRAPQFLGTEPAEPSQHEKPKEQRETKEKDKKPPKRRIPWWPPCHHPPPPLLLRTRETTQEKDSLVATVPPPPSTSSSEDEIEILASPKDLAIAESSEAMDVEALPQKSATAEDTVNPTAPSKSACPSDQEALGSASKIGDSRRYCEPYCAVEISMSQRPGGSLILQWNCGGFFRHLPELQQLLCRFPAFCVALQETWFPAQRTPALRGYRGYYKNRTIYDRVSGGVCIYVLATVPSAAVPLHTTLEAVAVRIRAMQELTICSLYLPPDEMVSLTDLAALVAQLPPPFLLLGDFNAHHPLWGGAQTTGRGRDIETLLSQQDICLLNTGAPTHFCVARGTYSAIDLSVCTPSLLPHFQWRVHDDLCNSDHFPVLMSLLDFQPAGCPSRWSLRRADWQRYTSAATVPVPLPADVDMAVHQVTSTIVSAAEATIPTTSGSPRRKSVPWWTPEIAEAIRDRRRALQRHRRHPTVENLIAFKRLRARARQLIRQSKRDCWETYVASIGCRASRLGCGPRLGAFLGTVHQQRSPAYTSRVTYAPTRLPSLNTWQSIMLAPLQHQNTVQTSVHTGLLGDVCRQHRLSRLASRVWSQVRRLFGYRPPAAIPSLHIQGNLCTDPTSIAEHMAEHYARTSATSEYSADFRALKERVEQRRLSFHTRRAESYNSAFSEWEFETALAACPDTSPGPDRIHSQFIRHLSVEFQHQLLALFNRLWAEGAFPSQWRESIIIPVIKPGADPQVADSYRPISLTNVLCKLFERMVSRRLMWILEHRGLLSPHQGGFRQGRSTVDNLVALESAIRTAFSRREHLVAVFFDLRKAYDTTWRHHILATLHGWGLRGPLPLFIQNFLSFRSFRVRVGTSLSSPRVQENGVPQGSVLSVPLFLIAINGLAEAVGSSVSSSLFADDFCLYFSSASIGVAERRLQGAIRKAQSWASTHGFQFSPSKTRVMHFCRRRAVHLDPELYLYDELLRVEETHRFLGMLFDSRLTWVPHLRQLKQKCVRYLNILRCLSHTSWGAARTSLLQLYRALIQSRLDYGSFVYGSAAPSALQVLDPVLHSGVRLATGAFRTSPVASLLAEAGVLPLRFRRQQLLSS	cd01650	1172	1430

				CGCTAGCGAGCAACCTCTGAGGAACCTGCCGCACCTCAGTTGTATAAGGCTTACCCAGGCAAGCGGGGCTCTGCCTGGGTGGACCTTTAGTTCCCTAGCTGCTCGTGGGACGAGAATGAAACGAGCAAATA	AACAACAACAAC	YAAHVHATPHHPNRRLLFPAAALPSPRQRPRSGLPIAVRVRSLLSSLESLPIPPTFHCPYRTAPWSLPRPQLRLDLSHASKASVAPLSLRRQFQAVLESFCGAEVVYTDGSVVDGRTGYAYVHGEHIEQHSLPCGSSVFTAELVAIYRALEHLRACPGESLVICSDSLSGLQALAKCFPKHPLVQSIQESVSRLVRSGRSVVFVWTPGHIGIPGNERVDQLAKQATTSSPLAIGLPESDLRLEFHRKILGLWNSEWANSAQPNKLRAVKLSTVAWKSSLRSSRRESVVLCRLRIGHTWLTHGHLLRREDAPQCDCGAPLTVVHILVDCPLLVALRQSFNLPDALSFMLGDDAAMVERVLDFIRAAGFYHVI			
Penelope	Penelope-1_LMi	.	Locusta migratoria	TCTCCTGGCGAATAACAACTTCGAAAAACTACGGGTATGCTGCCAGGTATCAAAATCGACGTCAAATTTTTCGAAGCTGTTATTCGCCAGGAGAAATTAAGGAATCCATTACTCTACGGGGAGGAAATGTATCGTTTTTGCGGAAGTCA	TGATGTTTATTATCGATATTCTGACCGGTTTCATCTTTGGTCGTGTCGGCGCTGTGTGCTTGCTTGCTCGGTGTGTGTTGTCTTTCCTGTTTACTCATTGAAAACCGAGGTTTTAAAGATCGTTGCACATTGCTTCCTCACTGCAGTTTCGCCTTGAAAATGGCAGGGTGGTCTCCTGCCGAAATA	MDPDIVILPADKGNATVVLNKQDYLQKMQCLLSDSAYRRISADPTKSVERKTNNLLKKSSLSQETIKSLNSYCAVPPRLYGLPKVHKDGCPLRPIVSNIGAPTYRVAKHLAGLLSPLVGRCEHHIKNSVDFLRRLEGLHLADSDILVSFDVVSLFTRVPLSDSLQLIETKFGTDLTNLFRHVLTSTYFLFNDEYYEQTDGVAMGSPLSPIIANLFMEDFEDRALESAALKPACFFRYVDDTFVVWPHGSENLNYFLEHLNSIHPNIRFTMEVEKNGCLPFLDVLVRRKADGTLGHAVYRKPTHTDLYLQADSCHHPAQREGVLRTLVHRAHVISDPESLPTELSHLEVTFRQNGYGEGQIKRALRYRPHVQRASDDSTESAPTSVAFLPYVGSTSNKIGRILRKNNVKCVFRPPSKIKALLGSVKDDLGLRKAGVYRIPCSCGMSYVGQTIRTVEDRCIEHKRHTRLQQPNKSAVAEHCLDTGHPMEYNKTEILACTSNYWDCVIKEAIEIKLGDNLMNRDGGFCLNSAWDPALSLVKKQRDKTNPTSSVEN	cd00304	201	285

				GATAAGC TTCGTAG CAGCAGG GTCAGGT TGCAATG TGCATGT CGTCTTT GTTGGAA TGTCGTG ATAATAG TGTTGTTC CAACTTT TGCCAAA GTTGTGC ACATATC AATTCTC CTGCCGC TGGTCGC ATTAAAC AACGTGC AGGCTTA GCTCTTG TCCGCGA GAGGATT CATTACA CCCGCCG AGAATTG GATTTTG TGTCCAA GGAATTA TTTCGGC TTCATTTG TACATTG CCTCTGA AATTTCG GAATTTT CTTGGGA CTGGTTG GACGGTG CTTCCTG	TCGGCCGTC GTCGATTTT GGTACCTG GCAGCATA CCCGTAAGT TTTTCGAAG TTGTT				


GTCACAA
TCCGATT
GGGAATT
TCAAACT
GTAACGG
CCCGGCA
TTCGGCG
AAGTTTG
AACGGTT
CCGTAGC
TCAGGGT
CTGCTTTT
ACAACTC
GACGTTC
TGTGATT
AATCTCA
CAGAGAA
AACCTTT
GACGACG
CAGCTCT
ATCCGTT
CTGGGGA
AAGGTTT
GAATTTT
GCACCGA
CGCCTAA
ACGTTTA
CCTGTTG
TTGATTTT
ATTAGTT
CTATCGA
ACAGGCT
GTTTTTA
AACTACC
TTCTGAC
ATTAAAG
AGGAGGT
TAGGAGG
GAAGCTT
GTCGTGT

Dap hne	Daph ne-26B_LMi	.	Locusta migratoria	GTTGACT AGGGCAC GTCCGCC TGTGAGT AATATAA CAGCAGC TGAGGCA GCTGCCT TACGTTC ACTTAGG	TGCATGA ACAGTGT AATACTT TAAGAGC AATATTT TTGGACC TAAATTC ATTTACA GTAAGTG TCTAGCA CAAGATG GACTGCA TCTAAAT AGGCTAG GCACAAA TACTTGT AGCAAAG TATTGGA GATGTAT GTGTGTG ATAAAAA CAAGGGA AACTAAT GTATAGG CAGGGGG TGAAAAT CCAGACA AGAATAC AAAAAGA CTACAAC	TAGAAATG TAGGTCTAT GATAAAGT GTGTAGAG AAAATGTA AATTTACTA TTAGTTATT ACGATGTAT TTATTTATA TAAAAAAT GTAATCATG CTGTGTAAA AGTGGCAA ATATCCATT TTATTCAAT ATGTAGAG TAACAAAA ATTCTAGCA GTGAAAAA CATTATGTG GTAAATCGT ACCAAGTA AAATGTTA ATCTACTTT TAATATTAC AACACAAG AAAAGAAA TATGTTAGG CCAGAAAT GGTCCAAT	MLFHQNVQSLNNKLSEIKIMLSNELIDVSVLCMSEHWL RDDALQTVSILNYELAAYTCRQQILGGGICIFVKKGLKK SFFFFLFFFFFFEISAIELPSLRDNFNMHLQIPNGDIKTFY QNLENTLTQTVKNKQTIIICGDFNIDFKGSAKEKTDLMNI LVSHTLKITIHSPTREIVTLSLQVYNRSNNNKPTKEQYKA EAVKTGLGDHHGQTISFHIKGIEVNKLTSTYKTRTYNDQ EIEIFKNLLKSESWEQIYDTSTENISEKFDIFIEKIKYYFDV VFPVKTRRLKDTLPKKIPSQNKWITTGIRISCANKRFLYR LTKTCDTTPAFDRYFKNYKTVLRKVIKSAKKLENDRYIS NSSNKMKAVWNIVKKETGNNKTKNKNIQLKENQQTTS KPKEVANMFNTYFINIAEKLIEKNVIKNPSQLATKIKHID KTIFLTPTNTKEVQQEIKLLKSKHSSGIDNIPDFIVKKCAD FLSEPLTYLINLSFSTGSFPKCLKIAKVKPLHKKGSQEEV SNYRPVSQLSAFSKJLEKLFIKDCPISQNPNMVLEKNHST ETAVFSFLYETLQALDRREHTTSIFLDLSKAFDTINHNKL LSKIQNWGLGAAHXXIKSNPHDRKQFQKAVLIEITYESN GTVSTYHSDSLVVKHGVPQGSILGPILFLLYVNDLYINN TNGKLYQFADDTSILITGSTKEIFESNIAAITKNIRDYFDA NGLIINVEKPLQWIYLRHRIDHQLTQILICKWPDNKCGK TVAMDIFTTQNRSPANPNIDINGKTIEKTECTKFLGIWIQ QNIKWDSHVDYINRKLSKTCYIMRTLKNSISYNNLKSIY FAYAHSQMKYGIVFWGNSGASMKTFRIQKKIIRIMEGV DQRTSCKPLLKRNTPTSKHVHNGKPNYYKRKIKKEILPL PSMYIMENLIIIKEKLNSKSPCITRNETVHCHNTRQKNNV HVKQTNTALFQKGVLHPGIRLYNALPKAIKSKTDTTKF KTAVKDYLVMHCFYSVDEYLEFQNYSIS	cd01650	503	794

				CTCTGTG AACTAGA GCATGTA CTAGAGA TATTTAA GGAAAAC AAAGTAA GTGCCAA CCAAGTA ATCAGAA AAAAGTG TATTATT ATTTT ATCTATTGT ACCCTGTAC AGTTCTGGA TTCACAGG ACCAATAA ATAAATAA ATAAATAA				
Penelope	Penelope-8_LMi	.	Locusta migratoria	TCACATC AGAAACC TCTACGG GGAGGAG ATGGTTG AATCGTG GGAAGAC TGGATAA ACTTCGC AAGAAGA AGGGTAA GCTGCTG AGCGCTC TTGCTTTT TTATTGA GATGTCG TAGTGAA AGTATCA TACCGAA GTTCGCC AAGTTTG TGCATTA CATTGAT CGAAAGT GGCCAAT AACATCA AAAAACG TGAGGATG GCGGGCAG TTGCGCCGC TGAAATATT GTGTCAAGT TGACTTTAG TATCCGGCT GCAAACCC GAGAAGAT TACA	MSDWAHRSAVDKQSSKFQRLSRPSISHNTNRRTVINFTD KELDDAMLSVLEKGLNFAPTPKNVPLTEFICSVEQAVSK LPGDQAEEIRQEVAHTLRRAPPPRSNISPLERAAIRSLRE DQDLIILPADKGNATVLLSRDDYLRKMDCLLDDSAYRR LQKDPTERIKRMTLSLLKKSSLSDDVLKKLHPGAAVPPR LYGLPKIHKQEVPLRPIVSNLGAPTYNLAKHLASVLSPH VGKCEHHIPNSMDFIQRLKSLRLSPSDLLVSFDVVSLFTR VPLEESLQLIGEKVDKETAELCRHVLTSTYFLFNDKFYE QTDGVAMGSPLSPIVANLFMEDFEDMALKTAALKPTCF WRYVDDTFMVWPHGEEALNRLLDHFNCMHPNIRFTME VEKEGKLPFLDVLVYRRDDGKLGHSVYRKPTHTDRYL HASSCHPPVQRTGVLRTLTRRAYAISDAESLTQELDHLK VVFKQNGYTDKQIRCAFQFGPSPEPLNETHKSVAFLPFA GSISSKIGRILKKHDIKSVFRPPVKTRALLGSVKDDLGLR KPGVYKIPCQCGKTYIGQTSVPFRTDVLNINATHVCSSP RSRRDSLSMWENVYWSDIRTIQDRCVEHQRHTRLLQPE KSAVAEHCLDEGHRMLYNETHMVASASRYWDCVLKE SIEIRLSDNLINRDKGYLLSKSGTLCYQMLKNNGWYFRN LNNNFLIIDYRFRRILPPERCMSCAPEGSYCHFLRTRYGS SAMYKGAADCTGNSVPARLCTSTLT	cd00304	318	402

			TGCTAGT						
			CTTGCCC						
			TAATAAG						
			GGAGCGC						
			ATTCGTTT						
			TACTAGA						
			AGACAAT						
			TGGATCT						
			CGTATCC						
			AAGGACT						
			TATTTTAT						
			TTACATC						
			TAGAGGT						
			ATCTTCA						
			CGGCTGT						
			CGGCTGC						
			ATCTTGG						
			GAATGGG						
			TTGATAG						
			CTCTTCGT						
			GGGCT						
			TTGCTTGT	TAATTCATA	MKRPFREKFPPSNPIGKEKATSDSVGSQLEDSNESEHLK				
			CTCCTAC	ATATTTATG	SPNYHQARNAKIKHKHINYIATHNINSLLQTGKLKNFTD				
			CGAGCGC	TTGGCCATA	ELNRQNIIIAGLQEMRNTIEDPFEXQGYRIYNGKPGQRV				
			CTCCCCA	AAAGTAAT	MKQCPQFGTGFIVNLKVIDSITDFKAISPRIATLTLKATN				
			GGTGGCG	AAATAAAA	KIYTIINAHAPTNQKNFLTRTMVEVEKFWDLLDQTINQV				
			GATAGGG	AAAATAAA	NKNHVKILIGDFNAQLGKERKYRDIIGKWAPHKHTNRN				
			GGATGCT	TAACCTCTG	GQRLVEICRTYGLISKSTYFKRKPSKLMTWKHPDWTKG				
			CACCAGA	TATTATTAA	EWQLDHVCMDKFFHKEIQNVKVLRGMDTGSDHYMVN				
			TATGGTG	ATTGTCATG	IKIKFTPLSKKRQRALEIKESMIHINKKTESIGDKRKYDP				
			GGTACCG	TTTATTTTT	HKLIKNDKYQKNTQNTVITDNLQELIPILKQHAEESAPV				
			GGGAAAT	GAACTAAA	NQTSKKHAWWTTECDEIHDERHQAWIKFQTHIRTEENA				
			AAAATAC	TTGTATTAA	YNLKNIRKKFSQIIRKTKRLYQKDVLSSIEENYHETNSRD				
			CCGGGGT	TGCATAAC	YFKAFDKQLRKFNPPTLLLRNKDGRIAHNNKENAEIMA				
			GGACCAA	AACAAATT	EAFNKLLNLTNLKSFKAFNNRHKHPCNNQTRKYSPPTV				
			AACTAGC	GTAATAAC	QEIETALKEMRNYKACGENQVFAEMWKYAGTTTRTSL				
			ACAAGGG	CTCTGTATT	HTILQKIWTTEKFPEEWTTAIITPLHKKGDKSDPDNYRGI				
	RTE-62_LMi	Locusta migratoria	CACGACG	ATTAAATTG	SLLDCTYKIFSKILHNRIREQLDKELGEYQGGFRPWRSC				
			GATCCAT	TCATGTTAT	AEQIITLKLIMAYYRKRNKPLSITFIDFKKAYDSIHRPSLL	cd01650	596	831	
RTE			ATGTCAG	TATTTTGAA	KILRHFGLHLKLIRLTELTLTNTNSKIKFEGNFLDHSIHM				

Dap hne	Daph ne-13_L Mi	.	Locusta migratoria	TGGCGGT ACCCCAA CAGCGTC TGTTCCA CATGTCA GTGGCGG CTGATCA GTATCCT CTGTGGC TAACTAC CATAGTT GTAACAC GGAGCTC GCTCCAC TCAAGGT TGATAGC CTCTGAC AGTCAAT C	CTAAATTTG TTAACAAC AAATTGTA ATAACCTTT GTAATATTG ACTAAAGT GGTCCAAT GAAGGCCA TAAAATAA ATAAATA	REWYKANPKNIKVGTQKDNIKLNCLGFADDLALLANNI QETKQQIRSLQEIAQCIGLKISFEKTEIMVTEPPLANKIRI GVHNIKIVEKFKYLGEIITHNLNEKPSWQSRINKLIKMHH VTKNTYNKKCLSIATKLKHYKTVTQPQITYASETIFKTT NTVAIDRLLKIERRIVRTCINKKYQKDGIWRIASNETVY KEVEPVTSTMKKKRISFLGHLMRTPENRISRKIIEKLWNS KSNIRWITEIKEDMMKLQITMEDLKDKTDTLKILQDKH VTLPIKINKQRLGRVISDEERRARSERMKKYWANKKKN KNKNNRRT	MIKKNVWFQTVKSMCVQEKKRVVPDCEIDVRPGIRTNQ LNNHFKNIQQAKHNSKNVTNSENYKGVFIHVGTNSLRS CSEEDLVNETRNVIRSAKNIYPYSRIIISSILRRRSVSDKYI DKINCNIRDQCNRLGAIFVDSNKFIRDDCIGQDGLHLNR LGSATFSKMFIDLCKIIRNRGNWQSVRGVNPVQNYTKQ GKLAVSEGGELKHVVDKDTTKEDVLKTVNRYNSSFLM HWNIHGLRATYLQNKIAKIDELQIILSEGRNIKVVCLNEH WITEDSITVLNKLENFNVASSFCRKNKTHGGSCILINSSL EYKERDDFNFLNEELIFESCCVELEQQNIVIISIYRIPGND VTKLFFDKFQCLLQKLTKEMRKKIIIAADFNIDTASETSIS TKFNDLIQQYGFRLNFTEFTRVSDKSATCIDNILTNYTLN ENAKFTLNLGISDHYALLMELLKTNKENGRKCFLKRQF NPDNIHLFCCRLNDIDWSLQHESSCSENFDRFLNSFLHVF NETFPPMSAKERQQVSNISTHVCQRKTTSKLKWITTGIKI SSKRKRELHAELKSNKSPEFIKYVKSYKKLFRKVVKAA KEMANTKFIRKHKNKSKAVWTVVKSELGIKPSNQEISK LKIQDKIIVNPAQISETLMNFSTKNTRQNNCKSCSDIRNF NEFFITVAKSDVNVPSYQDNVHLFGSSLSKNECLIKFDT VTVEDVEKTILNLKNKNSVGWDEIPSKVIKAASNIIAYPL CKIINQSLEEGYFPDSLKYAEVKPLFKKGSKEDMGNYRP ISSFPCFPKYLKYQKQFGFQKGKSTIYAINEFVNKVTSTL	cd01650	767	1046

				GCTAAAG TGAACCT TAAGTTC GTAAATG ACGAGTT KCCATGG TCGTGTG GCGAATG TTTATGT GATGAAC GGAATGA ACTGTTA AACACAA TAAACTC GAAAAAT GAAATCA TCAAGAT ACTAAAT AATGACG TTGAAGA ACTTAAA AAGGTAA TAGTTGC GTTGAAA GAAGAAA ATGCGAA ACTTGCA AGTGATG GAAATAC ACAAAAG TGCGAAA AGAATGT ATCACTG TGTGGAA AAAACCA AGAAAAT CCAAGTG TTTCTAAT TCTGGAA ATGATGC		DKSQKAAGIFCDLSKAFDSVNHSLLLYKMEKYGIMGTA LQWCTSYLTERKQRVVITSSKGKYSSKWRLSHRGYHKA QFVETIPQGVPQGSILGPILFLFYINDLPININYHSVLFAD DTSVLIENKKADEIPDNVLKTLENLDVWFHLNGLKLNIK KLRCLKDPRKPRCLVPSKWSKIKYKKTQVMQFKTKQSK LNDIQIKHNSQELHEADHVKFLGMKLDRIYPGPHTILGN ETGQNLSWSSHIDYLKNKLNSLAFAMSILANVTSLDTRK IVYRSYFESIVRYGIIFWGNASTMIRILTLQKKIIETCGAK PRESCRPLLRDLNILSVPSLYLFEIILSIIIQNYFFIITSSTVM KQEQRQFHAPTHRLNLYAQAPQYMGMKICNKLKNTHF RKVETESMKRKLYSTLVQKCYYSIDEFLNDDLSIEQDKT YNYVL			

				AACAGTAAACAAATCGGGTAGGCCTAGCGTGGAAAAGTGACAAACAATGCCACTTATGCGGATACAACAAACCAATACAAATGGACTGTCGTACAGCACAAAAAGAGGGACAAGCAGGAAAAGAAAGAAAAAGAGAACACTAAAAACGACTTCTTAATAATCGGTGATTCC					
Vingi	Vingi-7_LMi	.	Locusta migratoria	GGGGTACATGAAAGGAACCTCCAMGAAGATTGCAATAATAGAAGAACAAGAAATGAACAGCAATCGGGTGTACCTGGGCAACGTCAGAAG	TGACCGACCGGACACGGAAATGAATGA	MLRSALTADPFTTGLHSNGQHSDRSFSRPFVFISINVEGFSREKEVLVANLCQKYSCDIXLVQETHGGSTKNRPRIAGMELILERPHDKYGSAIFARPDVNITSASLTEDNNIELLTVQTPFFCVTSVYKPPVPILASKHQPTSTIRTRQQHRTPHSANSVFLCYISLQTTCANFSIKTPANFNDQDVNFVIGDFNCRSTSWGYPDSDTNGTSLETWAEAHDLHLIHDPSSHLHLTVAGGKEGTTQTTFCFFKGSCTMYEDCRRAHPSYATQWQVEKRVQPRQHFVSSKVAAQCTKTVGEPIPRTQHRPVICTVNPVVVPEEVPFKRRFNFKKANWQKFSEELDSKVINIPPQPEMYETFIKCVQQVSRKTIPRGCRTQYVPGMSSDDKLLLETYQTLYEEDPFAKETIQAGEILLETISATRKAKWSNLVENLDMXLNSRQAWKLLKNLSGDPVKSQDNFTSMTPDKVATQLLQNGKIKGRTPKEALQRVPETEHHFLRDPFSMAELEAAISSLKMHKAPGVDDLRVEQIKNFXCKTMEWLLSLMNKCVAKLHIPKMWRKSRVIAILKPGKEPTDPKNFRP	cd01650	561	816

						ISLLCHTFKILERLVLNRISATLEXKFIPQQAGFRAGKSCS				
						SQILNLTQHIEDGFERGEISGVAFVDLTAAYDTVNHRKL				
						LRKVYDLTKDYQLTSVISTFLQNRRFQVSLQGKRSRWR				
						TXKNGLPQGSVLAPALYNIYTNDQPIPANTRLFIYADDT				
						AVAVQGPTFKEVEKKLSLAMDELSKYYDANHLKPNPS				
						KTQVSAFHLRNREAEHKLDITWRGEQLKHCPTPVYLGV				
						TLDRALTYRQQCLNVKQKVSARNGILKKLVGSSWGAN				
						PHLLRTSAFALCISAAEYAAPVWDASAHAKQVDIAVND				
						TARIISGCLKPTPVQKLYPIVGIAPPNIRRAIAADYERTKQ				
						ETDHRHPLYGHQAARHRLKSRKSFLARTAPLEGAPEAN				
						RLATWKSSLPADVSPKEEIAPGANLQYPVWRSLNRLRV				
						GVPRCKSNLKKWGILPASADTSCDCGADEDASHLLVCP				
						LLDNTCTAQDVHEGNDKAIAAAIFWKC				
						MTSQHRTFSGPIMTIISVNIEGLSLCKEELIKDLCKKNKC				
						DVLCIQETHRDSILNRPKIQGMRLVAERPHRQYGSAVFV				
						KSEIQVLSTSVSDINDVEIVTVELKNCTITSVYKPPLSNFN				
						VNATHKHNSTKVNFVIGDFNSHSTRWGYPQDDENGMA				
						VISWAESNKLALIHDHKLPPSFNSGRWQRGYNPDLIFAT				
						ESVSQQCIKTVCPPIPKTQHRPIMCQVMTAVRPVCVPFR				
						RRYNFRKANWLKYQEILDREVAFLPPNPDSYDKFTDAV				
						KRSSRLSIPRGCKTEYVQGLTPETVADLHTYTELFEKDP				
						FDESTVAAGQQLLASISEAKRNSWVNLMEELNMSINSH				
						KAWRLLKRLSNDPTSLNNHANITADNIATQLLKNGKPE				
						NRSRRQTKNMSNTTEDETNSLSQPISIDELNEAIAKCKN				
						GKAAGLDDILVEQIKHFGPQTKQWMLNMFNKCMADC				
						KIPSIWRKSKVIAILKPGKDINDPKSYRPISLLCHMYKIFE				
						RILQTRLTYVIEPKLIPQQGGFRPERNCTSQILGLTQFIED				
						GFQKKRVTGAVFIDLTAAYDTVNTRQLIEKLYEITRDRK				
						FTDIIRNLLQNRRFFVEFQGKRSRWRNLTNGLPQGSVLA				
						PSLFNVYTNDQPLPVNTSSYIYADDRVIAAQGQSFEDVE				
						KSLNEALVELDSYYCRNQLKPNPMKTQTCAFHLRNKQ				
						ASRQLQVRWRGVDLEHCKNPKYLGVTLDRALTFKQHC				
						LNTKLKVCGRNNILRRLTGSTWGARPKVVRTSALSLAY				
						AAGEYASPVWCRSAHAKQVDVALNESARIVTGCLQPTP				
						VNKIYHLAGIAPPAIRREQAAYKERQKSIVSKWHPLHG				
						MEPAAQRLKSRNSFLTSTEKESDYRTTVERWKDAERHP				
						WMEPSEALAAGHNESWTTWKSLNRLRSGIARSKVNLS				
						KWGYQNDSTLCDCGAIQTTTHMYNCPLSPSHCTFEELM				
						SAQPNAIDVALFWAKNV				
									cd01650	471
										726
Vingi_5_LMi		Locusta migratoria	GGGGGAC	TAGAAACC						
			TTGAAAG	TTTCTTTAC						
			GAGCCTC	ATCTGCATA						
			CCAGCAG	ATTTTTCTT						
			AATGTGC	TTTAAATTA						
			TGTTAAG	ATATGTAG						
			CTAGCAT	GTATTTTAT						
			CACATTC	GATGTCTCT						
			GGGCGTC	GACACGAA						
			CCCTGGG	TAAGAA						
			CAACGGT							
			TTTTTCTT							
			GCCGGCC							
			AATCCTT							
			TCAAATC							
			CAGAGAC							
			ATCAAGG							
			CATAGCT							
			AGTGTA							

Daphne	Daphne-37_LMi	.	Locusta migratoria	TGGCGGCCTGTATAGATGTACTAATAAGTGGCTCTGTGAAAAACGATTATGTTGTAAATATAAAGTGTATTTTGTGCAGTGAAAATGTTCTCAGTGGCATTAGGCTGTGTTCAGTGACTCAAAAGTGGGTACATACGAAGTGCATAGCATATTTTGACACTATAAACACAAAATGTAATTGTGGCGTAATTCATTGTGATTTGCAATTGGTAGCTTTAGCAAAATATGATGCTGAGCGGGAAGAAAATGTGCTATTTACGATAGAACATGACATATTGATC	TAATAATAAATTACAAATAAAAAGTGTTGTTTATTTTATAACTATGTACACAGTATGGTACTATAACTTTGTTACTGTCTTGAACACATTGTAATACATCTGTATGTTTATTGTCTGATATCTATTTATGGTAGTGTACATATCTGATAACTATAATGTAATTGTTTATTATTCTTGACATTCATTTCATATTATTTTCTTCTATCTTATGGGATTTTCTAAGTCATTTATTGTTGCTATTTTGTTTTATCATTATTACTATTTTATCTAACCTTGACGACACCGATTGCATGTAAAAATGCTCAAAGGTGAATAAAGCATCTAATCTAATCTTA	MYIFSDSHGRGIASIMQTMQMKFKVTGITKPGAPLCIITNSKLLALPNLVPLYVLLHKTDNLKQISHVSNMWQSNHQGGFQNVMFVEATSSVRTISPNMVCTGIIMAKQFCARKYDNINSQRQEIHSPIPLSLVKEKEVTSHTQNEEKEQETSLSSPKKESPLLTLKRKDNWKRSVQKEKKNHHPVLRKRTRNITVFPQKRITIAHLEKKGQLEKKRAEREEESSSCSIEEVLSLSAPEEETFSSAEKLPAKLRRGERTKKDIPLKNQDFFIPAAGEKTTRYKLDLRKTKVVPRDQLHKFQGISCTNSIGQNVMFLHINVQSLKYKINELQFWLENLNCSILSVNEHWLNCEELKLYVPFGFDLASSYYRTNVEHGGVSMLKIMTWWSVYVKNNVNLIYEVTDVSELCIEAKFEAAAIKFPKQEIIVGSVYRPPKHNESLFVENLEKLILLLLKHKNYRIVIFGDINIDIRFKNRNVNHLLNMLRSLNFYCINTKPTRQAACLDNIICNFHLNKTDFDTISLCISDHDGLWLRQTLNYKILETSTEHRIIRPINDSKLNQLISKLNIVKWDEIITNVDANEATNRFLLLLAELFETCCPKHYRKSVKQRQCSKPQNLKKTGILHITTKSKKNWYTPHLNRIRILMLLYHDKAKHNNADKEMYMKIKKLYRSEIKAAKCKANDAYILNSTNKCKAAWEIIKTEMNYDHKVYQIPISPNTLNEHFVGPHSDSSTMQDVRKAEALLPETSSNLTNNFCWSRVTEKEVYESVAKLSNSRAEDYYGLSNFVIKSVRKQIVKPLTKIINKILEGGHYPECLKISVVKPVYKKGEVTSPENYRPISLIPIISKIIEYCMHKQMYQYFETNNILSASQYGFRSRLSTTKAVESIVSKIHSSFEKKEIVCATLLDLRKAFDMVSHNILIKKTPALWSERECYMPDGVLLGEQETVSQDKKKLQHYGVRGNAICLMESYLENRKQLVKINNESSECLPVLKGVPQGSVLGPFLFIVYINDLPTVVPCDAVLYADDTTLITCDTNLENVQHNMAVAMEKCNTWFEANVLQKNDSKTEEIVFNLTAPSQDNKTVKLLGFHIDQKLNWDTHLDKLCSRLARVLHLLYKLRSSVSKELLLYAYSAFFHSHLDYGIILWGNSVGSKTVFKWQKKAIRCIAGLSPRDSCRGYYKELNIMTVPSLYIYNCLVYVKENLNNYDLRMTIHDHSTRSGHTIDVPYARLAKTHNCYKYLGLLYFNKLPKEAYTVPVDQFKNTIVKWFKSKVFYSAQEFLECGTKDLNFL	cd01650	818	1089

GTTAAAA
AGTACTC
AAACAAC
TTGGAAA
CATTGGC
AATAACT
CTTAACG
AAGAACA
AAACACG
AAAGAGC
ATCCATT
TATTAGG
CCTAAAT
GGAGCTG
GTGTTCG
AAACAGA
ACGAGAA
AGAAAAA
CATCTTG
GAAAAAC
AAATGCT
TTAGCAG
TAGAAGC
AGTTGCG
CCTAAAC
AAGCAAA
AACAAGA
AGTATTT
CGAAGAA
ACAAGAA
AAAGCAA
AGCTCCT
ACAAAAT
GAATCCA
TCAACAC
AGAAAAA
GAGAAGG
AAGAG

RTE	RTE-26_LMi	.	Locusta migratoria	GTGGAGTTGCTAGTCTCCCATCGGGCACCTCCCCAGGTGGCGGATAGGGGAATGCTCACCAGATATGGTGGGTACCGGGGAAATAAAATACCCGGGGTGGACCAAAACTAGCACCTACTGCCTTGTAGTTAGGTATTGGATTACCAAAGGCTAAAGGAAGAAAACCTTGATCAAAAATTACCTGGCCCTCCAGGTTGGGGGTTAATGCA	TAATCTAATATAGGTGCTGGAAAAATGTACTGACATTGTAAAACCAGAATAATAAATAA	MGQLPIHWKKTLPVKKPIRLQQDGYNRRRRRLRKRTVLIGVWNVQGLQNKVGEVIKEMEARNQDIVVLPETKKKGHGIERIGPYTHFYSGVPKNQRAKRGISILMKNKFNRMVTNWEPIDENLIKININIQEHKITLIGVYGISEDEPIRKKDEFYNRLNGVIEDIGKTREIIIAGDFNSRTGSKIGNKIIGPFGENVVNDNGNRLIDICEQNSLKILNGFYQHKDIHKYTWHKDTLNQKSIIDYIIIKQQTRIKIHDIRAYRGVTLSDHYLVNSKFLFPFRGKNSEENQNCTTEDLQTEIKITKYNLNSFENESTVFLYQKRLDEKLIETEFDNTEDQYQYIISKLHESAKEALGETDEKHTXHSLYYWDEEIKKEVKLKKEKYLKWLNTRNVQDKIELNAQQAKIRKKVAEAKNKAWEQSCQKIESYIGGKRNTEAWKTIKNLRRNSKEXTHIKYITHDSWEKYFQNLLSENRRILGKPEGILKXQHILNILLMTLGKNIFKIYYQKTEEFLGNPEYNEITNELPDNYIHLEIETVKSAIKSLKNGRATGIGGIPAELLKFGTEKLIELLRNLFERCLNGEDVPNDLKVGYISVIHKKGRKDDCKNYRGITVTNTFSRLYGRIIKYFLEQEYKEKEAEEQAGFRAGRSTIDHIFCLQQIIEKNKKLGRAQPIHLVFVDIEKAYDSVPLPNLWKALLSIGINPTIIKAIQNLYKNSISKIKIGKFLSSGFQVTKGLRQGCSMSPTLYKIYTEVTLQNWKKKCQTMGIPINDRTIYSLQFADDQLIIAQDYEDIEYMTRKLIQEYKKAGLNVNMNKTKYMVIGGINEDLILEEGMGTITTTEEYKYLGVKITNDGKQEKEIRSRINAGKTTISLLNGILWDKHITTDNKIRIYKTIVRSIITYGSEVWTTKWQLKSKLLATEMDFWRRSARISRRERIRNEVIREKMKCKNSIIEFIEHKHLKWYGHVRRMGQERLPKCILDWTPNGRRKKGRPPTTWMQEVQSAMRKNNIPEDLWEDREEWRTIIXNLM	cd01650	592	851
RTE	RTE-38_LMi	.	Locusta migratoria	GCCATAGCCGGTCCCAAGCCCGGTTGAGAAAGGAGGAGGGGGAGGAGTAACACCTCGTTAAAAAACCTTG	TAATTACTCAAGTTCAATCGCTCTCTTTAAAGGGAATAATCGAAATAATAATAATAATA	MSTQFLKPVNKLTTWKSPNSTLGEFQIDHVAISNKNTKEILNVRTRKGFFESDHHLLQIKIKFQPNRKKPQMNKTFRIDPEYLKLNKKEISQEIRQSNATNWKELAEVLKSTMHLGKPPRKRKHRWWNSSCDQAIEERTQAWKTFSSNKTTEKWQEFLRIQKNTSKAIRREKRRYDINRLKEIEQEFLKNNTRNFYTIFKENLTGYQPPNLCFRRPDGSLETNTRNNCQILANIFTHSSTASRRQSNISKTPSNPDSKPPSLKEVKEIINSLNNKSPGEDGIIAEMWKLNDSTLTQATHRIISKVWETEQIPAEWKCALIHPLHKKVQKLIQITTGEFHCSQLQGNFTAPSNIQNFFKGPVKQTRRTGGPVNRRVSGRVSRSILRRTNMEP	SSF56672	393	550

				GTTCACC TGGCCCT GGTGATA GCACCTC GTGAGGG CCCCTGC CAGGGTA CGGTGAA GACCTCA ACGGCAG TGAAGGC GGAGATG AAGATCA TCGGTAC GGCGGAA CTGGCGG AAGAGGC AGCCCTA ATCCACT AGCGTCT GCCCTGG AATCCAG CGGTTAG TGGTCAG CGTCTGT TCACCTA GTTGGTG CGGCTGC TCCTAGT TCTTCTTG ATCTCAA CAATCAA GTCGTAA ACATGTG ACCTTCG TAGAAGA TCACCAA AACAAGT TTCAACT TGAATAA		KDILQMRRCRNTVVTFVDFKKAYDSIDRQTLFHTLEEFR IDRKTRSIIQQALTGTTSKVKFMGETSEAFHIHTGVRQG DGLSPLLFNLVLEKIVRTWEKEIKGIQLGIKKEQRIRVKC LAFADDLAILTNNREEARLALEKLHEISQKTGLQISYEKT QYMENKLQIVSPSQPSMEKWHKFLTSDIKDSVHGKQAA DSLPITTQYGKVAQVSHFRYLGEMIQPSGLNSIANQERV SRLQRAYKLTWNHYNKKSISVNAKLRHYNTVVLPEAL YAAETTVIHGHTKIREIEKQERKILRKIYGSVFREGIWIK RPTREIYKDMETITDTIRKRRMKFYGHISRMNETRLTRQI FEIANKSKNKTKWITETEEDIKELGITQDNINNREQFRNII KKHKLKQEHTENRTGKKWSERKREHSERMKRIWEERK KKKS			

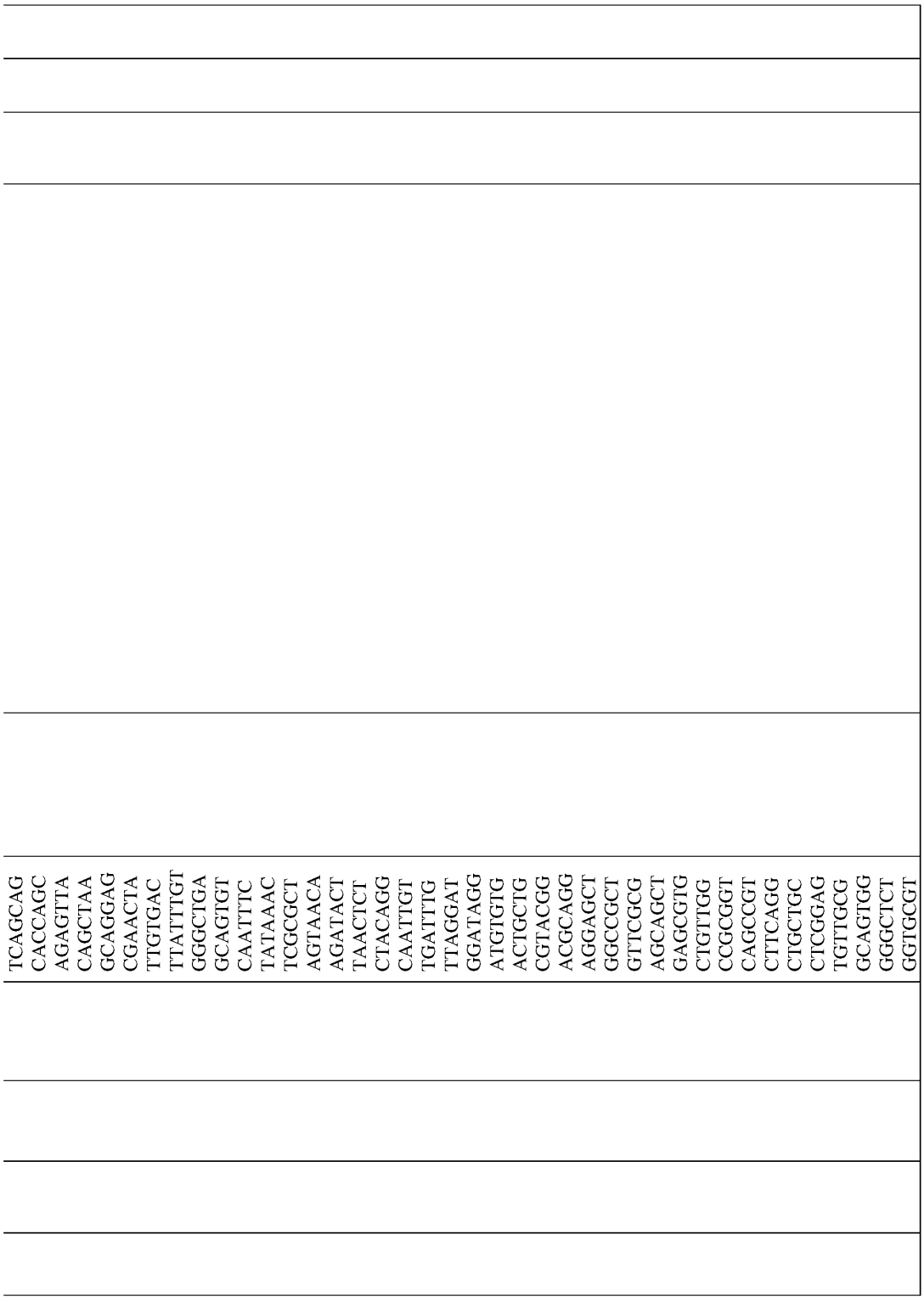
GCATGAT
GGAACAA
AAGGAAA
GAATTCC
ACTACCC
CAGGCAT
CAGTCTC
TAGACTG
AATTCTC
CTGGTCA
TCGGATC
TGGGGGA
CCAGGAA
CTTCATG
CGCAGAA
GAGTCGG
AGCTTCT
CAAAATC
TCTAAGA
CCAAAAC
AGAAAAA
CTTCATA
GGAACTC
TCAACAC
AAACACA
CTCTCAA
AAGTAGG
TAAACTT
AAACAAC
TCACAAA
CACTGTG
GAAAAAT
TCCACAT
AAAAATC
CTTGCGA
TTCAAGA
AACACGA
TTTACCG
ACGAAAA
TCATTTTG

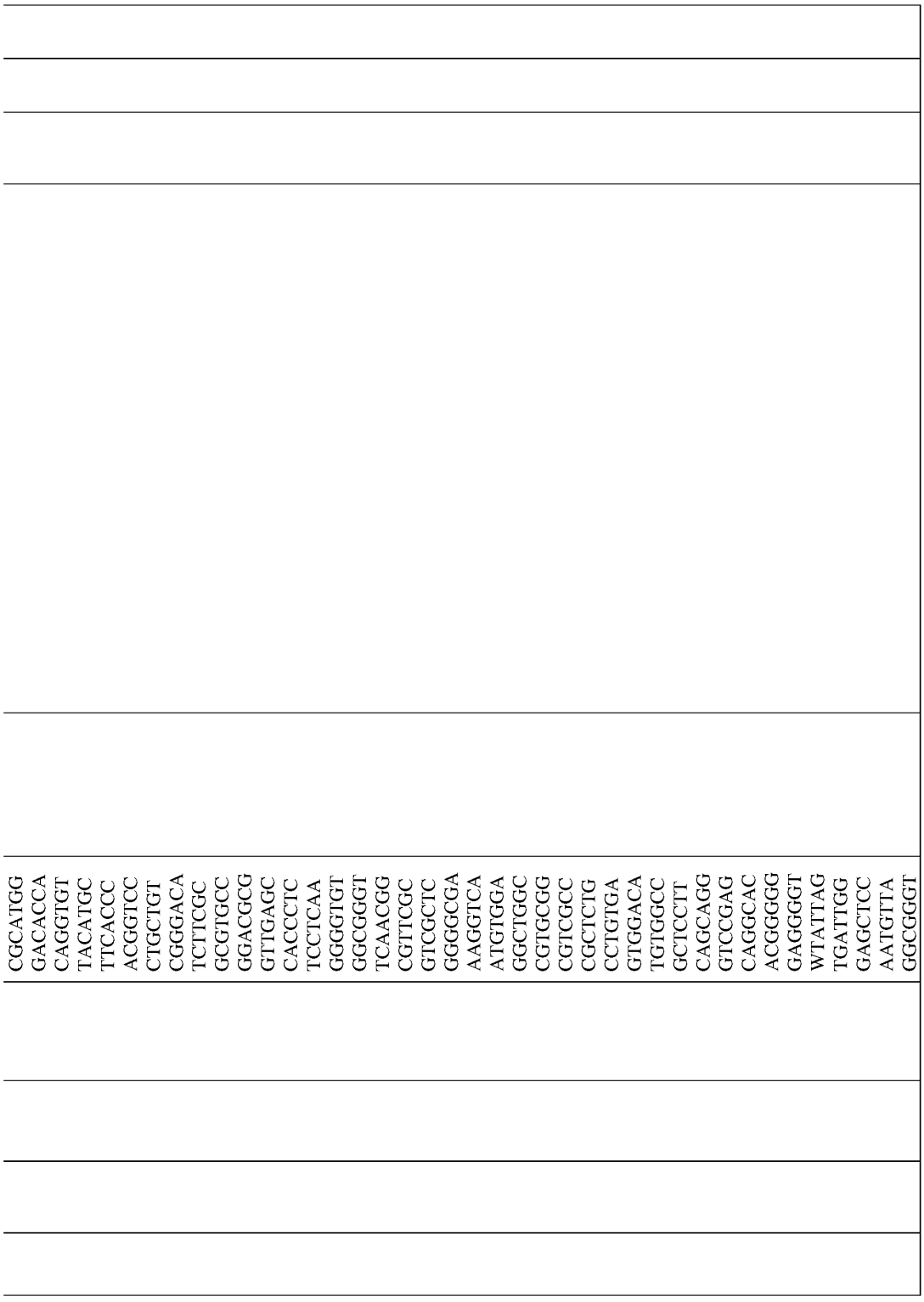
				TACAGGA AACTTCA GAATCTA CAAAGGA AAACCTG MAATCCA GTTGAAC AAAGGAC CCAAACA ATTTGGC ACAGCTT TTGTTGT GCATAAA TCAATTA CAGACTC AATAATA AACTTCC ATTCAAC ATCAGAA AGAATAT CCACAAT CTCATTA AAATCTG GAAACAA GGCATAC ACACTGA TAAATGC ACACGCT CCCACAA ATGATTA CAATAGA AAAGACC CACAGGT AATAGAG GACTTCT GGGAACT ACTGGAG GAAACCT CAAATAA AATTCCT					

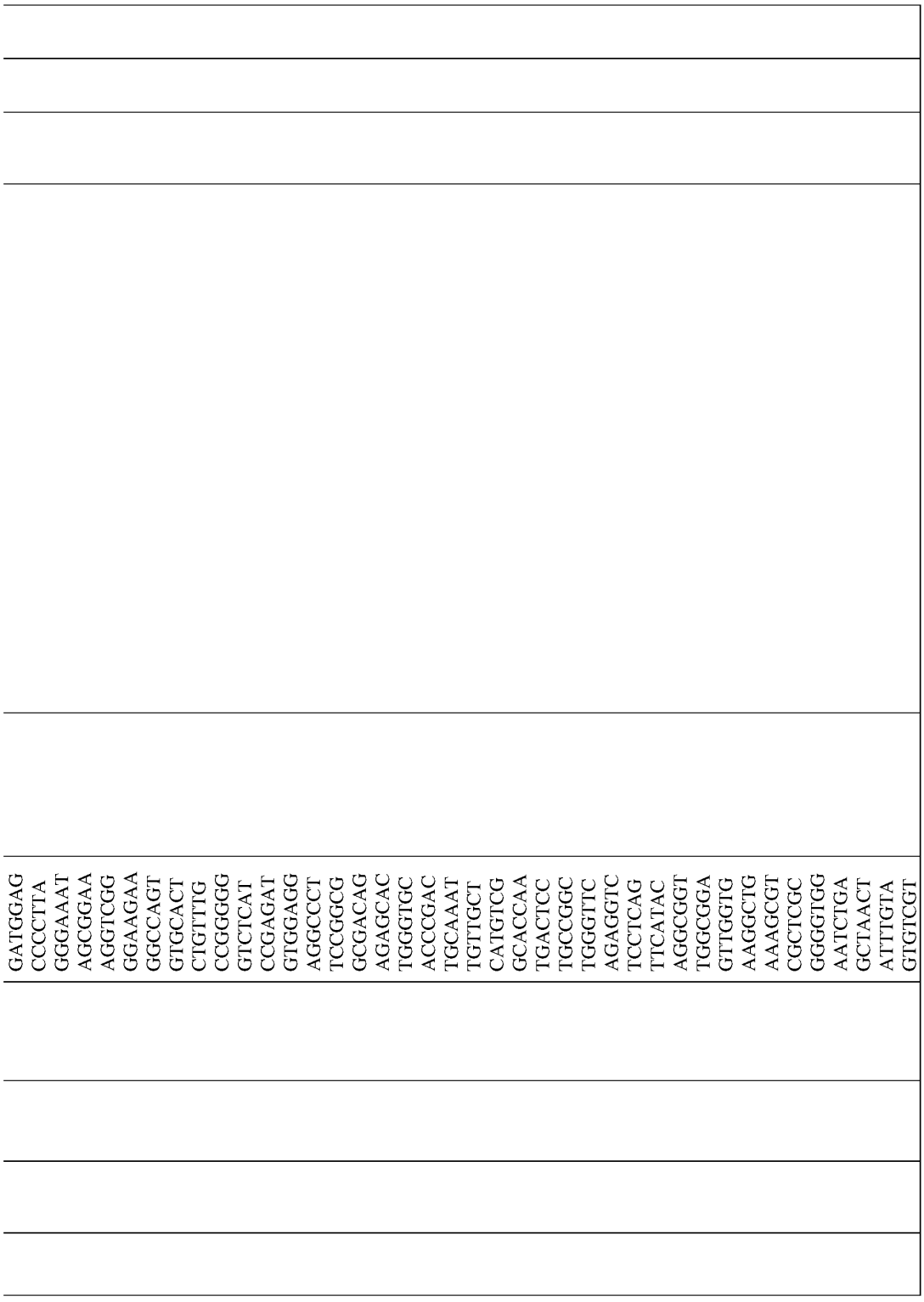
				GCACATCATGTGAAAATTCTTTTAGGAGATTTCAATGGACAACTCGGCAAAGAGAAAAAATTCAGGAAAACAACGGGAATTCATACAGCTCACAAGAGGACAAATAAAAACGGCGAGAGACTGATAAACTTCTGTGAAAATTTTGATCTTAAAATC					
Kiri	Kiri-1_LMi	.	Locusta migratoria	CTGGTTCACTTTAGGATAGTGCTGTTTTACTATTAATCTCATACTTACATTATTAATTACGGTACCATATTAGGTTTAAATTCAATTTTATATAGCGTTGGTCTCACGCATCATTGGCAATTGCTCGC	TAGTATTAATCCTATTAGAGTTAAGTCAATACCTTTTATTTACATCCTTTTACATGTAATCAACCTGTATAAATTATGTTATTAACTGTTTATCTGTAAAAATCATTGTTATGTTATAGGTAACTCACATGTAAAGAAATACTGTATGTGTAAAACCCT	MCPAFAEKDAATPQPAAVALPTSGIIARATTERTKLHVAHANVQSLQAHLVEFRHIFQDLGLHIILLSETWLKPSITTSSVNLDGYVFLRHDRLNKRGGGVGAYIRSELSPTVLHTSNNDETEQVEYMFIEIKPLKEKCLICVLYKPPKVGALDCFETALSDLESTYEHIIIAGDTNINFLFNSPSCMKFKRMLTSSNLTLLQLAPTHHTADSHTLIDIFATRFPDRVIRTSQTSAPGMSAHDVIFMTYSLECPRKKPLLTTYRDLRRVNLPELTADAXEIVFEPTSDTLPDVNEKLHKITNKLNQLYDKYAPIKTKKVRRKPAPWLTDELKQLMNLRDTAHRTYKLHPTPENHTAYKQKRNRVSQAVRNAKLRYAHSLTDNINRPAVLWKNLRGLGIGKVKTAAEPTVPIEELNQYFTTPTTTIDHQTKQTLIDHITRINNANHDKFCFKHVTGSTVKKAIMRITSASTGDDGITIQMIKLIIDPLLPIITDIFNQSLATSVFPDAWKRGLVKPLPKKDIATAPSDYRPICILPALSKALEYIVHDQLTSYLTINNLLDEHQSGFRKYRSTTSALIKVTDDLKQAMDRQEATIMCFLDFSKAFDTVDFDILLAKLSSLNLSQGVVQWFRSYLTSRLQCVMSGTEKSQWRKVLSGVPQGSVLGPLLFSLYVNDVSSVLSFCKYHMYADDLQLYLSAKPKNL	cd01650	510	773

CR1	CR1-1_LMi	.	Locusta migratoria	TCTACTA TCGATAT TGTTATCT TTATTTA GTGTTGT TCCGTGC AGATTAC TTTTACTG TGCTTGG CGAGATT CTCCCCC CCCCCTC AGCGCCC CGCCCTC GGCGAAC GGCCGGC GACCTTG GCTGCAG CGCCTCC CTGTCCC TCCCCTG TCCTGAA CCCACCT GACCTGC TTCGCTCT CTCTCTCT GCTGTCC TGCACTC CTCCCCC GACCCCC CGGACCG CTAGCCA CCCAGCC ACGCCAC ATTCCA	GGTCCGAT GTAAGAGA GGGCCTGA TGGCCCTAA TCTGATCAG GCTAAATA AACAAATA AATAAATA AATAAA	KTAIENFNSDLCALSKWAQNIGLKLNPSKTQAILVGHPR LINHKYRESQPSLILNDTPIYFSPSAKSLGVIIDDKLNWAE HTAAVCRKASASLHALQRYKKFFPLDLKKKLIQTLVIPII DYSDVILQGLSWESSRRLELIMNACVRYICNIRLFDHISP SYKQLSWLRADKRRDFHTLCFLYSLINVHCPSYLSSTLT LLSDQHNRTTRSCQSKTLSVPIHHSAGFSKSFSVAGTRL WNNLPQYIRELNNTASFKRQLMDYLLKQLRQSLIPYGP GTSHPFLLPDILLLPLLSPGTAFLTSFCQNSPCLNQSRYIQ KYILYLLISMLLLLLLLHQQQQKVSQWYC	TAGTTAT TCTCGCA TATTTGA TAAATAC TGTTTGA	TAGTGGTAC TCAAGGTA CCCTCCGCC ACACACCG TCGGGTGG	MVIGCFYRPPGSAAVVAEHLKENLENISSRFPDHVIVLG GDFNLPDIDWETQTFMTGGRDKESSEIVLSALSENYLEQ LNREPTRGDNILDLLVTNRPELFETVNAEQGISDHKAVT ASIISAVNRNIKKCRKIFLFGKSDKKRISEYLTAQHRSFIS STDSVEEQWSKFKAIVQYALDEYVPSKIVRDGKEPPWY	cd01650	388	636

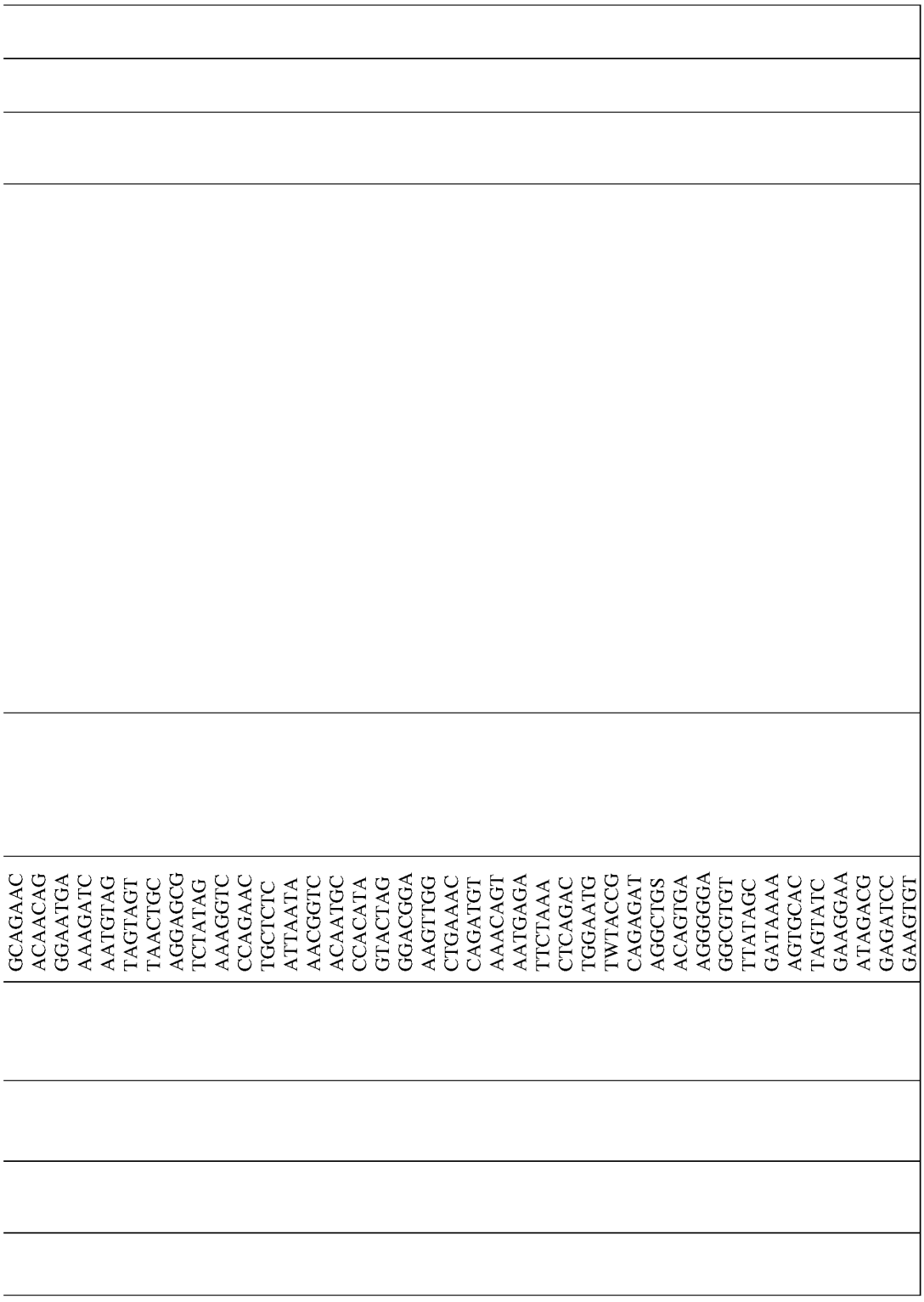
TAGCCTT	CTTGCGGA	NNRIRKLLRKQRELHCKHKHSQSLVDKQKLNEAKCSVR				
GTCTGTG	GTATGGTGT	RAMREAFNEFESKVLCTDLAENPKKFWSYYVKAVGGSK				
GCTGGTT	AGATGTAG	QNXQTLCDENGIETEDDTLKAELLNVFFQSCFTEEDCTV				
TTCAAAA	AGATG	VPSLDCRTDYKMADIEIIDRGIEKQLKSLKRGKAAGPDG				
CAAGTCA		IPVRFYTEYAKELAPLLAVVYRRSLEERSVPKDWKMAQ				
GAGACCT		VIPVFKKGRRTDVQNYRPISLTSVSCRILEHVLCSSIMTF				
TACTACA		LETRNLLCRNQHGFRKRRSCETQLALFVHETQRAIDTGS				
GTCGTTA		QVDAVFLDFRKAFDTVPHRRLMNKVRAYGLSDQLCDW				
TCAGTCA		IEEFLDNRMQHVILNGEKSSEVRVMSGVPQGSVVGPLLF				
GTCCAGT		TIYINDLADNIGSSLRLFADDAVVYREVVTMENCIEMQE				
GCCAGCT		DLHRIDAWCREWQLNLNVDKCNVLRIHRKKDPLSFSYT				
GTTGCCA		IAGQQLEAVNSINYLGVGIRSDLKWNDHIKSIVGKADAR				
GTGTCAC		LRFIGRILRKCSPKTKEVGYSTLVRPQLEYCSPVWDPYQ				
ATCGTAA		VGLIEEVEKIQRRAARFVTGSFGNRESVTEMIDKLQWKT				
ACAGCGC		LQERRSVARYGLLSKFREHTFTEESSSILLPPTYISRRDHE				
TAAGTAT		DKIREIRAHTEAYRQSFFPRTIRDWNRRRENR				
CTAATTT						
AATTTGT						
GTCAAAT						
TTAATTT						
GCCAGTC						
TAGTTCTT						
CATTTTC						
GTTAGTA						
CTTCTGTT						
TTAAGAG						
GTTTATTC						
TCGCGTA						
CTTGGCA						
GATTTCG						
TTTGTGCT						
ACTGTTTT						
AATAGTT						
TATTCTC						
GCGTATT						
TGATAAG						
TACTGTT						
GGGCGCG						
TCAGCAG						
CACACAG						


TCAGCAG
CACCAGC
AGAGTTA
CAGCTAA
GCAGGAG
CGAACTA
TTGTGAC
TTATTTGT
GGGCTGA
GCAGTGT
CAATTTC
TATAAAC
TCGCGCT
AGTAACA
AGATACT
TAACTCT
CTACAGG
CAATTGT
TGATTTG
TTAGGAT
GGATAGG
ATGTGTG
ACTGCTG
CGTACGG
ACGCAGG
AGGAGCT
GGCCGCT
GTTCGCG
AGCAGCT
GAGCGTG
CTGTTGG
CCGCGGT
CAGCCGT
CTTCAGG
CTGCTGC
CTCGGAG
TGTTGCG
GCAGTGG
GGGCTCT
GGTGCGT


CGCATGG
GACACCA
CAGGTGT
TACATGC
TTCACCC
ACGGTCC
CTGCTGT
CGGGACA
TCTTCGC
GCGTGCC
GGACGCG
GTTGAGC
CACCCTC
TCCTCAA
GGGGTGT
GGCGGGT
TCAACGG
CGTTCGC
GTCGCTC
GGGGCGA
AAGGTCA
ATGTGGA
GGCTGGC
CGTGCGG
CGTCGCC
CGCTCTG
CCTGTGA
GTGGACA
TGTGGCC
GCTCCTT
CAGCAGG
GTCCGAG
CAGGCAC
ACGGGGG
GAGGGGT
WTATTAG
TGATTGG
GAGCTCC
AATGTTA
GGCGGGT

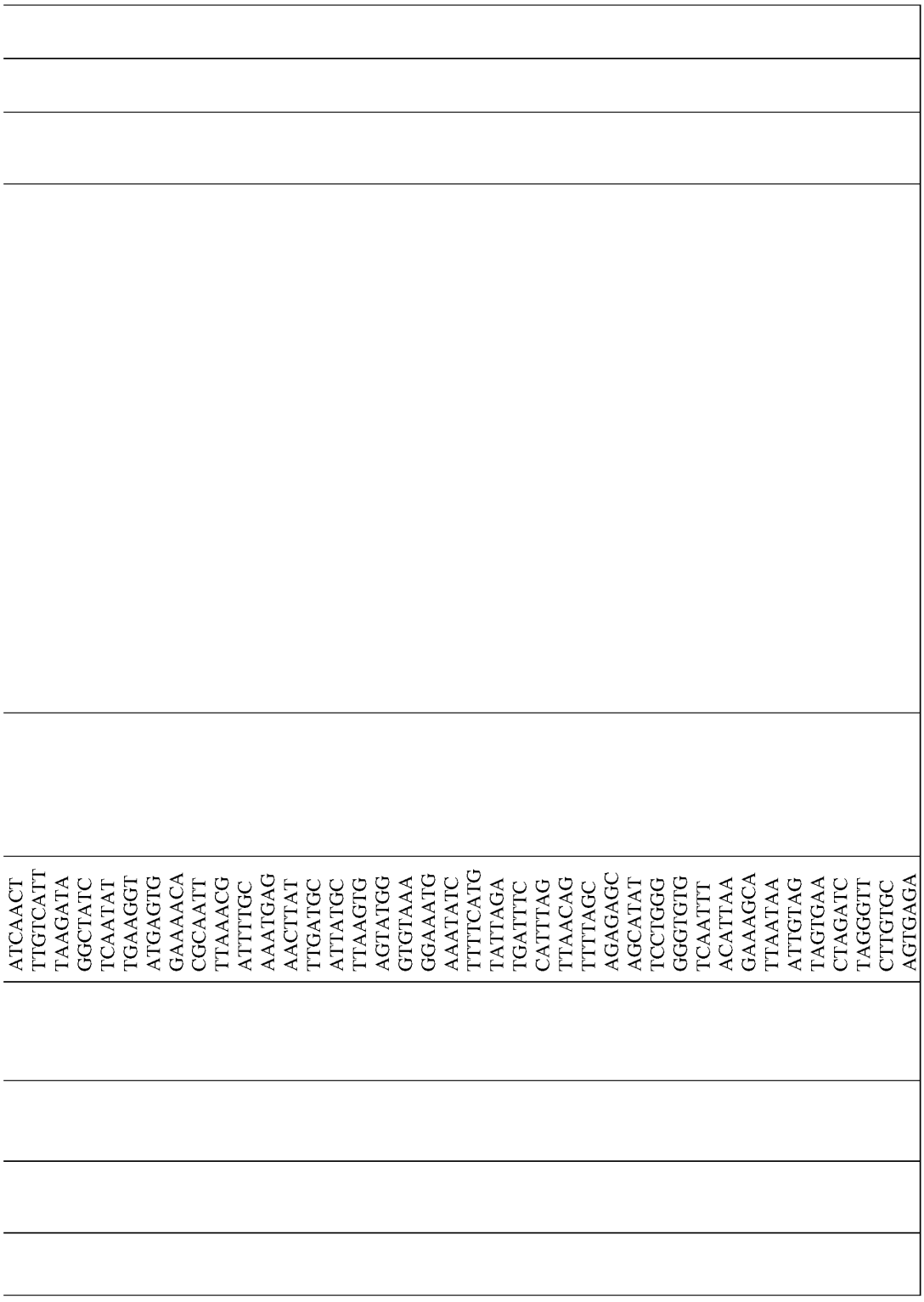

GATGGAG
CCCCTTA
GGGAAAT
AGCGGAA
AGGTCGG
GGAAGAA
GGCCAGT
GTGCACT
CTGTTTG
CCGGGGG
GTCTCAT
CCGAGAT
GTGGAGG
AGGCCCT
TCCGGCG
GCGACAG
AGAGCAC
TGGGTGC
ACCCGAC
TGCAAAT
TGTTGCT
CATGTCG
GCACCAA
TGACTCC
TGCCGGC
TGGGTTC
AGAGGTC
TCCTCAG
TTCATAC
AGGCGGT
TGGCGGA
GTTGGTG
AAGGCTG
AAAGCGT
CGCTCGC
GGGGTGG
AATCTGA
GCTAACT
ATTTGTA
GTGTCGT

				TCCCAGA ACCGATC GCGGTCC TCTGGTTT GGAGCCG AGTGGAA GGCTTAA ACCAGAG GCTCAGA CGATTCT GCGGAGA TCTAGGC TGCAAAT TTCTCGA CCTCCGC TATCGGG TGGAGAA ATGTAGG GTTCCCC TGAATAG GTCAGGC GTGCACT ACACGCA GGAAGCG GCTACAA GGGTAGC GGAGTAC GTGTGGA GTGCACA TGGGGGT TTTTTAG GTTAGAG AATTCCC TCCCTAG GCCCGAC AAGACGC CTCCTGA GCCGCGA CAAGGTA GCAGTAG					


GCAGAAC
ACAACAG
GGAATGA
AAAGATC
AATGTAG
TAGTAGT
TAACTGC
AGGAGCG
TCTATAG
AAAGGTC
CCAGAAC
TGCTCTC
ATTAATA
AACGGTC
ACAATGC
CCACATA
GTACTAG
GGACGGA
AAGTTGG
CTGAAAC
CAGATGT
AAACAGT
AATGAGA
TTCTAAA
CTCAGAC
TGGAATG
TWTACCG
CAGAGAT
AGGCTGS
ACAGTGA
AGGGGGA
GGCGTGT
TTATAGC
GATAAAA
AGTGCAC
TAGTATC
GAAGGAA
ATAGACG
GAGATCC
GAAGTGT

			GAAATAA						
			TTTGGGT						
			GAAGGTC						
			ACGGTTA						
			AAGCAGG						
			CTCGAAC						
			AGTTTTTC						
			TGGMTGG						
			ATCAAAT						
			GATGTAG						
			CGASAAA						
			TGAAGGA						
			AAAGACT						
			TGTTGTT						
			ATCATTA						
			AGGAAGA						
			AGCTGCA						
			GGAACTA						
			CAAAATA						
			CAAAAAT						
			CCTCATC				MVSLSNYDWSHIFNNNAHDSANTFFDRFFHTFLNIFETN		
			TTTTCAA				CPQYKCKVNPKSNSNKHRDKNPWYTAQLANMKSRLLL		
			CACCATT				YRDSYRLQKTDMAKQRYSRAQKDYKHAVNEAKKQHN		
			TCGGCAT				LVSIESSKNKCKKAWTIINSVSKSKQKVDVVISADEFNN		
			GATCTTA				FCIKSVNQISENIVRPHETVADLIKCHNVVNYTLKDKFLLS		
			AATCCTG				EVTPDVVLTIVNNFRPSDSVDIYGISCNMLKKIIGYIVYP		
			GTCATGT				LTKCINKCLIEGVFPDVLKLSRVVPIYKKGDRNCPSSYRP		
			GTGAATG				ISLTPILSKFVLESIIYYQLCDYLTCNNIISAVQFGFRKGKS		
			TCTAAGT				TVHAIDSLIKKVLNAYEEKKFAQATFCDLSKAFDCVEHI		
			AACGAGA		TAATTCTGT		SLLKKLVYYGIINSEVDNIFESFLSNRKQIVCIGKQRSQL		
			GTAAATG		ATATATGCT		AEVKCGVPQGSVLGPLLFILMINDLPSYINTHSFSMPMIP		
			AAGAAAA		GTATATTTC		LFSISAQTLICFKLFTILYKYSLILYADDTTFFNISSDFDML		
			TTTGTAA		TATACATTT		QTVTNDTLSQASVWFRANGFLLNESKTQKIVFSLRDLPV		
			ACACGTT		GTAACTTGA		EDFIDSVKFLGIVIDSKLSWEPHIKYIXTRLSRVVYLLRN		
			AAATTGT		CTTTGTCAA		LKNHVPAAYVRSAYFAFFQSIISYGLLIWGNSSHVQDIL		
			TGCTTTTG		TTGCTGTAA		VLQKKVIRIITNSDKLAHCKPLFINLKILTVINLYIYTALL		
	Daph		TAAACAT		CAGCTAAA		HIKSNLSEYQRRGDVHSHGTRNSSHLDIPYYRLSKSLNS		
	ne-	Locusta	AATTAAC		CAATAAAA		YEVVGMKMFNKLPQEWVEAPLDLFKHKLYCWLAANP		
Dap	46_L	migratori	TTAGGTG		TTTCATTTA		CYSLNEFYECKISWYLLGECTIISLMESFMMQCFKLYYF	cd016	252
hne	Mi	a	AAATATT		TTTATTTA		KYLILWKTYSDLLIFALCKVICIYALPYVK	50	527

TCATACA
GSTCATK
GCTTCCA
TCTAATC
ACTGGAA
AGAAGAT
GTGAGTA
ATGAATC
ATAAATC
TAGCCAT
GAGTTAT
TATACAC
AAATCGA
TAGGTCA
GATCAAT
ACCACTT
AGATTTA
AGGGAAA
TTCTTTTT
TAGAGGT
AGAAAAA
GGGCTGA
GCCAACT
GGTTTGC
CCAGATT
CACACTG
GAAGACC
AGATTAA
AGATAAG
AATAATC
TAATAAA
TGTTTGT
AATATTA
GAGAGAA
ATATAGC
AGTAATG
ATGATTG
CCCTCAT
AACAACA
ATTATTC

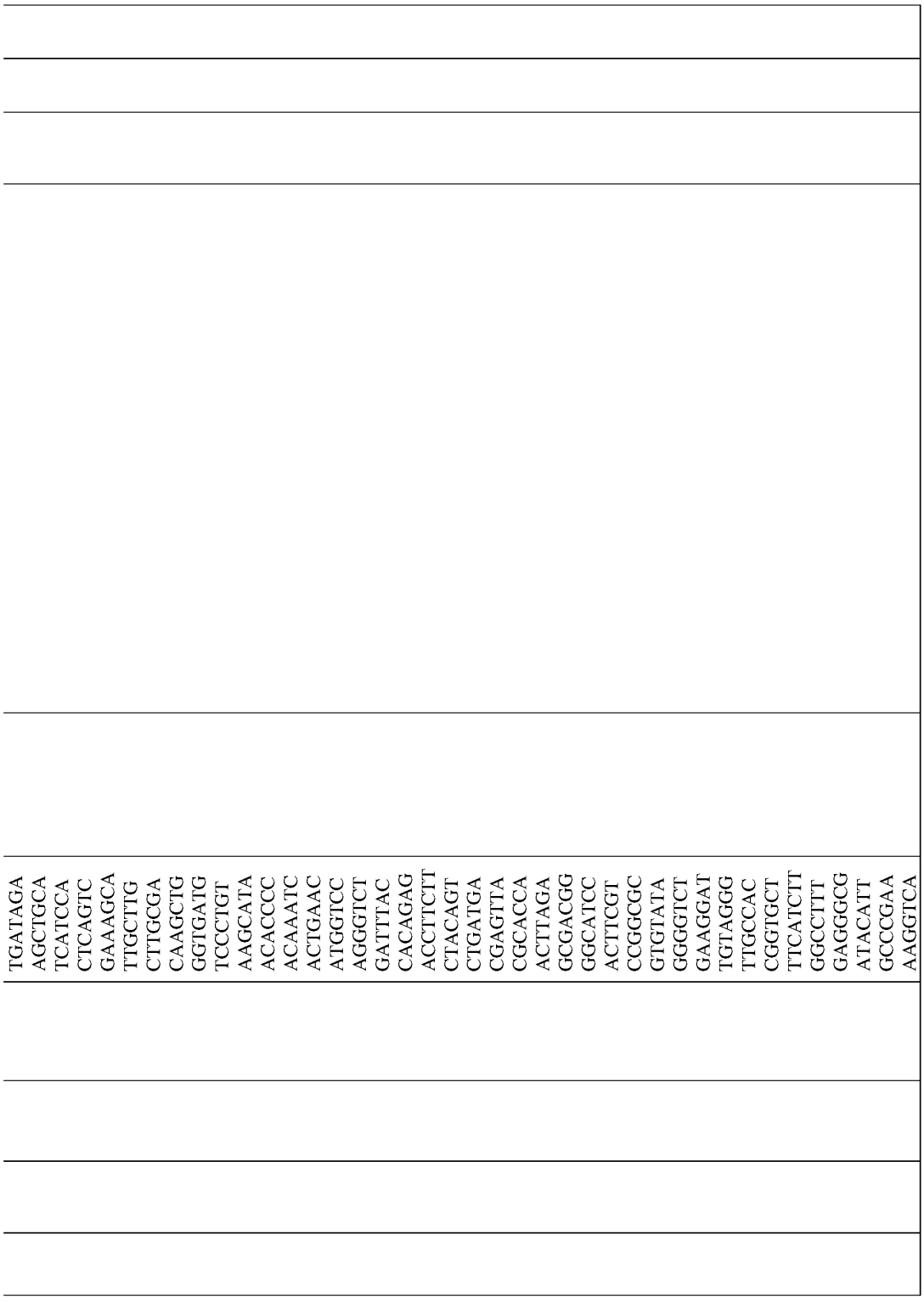

ATCAACT
TTGTCATT
TAAGATA
GGCTATC
TCAATAT
TGAAGGT
ATGAGTG
GAAAACA
CGCAATT
TTAAACG
ATTTTGC
AAATGAG
AACTTAT
TTGATGC
ATTATGC
TTAAGTG
AGTATGG
GTGTAAA
GGAAATG
AAATATC
TTTTCATG
TATTAGA
TGATTTC
CATTTAG
TTAACAG
TTTTAGC
AGAGAGC
AGCATAT
TCCTGGG
GGGTGTG
TCAATTT
ACATTAA
GAAAGCA
TTAATAA
ATTGTAG
TAGTGAA
CTAGATC
TAGGGTT
CTTGTGC
AGTGAGA

TACATTTT
GAAGCCA
CAGGTGT
ATGTATG
GATMATT
TAAAATT
GTATTGT
AGCTCTT
TATAGAT
CACCTAG
TGGTGAT
CCACTAA
TATTTTTA
AGAAAAA
CTTGAGA
ATTTATTC
AATTTCTT
TCTCAGA
CCTCAGT
GGAAGAA
ATATAGT
TAGTGAT
TGGTGGT
GACATCA
AGCTGTA
TTTGATG
TAAATAA
AGATGCA
CACACTG
TAAATAT
TTAAAAC
CGTTACG
ACAATTC
AACTTCA
ATTTTAC
AAACTGC
AACAGTC
TGCTTGT
CTTGACA
ACTACTT

				ACAAATGTAATAGAGAGTTACTGTCCTCAGATGTAGCTGTTTTTGATTTCTCGGACCATGACCCAGTTTGGGTAAAAATAAGTACCGATAGGCCTAATATGCAGTATAAGGAGTTCAAAGAGCCTAAATAGTCATCACAAGACCAATAACACAGGAAAAACTTTCCTAATTC					
Nimb	Nimb-41_LMi	.	Locusta migratoria	CATGGGCTCAGCATTAATTGCTGAGTATGGGCTTGGCGAACCCGGGGTCCTGAGCTGTGGGCTGTGTCGCGCCAGCAGTCTGCTACACCGTAAGCTCTGTGCGTG	TGACTGGCTCATCCATTATTATCCTAGTGATCAGCCAGCCACATCTTCTGCTGTGATTTTTAGCACAGTTGACTATTTGGTCCACTACTTTTGTGGTTTCCCTTTTCCTTTTATATGTCATTTAGCACACAC	MAPTVQPSSVGPKGHQTHSASLKVGGPSFIAPPVPWWGGNTSPPPSGTQVPIPQPEKQQPSPAPLAREEKSVGMLPSPVSKSRLEDPSRSLGDAHTTGRRASRSSSAPETSEKPSQWSKPKEQREDKQKKKRSKKKSLPTALPPSVSQTSDSEDGLEMLASPEDLALTDLPESMDIDATSTKAAAKGDTEAGINSMFMTLMASQYANSNILQWNCGGYFHRLAELRHLLCQFPVLCVALQETWFPVQRIPALRGYRGYYKNRAAYDRVSGGVCIYVLANVPSAPVPLHTALEAVAVRIRAGQEITICSLYLPPDKAVSSAELAALVAQLPPLYLLLGDFNAHNPLWGGTEITARGRDVEYLLAQLNICLLNNGAATHFSLAHGTYSAIDLSLAVQAYFHTLLGGSMTTFVEATISLSCRSHSHHLQAACPDGHSVGQTGRLTPRLLQQRSRHTLQLGPWHLLSHRSVPCSPSLLPHLTWRVHDDLCGSDHFPLLSLPQSPPAGCLPRWSLRRANWPAYTSAATVTSPLPDDVDLAVDDVTRAIVSAAEATVPKSSGLLHRKSVPWWTPEIAEAIRERR	cd01650	849	1106

				CTTCAAC GACCATC GTTCGGC GCGACGG TGGAATG TTGTACG TTTCAGA GCCCGGG GATCTTG GCTTAAT CGCCCAG GTCGCGA GGACAGT ASAAACC TCTATTA AAAAAAA CCTCAAT CTCAAGG TGTGCTA CGTGCCG AGGAGAT GCATGGC TGTTGAG GTTGAAC AGTCGCT AGGAGAA CTCTGAG GAACCTG CCGCACC TCAGTTG TATAAGG CTTACCC AGGCAAG CGGGGCT CTGCCTG GGTGGAC CTTTATAT CCCTAGC TGCTCGT GGGACCA	TGCCTTCCT GTACTTCTC GTGTACTGG CTGTGGAGT GCTTCGTCT TCAACTGTA AATCACAT GTGTGATCT CTTTGCTGG GAACAGGG GACTGATA ACCCAGCC GTTTAGTCC CCTTTAAAA ACAACAAC A	RALQRHRRHPSLEHLIAFKRLRAQARRLIQQSKRDSWER YAATIGSRASPTZVWKQVRRLFGHRPLAGIPSLHIRGTLS TDPAAIAQYMADIMLVARHLHLTVRTSAPSKSVLNSGS FPFTCAARKNTMRHSVSGSLRVPAPFRASSTGRNPQSSH SGYSVHRPSCYRAVHGRHYARSSASASYSADFRALKER VEQRQLSFHMRRQEKYNAPFSEWEFESALAACPDTSPG PDSVHSQFVRHLPAXFQQLLLALYNHLWVKGAFPSQW RDSIIIPVMKPGADPQVAESYRPISLTNVLCKIFEHMVSR RLMWILEDRGLLAPQQGGFRHGRSAIDNLVSLESAIRTA FTRREHLVAVFFDLRKAYDTTWRHHILATLYDWGLRGP LPTFIKNFLSLRTFRVRVGTSLSTHHIQENGVPQGSVLSV PLFLVAINGLAAAVGSSVCSSLFADDFCLYYSSASIGVTE RRLQGAIRKAQSWASAHGFTFSASKTRVMHFCRRVVHP DPELYLADEQLHVAETHRFLGMLFDNRLTWGPHLRQL KQKCWRHLNILRCLGHTAWGAGRQTLMHLFKALILSR LDYGSQVYDSASPSALRILDPILHGGVRLAMGAFRTSPV PSLLAEAGIPPLRFRRQQLLASYVAHVHATRLTIQIAGFC FQTPSFPLHNIDPDQVFQSRSGYVLSFPHSSPYLYLPHSP ANLISLLGPSLAYSSTCNCPAIQRRRFHHIHFINSFWTSLA RSVMQKSYTLMAPRSWHPSPEVSPATVVGFICRPRSCYT PHHPNRRLLFPNTVLPSPQHRPRSGLPIAVRVRTLLSSLE SLPLPPAFSCQSHIPPWSIPRLQLHLQLSCDPKASVSPHTL HQQFLDILGTFRDAEVIYTDGSQVEGRVGYAYVHGDHI EQCSLLPGCTVFTAELIAIIRALDHLRTCPRKSYVICSDSL SGLQALDQCFPKHHLVRSIQDSVFSLGXSGRSVAFIWTP GHIGIPGNERADQLAKQAIASQRLDIGTPERDIRLGLHHK ISMAWNSEWSTLSTPNKLCAIKPSTAAWKTSLRPSRRDS VVLCCLRIGHTWLTHGHLLHREDPPNCRCGALLTVVHI LLDCPLLAALRQSFNLPSALPSLLADDASMVDRVLGFIR AAGFYHMFNGGYLTFRRPPLNSFNLLFLLSCQSYFQLAA LVCTG			

				AGATGAA ACGAGCT ATAAACT ATGCCAA TGGCGTC TCTAAGC GAGGAAG ATCAAAT TTCAATG CCACGAA ATATGAC CCCCGAT CGTTCCC TTCCCTTG GTACATC ACGGGAG GAACATC GGGCTAA AGATAGG GATGAGT TATACTC TCCAAAA TATCTCA TATGCAG CAGAACA GACGGGG AMAGTTT CATGACG ACAAAAC CTCTTTTC TTCGTAG AAAATCT GGAGGAT AAGTATG GTGAAGT CGCGGGA ATGTCAA AGATGAG GAGTGGT TCAGTGC					


TGATAGA
AGCTGCA
TCATCCA
CTCAGTC
GAAAGCA
TTGCTTG
CTTGCGA
CAAGCTG
GGTGATG
TCCCTGT
AAGCATA
ACACCCC
ACAAATC
ACTGAAC
ATGGTCC
AGGGTCT
GATTTAC
CACAGAG
ACCTTCTT
CTACAGT
CTGATGA
CGAGTTA
CGCACCA
ACTTAGA
GCGACGG
GGCATCC
ACTTCGT
CCGGCGC
GTGTATA
GGGGTCT
GAAGGAT
TGTAGGG
TTGCCAC
CGGTGCT
TTCATCTT
GGCCTTT
GAGGGCG
ATACATT
GCCCGAA
AAGGTCA

				AGGTCAT GATTTAC CGGTGTG ACGTAAA GCCTTAC ATCCCTC CCCCTAT GTGATGC TTCAAAT GTTGGAA ATTTGGG CATATGT CATCCCG CTGCAMC TCTAGCC CCACATG TAGGGAA TGCGGGT GTCCATC CCATCCG GACGCTC CGTGTAC TCCCCCG CCAATAT GTGTCAA CTGTGGG CAGCACC ATCCACC CTGTTCA CCAGACT GTACCAT TCTCCGC AAAGAGT GTAAAAT CATGGAA ATTAAGA CCCTGGA CAGACTG ACATACA CTGAGGC					

RTE	RTE-40_LMi	.	Locusta migratoria	TACAACT	TAAAGTAA					
				GGAGAGT	AGTAAGTA					
				TCCAAAC						
				CCTTCTTC						
				GGCTGGA						
				GGGTGTC						
				TTTCATTT						
				GTATTAG						
				GTTTCGA						
				CTTCGAC						
				CCCAACA						
				TCTTGTTT						
				GGCCAAC						
				CCTCCCTT						
				CTGTCGG						
				TATCTTTC						
				CCTGCTTT						
				GCAGGCC						
				GACATTG						
				TGAGTCT						
				CCCGAGC						
				GCGGGAC						
				GGCGCCC						
				ACTAGGA						
				GGCTCCA						
				CCCAGGG						
				GGCGACG						
				CCACCCT						
				TCGCCAG						
				TACGTCG						
				CGACAGT						
				TCCTATG						
				GACAACT						
				GTCCGCA						
				CCACCCA						
				AAAGACT						
				AAATATG						
				AAGGAAA						

						QWNKVRQLFHFKGLYISSTGFERKNIYKGTWISPDGRTV				
						NQIDHVAIEDKLKKWITNVKTARGQSVVPITFCSKNSKG				
						AECGSDHFLLKIDLRVGWKPKLTKKLQKVVHFDYNKL				
						VTDEGLQDNFRINLIVLKPWQHCYRNPGNIATDSLDAGS				
						LWEGIRDLTETAKEVIPKRMNRKKVWFDXECELAILER				
						QRAKQVWIQSARGRYEEGKERFDEARRRTQATLRRVK				
						RNYFKNIMVEAQKDFHGNKSKELFRKVKAIKGEGKKH				
						GFIKDRNGRVLTNNNDIAIRWMEYFDELLNCEEPXETLE				
						FEEPVTVDPEYPLQHSKKGTCNCGSGIPPPTFEEIKKNIK				
						ELKNGKSAGEDGIVAEMLKSGGDSCIQNIHKLVRKIWQ				
						EEVIPEDWKVALICPIHKKNDKMLCDNYRGIALLNISYK				
						VYSKCVLLSRIKPLAEELIGEYQGGFQPGKSTTDHIFFIK				
						QVCEKMYEYDKEIHLLFVDFQKAYDSIHRTSLFKALIEF				
						GVSKKLVRLVSLCLDDTRAKIKLGNRTTDSFRVTTGLR				
						QGDSLSPILFNLVLEKIIREMDSFNHQGIQVANRNITHLA				
						YADDIVLMTHTKENLKCNANDDHGKHSKQIWTESEPH				
						KNQIHDSWKSPTPRNPGINCGNNTYKRVKDFKYLGSIIS				
						ENSSSGMEIKARIQAGSRCLHSIHNILTSRSISRALKLRLY				
						KTLILPVILYGCQCWSTRKQELNRLAVFERRCLRKIFGP				
						VKDNITGEWRRRRHNIELEQLYSEPNIVGLMRSRRLEWA				
						GHVVRMSEERWPRVAMEMIPTGKRPLGRPRKRWMDGI	cd01650	394	654	
						REDLQKLEAWKNWEQTAKNRNEWRALIISSRGPLGLIA				

AAACTCC
AAAAAACA
AAAAAAT
CCTGGTC
CTCCAGG
TTGGGGG
TTGGCAA
CAGGCTA
CCAACCT
GTTCCAG
AAAACTA
ACAAGGT
AAAAAACG
ATATAAC
AAGGCCT
CGGATAG
GACGGAT
GAATGGA
AATCGAC
CCAAGCA
AGTGAAC
GGACAGA
ATACTGG
ACTACGG
ATTGGAA
CCTGGAA
TGTTAGA
AGCTTGC
ATCAAGC
AGGAGCA
ATACGGA
GTATGAC
AAGGGAAA
GTAGAAA
GGTACAA
AATGGAT
GTAGTAG
CTCTACA
AGAAATT
AGATGGC

CGAACTG
CTCAACC
GATGTGA
AAACTCA
ACATTCT
GTATGGA
GTTTCAG
AAGAAAC
CATACAC
TGGGGAC
AGGTTTT
AAGTAAG
GAAAGCC
TTATGTA
ACCAGTA
CTAGACT
TCAGTCA
GTAAACT
CAAGAAT
CTCTTTTT
TGACTAT
AGTACTG
GGGGACA
CAAGCTT
GTCCTGA
TCAACTG
AATTCGA
AGGACGC
CTTTTATG
ATGACTG
AGACTGC
AAGTGAC
AGTGTCC
AAAGGAT
CACTTAA
GGTGCTG
TTAGGTG
ACTTAAA
TGCCAAA
ATTGGAA

Dap hne	Daph ne-51_L Mi	.	Locusta migratoria	AAGAACA AATTTTTT TGTGGAC TCCATGA TGTAAGC AATGA	TAATTTTGT ACATAATGT TAATACATT CTAGTGAAT TTGTGAAA AGTGACATT TGATTGTTA TATACATGT CTTACACAT GACATAAC TTGTATTAA ATTAAATAT GTATATGTT TGCTCTTGC ACATTAGG AGATAAGT GAAAATGA ATATATAAT GCTATTGAC GTATCCTAT ATTACACGT ACATGTATT GTACACAA TGTAATCCT ACAGGATC AAATAAAA ATTCAATTC AATTCAA	MAPIEKCSGTVLSSKRQKISSSSCKECEKRVMKGIMCVK CNYWLHEKCAKINLKYISDDFPWTCHACLSVETADLVS LVESKNEIIKLLQNDIEALKTEILCLKEENARLLQEKQCT VGEIHQEKQQDCANTCAREIQQTKKETRENTYARNTDV VNENISSAVDVTVNPTCKKPADKYKWKLVTYSKRKEK TTARPTQEDENEFLIIGDSLLKNVAVPNSKIDVRPGIRTQ QLGKHFKNLENAKSAEPDCEPKHNYRGVFVHVGTNSL RSSSEEEIINDTRNVIRSARSVYARSKLVISGIIKRRSVSD KYISQINCAIKEQCDSLGAIFIDPNKFLCENCLGKDGLHL NRLGXVTLSRMFTDVCKIIRNEGNGKLRWTDGGDKSDI LAEKANFKHHSKVKEIQNRIPEQMPQHLSQQXENYXKV LHQNIQSIGNKKLEVEILLSELQPDILCISEHGLTENQIDN LKISDYTLASYFCRSKQKSGGVSIYIKTGKLETANNKGV KLPTVKEKDFEITGVMIEGLSVFCVYRSPSGDISIFLKKIC MNMKRNIIICGDINIDMGKGSKIRQDFEELLLQYHMEAT ISTPTRVTSHSQSVIDNITNISRSDYESHVIQTAITDHDGQ MFYLCRRHGRACETKAKIKEIRVMSEQNINTFRTILSRET WTETLQAQTVNEKYDAFIATIKFHFEVAFPKVKRQIKVQ QQQQWITREVLEQRRKLQELRQVGDVEGYKTLQKTYR KTIRKAKANSINTILNSKNKIKAAWNVINTERKENAGP NKETEIRSMKIDNRIVTDGSKIANILNDNFINVVENLKLE RKSPKKKGTKVTKYSYSTMFLRPVTEDELRKTIQKLKA KKSAGDDEISNHVIQLVSEQHTHTISRYSKQLFQRWNFS REIEDKVSNTLTPLVDIANSSFRDGIFPEKLKTSKVVPVY KKGNKDDPNNYRPVSLISGFSKILEMLMYNRLVDYLDK HNILVPSQHGFRMGKSTESAIASLTEYVIQCLDEKRVVA AMFLDLSKAFDTIEHDILIKKLENYGIRGLSGEWIESYLC NRKQYVSIGNSFRSETKTIKYGVPQGSVIGPLLFNLYVN DIGGEEKEKKILYADDMTVLNHGQNMQQLERSAFISAN TTAQWLTENELVINLKKTMFTVFRNKEKSVQMNLKVN DMYLEEVESTKFLGVIVDNELKWKEHINYVCKKLSSVIF IMKQLSQYADKQLLCTVYHGLFEPYLQYGVTVWGNSNI QETKKIVHFTKESNQDHTKKKELGDLQCGETQTSKKRK KLFTLQKKAIRIILRRKNSETCRNCFKSLGILTFTSLYIYK	cd01650	933	1188

						TIMYITKEISLYDTNASVHTHNTRGKNEIRNIPHRLTILEK GPRYSGIKFLRNLPKNLRDNIHGNNFSRKLKDYLMEKEF YSIEEYLLN				
Ving i	Vingi-28_LMi	.	Locusta migratoria	CGTACCC TGGGCGT TAGGAGA AATCCTW ACGGCAG ATCCTTC CATCACM AGGACCA AAATCAA GACGTAT CTCAWCC ACAACGG CTTTTCTC AAAGCMA KMTCTCG MAGKAAT ATCACTG AATAAAG AAGTSMT ACTMTCC GACAGCT GCAGTAA CACAGCT TGTKATA TTTTACTG CTACAGG AGACCCA CAGAGGG CCACAAC AGCAAAG GCCAAAA GTAAATG GA	TAGCAGAA ACTAGCAG CGTCCTACA GAATTAGTC ACCTCTTTC CATATTTTA TGTTTCTAG TTTGTAATT TGTGTATTT TAATTGTAT ATTTATTGA ATTTTGCAT TCTTGGATT GTAAACAT ATATTTAGT GAATTTGGT GTCCCTAGA TTGTAAATT GTTTATTAT AATTGTATA TTTATTGAT CGAAAGAT ATGTCTTAG TCCTGCACA CGTAAATA AATAAAAA	MDLVLERPHTKYGSAIFVKSGLSIQSTGLTSENDVEILTV GFQSCTVTSVYKPPNAEFKFVEPLNFQSQPIKVVAGDFN CHSTSWGYREXDKNGEDLEAWAEQMGLVLIHDPKLPS SFNSGRWRGQYNPDNIFVSNHIAQQAKKGVLGPLPHTQ HRPVTCTVAAVVKPETVPFKRRFNFKRAQWEKFSIELD KEVIKLEPSPEMYDSFIQLVRTISRRNIPRGCRTSYIPGLD KDSKSLLQKYEHLFETDPFSEETTQAGEELLNAVSATRX QKWCDMVTQLDMKQNSRKAWKLLKRLNSDPTSAPPSL VNVTPDQIAHQLLLNGKTKRRTAAKKFQRYTDEEDYLG TPFTIEELEVVFKSMKNNKAAGLDDLRPEQIKMFGPVTK QWILNLLNNCISKLQIPKIWRKAKVVALLKPGKEPTDPK HFRPVSLLCQLYKVLERLILNRISEHVDKNLINEQSGFRP GKSCCGQVLNLTQYIEDGYERGQVTGVAFIDLSAAYDT VNHKKLMRKVYSVTKDYRLTQFIQCLLQNRRFFCNSAH SSSSVCCKIGGFFVTLHNKKSRWRTLKNGLPQGSVLAPL LYNIYTNDQPISSEVRTFIYADDTATAAQGKTFEEVEMK LTAALEDLGKYYDDNHLKPNPGKTQVCAFHLRNKEAR KKLKIXWRGQQLLHYDTPKYLGVKLDRSLTFKDHCQE TKMKVSARNSIIRKLTGSTWGSQPSVLRTSALALCFSAA EYAAPVWRNSTHARE VDTALNETGRIVTGCLRPTPVRK IYEIMGIAPPDIRRRTAAETEKKKQETDPRHPLFGHEPRK PRLKSRRSFMSTTTSTALRPGTRRTELWQNSSNERAGEN IKEEPAEGFHLPYATWKALNRLRTGVSRCKTNLRKWGF TNEDELCECGDVQDPQHLLVCRNLPNPCTIEDLHAAND VAIKTAEFWSLRSI	cd01650	408	681
Nimb	Nimb-40_LMi	.	Locusta migratoria	CCCCAGG GGGCCCA CAGCTCT TTTGTGG	TAAGATTTTT ATTGCTGAC GCTTGGAC AGGATGGG	MKRVKLDDNGRDTAVSEGRKKNNAPSYDKRTFPSLSTP WEEGQARRLGVKPFPRYLVCTRTDGGTFTATKPLFFVE HIEDKFGEVETMSKMRTGSLLIKTVSAAQSAALCACDH LGDVPVSITPHQSLNMVQGVIFHRDLLLLQTDEELRANLE	cd01650	101 2	129 1

GTACGTG	TGGGTTTGT	RRGVHFVRRVQRGPKDHRVDTGAFILAFEGDTLPEKVK	
TGTGGCG	GGCCCTCCC	VMIYHCCDVKPYVPPPMRCFRCMRFGHMSSRCTAVPV	
CGCACGG	TGCGAGCA	CGECGRPLHEGNPCVPPPVCVNCHEHHSSRSPECPVYK	
GTCCCCG	GCAGGGTTT	KEKKIQEYKSLDRLTYQEARRKFERLHPVQLSTSYASVT	
AGCCATT	GGAGCTCTC	SFTPPPSSSHSHPILPPPPPAPPLPPSQRPPSSSQPEKCSPSL	
GCAGCCT	CCCTCCCCT	ALAGEGAPSQGSSSRRPPDRRSSASSRQPDIQADNTKVA	
TTCTTCTT	CCCCCTGAC	RSSSVPDLTDSSSLSGTCPSRPAREKEKEKRKKKQKSQD	
ACATTCC	GACTTCCAT	KAPPVVLEAPPLSSQPNNPTELVSDPLFMDVVPPSSVTD	
TGGGCTG	TTCTCCTTT	VDPALPSFVTCISPFAFHMDPRSVLIQWNCNGYYRRLLE	
CATTACC	TAACTTGGT	LKHLIASYSAACVVLQESHFTDGHSPTLRGFRAFCRNRV	
TCCCCCA	TTTAATTCA	RPLQASGGVCTLVRTDVVSSWIPLRTTLEVVAVRVHLH	
GTTCCTCT	CGTTGTTTC	TPLTICNLYLPPDRPPTSAALTALLQQLPPPFLILGDFNAH	
CCTCCTC	CCCCTTTTT	HLSWGSAYTTSRGRLVDQFIADHDLCLLNDGSPTHFSA	
CCCATCC	ACATTTTTA	AHGTFSSIDLSLTSPPPPFSLSFAGHRMTTFATVTTFRCRS	
TAATTCC	GGCTTCTCA	RHLFLHRPFTHFSPPPLFPFFRWSPHDDLCDSDHFPLILSL	
TCCCCGT	TTTAGGAGC	PSHWTRDRIPRWVFRHADWPLYTSQVVFTPSPSDSIDEV	
CCCTCTC	ATACACTGT	ISDVSNAITRAASIAVPQSTGPPRRRPVPWWNATIASAIR	
ATCTCCC	TCTCCTGTA	ERRRALQHLKRHPSRAALVNFKRLRAKARYIIKQSKRV	
CCCCTCT	TCTGCTTGT	CWERFLSSLGSSVPLSRVWAMLRSIQGCQRPSSIPGLALP	
CTCGGTG	CATTAATTA	DGVCTDPLVIAEHLAAHFAKASATASYPVAFRLKKQQA	
GTTTCCTT	ACGGGGAT	EASNLCFTTHREEPYNKPFTEWELRLALSSSHDSAPGPD	
TAAAACG	TCTAGATTT	SIHNQLLKHLSPPQRQYLLQVFNHIWLQGVFPSQWRDSI	
GCACGGC	GTCGCGAC	VVPVIKPGKDPSSLDSYRPISLTNVPCKLLERMVARRLC	
TATTCCT	GTTCCTTGC	WVLESRGLLSPYQCGFREGRSPIDHLLRLETAIRKAFSK	
ATGGTTC	AGAATCCC	RQHLVAVFFDLRKAYDTAWRRHILCTLHDWAYEAPYQ	
CGATGAT	GCTTTTTCT	FLSVSFYPAACSVLGLRGSLPIFISQFLSRRMFRVRVGSSF	
GGTTGTA	ATTGGTTTT	SSPQIQENGVPQGSVLSVTLFLIAINGLVSSVGSLVTPAL	
GCAATTT	AGTCCTGTT	YVDDFCIWYSSHSVASAERQLQGAIRRASTWALSHGFQ	
TGATCTT	CTACTTGAT	FSPSKTRVMHFCRRTTVHPDPELFLGTQRLEVVAQHRFL	
CGTCGTT	TTTAACCCC	GLLFDKKLTWLPHIRQLKISCMRKLNALRFLAHSSWGA	
TTCCTCTG	CTTTTAGTT	DRTTLIHIYRALIQSRLDYGCQVYGSAAPSALKLLDPVH	
TTTTCGG	CCTCTATTT	HCGIRLATGAFRTSPVDSLLAEAGIPPLQTRRHQLLVSY	
CATAGTT	TACTCTCGT	AIAIRKFPGHPTYHVLFEGDGLRTPDTRPRVGLPVGLRL	
CGGTCCC	CACGGTCTT	ASLCRDFNLPSLDCMRRAIPRTPPWLVPRPQIRLDLYHA	
CCTCGGG	AGCCGACC	PKASVSPIIFRRLLHTVLEEFQGATIFYTDGSKFAGHTGY	
GTTTGAC	ACACCCTG	AFTSPIGVECHSLPESCSVFTAELTAISRALHFIKQVALDR	
CTCCACA	GCTTGCACT	VLICSDSMSSLQAIDRCYSSHPLVIDIHDLLADLRRTACS	
AAAATTC	CCATACAG	VAFLWVPSHVGILGNEQADQLARVAVTPCPFPWVIPGA	
ACAGTTA	AGGTCATG	DLRVQMRSLLIQDWNTNWQATALSNKLRSIKSTTAAW	
CTAGTGT	GGACTGAT	SSSLRPSRRESTVLCRLRIGHSRLTHGFLLRNEPPPHCGC	

				GAGCCATTTGGGGAAGAACTCCCGACCTAGCCCCTATGGCGTGGATCCCAATCCTTTCCTTTCCCACTTCCTTCCTTGCCATAGCCCCCTCCCCCGCTAGGTCCCCAGCACGTGTAGCCAGTCCGTGTGGGGGGGGGGGGGGGCTGTTATGTACCCTTATGGCTGAGCCCCCTGACAACGCAGGGATCACACTGCTGGTACCTGAGCTGCAACCTCCCCATGCAGGCCAAGGAGTCGATACCCGTCCTCGTGGGGTATCAGGGCTCCCGGCAATGGCCATCG	GACCTAGCTGTTTGGTCCCATCCACCCATCAACCAACCAACCAAC	GALQSVAHILVECPLLLALRAKYGIPHSLPSVLANDSRIADRVLGFIRESGFYSRL			

									752
								488	
							cd016 50		
						MDNYRILKGRPAIKVKKTEKTKNVVLILGTAFFIHNSITE SIVEFRSPSPRLPVVTLASGNKLYTIFNAHAPINEDNRKN PEKVEEFWSQLEEELFRTPDKHVKILMGDFNAQIGKERK FRNIVGKYVAHRRTNRNGERLIDVCESFGMKIMSTQFK RAPKQMKTWRSPNPMLGEFQIDHVATTSPKEIMNVRVQ KGANFDSDHYLTRVKVKPIPGNRRRKTPRIQKYRIEELK ISEEIVVKYQHELDTLNSNEWDKMAQKIQEAAKQTIFLA KKKKHPWWDDECEDALRKRSEAWNKWNSTKQNFEEF KTQRRTTAKVFRNAKRKFERDQLKELEENFQKNNTRSF YQTFKSKLRGYSSPNICLKMIRANRLLITQHLFKNDQGK LIMNNKHNCEHLARYFKELLNCDPPPEKLTFQKPQNKL PDSTPPDKKEIREIIRGLKNNKVSGKDSIVAELLKYAGDS FLERLERLFQKIWLTGEIPEEWRDAVIHPLHKKGDKSNV NNYRGISLLPVAYKILAIALKHRIEQHIDQQLGEYQAGF RTGRSCAEQIFTIKTLIKTKALAADKKVIMTFVDFKKAY DSVDRDTLIEVLKEFGIDDKTTRLIQQTLTNTRSQIKFLG ELSEPFEIKTGLRQGDSLSPILFNCVLEKVIREWRKEMSQ FNFPNGIRLGRKQSGLTFECLAFADDIVFFAENLQIATKQ IEVLKETAVKAGLQISFEKTQYMNINATDPVWTMRTKY GDIKRVQRFKYLGEILTPNNNEKAAIEERARKMEIAFRL CQNTYKSKSLSYQAKFRHYSSVVKPECLYASETIARKGL SDLEKKERKFLRKILGPRKISNEFTFRLRSNQELYQRFEN ITTTMKKRRLSFYGHIKRMKSERLANKILLFQENRKTQV PWVKEVHRDLEKCGITAEEIVNRQIFRKKVAEVKDFFEE KTRKNRVFSAEERARVSQRMKEYWRKRKEENFRK		752
				TGAGGGAA GTTGTGTAA ACGTAACC CATAGATG GCTGTAAC GAAAATAA ATAAAAAA					
		Locusta migratoria	.	TGCCAGG TGGCCTT TGCTGTG GCTGGGT GGCGCCC ATGGGGA GAGCCCC GGATCGG AGTGGGT GGTACCA GGGCGGA CGTTTCG CACG					
				GGTAGGG TTTCACC CTATCCC ACAATCT AAACCCA CGCAACT ACCATGA TGGCACA ACTACGG AAACTAG TACCCCA GCCGGTT CAATCCG GGGCCAA AATTTTG GTTCAAG TAGGGCA ATGGATT CTGGGGG CCCTACA CCATCAT GTTTGGA TCAGTCA GAGGATC CTTCATC CCGCAAA CTCAGAA					
	RTE-68_LMi								
RTE									

				CGAAGGA GAAACAT TTCTTTGG TACACTT AACGTGA ACACCCT ATTGAAA ATTGGAA AACAAAA GGAACTG TCAAAGT ACTTGAT GAATACA AAATAAA CTTCTGG CAGTCAA GAAACAA GATTCAC AACAGAT GAAGTGT CAGCT					
Ving i	Vingi - 3_LM i	.	Locusta migratori a	GGGGACA AGGAAGG AGCCTCC CAGAAGA TTGTACC GTACATG TACAGTC GGGCGTC CCCTGGG CAACGGT AAACACC GGCAAAT CCTTCCA CACACTG AGCTACG TCTAAAC CGACATT GTATTCA	TGAGCCTGC TATGGACA ATAATGAC GTATATGTG TGAAAATC ATGAGTAT ATATGAAT GTATAATTA GTACTTAGT TTTTATTTG TAATCTAGT ATATGTAGC CTATTAATT AACTACAG ATGTAGCTC AGACACGA TTAAATAA AAATAAA	MMKGPFSEPCLTAISLNIEGLSTEKEQLISNTCKLHSCDV LCLQETHRGPSQRRPRIPGMILVKECPHDQYGSALFIRPG IKVSNIRFNTVNNTEIISVIIGSITVSSIYKPPGERFAFSQHV CHPQLVMGDFNCHSKVWGYQDTNEDGELLESWAEEQE LNLIHDQKQPASFNSARWKRGYNPDLIFVSKTLSGSCNK FVLNPIPKTQHRPIMVKFRPVIVPQEVSFKRRFNFKRAN WKKFADILDKEIVYLPPNAENYEKFVEAVKRSARLSIPR GCRTLYIPGISNKMKQLYNEYIEQYNHDPFSVETIKLGE QLTETISKDRQESWKHMLESLDMKHSSRRAWNLLKRLS XDPAAPKDQVSVTANQIAHQLLLNGKPTRNSKIVTSYK RPQVDEIPPVPYFTSIELETAISSLKSRKAPGPDDVVNEFI QNIGPIAKEWLLQLYNFCWANKSVPKLWLKAKVVAVP KPGKSRDDAKNYRPISLLCCPFKLYERLILNKIGPQLESQ IIPQQAGFRPGKSCTGQVLCLTQYIEDGFQKGEKTGVVF VDLSAAYXTINHKILLENFSKKIFQNEGNYHLTEVVRSLI SNRRYYVEFQEQRSRWRNQKNGLPQGSVLSPLLFNLYT NDQPAGPDARSFIYADDVAIACQSRSISHIERNLSDALEV LTEYYNRNQLRPNPAKTQTCLFHLNNQEAKRELQLEW NSVKLKHCPNPVYLGVTLDRTLSYKSHVEKTRAKIYTR	cd016 50	463	722

Daphne	Daphne-28_LMi	.	Locusta migratoria	GAAGATG GCTGCCG ATCAGTA TACTCAT TTTGTGC GCGCCGT TGTGAAT AGTAAAA ATCGCGT GTTCAAG TGTTCTCT GTGCGAG AAATACA ATGCTAT ATGTGAT AATTTAT GCGTGCC GTGTCAC ATCTTAA TAAAAAA CAGTGGA AAAAATG AACAGGT AAATCCT CGTGTAG GTCTTTGT ACAAGTC TTAGTGA TCAGCGT GGTAAAC ATGGACA TCTGCAT CAAAATG	TAATTTCCA CTGTCTGTG ACCGAGAC ACCACTAA AACTGCCTA TTTTAGTTA CTTTCATGC ATAATGTTG TATGGCATA GTTTTCTGG GGAATCAG CCACTTTCC AAAAAAAA TATTCATTG CACAAAAG AGAGCGAT CAGGATAA TGAGCGGA GTCCAACCT AGACATTCT GTAGAAAC CTTTTTAGA AAATTGGG GATCTGAC AACTGCAT GTCAATATA TCTACTCCC TACTGTGCT TTATTGTCA AAAATGTG CACTTGTAC AAAACTAA TGATGTTAC	NGIIRKLTNSHWGANPEILRSSSLALCFAPAEYACPVWR RSTHAHKLDTALNDSCRLITGCLKPTNRDLLYMCSGIAP PQIRRQTWRWLKEADSARIEDIPCSDISQWCKIKIEERFH FEHPLNESQSTTRESMWTQKLEKKKKKKNSLQCWNIKEK LPAGSQLEWRVVWKSLNRLRCRMGRSKDNLFKWGYAE DELCQCGEKQTMTHLLSCPSLPAPCTFQDLWLANDAAV KCAEHWREI MPRKFFRPKMFTHGKPHETVLQVPPNNRFSVFEENTNT VDSFATVLNNERVSANTSRKIGKKCVKVNIHADSHGRG LAKILCEDXDVDAYSYVKPGAPLYEIGKNIKDGSPMNS NEYTVIVGGANDVYKNELSSAIQNLNIILESNNSNAMIIL GIPHRHDLIYESCVNKEIAKANKQFENICNLYPNTQFLPI DFLQRDSFTRHGLHLSRKGKKLLCFKIAEMVSANTKKG SNSQKTVQTSVPQQPTQSPSAQLYNINATTKTPTDDPILP ESTNSPPSVDSIEPTSVAAMPGLIVDTTVTPASHQEEDIW ASRIVPEPMEEASSTPEDSSQLLKIQGFFMVSSQGRKNSV IATPPSKPNVVNTKREINPINHPKTLVSVNKLTVFHQNIQG LRSKLDEISINILDKDCIDSADVICFSEHFMKLDTDSLQL DSYNLAALYCRQSLEKGGVVIYVKNKIVHRSLDLSDYCI DQHCEVCGVVISTNLSEITVLSCYKAPAGKICIFLKQIVS VNSYFLQKQRILLIFFAKTKNIVLGDFNVDYLIDNMGK NDLSHVMSSFNLTAQVDFPTRVTANSISRMTISSPNDNIF LDKSRNPNTIVKRVLNGLSDHDGQLLSIEGVGPLSKNKS YSKTFRAINSQSINTFCDLLQETDWDQIEHLDNVDDKYN AFLNEFICLFEAAFPKRTAKPKHETYCHKNWLTRGIKTS CRRKRNLYASLKSCFSLEKEKHYKLYCSILKKVIRAKSN SIKSEIDVSSNKIKTIWGIIKRETGKIPITENSLEIKKMMSF MIMLKKNDVIHDNVEITDIFNSHFLSSAAQTGKTCSINE AMLFLQKSIKKQSKLTCLLSQCMKLRLKTKQINMSPVT VHEIEVIIKEMKSKNSVGIDGISSRILKHCYKPVSKILCHI FNKSIQQGIVPERLKYAIVKPLFKKGDKTDVTNYRPISLL TIFSKILEKLVHKRFIDHLTQYKLLNSSQFGFQKGISTDQ AIFSFTNQILESFNNKLSPVGIFCDLSKAFDSVNHEILLEK AKHYGFIGSAGKWIKSYLTGRKQKVMLNNSEGHSVSSS WGSIGCGVPQGSVLGPLLFIFINDLPLCTSLKTHFTLFA DDTSILINESSNYNLEESANIVLREALNWFVINGLALNAN KTQIIHFQPSKKMEDPITIKCNGTDLLQAESTKFLGIHIDS NLKWSEHILQLHKKLSSATFAIR	cd01650	924	1185

				TGGACGAAATTCTGAGTGAATTAGCAACGAAAGAAAAAATTATCAACATCCTGTGTGACGACATAGCAAAGTTGCGACAGGAATATCATATTTTCAATCAACAAACAGTGTGAATACATCGCCAAATATAAATTGTGCTGAGTTTATTAGGCCTAAAAAGTTCACGAGTGGTAAACTACACGAAACGGTCGCGCAAGTGCCTTCAAATAACCGTTTCAGTGTCCTTGAAGAKAACACTGATACTGTTGATTTGTATTGAAAAACAAAAGTTTG	CATACTCACAATACCAGGAGGAAGCTAGATCTCCATTATGAATTAAAAAACCTAATACTGTACAAAAAGGAGTCCACTATATGAGCTGCAAAGTGTTCAATGCTCTTCCACTATCCTGAAATCACTTGTCCATGAGCCTCCCAAATTTAAACAACAGCTCAAACAATATCTATAGATAAGTCCTTTTATTCAGTTGAAGAATATTTCAATTTGTGTTAACAATTTATTTTTTAATTTACAAGATGTTATTTTTGGTAGCATGTATTGTAATTTTGTCAATATACTATTCTTGCTGTGAGATCTACCTTTTTTATGTAATTTTATTCTTTTAATTTTTTGTGTAATGTA				

				TCTACAAAT	ATTTTTGTAATTTTAATGTGTAATTTTTGTCTATACCTTTAATGTGTAATTTGTCTATTGCAAGAGGGATTCTGTATGTTCCTTTTGTAACAATTTGACACGTCCTACATCCAGTGAATGTCTCGCTGTATTGGATCCATGGGATTAAATAAATAAATAAATAA				
Penelope	Penelope-3_LMi	.	Locusta migratoria	AG	TAATGTAAACACCAATCCTGGACTTCCTTTTCCACCAATAGGAAGAAATAAAGCAAACACCAATGATAAGAGAGCTGAAACCACCAATCATAACTGGCAACATTTAATATTTAAGACACCAGCTTTCCCTCATCCAGCAGTAAACGTAACACCCAGAAGATGGCCGCTTGCAATAGT	SNRAKNICAQASKALLRERIRHIRSSLDSSNRKLIELHLALSSKLAEEDWEKVDRLTFQQAERTNDSVATRQMKKFDNLQQKTTGNGQTMDTQRTVVNLSSHELSDAATQILSKGMNFAITPRRVPKEDIIESVEASLRHLPQAEAEKIRQETVRVLAHAKPPKQNISNEERRAIIELKNNPDLVVVQSDKGNATVVMNTTDYDDRMEKLLSEPIYKKLKGDPTAKVIKQTQALLKKTRMNFDTARRLNPQVPQAPRIYGVPKIHRESFPLRPIVCSIDSPTYRLAKELVPKLQHLVGHTDSYVKDSTHFVALLEEKRISTTDLMVSFDVKSLFTNVPVSDTMNILEERVPPDVCDLVQHCLTTTYFRWKGEFYEQTDGVAMGSPLSPVAADIFMEAFEETALQSAPLRPDCWLRYVDDTFVIWPHGEEELQNFHQHLNQQHNKIQFTMEIEKNGVLPFLDVEVYRKPEGKLGHRVYRKATSTGRYLHATSHHHPTQKKSVLHTLTKRAHRISDEQHLKAELQNIRSIFEANGYGKRMIEQTMTSNDGNKREEQKEEQTIALLPYVKGVTERVGKILRREGIKPIFRSNNKIRDVLGSMKDKTDKLHTAGVYEIGCECGQVYIGETGRPINTRIAEHERYIRLGQHTKSAVAEHQEECGKQIGFQEVTVLAKQPYTLKRKIREAIEIAKRPGNMNREDGYRLPASWLPAVAALRDDTARSAVHRGPHSPETTAAARTQSKQHALAGARHQCYVTATTAPHSPVSTNGKH	cd00304	380	464

				GGCCGAAA CGTCGGTG AAATATTTG ATATATTGA CGATGCGG TCCAATACC CAGAAGAG TTTTATCAA CAA	MARNCDCCERMEAELVTLRSQLQAALASLATLETAVG				
					GRRCCGPLAGVQQTPCASVVPSDRPAAVASPLTAREGV				
					APQSVVDWEVSSARVRRRKTSRGTEVEASPVSLANRFE				
					ALSVAEDVPDADGVACPVLAPLPSRSGHSQRVGLLVVG				
					SSNVRRVMEPLRKMAAREGKKSNVHSVCIPGGVIPDVE				
					RVLPDAMKSTGCSQLQVVAHVGTNDVCRFGSEEILSGF				
					ERLSEVVKTASLACKMKAELTICSIVDRTDCGPLVQSRV				
					EGLNQRLRRFCDRVGCRFLDLRQRVVGFRVPLNRSGVH				
					YTQEAATRVAGAVWRGLGGFLGVRSPLYSGSGYTGSG				
					GCVAWTGRFFRLEGLGNTQNGFQSQRGQVRHRKKADL				
					GTIGVTVVNCRSCVGKVPELQALIESTEAQIVIGSESWLT				
					PEISSAEIFSKGQMVFRKDRINKKGGGVFIAVTNGLSCSE				
					IEADTSCELVWVEVIPDNRNKLIIGSFYRPPDSDDAVAE				
					MFKESLNHISNRYPKHTVIIGGDFNLPSICWHNYVFKTG				
					GKYKTSSELVLNAFSENYFEQLVHEPTRIVSGCENTLDL				
					LATNNPEIVKSIKTDTGISEHRVVVARLNVEIPKCTKNKR				
					KIYLFNKADKNSLHAFLSDSLQSFKTKNASVDQLWLTF				
					KDIISVAIERFIPNKLIKEGADPPWYTKQVRSLLQKQRKT				
					HAKFKKTQNPQDWQSFTETRNLARTSMREAYSSYHNEI				
					LSRNLAENPKKFWSYVKYTSGKTQTIPSLRDSNGNTTD				
					DSTAKAELLNTAFRNSFTKEDEVNTPVFNPRTAASMRN				
					LEVDVPGVVKQLKSLNKNKSSGPDCIPTRFLSEYADAIA				
				TGAAAGTG	PYLAIIYNRSLSERSVPKDWKVAQVTPIFKKGNRSDPQN				
				GTTAATGA	YRPSHSPKLQAISLTSICSRILEHILCSNIMNYLEENDLLT				
				ACCCTCTGC	NSQHGFRKYRSCETQLALYSHEVLSAIDKGSQIDSIFLDF				
				CAGGCACTT	RKAFDTVPHKRLLMKLRAYGISPQLCDWIHDFLSERTQF				
				AAATGTGA	VVIDGKSSSSKTEVISGVPREALNRSDIWRSQGSVIGPLLF				
				TTTGCAGAG	LVFINDLGDNLNSSLRLFADDAVVYRLVKSTEDQKQLQ				
	CR1-3_LMi	.	Locusta migratoria	TATCTATGT	SDLDKISTWCENWQLTLNKEKCEVIHMTTKRNPLNFNY	cd01650	871	114	1
				AGATGTAG	AINQSNLKVVNSTKYLGVTITNNLTGENTGRKFYQIPRS				
CR1				ATG	YNYEQLDWREHIENIVGKANQRLRFIGRTLRRCTSSTKE				

				ATTCGCC GTTAATA AACAGTT TAGTGTT AGGTTCA TATTTTTC GAAGCTC ACTGTTT ATCAAAC TTCTGTGT CCTCAGA CGTACTT ACTTGGA ACATACA TAGGTTT TTAGA		TAYITLVRPLLEYCCPLWDPHQVGLTEDIEKVQRRAARF VLSRYGGEASRKCCTNWGGNHEGSEKGSSFCFITLRGG SVTEMLHELGWKSLKERRYSLRRDLFAKFQSPTFTSECE NILLAPTYLGRHDHPNKIREIRSKTERFRCSFFPRAIREW NGREKL			
Dap hne	Daph ne- 41_L Mi	.	Locusta migratori a	T	TAATGTGAC AATCAAAA TTTAATTTA TAACCTACA CTCTTTTAT TGTACATAT TTATGTTAA AAAGCTAA TCATGTAAC AATATATAT TTTAATATG TAACATTTA TATATGTTA ACTTTTGTA CCATAATGT ATGCAATG ACGATGCCT TATACAAC ATCACAGTT GTCCACAG GCAAAATA AATCAATTT CAATTTCAA T	EQQCMPQGTCSDDQACPHKSWSVKSQPSFLTTGMNLC NYKPVQRLEVKDCGSMSFFHLNIQSVRNKLPELELFLLE FQCDILCLMSTGHTLSNEHWLKEHEHNMYTPLGYKFLT GFFRSNIEKGGVSILMKNDLNCKIRIVDLKHLASEGQIEI AAIHLPPSADTLKSHNFSCLNDKPTRQNACLDNVISNLS RDDVNLNVVKPMLSDHDGILLKLIVPKKDHRPTQPIVN RPVRQLKPINITNLTQRLLTFDWDATVMENTTLNNAFD YPPQLIIMLLIIRPNLFNEYCPPPPSTKVKYKEETKEKLTM VHPHLNSLRILLKIYYEKSCTDSEYRHSYNKLKNIYKKDI ISAKKEANGTYISQATNKCQAAWKIIKAEAGLQVVNPPI PRADELNKNFINSGQSAIDHSNLDDNADSFLSSLETELX VDKLKFRWHAIYPKDILDVARSLSNSRTEDIYGLSNFTIK RIIEAICIPLCQLINRMFLEGTYPDCLKQTIVLPIYKKGDR TSPQNYRPIAIVPITSKIIETIMKNQLNDFFCNSLLNKAQY GFRSGLSTVKAVQEVVHSVLHGFENRQHSYATLVDLSK AFDTVSHGILTKKLETYGITEKELALLKSYLANRNQMVL VNSTNKSDPLKVQYGVPQGSVLGPFLFIVYVNDLPNLLP FKSVMYADDTTFITSHNDVNAVKQMNNSMMEKATQW FQENRLTLNNNKTEHILFSLGNINNEQNHPVKLLGIYLD SKLCWNTHTEKLCSKLARVLYLLRKLKDCVSKELMLM AYYAFFHAPYIMEYFFGETPLERKKSSSGKRKPSDAFLA RTHAYGILCFFPCTLHYGILLWGNSAGAKEVFLWQKKA IRCIFGLSKLTSCKQYFVQHKILTVPSLYVLKALMHARQ NVDEYESRSDVHLYGTRNRNLLDIPWARLSKVKNSFLH	cd016 50	498	740

						MAKIFFNKLPYNIRELHENKFKIVIETWLKEKALYNTNE FIWTENKSCY			
Pene lope	Penel ope- 37_L Mi	.	Locusta migratori a	TTAAGAC GGCCACT GCACGGA AGATTAT GGAGAAG GGAGCCA TGCTATT GTGAGAG AGAATTA GTTATAC AGGCAAA GGTTGGA TACATTT CAGGAAC TGTTCCCT TGACATC GAAGATT GCAGCCT TTCCCCC CGGGACT GGGTTGA TGGAGTT ACTTCGC CAACTTT GATTGGG AATTTTG GATGCAA GTTCGGG ATGTCAA AATTGAT CGCATGG TGCCACA ACAAGCT TCAACAC AGGAAGT TACTGTA CAACGTT CCATTTT	TAAAAGAA TCCGTTGAA ATACGCATT GCGGAAAA TCTGATGAA CCGTGACA GCGGCTAC CAGCTGAA TAATGCGTG GAATCCCGT CATCTCCGA GATTTGTTC TAGACGAA GACGCCAG AAGACTCC GATAACTG CGGCAGCG GCACTATCG AAGACAGC TGATCCTTG CAGTTCCAC CAGCGAGG GCGCTGCC GCCGGGCG GTGTGGTCC TTCTGTCAC CGTGTCTTC GACGCATG CGCGGTAA TACCACCA GCGCGCTAT ATAAGGCG GAGCGGAG AGCGTCTTC GTCAGTCCT TGTTGGCTC ACCTGAAG ATGACTGG	KLLLNAVIQLLRKGLHYEIPRESGIVVLKADKGNSTVLL PRRLYLEKMYGLLSDTMYRKIEHDPTARVQRKTCTLLN SSSLAPETSSGPYSTCATEDVHTFEFFLVSSRDIKRLRPH GSVPPRLYGLPKVHKEGIPLRPIVSNIDAPTYDLAKHLAS LLRPLVGKCTHHIRNSADFINRLGALHLSSSDLLVSFDV VSLFTKVPLADSLALIGKRFQSDITGLFRHALSSTYFMFN QEYFEQTDGVAMGSPLSPLVANFFMEDFEERALESAVC KPTVFWRYVDDTFIVWPHGEDKLMEFLHHLNSIHENIRF TXELEKDGCLPFLDVLVKRKSDGSVGHSVYRKPTHTDS YLHSSSCHHPSQTMSVLKTLVHRAHTVSDAESLPKELA HLRTVFRDNGYSTRQINRAFSPKTRNREVDKEENAPAK SLAFLPFVGNVSFKIGRILNNYQVKVIFRPPSKISTCWDQ DFDLLGSVKDDLLLRKAGIYKIPCECGMAYIGQTTRTW KNVAQNINVALAFCNPASLQLRSIVFPTDIQWSMIKHQF WPQQHPSGTPL	cd003 04	246	330

			AAATTTC GGAGAGA ATTTGGA CGTTGCA ACTGTCT GTTTTGG GCAAAGG ACTGAAG TTTTGTAC CTACACC CAGTGAC TGGGTTG AAGAAGC TGTACGA CATTCCT ATGGCAC TGCTGAA GAAATGG GAAACTT GCCGTGT TCTCTTG A	CAGGTGTCC AGTTGAAA TATCGTGGA ATGAAGTTT ACGACGAC CGGCTGCA AACCCGAA ATCTTTTTG AACA	TAATATTTC TGTCTGTTT ATAGTTCGT AGTTTACAT GTATGTAA GTAATCTCT GTATGTTAT ACCTTTATA AACTTCTGT ACACATGA TGCTTCTGA CACGAATA AAAAAA	VLCHIFKLFEKVLHDRLSKAIDNYLIPQQAGFRPGKSTTS QVINLTQLLEDGYENRKITGVAFIDLTAAYDTVNLRLLM KKLYAITRDFKFTSIIRSLLQNRRFSVEFQGKQSRWRAQ KNGLPQGSVLAPLLFNIYTNDQPLPAGTENYIYADDRA MAVQGDSFEEVEKRLTDALDALCQYYADSQLKPNPSK TQVCAFHLRNKQANRKLRVSWKDTPLEHCANPKYLGV TLDRALTYKQHCINTKMKVSGRNNILRKITNSSWGAKP EVVRTTALTVCFSAGEYASSVWYNSAHAKQVDIALNET CRLVTGCLKPTPIDKLYHLAGIAPPAVRREAAAYSERAR AELSESHLLLHNIQPATRRLKSRRSFLRATDNLSDQPGSR VDRWKRSRERHWMEPKEELAPGHNEDWTVVWKSLNRL RTNTARSRDNLMKWGYSTVSNLCDCGDVQTTQHMYN CPECPSHCTLEDLMLATPNAVDVARFWAKCI	cd01650	1	232
Vingi-9_LMi	.	Locusta migratoria	CA	GTCTCCC ACAAGGC CAACCCC TCGTCCA CGTCGTG			MDRRKVKILGLSETKWKGSGKKTLRNGYILHWCGSQT ETRNGVALLLHKDVAIHTDIRYVSDRIIQMKLHVDNTYI TIFQVYAPQTGCTADEKEQFLENLEDLIVEDNVIIMGDL NAQVGKERQGCENIIGPHGYGKRNPEGEQLLDLCVRNK LIIKNTWYQKRESHKITRYSWDGQHKTVIDYIITDQKAS	cd01650	497	746
RTE-24_LMi	.	Locusta migratoria								
RTE										

				GGAGACG GGCAACG GCGCGAA GTAAGGG ATTGTCT GTAAGGG TGAAGTA CAGTGGA GACTATG CGAGTTC CGCGACC GCTACGG CAGCTGT GAAGGCC TTACAGG AACCCTG AAAAGCG GCGGCTA ACGGGGC TCTGGTG AAGCCGT GTTAGGC GTAGCTG GCCAGCG GTGGATT AAAGCGT GCTTTCA AGACTAA AACAAGC CTCGGAT GTATAAA AGAAATT ACTGGAC ATGACAC CTCTCCA GGGACCT CACGAAA GTATGAC GGCCAAA AACCTGA		QTVKVIPSECMDSDHRLLTADLKNVRVPLVKYKKMPKI KTWLLRDPEKKQEYQKKIKDRIPKTEKLGVEEEWKLFK DTLVTEAVHICGKTSAAQREKETPWWNDRVKTAIKERN VLKKKLDREKQNRVDVRNDTEIERLSQEYRNKKLAVK RLIQNEKDKCWDDFTTKLQEDCQGNKKLLYKLVNDKR YERTNIVVLETEEGKLLRTESEIRSEMRNHFNSLLNGNV DSSDDSDCPAEETPEVLQPLTWLEVESALRNMRNGKSP GFDELSSDMMKAAGTPGLQWIYRVLSVIWTNNKIPDD WTKGIIIPLFKKGDRKKCGNYRGITLLSHTLKLLEKIIETR LRKIVEPLLEEEQHGFRKGRSTIDLIFAVRILAEKYWEHG KNLVMVFLDLEKAYDSVPRNKIWKCLEDLKVPNYLIDR VKMLYQTCNSSVQIGNGRSEWFETKRGVQQGSALSPLL FIIVMDRMLKSIKETANESRSLVFADDVVIWGETEKEVQ KSLTVWNNKFKEFNLKISKTKTVEMTINREGTSSNIYLE GTKIKSVSHFQYLGSIFSKDNTIRQEILNRTQKASNFYMQ VRNLLWDKKVPQAAKLTLYNTYFTPILIYGMETCTLLK KDYSRIQATEMRFIRTMNQSTRRDKIRNEVHRETAGIRV PIANILGKRRLQWYGHVMRMDSTRTAKKYFRMEVGGR RPRGRPRRRWIEAVKSDIEKNGFNWDDVQDQMLYQDR MRWRALVSHTRETGDGK			

				AGCATAT TGGAGAC SCCTGAA GACACAA AAACCAA CAAAGAA ACAAAAA CTCAGAC TTGGAAC TCTAAAC ATCCTCA CGCTCAC TGGTAAA TGTGAAG AGCTTGT TGACCTA					
Pene lope	Penel ope-54_L Mi	.	Locusta migratori a	CCAGGCA CAACATT ACCAGAC GTGAAAG GGCGCGC TGCGGTC GCTGCGT GATGATC CAGACGT CGTCGTC CTGCCGG CGGATAA GGGGAAC ATAACCG TTCTGCT AACAGCT GCGGAAT ACAAGGA GAAAATG TATAAGC TCCTAGA TGACCCT GCTTACA GGAAGAT	TGATGATG ATGGCACG TTACGCTAT CGAAAATT CGTTCCACA ACGATCAA TCCATCCGG CTGCACGCC CGAGAGAT CTTCACTCA	MGIPLRPIVSTIGAPTYQLAKHLTSVLSPQVGKCEHHIKN SAQFVERIRQLRLSNSDLMVSFDVVSLFTRVPLKESLELI EKKFGPSFTKVFNFVLTSTYFLFDGEYYEQTDGVPMGSP LSPVIANMYMEEFEDKALNTAPLKPTIFYRYVDDTFVIW PHGEDELQVFLQHLNSIHSQIQFTMEVEQNGQLPFLDVL VRRKPDGTLGHSVYRKETHTDQYLHASSCHHPSQREGV LTTLVHRARVISDEESLKPELDHLKEVFQENGYSDRQIR RAFKHKSTPQPTPEEDPPAKIAYIPYAGNVSGKIGRILRR QNIKCVYRPPAKIKSLLGTVKDDLGLRKSGVYHIPCECG KAYIGQTIRTIEERCTEHKRHTELQQPTKSAIAEHCLETX HKMKYEDTEVLHNTSRYWDCLYKEAIEIRLRENLINKD SGFQLSHAWNPALKFLRSRRKQNTEHSEGRAQTEVRHS SPEQELPPPPQHPGVRTRGQWRGVTAGIYSTAAGPAPQS AGVT	cd003 04	114	198

				CACGACGGATCCAACCAGCCGAATACAGCGCAAGACAATTGAACTTCTTCGGAGTACTTTCACAGACGAAACAGCTAAGAAACTCCGTCCTCGTGCGGCCGCACCACCAAGACTCTAGGGCTTCCAAAGGTACACAAG					
Penelope	Penelope-40_LMi	.	Locusta migratoria	AACACCTACATACCAAGAAACAAGAATAAATGTTGAACTTTACAAAATTCATCTGGAACTAGGCAACCACATCTCTCAGACATACTCAATGATTTAATGAATACAGTCAAACAAAATACAGAACAACTAATCAAAATGGG	TAAAACACACACACACACACACTAAAACGAAAGCAACATGATTCAAAACTCCCGCCCCCACCACAGTCAGCTGTTCTTAACACTACTCCCAGAATTCAACTGCCCACCATCTTGTTTTCATAACATAAACATTGAGTGAAAGTGAAAGTGACACAGTGCCTGTCCACA	MSGTKSAEEKTCRRIRNKLKNEGALITKADKGTPWSGCPHHKGRQGNTMVIMKQDDYRNKTLTFIEDNNIHKLTSDPTQRFHKNTQTLLKSITHIFPKQKIKWLKQKNPKAPTLNCLPKIHKEHIPVRPVVNFRNAPTYYLAREIHSYLQKHYVFTNNRSILNTSELIQKIKHIIIPSTATLTSFDITSMYTNIPITETINIIKQQLQQQTDTPKSHIKETTNILKLITEQNYFTFDNQFYSQEDGLPMGSPVSGLLANIFLNHIENKIFDTIVKPKQYKIIYWHRYVDDIICLLDETPSRIQELHNDINNVHPKIKFTIETENDRRLNFLDLSIHRDKNQHKFSIYRKPTTTSTAIHKHSNNPITQKQAGFQYMLHRLNKTPLSDEQYKNELQTIKQIATENGYDTNIIDKLNHKIKAKLKGTTNTQKQTDQKHNTSTQNINPLNSHTPNTHTQEQKSNKTTNKWHTLTYDNKATHRIGNILKKQGLKVAYRTDNSIQRKLKTTNTNTDKHNTSGIYQLTCQDCNSVYIGQTCRNFNTRYTEHRRALKSNSHSTFAEHLINMNHKPTDINTDLKILKHSKSLPQKLTIEENYHIQKAIIERKEIINEYTALCNGTIFSALKELTENP	cd00304	247	335

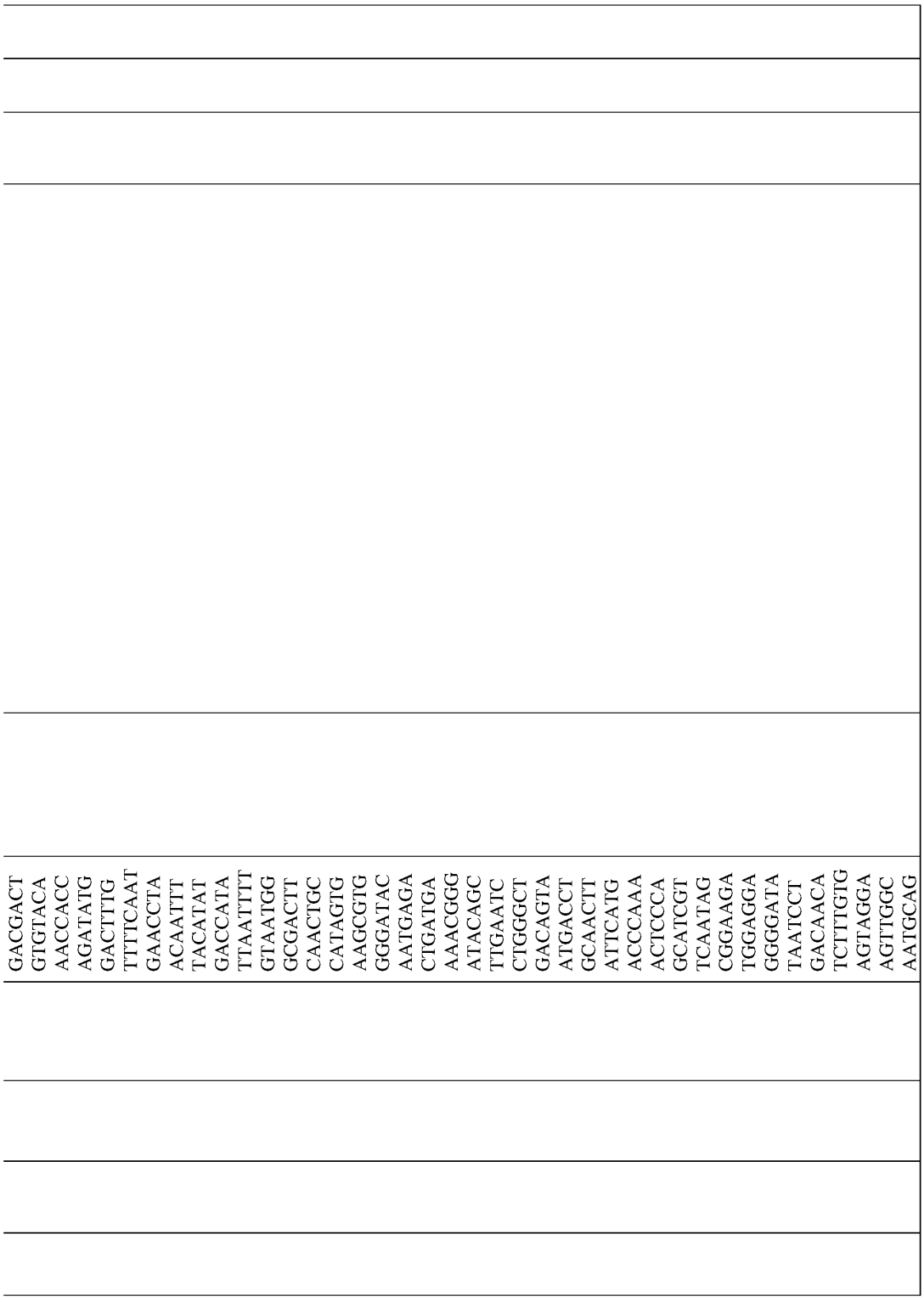
			AGAAACA ACAACAA AAAATAA ACACACT CCTCAAA CATAGCA ACACTGT TAAAGAT AGAAACA CACAACC AAACCAC ACTTTCA CAAACGC TAGTCAA CCTGACT GACATAC AATTCAC CAAACAG GAAACAA ACCTCTT AGAAAAA GGACCCA AATATAA TGTCAAC ACACATA TTACACA CAGAACA ATAGAGA ACCTTGT CACAGAA AGTGAAT TCATAAT AAAACAA CAAGAAA GAAACAC AAATTTC AATGCAG GCCTCAC AAGAGAA TTAATTG	ACACGAAA ACAGCGCA TATCCACCA AAGTAAGT GACAAAAA CTTGTTATA AACACATA GACATTTTG TTATAGAA GCAGTAGC TACAGTAA AATCAAAC CAAAACAC AAAAACTA CGTCCTGTA TTTTCGTAA AAATCCCG ATATTTACA AGCTAACC ATTCACCCC ATTAAATAC ACACGTTTT GGTTTGCAG ATGACAAA CTGTCAAA AACTMTCA GTGACGTTT TGTGAGTTG AAGAACCA CAAACACA ACACATGT GGTATTAA GGTAGGCC TAGCATACC TATAAAAA TAAGAAAC GATTTATAT TAGTCTCAC AATTATCTG TAGTTTTCA				

			CAGAGGA AATTAGA TGCATAA TAAAAGA AAACAGA AATGACA TA		CAGTTAATG TACAAAAA GTAACAAT CATTCCATT GTTAATTTTG ACAGACAA TAAACATCC CACTGATG ATGCTGCA ACGCCAGC GAAACATG TATGGGAG AAACTAAA AAACATTTG TGTTTGCTC AAGGCGGA CCATACCTA TAAAGCAA ATTTATTTG TAAGCAAA CACG	MMASSLKGSTTPGGNQSTIVKDNVSEPTDSGELTPTCKR LKSKKVSYFATFNVNSLLKIGKLKHLTDTLTNQNIAITA LQETRYTDEAAFESEGFRXLKGKPGKRVMKNVPQLGT GFIINRQILGAITDFSSKSSRLSTLTFRGANKMYTMVNVH APTNETNKTDPEKVEEFWXDLEETLSKIPTKNVIILLGDF NAQLGKERKYRNIIGNYPAHNRTNXNGERLVEICKTFD LIMKSTAFKRLPRKQMTWVCQILCKTDDMGMPNPLLG EFQLDHVAIPKRAGKEIQNVKVLRGANLDSDHFLXKIK1 KIIPKVHQPRNTIRTVKYDTGKLKTSTDFKTSLSQKDYQ GWEQLKDAAIEKRRKAWADWCSHKTLENLNKFQEVR KNTAKTIRNLKRKYEQARXTQIEEDFQKNNTRDFYRTF KTNLNRYDPPTLSFKNKEGKIVYGEMENSQILAEYFEEL LNCEAPADTFSYNDTEPNPDSEPPTTEEIKNIINTLKNNK APGEDGIIAELFKYANENVIIELKQIISKIWEEEKLPEGWT SALVHPIHKKRRYIPFTKKGDKTDPGNYRGISLIPVAYK VFSKALLNRVEKQLNPQIDEYQGGFRKGRSCAEQILNLK MIMAYQKQRCKKFVLTFVDFKKAYDSVDRQTLFQILKE MKLDRKSSELVRQTLNNTTSKVKFRGKLSESFKINTGVR QGDGLSPMLFNCVLDKVIREWRQGMKGVGGIRLGCKK			
									cd016 50	556	814
RTE	RTE-37_LMi	.	Locusta migratoria								

			AGTTGTA ATCACGT TGACGCC TAGGCGT CAACCSC CAATATA AATCATT CGTTTCA GC	GGTCCAAA GATGGCCG AAACGAAA GAAGAAG	KGIRVDCLAFADDIVLLSESLEEAVNQTKELQRQAAKT GLQISIEKTKYMTNITEAPGWMAIDRGKIQKVEKFKYLG EWIDTGLTEKDAMEARINKINLAYKLTMNIYNKKNISIN AKLKHYQTVIRPEALYAAETLNLIRTGVIKRLELVERKIL RRILGPRRTEDGQFRNRSNQELYSKIEKISNIIKKRRCEFF GHICRMDQGRLTQQISQFLLYKKSKIRWIQEVQRDLEE MGIRDQEIHNRVIFK			
Daphne	Daphne-29_Li	Locusta migratoria	GAAACCG TTGCAAC AAATCGA AAATTTA GGACGAT TTGTAGC AGTTTTG TAAACAC ACATTCA TTGAAGC AACCAGT TCTGCCA GGAGCAC TTTACCG GACATGG ACTGCAC AGAAACC ACCTGGG CAAGGAT GTGTTGG CCAAAGA AATTCTC CATGTGA TAAAAGC TGCAGGA ATGAACA GTCACGT AAAAATT GCACTGC CTGCAAT ATGTTCA	TAGTTGCCC ACCCATTCT ACAATTTAG AAGAATTTC TCAACTGTA ATATAATAT TGTAATATT ACTGACAA GCTGTTATC TCTTTTAAA TTAACAATT TATGTTATC TATGTGACG TTGTCTATT GCTGTAATG GCCTAATG ACAATAAA ATTAATTAT TATTATT	MSQSPVASSVAVVGNKAAVTVPLREDSPVASSVAAVG NKAAVPVTVPPQZASTATASTAPPNEEAAEKKKRSSREE EVVSSVAGRRKMVPPLRLKDFLVEGSKMKNLEISDLKH PIAPHQNVHSKGNSSSELKFATNLLPEHQGIKDKEFFVN DFGLDNNFDIFCFSEHWATENELTLVLDTYNIANSYCRK LHLRGGVAIYTNQSLSYSFTKLDLDYLCEEQNFEAAGIFI VSLKLVIISLYRSPKGIISVFLEKLEMLLNVFRESKWLHY DIVIGGDLNADFDVTTNKGVSLNCNDKQRSVTELKNLL RQHNLHSINSRPTRDKACLDNIFVNFKTAEDSCMNCRRL MYVTVFPFSDHDAVSVNYISRFHPDKCDIPVPKVVITRPI TDDKISRFRMALADREWFDQCYGDTHCSLSSDLSAKVV FDRFFKTFLTLFNDSIPLKNAKQHSPKKCKVIDKGFKTK VTNRQNPWYTEQLTNLKHQLMVLYNISNALKTEHARS AYVKCRNEYKKAIVEAKKAHNFRSIENSTNKCQAAWK LINSVAKDPRNKSVNVSPQKFNDYFIKSVEEVGNAIIKPD IRSSELLSKNVSRSTTNINVFTFSEVSPSLILKIVKQLKSSD SADIYDISCNMLKRVIDCIVYPLTFCINKCISEGYFPDELK VSRVVPVYKKGDTESPSSYRPISIVPAFSKVLECVIYQQL SVYFENLGLISESQYGFRKNLSTVDAIDSVVKYLHQVFE DKGFAQVTFCDLSKAFDCVEHTPLLDKLEFYGIRDISLK LLKSYLQNRRQVVCIGKEMSSTEQVKVGVPQGSVLGPF LFLIMINDLPSFIKSSTVLYADDTTFLHSSNDXSSLKTCV AKTLDQASYWFKANGFLLNETKTQQMVFTLKDKLPSD DPSYVKFLGVYLDDKLSWGQHVKYISGKLSRVIYLLRR LMDYVPEKYVRITYFSFFQSIITYGIILWGNSSCVQDILIL QKKAIRVITGSPHKEHCKPLFCKLKVMTVINLYIYYVLI YTKRKLPETKLRENVHYYNTRRNKCIFTSYHRLSKSLNS YELVGINYLISCHNMYKIYPNKHLNKRCMTG	cd01650	665	908

			TTGCCAA						
			TTACAAA						
			MGAAAA						
			GAAGACG						
			GTACCAA						
			CAGCATA						
			TGAAGAA						
			GAAACAC						
			CTATGGC						
			TWCATCT						
			AATAAGT						
			AGAAAGT						
			ACATCAG						
			CACCAGC						
			ACATGAA						
			GGGGTAC				MQAIAEDKRSRWCNLVENLEEITVVQSCREPRMKQNS		
			GAAGAAG				MEIIKKLEWDPTDHKQKNNGVTANQIASQLLKNGKTGG		
			GAGCCTC				KKKGKELYPKNKKMMASSIHIHSNRTGRRHSEYETEWP		
			CTAGTAG				PQYTFTATELEDAIQSMKLNKAAGTDDLRTEQIKHFGPK		
			ACTGTGA				VRNWLLQMINACIKIPKLWRQARVVALLKPGKEIRILKV		
			ACCAATA				TDQCHFYAISTRCWSGDPKTMAAGEGGGLTKTWQGDT		
			AGCTGAA				DPKSYRPVSLLCHLYKMLERMILNRVAQIIDGKIIKEQA		
			GGACACA				GFRPGRSCCGQILNLTQHIEDGFERKLITGAAFIDLTAAY		
			GTCGGGC				DTVNHRKLLCKLYHLTGDCRLTMVIGSLLQNRRFYVSL		
			GTACCCT				NGKNSRWRLQKNGLPQGSVLSPILYNVYTNDQPIPQDT		
			GGAAACG				RLFVYADDTAVAVQGSSFDEVSQKHPVVLKNLLNTTM		
			GTGAAAA				PITSIPKTSSGLEELAQYYDANHLKPNPCKTQVCAFHLR		
			ACTTCAC				NRDAGVDLDVTWQGKKLDTVQHLNTLGYTIPWGTLDR		
			CGGCTAA			TGATCTGGA	TLSFKQHCRNTKAKVCSRNNIIRRLTNYEWGAQPSVLR		
			TCCTTCTA			CACGGAAA	VCKWKNEQTGLPITPKEEMAPGNKLPYTVWRTLNRMR		
			ACACCAG			TGAATGAA	AGVPRCKTNLRRWGLLTNDDDALCECGEVQDEAHLLT	cd01650	226	489
			ATCAAGT			TGA	CRLLPEPCSFSDLLQANDKAVAVANYWKNKT			
			TAGAGCC							
			AAAAGGA							
			ACATCAA							
			AGATGCA							
			ACGGCCT							
		Locusta migratoria	TTATCAA							
			GACCAAC							
			CCTCGTG							
Vingi	Vingi-49_LMi	.								

				ATAATCT CTATAAA TATTGAG GGATTTT CTAGTCA TAAGAGT ATTTTATT AGCAGAC ATATGCA ATAAGCA TAAATGT GATGTAT TATTGGT GCAGGAA ACACATC GGGAAAC CACCAGC CACAGAC CTAAAAT TGCAGGA ATGGAGT TAGTCAT CGAAAGA CCTCATG ACAAGTA CGGGAGT GCGATTT TCGTTAG ACCTGGG CTCAATG TAATATC AGCTGCA TTGACAG AGATAAA TAACATT GAAGTAC TTACCAT CCGGGCC CAGAAGT TCTCTGT					


GACGACT
GTGTACA
AACCACC
AGATATG
GACTTTG
TTTTCAAT
GAACCTA
ACAATTT
TACATAT
GACCATA
TTAATTTT
GTAATGG
GCGACTT
CAACTGC
CATAGTG
AAGCGTG
GGGATAC
AATGAGA
CTGATGA
AAACGGG
ATACAGC
TTGAATC
CTGGGCT
GACAGTA
ATGACCT
GCAACTT
ATTCATG
ACCCAAA
ACTCCCA
GCATCGT
TCAATAG
CGGAAGA
TGGAGGA
GGGGATA
TAATCCT
GACAACA
TCTTTGTG
AGTAGGA
AGTTGGC
AATGCAG

				TGTGTGAAGCTTGTTGATCATCCCATCCCAAGTACCCAACACAGACCCATTATTTGTGTTGTGAATGCCATCGTGAAGCCAGATGAGGTGCCCTTTCGAAGGCGGTTTAATTTCAAAAAAGCAAACTGGCAACAATTTCAAGAGCAATTGGATGAAGAGATTGAGAAAATTCCTCCTCGACCAGATGAGTATGGTAGATTTATCGATCTTTGAAAGAGGTTTCAAGGAAAAATATCCCAAGAGTTGCCACACCCGTACATCCAAGGACTGACTGGCAGCAGTA					

				AAGCCTTGCTTGACAGGTACCAGAAGTTGTTCAATCAGGACCCTTTCAGCGAAGATACGTTGGAGGCTGCGGAAGACCTA					
Nimb	Nimb-20_LMi	.	Locusta migratoria	CCCCTCAGGGGCTCAGCATTAATTTGCTGAGTACGGGCTTGGCGAACCCGGGGTCCTGAGCCGTGGGCTGTGTCGCGCCAACAGTCCGCTACACCGTAAGCTCTGTGCGTGCTTCAACGACCATCGTTCGGCGCGACGGTGGAACGTTGTACGTTTCAGAGCCCGGGGATCTTGGCTTAATCGCCCGGGTCGCGAG	TAATGGTGGGTTCGTATGCCTGATCTTCAGGCATACACCGCTTATAGTTTAATCTGTTGTCACTTTTGTCTTGTCAGTCCTATTTCCGACTGGCAGTTTTAGTATGTGCAGAATGGCTGGCTCATCCATTATTATCCTAGTGATCAGCCAGCCAACAACATCTGCTGTGATTTTWAGCATAGTTGATTATTTGTTCCACCTCTATTGTGGTTTCCTTTTTCTTTTTATGTTGTAATTTTCGCC	MALPGAKPASLSVXGPNLMLRHMTPRRSLPWPRHGRNAGREIEVRLMPPXTWYVPGLTANRLPPRSPCFLWKTLRISSGRWMACQDEIWXCXHXDVVPSPVTGTACLVRGGGWHVKMRSGXVLITTSSPAQSRALLACDKLGDVPVSITPHKSLNMVQGIIFHKDLLLQXDDELRKNLERRGVHFIRRVHRGPKDNRVATGAFILAFEGDSLPEKVKVMIYRCDVKPYVPPPMRCFKCWRFGHMSSRCXGNPTCRECGRAXHPDAPCAPPPVCVNCGQHHSPCSPXCTVLHKERKIMEYKTLDRLSYSEAKRKYXKLHPVSMKTSMPQLHHTASAISATALHCLVFHSGPPGXTAHJCPHGLWGHLNLHRPQGPCFGSTTISTVGXSVPTPEPESQQPXLAPXARKGSLGSLPSKASPADASGWLKXPPTAGRRASRASSVPGDGSAEPSQHEKPKEXREKKKKKNKSKTKDSLVATVPPXHSTSSSEDEIEILASPKDLAIAESSEAMDVEALPQKSATAEDTVXPTAPSKSACPSDQEALGSASKIGDSRRYCXPYCAVEISMSQRPGGSLILQWNCGGFFRHLPELQQLLCRFPAFCVALQETWFPAQRTPALRGYRGYYKNRTIYDRVSGGVCIYVLATVPSAAVPLHTTLEAVAVRIRAMQELTICSLYLPPDEMVSLTDLAALVAQLPPPFLLLGDFNAHHPLWGGAQTTGRGRDIETLLSQQDICLLNTGAPTHFCVARGTYSAIDLSVCTPSLLPHLSWGGSTMTFAGVTISPFFCHSSTWTSGSHPMVSIAADWRGYTSAAVVSVPLPPDIDTAVQQVTAAIVSAAEETVPITTGSPRRKVVPWWTPEIAEAIRDRRRRCNATGAAPQSKIGHPGPSAALQRHRRRPTIENLIAFKRLRARARRLIRQSKQDCWERYVASIGSRASPSRVWSQVRRLFGYRPPAGIPSLTVDGTLCTDPAVIAECMAEHYASCSATANYSASFRSLKQRVERNPLSFHTRRRELYNARLRSGSLTLLSLHAPIPLRDQTAYTASSSVTCQRTFITNCSPYLIGSGQGVFSHPSGGIVL	cd01650	1160	1418

				GACGGTA CAAACCT CTATAAA AAAACCT CAATCTC AAGGTGT GCTACGT GCCGAGG AGATGCA TGGCTGT TGAGGTT GAACAGT CGCTAGC GAGCAAC CTCTGAG GAACCTG CCGCACC TCAGTTG TATAAGG CTTACCC AGGCAAG CGGGGCT CTGCCTG GGTGGAC CTTTAGTT CCCTAGC TGCTCGT GGGACGA GAATGAA ACGAGCA AATA	CACACTGTC TTCCTGTAC ATCTTGTGT ACTGGTTGC GGGGTGCTT ATTTTTAAC CGTAAATG AAATACTGT GATCTCTTA GCTGGGAA CAGGGGAC TGATGACCC AGTAGTTTA GTCCCCTTT AAACACCC AACACCAA CCAACCA	LSLSTTGGAKSSIFPHSPPGTVQCSFTEWEFNTALAACPD TSPGPDRIHGQFVRHLSADFHHKLLALFNRVWTGGVFP SQWRDSIIIPVMKPGADPLAVDSYRPISLTNVLCKLLER MVSRRLMWILEDRSLLAPQQGGFRQGRSTIDNLVSLES AIRTAFSRCEHLVAVFFDLRRAYDTTWRHHILATLYDW GLRGPLPIFIKNFLSLRTFRVRVGTSLSSXRVLENGVPQG SVLSVPLFLIAINGLAAAVGSSVCCSLFADDFCLYYSSPSI GVAERRLQGAIRRAQSWATAHGFTFSASKTRVMHFCRR RAVHPDPELYLAADQLHVAETHRFLGVLFDNRLTWGP HLRQLKQKCWRHLNILRCLSHTAWGADRRTLMHLFKA LIQSRLDYGSQVYDSASPSALRILDPILHGGVRLATGAFR TSPVPSLLAEAGIPPLRYRRQQLLASYIAHVHATPHHPN RRLLFPNTDLPSPQRRPRSGLPVAVRVHSLLSSLESFPLR PAFSCLSHIPPWTIPRLQLHLQLSCDPKASVSPQLLRHQF LDILGTFRDADVIYTDGSQVDGRVGYAYVHGDHTEQRS LPPGCTVFTAELIAITRALDHLRACPRKSYLICSDSLSAL HALDQCFPKHHLVRSIQVTVFSLGKSGRSVAFIWTPGHI GIPGNERADQLAKQAIACQRLDIGTPEGDIRLGLHRKIFE AWNSEWATMSTPNKLRAIKPSTAAWKTSLRPSRRDSVV LCRLRIGHTWLTHGHLLRREEPPECRCGALLTVVHILLD CPLLAALRQSFNLPTALPSILADDASMADCVLGFIRAAG FYHKI			
Pene lope	Penel ope- 1_LG	.	Lottia gigantea	TTAAGTT TAACCAC CTCACCC TCATCTC CAGCCTG TAAACCA CGACGTC ATAATAA TGAGAAG	TGACCCCCC TCTCTATTT TCTAGCTCA TTAAGTATA ATATATTCT AATCCTTAT ATATAGAT GTTTATCTT TTCAGAACC	MGLPFKLNAGGNALKLGETVVNLSSRKLSPLEERVLDR GLTFIPTSKTLSHEPILSSVEHVNRTFKLNFFFRNRLGRG KRIPFIEKSGWTPPDSAIDSDILATCDKLHEEISHLPLVNE KHNLTKFERKALNSLKNDQNIIIRKSDKGSCCVIQNRED YIFELEKQLNNANHYRVLDQPIFMETAQKINNILKTLVN KQFISDSQYDYLSAKPEVRQRHIYTLPKIHKNPETEWSL ANRIPKGRPIISDCASESYRVSEYIDHFLLPLACKHPSYLK DTTDFVEKIKSLNVPPNARFISLDVTSMYTNIDNKTGLD AVRKAFLRNPCVDRPDKEILELLELCLTRNDFAFNEKIY	IPR0 00477	204	450

GATAAAT	TATWAATA	LQTSGTAMGKRFAPSYANIFMAEWELEISSISPFEPFFYG	
CCAAATG	CTCTTTTCT	RFLDDIFIIWTESLTNFWQYFEILNNHHPSVKLTANISEQS	
GTCTATT	TTTAGAACT	MDFLDTTVFKGPSFSQTNKLDIKVFFKPTDTHQLLYKNS	
CCCACAC	GGTAACAA	FHPKHTFTGIIKSQILRFFRICTQKVDFDAACRLVFSKLRE	
TCACCAA	GAAAACTTT	RGYSKRFLRKIKNDTVLDLLKINEPPPPINNFGAIPCRGS	
ACCAACA	CATATTATA	RCRTCEHIHMSHYFQNKKNGKIFEIKERLTCTDKNIIYLI	
GTTATCA	GACTTTTGT	SCKKCDKQYVGETQNALRTRFSGHRFDIIHSKNRPISHH	
TTGGATC	TTATTGTAG	FNLPDHSIEDLLIMPIESIPDSDLIKPKTKRLQRESYWMD	
TTCCAAC	ATTACTCGT	QLQTVAPNGLNIQHDNQMIPFVATFNHTAAKTSSIVKQ	
ATCAGCA	GGTTGTTTT	YYSELQNRFPNVYTKRLVIAYKRNKNLSNFLVRGRLDPI	
ACATCCC	TTTAACTTA	ESNNT	
GTTCCTA	ATACTAACT		
CATCCAG	ATGCTAACT		
ATACCCA	TCACCTTAT		
AGCCGAA	CTTGTATTG		
TCTTACC	AACATTTAT		
CTGGTGC	CTCTTTATT		
CCGAATT	TGTGACCTT		
CAGCATA	GGCCAGTG		
TCAATAC	CTTTACATA		
ATTACTC	GAAGTTTTT		
TCCAAGA	AGGAACTT		
TCTCTCC	CTTCACTCT		
AACAACC	TAACACCA		
ATTCCGA	ATAACCA		
AAAAGGT			
GTTATTTC			
ATGTTGG			
CCTAAAT			
AATAAAC			
AGGAATC			
AGGAGAA			
TCCATAC			
ACAGAAG			
AATGCTG			
GAACTCC			
TCAAACT			
TGCCAAG			
TCGAAAT			
TTCCGTCT			

				GCTACCC						
				TTTACGC						
				CACACAG						
				ATTCCTT						
				ACTCAAA						
				GGAGCTA						
				CAATTAG						
				CTGAACC						
				TCATTTA						
				GTCAATA						
				ACCTCCA						
				GAAGTTC						
				AACAGTC						
				TATTATA						
				TCATAAA						
				CCGTCTA						
				CCAATGC						
				CCATACT						
				ATTCCTC						
				GCTTCAA						
				TGGAGTG						
				GTCAGTG						
				TCATTTC						
				GGATAAC						
				ATCCACT						
				GGATGTC						
				CCCCACT						
				GCAAAAA						
				GTATTTA						
				TTTTTC	TGAGTTGG					
				TAATCCT	ATCCAAATT	MFSCRFFLIRKSDRKICLDCKVIPGESLTTQHRVMVMDV				
				GCCAAGG	TGGTGTATG	RVKRRAKRRSQNGAPRIKWWHLKGEKLRIFQHKILEGG				
				CCGGTCC	ATAGAACA	FRQPHGSANDMWERMAHEIKKVAKETLGESRGFGPRG				
				CAAGCCC	CTATGGCGT	KESWWWNDSVQSKVRIKRDCFKDWSRCKNVETWDKY				
				GGATAAA	CATTTGATC	KIARKEAKKAVSEARTQAFEGLYQSLGTKEGEKSIYKL				
				AGGAGGG	GACCCCACT	AKGRERKTRDLDQVKCIKDEEGRVLVQERDIKGRWKK				
			Medicago truncatula	TTGTGTT	TAGTGGGA	YFHNLFNEGYEILPDSNRLDIGEEDRNYNYYRRIQEHEV				
				AGGCCTC	CATGTAGCC	REALKRMSSGKAVGPDNIPIEVWKSLGDRGIVWLTNLF				
				GACAGCC	TAAGGCTTG	NEIMRTKKMSDEWRRSTLIPIYKNKGDIQNCANYRGIKL				
RTE	RTE1_MT	.		AAAGTAA		MSHTMKLWERVIERRLRKETRVTDNQFGFMPGRSTME	cd01650	317		567

				AACTTTG TCGAATC TCTATGA CATGGAC CAAATGC AAAATAG CGTGAAT GCTAGGT CGTTGCC CGGAAGC AACGCGT TGTATGG CTCGAGT ACAGTGT CAAATGA GCAAGAG CCGCTGC ATCACCA CCTGGAT GTAGTGA AAATTAA GCAAGGG TTCCCGC ATCTTCG TGAACGG GTGCGGG TAAAGAA GCTAGTT CATGAGA ATAGGAT TCGTTTTG GAACTTG AAATATA GGTACAC TTACAGG AAAATCT ATGGAAG TAGTAGA CACAATG ACTAGGA	GTTGTTGTT GT	AIYLLRRVMERYRTDKKDLHLVFIDLEKAYDRVPREIL WKALEKKGVRIAYIRAIKDMYEGASTSVRTQDGTTEDF PITIGLHQGSTLSPYLFTLVLDVLTEHIQELAPRCMLFAD DVVLVGESREEVNGRLETWRQALEAYGFRLSRSKTEY MECNFSGRRSRSTLEVKVGDHIIPQVTRFKYLGSFVQND GEIEADVSHRIQAGWLKWRRASGVLCDKKVPLKLKGK FYRTAVRPALLYGTECWAVKSQHENQVSVAEMRMLR WMSGKTRHDRIRNDTIRERVGVAPIVEKLVENRLRWFG HVERRPVDAVVRRVDQMEESQVKRGRGRPRKTIRETIR KGFRGQ			

				GGAAGAT CAATTTC ATGTGCC TACAAGA AACTAAG CGGGTGG GGAAAAA GGCGAAA GAATTAG ATAGTTC AGGATTT AAGCTTT GGTATAC TGGTGAA GTTAGAT CGAGAAA TGGGGTA GGCATCT TTGTGGA TAAGGAG TGGAAGA AAGATAT TGTAGAT GTTAAAA GGATAGG AGATCGG ATTATAG CCCTAAA GATTGTA GTGGAAC AAGACAC CTTTAAT GTAATTA GCGCTTA TGCACCT CAGGTAG GGTTAGA AGCACAC CTTAAAG TGAAGTT					

RTE	AviRTE_MUn	.	Melopsittacus undulatus	TCTCTTTC ACGGACA CCCCCTT GTCGTGG	TTGGGAG GAGCTAG AAGGCTT AATTCAG GATATCC CGTTAGG AGAAAGA TTTTTCTA GGAGGGG ATTTAAA TGGGCAT GTAGGGA GTGTTTC GAGAGGT TTCGAGG GTTTGCA TGGGGGG TATGGTC TCGGGGA GATTAAT GCGGAGG GTAAATC CATCTTA GATGTTT CATCGGC TTTTGATC TTACTAT AGCTAAT ACTTGTTT CAGGAAA AGAGAGG AGCATCT TATCACT TTCAAGA GTGGGGT CTC	TAAAACCCT TCCCAAATC TTCGTTCGC	MGIFCCGRLKTKRQRHTPASKGVSRLTLACWNIRTMLD TANSGRPERRSALIAHELSRLNIDIAALSEVRLHEEGSLK EHGAGYTLYWSGKSKTESHLSGVGFMIKNSIVSKLENLP IGHSDRIISLRLPLHNKQHVVLFSIYAPTLQADPAEKDKF	cd01650	563	847

GGAAAGA	GAAGTCAA	YSDLRRLTQKVPIDDKIIILGDFNAQVGKNFEAWKGVLG				
GCTTGCG	GCCATG	KHGVGSCNDNGRLLLEFCAEQQLTITNTIFQQKDSXKTT				
TGCCCTT		WMHPRSKHWHLIDYVLVRQRNVRDVCHTRVMPSAEC				
GTGAGGT		QTDHRLVRCKLNFHFKPKPKRGGIPRRRLNVNNLQIAA				
TGGGAGC		VRDSFQANLQTRLEDHLIDSSPEALWQHIKNSILQSSVES				
TATGCTG		FGFSLKKNKDWFDDNNQEVQELLTKKRSAHQAHLAQP				
GTGGTGG		LCHVRKAAFRLACSKLQQKLRDIQNKWWLNLAEKTQL				
CATATGC		CADLGDHRGFYEALKAVYGPTYQVQSPLLSADGQTLLT				
CTCTAGT		DKISILNHWSEHFQTLFSANRVVQGXGIQHILQHLVKNE				
AGGGTCT		LDIAPTMGETLKAIQQVKTGKAAGVDGIPPEIWKHGGQ				
CCCATGC		ALHAKFHELVVRCWEQGKLPSDFRDAIIITLYKKKGVKS				
CAGACAG		DCSSYRGITLLSIAGKILARILLNRLVPAIAEELLPESQCG				
GTCTCAA		FRXNRSTTDMVFVLRQLQEKCREQNKGLYVTFVDLTK				
CTGAAGG		AFDTVSRKGLWLILERLGCPPKFLKMIILLHEDQQGQVR				
GTCAGAC		YGDALSEPFPITXGVKQGCVLAPTLFTIFFSMMLQRAMV				
AAAGTGT		DLDEQNGIYIRYRTDGSLFNLRRLKAHTKTLNHLVRELL				
GTCCATG		FADDAALVAHTEAALQRLTSCFAEAAELFGLEVSLKKKT				
GGCAGGA		VVLYQPAPQEDYHHPHITIGESELKSVQQFSYLGSIISSD				
TGGGCTC		ATIDKEIDNRIAKAYRAFGKLHKRVWCNKHLKKSTKIS				
ACTAGCC		VYRAIVLSTLLYGSESWVIYRHHLRLLERFHQRCLRTIL				
TTTGGGC		NIHWTDYVSNVTVLEQAGVTSIEAILLKLQLRWAGHVS				
AACCATC		RMKDHRLPKIVFYGELATGCRKRSAPKKRYKDSLKQYL				
CTAGGAG		SLGHIDYHQWSTLASNQETWRHTIHNAVVSFENTRRISL				
AAGGACA		EEKRQRRKNRAPPIPPKETFRCAFCNRICLSHIGLFSHQR				
ACTCTAA		ACSKRG				
CTTCAAA						
CCCAGGC						
AGATGGA						
GCTCGTT						
TAGCCCT						
GTAAGGC						
CATCCAT						
CTAAGAG						
AAGGATA						
CTCTAAT						
CAAACCT						
ATGACCT						
GAGGACT						
TTGCTGT						
CACCGTC						

				CAAGCTCGCTAGGCCCTGGCAGATAAACCTCAGGTGTAAAGGGTGGGGCCAGTTCTGCGCATGCTGTGCCTCACCTAAAAAATCCACTGCACAGGCCCGAAGGGTTTACCCACACTTGCAAAGCCCTGTAGCGACAGGCGAGGGACGAAACGGCAGGTGATAAGGTGCACTGGAAGYCGCAGACCCAACCCTGCATGCAGGCGGTTCAGGACATTGGTCGCTCGAGACTGACCCAGGAGATGACAGTCTTGTTTGGCAACATCCTGAATGACCAAGCAGCC					

RTE	RTE-2_Ma	.	Mesobuthus martensii	GCATA	TGATTATGATGA	MNSSTNLRPSVPGGWNVPADGSKQVYRKAEGAKNHRV	cd01650	561	814
				CCTTCAA	TGATGATG	SCHGFYLATYNCRTLSTEGKLLEIEEELKHVKWDILGLS			
				GCATAA	A	EVRRKGEELLELKSGNLLYYRGTNDGRCSGSGFLINKRI			
				GCTAAAA		KDNIIEFNSSSDRVSRLVLKISKRYEFQIIQVYAPTTSHTID			
				TGTTTTG		EEVEEFYEEIEKIRGRKQQFTVIMGDFNARVGIQLQGES			
				TGGCCAG		AVGKFGEGIRNERGQRLVEFAESENLFIMNSFFKKNPKR			
				CGTGGTG		KWTWRNPSGGTAEIDYILASKKYLIEDCTVLNKFNTGS			
				ACACCAA		DHRMVRCKVHINVKNERKKMFKGPRSGIDIGKLKANQ			
				CCCGCAT		NLFQIELKNRFAELKANQNLFQIELKNRFTELTNLTEDID			
				CCTTCAA		IKNLNTNITNILIQTATEVAGKQRQVKNDKLSDKTKLLL			
				GTTTCTTT		NKRREMKINGNAVSKKIEYAELCKTVRKCITDDIRKYN			
				AACATGA		CNLVKKSIEENKGIKSVKKHLCCGRKRIFALKAKDGKII			
				GGAACGG		KNQEDILKRVEEFYSDLYRDTNNSCLVQNSKSEIPPILKE			
				ACACGGA		EVHLAIKKMKKGKAPGEDGITVDILHEGGGEQIEDILTDL			
				TGGGCTC		FNRCLKQKDVPEDWKNATVILLHKKGEREDISNYRPISL			
				ACACCAA		LSVLYKVFTRILADRLKRSLDEAQPREQAGFRAGYSTM			
				CCCGCAT		DHIFVLREIVERTNEYEIPLCLAFIDYEKAFDSVMTSAVL			
				TGGCCAG		NALANTDTDSGYVDIFSKIYSQASARIRLHKYSEKIKIGK			
				CGTGGTG		GVRQGDTISPKLFSAVLEEVIKKMDWDKTGLKIDGEYL			
				TGTTTTG		SHLRFADDVVLLASNPNELQNMIQSLATESKKVGLRINI			
				GCTAAAA		SKTKVMYNNFADQQKIKLDNIVLEEVSGYIYLGQLLTL			
				CCTTCAA		DGDLMPEINRRIKLAWSAFGRNNSILKGKMPLCLKRKIF			
				GCATA		DQCILPVLTYGSETWTLTKRAIQRLQTTQRSMERCMLGI			
				GCCTTAG		TRRDRKRNTWVRSQTKVIDVIQKIKKQKWTWAGHIIRI			
				AGGAACC					
				GACTAAG					
				ATCAAAA					
				CCTAGTG					
				TCCAATC					
				CGTACCT					
				TCACGTG					
				GTGCAGC					
				TGTTAGT					
				ACTCAAT					
				GTTTCTTT					
				AACATGA					
				GGAACGG					
				ACACGGA					
				TGGGCTC					
				ACACCAA					
				CCCGCAT					
				TGGCCAG					
				CGTGGTG					
				TGTTTTG					
				GCTAAAA					
				CCTTCAA					
				GCATA					
				TTTTTTCA					
				AGGAGAG					
				CACTGCT					
				TGCTCCA					
				TATGGAG					
				AGAGGCC					
				TAGAAAA					
				GGTGGCC					
				TAACCAA					
				AGCTCAT					
				CTCCCCC					
				TCCACCT					
				GGTTGGC					
				TAGCCGC					
				GGCCA					

						NDNRWTTKVLHWIPRDKKRPRKRPHYRWRDEIVNFTS KCWIREAVDRKNWKLLGETFILQWIDNG				
						MMSGTHDSDNARAFPGVGDAQSLRRPRAVRSRLGLSH QERVRLKKLARRPLLVRIASLNVGTMTGRGREIVDMME RRKIDVMCVQETRWKGDKTRELGEGYKIFYSSASKEGR NGVGIITSRKMKENIIEVSRRNDRIMRMKLDIGETIHVVS AYAPQVGTSMEMKDTFWKTMDEELRQLEEKDRVIIGG DLNGHIGTNKRGIERIHGGHGFGDRNQEGEQIIDFAVAF DFALANSFFNKKEEHLVTFKSGGNSTQIDYFLYRRKHLK EVKNCKVIPSDHVATQHRPMLMDLHLLRPAGREKVKT RVKVRKTKWFKLKDTVLRSRFKETALQRVSMDFEDVQ EWWNRNAEVIRCTGREILGETTGKCWENKESWWWNQ EVQAAVQEKRKAKKKWDSTHAPADLEVFKVRNKQAK KKVAQAKAKACQELYDDLETPEGQKKIYSIAKHRSQTA KDITHIKQIKNVNGEILRNTGKIQDRWKEYFEELLNVEN ERKYHEDGIPNEGLTRDITKSEVKVALKQMKNGKAVGL DGIPAEAWKALGEEGLDLLWDLLGKIYREEEIPDEWRE SIIVPIYKGKGDIQMCGNYRGIKLMSHTMKILERILDKRL RAEVKIGGQQFGFMPGKSTIDAIFCLRQMMEKFREKQQ ELHLVFIDMEKAYDRVPRQEVWRSLRLRNAPEKYVRLI QEMYRKTTTKVRTCLGETSSFEVGVGLHQGSALSPYLF NTVIDVLSEKVREEAPWNMLYADDVVLCARSRAEVEE KLENWRKALEDRGLKISRSKTEYLACNIGTQENEELKL METTVRKTDHFRYLGSIVDELSETGREVNNRIQAGWKN WKEVSGVLCDKKIPVRLKGKVYRTVVRPAMLYGTEAL PLKKSNLQRMKVAAEMRMLRWMCGVTRADRIRNEHIQG TVKVTDIASKIQESRLRWFGHLMRRDEDYTGKRLMNIK VEGRRKRGRPKRRWQDIVDDDLRSKGLRREWTKDRTQ WRKLIKNSDPI	cd016 50	567	814
					AACTCCG CCCAAGA AGCCATA TAGCTGA CTGGCCA GTCCCAA GCCTGGA TAAAGGA GGAGGGT TGGCGTC AGGAAGG GCATCCG	TAGAAAATG GGAAGGAG CTTGATGAT GATGATGA	MKNFTFDVLGVSEVRWTGKGELSTGDFIWSGEEKHHFK GVGFLLSKKAQKCLISYNPISSRIISARFSATPFNISIIHVY APTTDASDEDIESFYGEIKNTLRETPKKDFTIITTGDWNAK IGSNNTGWESAMGKEGYGDRNERGERLLEFATEHNFYI CNSRFRQKPCRKWTWTSPDGIHHNMIDLILVQSKWKTS VQNCRTFQSADIQLDHSLVLCNIKIRLKKQSTTSQGTLD FRKRNSNDLKQPPIKQAYEKSLNEQLKKITPTMSLDEHA SILESAIKTAIKSSLAVTEKPRKPWITENTLKVANEKRKW KLKRHISEENNNIYKAHCIQVKKLAKEDKENWLNEQCE AAENGLKTGKTHQAYKLIKMLRKPFTPKLNILKDDAGK ILQTKTAITNQWTKYCSTLYNGSNNNQSQATDAELEEIT	cd016 50	498	754

						PPADEYPCPHILHSEVEAAIRTLKQRKSPGFDEITAEMIKAGGPKLTASLHELCQKIWLEEKIPTNWGKTILIPIPKKGDLARCENYRTISLISHLEKVFLIILLRRLQHQLEPYLAEEQAGFRKDRSTVQQILLVRLLAEKAKRKGKKIYNCFIDFQKAFDSINQQRIWAVLASFGVDRKVVKLLQNIYENTRLAVKTGNDMGTWFKMLIGSRQGDPLSPLLFITYLERTMDSVTKLNLGVTIQGSIISNLRFADDINLMNEDLQALQDQATALQASAEKSKLRINTSKTKTMVFGARNIEQQIEINNQPVENTEQFEYLGSIITWDNNCSVEIKNRIAKATGALASLKHIWKSSKITVPTKLRILSVCVFSVLLYAAETWTLKENDKKKLLAFEMRCYRRLLKIKWQDKIRNEEIRNTLNIHTTIIDLIRQRKLRLFGHICRMDNNRJIKHVVTSXMNGKNKQGRPPRAWMDDITDWCKCSAQEAFHRAQNRNQWKTATRTND			
RTE	RTE-1_MMa	.	Mesobuthus martensii	TTCTTCGAGAAGAAAGAACAGTGAAAATGGACATGGATCAGCCCCAACGGACGGACAAANA	TGAAATAAGAAATCTGATGATGATGATGA	MRLTVXLANSNCTIKDVSVLNQFNTGSDHRLVHAKVKXDIQCERXRLXQHGGQKVKKEEYSLKLQTDSKCSQKTKVXYGIHTMRQCRRNRGNSTKKSRQENIHQNTGADRKEKEDEDQIHYQPDRIEDEVRKSPRRRQFNIKLLKEGGVPLYKQLAGIFSDCMRTGQIPDSWKNAVIVPIHKKGDKTDLRNYRPISLLSTIYKVFTIILTRRLEQVLDSNQPREQAGFRKGYSTMDHIHTINQLKEKCKEYHISFCMAFIDYEKAFDSVSTEAILTALSNQGIETTSTQNXCCKDSGIAPRPPSVRRNLLRIIYEEATAIIRLHKDSRKIKIQRGVRQGDSISPKLFTACLEEIFRQINGWEHKGININGEYLNHLRFADDIVLFAQTPNELQEMLSDLKTESEKIGLQINYEKTKVMFNAEVIPQPVTIEGRTIERVEEYKYLGQVITTQGDQMREIQHRTQLGWHACGRYSNFLRDNRIPICLKRKLFNMCILPVMTYGCETWTLTKKMEQKLATTQRSMERYMLGFTRRDRKTNIWIREQTKVEDILVTVRSSKWHWAGHIARYTDNRRTKRITEWQPRQGKRYPGRQYKRWRDDIRKVIGSQWMQIAQXRYIWKQKEAYISSSG	cd01650	175	448
RTE	RTE-4_MMa	.	Mesobuthus martensii	TGGGGAAAGCAATGGTGAACTACCCTGTAAAAACCTCACCTCGAAATTCTATGGATCAAGACCAAAGTGT	TAGGATTGGATTGGAT	MIRSSQAENDARGANPPGGPIADVATGKKDHLSCKDLFTIGTWNVRGLNIGKLEIIKREMTRTGLDLLGISELHWTGSGDLKSDDYTVHFSGNDHNRRRGVAFVANKRLAKYIENARCISDRIMTIRLRGKPLGVTIIQVYAPTADADKQDIVDFYTAIQNALDQTPKKDIIYIMGDFNAKVGDGEDGQVVGRHGLGDRNDAGDRLVQFCQENKLRITNTFFKQHKRRLYTWTSPNGQHRNQIDYILCQQRWRGSITAAKTLPGADCGSDHQLLIARIKIRLCKVKKTVVAKKFDVSRIPTQYTIEVRNRFEALSLDGGQTEELWAMIKGTIIETALEYVPYKMPTKGAVWLTEKTLQIAEKRREAKAKGEIDKARRLHADFQRE	cd01650	537	789

			TATCGAT			ARIDKNNALNARCISLEEAFRKGRTRELFATVKQVKTPF				
			ATGGTGT			SARQSTIKDRDGNELTDQQDIKERWREYTEDLYTSIDDQ				
			TGGACGG			QNEDQTDQGLQEPDLLESEVEWAMRQLPSNKAPGIDGI				
			TGGGCCC			PAELLRPIPRAALTALCQQIWKSCTWPADWRRSVFVPIP				
			CCGGGCT			KKGDTRDCANYRTIALIPHASKILLKVIQKRLGNVIDREL				
			GGACAAC			PDVQAGFRRGRGTRDHIANLRWIMEKAREYQKKLYMC				
			ATCGACC			FIDYTKAFDCVEHSKLWVALRELGVPEHLIRLIRSLYQN				
			ATGCTAC			QEATVRTRYGDTDWFKIKKGTRQGCTLSPFLFNLYAEVI				
			TGGGGAA			MRKLDLEESEVGVKIGGRTISNLRFADDTTLLAESEKDL				
			GAGCAAA			KRMIMKLKNESEKMGLQLNIKKTKIMTTAAKEVHIKIN				
			GG			NEEIEVVDSFIFLGSKVDCNGDWRPEITRRIALGRTAMV				
						NMDPIWRCKDVSITTKRRLVTAIVFPIATYGCETWILTK				
						ADRRKIDAFELWCWRRLLRIPWTARVTNKSVLERINPD				
						FSLEGKITRQKLTFFGHVMRANSLEKDVMLGRMSGTRP				
						RGRPKTRWLDTIKKDTGKNIEELRELVRDRKTWRDMT				
						HRIAKNRTRLNG				
				TAGCCATCA						
				TTGTGGTGG						
				GAAAATTTC						
				CCAGGACT						
				GTATCTGAG						
				GTCCTTCAG						
				ACAGCTCTC						
				AATACAGT		YDLALIQEPWHYKGQIRGLNCKEGKLFTFTSDSGPRTCV				
				CTCTAAGTG		FVRNNINVSPVWDFCSRDITTVRVKLDEGGVSRELIVSS				
				GTGTGATA		AYFPYDKDELPPTPEFXLLVEFCQKNHFDLLVGVDANS				
				GGGTTGGA		HHVVWGSKDINRRGEYLLEFLMGTNLFVLNKGNEPTFV				
				TTGTCTGTT		VVNRREVIDLTLGTQNISDLIENWHVSGEPSSSDHRYICF				
				AATCCAGCT		NLQGMVPSQVTYRDPRRTNWASFREDLKFNLGGIQSRF				
				AAGACTGT		CSVVDLEMTVDQIQQAILLSYHQNCPVRTARSSRQVPW				
				CATTGTACC		WNEQLGGLRXNVRRLFNKAKRTGDWDTYRSALTCYN				
				TTTTACACG		KAIRKAKRSSWKRYCQGIEDIPDVSRLVKIMSSGNFCKV				
				AAGAAGAC		GSIRLPDGSMTSTCKDSLKELFRVHFPGSTLIEGSSCLQV				
				AATTGGCG		QSDLDGHVYHAKREDWGLVRGVINQSNIRWAVSVFKP				
				GGGCTAATT		FKSAGPDGILPVFLQQGIDILASPLCRVFRACLAFGYIPW				
				GAACCATC		SWRQVRVSFIPKPGHSSYTEAKSYRPISLTSFLLKTMERL				
				ATTCTTCGG		VDRFVRDGALLKTPLCRSQFAYQSGKSTENALHEVVSLI				
	Loa-2_Ma	.	Mesobuthus martensii	CAAAAGGT	T	ERAIGYKEIALGAFLDIQGAFDSTSFEQISLAASRHGVCS	cd01650	468		
Loa				TGCAGTTTT		TIIRWIGSMLKHRVVYSTLYDEEVKILTDRGCPQGGVLS				
				CTACCCAG		PLLWCLVVNELLTLHSMGKVSSLLGMLMT				649

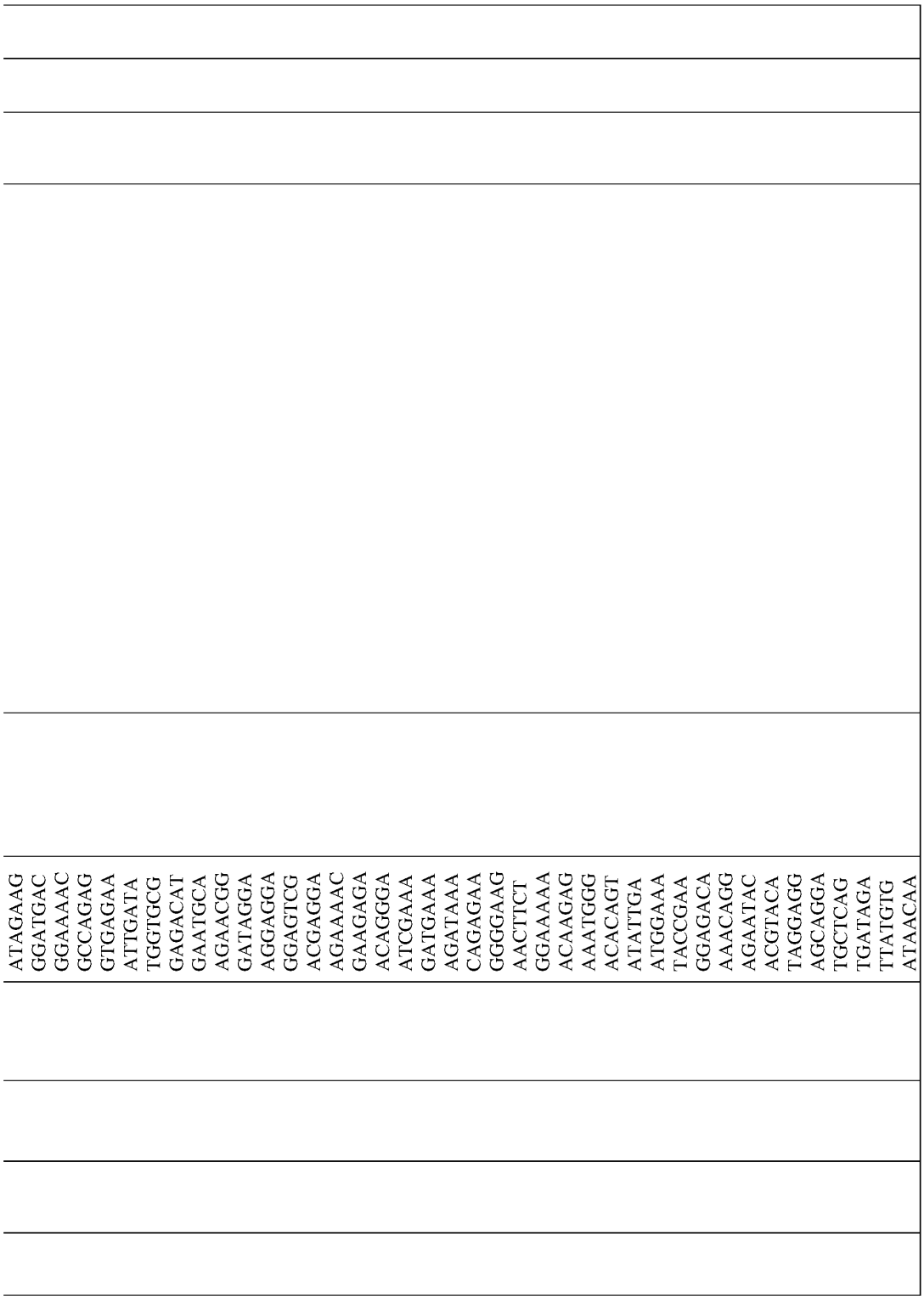
				GTCAAATA CCTTGGTG			
			AGAAAAG	TAAAAGAA			
			TAAGAAG	AATATAAT			
			GAAAGAA	ACAAATAG			
			CTGGAGA	AGAAAAGA			
			AGAGGCA	GCAAACGA			
			GAAGAAA	AAGAACGW			
			GAAACGC	AAAGAGCA			
			ATTTGCA	ATCGAAAG			
			AAAGGAG	GAAACNAG			
			GGAATGT	AGCAATAG			
			GAAAAGA	AGAAGAGG			
			TCACCAA	AAAGGAAA			
			TTGAGAA	AAACAAAY			
			AGAGAAG	AAGGAAAG			
			ATGGAAG	AWAGAAAG			
			AGGAGGC	AACCAGAC			
			AATCGGG	GAAGGAAC			
			TCGTTAA	KAAATAGA			
			AGTGCAC	GCGCAGGG			
			TATGAAT	ACATTTAGA			
			ATAATCC	GAGCAGGG			
			TACAAAA	GAGACGAC			
			AGTGGAG	GGAACAGG			
			AAGATGA	AGGAACAG			
			GTATGAA	TGGATGAA			
			AATGGAT	ACAAAAGT	MRAAFDKVNRREMWEMMEKLGLGVRIRNRVKELYRG		
			AAGGAAT	CACTAGGW	TRSEIRIGEDKIGGFELHKGVRQGCPLSPTLFNVAMADT		
			AGGTAGA	AWGGACAA	EKELSKVQEGGIVLGKKKFWSISYADDVVMVANNATG		
			AAGGGAT	GAAAACGA	LKQMLKKFKKIIERKGLELNTEKSKIMVFKNGGRRKKE		
			AAGATGA	TACAGGTG	DSFFWEKKEIEVVKNFEYLGYMLKENGKEEAQIKKLKE		
			TCGAGAA	GARAGAGG	KASTVMSSMWGLGEELFRDNWELRMRLFDTMVKSVM		
			CTGGGAA	AACGTGGA	EYGAEVWGWKGKEELEKIQRKYMKWVMKLDRTTPEH		
			CAGAAAT	CGAGACAG	VLHLETKRYKLETRTRRAIKYEVKLSKAKEGTLWREC		
			ATATTGG	AAAAGAAA	WRTLEKEEEKERRRGXRRXNREEELNKRGWALSEYHR		
			ACTGGAG	ACACGTGG	RLRAGEQVWSELERIEKDIQWQKWESEMRDSRQARDV		
		Nasonia vitripennis	GAAAAAA	TGAGCCAC	KELIKPQGETPEYLRRTGNTGKKGLAMTARFRMGNEAR		
	NAVIRTE1	.	TGAACAG	AGATAGAA	GYKYWMKEEERICRLCREEEETFEYIFSRCNRTGSRENN		
RTE			AAAAATA	AAGAGGTT	WKNILNGERKNLAKLYRVIWVRKEKEKEEEX	NA	NA

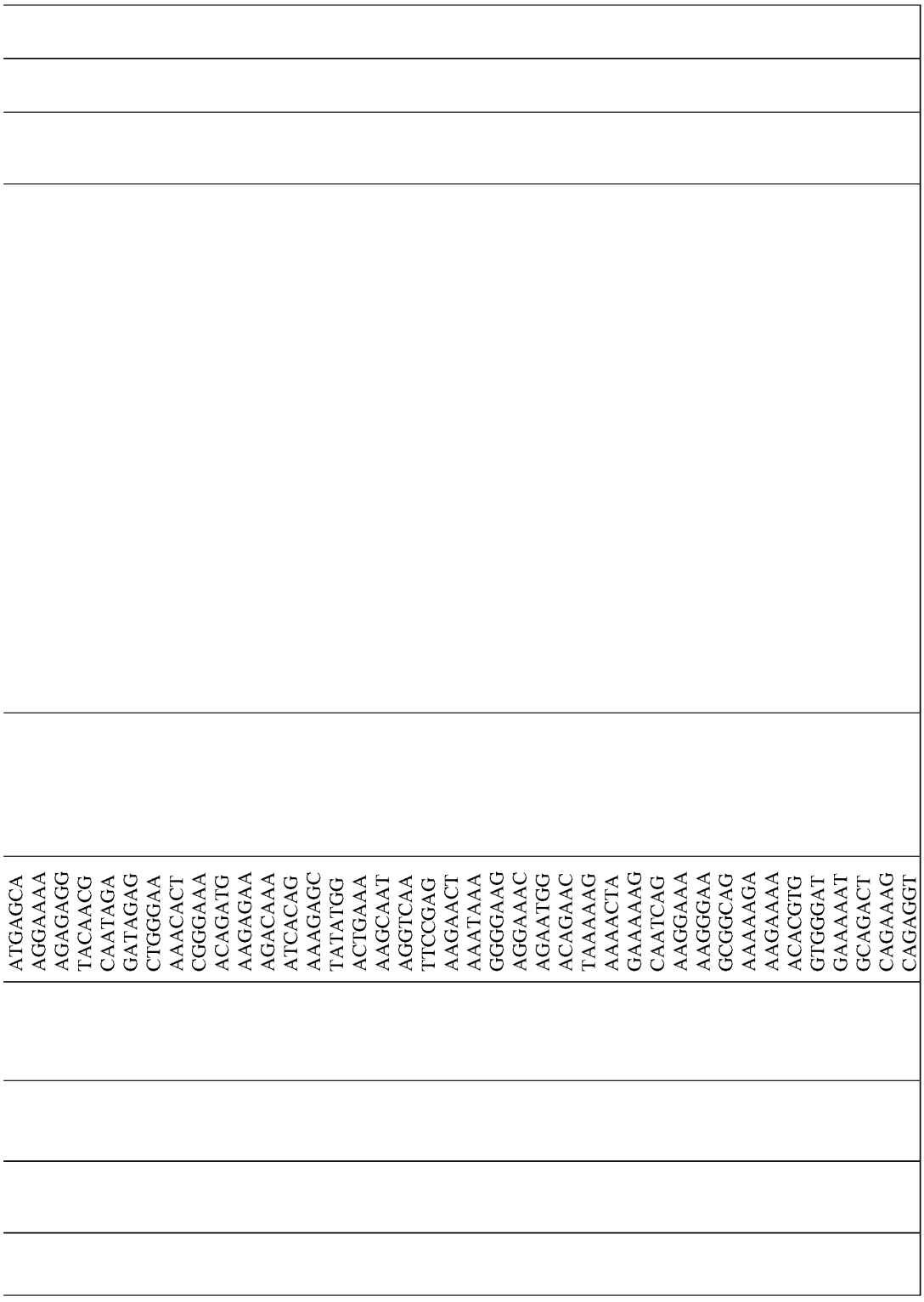
				GAGGAAG	AGTAAAGA				
				GGCTAAA	TAAAACAA				
				AGGTATG	GTAGTGAA				
				GAAGGAA	AGGAGAAG				
				GACTGTG	ATAAAAAA				
				AGAAGAA	GTGAAAAG				
				GGAGAAT	GAGAAGAT				
				AAAGGAA	GTACGGAG				
				AAGAATT	ACAAGAAG				
				AGAAGAG	AAAGAAAG				
				AATGCAA	AGAGATGA				
				GGAGGCG	GGGAACTG				
				AGAGAAA	AGAGAGGA				
				AGGAGGA	GGTRGAAG				
				AAGAGAG	GATGGAAG				
				ATGGAAA	ATAAAGGC				
				GAAAGAA	GAGGTTTG				
				TGGGAAT	AGAGAGAA				
				GAGAATG	AAAAGAGA				
				CGAAAGA	AGCCCTGG				
				AAAGAAG	AAAAGAAA				
				AAAAAAA	ATGCAGGA				
				GAAATTG	TGTGAGAG				
				AGGAGAT	AGGACGTG				
				ATTAGGG	GAGTTTTTA				
				AAGATAG	AAGGGGAA				
				AAAGTCT	GATGAAAA				
				GGAAAAA	ACATGCAG				
				GGGGAAG	AAAGAGAT				
				AGATAGA	AGATGAGA				
				CAAGGGT	TGAGAGAA				
				AGAGAGA	GGGCTTGC				
				GTCAACC	AGGTAAGA				
				AGGGGAG	AAAATAGA				
				AGAGAGG	GTGGAGAC				
				ATAAGGG	GGTAAAGA				
				AAAGGTA	AAAGGCAG				
				ACGGAAA	GAGATGAG				
				AGGAGAG	AAGAGGGA				
				GGACAGA	GAAGAGTA				

				AAGCAGA TAGAAGA AATGGAA TGGAGGA TCGAGGA AGGAGAG AGGATAG GAAAAAG AATAATA TTATAAT AGCAGGG TGGAAAG AAGAAAA ATGGGAT AGAAAGA CGGTAGA AGACTGG ATGAAAA AACTCGA AGTATAG GTTACGT TGAGGAA AACGGTC ATTGAAC GGGAAGA TCTAAAG AATAGGG GCGGAAT GTAGAAA CTGGGAA GAAAAAG GAGATAT CATGAGA GAGAAGG CTAGGTT ACAAGGG TCTGATA TATTTAT AGACAAC GATCTAA	ATAGGAGA AGAGAGAA TTTTGAGAG AGAGAAAG GCTGATGC GGCCAGTA CAGATATA GGAAAGGG GACAGCAG GTGCAAGG TGTGGGAA AGAGGCCT GATAGTAC ACCTAAAC AATTGGAC GGAGGAGG ATTCTATAA AAGACTGG CTGGAGGA TATAATGG AAGGAGTG GAATGGGA GATGGAAG ATACAAGG ACGGAGGA CACGAAAA AGGTAATTT TTAAGAGA AARGAAGA CATGGAATT GATATGGG AGAAGAGG GCGGAAAT AAGAGAAG GAGGCAAA ATCAGGCT GGAGCAGT GGCTATCM TTCGATGAG AGAAGGGC				

				CATGGAA	CAAAGCGC				
				AGAGAGA	TGATTATGA				
				AGAAATA	GAGAGAAA				
				AAGAAAA	AGGAGAAG				
				ACTGAGA	AGAAGAGG				
				GAATACG	ATGAGATC				
				CGACAAA	GAGGAGGG				
				GGAAAGA	CGAGGTCTC				
				AAAGAAG	GCACAAAA				
				GAAAAGA	AAGAATGG				
				GGTCAAG	GTAGCTGA				
				ATAGGAT	AATAGAAG				
				ACAACAA	GCGAACTG				
				ACTATGG	TACAAGTG				
				ATAAAAG	GGACGAAA				
				GCGTAGA	AGAAGGAG				
				GTATTTA	AGAATTGT				
				TGGAAGG	AAAAAAAG				
				AAAGAGA	GAAAAGGG				
				ACAGGAA	ATAGAACK				
				TTTTTTCG	GGGAGAAG				
				GAGAAAG	AGGRCAGA				
				AAAAGGA	TATAAGAA				
				CAGGAGA	ATGGAGGT				
				GGAAGAA	CGGGGGAG				
				AAGAGGG	TCAGAGGA				
				GAGCAAA	GGAAGAGG				
				GGAACTT	AAGAWTAT				
				AAGGATA	CAGACAAA				
				ATTAATT	GATGGAGA				
				GGAATGT	GAAAGGAA				
				AGCGGGA	GAGTAGAG				
				ATAAAGT	TTTACTCAC				
				GAGTAGA	AAACACAC				
				TGAGAGG	ATAAACAC				
				ACCTGGG	ACACAAAC				
				GCTATCT	AATCAAAC				
				AAAAGGG	ACACACAC				
				TTTGATG	ACACGTAC				
				TAATTTG	GACTTATAC				

				TCTGCAG GAAACAT GGCTGGA AGATAAA GAAGGAA ATTTCTG GGAAAAG AAATTAG GAGGTTA TGAAGTC AGAATTA GAAATGC AAAGAAA GGGGGAG AAAGAGG GAGAGTG AAAGGGG GAATAGT GATGGCG GTAAGAC AGGGAAC GAAAACA GAGAGAA TAGAGTG GATAGAA GAAAAGA CAAACGA ATTCATA GGAGCAA GAATATT AGGAAGG AGAGATA TATGGTG GATAGGT ACGGCAT ACATGAG AGAGGAA AAACAAG AAAACTT TAAGGAG	ACACGCAC ACACACAC ACATGCAC AAACRGAC ACGAAAAC ARGAAAAG MAGGAGAR CRCACAYA CACAYACA CACATRTAC ACASGAAC ATTACCTAC AGAGACAC ACATAACA CACCAGATT AGGTTAAA GTAGGATT AAAGCAAG AGGTATGT AAATATAC GAATGAAT AAATATTGT AAACACGA ATAGTGAA ATACAATA AAGTTATTA TTATTATTA				


ATAGAAG
GGATGAC
GGAAAAC
GCCAGAG
GTGAGAA
ATTGATA
TGGTGCG
GAGACAT
GAATGCA
AGAACGG
GATAGGA
AGGAGGA
GGAGTCG
ACGAGGA
AGAAAAC
GAAGAGA
ACAGGGA
ATCGAAA
GATGAAA
AGATAAA
CAGAGAA
GGGGAAG
AACTTCT
GGAAAAA
ACAAGAG
AAATGGG
ACACAGT
ATATTGA
ATGGAAA
TACCGAA
GGAGACA
AAACAGG
AGAATAC
ACGTACA
TAGGAGG
AGCAGGA
TGCTCAG
TGATAGA
TTATGTG
ATAACAA


ATGAGCA
AGGAAAA
AGAGAGG
TACAACG
CAATAGA
GATAGAG
CTGGGAA
AAACACT
CGGGAAA
ACAGATG
AAGAGAA
AGACAAA
ATCACAG
AAAGAGC
TATATGG
ACTGAAA
AAGCAAT
AGGTCAA
TTCCGAG
AAGAACT
AAATAAA
GGGGAAG
AGGAAAC
AGAATGG
ACAGAAC
TAAAAAG
AAAACTA
GAAAAAG
CAATCAG
AAGGAAA
AAGGGAA
GCGGCAG
AAAAAGA
AAGAAAA
ACACGTG
GTGGGAT
GAAAAAT
GCAGACT
CAGAAAG
CAGAGGT

				AAGGAAG TTAGGCA AGGCGAT TAAGAAG GGAGCGG GAAGGAA AGAATAC GTCAAAG CAAAGAG AGACTGG GGAATAT TAGTAGA AAGAAAA AAGCAGG AAGAAAT GGAAAAA AGAATAA AAGAGGC TGGAGAT AGTAAAA CCGGGAG AAAATTC TGGGAGG TGGTAAA AAGTAGG AGAAGAA AAAAGAG GACAAGA TGCAGCA GTAAAAT AAAGAAA GAGGAAT GGTTAAC TYACTTC AAAGGGC AACTAGG AGAAGAG ACAGAGG AAGATTT GGGAACA					

AGCAYAC
AAGGAGA
GGATGAA
GAAGGAG
AKAGAGA
GGATAAA
GAAGRAA
GAATAAC
CGAAAAC
GAGGTGA
GGAAGGC
AATTAGA
AAAATGA
AGAAGGG
AAAAGCG
CCAGGAG
CAGACGG
ACTACAA
AATGAGG
TATGGAT
ACACGGG
GAGGAAC
AGGTAAT
AGGAGAA
ATAACTA
AAATACT
GAATAAC
ATATGGA
AAGGGGG
AAAAGTA
CCAGAGG
AATGGAA
AACGGGA
CTAATAA
CCCCGCT
TTTCAAG
AAAGGGA
AGAAGGA
GGAAGCA
AAGAACT

				ACAGAGGGATAACTCTCATGGACACAGGGTATAAATTATACGCAGAAATAATAAGGGAAAAGATGGAAAGGGAGTTAGAGGGAAGAGAGATGCTTGACGATACACAAATGGGTTTCAGAAAAGGGAGAAGGTACGGCAGACGCAATTTATATTTTGAGTAAAGCAATAGAAACTGAACTGGATAAGAAGGGAGGTAAAGTCTACGCGTTTTTCGCGGAT					
CR1	L2-3_NVi	.	Nasonia vitripennis	TGGTGAGTACAGTTGACAGCTCTGCTGATCCATCATTTTTCTGTACTGCTTTTCTGCTGCTTCTT	CTCGTACACACTCCACGCACACGCACACACCTACTTTCTCGCTCATTCCACCTTCACTATCGCCAYACTYTCACACTATA	MEALKTQNIVSERISKIEQRLGPLEKRLKALDELPALKTRIHNAESTITELQAQIQDLSSRSPTMQQDNGSTVPNTAEICSLRSELAEVKRRQEQTSNCVVVVTGLHYTRETSLHLLAFSVVNALDPTVLRRDVASVRTMGRLDATNSSARGDGRLPPLAVTLSSSALARSIVIAKARKRKLHTSELDATLLEEAKALSPDHQGLININELLPSDVHKLRTRARLEAKKRQGCRTFVREGRLYTCAATMTASVLRSSTPTPSWRLFPGFRPLPTSSNTEATXQYTHQTLSSSAQERSKVKSEYLFCEVSAKGVSPIFVGVVYRPPHAPFFQGSNFIDQLTTHMHNYSTKVI	SSF56672	671	971

				GCCTTCGGGCCCTCTTCTGTTGTAGATAGTTGCACGCTCCTTTTTGCAACCTTCACGTCGATACATAAATCGCCGAAGGCTGCAATTTTTCTCTGTCATCTTTTCTACTGCAAACTTTCTACACGGCTCCTAACCTACAAATCTGCTTGGCTTGGCTTGGCTTGTTTCCTCACTTGCACGGCGTCTCTTTCTCTCCGCAATCAACATCGTGGCGACATCTGCTGCATCGTGGGAGAACTTTCAAGGCGACTTCACACTGATCGAGGTCGGTACTTTCACGCTCTTT	CATACTCTAAACATACCTCAATCTATTGTTATTTCTTTTCTTACTTTGTAATGCTGTAAACTTACAATCGTACAATATAATGTACGCAAAATAAAACTAATCT	MGDFNSDQLSSSEDAKFIKAFIDENSLSSVPYGATHHKQGSDTWLDLCLIDEQDRLLSYWKTNSPFINGHDLITATLDVQIPRYVPNTYSYRNFKGISAEKLRDFLSACDWSSLTSSSLDECISTLNANLTNAINHLAPLRTVTPRRQRHPWFTTVSHVXXXXRILVALRDLVSERDRLYRRFRDSRLDSDLRIYRLARDNAHKQVEEARLNYYYSRLSTLTDVAEIWRELEKLGISATKAPSPSRFTTDELNKHFSSISNDPLAPAVEDYLLTLESLDLPEHFEFSTITESDVLAAVSHFDTQARGSDGIPQVVISKALPVLAPLLSHIFNLSLSEPCFPSAWKLSLVRALNKVSSPTALTDYRPISLLCFLSKALEWLVHRQVSEYLESRLLLDNFQTGFRTGHSTHVRVGINKKKVTLLLLFDFSKAFDTVCHVRLLRKLSTFGFSKQVIRWFASYLTGREQAVVGDNSERFSPRRPAGSVWSLLLHCTSTTSVSALIPMFPILSMRMTCKFTVNATLRSSILYQTRRAFLSSASRRVRLVLAVALYINDIGFCLDSDVSHLIYADDLQIYSQCHLEELDSLSNKMSANAERIMGWAAQNRLKLNVNKTKAIVLGSPYYINALPKVTLINMGVPGQLHLLIWGCQVSFESSVRNLGLVLDSKLTWKEHVTQVCKRAHSLMYRLYFFRKSTNLRLRKHLVQALLFPIIDYCSLVYCDLTQELDLKLQRLVNTGIRYIYGVRRDEXISPYRRELQWLTTAGRRKYFTACFLRKMFNSVVPSYVLAFFDFRVTLRPVRGEVTPLEIPAFATETLRNSFHISASYLWNNLPSHIRNTSSTVTFRKLAKDHFFQLENT			

				CCACTAA CGCTCCA TCTAGCC GTAAGAG TCACAAC ACCTACT CTTTCACT CCGTCGA TAACTTA TCATCCT CCTACAA ATATCTG CTTTACC GACTGCT AGTAAGT TATCGAC GAAAAAT GAGTGCA AACGGCG AGAAGAG TGTGGAA GTGCCTT TTCAGCC ATGCAGC AGCTGTA AACAACC TATACAG CACAAAG ACCCCAA GAAGTGT GAATTTT GTGGTCT GCACTAC CACGTCA AGTGCCA AAACACC GTTAACA TCAAAAC TATCGTG GTGAACG					

				ACAGGCA GATTATA GCCTGTC CCTCATG TGCAGAC AACAGCA ACGCCAA GGGTGCC AAACTGA GGACTAA ATCCACC TCTGCCA CAACATC AGCAACG GGAGTAA CTGCAAC AACAGCA GCAGGCA AGAATAC TCCCAAA CAACAGC AGCCGGT GAATAAG AAGACGA GGTCAGC ACCACCA TCAGCCA ACACGTC GAGAAAT CCCTCAC CAACCCG CTCGGCC GGAGTAA CAACACC TACGCCG TCCGTTG AGGCCGC TTTGAGG GACATTC GCGGAGC						

			ACTGCTC TCGACGA CATACGC AATGCTC AA					
						MPNTRTLASRGEIRGLGGAGRIFRDPTCSVPRACILVKG		
						FDALLLPARCSRDLTAIKIRLPVGEGSEREVVVASGYFP		
						YDSQEEPPPREVQDLVEYCRQRSIPLILGCDANAHHIVW		
						GSSDTNGRGDALLQYLVTTSLCIMNRGREPTFYNSVRSE		
						VIDLTLCTVGMEGWVGSWRVSNEPSLSDHRYIQFEWKE		
						RCSETRAFRNPRKTDWAFFRENLRNELQSFKPSFGTTDE		
						LDHWAFELGEIVNISFQRSCPWTIPRGTWGTPWWNREL		
						EDLRRETRRTLNRAKNTRNSIDWRIHREAQRLYKNRIN		
						AVRIKGWRDYCEDIERYPDAARLLRILAKNPEVWLEAIR		
						LPTGEYATSEEECLKLLLEANFPGFRLSHEMGDESSGRN		
						RQQRAAWDLAAKVVTPEKVKWAIRNFQPFKAPGIDGIY		
						PAFLQEGLEELVGPLVKLFRASVALAHVPEIWKTAKVV		
						FIPKTGKPSHTTVKDYRPISLNSFVFKTLERLVDRFIQDDI		
						LTVFPLHPNQHAYRAGFSTETALHSAVVWRIEEQLERGE		
						VMVGVFLDIEGAFNCTSIEAVVKEAGLHGMPGPLIKWL		
						QGMLTRRTVISSLGTVTVSGEVSKGCAQGGVVSPTIWC		
						LVANGPLEALNGWGCYAQAYADDFLILIKGSEIGAAMD		
						AMQLALRKVERWCCSTGLSVNPDKVEMVIFSRKYKLD		
						AYRAPKLSGVALQVKDSAKYLGIVLDKKSTWEKHLTA		
						QCDKFLVALWTCRRAFGSTWGLSPRIILWLYRAVLVPR		
			ACACAGA	GGGATGTC		LAYAALVWWPRAELAGARAAPEKLRGQVLRGATGAY		
			TAAACTT	ATGACGAA		RTTPTKALGILVKVEPLHLTIIGMAAKAAHRLNAYGQW		
			GCATCAT	GGGTGGTA		TQGTRHTRLPGGVGLLPEFSIKTDMMIPRYFFGRRYGVV		
			AGCAAAG	CAATGGGTT		IPTREEWKARRDSLPGTGDVWYTDGSRAETGTGSGYYC		
			GCGCCTC	TACAGCCTA		RRDGRGTFFSLGRYATVFQTEIYAILTCAQRNIELGARD		
			GGCCGCA	AGTGCCAG		RIITICSDSQAALRALMAHRTTSRLVWECKVVVNQLTA		
			CTGGTCA	GGATCTCG		HNNKVRLLWVPGHTGIRGNEIADRLAALGAKHPPIGPEP		
			GGCGCAT	GAGGTCCC		YTGAARCLLAGEIRDWVEREHTKEWQGTQGCRQAKAV		
		Nasonia vitripennis	GGCTAAG	CCCTTCATA		MGQDTNVGWTKCIAGGSRNNSRLLTQIVTGHIRLRYHG		
			ATGCACA	GTAATAAT		LKMGKEATGVCRWCEGAEETPTHVLTVCPKFAQLRHQ	cd01650	459
Loa	Baggins-2_NVi	.	CAGGCAT	AATAATAA		CLGELFPTYEEIRALDAGAILTFWRRAGLPA		718
			GACTAGA	TGACCTGA		TFAPTHPMVRSGPCRKTKRPGSDYRESLIMDSGNNVAS		
		Nasonia vitripennis	CTATGGG	ACAAAACG		EPRGAVDVTSAAPIGAELNAEPCEGRNQRREAALSAQT		
R2	R2B_NVi	.	TTCAGTC	TGTTGTCTT		RRRNXARRARNAQQADEPGDDEEIETHGPLTIRTXEPM	cd01650	626
			AGTCCCA	GTCTTGTCT		EIVAIAKNPQACPKCLQGGTQLLCMGSWELSRHINKEHP		885

				AATAGCC GATCCTG GCGCGTC CGGCAGT AATGCCA CGTATGA GTCGGTT ACCCATC TCTAAAC GCGTAGA GGTGGGG AGCTAAA GGCCAGG CGGTTTA CCCGACG TCGAATT TCTCCAG GTCTGTG TCAGTCG ACGGAAT AAAGGTA CTACAAC ATCTACT ATCTATC GGGATCG GAAGACG CCTTACA GCGTTTT CCGATTT TTGCTCTT TGAGCAT TTTTCTTC AAATTGC GATAACC GACCCGA TCACGCG GGGCTTT GACAAAG CAATGCG TGGTCGG	AAAACTATT TATTCGAAA TAAGGGGA GGCTAACT GCCTGCAA GTTGAACG CGAAAGTT AGACCTTCC CACCTAAA GCCCAAAA GTGATCGG GGAATGAA TCCGCGGGT GACCCCAG AGTTGGGT AAACCCTTG AAACGTTG GAGAAGCG GAAGAGAG TCCCGCCAC CGAGCATC GAGTGCTG CGGCGCCC GAATGAAA CCGATCGC GGATGGTG CAAGTCGT AGGACGGG GCACGACC TAAGCCTCT GTCACGGC GGCGAAGC CAGGAATC ACCATGCA AAGGTGTG AACTGGGG CGGATACCT CCACGGGG TTTCCCTGG GCATCGCG	SVDVTWVCGACQRRCTTLRSWSCHVLHCKGRQEPKDL PFKCEHCSLSFDSQIGLSQHERHVHPEVRNDKRAAEAN KPKGKSGRRPSIWSDEDLLLIRELESEYHGARNINEKIAE HFPDRTGRQVSDARRRKDYAALRGRGGPQGPAEGVEAI EEVDEGEIPEGEELVATDGAALESGPPENGGSAPAEQVN APALESSSQQDRECSPAVGSDEQIEDSSDDDEFSDALGEI SLPEPLSVERTTISPPPRDDWKGPMRWEICNASEEAGSY ANWVTGLQELVRNNALSEIGLDSLYDQLIQIMRHPSDD NEQDRLQLNARGPPRRGHRKNRRRRRLTAADRKRFAF ARCQDLWNNNPKKLAELVIANDLSILQRRQAPGRTETQ TLYNELWGRVGPNIEAPRRTEDPIPVSRIFTPITPQEIMGR IRRIKNDSAAGPDGVTKDDLRGRGVSIALSKLFNSILLAG YYPKAWRENRTTLLPKPEKDPADVKNWRPITISSMVSR VYSGLLDQRVRAVIKQCDRQKGFTEENGCFSNIQLLDD AVSNAKKAGGVITILDVSKAFDTVPHAVIQGCLEKKGIP ETVAAYISSMYRDCSTAIRTRSGDVKIGMKRGVKQGDP LSPLIFNLVLEPLLERLQETSGVEIEGMNLSCAAFADDIV CFANTAPEAGRQLRMVADYLGRLDMSLSVSKCIAVEY VPHRKTWYTKNPGLEVNGNAVPSISPSETFKYLGAKVS PWKGLLEGFESDAFREVISRVQRLPLKPMQKVDLLQMY IFPRYTYGLITSPPAKAVLKTIDRIIRTRIKEILHLPESVSSS FLYTPRKQGGLGLLEVEKMVLIAALRNGLRARQSHDPV TRAAMNSNAADDRLKSYADALRLHWPLTTKELDTYKY QLRLSYAQKWAEQKWQGQGVEEFAQDPVGNSWLQRY DLLPASRYIDAIKLRTNTYPTRALMKIIDGRVDSSCRKC QGSSETLGHILGRCRYTKDKRISRHNEIKDLLKARLAKN HQVMDEPQITVRGQRFKPDLVVKTNEGRVHVIDVTVRY EHRTYLDEGRTEKIGKYRQILSTLRRDLHSNAEEVIPIVI GSRGAIPRETRKALSKLGIGKSDWLTISLIALRSSLEIVNA FMDD			

				TAAGATG GTTGCAA TCTTTTCC ACCTCGT TTCTTTTA CGGAACG AAAGCAA TGCGTGT GGGGAAC GTTAAAA ACTCCCT TCATGCA TCCCAGG ATTTATC CTGCTTA CTGCAAA GCAATGC GTGTGGA GCGACTT TACCACG AGTCGCT CCACCGC AAAGCAA TGCGTAT CGCGCAA AAGCAAT GCGTGTG GGGGACT TGTCAAA GATCCCC CGCCGCA AAGCAAT GCGTGTC GGCACCA CGTAGAG CAAAGCG TGTAGGC AGACTTT GTCAAAA GTAGTTC	CGAGCGAT GGCCAAAG TCCGCTTTC TCAGCTACA AAACAAAA ATGGTATG AGACTTCGT TAACACTA ATTTTTCCG AGCCTAGC AGGCTCCCT TGACAACG CTTATGAAT CTGGAAAA GGACACAA AGTGGAAA AAGCGCTG ATGGTGGA CAAAAGTC AGTTGAGA CTTGATATC AGTTGTTTT GACTAAGA ATTTTATTA TCGTTGACT TTTAAATAT TTTATTATT GACTGTTAA TATACTGAC TTGGGACC AAGTCATCT CTGTTACCC GGTACCGG TTCCTGTCA TCAAACCG GAAAGTCC GTCCCACGT AATGTGGT AGACGCAG GAG				

				TGCCGCA						
				AAGCAAT						
				GCGTGTG						
				GAGATCT						
				TCGCCGG						
				TGAAAGC						
				AATACGT						
				GTGGGCG						
				AACTTAG						
CR1	L2-5_NVi	.	Nasonia vitripennis	TGCTGTA	AA	GILPHSPVLPVRMNTPPARPRPRRRELNQQQRQQQQIHL	cd01650	901	1157	
				ACCTAAC		PRARLERMCPKRINRHPHHRLLAPEPNRLREHRLRLSRRP				
				CTACAAC		LRTSSRRWMTFGRSRMRPFRHKTPCLKGSLELRKGLAL				
				CTCTCAT		LQSRLKALDDLPALKTRLNNVESSISELQAQYQELSSRS				
				GGCTTTT		PAEQQNISIAPASEEINNLRSELAEVKRRQERSANNVVVI				
				CTCCTTG		SDLAYTQETSLQLLAFTVLAALDSTVLRRDVATVRTMG				
				CTGCACG		RLDATTTNAQGDSRFPPLAVTLSSSALARSIIVAKARKR				
				GCGTCTA		KLHTSELDAALLEEARALCPDHQGLINNELLPPDIHKLR				
				TTTCTCCT		IKARLEAKRRRGCRSFVRDGRIYIRSGEDSERATIISTDA				
				CAAAGCG		ELEAFLAHTPPIANTDITHGFFSPYAANSQHGHHTLRHTP				
				ATCCACG		IIPPKPPSKSSNSSTSASKASLSEGLRVCHFNANSLTGHIE				
				TGGAGCC		MVRLFLSTRSPFHVIAVTETWLSDKITSIPSLDDYILYRR				
				ACCTGTC		DRNRNGGGVALYIHHSLTASVISSSDGEWSGKPGKPEY				
				GGGTACA		LFCEISAKGVSPIFVGVVYRPPHSPFIQGSNFIEQLTTHMH				
				CGGAACA		NYSTKVIMGDFNSDQLSSSEDANFIRAFIDENSLTSVPYG				
				CCACCAG		ATHHRQDSDTWLDLCLIDEQDCLLSHWKTDTPFINGHD				
				CAACATC		LITATLDVQIPRHVPATYSYRNYKGICAEKLRDFLSACD				
				ATCAATC		WSSFTTSSLDECITILNANLTNAINHLAPLKTVTPGRKRH				
				ACAGGTC		PWFTAALRDLLTERNRLYRRFRKSRLPWDLYFYRIARD				
				GGTCAAT		DAHKQVEEARLNYYYSRLSTLTDVAEIWRELENLGITTS				
				TCTCACC		KSPSPARFSSDALNKHFSSISNDPPQAPSVEEYLRTLESLD				
				TCTCTCG		LPEHFIFRAITESDVSAAVSHFNTQARGSDGIPQVVISKA				
				ATTAGCA		LPILAPLLCQIFNLSLSEARFPSAWKSSLVRALNKVNSPT				
				CCCCATC		ALTDYRPISLLCFLSKALEWLVHKQISEYLETRLYLDNF				
				TGGCTTC		QTGFRTGHSTQSGLIKLTDDVRLGIDRKKVTLLLLFDFS				
				GAGAGTA		KAFDTVCHVRLLGKLSSFGFSKQVIRWLASYLSGREQA				
				TTAAGTC		VIGDNNELSTSLPLNTGVPQGSVLGPLLFALYINDIGFCL				
				TCTAGCT		DSDVSHLIYADDLQIYSCHLEELDSCSVRMSANAERIMG				
				CTGTCGA		WAALNRLKLNVLKTKAIVLGSPYYINALPSIANTFINIGG				
				TAACTTA		ARVDYESSVRNLGLVLDSKLTWKEHVTQICKRAHSLM				
				CCACCCG		YRLYFFRKSTNFRLRQHLVQALLFPIIDYCSLAYCDLTQ				

				CTACTAT			ELDTKLQRLVNTGIRYIYGVRRDEHISPYRRELQWLTTA			
				CTTACAA			GRRKYFMACFLRKIFNTSVPTYILAFFDFHVALRPVRGE			
				CACTCTC			VTPLDIPSFKTETL			
				TACGTTA						
				TCGACTG						
				CCAGTAA						
				GTTGTCG						
				ACGGAGA						
				CTTAACA						
				ATAACAT						
				CGAAAAT						
				GAGCGGA						
				ATTGGTG						
				TGAAGAA						
				TGACGAG						
				TCGCTCC						
				TCATGTG						
				TGCTCAC						
				TGCAAAA						
				AAACAAA						
				AAGAGGG						
				AAAATGA						
				AGACCTG						
				CGAGTTG						
				TGTGACC						
				GGCAGTT						
				CCACACC						
				ATATGTG						
				TCGCCTT						
				CAGGTCT						
				ATAATGG						
				TGA						
				GGTGGGA		TAATGAAA	MKIISWNVRGLGSSRKRSVIREVLRKEKPTIFLIQESKLT			
				GTAACAG		GTTACCGTT	EVDRRWVRSVWSARGLSWAFTQSWGASGGIITIWKNE			
				AAGAGAT		TCCTATAAA	EMDCTEELIGNFTVTIKLRRKEDGFEWALTNVYGPNSY			
				AGGGGWT		AAAAA	RERDDFWXELAAIRGLWNGPWCLGGDFNVIRFLNETTN			
				GGTGAGG			PNHRVTSSMRKFNLFVDSYELRELPMLQANFTWSNFQE			
				MTGTAGA			EPLCCKLDRFLFSSDFESAWRGLAVRALPRPTSDHCPVR			
	L1-11_N		Nelumbo	AGGGTCT			LSQENLGGGPCPFRFENMWLLHPDFIKKVQLWWEEDQ	cd01650	496	761
L1	N	.	nucifera	AAGGAGG			VEFRWKGYKFYKKLQALKARIKAWNKEEFGRVEERKE			

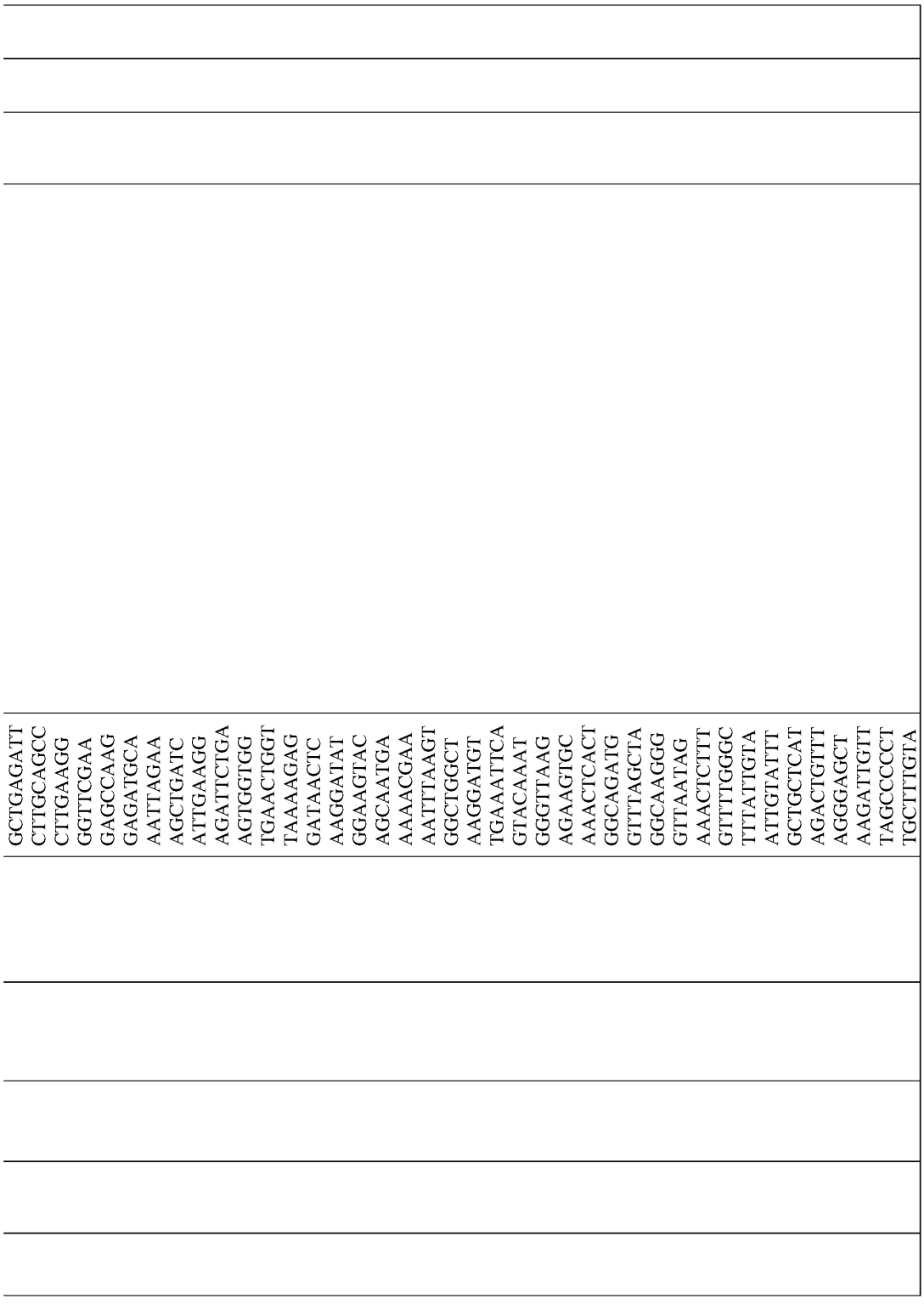
				AAGGGGA GMGAATC TGGGAAT TAATAGT GATCCTT TTCGAGC AGTTGGT AGAGGTG GAAATTC ACTCCAG GTTGGGA TCAGAAG AGGGGGC TGAACTK ATTGGGA CTACTTCT AAGGAAA AGGAGAG GGAGATT AATAAGA GTGGTGC TGGAACT AAAAAGG CAATTAA AAAACTT AGCAAGA ATAAGAA AAAAGGA CAAGAAG TAACTCA GTTAAAA GCTTCAA ACAAGGG GGAGGGA GAGAATT CTAAAAG AAGAAGA AAAAAGA TGAGTGA AGTCGAA		WIANRIQEIDGGEAERRLTVGEREERWKLKQELQEIIRK EEISWRQKSRCKWLKEGDGNTKFFHTLANGRKRKNFV GRLEWNGEVITSEEEIARAFVEHFKSIYDSPRREREEIPD LDWLPGNRVLMDQLEKNFEVQEVKDAIFSMERSKAPGP DGFTMAFYQDCWDIIGGDLMEVFLEFFERGKINSGANA TFIALIPKKVNANHPKDFRPISLIGSMYKVISKVLAERIKK VLPSLISPNQGAFVSGRQIIDQILIANECVDYRKRQGEKG LICKLDLEKAYDLVDWGFLLEVLRIRGFGERWCMWVK GCISSCHYSILVNGVAKDFFPAKRGLRQGDPLSPFLFTLI VDVLSVQITRADARNLLRGWRVGREEIVVTHLQFADDL ILFCEGDPEELENLKVLVENFEKVSNLRVNWNKSCVAGI NRQEVEVASLARSLGCEVSDWPLKYLGLPLGGNPRSVS FWYPVLERFSKRLAGWKRNYISLGGRITLIKSTLANLPV YFLSTFKLPVKIANQVEKIMRDFLWEPGDTKKDHRVKW EVVSRKKDKGGLGIGNIRIRNAALLGKWLWRFPKEQES LWHKVIRSIHGMAANGWDAGLDPRASRFSPWKDIVGIL PVFSPFIRYDACVGNKILFWEDLWIGESPLKDRFPNLYSL SLKKSQFISSFRASGDSWDLGLRRGLRDSEVDSANDIFSL LSSVFLDHSSPDSRIWAETSTGSFSCSSFFNALVLPSNPVS SFPVKEVWSAPVPPKVRAFCWEVVLQKINTLEVLQRKR PNWCLSPHVCSLCLRNGETVNHLLISCPFSALIWDWFSS LFDCRWTPPTSVADLFDPSFFRQPKKKGRVMWRTTRAA ILWAVWEERNNRIFQGKEKSAPEVIEGITIKAVHWLKTL DLFKDVSIDDLIRGWSILARSSTSSPTIQAVWSPPLQCNV KLNFDGSSIGNPGPSGIGGVFRDAEGKILALYSGPIGLGD SLRAEVLALLQGLRWAKQLGFQQLEVEGDSKLVVGWL QSNSLGTWSLRQHVRELLSLTRELGAQVYWSRRSANAL ADNLAKQGVNREVFVLCSVVSYLLFRGHLLLWKVTVV SSGFI			

				AGACTTA						
				GGTGTGG						
				TGTGGAT						
				TATGAAA						
				GAGGAGG						
				CACAGGA						
				AGGAAGG						
				AGAGAAA						
				GCTGGGT						
				TAAATTA						
				AGCTGAG						
				TCGTTTG						
				GGCCTTT						
				TTGCTTA						
				GTTTTCC						
				GAAGTTT			MDMCGCLQMSMGQMTTEREELWEELADVRGLWDGP			
				GGGGCAA			WCVGGDFNVIRYPEETNKTNQRVTTSMRKFNRFVDAFE			
				AAGGAGA			LKEIPLAQKTFTWSNLQARPICSKLDRFLFSVDWEEHWG			
				KAGAAGG			NLTSRALPRVTSDHWPIVLCKEERRSGGPAPFRFENMW			
				AAAGGGG			LLHPGFTNRVREWWNECNPLVEWEGMRFHKKLQFLKA			
				GAAAGTA			KIKEWNILEFGRLEERKRDLMDQLESLDSKEERGEFGEE			
				GTGGTCA			DNELRRGLRRQLEDCLHKEEISWRQKSRCKWLQEGDR			
				AGGGTAG			NTRYFHSLADGRRRRNFISKLVVDGIELTQQEEIAGACV			
				AAGGGAA		TAACTTAAC	EHFKSILRRPVSGRVWVEGLDWPDVGGDNRRSLERXFD			
				GTAAGAA		TGTTTCTTT	LEEIKEAVFSMDRTKAPGPDGFTMAFYQDCWDILKEDL			
				AGAAGTG		TTTCCTTTT	LMVFEEFHRRGIVNPGVNSTFIALIPKKKEASHPRDFRPI			
				AGAATGA		CTCTCTTTT	SLVNSLYKIIAKVLSLRLKKVLPDIIEQNQGAFVSGRQIT			
				AGATACT		GAGGGTTT	DQILIANESVESFKKDKKKGVVCKLDLEKAYDMVDWE			
				AAGTTGG		AAGTTGTAT	FLRKVMTKMGFGARWCKWIMGCVRNCHYSVLVNGTV			
				AATGTGA		TCGGAAGA	VDFFPATRGIRQGDPLSPFLFLLVADVLGKLIHRADSRG			
				GGGGGAT		TAAGAGTG	LIRGWRVGREDIPLTHLQFADDTILFCEGGNVSFIRNLKL			
				AGGAAGT		TTTGTCTTTC	VIQLFEQMSNLKVNFGKSCVVGVNLEEGEIRESAEVLGC			
				GCGAGGA		CTTGCTGTA	EVSGWPLSYLGLPLGDNPKRESFWNPVVERFEKRLVSW			
				AAAGAGC		ATTATTCTT	KKNYLSLGGRTTLIKATLANLPIYFLSVFKIPXKVSSRLE			
				AACAATA		GTGATGTTT	RLMRDFLWEPGEEKRDHRVSWKVVTRSKNYGGLGLG			
				AAAGAGG		CAATGAATT	NIGKRNSALLGKWLWRFPLEQEKLWARIVASIHGMEQN			
				TAATAAA		TAATGAATC	GWDAGIRARTTSTSPWKAIHDLLPVFSPLVRYDLNSGNS			
				GAGGTTC		TGTTATTGA	IKFWEQIWVGNVPLKDQFPSLYALSDFKNEPISSFITPSSS			
			Nelumbo	GACCCAG		TCAAAAAA	SSGFGWNLGLRRAVRDSDLDSISSLLELLDNVFLDPSKV	cd01650	398	
L1	L1-5_NN	.	nucifera	ACATTGT		AA	DSRIWTPTPTGSFSCASFFKVLCQSTSPATSPFPVNEVWC			663

						RPIPPRVGSFCWEAYLERFNTGDLLQKKRPNTCLSPHWC VLCKADGETGFHLLVSCQFAKSIWEWAFSLFGRRWNPP SDLQSLFSPSFFRWHMKRGRTMWRGLRAAVCWSLWEE RNSRIFSEDERGAEELRAVIHIRAVQWLKSLELFTDYTV EDMLRKWGEIANCTQTKPVIRVSWTPPPAGFLKLNFDG SCRGNPGQSGIGGVFRDENGVISALFSGPIGVGDSLRAEI LAVMEGVQKAKEMNVRKLIIEGDSEVVIGWLIKDNCGL WNYXNERRKFRWLVREMEIRMSWVKRSANSLADGLA KQGIEREAMFWAKL				
									cd01650	
L1	L1-10_NN	.	Nelumbo nucifera	ACTGTTG CAGGAGT CCAAGAG GATGGAG TTTAATA ACAGGTG GGCAAGG AGTGTAT GGAAAGC CAGGGGG ATTAGTT GGGTGGC TGCTCCC TCTTGGG GTGCTTC AGGGGGG ATAGTGA CAATGTG GAAGAGG GAATGTG GTTGAAG GCTTGGA GGAACTG ATTGGAA GATACTC AGTTTCC ATGAAGT TCAAACA GCAGGAG G	GGAGATT TGTAGGG AAGATTC AGTGGCA GAATTTG TGGGAAA GTTGGGG CGTAAAC TGAAGAA CCCAAAT	TAGTTTCAT ATCTTTAAT AAAGTTTCT TGTTATTGA TCAAAAAA A	MRIXSWNVRGLGSARKRATIKEVXKKFNPDVVLLQESK RMDFDCRRARSVWRAKGIKWAAAPSWGSSGGIVTLWK EDVLIGLEELIGDHSVSLKFRQVEDGFEWVLTNVYGPN DYRQRKKLWEELSDIRGLWEGPWCLGGDFNVVRLTEE TNNPRQRVTTSMRSFNQFVDXFELKELPMLNNTFTWSN LQSSPICSKLDRFLFSTDWEAHWGNLTSRVLPRLTSDHW PIVMCKEXLRGDGPAPFRFENMWLTHPGFIEKIKEWWQ HYNPPVQWEGMRFHKKLQFLKSQIKEWNRLEFGRVEE RKQVLVEQIEELDMKEGRGLFGEEDRERRWSLRKQLED LLIKEEISWRQKSRCKWLKEGDRNTGFFHSVANGRRRK		498	763

				GATTTAA AAGCATT TCTACGA AAAAAGG AGGATAG AAAGNTC GAGAAAA GAAAATT GAGCGAG GTTGAGA GATTAAG AAGAGAT ATTGAGG GGGATAA GAAGTTT GGGGCCA GAGGGGA CAGAAGG AAAGGGC CCAAGTA GATTTTG GTGGTGA TAGGGCA GATTCTT GCAAAAT TGTG		NFISKLEVEGTVISHPVEIAEAFVGHFKSIFRRPLNERIWL EGLDWPTLSPENREWLERRFXLEEIKEAIFSMDRTKAPG PDGFTMAFYQDCWGIVEEDIGSVFEEFHRRGIINPGLNST FIALVPKKKEASQPRDFRPISLVNSLYKIIAKVLSLRLKK VLPSIIAPNQGAFVMGRQITDQILIANESLESFRKSKKKG VVCKLDLEKAYDXVDWSFLLEVLKSFGFGDSWCXWIW GCIRSCHYSILVNGTVVDFFPASRGIRQGDPLSPFLFTLIA DVLGRLIQRADSRNLIKGWRVGREEIPLTHLQFADDTIL FCEGGDIEWIRNLKLAIQLFERVSNLKVNLDKSSLAGVN LEERELREYADVLGCATSDWPMAYLGLPLGDNPKKVSF WNPVVERFGKRLASWRKNYLSLGGRITLIRSTLANLPV YFLSVFKIPTKVSNTLEKLMRDFLWEPGEEKRDHRVSW EVVTRGKKLGGLGIGNLRKRNMALLGKWLWRFPLERD KLWTKIVASIHGLEDNGWEPGILARTTAASPWSAIHKLL PVFSPCIRYGLNSGNKIKFWEQLWVGDSPLMDQFPSLY ALSDLKNEPISSFLSPSDSPSGFGWNLGLRRAIRDTDLDH ILSLLAILENVFLDPXKEDTRFWVPSSSGVFSCASFFETLT QPSDPIGLSFPVKEIWSRPIPPRVGSFCWETFLERINTNDL LQKKRPNSCLLPNRCTLCKADGETGYHLMVTCSVAKSV WEWVFSLFGCVWXPPGSVQSLFSPSFFRWPKKRGKTM WRSIRAAVCWXLWEERNSRIFSDKERTIGNLLDLILLRV AHWLKTMELFSDYSVADXVRNWKEIASSALCKRTCSPS WSPPPSGFIKLNFDGSCXRNPGPSGIGGVFRNEAGAIIAM FSGPIGEGDSLRAEILAALEGIQRARELNIRCLVLEGDSE VVIRWLKGGGSGLWSYNNEKNKFKWLSRDIVVHYNW VKRTANSFADGLAKQGVNRDSTFWANFLXFLVFFGGEL SLFTSPLYLSFT			
L1	L1-17_NN	.	Nelumbo nucifera	AGTGTTT GGAGAGG TAGGGGG ATCAATT GGGCGTT TCTGCCC TCATGGG GGTCTGC AGGGGGT GTTGTGA CA	TGAATGTGC AGAGTTTAT TCAGGTTCA TGCCGAGT GGGCCAAC AAAAAGGG CCAAGGCT ATGTGGAG ATGTATGA GGGCTGCA GTATGCTGG ACTGTTTGG GAGGAGAG	MWKKEVLEGKEELIGQFSVSLKFRQVEDGFEWVLTNV YGPNDYRERRQLWEELADIRGLWEGPWCLGGDFNVIRF LEETNSTRQRSTMSMRKFNQFVDAYGLKEIPLSHSCFT WSNLQVNPICSKLDRFLFSNDWEEHWGSLNSRTLPRVT SDHWPIVICKEEVRNAGPTPFRFENMWLGHPGFIKKVK DWWLESNPTVKWSGMRFHKKLQHLKGRIKEWNLLEFG RLEERKQMVTAQIEALDSKEGRGCFTEEDKEQRWNLRK QLEDCLLQEEISWRQKSRCKWLKEGDRNTKFFHAMAN GRRRRNYISSLEIEGATISNQSEIAGAFANHFRSIFRRPLN NRVWIEGLEWPTLKLEEREWLERKFXLEEIKEAVFSMD RSKAPGPDGFTMAFYQDCWEILKEDLRIVFDEFHSRGVI NPGLNSTFIALVPKKRNANHPRDYRPISLVNSLYKIITKV LSIRIRKVLPTIIADNQGAFVKGRQITDQILIANESLEEFK	cd01650	426	691

					GAACTCAC GGATATTCT CAGAAAAG GAGAGAAC CTTGGAAGT AATATTGG AGGTAATC CATATTAGG GTGATACAT TGGCTGAA AACTTTGGC CTTGTTTGC AGACTTCTC AGTAGATG AGATGGTA AAAAACTG GAAGGAAA TTGCAGAAT TAGCCTCAG CTAAGAAG GCATCACTC TCTTGCTGG AGTCCCCCA CCTAATGGC TTTMTTAAA CTGAATTTT GATGGCAG CAGTATGG GTAATCCA GGTCCATCT GGAATAGG GGGAGCAA TAAGAGAT GAACATGG AAATCTGTT AGCTATGTT TTCAGGACC AATTGGGG AAGGAGAC TCCTTGAGG	RSKKKGVVCKLDLEKAYDMVDWSFLLEVMVNFGFGD RWCKWVMGCIRKCHYSILINGTVVDFFPALRGIRQGDP LSPFLFTMVAEVLGRLISRADSRDLIRGWRVGRESVPLT HLQFADDTILFCEGGDLNRIRNLKLTIQLFEQMSNLKVN FEKSCLVGVNMGEGETSECADILECAVSNWPLNYLGLP LGDNPQKDEFWIPVVERFEKRLATWKKNYLSLGGRITLI RSTLANLPVYFLSVFKIPAKVASRLEKQMRDFLWEPGEE KKDHRVRWEMVCRSKREGGLGIGNIRSRNLALLGKWL WRFSLEKDKLWVKIVGSIHGVADNGWDSGIVARTTRA APWKAIQAILPLFSPWIRYDLHSGNLINFWEHIWIGDSPL KDQFPSLFTLSTLKNEPVSSFLVPSNSTSGFSWNLGFRRE IRDTEFDCILSLLSILEGVFLEPTKTDSRIWSLTPSGSFTCA SFFKALTQPPSLGLSFPVKEVWCLPIPPKVGSFCWEAFLE RVNTCDLLQRKRPNSCLCPHWCALCKADGESEYHLLIS CRFTKSIWEWCTSLFGCRWPPL			


GCTGAGATT
CTTGCAGCC
CTTGAAGG
GGTTCGAA
GAGCCAAG
GAGATGCA
AATTAGAA
AGCTGATC
ATTGAAGG
AGATTCTGA
AGTGGTGG
TGAACTGGT
TAAAAGAG
GATAACTC
AAGGATAT
GGAAGTAC
AGCAATGA
AAAACGAA
AATTTAAGT
GGCTGGCT
AAGGATGT
TGAAATTCA
GTACAAAT
GGGTTAAG
AGAAGTGC
AAACTCACT
GGCAGATG
GTTTAGCTA
GGCAAGGG
GTTAATAG
AAACTCTTT
GTTTTGGGC
TTTATTGTA
ATTGTATTT
GCTGCTCAT
AGACTGTTT
AGGGAGCT
AAGATTGTT
TAGCCCCCT
TGCTTTGTA

RTE	RTE-7_NV	.	Nematostella vectensis	TCCTCGA TGGGATC GCCACCT TATCGTG GTGAGGG TGTTTGT GTGTCCC AATGACC TCTAGAG CTATGTC GGCGGGA GTATTTT ACTCCTG GCAGGGA CACCCAT GCCGAAC AGGTCGA AAGGTAG GGGCCAG ACGAAGA GTGATCC ACTGGTC CTCCAGG TTGGGGG TTGGGCA AAGGGTT GATAACC CTCTCCC ATAAAAA ATAGCTT ATCACAG AAACCAG AAGCAGA	ATTCTTCTT AATTTTAAT GAATACATT TCACAGTTA TTGATAAA AAAAAAAA A	TAACAGGT AAC	MTRRGQSHPDAPGLTKPLVHPKHAMRIGNWNVRTLYS SGNVAQAAREMGKRDIDIMGISETHWTGQGKMQLRDG ETMVYSGREDNIHREGVGILMSKNAAACLIEWTPWNER MIQARFYSRHIQLTIIHIYAPTEDADEHLKDEFYTRLQEV LDSRNQHDMLIITGDMNAKVGDQNWDYERVVMGKHGL GVRNDNGERLCELCDLNELVITGTLFPHRTIHKATWISP DGKTKNQIDHVLVNKRFRNSVKDTRVFRSADVGSDHY LVCATVKLRLRKDPKRKGNIRVKYDTSKLTEKEVRNTF NIKLRNRFQALEQEDEEERGEEVVDAKEEVERDFTIMK KALLEVAETVLGKPRKKNKPWISGESWSLIDERQQMNE KILSTRSERVKGQLRKKYREKDREVKRSLKADKKKWLE EIASEAEKAARSQHMKTLYGLTKKLCNERPRHSTAVLD KDGNLLSKKEEIQRRWTEHFREVLNREQPDNPIEPEDEV EFEFGELIEEISVNEPTFGEVKDTITTLKNGKSPGIDNITA ELLKADPEFSAERVHKLLVKVWRHETIPNDWKKGLIIKL PKKGNLKECKNSRGTTLLSIVGKILGRIVIDRIRNGVDKK LRKEQAGYRRGRGTTEQVFILRNIIEQVNEWQATLYVN FIDFEKAFDSVHRESLWVIMEKYGVPEKIIRIVRLFYEDF QCAVEDQGETGEWFNIKTGVKQGCNMSGFMFLLVMID WILRRSVGKGENGIRWKFTSKLDDLDFADDIALLSSTKH HIQEKTAKLDGEARRAGLKINLEKTKMLRINAKQHGNI VINGKEIEDVSEFTYLGAKVCREGGGMKDLRNRLSKAR SVYTRLKRIWNSKKIMRRTKIKIYKSLVLSVLLYGCETW KMNKKDDKLLDVFQQKCLRRILKISWEDHVTNEEVQR RAGVHQLSSEVKRRRWKMIGHVLRQGRENDCSVALTW TPEGRRKQGRPKTTWRRTVERERGEAGWRNWDEVRSK AADREKWKSTVKALCATRHEEDR	cd01650	567	824

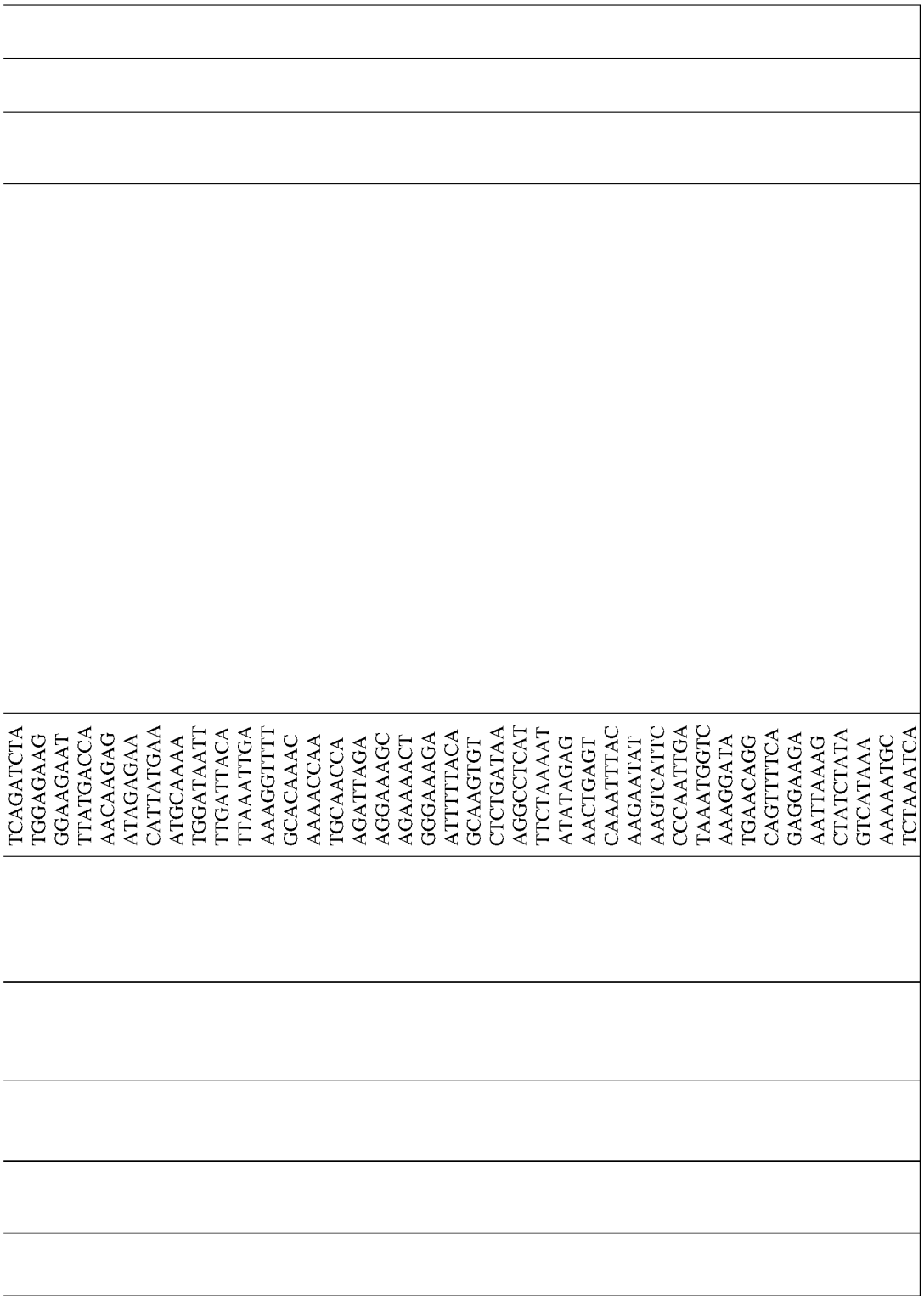
			GCAAATT CACTTGG GAAGACC GTGGCTG CATCTCA TGAAAGA GATTGT		TAATAATAC CTATTGATA TATATATGC ATGAAAAG ATATATTTA TTTTTATTT AACATTTAC ATTAACAAT TAACATATT TTTTACTAG TTTCTGATC TTTAGGGTG CATGCCATG GATGCATG GAGCGGCT GCCAAACT AGCGGGAG TTGGCTCAA GGGTGAGG CAAAGATT CTTTGTTTC GCGCTTGG GTTGATTCT TGTTGGTTG GCTGTTTCA TGCATTTAT GGCATTCAC TCTATTGAT GATAATTTT AATTTCATG TATATATAA TTAATTATT TAATTTTAT	MAANFLTPVFIAFLAILFNQKLHYSPNLTSGINGFGLNCI LANSESILTGLEAFYSHNIRANKMNVGRSSSEHKHLLLC LLLLGGDIQLNPGPKWKFPCGSCNKPVKSNQKGIQCDN CNLWYHTRCCSINDDLYYTLAASSCTWLCVSCGFPSFSS SLFESSSIELQNSFSSLDSLHQDTQNISSSTIRSPETSPLDSS SRNSSNTRKQSKEKLTGMILNCNGLKGTDHITKFQALIN LHDPDFILGTESKLCPDISTYSIFSQNYSVFRRDRNKFGG GVFQAIKSDLVCTEEPDFEVDNCEIIWTSLKLANRKSIY MSSYYRPPNQPSDSIDNLNESIRRVFNKVKNFPNIIIGGDF NLGDIEWNSSVPTPTNPATASLHQKFLQTIDDFSLSQHV KTTTRPSSGKTLDLLLTTYPNVIVNTSTASGLSDHLAVIF EVNLKPSRSLKPPHKVYQYKKANFPDLCKFMSDSSTTFF STNPEKRTVEENWVLFKTSLSTGISQFIPQKSSRPKYKLP WINTNIKREMRKKDRMHKKAIKSKNQQHWEAFKRQRN LVSKLIKDSHNDYLNNIIGNSLKDNPKKFWSYVRSSKSE SIGIPPLKNNDSSVSVTDAHKAETLNTFFHSVFTQEQLPT PVLGNSPYSSIPDLHISAEGVCKQLSELNPKKACGPDEIP ANVLKEVSQSISHWLAFIFQQSFDCNKIPSDWSNALITAI HKKGLKSEPANYRPISLTSICCKIMEHILSHMAKHLSSH NILINQQHGFREKFSCETQLISAIHDWAKTINFRGQTDVL LLDFSKAFDSVPHQRLLIKLDFYGIRGNMLNWIKAFVSN RKQSVSVNGTVSSSKSVVSGVPQGSVLGPVLFLLFINDIS TSIQSNLRLFADDCVLYREVASLQDCQALQQDLDRLSV WSKTWQLKFNVSKCYHLGITRKKIPVKFTYSMDGKCIT QASSTPYLGVTIADNLSWNEHCDNICKKANSTLGLLKRI LSGCSPEVKDSAYRTLVRPKLEYATSAWNPYTQCNINKI EMIQRRAARFVFNDYSRYSSVSPMINSLGWESLEQRRQL NQAHMFYKIYNGLVGISFPPEVTPITRVSRQPNCAPFHQ LVTLNDTYKYSFYPRTIRTWNSINLSTIPDSLAKFKDSLL LKVQTN			
CR1	CR1-20_NV	.	Nematostella vectensis	GGTCAGC GGTCGAC CAAG		cd01650	707	959	

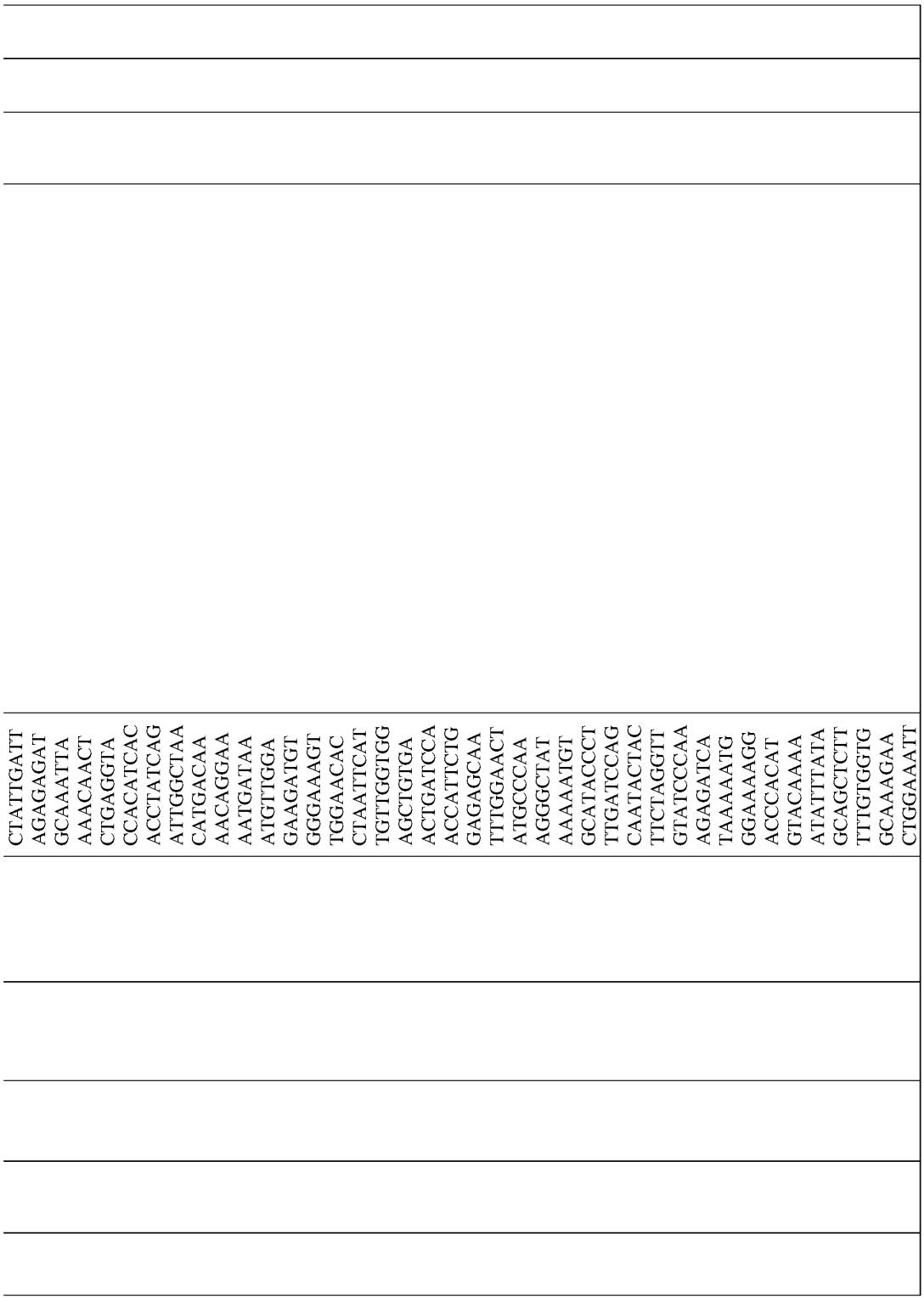
										864
										612
									cd016 50	
CAAATATCT AATTTCAAT AATTTTCAA GTAATATAT ATTTATGCT TCATTTTCA ATTTAACCT AAATTATTT TTCCGATGG CCGATAATT ATTATTATT ATATTTTTC TTTTGTATT TACATATTG TATTTTCTT TTTCAACTC CTAGTATAG TCTTAACTG TTTTTTAGG AGTATAACT ATTAAAGA TTAAAGATT AAAGATTA AAG									MINNPSNVVPVYFCLPILAFWSFLPCYFHLKAVDRGSHV ATLTSEPVASTGWSYSNQELKTIRKSQPKQRLEPKLWQ KLGDLGIRKPFRSRRKPRQSRHVCTGLPVVITTVDSTGHP QLSSGCSTNSSSSQTARFPVIPDTIIANMQSLLPKLDELY AAIMQFPAGIICLTETWLNESHPDPLIHLPGYLCFRNDRT TSRGGGVCTYINSQYPCRRLPEFEDLSIESLWISVRPYRL PRSVSTLLIGVIYHPPSNGEMNRQLVDHIIDKNCESFLSKH PDGLVVVTGDFNPSSTGIKAQYVTHKTGLTQIINVLTRD TGILDWCLTNKPKYFQQPLQLSKMGSCDHYPILIKPNLS VTNPKAHKSSVWRRNLRRSNLDAFGRWITQFCWNTILS LQTAREKFERFYTVITDAVNTYLPLQRVQVCATNKPWI GTSLCSLIARRQKALTTLGKDSRLYKELRNQVQRECKV CRERYYPNKVAALKESNISRWWQEVKCLSGSVVSNEW WHQLMDDSTPSVQILAERFNEFLLGLTSNFAPISQQVTC SGPVPPELLITPGKAFKALSRIKVKKSSGPDPVPSVIWKE FAPELSDVVCNLYNSSLMEGFVPQELKESIVHPVPKCSP	
CR1	CR1-18_NV	.	Nematostella vectensis	AGGCGTT TGACTGT CAAAG	TAATAGTAT AACGCTAT ATGTGCTGG CCCTTTGCC TTAATTCAG TCTGTAATA CTGCCATGG CAATTGAAT AAAACCCCC TATCTATCT ATCTATCTA TCTATCTAT CTATCTATC TATCTATCT ATCTATCTA TCTATC	CCAAAAT GGCTGCC ACGGCTT TTGTAGC TCTGTGA ATCTTGT CTCTGTTT TCATGTG TCATTGC TAGTTGA				

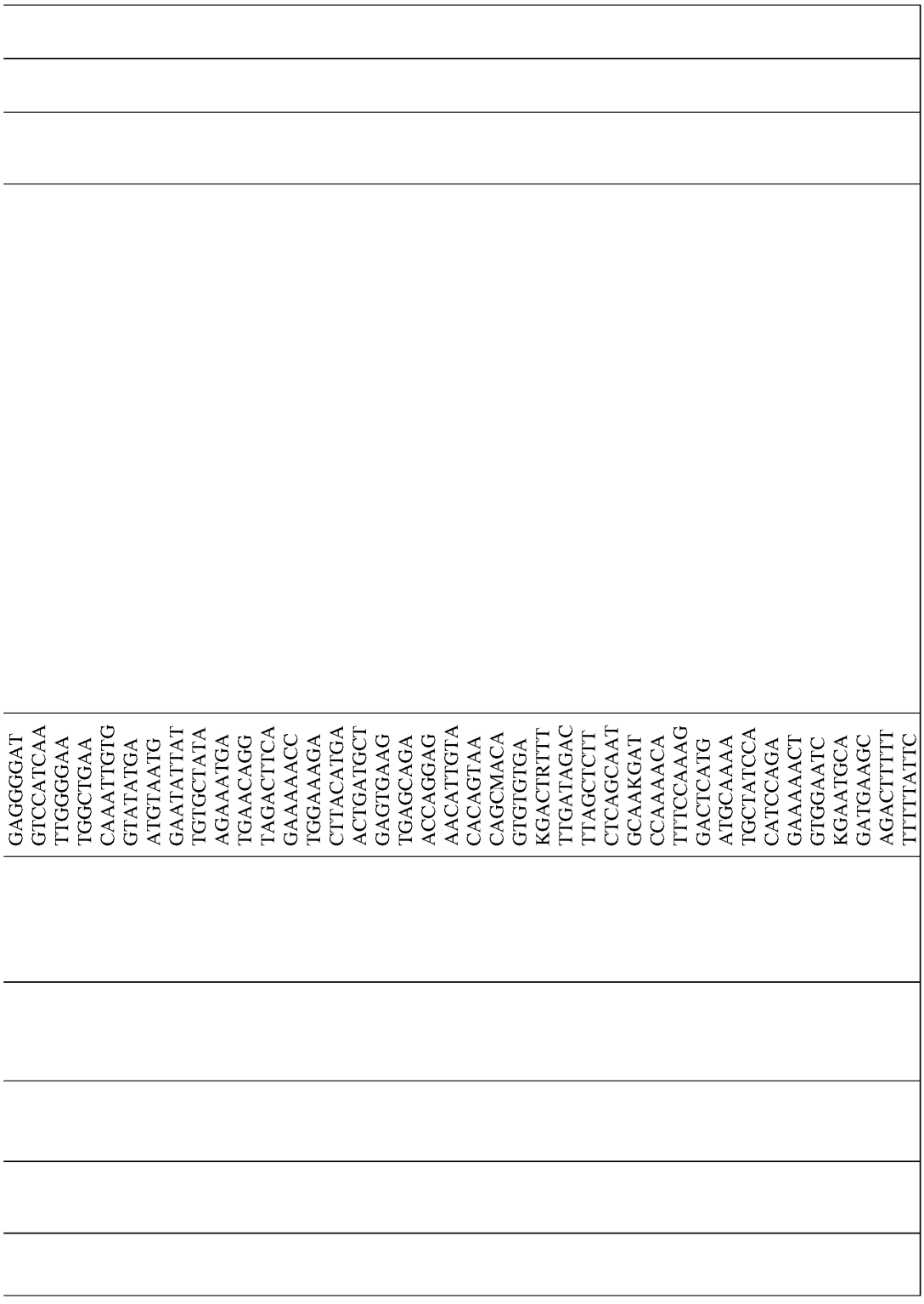
						PKTLSEDLRPITLTSQLAKVMEGFALDSLFACVKDKLDP			
						KQFSVSGKSTVLALVYFLHCILESLERGENYVRIFFADFS			
						KGFDLVDHNVLMSELDLLGVNEYLQNWIAAFLTARPQ			
						RVKISGHTSSPVFPNGGIPQGTKLAPLLFAILVNRLASTW			
						PYRIKYVDDTTIFEVIPRCSPSYLPFIANDICNFATERGMR			
						LNPKKCRELVINFLQYQPSPVNALQLMGSVIKRVESYKI			
						LGVHISSDLSWNKHVDYVVKKANKRLYALRLLKKSGV			
						PVQDLIAIYCVLIRSVVEYAGPVWANITGFLVDIVEGIQK			
						RALRIVLPHLRYKEALETSGLQSLADRRRELCISLMTSA			
						KELEPLRSVITGGVINNVHGYSLRSGSMNTYKHISVTKR			
						FGDFVTYRFL			
					CTCTCAA	MJKALHGGPWFILSHFLTVRGWEPKFKASSTQLTYSAL			
					TAATGTG	WVRLPELPTEFYDLEILSRIGSKLGRLLKIDTCTSSTTRG			
					AAAACCT	RYARICIEVPLEKPLKTHLYIGNHRQSLLYEGLNLLCTYC			
					TCCTCAC	GRFGHPTTSCPEVVRAPNEQNXRNXQXESNSTPDTIPSK			
					TACCGCT	ENEWKTVTFTKKSIHPPKHLNFDKGKAATMAATPAAQP			
					GCTCCTA	LXGQSSNPTSAVLGRVTSTRQVXLQGPVIEFNNKFQXLS			
					TAGAACA	NAGPSDLQTDHFYXGPIPKNJGTEPLLTQQXAKAHFHTP			
					CACCTTC	DKLPTSLPNXLPNDPKPRLSEPNPHLTKSPPLVGYNTFSQ			
					CTTTACCT	DYXTPHZPLPCNHLLTSSTQSSSNNSEGPLHILDTLXPTQ			
					CAATCTC	NKAIXTMHATVTSLTSDSRLPQNPCKQHDEFPCEHHHE			
					TTCAACA	XPTTLFPSLPNIYXHSIQTXKPNTTLXAEPQXLPNPILNAE			
					CTTTACA	TNLHRHEPESVSEGKYTNQQPVFSSPSSHNGGXSTASEFI			
					CWCATCA	TSNLHISAHDHGPPSPILVGNGTAGICSNIHSATQQPRGG			
					TGTCTGA	TVRKKFILKCPPEMVTCSLNFRQTTNPEVGKYAMVIIPT			
					CTCTCAG	LLQLPLSLDPMTSSVVHPVTQAPMSQPRNPNSYMNFLV			
					CAACAMC	WNCRGSHNPEFRRHFRSLLDNHRPTLAILLETHMQDHA			
					TGCAACA	SLREDFNFGNMAQAPANGLIGGLVVLWNADLINVTEM			
					GCTTAAT	RISDQEIHCMVQVSPTKPPWILSAIYASRHYQLRDILWNS			
					CAAGTTW	LINLHDNTAMPWLIGGDFNEILESKEKFGGLPMNNNRA			
					CCGACCC	DKFANFMHXCSLIDLGYQGSRFTWTNKQRNGVPYTKK			
					TCCGGAT	WRTILERLDRFLANYDWLNLXPNTTVIHLPRTHSDHCPL			
					ATAGGAT	LINLEPPQPIQNRIFRFETMWASHPQFQQLVHATWSNDV			
					CTATGCA	PLLATTDNFKRVVTRWNQNTFGNIFQQKRKLLARLAGL			
					AATCGAT	QSSVHYPTSVFLQNLETQLTLEYSHILRIEEEFWQLKSRI			
					GATACAA	NWLNEGDANTRFYHLSTLHRRRRNRINALQDNVGNWV			
					GTATGCA				
					TACACAA				
					CCAGCTA				
					CTTTCCTA				
				TAATATTAA					
				TATATGTTT					
				ATCTTTTTAA					
				ACCAAAAAA					
				AAA					
			Nicotiana sylvestris				cd01650	126	150
L1	L1-17_NS	.						4	3

				CAAACTC TACTAAC AGATAAA TACCTTCT GCCAWCT CTAACCA ACCAAAT ACCTTCC TTAAACA CTCTAGA TCTAGCA GATGAAG TACTCCA TAACMCT ACTGACG ATGACAC ATTTATA CCAATTA CATCAAC TGATAAA TCACGTT TATATGA GCCMTGG AAATACT CTGTGAT AGTTAAG CTWTTTG GCAAAMG GATTACA CACCAAA TCATGCA TGATAAA CTGATGA CTCTGTG GAAACCA ACAGAGG TCCTCCTC TCATTGA TTTGGGT TCTGACT		ATTXEIHDLIXNYFXDLFQTDQLISENMTGNTYNCEISSD DILXLXSPLHQDEILTALNSFKAYKAPGPDGLHPIFLPKIL DGCRIVLRHTKPQDQMGYIQFFYQKYWTDVGDKVSSF CQDVFCSKIIPPELNKTLICLIPKVKHPTRIQQFRPISLCNT VYKLITKIIVNRLKPIISKIIGPTQTSFQQGKRACDNAIIVQ ESLNYLAKMKGKYPSMLLKLDLEKAFDRLEWSFIRRAL FYFNIPPSMISLIMSCVTTTSISILINGSKTQYFTPTRGIRQ GDPLSPYLFIMCMELFSRSISLSVDYLKWQPIRLSNKGPL LSHLFFADDIILFSKITTKSCHAIIDVLNNFSKYSGQKINF EKSKLFFSKNCSSLIKILLLLPLTLKKEPLLVNTWDFLCS KQDPKMQISSSSLIISKINLLDKDFVITSFNIKEGTSFGKY LGFPMFQARPKNADFQFLLDNFKNKLAGWKSRFLTLAG RATLIKSTLNQLPNHIMQYIHIPAHILNKMDQYQRNFLW GTTLEKKKLHLVRWNEITRPKEAGGLGIQKLRQKNHAL LASLAWRLFKTPQAMWANILLTKYLNTSPNNIHSQIWS NVINGWHQCQLGIKWNIGKGTHIKFWEDPWIGPGTTIRS KIHGPLPLDQQNITISSLINNKGWDWNKLPFELPITLLKV AHNITQIIQATYISKYSNKVDNIFWGLTPTGHFTTRSMYS ILEAQHDVSTNSSTKYNWIWKLPVQPKIKGFVWLLHKK RLPTKDYLRHINLIQDATCQFCHQEDETIRHLFLDCTNA KTFWQDLGILDFINQKLHYTNSFNWVHDLITCKDFHFPY QMNPATYLPLSLWNIWLARNNNCYKHTMSTASAKLTT QQAAKFTYLVAKKHTXVKKRQQNLKWKPPKEHFYKL NTDGAHTTSKAGIGGLIRNSKAEWILGFAAAVPXDSSTS AELRALLQGLQLAYSHNLVPLEVEVDSQDIITLLYSNNV SFDNLTTDCRYFLDRLGKPVVLHAYREQNRVADKLAK AGCNFDIHSVVRILEDNPDFAWSVFEQERSAVQGHVLE HFRPHSDNYDTVNRYCNSDAVGSSINIGLYNTLAPTTGA SACNRMQYRHPPSVRNT			

L1									
	L1-4_ME								
		.							
			Notamacropus eugenii						
	TTTTTTCTC ATTAAGT TCCAAAG AGAAGAA AAT			AAATTCC TTTTATG ATACAAA TATGGTW TTGATAC CTAAACT AGGAAGA GCCAAAA CAGA	CTTGGTACT GGCTAAGA AATAGAAT GGTAGATC AGTGGAAT AGGTTAGG TACACAGA AGTCAATG ACYATAGT AATCTACTG TTTGATAAA CCCAAGGA CCCCAGCTT CTGGGATA AGAACTCA CTATTTGAC AAAAACTG CTGGGAAA ACTGGAAA ACAGTATG GCAGAAAC TAGGCATA GACCAATA CCTGACACC ATATACCA AAATAAGG TCCAAATG GGTACATG ATCTAGATA TAAAGGGT GATACTATA AACAAATT AGGGGAGC AAGGAATA GTTTACTTG	ERKLYNFPNEYRCKNFKYQKDYNNISQEYIMXKXDLYQ ECRXGSILGKLXALIISXTKLTKIILSQMQKKAFDKIHPFL LKTLKSIRINGSFLKIISSIYLKPXASIICNGINMPFQDGNK DVHYHHYYSILYKCLQXKKKKLKELEAKKKQNYHXLQ MIYTRILENQLKNYLKLSTKKLLEIINNFSKVAGYKINPH KSSAFLYITNKTQQQEIEREIPFKXTVDNIKYLGVYLPRQ TQELYEHNYKTLFTQIKSDLNNWKNINCSWVGVNIIKM TILPKLIYLFSAIPIKLPKLFYRARKKYQNSSGRTKGQEY QGNKEMLGKVAPPXILNYIIKQQSSKP	NA	NA	NA


TCAGATCTA
TGGAGAAG
GGAAGAAT
TTATGACCA
AACAAGAG
ATAGAGAA
CATTATGAA
ATGCAAAA
TGGATAATT
TTGATTACA
TTAAATTGA
AAAGGTTTT
GCACAAAC
AAAACCAA
TGCAACCA
AGATTAGA
AGGAAAGC
AGAAAACT
GGGAAAGA
ATTTTTACA
GCAAGTGT
CTCTGATAA
AGGCCTCAT
TTCTAAAAT
ATATAGAG
AACTGAGT
CAAATTTAC
AAGAATAT
AAGTCATTC
CCCAATTGA
TAAATGGTC
AAAGGATA
TGAACAGG
CAGTTTTCA
GAGGAAGA
AATTAAAG
CTATCTATA
GTCATAAA
AAAAATGC
TCTAAATCA


CTATTGATT
AGAGAGAT
GCAAATTA
AAACAACT
CTGAGGTA
CCACATCAC
ACCTATCAG
ATTGGCTAA
CATGACAA
AACAGGAA
AATGATAA
ATGTTGGA
GAAGATGT
GGGAAAGT
TGGAACAC
CTAATTCAT
TGTTGGTGG
AGCTGTGA
ACTGATCCA
ACCATTCTG
GAGAGCAA
TTTGGAACT
ATGCCCAA
AGGGCTAT
AAAAATGT
GCATACCCT
TTGATCCAG
CAATACTAC
TTCTAGGTT
GTATCCCAA
AGAGATCA
TAAAAATG
GGAAAAGG
ACCCACAT
GTACAAAA
ATATTTATA
GCAGCTCTT
TTTGTGGTG
GCAAAGAA
CTGGAAATT


GAGGGGAT
GTCCATCAA
TTGGGGAA
TGGCTGAA
CAAATTGTG
GTATATGA
ATGTAATG
GAATATTAT
TGTGCTATA
AGAAATGA
TGAACAGG
TAGACTTCA
GAAAAACC
TGGAAAGA
CTTACATGA
ACTGATGCT
GAGTGAAG
TGAGCAGA
ACCAGGAG
AACATTGTA
CACAGTAA
CAGCMACA
GTGTGTGA
KGACTRTTT
TTGATAGAC
TTAGCTCTT
CTCAGCAAT
GCAAKGAT
CCAAAACA
TTTCCAAAG
GACTCATG
ATGCAAAA
TGCTATCCA
CATCCAGA
GAAAAACT
GTGGAATC
KGAATGCA
GATGAAGC
AGACTTTTT
TTTTTATTC

RTE	RTE-2_ME	.	Notamacropus eugenii	GGTTACT	TGTGTCTTC	MIPLTISTWNVLTLIDNTKSSRPERGTALVARELNRYHIQ				824
				GCAGCAG	TTTTACAAC	IAALSETRLANEGQLTEVGAGYTFFWSGHSEGECREAG				
				GCAAGGA	ATGACTAAT	MGFAIKTNLISKLVCLPKGVNDRLMTMLPLAGKHHATII				
				ATCCCAC	ATGGAAAT	SAYASTMTNPDEVKEKFCEDLETLIINVPKEDKLIILGDF				
				AAAAAAT	ATGTTTTAC	NARVGSNYQTWQGVLGRNGVGNSNSNGHLLLKTCASH				
				ATGGGAGA	ATGATTGCA	DLLITNTVFXLPKCNKTXWTYPRSKHWHLIDYVIVRRRD				
				CACTGGCA	CATGTATAA	RQDVRVTKAMCDAECWTDHRLILSKLNIHIHQKYCPQG				
				CAGGACCA	CCTATATCA	KMTTRRINVNKLELFSESEQFVADLEGKLSQHIVGNSGA				
				CTCAGCATG	AATTGCTTA	EKEWAAAFRDLVYSTAFAHLGQNTHKHQDWFDENDGEI				
					CCATCTTAG	QKLLNEKRELHRIYQQDSSSTSKKAAFNSIKSKVQAKLR				
					GGAGGGGG	EMQDSWLSKKADEIQFYADSNNPKCFYDSLKAIYGPKT				
					GAGGGGAG	YGASQLLSADGATLISDKDMILERWAEHFHSVLNRPSSI				
					GGAGGGAG	NAEAIDHLPQVEVNPSLAELPTEEEVLRAIRLLSCGKAP				
					AAAAATTT	GADSIPAEIYKVGGPLLIQKLTEIFQVIWQEEVIPQEFKD				
					GGAACTCA	ASIVHLYKGKGNRLSCDNHRGISLLVIAGKILARVLLNR				
					AAATTTTAA	LILHLEDGHIPESQCGFRKGRGTVNMVFAARQLQEKCQ				
					ATAAAAA	FSAACANFGLIINTKKTQVLHQPPPHHPYVEPSVTTNGE				
					TGATGGGCT	VLNAVDKFTYLGSVLSRDVHIDNEVDTHIARASSVFGRL			cd01650	542
					GGCCACGTT	RRKVWERRGIRLTTKLKVYRAIVLTSLLYACETWTVYQ				
					GTTCGAATG					
					CAAAATGT					
					ATGCTTGCC					
					AAAAAGAC					
					TATTTTATG					
					GAGAACTC					
					GCATGGGG					
					CAGGTGAT					
					CACATGGT					
					GGTCAGAA					
					GAAACAAT					
					ACAAGGAC					
					ACTCTCAAG					
					GTCTCTCTC					
					AAGAACTTT					
					GGATTTGAC					
					TGTGCAAC					
					ATGGGAGA					
					CACTGGCA					
					CAGGACCA					
					CTCAGCATG					

L2	L2-1_OL	.	Oryzias latipes	A	GTGTGCCCA CATCAGAA AAGGTGCT GTACTCTAT GAGCAAAG CAGAATTG AGATAGCA CAAAGTAA ACACAGGA TGTGCAAAT TTGGAATAT CCACCCCA AATATTCAT ATGGACTAT CTGTGCCCA ACCTGTGGT AGAGCATT CTGAGCTCA TATTGGTCT GATCAGCC ACAGTCGG ACACACTG AAATTTCAC TTTATCATG GTGATGTCA TTTTGGTCC TCTTTGAAG ATGAAGGA CAACAACC AACCTATTA T	RHARKLNXFHLNCLRKILRITWQDKVPDTEVLAQAELP SIQTMLQRVQL	TGAGTTGAT GTTTTTATT GGTTTTATT TTTTTATTG GTTTTATTA TTTTTATTT AATGTATCG ATTTTTATT GATTTTATT	CLPPSLLVSLSSFELELFELGCSPPVLCAVVYRPPKYNKD FLNEFSGLLAEILLKYDRVLIVGDFNIHVCCPDHTMSKD FLSLIDSFNLVQHVSEPTQERGHTLDLVLTYGLSVSNLVI CDAVFSDHMPVLFDITLPCSPCRPRAPARWVRIINPSTAA LFSASFNQTSVIPQSTCSSTDELSSWLHSTCLTVLDAVAP LKSRQPKVKTEPWLNNVTRAARRECRKAERKFKKDKL QVSLQMLKDSWRHYQATVKEAKREHLSNVILLNRHKP RVLFSTIDAALNAPKSVCVQPTLETCETFLKFFVDKVNN VRALISPPGFDPSVVVSCPVTFDQFEPVTLSVLYETVKHL	cd01650	403	669

L2	L2-4_OL	.	Oryzias latipes	AGATTTG	T	CTGTTTTAG	KPSGSPSDAVPPRLFKEVITTLGPAVLAVVNGSLSSGVFP				
				TTTTTATT		TTTGTTGTT	RNCKHATVQPLIKKPNLDPSVLSNFRPISKLPFLSKILEKI				
				GACAATT		TTGTTTTTA	VYVQLKDFLDENGNLEVFQSGFKSLHSTESALLRVFNDI				
				TTAATCTT		TTCTGCGTT	LLATDAGDHVVLVLLDLTAAFDTLDHNILISRLQQLVGI				
				TTCTGGA		GTTAAGCA	RGTALQWFRSYLEDRTFCVNLNNVQSSSAPLLYGVPQG				
				ACAGAAC		CTTTGGTCA	SVLGPLLFSLYLLPLGSILRKHNCAFHCYADDTQIYVPIT				
				CTCATTCT		GCTTCTGCT	KKDAHSISPLIKCLEELKIWMALNFLHLNEEKTEVIVFGP				
				ATAAAAC		GTTTTAAAA	SGNSESHPSHTLVTPESHPIALGSLTQSVKSTVTNLGVK				
				TGATGGG		GCGCTTTAT	MDNGFNLDKQISSVVRSSFFQLRQLAKVKSFLSRKHFET				
				ATTCACC		AAATAAAG	VIHAFVTTRLDYCNSLYFGAPHSSLSRLQLVQNAAARLL				
				CCAGTCC		TTGACTTGA	TGTWKREHITPVLASLHWLPVHFRVRFKILLFVFKSLNG				
				ACTGGGC		CTTGACTTG	LAPPYLSELLHLHSPSRSLRSADQLLLDVPRTRRKLRGD				
				AGCAATG		AC	RAFAVAGPSLWNALPMQVRQARSLNDFKSKLKTHLFSL				
				TTTTAGCT			AFNSV				
				GCMAACA							
				TCAGACA			MRGSAGRKANTNWLRPLKEAPKPEEAASTTTPPPSFNM				
				CTCTGTC			ALFNVRSLSNKTFIINDFIHDYNIDGLFLTETWLSTDAST				
				CAGACAT			VLTEASPPNFSFVFSTREGKKGGGTASIMSNSFAVKQITL				
				CACCGGC			DEYPSFEHHAFVFNSPPVLCITVYRPPKPSSVFIEEFSEFL				
				CCCCATC			ASVHTSFNMILITGDFNLHIDGKSDYCTTQFLDLLSSLDF				
				ATGCCCM			KQHVMQPTHNRGHTLDLVISHGLPTSVSAVVDLAVSDH				
				CACCAGT			YCVLFNITGFIKQETSVRTVKKRYLTPEVVEKFITSVRSY				
				CCAGACC			SPPVLPAPCDLLVESFNSKLKSSLDSVAPLXSKKIRIKAT				
				TCCTCCTC			PPWRTEEIKQLKRRCRSAERLWRKHKLEINYQLYQNQL				
				ATCCGTG			KLYNQTVKQTRNSYFANIISVNKNNPRVLFTTIESLINSN				
				CAGACAT			VNKNTMPASASLCEDFADHFRSKIDSVRSNIIKGQKSLFI				
							GNDGLILPEEALDSFDLVEAATLGRVFSQVNPTTCFLDPI				
							PTSLFKSFYGSFESEILNIMNSSLQTGIFPAAFKTAVVKPL				
							LKKTNLDPSVFNNYRPVSNLPFLSKVLEKLVFIQLNEFLL	IPR000477	521		582
							QNNIFEINQSGFRTNHSTETALLKIVNDXRCA				

					CGGTCTG TGGAAGT GACTGAC TGGATGT GGTATCA CCCCCCC ACATTCC CCAAAAA ACTGCTT CAATCGA CAGTGCT TCCTTGC CACCACC TACACCC CCCCCAG TCTATCC CATTAAG ACCATTA TTACAGA CAGGTCT TCC		TAAGGGAT AAGGGCAA GTGGATGA TGATGATTG TAGA	TYVRSAAGDSDRFGVSVGLHQGSALSPYLFLLVMDALT LNIQEEAPWCMLYADDIVLVGEDECEVQSRLEEWRVKL EGAGLKISRTKTEHLFCDFGGPTSFSPIALNGVPLPTCSD FKYLGSLVQSDGGVDRDVKSRINAGWMKWRQVTGVT CDPRMPLRLKGQIYKSIIRPVVLYGSECWATKVADERRL HVAEMRMLRWMCGVTRLDKVKNVYIRGSLKVAPVVE KLRSKRLAWYGHIMRRNDGDINQKVLRLNVDGYKGRG RPKKVWMDCVKEDMLKKGVDSDMTADRDVWKSSTY CADPP	SSF56672	18	134
RTE	RTE-1_PPo	.	Papiliopolytes		ACCGGAG GACGGGC AGCAGCG TCCCTTTC AAAAACG GTGGAGA AGTCTGC GATCGCC AACCCGC CTGCCAA		TGACAATG ATGATGAT GA	MMATRKVRPPVPGGPAVPGYGSGQGNGGAGGAKNPR LGYGYHTRRPLALATLNTRTLRTDERIAELEEELSRLRW DIIGLSEVRREGEETIILKSGNLFFFREGDQQSQGGVGFIV HKSLVNNVVKIESVSTRVAYLILRISKRYSLKVIQVYAPT STHSDDEVETMYEDISRAMHSSKTFYTVVMGDFNAKLG IRDGDELRMGKFGYGRRNHRGHLLAGFMEKEGLYMM NSFFKKQGHRKWTWMSPDGATKNEIDFIMSTKKQIFND VSVINAVKTGSDHRIVRGTLNIDVKLERCRLMGSTLRPA PVHIHNAETFQLELRNQFECLGSSVDDLNDGFVHALQS VGSKFYKTHRKSSPKKLSDHTLELMKERRDMKLHSSDD	cd01650	549	822
RTE	RTE-4_PPo	.	Papiliopolytes								

				GCGTGGC			AERYRRLNRQISTSIKRDIRRFNTNRVKLAIERNQGSKVF				
				GATTATG			SRDLSIGQSQLLKLKTGEGRVVSSRPELLSEVERFYGQL				
				GCAAATC			YATQEPVTDLVEDPRARLTRHYTEDIPDVSLDEISMALK				
				CCCCCAA			QLKNNKAPGNDGITAELLRAAGEPALKVLQKLFNSVILE				
							GHTPKAWSRSVVVLFFKKGDKALLKNYRPISLLSHVYK				
							LFSRVITNRLARRFDDFQPPEQAGFRKGFSTVDHIHTLR				
							QVIQKTEEYNLPLCLAFVDYEKAFDSVETWAVLQSLQR				
							CRIDYRYIEALRCLYENATMSVRVQSEDTKPIHLQRGVR				
							QGDVISPKLFTAALEDVFKVXKYCHLPETLHRGVGRRF				
							QGPEILVVTYTFLHVTFRVQVIGSLLLLRILIKIGTMLEDL				
							NRASERVGLRMNMDKTKVMSNVHVSPTPITIGGHTLEV				
							VDEYVYLGQTIQLGRSNFEREVNRRIRLGWAIAKXNNR				
							RSGSCGISSRPKTYRSVSRQKSLTSVCCQHTGQRRGHSL				
							RATSGGSKSLKGQWSGQCSESLCEIASGTMRSAAESVT				
							DIARRIANLKWRWAGHIARRTDGRWGRRVLEWRPRIG				
							RRSVGRPPARWTDDLVKVAGVHWMRAAQDRALWQS				
							MGEAYVQQWTTYG				
							KSEHSRQVRSRRNRQVQCGCGPAPGIGXSPYLFLLVIDA				
							LTVEIQEEAPWCMLFADDIVLVGEDALEVQSRLERWRD				
							KLETAGLKISRAKTEQLFCDFGGLSSFIPIALDGAHLPVC				
							SDFRYLGSLIQSDGYIDRDVTNRIKAGWMKWRQVTGVI				
				GTACGTG			CDPQIPLKLKGKIYKAVVRPVMMYGSECWAVKGTDER				
				CGACTCG		TAAGGAGA	RLHVTEMRMLRWMCGVTRMDKIRNEYIRGSLRVAPIV				
				TGAKTGC		TGATGATG	EKVRSKRLAWYGHVMRRDESHVTRRTLVMSVEGYIGR				
			Papilio	TATGTAT		ATGATGAT	GRPRKRWMDCVKGDMIKKGVMTEMTYDRVMWRTKT	SSF5	25		135
RTE	RTE-6_PPo	.	polytes	GA		GTATG	CCADPTGVG	6672			
				GGAAAAG			MMRPRSSSLSLTKPNQIKQLNTSNQKRNNPNPAGHRGV				
				AAGGGAG		TGAAAATG	ARPLPPKSMSSGSKLKSKLYLATLNVRTLMEDHRLIELE				
				ATATTAA		TACATACAT	HAISQIKWDIIGLCEVRRSGDEIREYNKYIFYYKGEKKGR				
				AAAATAA		ACTTTATTA	RGIGFLINKYLKDKIISFEGISDRIAVLNILNGFKHPMSIIQ				
				AAGGAGT		AAAAAAAAA	AYAPTEQDTNAEKDKFYEKLTEVIQKSNKNIIVMGDFN				
				TTGAAGC		ACTTTTTAT	AKTGVKSNKDEFVLGKNSCGERNENGQRLIDLCFASNL				
				AGGTATA		ATTTAACTT	KIANSFYKKEKSSGRGCLQTEGTEMKLISFQFLQKRKVK				
				TGTTACT		TAACTATTT	WTWMSPNGRNRNEIDFILTRETRILSDVSVINKLNFSTD				
				GAAGATT		ATATTGTTA	HRMLRGTLICEDKNSRKHIISNKNGKTITVPHNILEDIEK				
				TTCCGAA		TTAATAAGT	TLQDIKKQQTIQDKYDFLEKTLKEVNIFGTKRGKRNKLN				
				AGAAGTA		TGTAAATAT	EETKKILEERNELLDNRRENRHSIANISKKIQSSIRKHRKE				
				TTAGAAA		TGTAAATTG	ERLKTLKEQIEKTGGMTKGLKQLKEYSQWIPNIKLQKK				
	RTE-19_PPo	.	Papilio polytes	AGCGAAA		TCAGCGAA	KETYRITSRRSSIAKTATEFYKQLYTESGESIVASKNIIEE	cd01650	581		833
RTE				AGAATTA		TAAAGGCTT	GDSNEEEIPTILETEVAGAIRSQKNDKAPGDDKIPNELLK				

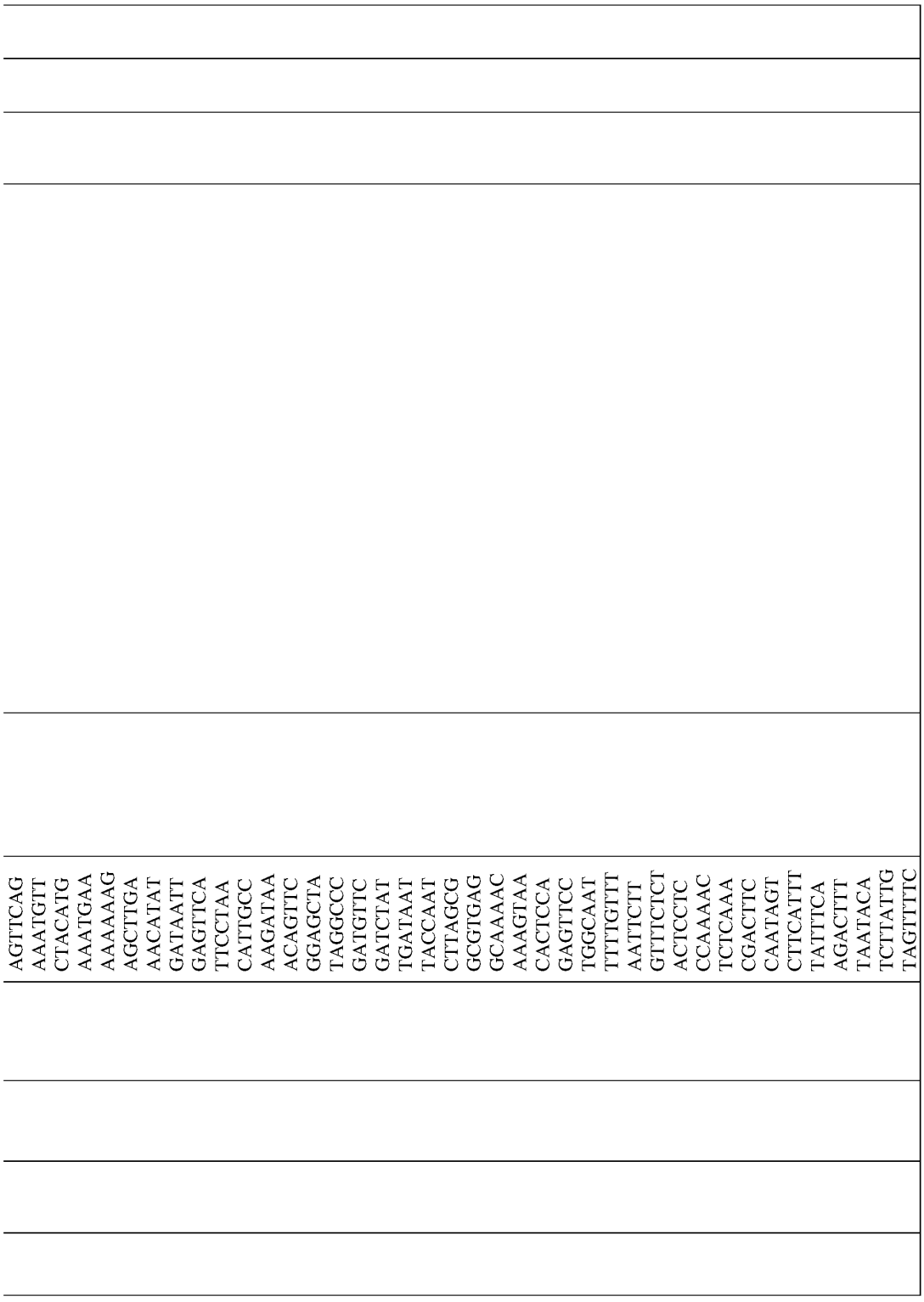
CR1	CR1-2_PH.	Parhyale hawaiensis	CAAATAA AGGTTAA AGAGGAG CGTGCTA AAGGGAA TATAGCG TATATAA AATATGA TAAGATA GTAGTCA AAGAAGG CCATTTA GTCAAAG AAAAAAG GAAAAGA GATCAAT CAACATC ACCAGAA AATCAAA GTTATGC CAAAAAA CTGGTTA GCAATAA TTCATTG ATAAAGG AAAATAG GAAAAAC GCTTTCG AC	TATCTATCT ATCTA	STLSSITKALTILFNEIINSEQIPVQWTKSTITLLHKKGDK QDINNYRPISLMSNIYKVFSKVILGRVTQTLDANQPREQ AGFRADYSTMDHINVVVKQIIEKCNEYGQNYYLAFVDYS KAFDSLSHNYIWKTLASQGIERKYINIIKEIYKNSTAKIQL EQVGEEFKIERGVRQRDPLSPKIFTAVLENIFRNIDWDKF GINGTKLNHLRFADGIVLFAETPETLRTMLQQLDQES RRAGLSMNPMKTKLMTNAEKDIIELRHEIIEYVDEYTYL GQIISVNQHSKEIDRRISNAWKRYWSLKEVVKKKSIPMS IKTKLYNTCILPCLTYGCQTWPSTTKHNLKLANCQRHM ERSMLDIKLRNRWRNVTIREKNKSKEKKTKVNDVFKKI KKLKWNWTGHIIRTNKGKWTKDIMDWYPREGKRQRG RQIKRWEDDLPKGWRRLARDRETWKKLGEAYVEGQPD TQLDAGK						
									cd01650	959	1215

CR1	CR1-4_PH	.	Parhyale hawaiensis	GCTCTGG	TCACCCGCT	SEVHRLLLLDLDPYGGTDPSGMFPFFLRRVAGVLAAPLA	cd01650	514	772	
				GTTTGGG	TTGTTTTCT	AAKADLLLRYFDGKQSRDRVRIPLACDPEPRFSGFAFRS				
				GCTTTGG	TTCATTCTT	ASSVSAHKWWSTLKSAVLGVSSSVPPLIGTGGDLVADP				
				CCGCTTT	ATTCTGTGA	SRVNWDAVSGAILALPWGVVRESTDPGSTLDRLLSDMV				
				CCGGACG	ACAGAGCT	VGVSVVSPIGSSDHGALSATLQVRQRVPRVTVRREIFLK				
				TGAGCTG	TTCTTTTTT	RTSRSAVDFQLFREQRRRAGATYEEAEREHNLRSRELL				
				CCGGCCA	GATCTGTAC	HAAFDFASASGCEQLVEGPTHLMGGVLDLLMTDVSDL				
				CCATTGT	GCTTTAAAC	AMARIQSSDRRSAFAFMGDFNCHHVEWLGSGRTDCHG				
				GCCGCCC	GTATTGGCA	PCCEMMVLRVCGSFRNFYLFSVYRSPSADDRVFDCLLD				
				TTATCAG	TGGTGCGG	IPNFEKPVLRLRDSCPGFRGLAAYVVRKGVVVSRQSEYEC				
				CTGCGCG	TGTCTTTGC	VRGLYGNLAELAVASSNYDLVLCSETLVSSHRHVSELM				
				GCGTGGC	TGATGGCTC	MVYVYLLFFACLITRYPDVEENPGPRGFGSRSCRILCTN				
				TTTTTCTCG		QLKTQTAIHHAKQNEHQELKSNKDEILSKVHVINIRGLL				
				AAAGCAA		TKNCRKQMVLQDIINDAGSKLLAVTETHLDDAVNDAE				
				TAAACTA		VKIENYVLHRSDRKDRKCGGVALYVVRDHLGSHTKQKF				
				ATTAAGT		SNGVCDCIIVNIEKLRLMCVVCYRPPKCTTEKFAEMLQV				
				GTTAAGC		INTHCKEVQSKYQLYITGDFNFPFLKWLTTDLPDETGSF				
				AGTGTGC		VLSRTSSATRDEKDQALSLVRIMEEFSLTQMISSPTRNEN				
				CAGCCTG		VLDLLFVDNAQLIKNIEIQDTTLSDHRLVEITTQLGERDS				
				GCTATAA		HRPQKNSQINSISSLRDLNVHDKGINWDGLNAELQSVN				
				ACTCGTA		WDVLLNNSIESESINETFENILLEKCMKYLSPAKRKFNKS				
				CGGACCG		RHHKFRKGLLKKRAKLIKCQQSTKLLQRKSECKKKLDE				
				TGTGACC		IERQLARSHDEEKQEKEKKAIEKIKSNSKYFFAYASHHS				
				AGCGGGA		SVKSKIGPLVNDTGSIVXDDQEMSEILSRQYQKVFSDKIP				
				CCAGTGA		NPEIPRIDLTTNYLTDFEFTKLSIELAIDELSGNSAAGPDD				
				CCACACC		VPAKILKMCKLSISYPLYLMWRKSMDDGVVPLTYKHAI				
				GTAGCTC		ISPIPKAGNSMDPANYRPVSLTSHVVKTFERVFRKHLVH				
				GCTTAAT		HLESNNLLTSNQHGFRHKRSTTSQLLEHYDSILDLLCQG				
				AGGGGAA		KSVDVIYLDFAKAFDKVDHGILLLKLFNMGIRGKCLRW				
				GCGTAGA		IEAFLTNRTQSVIINNKVSKRQPVLSGVPQGTVLGPILFLI				
				CGTCAGT		MINDMGHQCKFAVPSSFADDTRILAAVQSTADVRNVQL				
				T		DLTNIFHWTSVNNMTFNESKFELLRYGKNFSLTESTSYT				
						TRDGSEIERKQILRDLGVRIQDNATFEDHVARVVTVGK				
						KWSGWVLRTFSSREADTMLTLWKQLILPRIEYCSPLWM				
						PSKRKDLERLEAIQRSFTAKIRGMKDLNYHERLAQLQM				
						YSIQRRFERFTIIYVWKILEGITVNVNGRIQARHSNRIGRT				
						CEVRIPIKTGGDAIATKYSNSFVVRGARLFNKMPRHIRN				
						VTGTETAKFKHLLLDACLLKVPDKPAVSGYVNCGDNSLL				
						ATPLVTDCSITSAQDGWRA				

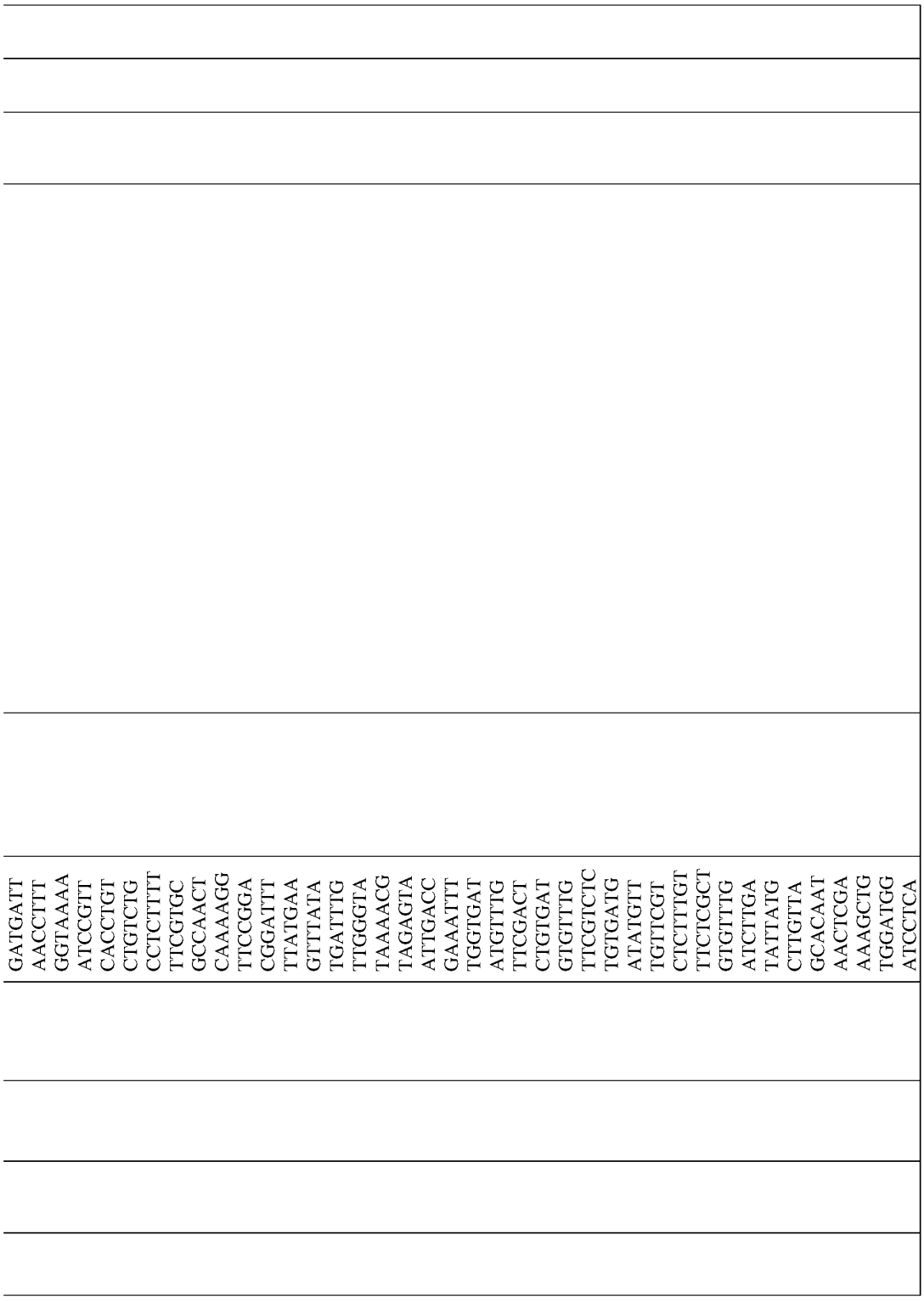
				GTAGCCC CTGAGTT GGAGGAA ATGCCAC TCTAAAG GCGGATG AGACCTT TTTGGTC CAACGGC ATCCCTG AGAGGAG ATGAACA GGCTGGT GCCATAG CTGGGTA CTCAGTC TTCCGCA GTCTGCT CCTCTCTT CCTTCCCT CGGGTAG TACCACC TGATGCT CTTTTCCT TTCCTTTC TCTTCTTC CTCTTTTC GGCCAAA GCCCGTA CTGGTGA TGGCAGG GTCTGGG GCGCGTG GATAGCT GGCCGCT TTCGTTG CTATCAC GACTGTC GATTTGC TTTTTTGT	GGGCTGGTT GCGTGGGG GTTTGTTAA CCTCATGGA GTTTCTCCG TTCCCACGT CGCCAGTCT GGTGTGATT TAATAATA ATAA	SVFRRMLRLGSFPECWRVAHVTPVPKGSVSSRTSDYRPI SITPVLSKVYERLVASRLGCFLEKQGVLPPTQYGFRKGL GTCDALLHISHVLQLALDGGHEARLVQLDFSAAFDRVN HLGLLHKLRLMGIGGSVLAVISQFLSDRRQCVVVDGKR SEFLDLVSGVPQGSVLGPLLFIVYTADLFHLVNNLLVGY ADDVTLLAVCDKPSDRLEVSESIQHDLEMIDSWCSRWG MRLNAGKTKTMLVSRSRTVGPPFPALMLDGVSLPVVD DLVILGVNFDPKLTFEPHIRSVASRASQKIGLMRRAWRV FGSVELTRRCFNCFVLPLLEYCSVVWGSAADCHLCLLD KVVRRCAFMLGGVVPCDLAHRRRVGGVCFLYKIYNCP SHPLSEFLPDPFVRVRYTRRTGALHEFALEPLRCRTGQY SRSFFSIVCGCLEWS			

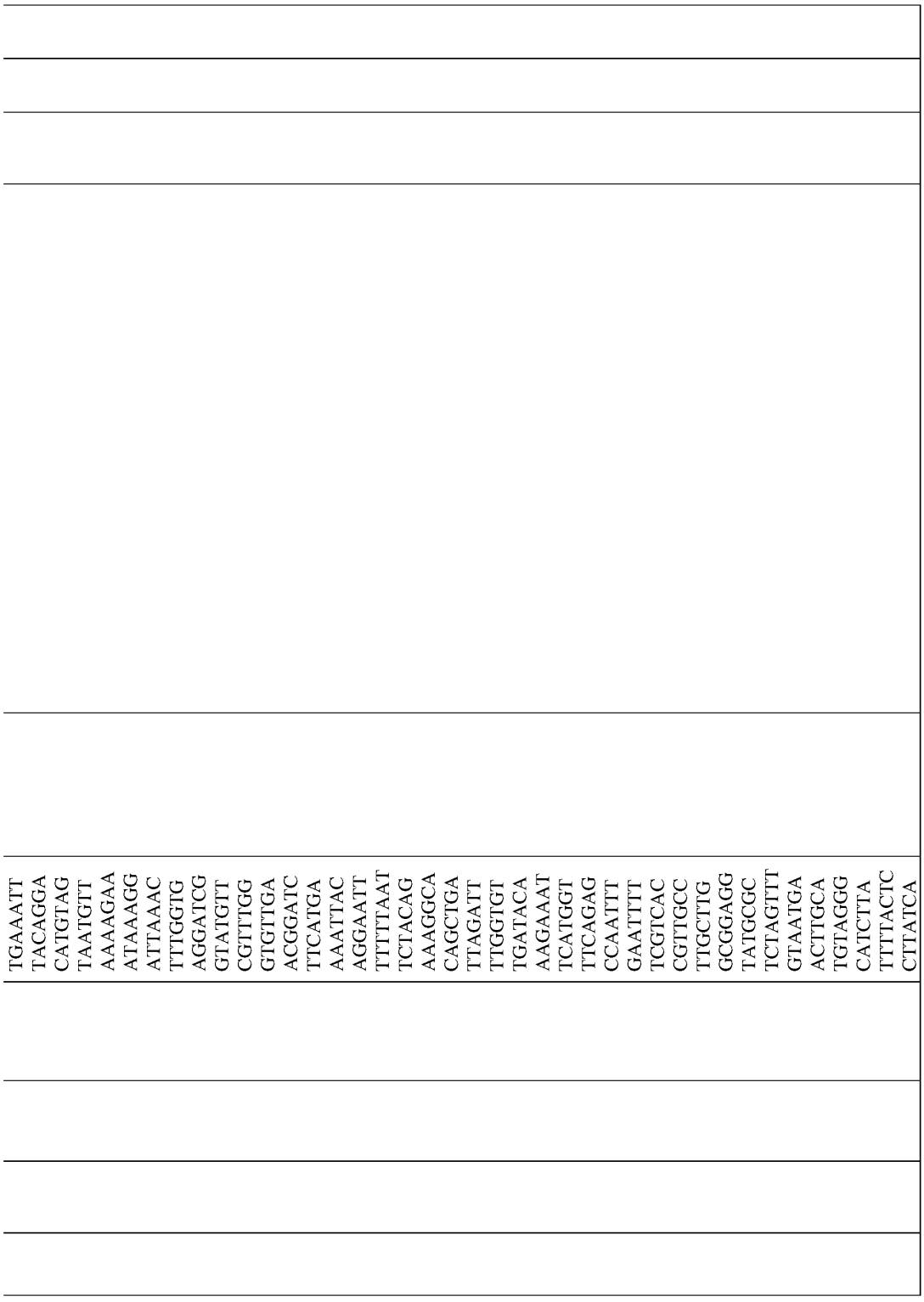
			AATTTTTT TGATGAT TTCCTTTG CTT						
CR1	CR1-3_PH	.	CAGAAAA TAATCAA CATTTGA CTGTGGT GTATATA CAAATAC AAGTGTA GTGCTAA AATGTGT TCGATTA WACTCTT ATTCATT GCTCTCG AGTGCTA CCGCTAG ATATTTTC TCCATTG CTTTTGGT TCTTTTGG TTGGATT CTGTGGT CGAACTT AGAAGAC GCTGATC CTGAGTT ACTAAGT AATATTA TATTCTAT CAAATTT TCCTCTAT CAATATT TTTTCCAT CTTCCTCC TCATTGA CTTGCCT GATCCTC	Parhyale hawaiensis	TAAATGGCA ATAGCTCAT AGAGGGCC GGGGCGTA AGTTCCGGT GTCTCTTTG GGAATGGG CCTATTGGT TCGTCTACT GCCGCCTAT CTAATCTAA T	MKQCDECHSVFPLEFFYYEEVELESCKYCCKIQALQAEI EKLESDREAMQRGDGGEKEKAPREAHLRLSSPSPWKTV MRNKRQRRSTGAILEQTNHIKVNNRFAAFNDGDDNVA TSVRPKIKLKTVNGEQHRGPPQSRDQKKSETSVQPPSHE RERNPAPIPGQRAIKTFLIGDQIIKGQYYAFKRGCNDRKV ICRPQASTSDATAILKDLRCKKMDHVIIHVGTNDLRTYS ETQLQLVKDLIAAAKKKSDHVTLTSILPRLRAGNRLYSR IRYINGQIKREAEXNGLEFIDLFPSFDDDWRLFLPDGVNL NGAGHHVLGEMLLAAAASETAALEEGLRTRPKTIIEQPR HPAPFHALPADLPAMEQTATNVPAAVPCNPQISSPSGM KVGDPKGNEGTIITNSLDTDEEITPTPDASVQMSRLGPP PLLPSIVSNLVSTNSNMSEGTDVPQPTNARQTDLDSSNT ERGSPVSVSKEKSQSQTPVAKKRGQFPRTSSIPAHEGNR STRKSGARNIPPKTNSCQPPQTRITRSKANSAGNDLLSAR LGGKRPIVRQIAADVKQSDNCQAEIQNISLAXLQTGALA ATDNAFADLTPPIFSQQSDAPSVSVTRRSNANTSVKFLF GNVQSITNKMNELQAIAATLAYDVIALNETWINLSNKN LEAEVALKGYKLFNIDRPTETGRGGGSIIYIKEHLNPLCK GFAATKEYEFITTHARLGIRQSLNIMLAYRSPRSTAETDA ALYNKIENALNTPNETILMGDFNLPQIDWSTLCSPAPGS KLLELLSHHHLTQHVNEPTRYGNILDLVFSTPDCEIVDL AILDRLGDHNMIEFTLKTGNEFTTNASRLKLPDYRRANF QGMATEIETTDFETPMRDASVEQCSHILKTSILRAAERYI PNKRVRISNDPLWMNRTVVKRCIAGDKCVTKPGSVIQQR RTGPVTSPLDESDSQTLHCRRQVRYKAWKRNPTEENRS SHIRACRDVKKIVRTTKREKEVSVAAAMAKRNPKAFYKY VNERRLKKETIGPLKRPDGXITNDLRDKATILNDYFTSV FTVENVTNIPEPAPMQVAPINTCVFTEIEILQELTHLNVY KSTGPDNLTPRVLREMAPVIVKPLLEIFNRSLRTGEVPQ DWKLANVIPIHKKGDKKSPSNYRPISLTSVVSKTIERLLK KRITDHLRMHQLISDTQHGFRNGRSCLTNLLDFFSKVTD IKDEEKGVDLVYLDFQKAFDKVPHQRLMKKLEAHGIQ GDYLRWIRAWLTGRQQRVVKLEGETSDWTQVSSGVPQG SVLGPLLFSIYINDLDLNIRSKISKFADDTKLCHPATKRA ELQQDLDHLMEWSTTWQMSFNVDKCSVMHVGLVNPS TPLKMNGTELAAVTQQRDLGITITNKLQHSQQVDKSVK	cd01650	113 6	138 3

				GCCAAAG GGACAAG TATGACT CCTGAAC TTCCCGA GTGGCTA TAAGTTT GCCTTAC TATAGGT GGGATTA AAAATTG TGTGCTT CGGTCTC AACTTTTT TTTAGTTT TCAAGTA GATAACA CCTTTTCA AGAACGC TAATATT TTTTTTTT CAAACCT TTCTAGA GACTAGG AGGTACA CTAGCGA ATATTTA TTCTCAA ACTCCTT ACGTTTT GGATCTT CATAAAA TATTATTT TTTATTTC TTTTTTTT CTCTCATT CATTTTCT TTTCTAA CAGACTG TAGGTTT		MASRVLGFITRNFEYKSKDIIVPLYKALVRPHLEFAVQF WSPHLCKDIDNMERVQRRATKLIPQLRNKPYECRLKEL GLISLEQRRLRGQLIETFKYLNGFNDVPTTGLFERDLND RLRNNGQKLKHKARRTTLSQQFFPYFIVSIWNRLPSDVV SAESVNAFKNRLDIHWERHPPQLRTPPPGVASSELLLQE SQALNGQSFRGHLNPSSSLSRSNYTSLSSAYQRLLIVLAC IHDRLFGWDSSPPSFSFFFLFLPFFSSLSPVLYILLT			


AGTTCAG
AAATGTT
CTACATG
AAATGAA
AAAAAAG
AGCTTGA
AACATAT
GATAATT
GAGTTCA
TTCCTAA
CATTGCC
AAGATAA
ACAGTTC
GGAGCTA
TAGGCCC
GATGTTC
GATCTAT
TGATAAT
TACCAAT
CTTAGCG
GCGTGAG
GCAAAAC
AAAGTAA
CACTCCA
GAGTTCC
TGGCAAT
TTTTGTTT
AATTCTT
GTTTCTCT
ACTCCTC
CCAAAAC
TCTCAAA
CGACTTC
CAATAGT
CTTCATTT
TATTTCA
AGACTTT
TAATACA
TCTTATTG
TAGTTTTC

TTTTATTT
TGACATG
TTAAAAT
GCCTAAW
TCGTTGC
TGCAATG
TACGTAA
AAATTAG
AACTTGA
TCTTTTTT
TGTTCAT
CTTTCTA
GGAACAT
TTGTGTA
CGGTGCA
ACGGGTT
ATATGTT
ATATATT
ATATACA
GTTCCTTT
GTTTACC
TCGTCAA
GGAGGTT
ATGTTTTT
CTCGTGT
TTGTCTGT
CCCTTGT
CTGCCTC
TTGTAAG
TCACCAA
GATTACT
CTAAAAC
TAATGAA
CGGACTC
TTATGAA
ATTTACA
GGGATTG
TTAGACA
TGTTGAC
AGTAACA


GATGATT
AACCTTT
GGTAAAA
ATCCGTT
CACCTGT
CTGTCTG
CCTCTTTT
TTCGTGC
GCCAACT
CAAAAGG
TTCCGGA
CGGATTT
TTATGAA
GTTTATA
TGATTTG
TTGGGTA
TAAAACG
TAGAGTA
ATTGACC
GAAATTT
TGGTGAT
ATGTTTG
TTCGACT
CTGTGAT
GTGTTTG
TTCGTCTC
TGTGATG
ATATGTT
TGTTCGT
CTCTTTGT
TTCTCGCT
GTGTTTG
ATCTTGA
TATTATG
CTTGTTA
GCACAAT
AACTCGA
AAAGCTG
TGGATGG
ATCCTCA


TGAAATT
TACAGGA
CATGTAG
TAATGTT
AAAAGAA
ATAAAGG
ATTAAAC
TTTGGTG
AGGATCG
GTATGTT
CGTTTGG
GTGTTGA
ACGGATC
TTCATGA
AAATTAC
AGGAATT
TTTTTAAT
TCTACAG
AAAGGCA
CAGCTGA
TTAGATT
TTGGTGT
TGATACA
AAGAAAT
TCATGGT
TTCAGAG
CCAATTT
GAATTTT
TCGTCAC
CGTTGCC
TTGCTTG
GCGGAGG
TATGCGC
TCTAGTTT
GTAATGA
ACTTGCA
TGTAGGG
CATCTTA
TTTTACTC
CTTATCA

				AATTATG TCCTTTCT TGCATGT AGGTTTC TTTTCCTT TAATACA TTGTAAC TTACTAT AGCTATA TCTAGGT AGCCTTA TTTGCATT TTTTACTT GTTTCAT AGGAATT GCTTGGG GGCAGGA TAACACG TTAAGAT ATTTTCTT TGTTTTTC ATCCTTG TGTTTGT AATTTAA CTGCCAA GTTCTTAT GTACACT GTTTAAT CAGTTAA GCATATA CAGTATT TGCGATG TACAGTA TACGGAG GTGCTAC TTTGCAT GAGGAAA CTTATTA ACAAAGT TCATTAA					

Vingi_1	Vingi-1_PH	.	Parhyale hawaiensis	TCTAGTT CAAACTG TTCAGGC GATTTCG CTCGAAT CTTGGGC CATAGCT TTAGTAA CGGCCCT ATCGATG GGGCACG CGTGACG TTTTTGCA TTGATGC CCCTGTC CCGAATT CCCGGTA CCGAGTA ACTGTTC CATTTCT GTTTTTCT CCCTTTCC TCTTCTCC TTTCATGT CCTCTTGT GCCGGCA GCACAGG CCTACAA T	CCACCCG ATGCCTG TGCGTCT TCGTCGT GGGTGGA CCCTGCG GGAGCTC TCTGCGC GGCTGCG ACGGGGA CTTCCGC	TAACTACA ACAACAAC A	MPCSGLFSAASYHPDWRCPSCAPQPPRPPSPPPASPLSPH PTNAPPLIVMQFNCNGLLHCRSELSAFLHSSNIAIAALQE TKLTPSSSLPVFPGYTPVRKDRPAANGGGGLLFLIRQDIP YQQATTNHLFPNDHTVEHSGVTVTLCSSSLTLINIYVPP ATSCPPGYRLDSSALFADSSADTLIMGDFNAHNTDWYS VTQDARAAARGEEISASINNGPYCLLNGDSPTRVPRDGP GSSPDLTVISAHLALDYSWSTTTTLSSDHLPILITAESSDS APSGPRRTFVNFRKANWEEFTRECEELITSLPLPSSCSGG LRALNRALLSAAKHHIPAGHRKEYIPRLTPDARDAISRR DHLRSTSPHDPSLPQLQQEVDRLICGAAQEAWKQTIQD CDPKSNPSKFWSLLRSLSGKQSSVPRNQPISFGTHTFTHN	cd01650	536	795

				GGGCCAA ACGCAGG CACTCAT CGTTGGA GTGGCTA CGCGGGT CGCTGTG GCGCTCC CTTTCTTC CTCTCTTA CTCTCTTC TCCCGCT ATTCTTTC TCGCTGT CTCGTGT TTTTTTCC TTCACGG ACAGGTC CTTTCTGC CGACGAC CTGCCCG CTCTTGTC TAGGCAG CCTTTTCC TTACTTGT TTCCCCCT CCCGCTG TCCCTTTT CTCCCCC TATCCAT TCTACTA AGCGGTG ATCATCC ACCCCAA CCCCGGC CCCGCCT GTCCCTG CTCAGTC TGCTCGC GGAACGT		GDIARGFCKQFTATSPPALATPPRLARRLLRKRHPIAHGF SPFTPAQVREAIVASGSSTAVGPDQLSILHLRHLGPQGII YLTRLYNLSLAHADIPALWKLAVIIAILKPSKPADKGSSY RPISLLCPAAKVLERLLLPFINEAYQPAPHQHGFRPMHST TSALLPLAFQATRGMNKPKPHDRTVALLIDFSKAFDTVP TEGLLSSVSGTTLHHNVVRWLAAYLRGRAACCLYNGS RSQFRAVKAGVPQGSVVSPSLFNLYVSDFPIPRDQLCYP SVTSYADDFTVVASHRDPLTAAAALSDHSASVSNWAA NKGLTISTSKSNSTLWTTDTHQSTLDPGITMNDSALPLE RRPKILGVTFDTHFTFTPHITSTAARASSRLRLLKALAGT RWGQCKETLLLTYRAFVRPLITYAAPVWYPGARQSAVR KLQTVQNAALRTATGCLRMASVDHLHAESKXLRIDQSL DLQCRQYLVSALRPEHPAHSVVTAQAGPRLIRHNLATK FQPDINHLLTDNIMPSPEYRSAISSLHRSAVQASLDRLPV NRVLGFAPPEISPDEAVLSRAARSTLSQLRSGFCSALMS YRHRIGLADSDSCPECHRAPHTVNHLFECPAHPTALLPT SLWDSPLSVTDFLCSLPSFSYLPPHARPPPEPPPT			

				CACCCGC CGGGGGT GGTCTTTT CTGTGTT CCGCCTG CGCGCGC TGGGTCC AC						
CR1	CR1-2_PSi.	Pelodiscus sinensis	GGGTGGT GGTGGTC GGGGACT CCCTCCT AAGAGGG ACTGAGC CATCCCT CTGCCGT CCGGACC TGGAATC TCGAGAG GTGTGCT GCTTACC GGGAGCT CGCATTC AGGATGT CACTGAG AGGCTTA CCAAGCT GATCAAA CCTTCGG ATCGCTA CCCCTTC CTGCTTCT CCATGTG GGAACTA ATGATAC GGCCAAG AATGACC TTGAGCG GGTTACT GCGGATT	TAGATGGG ATTGGTCCT GCCTAGAG CGGAGGGC TGGACTTGA CGACCTTCT GAGGTCTCT TCCAGTTCT ATGATTCTA TGATAAGG GGATGGTT GGATGAAA TAACATGAT CTTGGTAAC TAATTGACC ATTCATTTC CAGTTGGA AATAGGTC AATGGAGG GATGATAG GAGTTGCTA TAGGGAAC TTTCTGGGT GTCTGGCTG GTGAGTCTT GCCCACAT GCTCAGGG GGAAGGAA TTTTCCTCC	MHRARSEKGGPQFRMEKIVQSTGYLKCLYTNARSLGN KQEELEALAQAKKYDLIGITETWWDDSHDWSAVMEGY RLFRKNRQGRKGGGVALYVREHYDCSELQYKEGEKLV ESLWVRFKGANSSDVVVGVCYRPPNQVDEVDEAFFEQ LSEASRSQALVLVGDFNHPDICWETNTAVHRQSRKFLE NVGDNFLVQVLKDPTRGRAQLDLLLLTNREELIGDVEVG DNLGSSDHEMVDFRILTKGRKESSKIHTLDFKKADFDSL RDLMERIPWDVVNMKGKGVQDSWQYFKEALLKAQKETI PTRSKRGKPGRRPDWLTGEILGVLKHKKEAYRKWKLG QMTREELKCIVRECRGVIRKAKAQMELRLAKDVKDNK KGFYRHVNKKKVIREGVQSLMDEGGNLVTEDVGKAEV LNAFFASVFTDKVGSWTSVLSDARWEEDGQPLVGKEQ VRNYLEKLNVHKSMGPDLMHPRVLKELANVIEEPLAIIF EKSWRSGEIPDDWKKANVVPIFKKGKKDDPGNYRPVSL TSVPGKIMEGIIKESILRHLNERKVIRNSQHGFTKGKSCL TNLISFYDEVTGSVDMGKSVDVIYLDFSKAFDTVSHNIL ASKLRECGLDKWTVRWIERWLEGRAQRVVINGSMSGW RSVSSGVPQGSVLGPVLFNIFINDLDEGIDCTLSKFADDT KLGGEVDTLKGRDRVQSDLDKLEDWATRNLMRFNKD KCRVLHLGRKNPKHSYRLGTNQLSSSSAEKDLGVTVDE KLDMSQQCALVAKKANGILGCIKRSIASRSRDVIIPLYSA LVRPHLEYCVQFWAPHYKKDADALERVQRRATKMIRG LEHMTYEERLRDLGLFSLQKRRVRGDLIAAFNFLKGGS KEDGERLFSVVTDGRTRSNGLKLWWERSRLDIRKNYFT RRVVKHWNGLPREVVESPSLEVFKSRLDKALAGLI	cd01650	502	753

				ATGTAGC GCTGGGA AGAAGGA TCCAAGA ATTCGGA GCGCAAG TGGTGTT CTCGTCC ATCCTCC CTGTTGA AGGGAAA GGACTGG GCGGGGA TCGTCGA ATTGAGG ACGTAAA TGCGTGG TTGCGCA GGTGGTG TCGCAGA GAGGGCT TTGGTTTC TTCGACC ATGGGAC TCTGTTCC GGGCGCA AGGATTG TTAGGAA GAGATGG GATCCAC CTAACGA AGAGAGG AAGGAGC ATCTTCG CGGGCAG GCTTGCA AACCTAG TGAGGAG GGCTTTA AACTAGG	AGGGTAGA TTGGCAGA GGCCCTGG AGGTTTTC GCCTTCCTC TGTAGCATT GGGCACAG ATCACAGCT GAAGGACT CTCTGCATG TTGGGGCCT TCAAAGTAT TTGAAGGCT TCAATATCT GAGACATA GGTGAGAG GATTATTCT AAGAGGGG TGGGCGAG ATTCTGTGG CCTGCACTG TGCAGGGG GTCAGACT AGATGATC ATAATGGTC CCTTCTGAC CTTAAAGTC TATGAGTCT ATGAGTCTA TGA				

				TTCGTCG					
				GGGGACG					
				GTGACCA					
				AAACCCT					
				GAGGGGA					
				GTGGGAA					
				AGTCAGA					
				TACCGGG					
				AAGAA					
RTE	RTE-14_Psi	.	Pelodiscus sinensis	CTGGTGC	TGATCCTTC GG	MIPGLTEDLRSISNPRKTAVINNELRRLRMDIVALQETRL	cd01650	525	802
				CAAGCTG		SSSGCLREKDFSFFWHGKPPEETREHGVGFAIRNNLVGS			
				TACTGGC		VIPPNVGTERMLKIQLQTSVGMANIFSVYAPTLTSAQED			
				TTTCGCCT		KDKFYDELSIAISEVSSHEPLFILGDFNARVGSDHHSWPS			
				TTCCCTTG		CLGQFGTGKMNENGQRLLELCSQHDLCVTNTFFNVKPQ			
				GGTTACA		HKVSWRHPRSKHWHQLDLVLTKRSRLGSVRVTRSYHS			
				TCAGTGG		ADCDTDHLLVCSKVYIQPRRFFRSRKEGRPRIDTGKTRD			
				CGTGGAG		PSKVCEFVQALEEALPRRDDTDISKRWEQLRDAIYDTAI			
				AGGGGGA		SIFGKKTVKSADWFEAHTEVLTPLIEKKRQALAAYRSSP			
				TCCTGCT		SEENLQALRFARSQVQQVARRCANDYWLQLCSKIQQA			
				GCTGGGC		ADTGNMRGMYDGIKQVIGPSQRKTAPLKSATGEVLKD			
				ATCAGCC		RTKQMERWVEYYSQLYARETVVTERALNAIEPLPTMTE			
				GGTTCTC		LDAEPTMEELSDALKALSSGKAPGKDGIPSEILKCANGT			
				AACTTAC		LVSELYGILCQCWREGSVPQDMRDANIVTVYKNKGDK			
				TTGCCCA		SDCNNYRGISLLSTVGKLFARVVLKRLHVLAERVYPES			
				GGCCTAC		QCGFRAERSTIDMVFSVRQLQEKCREQNKPLYIAFIDLT			
				GCCCTGG		KAFDLVSRKGLFAILAKIGCPPTLLSIVRAFHDGMRGTIV			
				AGAGGAC		YDGSTSEPFEILSGVKQGCVLAPTLFGIFFAVLLKYAFGT			
				ACTCCGG		ATEGIFLRTRTDGNLFKLSRLRAKTRVLTKRLREFLFAD			
				CATCGCT		DAAIAAHTLQELQSLITRFADACMEFGLTISLKKTQVMG			
				TTCTGAG		QSVGNEPFISVLDYELEVVHEFVYLGSTISDNLSFDSEIN			
				CAGTGAC		KRIGKAATTFSRLTKRVWANNKLTVHTKVQVYTACVL			
				ACAACAC		STLLYASETWTLCSKQEQRLNTFHMRCLRRILGISWQD			
				TTAAGCC		KIPNSAVLERAGTVSMFTLLKQRRMRWLGHVSRMLDG			
				ACTGCAC		RIPKDLLFGELASGKRPKGRPKLRYKDTCKRDLSALSIS			
				TTGATTG		ANTWHSLALDKYAWKQGVKEALAMYETTLVKEAEDK			
				GCGTAGA		RALRKIRARDTRATSAYCCSQCGKDCHSRIGLHSHRRR			
				GTGAGGC		CSNQSNRGANP			

				GCCAGGGGTTGCTTCCGACGGTGGGAGAGGTCTTCGGATCTCATTGGGCAGCTACCGCCCGCCTCAAGCTGGGCAGCCCCCAGCCAAATAGGTGCTGCCTCGCCACGGTCCACTTGCTCCATTGGGTGTGTGGGGCTTAGGGAATCACCAACAGGTGGATCGTTAACCTTACACCAGGCAAAACAAACAAGCGACAAAGTAATCCAGCCTTCAGGCTGGGCACATGGAATGTAAGGACC					
L2	L2-1_PSi	.	Pelodiscus sinensis	GGTTTAGACAGTTATGKGCTATTCCTKTTCCGGGTC	TGAGTCTGTGTGTGTGTGTTCCTCCTCTCGAGATGTTTTTAGCAG	MLLVYRPPRCTADALLELADVVSDVALRTPRLLVLGDFNIHAEAVLSGPARDFMDTMTTLGLSQEITAPTHVAGHTLDLVFWTKGDRVLGIRDVATLPLSWTDHYLVKFRLSADLPLHRGGGPIKMIRPQRLMDPEGFLRSLRDCSAIMANGPVDALVDTWNTEVARAVDMIAPKRPLRASRAKFAPWFS	cd01650	379	640

				CTGTCCC CAGCCTC AAWAACT CTGGGAA CCTTAAC GCCTGKC CATCAGC CCTGAGC GTGCTGT TGCTGAA TGCCAGG TCAGTAT CCCAGAA GACAGCT GTCATCC ACGATTT AATTCTG AATGAAG GGGCTGA CCTGGCT TGTATCA CTGAGAC CTGGCTG AGAGTGG AGGGAGG CGTTGAC CTTTCCC AATTGTG CCCGGCT GGGTTCT CAGTACA GCATCAG CTGTGGC TTAATGG TCGGGGA GGTGGGG TGGCAAT TGTCTAT CGAGATA CCATCTC	GTTTTATGT CCCTCTGGG TCTCTTTTG GGGTCTCGT ATTTTAAAT TTTTATATA TATTTGTAT TGTGTTTTT AAAAATTTT GTAAGCCG CCTCGAATA TTTTAAATA GAGAGGCG GGGTAAAA ATATTTTAA ATAAA	GELMAAKRSVRRLERRWRRTRAESDRLRVRAELRAYS VAVLTAKKSYFSTRIASAQCRPAELFRVVRGLLQSGTRD EMEEPLAPRCDLFAKHFADKIARIRLDLDSGCCVSGAQN VLSAPPSLFSLDQFQLLQTEDVDKVLGGVRPTTCSLDPC PSWLIKSSREGMTGWVQGIINASFRSGMVPAALKEAVIK PLLKKPSLDPRDLNNYRPVANIPFLGKVLERAVALQLQA FLEETDYLDPFQSGFRPGFGTETALVALVDDLRREIDRG SASLLILLDLSAAFDTIDHGILLERLAELGIGGTVLQWFR SYLSGRFQKVELGGCYSTPWRLCYGVPQGSILSPMLFNI YMKPLGEVIRSFGGRCHQYADDTQLYFPVTSEPGVAAT DLNRCLEAVMDWMKSNKLKLNPDKTEVLWIRRPSVRE CGVVPVLDGVTLPLKEQVRSLGVLLDPFLTFENQISSVA RSAFGQLRLIRQLRPFLNKHDLATVIHALVTSRLDYCNA LYVGLPLKSLRKLQLVQNAAARVLTNTRYTEHIMPVLR QLHWLPVCFRAQFKVLVMTYKALNGLGPGYLKDRLLP YEPARGLRSAGEALLRVPPLMEIRLTSTRSRAFSAFAPRL WNSLPRDIRMAPTLALFRRQAKATLFTRAFIGV			

											548
											464
										cd00304	
Penelope	Poseidon-2D_PM	.	Petromyzon marinus	ACTGGCT AGGGTCT TGGTCCA ACAAAGT GTCAGAT TCGAGTG TCTGTAC TTGACGT TGGGTGG CCGGGAT AGGTTGG GT	TAGCTGCCT TATAATTGT ACAGCCATT TCCATTTCT TCCAGTGG GTTGTACTC GCTCTCGTG CAGTGCAC CACCGCTTC CCCCGAAC CCCCTGCAG CATAACAA AAGGCGGC TTAACTTAG CCCCCCCCC CCCACCCAC GTCCCTTCC TGGCTCCTC CCCCCCCCC CTTAACCTG TATATAATC TCGGGTGTC TCAGAGCTC TTCACTTGA TACAGCTAT TTTGCTGTG ACTGTTCCA CCTGAAGA	YLRDFTEKPKSMNPCSHESFKSRINALYGEXFHLSVRAF NRQRDKKSQLIETLTFLKRCRDNKVIPIAFTLTSAAVSSP NSKRILFNASLSLISERIRVLRFELDSLDKTLLQSHLMLT QKMSPSDWNHVDNMTAASADSLSAKCKANLIAKFNLL SKKKPEPHITKLDVNKIVINQSNIPISDNLKTVLAKGLNF AVTPKHIPVEDIISQVDVALKGIGHIQADVARAEIATILK NAKKPISNISKNERSELANLKKRDDIVILSADKGNATVI MDKNSYNNKIQDLLNDGSYSKIKSDPTNSLTIKTNKIITS STSLIEPTKRPSLKPSAARPPRFYGLPKIHKEGTPLRPIVS TINSPTYKLARFLSDTLLPFTGKTTSAVLNSTQLVSMVK DIVLSPDDILVSFDVVSLFTKVPVPDTITIIEDLVQHGLNP DVPALVRHCLTNTYFTWNGTFYKQNEGTPMGSPLSPVI ANVFMEKFESDAIQTSTLRPSLWKRYVDDTFVIWPHGL PALNHFLDHINSRHPSIKFTMEVEKDGKLPFLDVLFERR TDGSLGHSVYRKSTHTDSYLKPDSHHHPAQKNSLIKTL YHRAHCISDQHHLASELKHVHLALKMNGYSDRTIRQAV SPTIQKQPPDTPIKIITLPYIAGVTKKISKILNKHKDPGPVQ YSTQNPTNTKIQVRFNTVHKIRDLIPSVKDVIPDLQYPGV YKLSCTCGDCYIGETKRHISDRIREHKSDLKHGRTNSSA VADHCFNTAGSHDIVFSDTKVLCKPPGYYQRFVHEAIQI YKHRHNFNREDGFKLSNHWKDVINQVKL				

					CGGAAGCT TGTGTTGCC TCCGAAAC GTCGTGATT TAATTTAGT TGTTTTCAT TTATTTTAT TTTACCTAC GTGACTTTA CCGAAAAA AACCAAAG AGTATGAA TCCTTGCAG TCACGAAA GCTTCAAAT CTAGAA				
Pene lope	Posei don-1F_P M	.	Petromyz on marinus		TAACCTGCT GTGATGTTT CCATCCTTC CTGCGATTT TAGCTGACT CTCTGAGTT GCATTATTA GGTTGGAT GTCTTTGTA TTGACGGG CTGGATGTC CTTCTCGCT CAGTCACC GATGCTGC GACATACT AGGCAAGT CACCGATG CTGCGGCAT ACTAAGCTC GTGCATTGA CATGCTAAT CTCTCACAA TACAATAG CCACTTCCA	IHKEGTPLRPIVSTINSPTYKLTRFLSETLLPFTGKTTSAV LNSAHLVSLIKDISLSPGDLLVSFDVVSLFTKVPVPDTIAI IEDLVQQGLNXDVPTLALPYQHLLHLEWYILQTRGGYP QWVLPYRQWCANFGTALPTLTSPGMVHSTNKRRVPPM GSPLSPVVANVFMEKFETDAIQTSTLRPSLWKRYVDDTF VIWPHGQPALNQFFNHINSIHPAIKFTMEVEKDGKLPFL DVLIEKRPDGSLGHSVYRKATHTDAYLKPDSHHHPAQK NSLIKTLHHRAHCISDQHHLASELEHVHQALRMNGYSD RTIRKAVSPAMGRQPSEDPIKTITLPYIAGVTKNISKILNK HKIQVRFNTVHKIRDLIPSVKDVIPDMQYPGVYKLSCGC GSSYIGETKRHVSDRVREHKADLKHGRTNTSAVADHCF NAVGTHEIDFSKTKVLCKPTGYHQRFVHEAIQIYKHRH NFNREDGYKLSNHWKDVINQVKS	cd003 04	156	239

					CAGCCTCAGGGAGACCATCTCTAGCCCCGCCCTCTTACCCTGAGGCTGCACCCATCTCAGGCCTACATAAGACCTTGTTTCAGTAGCTTTCTCTCTCTCACCTGAAGACGGTCGCTTGTGTTGCGATCGAAACGTCGTGATCTATTAATTTTCTATCTACGTGACTTTACCGAAAGAACCAAAGAGTATGGATTCCTGCAGTCACGAAAGCTTCAAATCAAGAT				
Penelope	Poseidon-1G_PM	.	Petromyzon marinus	AC	TAACTGGCTGTGATGCTTCCATCCTACCTGCCGATGTTAGCAGTTATTTCCGACTCTCTGAGTTACATTACCTTGTTAAATGTCTTTACATTAACAGGCTAAATGTCTCACTAACA	QPPGYENQQYITDSTSLIEPTKRLSLKPSAARPPRFYGLPKIHKDGTPLRPIVSTIDSPTYKLARFLSETLLPFTGKTTSAVLNSAHLVSLIKDISLSPGDLLVSFDVVYLFTKVPVPDTIAIIEDLVQQGLNADVPALVRHCLTNTYFTWNGTFYKQVEGTPMGSPLSPVVANVFMEKFETDAIQASILRPSLWKRYVDDTFVIWPHGLPALNQFFNHINSIHPAIKFTMEVEKDGKLPFLDVLIEKRPDGSLGHCLQESHPHGCLSKTSKRDQMAHWDTVYRKATHTDAYLKPDSHHHPAQKNSLIKTLHHRARCISDHHHLASELEHVHQALKMNGYSDCTIRKAVTPTVGREPPEDPIKTITLPYIAGVTKXISKIMSKHKIQVRFNTVHKIRDLIPSVKDVIPDMQYPGVYKLSCACGSSYIGETKRHVSDRVREHKADLKHGRTNTSAVADHCFNAVGTHEI	cd00304	155	246

					TAGCTGCTGCTCCCTTTTCCTCGATGCGGCGGTGGCGCTCGTTGTTTAATAGGCTGAATGTCTCTCATTAACACAGCCGCTGCTGCGGGAGCCGTCTCACATTGACATGCTAATCTCTAATCTCGAGCAATACAATGGCTTCTTCAGCACCGTAGCCGTGGGGTTACATTAACAGATTGACATGCTAATCTCTAGCAATACAATGGCCGAGGTATACCTAGCTCTGCGATCTTACATTTACTTGCAAATTCCATCTCTAGCCCCGCCCACTTACCCTTGGGTTGTACCCATCTCAAACCTATATAACCTCTGGCTGCAGTAGCTCTCTCGCTATATTTCTTCCTGCCTCAGC	DFSKTKVLCKPTGYQQRFVHEAIQIYKHRHNFNREDGYKLSNHWKDVINQVKS			

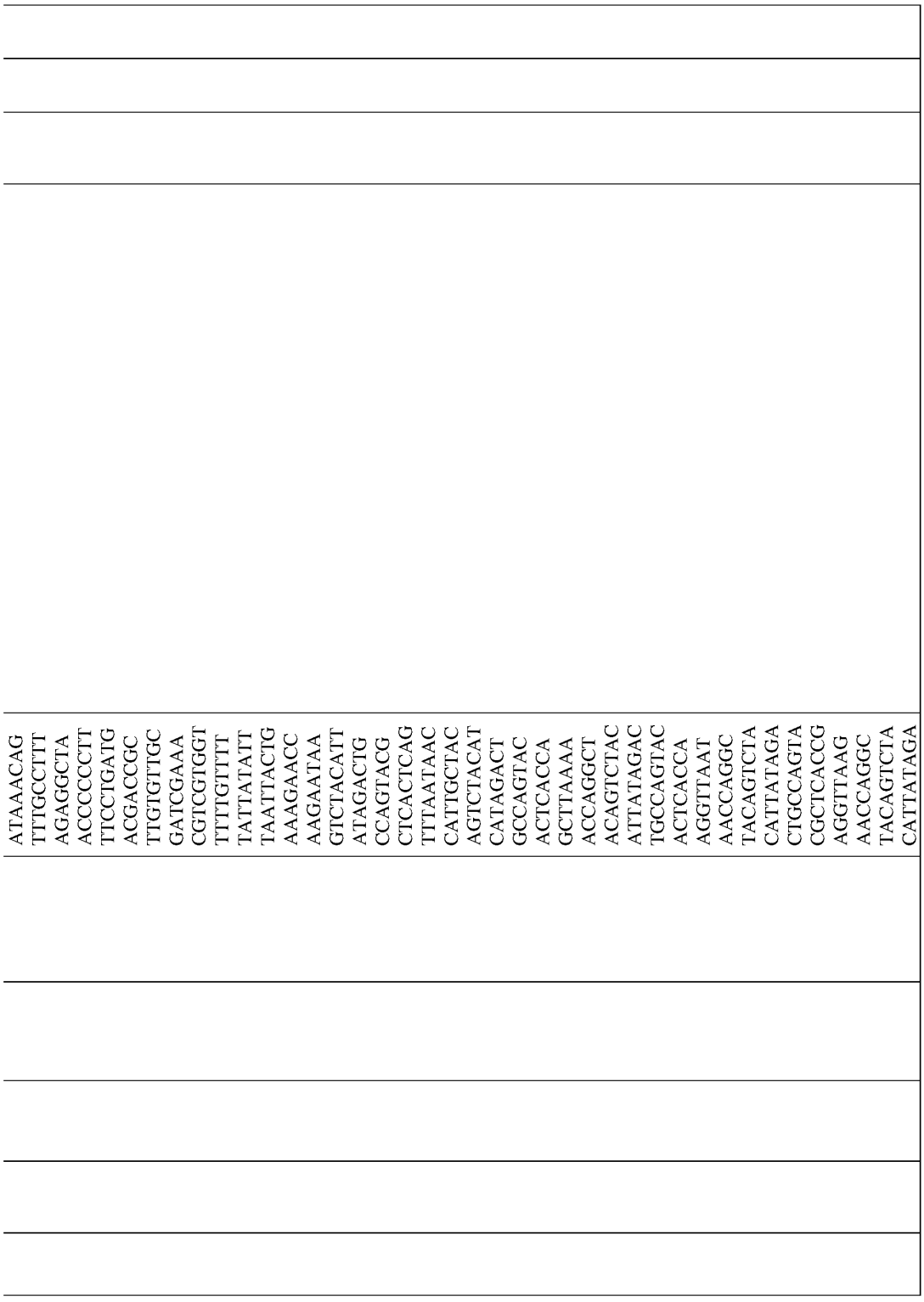
					AGTCCGCCT GCTGTATTT TTTCACCTG AAGACGAT TGCTTGTGT TGCGATCG AAACGTCG TGATTTTAT TTTATTATT TTTCTATCT ACGTGACTT TACCGAAA GACCCAAA GAGTATGA ATTCCTGCA GTCACGAA AGCTTTAAG TCAA				
Rex 1	Rex1-1_PM	.	Petromyzon marinus	G	TAATTATTA AGTAATACT GTTTACAAT TGACTGTAA AATAGGGA TGTGCAATG ACAAAATG TACCTTTAT GTGGGTAT GTGGGTAT ACTTGGGC AGTTTTTAT ATATAAAA TTTATTGAT TATCTTTTT TTATTTTAT TTTTATTTC TTGTTTTAT TTATTTAAA TTTATGAAT GATGCACT GACTGGTG	MAAHTVAAAQCLPGRCLVWLFMYALCLVLCTFCRASN TFSRHDLLKIGLCYERSVTSEFLRLHKIPVELARSPGSPW IIIPAGRRRRRRRERRQKRGCRAGALARLRRQPLKPPLPS LFLTNARSLANKMDELRLQTTANNVVKDSCILLFTETW LHSSIPDSAIELAGYTAQRHDRTVDSGKSRGGGLCVYV NNSWCTNTVTVDSHCSPDLEYVTVKCRPIYLPREFTVV MITAVYIPPDANANSAIGLLHGSISIQQSKYPDAVQIIAG DFNHADLKAALPKFHQHVKCATRGTKTLDKVYTNIKL GYRARPLPHLGQSDHLSLLLIPAYAPIKKTAPTITKTVAT WPEGATQQLQDCFDRTNWGVFEHQDLEVFTDSVLCYI KNCIDTVTVDKRIRVYPNQKPWMTREVQQLLKERNIAF KSRDKALYSTARANLKRGIREAKADYRRRIEERLDSND SRQVWQGVQHLTNYRANLGAADGDTALAEELNLFFAR FEVTPPGTASPNPTVHSSLNLTVEEHEVRRTLRAVNPRK AAGPDGVPGRVLRDCADELAGVFTRIFNLSLEQSTVPPC LKSSTIVPLPKKTHISSLNDYRPVALTPVVMKCFEKLVRS HIIATLPRSLDPHQFAYRANRSTEDAVATALHAALSHLE QQGSYVRMLFVDYSSAFNTILPHKLVVKLGDLGLPHST CMWIHSFLSGRRQRVRVGRHTSTALSLSTGSPQGCVLSP LLYSLYTHDCAPVHHSNTIVKFADDTTVVGLISGGDESA YRDEVERLTAWCRGNNLLLNTAKTKELIIDYRRKKTDT PPLTINGDCVERVADFRFLGVHIGEGLTWSANTSELLKK	cd01650	580	832

Penelope	Poseidon-1D_PM	.	Petromyzon marinus	ATGGATA	TTCTCTCCT	KTKVLCKPTGYYQRFVHEAIQIYKHRHNFNREDGYKLS	cd00304	186	270
				A	CCCCCCCCTA	NHWKDVINQVKS			
				CCCAGGT	TAACTGACT	AQQRLYFLRILRKNNITQKLLVSFYRASIESILTYCICVW			
				CGACGTA	GTGATGTTTT	YTSCTVAQRKALQRVINTAQKIAGCPLPTLEELHSSRCIK			
				GCCCTSA	CCATCCTAC	KAQNIIKDSSHPGHSLFELLPSGRRFRSVKTRTNRFKHSF			
				GAGKCTT	CTGCCTATG	YPTAITTLNTAN			
				GGGACAC	TCAGCTGTT				
				ATCCCCG	ATTTCCGAC				
				CAGACGT	TCTCTGTGT				
				TSCTAGA	GTCTTTCTC				
				TCTGAAA	TCTCTCTCT				
				TCGCTGC	GTGCATCTC				
				TACCCTC	TCTCCGATC				
				AAAAAACG	TCTACTCCA				
				CTAAGAA	GTGACCCG				
				ACCCATT	GGTGCTGCC				
				AGCAACA	GCGGGGTC				
				TCAGCAA	CACTTTCTG				
				GGACGAG	CCCCCCCCC				
				CGTKTTG	CCCCCCCCC				
				AACTCTC	TTCTCTCTC				
				TCAATTA	CCCCCCTGC	TTYNNXIQXLJNDGXYTKINRDPTNRLATKXNNITNSTS			
				AAAAACA	AGCCGATCT	LIDSNKQRSLKPSAARPPRFYGXPKIHKEGTPLRPIVSTIN			
				GAGATGA	CTCCAGTGC	SPTYKLXXFLPETLLPFTGKTTSAXLNSAHLVSLIKDISLS			
				CATAKTC	TGCCGCGGT	PDDXLVSFDVVSLFTKVPVPDTIAIIEDLVQQGLNPDVPT			
				ATCCTCT	GGTCCACTT	LVRXCLTNTYFTWNGTFYKQNEGTPMGSPLSPVVANVF			
				CAGCCGA	TACTCCAGT	MEKFETDAIQTSTLRPSLWKRYVDDTFVIWPHGLPALN			
				AAGGGCA	GACCCGGG	QFLNHINSIHPSIKFTMEVEKDGKLPFLDVLIEKRPDGSL			
				ATGCCAC	CCGCCTCCT	KIRDLIPSVKDVIPDMQYPGVYKLSCGCGSSYIGETKRH			
				TGTCATC	CCCCCCCCCC	VSDRVREHKADLKHGRTNTSAVADHCFNAVGSHEIDFS			
				ATGGATA	TTCTCTCCT	KTKVLCKPTGYYQRFVHEAIQIYKHRHNFNREDGYKLS			
				A	CCCCCCCCTA	NHWKDVINQVKS			

					CTGCCGAGGCCTATATAACCTCGTGCCACGGCAGATCTCTCGCCTACTTCAGCAGTTTTTCTTGCTGTGGTTGTCCCACCTGATGACGATTGCTTGTGTTGCGATCGAAACGTCGTGATCTATTATTTTTCTACCTACGTGACTTTACCGAAAGAACCAAAGAGTATGGATTCCTGCAGTCACGAAAGCTTCAAATCAAGAT				
Vingi	Vingi-1_PMa	.	Petromyzon marinus	GGGGAAGAAG	TGACGATCATTAAGAAAAAATAAAAACTGTTGCTTTTCACAATGCCATACGAAAGAAGAA	MTTCSRSSLLRAKAQRMSSHRRPNARQRYPGLKSNSTESISQGQSNVNVGPSLRIMQLNIEGLSSAKCDYIAKILQFQDVDVLALEETHLDRQCPRSKIIGYTLIAASHHPKHGLATYVKNDLKDTVQVVPHTSPHYIAIKINDVTVVNVYKPPGMEWPCPALPQPSHPIIHIGDFNSHHTLWGYSANDRAGEQLTSWAAAEDQHLLYDAKDRGTFKSARWKKCYTPDLCFVSRDQNGNPMPSTRTILDDFPNSQHRPVVIHIGLQIPLFRCAPKPRWNLRSAEWVNFTADLEESVGMIPPRAVNYRRFVGLIKAVAKANIARGFRKVFTPCWTPESEQLFRKFNADKKQETAKALLESLNLQRRERWRETVEALDFAKSSRKAWRLLRLMGAAATPAYKPPQVSPDEIARQLINNSRTSIDKSTRRQVYKELYCTLDNCQETSEFSRPFSVEEVDAGIAQLKCGKASGLDDIYPEFIKNLGPRARKWLATFFTDIHSTSIVPKEWKEARIIAILKPGKDASDAQNYRPISLLSVCYKLMERLLLKRLMPTLEEIVPQEQAGFRTGRNCCDQVLALTSHIEAGFERQLKTATAFIDLSSAYDTVWRKGLLLKLAKAVP	cd01650	508	762

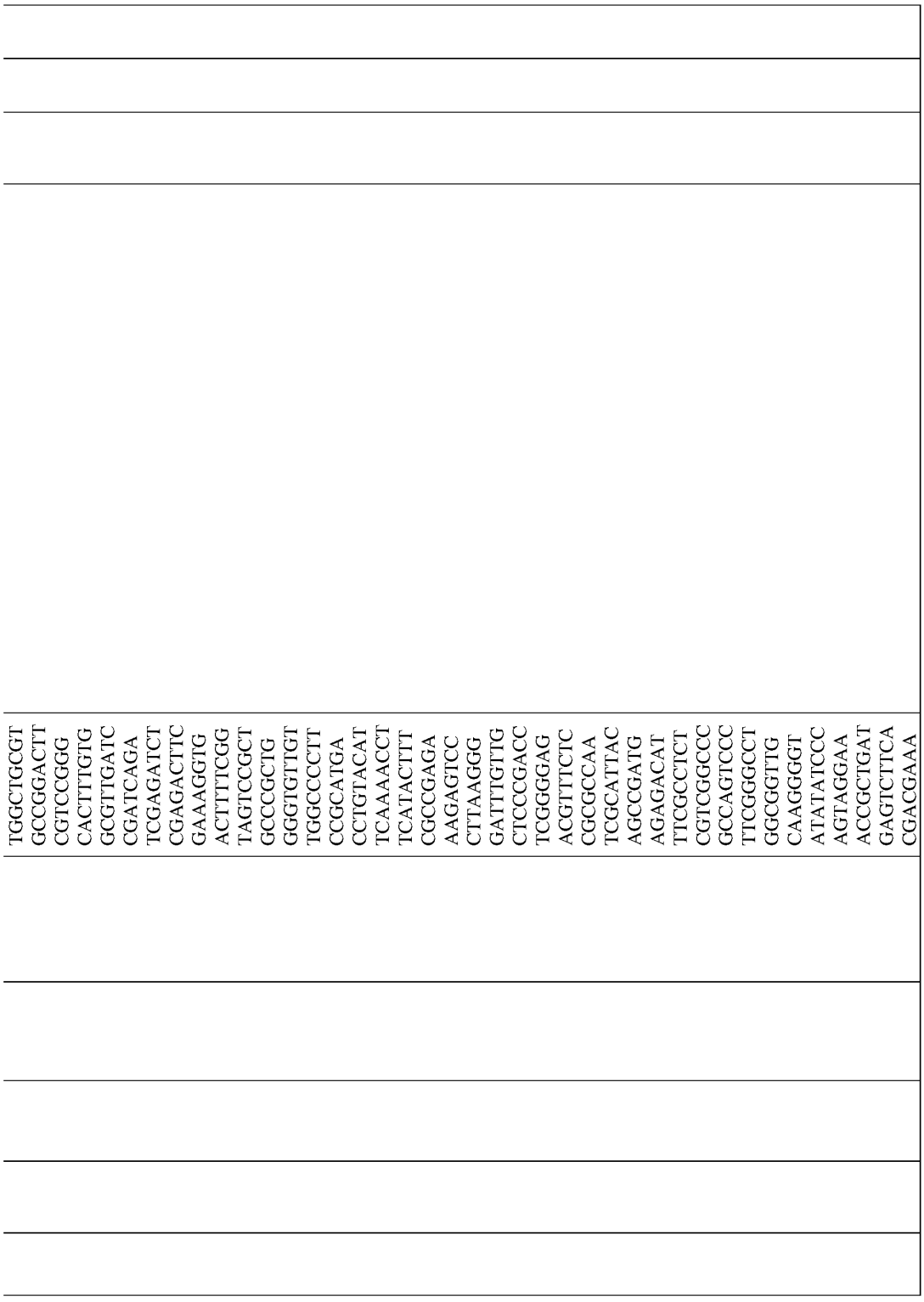
						CRPTIRLITEMLSNRWFTVHLGEAASSPRRLNNGLPQGS				
						VLAPALFNLYTSDIPATSCRKFTYADDIALATQARKMED				
						TEAVLEKDLKTLEEYFKNWRLKPNPSKSVVSAFHLNNR				
						QASKTLNVNFCGAKLRHDEHPRYLGVTLDRTLNYHHH				
						LKQAKGKTKARVNIIQKLAGTTWGSDASTLKISTQALV				
						HSTAEYCAPVWGASPHTNLVDAEINNALRIISGSLKSTP				
						LPWLPVLANIAPANIRRDDATTREYNKIMKNFTLPLHQD				
						LKEPPRPRLKSRKPFWIRASALAAENKDTATRWSESWA				
						AMDVANKNIIHDPTVEVPGFSLPRRLWTTLNRARTQHG				
						RCAYTFHKWKIKESPNCDCGEVQTIKHITDECSIHAHPG				
						GIAEIHTASTTSLDWMRNLCIHL				
					TAAACTAAT					
					TTTCAATGG					
					TACAGTATA					
					TTTTTTCAA					
					GTAGTGGT					
					ATCCCAAA					
					CTTTCACAG					
					CAGCTTATT					
					TCCCTTCAC					
					ACCAACATT					
					ATGTAAACT	SLSHTSAAVSSPNSKQILFKASLSLISERIRLLRFELDTLD				
					CCCCCCCCC	RTLLQAHLTLTQKMSPSDWNHVDNMTASSADSLSVKC				
					CCCCCCCCG	KANLISKFNHLSKKNPVPHITKLDVNKIVINQSSTPISDN				
					CGCCGGTCC	LKTVLARGLNFAVTPKHIPVDDIISQVDVALKGIGHIQA				
					ATCATGTAC	DVARAEIATILKNAKKPISNISKNERSELANLKKQKKNR				
					CTCCTACGT	SDIVILSADKGNATVIMDKDSYNNKIQNLLNDGSYSKIK				
					ACATCATGC	NDPTNSLTIKTNKIITSSTSLIEPTKRPSLKPSAARPPRFYG				
					TCATTTTTA	LPKIHKEGTPLRPIVSTINSPTYKLARFLSDTLLPFTGKTS				
					TGCTAATTC	SAVLNSAQLVSMVKDVVLSPDDILVSFDVVSLFTKVPVP				
					CTTTGTCAA	DTIEIIENLVQHGLDPDTPALVRHCLTNTYFTWNGTFYK				
					GGCGTTTTT	QNEGTPMGSPLSPVIANVFMEKFESDAIQTSTLRPSLWK				
					GTTTCCTCC	RYVDDTFVIWPHGLPALNLFLHHINSRHPAIKFTMEVEK				
					CACACAGC	DGKLPFLDVLFERRPDGSLGHSVYRKPTHTDSYLKPDSH				
					TTAAAGGA	HHPAQNNSLIKTLYHRAHCISDQHHLTSELKHVHLALK				
					CGCGACCT	RNGYSDRTIRQAVSPSIQKQPLDTPVKTITLPYIAGVTKK				
					GGGCTTGG	ISKILNKHKILVRFNTVHKIRDLIPSVKDVIPDLQYPGVY				
					CTGTCTTCA	KLSCTCGDCYIGETKRHVSDRIREHKSDLKHGRTNSSAV				
					CTCCAACGC	ADHCYNAVGDHDIVFSDTKVLCKPTGFYQRFVHEAIQI				
					TGTCCTCTG	YKHRHNFNREDGFKLSNHWKDVLNQTKH				
									cd00304	401
										485
Penelope	Poseidon-2_PM	.	Petromyzon marinus	CA						

					GACAGTCCGGTCCCTTCACCTGATGAAGGCCACTTGTGTTGTGGCCGAAACGTCGTGATTTTTGTTGTTTTATTTTTTACCTACGTGACTTTACCGAATAGAACCAAAGAGTATGGATCCTTGCAGTCACGAAAGCTTCAAATCTAGAA				
Penelope	Poseidon-1J_PM	.	Petromyzon marinus	A	TGACTTGATTTGTGTTTTCTTGTCAACCTCCACCGTCAGCAGTGTGGGTGTGTGGAGCGAGGTGCAAATGATTGCCTGGCGGCCCTTTGATTGGTGTAGACCCACTGCGTTGGCGTTGTTTTCTCCTACATTAACATTAGCTGCCCCGCCCATGCTGCTGACACACTTGGGCTACGCCCATCTCTAAACCTTT	KSKFHRLYGEGFHITIRSFNKLRDKKAHLIESLTFLKRCHNSKVIPISFTLTSAAVTSSNAKRILFRASLALISERIRLLRTELDHVDKTLLKSHLTLTEKMSTSDWEHVDNMAASSAATMSAKCKDNLISEFNLLSKKKPDPHIKKDVDKVVINLSNSPISDGLKSVLAKGLNFAVTPKHIPVDDIIAQVDVALKGLGHIAADVTRSEIAATLKSAKIPVSNISKDERFELSQLKKRDDIVILSADKGNATVIMDKTVYNDKIQDLLNDGSYTKIKRDPTNSLVTKTNSIITDSASLIEQTKRFSLKPSAARPPRLYGLPKIHKEGTPLRPIVSTIGSPTYKLARYLSETLQPFTGKTTSSVLNSAHLVSSIKDISLSPGDLLVSFDVVSLFTKVPVPDTIAIIEDLVQQGLNADVPALVRHCLTNTYFTWNGTFYKQTEGTPMGSPLSPVVANIFMENFETDAIQASTLRPSLWKRYVDDTFVIWPHGRPALNQFLNHINSIHPAIKFTMEVEKDGMLPFLDVLIKKRPDGSMSHSVYRKVTHTDAYLKPDSHHHPSQKNSLIKTLHHRAQCISDHEHLTSELKHVHQALKANGYSDYTIRKAITPTVRKEPSEDPIKTITLPYIAGVTKNISKIMNKHKIQVRFNTVHKIRDLIPSVKDKIPDMQYPGVYKLSCICGSSYIGETKRHVSDRVREHKADLKHGRTNTSAVADHYFNAAGSHEIDFSKTMVLCKPTGYQQRFVHEAIHIYRHNFNREDGYKLSNHWKDVINHFGS	cd00304	443	527


ATAAACAG
TTTGCCTTT
AGAGGCTA
ACCCCCCTT
TTCCTGATG
ACGACCGC
TTGTGTTGC
GATCGAAA
CGTCGTGGT
TTTTGTTTT
TATTATATT
TAATTACTG
AAAGAACC
AAGAATAA
GTCTACATT
ATAGACTG
CCAGTACG
CTCACTCAG
TTTAATAAC
CATTGCTAC
AGTCTACAT
CATAGACT
GCCAGTAC
ACTCACCA
GCTTAAAA
ACCAGGCT
ACAGTCTAC
ATTATAGAC
TGCCAGTAC
ACTCACCA
AGGTTAAT
AACCAGGC
TACAGTCTA
CATTATAGA
CTGCCAGTA
CGCTCACCG
AGGTTAAG
AACCAGGC
TACAGTCTA
CATTATAGA

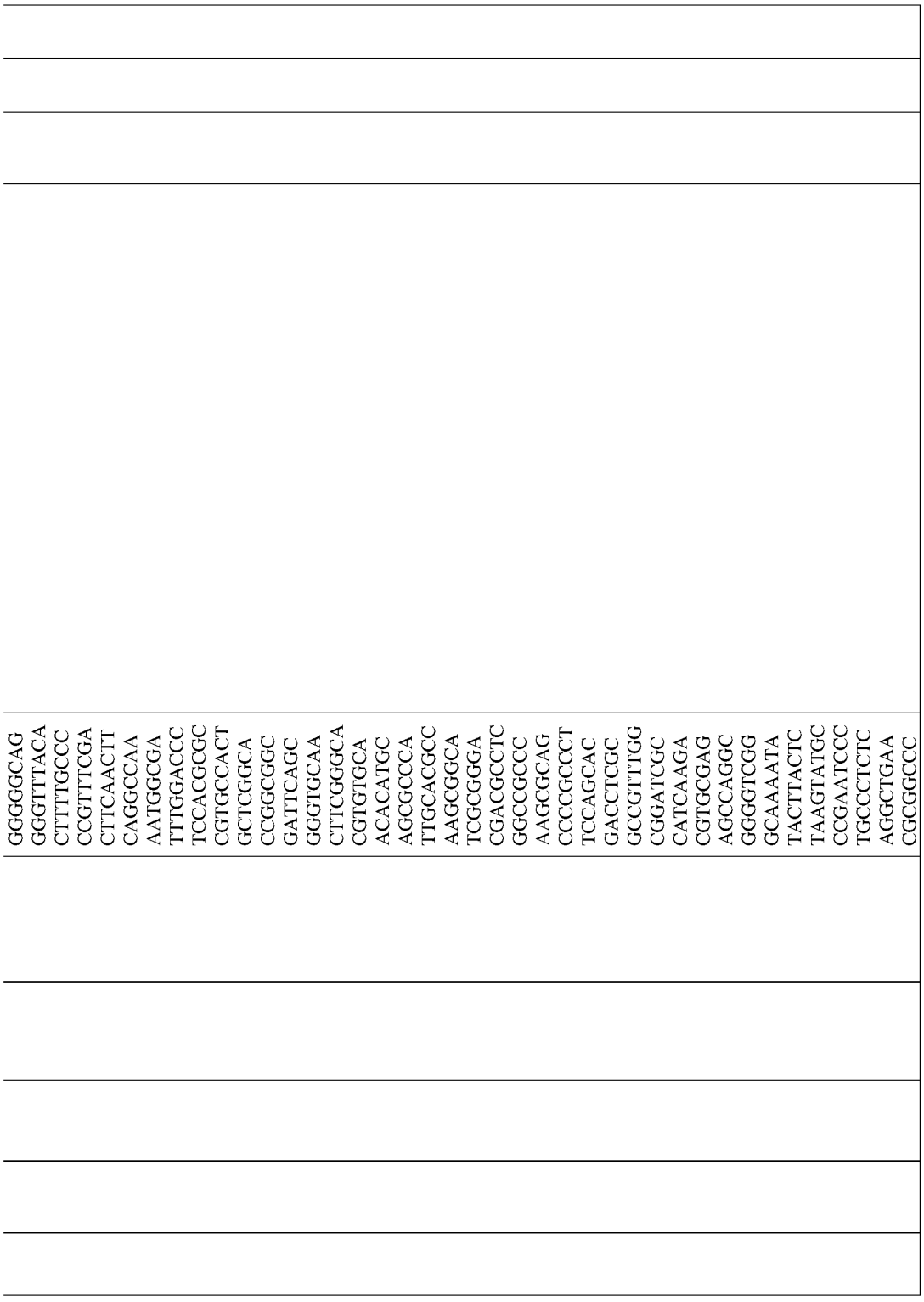
					CTGCCAGTACACTCACCCAGCTTAATAACCATTGCTACAGTCTACATTATAGACTGCCAGTACACTCACCAAACTTAATAACCAGGCTACAGTCTACATTATAGACTGCCAGTACACTCACTCCGATTAAGAACCAGGCTACAGTCTACATTAGTGGTTTTTTTGTTTTTATTTTTATTTTTATTCACTTTATTTTTTATTATTATTTTTCTATCTACGTGACTTTACCGAAAGAACCAAAGA				
Penelope	Coprina_Pc1	.	Phanerochaete chrysosporium	ATGATCTGTGCATACAAAACCTTCGTACTTTTGCCGAGAAAGAGTCCCTTAAGGGGATTCGTTGCTCCCGACCTCGGGGAGACT	TAATGCGATTCGATAACTGCTGGCCGCAATGGTCAGAGGATATTTACTTAGCGGCATACACTTTGGTTCAAACCCAAAACCCCTAACCTAAGCCCTAAGCC	MPAQDTTRRAVRATASSSTGHFPSFTRALSPSQASVVLGEQISFGTFDERCAAALAHCTLLDEVIATLPKLYQVALSPVLTDIHHDATKRVEIAAQLERLRQHKAAGTLPSPLKGLGSTPAIQFMKGFLESQKPDFPGKAPDIVKACAAKLLDGFIETKAAEHAWYESRLQAMEVTRRVLTAADTHWKDELEQRYQVLEGFVRPPVEGQVPATVPSAEELHPLKFVTDPVMVAQHENIRRSIPRLASTVIMLVVNRALTQQQRNKAKRDLKDQADVEMATDDTAVDSAVKRLEKQVEGLVSGMKALTNGKGKVSFSSSSRTHVEDAHPFVSGQSPSSRPFSLVDHRARHEAQGVQLGKRLRQRQDDGHQGPSRRLRCSSGEAGSPQEGQGEGTREGQEDPEEIVAYVRAAPWRYDAPHSYPDVILMLPAPRAIQLLLERVPLEVKRAARFRSRVFLGP	IPR000477	664	917

TTTCTCCA	TAAATATTC	HVTVPEHLQMHVSSSLRYLMYSVPSSNVVREAWEDFK			
CACCAAT	ATTAGCGTC	DRFRWRVFFTEKLIDSPNDDDPDFDPDYVVPKARKRFD			
CGCATTA	TCCCTCACA	DGVRTTPAYIELGLRKGDKFVDDYISNVMPSVKATSRSL			
CAGCCGA	GGCCTAAA	NLVWLDELKDYLSANKYIVSPSDKNLGVCILQADWAT			
CGAGCGA	CCCCCTGGA	DQTRTLWNDPSNYRKLGSLLEADEIMEEKHERIFRAAR			
CATTTCG	GCGGAGTT	LARRLGKRQLAEFLVSKIHGKDAKTAGYTLGQFYGVPK			
CCTCTCG	GTACCCGTT	VHKRPVKMRPIVPCHSVSQAPSATYVSKMLKPLVSRQP			
TCGGCCC	GAACGGTA	HVIHGSKHLVRQLESLQLDKHRKAWIVSGDIVAFYPNIP			
GCCAGTC	CAGCCAGC	LENCIRVVTQMWRNSPEAEKLSPEERQLFLMSFMTANK			
CCTTCGG	AGTTATAG	RLVIQFGDEYAEQIRGLAMGIACSPDLANLYGSHYENPI			
GCCTGGC	AGTATTTTG	LARPDMKSKFAFFGRYLDDVLGIVYASSAEEAMSYANE			
GGTTGCA	TTGTACAGC	IRYDGVEIEWSVSEYHTPFLDLFLYLDPVEKRLHWKPYS			
AGGGGTA	GTCATTATT	KPLNHRERIPWASHHPKDVKKGTFIGEMSRLATLCSKLE			
TATATCC	TAATGGAT	HYRDAISDLSNLYLARGYPDDAVRSWVVKAHFSQRWER			
CAGTAGG	GCAGTGAG	RLGEPIVRDEVFVLKSHFNPAWSAFNVHGLSEVVVESW			
AAACCGC	CGCCCCGCCT	LSSLSKLDGANQSQDRTWHAAESTDGSELKIFRQTTLG			
TGATGAG	TCGTGGCGT	QFQGEQSSLYPGESRPEGVAGRVGVANVVAKPRVTYDPR			
TCTTCAC	GGGCGCTTC	DVMDVDELVALSEPSFSAHDLSVVWTDNALRVPKSLRI			
GACGAAA	GGTGATCG	RYAMAELRLDADPQVNPALHVPLGVEDQVGDLGPASD			
GCTCAGC	CCGTGATG	AESVASGDDGTTNSKPYRICTAKKDSGIVQEDLDIRSVG			
GCGACAC	GCCGCGCA	FHDRAWLVSRKKIRNLGDIVNSWKKDIIQAALNSDSVP			
AACAGTG	CGTGTCCCA	HMHVDEWQ			
TCGGCTG	CGTGTTCGT				
CCTTCGG	CTCGTGGCT				
TTTTCCTG	GCTAATAG				
TGAAGCA	CGTCTGTGA				
TTCCGCG	TCATGTGAT				
CTCGCCG	CTTCGACCC				
CTCACTA	AGGGCTGC				
CGCTCGC	GGTTAACTC				
GTCTCGC	TGGATTCGA				
GCCCCCC	TGCGCATGT				
AAACCCT	TAATAGCC				
AATCAAA	GGCCCGAA				
TCGAGTA	TCCGGGTCG				
AGCACTC	GCTCCGAA				
TTCACC	CTGACTCCG				
	CTAGGTTCG				
	GGAGACTTT				
	CGCCGGGGG				


TGGCTGCGT
GCCGGACTT
CGTCCGGG
CACTTTGTG
GCGTTGATC
CGATCAGA
TCGAGATCT
CGAGACTTC
GAAAGGTG
ACTTTTCGG
TAGTCCGCT
GCCCGCTG
GGGTGTTGT
TGGCCCCTT
CCGCATGA
CCTGTACAT
TCAAAACCT
TCATACTTT
CGCCGAGA
AAGAGTCC
CTTAAGGG
GATTTGTTG
CTCCCGACC
TCGGGGAG
ACGTTTCTC
CGCGCCAA
TCGCATTAC
AGCCGATG
AGAGACAT
TTCGCCTCT
CGTCGGCCC
GCCAGTCCC
TTCGGGCCT
GGCGGTTG
CAAGGGGT
ATATATCCC
AGTAGGAA
ACCGCTGAT
GAGTCTTCA
CGACGAAA

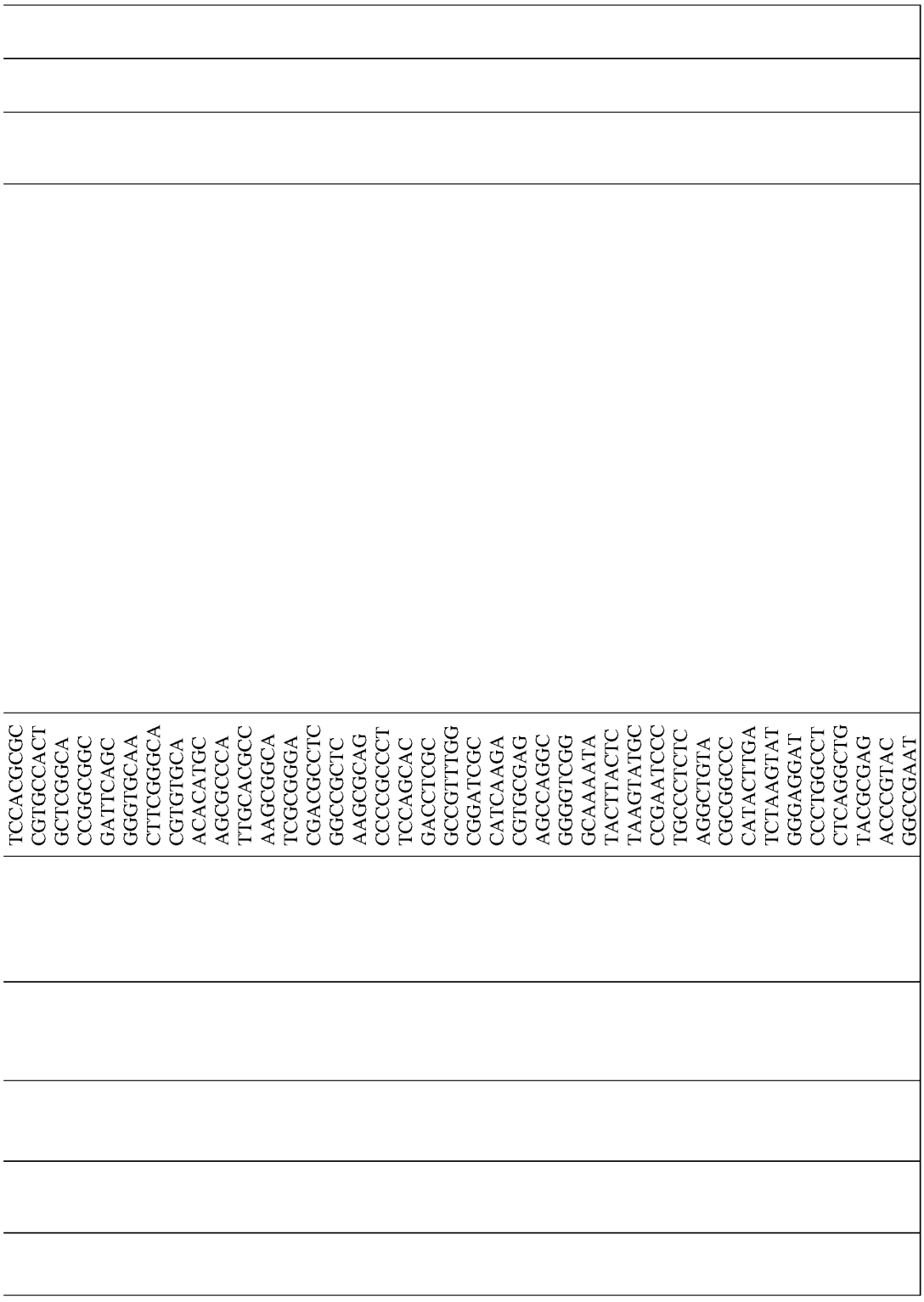
					GCTCAGCG				
					CGACACAA				
					CAGTGTCG				
					GCTGCCTTC				
					GGTTTTCCT				
					GTGAAGCA				
					TTCCGCGCT				
					CGCCGCTCG				
					CTACGCTCG				
					CGTCTCGCG				
					CCCCCCAA				
					ACCCTAAA				
					CCCTAAACC				
					CTAAACCCT				
					AATACCC				
					TAAGCTGGT	MLADPAALAAGLARAPPPPSAPQDPSPAFPAGPAGQNP			
					CATCATGAC	RAAAPARVEVHTVVAPPGRAGGMLPDPGLVEEPIQATY			
					CGACAGGG	AHDAAQFECALCPYVAESMAVLVQHRRSAHRGTRFKD			
					CACTACCCA	IFTSGCQCSLVFYARIVAASHAVACARRNQRAVPPAPTP			
					GGTAGGGA	VAPTRPEATPQPTGYLAAAMTAAAAAASSDTVVAAAT			
					ACCGCCCTT	NMQSAVPAAAKTTGLQLVPPELEPALPQRASCHAGKRR			
					AAAAAACC	RLNADEAVTPCTPTARVSPQTEVAMAPHDAPQDDTVLQ			
					CAGGAAGA	REAAEPQPDPAATPGAQVQRVEDTTAAQDDTVQQDHD			
					CACAAACA	ADTAQVSPPRRTPTRWGPRPSSTQEPSPMTGEPAATLAA			
					CCCTCCACT	RRPLTPAATGTRATRWGPCHRAIGAAAIARLVTGLSTEP			
					TAGTGACAT	AQPQRRQPPPPQEPPSQPEPLAAAATAAADIAATVAADI			
					ACATATTTT	AAAAANAAMDVDGGPAADETWLLRFDGACRRNPGPG			
					AGCCTAGA	GAGAALFAPSGAVVWTCSHYMPSRSETNNTAEYTALLL			
					TTTCAGTTA	GVQSAVHHGASHLEVEGDSSLVIAQVKGTFACRNARLR			
					CGGAGAGG	QLRNRVRHALRSVDTHKLRHIDRQANAHADRLANRAL			
					TTACTAACT	DQRRTSSECGTHGSCMDSCLAVPTALAAQETPPAAPPST			
					GGTAAATA	SATPAEGNAMDDIAAEIAARDEGETFPVLPIGPGSAPER			
					CGAACACA	QPRLRLRQLSDEERDAAADALQELADTMASKIEDADSW			
					TATTCTGTT	TSGEGYISSIPERIREVLQPYATAPPQPGRSRRQQRRRPPR			
					CTAATCAGT	VTRNQREHRLDEALDDMAATQQATPRDQRAVRRARRR			
					GTGAAAAC	VGRVRASMAQQELRHEFAKDESKCVAKILKTASTETAA			
					TGGTTTTCG	EDEHPETCPIDAATLHAHFTGVNAPRTDFDYDATSGREF			
					CCTTTTGGC	RAAMSDLPPATVEIDAFDAELTIDEVEDQLTRAAKTSSP			
					GGACTTTTT	GHDGIDYGIYSRFAAQLVPLLHAVFQFCWRHRRVPRLW			
					CACTCGCAT	KVGIVRLIHKKGDPRQPTNWRPICLQPTIYKLYSGLLAH			
NeSL	Utopia-1_PI	.	Phytophthora infestans	AGCTCGG			cd01650	919	1187
				CCTCGCG					
				GCTGCCT					
				TCCCAGG					
				CGCCGCC					
				GACTTCG					
				CGCTCTG					
				GCGCGGC					
				CCACACG					
				CCGCCGC					
				CGAGCCT					
				CCAAGCG					
				CGCCCGT					
				TGGCTTT					
				CGCAGAC					
				GCAGGGC					
				TCGCGGC					
				GACGGCC					
				CTCGCCA					
				GCCCCCA					
				AGACCCC					
				CCCTACG					
				ATGTGGC					
				ACCACCC					
				GGCAGGG					

CGGCCCGG	TTTTGGGCA	RLSRWLEGNDRLPMAQKGFRAFNGCHEHNFMATTLLD		
CAGGCTG	ATCGTCTGC	QTRRQHRKLYQVWYDLRNAFGSLPQQLMWRVLRHLG		
CCCGACT	GGCTAGCTT	VDSGFIDRCRDIYRDSAFVVANAADGATDPVRQEVGVY		
CGGTATC	GCTAGCGG	QGCPLSPLLFVAALVPLVRRLEKLDGVGVPLADGVRPC		
GCCGGGT	CGGACGAG	TTAYADDLKVFSDSAAGIRKCHDTVAGFLAWTGLRAN		
GCTACAC	CGGTCTCCG	PGKCASLAVTTNARGNPTRDSSMRLEVHDAAITTLSLH		
TCTCAGC	GGGGCGTT	ESYRYLGVGDGYDHVRHRLQLEPKLKQLKREAVALLT		
CGCCACA	CACCTTTCC	SGLAPWQVVRALKVYVYPKVEYALRHLRPLQSQLQAF		
GCTCGGG	CCCGCGAG	DRVVSKGLRHLLSLPRSATSEVLYAPTSSGGLGLQPLVE		
CCTTGGC	GCCAACTA	LHRALQLAHAWQMLHSKDPAIQAVARAQACQVVRKR		
GGTCCGC	CACCGATCT	YRLQEDHWRGRDDELVRSFLNSELAASPHAEVLRRNG		
CATTGGC	TCTCTACAC	DIASLWSDVQRWLRIYHLRFEHCDETEAHGPLSFRVPH		
CCTTGGA	TTTTCTAAT	HNKWLTHKTVLRHVKLHLKIRHQTRWKGMVDQGKTV		
GCTCGAC	TCGCCTCCG	RVHGGVGAKFMTTGAGLSDDDYRFGVKGRLNQVDTN		
AGCGACA	TCTTCGGTC	SVLKRKRLRAHTTCRDPTCSSAETLAHVLNHCESNMDA		
GCAGCGA	TTCGGTTGT	IRQRHDDALEQIGSKIRGALDRAKSPTELRLNQTVPEYT		
CGACGAG	CGGGCTTTT	GAALRPDIVLRNVAAKTMVIADLAVTFEDQAARARHSS		
GACGCTC	TTCTTTTTG	LQLSHDHKTLKYQPIVAELQHKGWRVQTAAIVYGTLGS		
AAGACCC	ACCAATCA	VQPSNFKAYTEKFKLHKREARQLDLQLSSHCIRASHRIW		
CCACGCC	GAGCGCGC	GWHCRQHRDRQRSGTASRASRGSGGAPRRTSQAPARR		
GCCGCCC	CATGCGCCT			
CAGAACC	CTTCTGGCC			
CCCAGAA	AATCAGAG			
GACGTCG	ACCGGGCC			
CGAGTGT	CTGTCCTCG			
GCTTGCC	GACAGCGA			
CCACCCG	GGCCTCCAC			
GCAGGGC	GGCCAGCC			
AGGCAGC	AATCGAGT			
	CTCGGCAG			
	CGACGCGT			
	CTTTCTATA			
	GCGCAGCT			
	GACGAGGC			
	CGATCTGGC			
	GGCCCCCG			
	ATTGGTCCG			
	ACTTTCGGC			
	CAATCAGC			
	GACGACGA			


GGGGGCAG
GGGTTTACA
CTTTTGCCC
CCGTTTCGA
CTTCAACTT
CAGGCCAA
AATGGCGA
TTTGGACCC
TCCACGCGC
CGTGCCACT
GCTCGGCA
CCGGCGGC
GATTCAGC
GGGTGCAA
CTTCGGGCA
CGTGTGCA
ACACATGC
AGCGCCCA
TTGCACGCC
AAGCGGCA
TCGCGGGA
CGACGCCTC
GGCCGCCC
AAGCGCAG
CCCCGCCCT
TCCAGCAC
GACCTCGC
GCCGTTTGG
CGGATCGC
CATCAAGA
CGTGCGAG
AGCCAGGC
GGGGTCGG
GCAAAATA
TACTTACTC
TAAGTATGC
CCGAATCCC
TGCCCTCTC
AGGCTGAA
CGCGGCCC

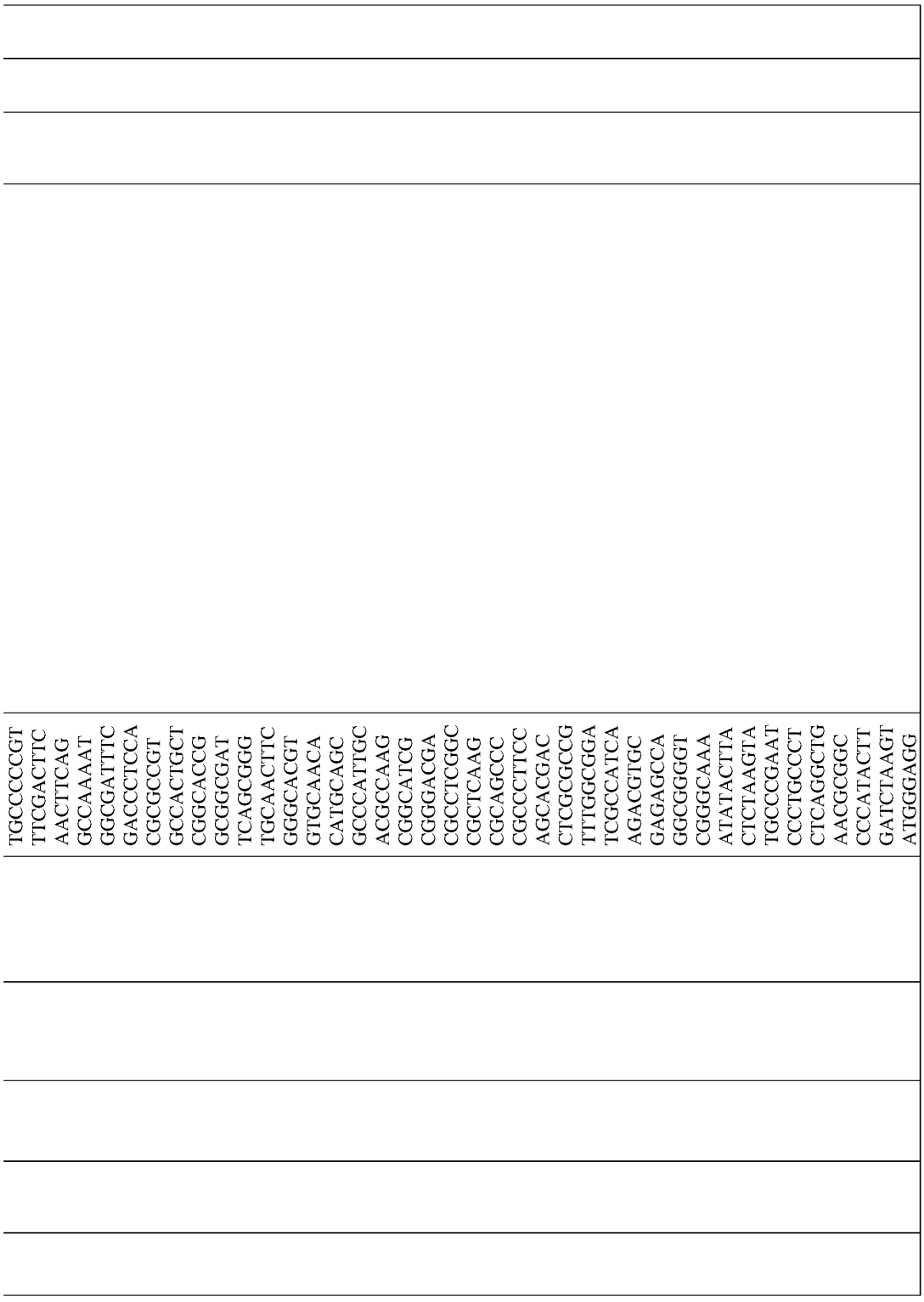
NeSL	Utopia-4_PI	AATU01001281.1	Phytophthora infestans	AGCTCGG CCTCGCG GCTGCCT TCCCAGG CGCCGCC GACTTCG CGCTCTG GCGCCGC CCACACG CCGCCGC CGAGCCT CCAAGCG CGCCCGT TGGCTTT CGCAGAC GCAGGGC TCGCGGC GACGGCC CTCGCCA GCCCCCA AGACCCC CCCTACG ATGTGGC ACCACCC GGCAGGG CGGCCGG CAGGCTG CCCGACT CGGTATC GCCGGGT GCTACAC TCTCAGC CGCTACA	CATACTTGA TCTAAGTAT GGGAGGAT CCCTGGCCT CTCAGGCTG TACGCGAG ACCC	TAAACGGG TCACTTGAC CGACAGGG CACCACCC AGGTAGGG AACCGCCCT TTAAAACCC AGGAAGAC ACAAACAC CCTCCACAT AGTGACAT ACATATTTT AGCCTAGA TTTCAGTTA CGGAGAGG TTACTAACT GGTACATA AAATTACA CATTCTGTT CTAATCAGT GTGAAAAC TGGTTTTCG CCTTTTGGC GGACTTTTT CACTCGCAT TTTTGGGCA ATCGTCTGC GGCTAGCTT GCTAGCGG CGGACGAG CGGTCTCCG GGGGCGTA CACCTTTCC	MLADPAALAAGLARAPPPPSAPQDPSPAFPAGPAGQNP RAAAPARVEVHTVVAPPGRAGGMLPDPGLVEEPIQATY AHDAAQFECALCPYVAESMAVLVQHRRSAHRGTRFKD IFTSGCQCSLVFYARIVAASHAVACARRNQRAVPPAPTP VAPTRPEATPQPTGYLAAAMTAAAAAASSDTVVAAAT NMQSAVPAAAKTTGLQLVPPELEPALPQRASCHAGKRR RLNADEAVTPCTPTARVSPQTEVAMAPHDAPQDDTVLQ REAAEPQPDPAATQGAQVQRVEDTTAAQDDTVQQDHD ADTAQVSPPRRTPTRWGPRPSSTQEPSPMTGEPAATLAA RRPLTPAATGTRATRWGPCHRAIGAAAIARLVTGLPTEP AQPQRRQPPPPQEPPLQPEPQAAAATVAADIAATVAADI REAAEPQPDPAATQGAQVQRVEDTTAAQDDTVQQDHD AAAAANAAMDVDGGPAADETWLLRFDGACRRNPGPG GAGAALFAPSGAVVWTCSHFMPSRSETNNTAEYTALLL GAQSAVHHGAKRLNIEGDSHLILSQVRGAFACNNKRLR SLRNRVQASLRQLDWYRLQHIDRKANQHADRLANRAL DLRRTVTECGPHAETRNRCFQTPQPLVEPGETHCVPGSD EVLAANTAMEDATAVPTEDDEAEVAARDGGEVFPTIAI GPDSAPAKQPRLRLKKLDEDDFDAAAAAVTRVSEELAS KIVDAGDWTSGEGYISAIPERLRAALRPFALPTQPARPQP REPRMQQPPRRPPRVTRDHLEHRLDEALDTMENVQRST PQNQKAVRRARRRVGRLRSAMDRTRLRKKFATHEREC VAEILRRASTEEAANPSQEKCPIDRATLHEYFTATSTQRT PFDYDSAKGTEFRTFLEVMSTPSHETSALTAEPTLDEIED QLAHVKAGSSPGHDGVGYDVYRRFQVQLLPLLHAAFR FCWRHRRVPALWKVGFVRLLLHKKGDPQQPNNWRPICL QTAIYKLYSGLLARRLSKFLEANELLPMAQKGFRAFNG CHEHNFVATTLLDQTRRMHRRLYQVWYDLRNAFGSLP QQLMWGVLRQLGVTEEFVARCSGIYEDSYFVVVGNASD GATEPVRQEVGVYQGCPLSPLLFITALVPLLRALENQDG VGVPLADGVRPCATAYADDIKVFCDSATGIQRCHALVT KREAVALLHSELVPWQILKALKVYIYPKVEYALRHLRP	cd01650	934	1198

				GCTCGGG CCTTGGC GGTCCGC TATTGGC CCTTGGA GCTCGAC GGCGACA GCAGCGA CGACGAG GACGCTC AAGACCC CCACGCC GCCGCCC CAGAACC CCCAGAA GACGTCG CGAGTGT GCTTGCC CCACCCG GCAGGGC AGGCAGC	CCCGCGAG GCCAACTA CACCGATCT TCTCTACAC TTTTCTAAT TCGCCTTCG TCTTCGGTC TTCGGCTGT CGGATTTTT TTCTTTTTG ACCAATCA GAGTGCGC CATGCGACT CTTCTGGCC AATCAGAG ACCGGGCC CTGTCCTCG GACAGCGA GGCCTCCAC GGCCAGCC AATCAAGT CTCGGCAG CGACGCGT CTTTCTATA GCGCAGCT GACGAGGC CGATCTGGC GGCCCCCG ATTGGTCCG ACTTTCGGC CAATCAGC GACGACGA GGGGGCAG GAGTTTACA CTTTTGCCC CCGTTTCGA CATCAACTT CAGGCCAA AATGGCGA TTTCGACCC	LKSQLQGFDSAIVRGLRHLLRLPENSHDGLFFSPTSAGG LGLLSLVELHEALQVAHAWQMLHSKDPAIRAIARTQVG QVARKRFKLVEEHWRGREDDLAQRFLNTELAASPHAT ETRRNGDIGSLWNDVRDTLQTLGLKFAAGDEEEAPGLL QLRVPHHTKWLSHSTVLRHVKLHMKLRRMDTWKSKV SQGTTVREHGGVGSRFITAGAGLSDAEYRFAIAPRAHLI DTNSTLKRRRLRANDTCRAPGCSYTEPPAHILNKCSPNM DAIRKRHDDALERIADALRRKVEKSGGRLEVAINKTVPE YDGAALRPDIVLRNTETKRAIIADLAITHENQPTDATTSS ALQQSRDNKITKYQTVAAAMMRAGWRVRVTGIVYGSL GSVLPSNFKVYTELLALLKRDARRLNRQLSSHCIRASAR IWSAHCRRHRERQRSGNASRASRGSGGAPRRTSQASAR R			


TCCACGCGC
CGTGCCACT
GCTCGGCA
CCGGCGGC
GATTCAGC
GGGTGCAA
CTTCGGGCA
CGTGTGCA
ACACATGC
AGCGCCCA
TTGCACGCC
AAGCGGCA
TCGCGGGA
CGACGCCTC
GGCCGCTC
AAGCGCAG
CCCCGCCCT
TCCAGCAC
GACCTCGC
GCCGTTTGG
CGGATCGC
CATCAAGA
CGTGCGAG
AGCCAGGC
GGGGTCGG
GCAAAATA
TACTTACTC
TAAGTATGC
CCGAATCCC
TGCCCCTCTC
AGGCTGTA
CGCGGCCC
CATACTTGA
TCTAAGTAT
GGGAGGAT
CCCTGGCCT
CTCAGGCTG
TACGCGAG
ACCCGTAC
GGCCGAAT

					CCCTGGCTT				
					CTCAGCCTG				
					TACGCGGG				
					GCTATACTT				
					GGTCTAAGT				
					ATGCCCCG				
					GTCGCTGGC				
					CTCTGAGCC				
					CGTACACA				
				GCTCGGC	TAGGCGGA	MLADPAALAAGLARAPPPPSAPQDPSPAFPAGPAGQNP			
				CTCGCGG	AACCAGGC	RAAAPARVEVHTVVAPPGRAGGGMLPDPGLVDSSPAAA			
				CTGCCTT	CCAAGACG	TAATPAPVAATATTARAAAARVAVEHHAHAEPNQEHLP			
				CCCAGGC	GCCGACAG	MARVLVEPMQVDECSSCDRSTLTADDGSGDDVAAPSS			
				GCCGCCG	GGCCCCAC	MLSNDVAAPMDVDSGTSCPPTLQQPLQRPRALHVGSKR			
				ACTTCGC	CCAGGTAG	RRLDADDGEEAHQLQEEEEAGIHAPALRLSAASAQPAS			
				GCTCTGG	GGAACCGC	VLAVYTHNASRFDCTLCAYTAGSFASLLTHRNSRHRRT			
				CGCGGCC	CCTAGAAA	AFLDRFSAGCACGVPFASRLAAARHAQACASLSSAPSA			
				CACACGC	CCCATTTCG	EASSAAGTSSPTADGADSTVSAVAHAEPGLPHHNDTEL			
				CGCCGCC	GTGGTCGA	TASPPLVSSSDVEVQATKTEATDNRWGAPLPRVLVASRI			
				GAGCCTC	CTCGAAGG	AGRLAQVPPPRWGPPLPRTTIAARIATRLAATPAPRWDP			
				CAAGCGC	CCTTACCTA	PLPRSLVVSRIAARLLPALPDAPACEEEAKDSDTMDWAP			
				GCCCGTT	TTTTTTCCT	TWTNEETKESEPHDEAPGQVDEETIDDADGEWLLRFDG			
				GGCTTCC	TAGACATTC	ACRANPGPGGAGAALFKPSGPVVWTCSHYDPSTTATNN			
				GCAGACG	AATTAGGT	TAEYTALLLGARAAADHGVTKLRIEGDSTLVIQQVRGIF			
				CAGGGCT	AGCGACCA	ATRSTRLRALRNKVKLELARVGSFSLHHIDRQANGHAD			
				CGCGGCG	AATTACAA	RLANAGLDRRRTKLECSVHPDGRGCTNTSVATAAPTAP			
				ACGGCCC	ATTTGGTAA	AAPLPSARPPASTAAPSPDDDHSDQGDIDDGEVYAAMCI			
				TCGCCAG	CGAGTAAG	SPDAVPHRRPRLRLRRLTDEESEEAGNVVERLAASLAA			
				CCCCCAA	CCAAATGG	KIADAPDWETAEGYITALPYALYDKLQPYSQSQHQPPR			
				GACCCCC	TAATACAC	QQQQQQRQRPRQQQQTRQRRQRRCKRGGGSQHRQRK			
				CCTACGA	AAAACTTTT	TRRRRPPPRVTRHHHREHRIDEALDDLHALESRRPQDRTAI			
				TGTGGCA	CTGTTCTAA	SKARRRVGRIRSALDQHQLRHRFDTDEKACVDGILAAA			
				CCACCCG	TCAGTGTGA	RDKDRAASVTTTAQTAAPPHSAPASAPSSAVDDGICPIP			
				GCAGGGC	AAACTGGTT	GDLLHAFFTDVNTPRTEFDADSPIGARFREALAQLPAAI			
				GGCCGGC	TTCGCCTTT	AATELLMEPPSPDEVEDQLQRVRGTSSPGLDGVGYDVY			
				AGGCTGC	TGGCGGAC	KTFTQQLLPALHAAFSRCWTDQRVPQSWKLGVVRLLF			
				CCGACTC	TTTTTCACT	KKGDRQDPANWRPICLQQAVYKLYAGILAHRFTRWLD			
				GGTATCG	CGCATTTTT	ANTRHADAQKGFRAVNGCGEHNFLAATLTDNARRRRR			
				CCGGGTG	GGGCAATC	ELHVVWYDIKNAFGSVPHELLWEVLRRMGVPAQFIACC			
NeSL	Utopia-2_PI	.	Phytophthora infestans	CTACACT	GTCTGCGGC	QGIYDAAAFTVGNAADGTTAPIQLRLGVFQGCPLSPHLF	cd01650	1026	1289

CTCAGCC	TAGCTTGCT	TAVISPLLHALKRLPGTGVQLSAVDRPGASAYADDLKV	
GCCACAG	AGCGGCGG	FSDTKDGITRQHQLVTDFLRWTGMVANPSKCSTMSVQR	
CTCGGGC	ACGAGCGG	DNRGVLKTANLTLQLDGAQIPALGMTEAYAYLGIGDGF	
CTTGGCG	TCTCCGGGG	DHVRRRVELAPKLRELKADTTALMQSGLAPWQVVKAL	
GTCCGCC	GCGTTCACC	KVYIYPRVEYALRHLRPFQQQLQGFDRHLARGLRHLLR	
ATTGGCC	TTTCCCCCG	LPTSATTEFLYAPTSRGGLGLLPLTEVHGALQIAHAWQT	
CTTGGAG	CGAGGCCA	LHSPDPAIRRIARVQLRQVADARHRLDAEHWKERGEEL	
CTCGACA	ACTACACC	CERLLNSQLGTSAHAPPKRRNCDIGSLWVDVQRHLRSL	
GCGACAG	GATCTTCTC	GLQLQTAPADTHTGAPAQPLQLRVPHHDKWLTHKDVL	
CAGCGAC	TACACTTTT	RHVKLHIKNNHWHRWTSMRDQGKTARAHGGEGSGFL	
GACGAGG	CTAATTCGC	TQPRGMWEADYRFAVAGRLNQVDTYSVLKRRRLRSHD	
ACGCTCA	CTCCGTCTT	RCRQPGCHRAETLAHVLNHCPGTMDAVRGRHDGALKR	
AGACCCC	CGGTCTTCG	IERELHASATDRRDRVELRVNQTVPSLAGPALRPDLQLY	
CACGCCG	GCTGTCGG	NHTKKTVAVVDLAVAFEEQASDDASSSALSLIASHKRA	
CCGCCCC	ATTTTTTTC	KYDRIKRHLERQGWKVHLSALVYGSLGAVASGNYQVY	
AGGACCC	TTTTTGACC	TTHLGLLKRDAKRLDRQLSVECIQSSRRIWNLHCSQHRT	
CCCGCAG	AATCAGAG	RQHQARPSQGPRGSRATETGGTPSQTSRR	
ACGTCGC	CGCGCCAT		
GAGTGTG	GCGACTCTT		
CTTGCCC	CTGGCCAAT		
CACCCGG	CAGAGACC		
CAAGGCA	GGGCCCTGT		
GGCAGC	CCTCGGAC		
	AGCGAGGC		
	CTCCACGGC		
	CAGCCAAT		
	CAAGTCTCG		
	GCAGCGAC		
	GCGTCTTTC		
	TATAGCGC		
	AGCTGACG		
	AGGCCGAT		
	CTGGCGGC		
	CCCCGATTG		
	GTCCGACTT		
	TCGGCCAAT		
	CAGCGACG		
	ACGAGGGG		
	GCAGGGGT		
	TTACACTTT		


TGCCCCCCGT
TTCGACTTC
AACTTCAG
GCCAAAAT
GGCGATTTTC
GACCCTCCA
CGCGCCGT
GCCACTGCT
CGGCACCG
GCGGCGAT
TCAGCGGG
TGCAACTTC
GGGCACGT
GTGCAACA
CATGCAGC
GCCCATTGC
ACGCCAAG
CGGCATCG
CGGGACGA
CGCCTCGGC
CGCTCAAG
CGCAGCCC
CGCCCTTCC
AGCACGAC
CTCGCGCCG
TTTGGCGGA
TCGCCATCA
AGACGTGC
GAGAGCCA
GGCGGGGT
CGGGCAAA
ATATACTTA
CTCTAAGTA
TGCCCGAAT
CCCTGCCCT
CTCAGGCTG
AACGCGGC
CCCATACTT
GATCTAAGT
ATGGGAGG

				GATCCCTGG CCTCTCAGG CTGTACGCG AGACCCGT ACGGCCGA ATCCCCTGG CCTCTCAGC CTGTACGCG GGGC					
			CACGTTG	TAGGATGC	MWGGHRENLEGRALAELIDSYNLCTLNDGSPTMIPRPG			
			TCAGAGA	AGCTAAAC				
			CGATCGT	TCCATAGTT	ARAQPVDITLCSPDLMSELSWEVIQEPLGSDHLGILVKSF			
			GAGGATG	GCGATTTTA				
			GATATGG	GTTCTTAAT	NGRQATSGARIRHPKAGGWNLSKADWEKYRMVLDLK			
			TGGAAGT	TACTCACAC	FSTSESTSCTYDYLTQEIRSAALGSIPRFPQWLYLNPNRA			
			GCCATAT	AAGTACTAT	IKPWWDDDCNRVKQESIIAYESYKRAPTLENYVNLNRK			
			TTATCAG	CATAACTAC	LAISRRTYKEKKRIMWYEFCNSLDRQTPSNVIWKRIRSL			
			TAAAAAA	ACTACTAAT	KKGYSPQVEVTDDGHWVEEFHKSLTPDYASQFIDLSNIP			
			TTGCCTTT	TACACCACT	IGDEDLNXPFQLWELKYALKKGGSTSPGQDGIHYYQMLQ			
			CGACCCA	CTAAGCTCA	NLPfPALLSLLRIMNEMWVGGAMIESWRTHLIFPLLKPG			
			ATAGTGC	CATCAAAT	KDPNSYQSYRPIALSSCVCKTYERMIKHRLDWWVEEYN			
			TGAATAA	AACGAAGT	KLAPHQLGFRRHRGTLDCLAAFVNEVHLNRTENKPLLA			
			TCCATTA	CACATTCAG	AFVDIHKAYDNVNIAILKDKMLVMQVPGRVIVNIINML			
			AGGGGCA	AGACCATT	QQRQLIIRRGVLVTGPRTTGIGLPQGAVLSPLLFNIYTSD			
			GAGTTCA	GTGTTGTTC	LRDWLVPEIRIVQFADDFTVYTGGKSEENLVEIMGQAL			
			GATATGT	TAGGAAGG	HDLNAWCNHLGFRISGNKSSVTMFSRKRQNILAGTIQL			
			GGGGCTA	AATGTCAGT	GPFKVSQSPIVKVLGLYVDHKLTWRPHIKRIIQSCEKAL			
			AAATCCT	ACGCAGCG	NVLRYVAHVTYGTDPKIALLLYRNLILSKIDYGCGIYGD			
			ATTTAGG	CTCAAGAT	ACQTSLSGLDKVQYKAIRICLGAMRSTPTNVLLVEAKIM			
			AGCGTCA	ATCCTTATC	PLHIRRELLLGKFLSKRLIKFKEVSLMRVSELSEVVRSHR			
			GTATTAC	TCATGGAA	FWLKKKTPLMCEMFRVTEKLYSGLEVLHASPRRSLHYE			
			GGTCCTT	GCCACGAC	VSCKESPYVHLDLYDKVKEEWGDLTYLAFIARMREIIIN			
			AGCACTT	GCAGAATG	PYLLFTDGSKSEQGVGCAVLDPQTGSFRLYKLPNNASIF			
			ACTGCCG	TAGTAGAA	TAEIVAIWNAVILAEQIQSNDIIVILSDSKSALQKIKRPGD			
			TCCTGGG	GAGTGGCT	VFDPIIIEIISHLINAMGRLHLVWLKGHAGIVDNEKVDVL			
			CATAAAG	GGCATGAC	AKRAILLGEHTPVLNLADFEGALRKRCYDAWQKLYDL			
			CCTCAGT	GAAGAAAG	EKEIKGKYYADLQPRIPNKPWFVEVLEGERQFYTTIMRL			
			AGAGGAA	AAGGGAAA	RFNHNLTNPHLHKIKLLDSPKCDCDSGEDGDIDHLLFRC			
I	I-2_RPr	.	TGGAGTA	GAGGCTGC	SRYNRNRDKFLKELATHVRFPCSAKSILAIMNPDALRIIY	cd01650	340	595
		Rhodnius prolixus	ATGTACT	GCTATTCCT	KFIRENKIKI			

				AAATCAG GTTTCGG GTCCTCTT ATTCTTG GTGGCGA CTTTAAT GCGCTTC ATCCA	CAAATTGA ATATTGGA AAAGATAG GGCTGGTTT TTTTCCACC AACATATTG GTAAATTG AGGCGCTG GGAGCCAA CTGTAATTG TTCATGCCA ATATGTTGG TTTTTTCCA GCCCTCATA TAAAGTAA TTGGTTAAT GTGCAAAC AAGACCAC TATAACAAT AAATAAAT AAATAAA				
RTE X	RTE X-2_RPr	.	Rhodnius prolixus	GAATCCA GAAATAG AGTATCC AAAGATA AATTTAT TAATAGT AACGGTT ATAAACT TTTGGAT TTTTTCGA TAGCTTC AGTCTTA TAGTCCT TAACGGT AGATTTG AAAATGA TTTGGAA GGCGAGT TCACCTA	TAGTTTATG TTTACTAAA ATTTACTTA TTTATTAAA ATATATAAT ATA	MVSSFRVDHQLYSDHMPIITTLLIHDKQQIYNDNTLPLLP KLNWSKQNLQKYQQSISFYLNHIVYKYDTLDAEVNYLN DIIRSSAIGNNNKKGPKVWKADWFDVECDRARRKSFKQ LRLFRRTNEVEDKHLYLKYNFDFKRLCKNKKQKYYVD CARRLYNVKSSKEFWQQVRAFQKRKPWVLGNVSPEQL FEYFCSLHNSSVTARLVFDVSHRVDAFLDKPFSFDELEL VLSKSKENKAPGTDRICYEFYKFAPIVFLNRLLLLFNKIF ETSKIPASFKESIIFPLYKKGNKNSASSYRGISFINVICKLY SGLILHRLNNWLESNNILKESQAGFRSNYSTIDNLFNLTN IIKIKLSKKKQKVYAFFVDLSSAFDSIDRFSLIYKLHRTGL STKIVTSIKNLLSGTTASVMCRNGLSPNFGTNIGVKQGC LLSPTLFSIFINDLEECLGGGVVISDNCKIKLLAYADDVV LLASNPISLQLMINEMESYCYRWKLKINLQKSKIMIFRN GGRPANNEVWYLGTQKVEVVNEFKYLGIILTPTLSYTK HFQNKLSKAKMSLNTIWHHFMDNTSIPFSTKIRIFNSVN RSMMCYGAQVWGWQRFVEVERLFYFFIKKLFKSPKTTP SAFIHTETGIDPIYLYTFEMHFKYLVRVLQLPDNRYPKIL CHEIIKSKILWFLEIKNLITKFNLPNDLNWQNIGNFTLWE ANFNILIKSLQADMNNTFQLKLFNSGSHPLYKNLQYNL	cd01650	280	540

				CATTAGC GAGGTTG GCTGTTC GGTAATT GACTATT CCTGCAT TAGTATT AAAGGTA TTGAC		GDNPYLSDSHPFHLIRMIFKVRGDLLPLNYKFYKNRDLD CLMCASGEVEDLFHFLAVCPALSNMREQWFGSRTLPAQ HVAAYLNGKDWEALAKYCLAALSRRALLT				
					TAGTACTGT	LNNKVNVEDKSFSQTIVNDISSMESLDDYKHIINQDFTN				
					AAGTTTTTG	KEVLQHIKLLKSGKAASGTDMISNEMLKHGAPFIINSLTK				
					TTGTTCTGA	MFTQVLNSGIMPLHWNSSLIIPIFKSGNPKDANNYRGIAI				
					ATGTTTTGT	SSCVSKLYTMMLNTRLTKYLQETNQIVDNQFGFKKNHR				
					TTTGTTATT	TTDNIYILKSLINKYVVNCKLKKSGNYLFSCFIDFSKAFDS				
					GTTATTTTT	VWREGLLYKLLRLGVGGNFYKAIKAIYKNTKYSIKLSA				
					GGACTTTAC	GITPYFNSCRGVKQGCNLSPQLFNIFVNGICDDVKNFDP				
					TCAAGATTT	VDLNGLKLNCLLYADDLLLFSTTQEGLQKCIDAVQNYC				
					CATTGGTAC	DNWHLKVNINKSKIIIFNRLGKNLENKFQFYLDATKLQC				
					GTTCCAAAT	VNYYNYLGIVFQTNGKFHNAMSRLQDKAQKALFKMY				
					AAAAAATT	KMLWISEECYSVSLAKKLFNSIIKPTALYGADVWGGVM				
					CTATCTATC	LNIYKGSGKNKCTVDNCNFETKLFSDKLIPEKLHIKFCK				
					TATCTATCT	RVLGVHRKASNDPCRGELGQYPLILDICVSVMKYYQRL				
					AT	VCKKKTILSLMQPMKNPKLLCMTHGLTLYKNYKKNNSI				
						IDAAYEESKTLVHDPWLNFIQKLLVITDLTNSDPHQNNF				
						CDIFKEKLKCLYRMKWKNNLLNNCKLDFYRCAKENFE				
						FEHYLEKIKNKNHRIAYTKFRTSAHFLPIESGRHKGILKE				
						NRLCNLHECNEVGDEKHYILYCNNCQIKEQRSRFLAAL	cd016	96		361
						YIYAPKLEDFSMDELFPIIVSSNDFFILSIFAKFIFYILDMF	50			
					TAATAGTTG	CINDSLLETFVSKHDIVCLLETWLDYTDSVSVPNGYLYF				
					TTAGTGTTA	RVDRKKHKNARKASGGMIIFYKSKFRKGINKFISKYKES				
					TCTGTTTTA	LWLKLDKHFFGLENDLYINSTYIPPIGSSYYNSSDLTGED				
					ACATTTTAT	IFNIIAEESLHYSNHGNVCILGDLNSRTVNLNDTLDDINIS				
					TGTATTTTT	DIDIPYTELMDNNRQSMDTMNNKYGQYLCDLCLTTDMI				
					TATATCTGT	IANGRIVGDLMGKITCFQYNGVSQVDYCVVQKSFLPKIN				
					ATTTCTTGT	YLKINELSHISDHCSLSLAINTNYVNEHSIDISLTPYKKYN				
					TATTGTGTT	WNNSSKENFVTQLNSKDSVDKMNTFLDNHFNLEESDK				
					GTTGTTGTT	AISDFNDIMHFTADKALKFTAIKSRKFSRKNKRKNKQK				
					TTTGATTAT	WFDNDLAHVRKCLKDISRLVDKFPFFDQNLRMQRMSIN				
					TGACCATG	KRYKKLVKLKKNKFKEKLLHDLMNLETKNPSEFWDIIK	cd016	551		812
					GTTAATGTA	QIKEYNDLDTSKHPPSIASTEWTNYFSKLYNISTTDSNSV	50			

RTE	RTE-1_SK	.	Saccoglossus kowalevskii	AACAGTG	TTCCCAATT	LDSEVTNFVDNMDHNATSNPDLDKEISIAEIIKAIKKTKV				
				GCTATGC	CATTAATTT	NKSSGPDSILNEMFKCYSPLLLACISKLFNHILSSGCFPV				
				AATAAAC	GAGTCAAA	HWSTSYLVPIHKSGHIWDVNNYRGIALSSATSKIFNTIIHI				
				AGTGGCT	TAAAATAT	RIQHFLDKNNFFQDTQFGFRKNHRTTDNVYILKSLINKY				
				ATGCAAT	ACTATACTA	TNKVGKHYLYACFVDLRKAFDSVWREGLLFKLANAGV				
				GAATAGT	TACTATACT	TGKIIHLIKSMYSNTKYAIKTNQGYCKPFMSSSGVKQGC				
				GGCTGTG	ATACTA	NLSPLLFNIYINDLSESLVQTEPVRLHDMSFNCLLFADDI				
				CAATGAA		VILSNSGSGLQDSLNLFNDYCKKWKLTVNIKKTKVIIFN				
				CAGTGGC		KRGRKINDIFFLGDEQIEVVSSYTYLGITFQNSGNFTVAV				
				TATGCAA		NNLVKKAQRALGVLKQTLLLNNIASPINAIKLFRSSVM				
				TAAACAG		PILLYGSEIWAAFNFKPNITPELFIQKLFSDSSKIESIHIKFC				
				TGGCTAT		KQLLGVHKRASNIAVRAELGSYPLITEICTQMYKFWLR				
				GCAATGA		VKNKECSNIVHCAYQSECTINSNSWANFVSKLLEITRLN				
				ATAGTGG		ETDHNANIFRNNLKRIYYEKIWLPTLSKKEHNKLRIYNK				
				CTGTGCA	TGAATCCG	YKSDYEFEPYLNKINNLNHRVSVTKFRISAHKLPIEKLR				
				ATGAACA	GCCACCCTC	YQKIPSEQRFCTLCSDQSVGDEMHALFDCKLDEIVKARS				
				GTGGCTA	GTCTTCGGG	YFWNTLSLYNDNILKLPKEDMFVHIMGIKDSFIANSLGK				
				TGCAATA	GACCAAAC	FINTILQCNT				
				AACAGTG	CGGATTTAA					
				GCTATGC	AACTCTGG					
				AATGAAC	ACTACTACA	RGEGSNSKKERCIQENAKSGTEENKVRYKRLKKQAKK				
				AGTGGCT	T	VVAKAMREENEKVLSELKENPNKVYKIVKAMKGDGRD				
						VTGGRCMRGSDGKLNFSERDRSKVWKEHMEKIMNEEN				
						EWDQDVEADLVEGPVERVSREEVVKAIRDMKTGKAAG				
						PSEVSAEMIFASGDIGIGVIMELCQRVLDGRGMPEEWEV				
						SVVVPIFKGKGDAKSCGAYRGVKLLEHAMKIVERVIVN				
						RIRKLVEVDEMQFGFMPGKGTIDAVFILRRMQEEYRAK				
						GKKLYMCFVDLEKAFDRVPRKVMEWAMRKRGIPEAM				
						VRAVMSLYEGAKTRVRVGSELSEEFEVKVGVHQGSVLS				
						PLIFAVVIDVVTESAREGLMREMLYADDLVLMDDTME				
						GLREKFVKWKKAFESKGLKVNLGKTKVMVSGTEEKAL				
						VSKVDPCGMCGKRVKVNSVRCVKCGKWIHGRCAKIKR				
						VTQRLARDFACGRCKNGVGRVVEQEEKLCDEVESVKE				
						FTYLGDKVSVGGGCEAAVTARARFAWVKFRECGELLY				
						GKRFPLEMKGIVYKSCVRSALLYGSETWCLSENELGILR				
						RTERAMVRVMCGVKLMDRKITEDLMQMLGLKEAVDR				
						LAKANGVRWYGHVLRREDGHVLRRALDLEVSGPRKRG				
						RPKKTWKMQVEEESGKVGMRMRDALNRAKWREGVR	cd01650	188	431	
						AIATSLR				

			ATGCAAT AAACAGK GGCAGAG GAGACAC AGGCTGG TGGAATG A							
RTEX	RTEX-10_SK	.	Saccoglossus kowalevskii	AAAGCAC AGATACC ATATAAT TAAAAAA CAATATA CTAG	TAATTGTAA TACGTATTT TATATATAC CCTTTGATT TGTTTGTAA TTTTGAAAC TTTTGTTGT TTTTGTTTA TTTGTGTTT GTTTTTTCT GATCTGTAT CCTTACCAT ACAGCACC CACTGATG GTGCAGTTG TGTAATAA AGTTATTAT TATTATT	INKNKKNFFKNEMINKLETCDKTDQNEYWSLLKQLRCQ QNSLNPDSTINSDEWLDHFSKLYNNNNNNADFKQSLKL LEDTKLSNHELNNVISYNEIKKAIKNLKNKKSTAEDSISN EMLKTKNSLLIPCIQKLFNFILESGTYPTSWKSTILVPIHK SGDRSNPSNYRAIAISSCLSKLFTGIMNCRLTNHLEKNKL MNEVQFGFRKNRRTSDAIFLLKTAVEKTLYGKKETGAK CKRLYSCFVDLKKAFDSVSRDALFLKLQKTGINGNFYNI IKSMYTDQEYKVKLNSGITKGFHPTSGVKQGCNISPTLF NIFMNDLPGIFGENVDKPIMIDTFPLNCIQYADDLAIISES VDGLQNCLNKLNDYCTKWYLNVNVRKTKVMIFNNSN RLIKNMNFKINNEEITSVKEFCYLGTIFTPNGKFKINQSN LKNKAMKALFSIRHTILNHKMISTKLSLKLFDQLISPILL YGSEIWANEITNENNYLEQACNHFYRFILGVSKVTPLIGI RGEIGRIPIRIKATLSLIKYWHRLSLLPENHILKSALKETV SSNSSWIKHVKTILGHTDLSEVNKIYKVGNIAELTKLLEE KLKNKYIQTWYQQLSHVPKNGKSGNKLRTYNLFKMEY EFEGYLTEIQDFNLRRMVTKFRLSDHNLEIELGRRKFPR MEANYRLCKKCNTGTVEDEKHFILDCCFFTQERNTLLE DCKLEMCSTTNRNLIFTCIINSSTALAKYLKNTIYKT	cd016 50	150	418
RTEX	RTEX-6_SK	.	Saccoglossus kowalevskii	AAGTTAA ATCTAGT GAGTACT TATGGTG TAAAATT GATAGTA ATTTAG	TAGCAACTC CATCAAATT AGTCTGCA GGTAGTTTT AATGCAAT GGCATATTT AGTAGCTTA GTAGTTTTA TTGTTATAC TAATTTTCA TGTATTTTT TCCTCCATT CTTAGTTAC ATGTATTCA	IFIYVCLYPPENSVIHNRQDIDIFEKLSQDVIKYSNIGEVL LMGDFNARTGTLADHIINDNSTYMPIPDNYVDDEYNNR NNKDQKVNSFGKKLLELNKNNNLLILNGRTPGDLLGKF TCYNFAGGASTVDYALVSKSFLSNIQYFRVNEINNLSDH CAISLAITTRDLLSSDNNEIKCTPLPDRYLWDTKSIKRYQ DALHSTEIINKLVSFNYYENNNIXNSVTELTNIICNAADNS LKKIKHKISSSNKMKNTKRQKWFDDDCWNLRKKFREA AKLVEKYPKDCHIRGQFMVLRKKYNKILKIKRKQYKQN LWDSVTSQESNPKQYWKNFKIFSKNDTSISANNDIIQAD EWYNHFKNLNNLKNSKTCKYDSDFQNKIINETREMETN KEINKELNEPITIGEVKAAVKKLKLKNKSSGEDKIINELIK YGQNYLMEPITAVFNNVLKSGTFPKGWKGSIIVPIHKKG SRSDPDNFRGVALSSCLGKLFTATLNNRLIKYLEKNNLY SPYQFGFRRNRRTTDCLFILKSIVDKYIHSEYSKNSKSKK	cd016 50	458	726

					ATGTTTTATTCGTGATTCTGGTGGATTGAACTGTACTTCCGTCTTGCTTTGTACTAAACAATAACATTAGCTTATGTTTAATGTGCTATTTGTGATATTAGTTGTGTGAAGTATTATTTCTGTTCTTGTTTTACAACCATAAAATTACTCTCTGTAGTAGATTTTCGTGTTGAATAAAATTATTATTATTATTA	LYACFVDYKKAFDSMWRDGLYWKLQKYGITGNIYNLIKNMYSGVSYKVKTSKGLTNEFDSFMGVKQGCNISPTLFSLFINDLPEKIKSANGTPPELNSQKLNCILFADDVLILSQTQEGMQCSIDALAKYSNSWKLSVSLNKTKLMVFNKRNHTLNKKSVFYEGNEIATVDSFNYLGIVFTPNGKFKTASNTLKEKAMKAMFPLKQSLLTLSHYKISTALNLFDKLIVPIINYGCEVWGEESAKHSNIIESVQFAFCRFLLGVSKNSTKNAILGELGRYPLNVKIETQMLKYWLRLKSPVIPNIVATAFMEINTWKTGWVERIEEITKNNNLSELSKQHEKKIGEVLRENYVKTWRRELFNDSRMNGATNKLTSYRKFKTTFNMENYLNDIKXLKLRRLFCKFRISDHKLLIETGRQKKIPRNERICKLCTDDKIEDEYHFLIGCSKYDIPRNKLFKYIQGKKSDFLKMNEDEKFRLIMTSFAGIETAQFIHTSMNLRDET			
RTEX	RTEX-13_SK	.	Saccoglossus kowalevskii	AT	TAATTAATATAACTTATTAAGATACCTTTTCTGTTTTGTTTTTTCTAACCATTTCTTGTTCTTCGTTTTTATTTGATTTTATTATGTTTTCACTWTGTCTTCATTTGTATTTTTCGTCTTGCCATTCTATATGTATTTGAAACTTTATTGTATTTKAAA	TDSQDINTNXKHESNILNKLSTDSKELNEPISQLEIQKAINNLKNGKSPGKDNISNEMIKAGRHTLIIPVQKLFNMILTSGFFPTNWNTGIITPVFKKGDKSDPSNYRGITLLNTLSKLFTAVLKQRLIKYIDVNNLIVKEQIGFIPKFRTSDHIFVLKQILAKYLEKGKKVFACFIDFKKAFDSVWHQGLFHKLAQNGITGKFYTIIRSMYSESLSCVKTKNGLTSEFPVKRGIRQGDGLSPLLFNIYVNDIKNEIEQSNNDPPPPSLTVSQLSCMLYADDLLLLSVSESGLQNALNSLGTYCKKWKLEINTTKSKIMIFNKMGKHIKKHKFCIDKQELENVPTYSYLGIKLSASGKFYKAQEDLANRAMKACFKLKQQLFSETNIPIKIYLKMFDSVIKPVLLYCCEIWSNENINIKNLSNFNLPMEKLHSKFCKFLLKVNCKASNIACLLELGRYPLAFSSIVQTIKYWLRISDKPCNTLIYEAFRLGKDMDSKGIKNWVSHAKNIFNITNMDHMLDTVTADKKQTIITSLESNLKSMFKNHFKETLYNDNKGKNRKNKLRTYRMIKHNFDFEKYLEDVKNPSHRSALTKLRISAHTLQIEAGRYINQTVEERKCHYCQEKQIEDESHFLLKCSLYADLRNKFLHTTKVITTIDKSDIENTVDILKTKDKQSLIHLAKYIFECFNLRNDTHKQKIPAAL	cd01650	89	353

						AATCCAGA TCCTGAACA GGATATGA TTTCAATAA AGGCATTA CTTACT	LYNXYLISIIDKGHRDILGVRFENKLTGFTFLVFSCYLPPE NSPWGRDANAFFSHLLSQLYMTDDVDTIIICGDLNARLG NLNDNIDVVDNIRNRHVIDTKINQHGHSFCEFLQESKLCI LNGRFDRHQNDFTFISSRGRSVVDYVCVPQDHLSYCNN FNVITMSSLVEKYNLYDALTDRSKLPDHSLLIFDFLDRPL GSIVEQSLPNDINIGNNRHSATVRKYCLNQTPINFLESEV AREALLHIISKIEICRETQQEIDEIYDDLCKCILNEMNSTIP NYNCSSGARKRFRHHKPYWNDELYSMWRVMRERERS YTRFGGSKNIRTILRNDFKAARRAFDKKLRCCERNYRN			
						TGAATTTAA TGAACTAC ACGATTGTA TTGTTGATA TATTATTTC AAGCTCTCT TTCATAACC TTTATTTGT GAATGTAA TCGTTATTA TTTAAGGTG ACTTATAAG TCCACTGGG CTGGGCAG TTCTCTGAG CTGCAGGTC ACTATAATA AATTACTTA CTTACTTAC T	DLVSSIEKVCTENPHQFWEHINKLGPVKRKDIPMEVYD DGGHITSNHNIVLNKWKHDFSSLYNQSDLDGFDANFYE HILHHKLVLENDLDDPLYDSNSELNKNFSTVEIKTVVAK AKSGKAVGVDKIPYEVLKNDSSVAILTHLFQLCFDSAK VPCIWKKALICPILKVRSSDPRVPMNYRGISIQSTLSKLY SALLNNRITRFLENNNILVDEQNGFRKDRNCEDHIYILTS IIRNRISHDLSTFAAFIDLRKAFDSVDRDALLYKLLINGIN GKMYKSIRAMYSDTNSSVKLNDFKTDWFNTEQGVRQG DNLSPTLFSIFINDLAVEINLKQKGITLDNNVNVSILLYA DDIVLLADNEIDMQMMLSHLNSWCEKWKIRVNNVKSK MLHFRKASEARSAFQFCIGDDVLDYTDKYRYLGIVLDE FVDYSVTAGIAADSASRALGALFVKFKTLKNLGFSTFTK LYDASVAPILDYCSGVWGFSKFEKCEMIQNRALRFFLGT HRFSPLPAVKGDTGWVSCRDRRWAHMVRLWKRLTEL DDNRLVKRVFLWDHSLNYNNWSSEIYQIMELVDLNYE HLIDVNIHALRENIARRSESSWSEDVSHKPKLRVYATFK FEYKPEKYVLINLSRGQRSLLAQLRMGILPLRIETGRFRH EPVSERICKFCDANVVEDEMHFLFDCQQYNELRHEYLD SILNEITGISREQILVRLFSGPIRKLALFIEKAFYKRRNLLY V	cd0165 0	514	778
RTE X	RTE X- 22_S K	.	Saccoglo ssus kowalevs kii	A	CTCTCCC GAGCTTT CGCTCGC	TGATGTATT GTTGTCTCT ACCTTCTTG	MTPACFVDYKLSLPNRGTDYVCSHTRDSDSLLVTQTFG LPSYSNHLSPALYSCYLMKLLYNMILLPSDAHSDVINQH CRALPVLCRALIVLLLMISGNVHVHPGPSTVASPNSDLC			
Crac k	Crack - 3_SS a	.	Salmo salar		CTCCAGC ACACACA	CCCTTTGTG CTGTTGTCT	SDICFTDFCSRKSLGFLHVNTRSLLPKMDQLKVWVHSS NPDVLVITETWLRKSVLNTDVNLSGYNLFRQDRSSKGG	cd0165 0	627	884

Tx1	Tx1-2B_Sa	.	Salmo salar	GT	CACTGTT GGGGCGA GTGACTC TGAACTT ACAATGG CTAGCCC CAAGTCG TTATATA TTTAATTT TTTCTTGT GCATTTT GCTATTT GCCTC	GTGCCCAAT AATGTTTGT ACCATGTTT TGTGCTGCT ACCATGTTG TTGTTATGT TGTGTTGCT ACCATGCTG TGTTGTCAT GTGTTGCTG CCTTGCTAT GTGTTGTCT TAGGTCTCT TATGTGTGT TGCTCTCTT GTTGTGATG TGTGTTTTG TCCTATATT TATATTGTA TTCATTTTT TAATCCCAG CCCCCGTCC CGCAGGAG GCCTTTTGC CTTTTGGTA GGCCGTCAT TGTAAATA AGACTTTGT TCTTAACTG ACTTGCCTA GTTAAATA AAGGTTAA ATAAA	GVAIFTKDHLQCSVVSTKSVPKQFDLLVLSIKLSNSSLLT VAGCYRPPSAPACTLPALSSLLAPYTKSEFVLLGDLNWD MLKPPDQVLKQWDSLNLSQIITNPTRYDSKHPEKATLLD VILTNNPDRYQSGVFCNDLSDHCFTACVRNGCSVKRPV LICHRRLLKNFNEQAFLHELASVKWYRISLIPSVEDAWT FFFDIFSGIVNKHAPIKKMRIKNRFSPWFDRDLAELLHLK NCIWRKARHTHTQADWLSFRQMRNKCTQAIRKAKVSY FKEQFSLCGSNPKKFWKTVKDLENKPSSSQLPMSLNVD DVVVTDKKHMAELFNHHFIKSGFLFDSAMPPCPSNISSS PTPSNATIPDASPSFSPAPLQSFSLQAVTESEVLKELLKLD PKKTSGSDGLDPFFFKVAAPIIAKPISNLFNLSLLSGEVPI AWKAATVRPLFKGGDQADPNCYRPISILPCLSKVLEKLV NNQLTGFLDVYSILSGMQSGFRSGYGCVTATLKVLNDV TIALDSKQCCAAIFIDLAKAFDTVDHSILVGRLRSIGVSE GSLAWFANYLSQRVQCIKSENLLSQPLPVTKGVPQGSIL GPTLFSIYINNIAQAVGSSLIHLYADDTVLYSAGPSLDFV LNALQQSFLSVQQAFSTLNLVLNTSKTKVMWFGKKNA PLPTGVITTSEGLELEVVTSYKYLGVWLDGALSFSQHIS KLQAKVKSRLGFLYRNRSSFTPAAKLTLIQMTILPMLDY GDIIYRSAGKGALERLDVLYHSAIRFATNAPYRTHHCTL YSSVNWSSLYTRRKTHWLMLIYKTLLGLTPPYLRYLLQ PSSSTYNTRSASHILLKVPKAHTSLGRSSFQFAAASDWN ELQQTLKLDSFISISSFKDSIMDTLTDSCGCFV	QRNSKGVMVLINSNFDPNVVQIVQTDHQGRWIILNMLLD NKQIWLINLYGPNNDDPSFFENIYKNLSTLQATLDSIIIMV GDFNTVLNTSMDRKGNHTTNYHPQALKEIMNVMDILE LVDIWRLKYPDLVRYTWRRLNQASRLDYFLMPFSLAPK VKVLIGDRMRSDHHIIGIYITLTEFPRGRGYWKFNQSLL DDKLFRTRTEEFITDFFRHNIGTADPLIVWDTFKCAFRG HAIQYSSIKQKQFRSKESILTKEIEGLTVQLDSNKNGTIE	cd01650	457	723

			AQNKLEEKQKEMEELIQERSSVIYYKNKANWMEYGEK	AACTTTTGT		
			CTKFFFNLQYRNATKKMYKCYQKNVLKLVTNDGVTHD	GAAATAGC		
			SPNXILKEEVKYFKNMFSFQAPPSPLTETNCMDFFPNNN	ACAGTTAC		
			VKLTSVQKDSCEGQITEEELLDAIGAFKDGKTPGLDGIP	AAATAGAA		
			VEVYKTFFDILKGPLLACFNHSYINGRLSDTQQEGLISLL	ATCAAACT		
			LKQDPSGIYKDPVHLKNWRPLTLQCCDAKILAKCLAHR	GGATGGAC		
			IKKVLSDIIHPNQTGFLHGRYIGDNIRQVLETIEYYEISGT	ATCAGAAA		
			PGLVFIADFEKAFDKVRLEFIYKCLEYFNFGESLIKWVK	TAGAGGAA		
			VMYSNPRCKIVNNGYISESFKLSRGVKQGCPLSAYLFIIA	GGACTAAA		
			IEMLAVKIRSNNNIKGLEIRGLKTKVSLYADDSCFLLKP	AACAAACA		
			QLESLHGLIEDLDTFAILSGLKPNYDKCTILRIGSQKNAN	AAAAATAA		
			FTLPCSLPIKWSDGDVDILGIQIPKERNDLTPIHFYRKLA	CTATTGTAA		
			KIDKILLPWKGKYLSICGKITLINSLVISQFTYLLMVLPTP	AATAGACT		
			SDLLFKLYEQKIFNFIWNGKPDKIKRAYLYNEYEFGGQK	GTGTCTGTA		
			LNIKALDLSLKASVIPKLYLNPNWFSSKLVRMSHPMFKK	AAATGTGT		
			GLFPFIQITPAHFRLFEKEIISKISLFFKQALESWLQFQFNP	ATAAGATG		
			PEKTEQIIQQILWLNSNIVIDKKKIFEEMFKKGIILVNDIIN	AATAAACT		
			RTGGVMSHMQLTQTYGNVCSTQNYNQLIAALPQKWKR	GAAGGTAA		
			QVEGEKSKELVCRPCIKEHKWLKKSVINKNIYQFHLRTK	AAGCCGAA		
			KLTAVPYKLQNSWEEIFDVPIPWHMVYELIRKTTPDSKL	GTGTTTATT		
			QIFQFKLLYKILATNRMLYIWGIQSSQLCRFCCEEAESLD	AGTTTACTC		
			HLFWYCPYVARFWSQVQEWLKNCNICLELTLQIAILGD	CAATTGGG		
			LKSHSQSINNIIIILAKMFIFNLQSVEAMRIERFNCFVKHH	GGATCGGT		
			STVEKYMANRNPKWMMLRDRWEGLNGAEGWD	GGTAGGGT		
				TTGCGGGG		
				AATAATAA		
				TAAAGGTA		
				TATTCTTTA		
				AAAAAGTA		
				TGTATGTCT		
				ATATAGGT		
				ATGTGTATG		
				TATATATGT		
				GTATATGTA		
				TGCATGCGT		
				GTATGGAT		
				ATATATATT		
				TACCCCAA		
				AAATATAT		
				GGGGGATT		

Tx1	Tx1-2_SSa	.	Salmo salar	CTGTATC CGGTACA AAGACAA CAAAGGA AGGATGG TAGGCGT GCCTTTA TCATTGT GGACAAA CTCTTTAT AGATGGA CAGCTAT TCCGAGA CAGCTCC ATAACAC CGTGGCT GTACTAA ATTCTAC AGGGACT TCAGGTT ACGAAAA AAAGAAG GTATGGG AACTGTA AAAATAT CTGAACA TTATGAA GTGGATA TGAACAC ACACGTA CTCAATT ACATGCT TGCTTAC ACATACG GACTTAT	GGAAATGA TGCAGACA ATTACATTG ATGGAAGC AATA TGAGAGAT AGAATGAT GGTCAAAG ATAAAAGT AAAAATTT AAAAAAGT AAAAAAAA CATAATAA AAAGTACA TTTGAATGA CACTAAGT GGCAGTGTT TTTACAACT AATGCCGG TTTGCCTGA GGCTGATG CCGTGCCG GTGTTTGTA CACGTGCAT ATACACAC ACTCTCATT CAAATAAA CAC	MVRGEISIASYNCNGLADYKKRRSVFTWLKEKEYNIYC LQETHSTSLDEVAWKKEWGGEIIFCHGQRNSKGVMILIN KNFDLNVQIIRNDSQGRWILLNMKVDEKEIWLINLYGPN QDDPHFFENIYTNLLNLQATNDLIIMVGDYNTVLSTSMD RKGNHSTNYHHRALKEITNIMDTLEIVDIWRLKNPDLVR YTWRRLNQASRLDYFLVSFSLASKVKKVLIGDRMRSDH HLIGIHITLIDFPRGRGYWKFNQSLLEDNLFLTKTKFITEF FQYNIGSANPLIVWDTFKCTFRGHSIQYSSIIKKQFLAKE TRLTREIHELIVQVDSNKNDTTEIQNKLEEKQKELEELIQ ERSNVIYYKNKANWMEYGEKCTKFFLNLQYRNANKKN LQKLVTEDGVIYDSPNXILKEEANYFRQMFSFPSHPLPL NEDYGKEFFPNNIKNGKLTNVQKDQCEGQITEEELFEAI KSFQSGKTPGLDGIPVEVYQVFFDILKAPLLDCFNYSYR NGSLSGTQQEGLISLLLKQDPDGKYKDPVYLKNWRPLT LQCCDAKILAKCIALRIKRVLPGIVHPDQTGFLHGRYIGD NIRQLLEIIEHHETYKKPGMVFIADFEKAFDKVRLDFIYK CLDFFNFGNSLIKWVKIMYSNPRCKIVNNGYFSESFELS RGVKQGCPLSPYLFVMAIEMLAIKIRSNNNIRGLEIQGLK TKVSMYADDSSFMLSPQARSLQCLIKDLDNFSVLSGLKP NYDKCTILRIGSLKNTTFTLPCSLPIKWADGEVDILGIHIT KDINKLSTMNFNRKLVKIDKILQPWRGKYLSIYGKIALI NSLVISQFTHLLMALPTPDDSFFKSYEQKIFRFIWDAKPD KIKHAYLYNEYELGGLRLLNIKALNLSLKASLIQKFYLN PKWFSSRLIRKAHPLFKNGLFAFVQIAMSHFRLIENYTFF KVSVFFKQALQSWLQFQFHPPEKIEQILWLNSNVLVDKI PVFMGKMFEKGILFLNDIVNWNGRVMSFMELSELYGK VCSIQEYNQLITALPQKWRRRVAAGGGRELVCLPNIKD QNWRRNKNSINRKVYQFHLRTRMLTTVPYRLQNSWEEI FDVPIPWYRVYELIYKTTQDSRLRAFQLKLLYRILATNK MLNIWGIKSSKLCRFCCEDTESIDHLFWYCPQVACFWS QVQEWLKMHSIDLKLTLEIVLLGDLERPGQSITNILILLV KVFIFNSQSVDSIRLDRLKLYVKHHSIVERYVLRRNPKW VASRDRWDGLREAEGWYVVELETSGSGVAV	cd01650	517	783

ACATACA
CACACAC
ACCTGAT
CTCGCTC
TCTTCTCT
CACTCTC
TCTTTCTC
TCTTCTCT
CTCTCTTC
TCTCCCTC
TCTCTCTA
TTGTCGA
GAACTGT
TTTATAA
TTTAATG
TGTTTATT
GTTTGTCT
ATCTTTTT
TATGTAT
TTGTCTA
CAAGTTT
ATATATG
TTTACAA
CAAGCAA
GGGGAAG
ATATCAG
AATAGGG
GCATTGG
GGAATAA
TAACATA
TTGTACG
GAATACA
TTTGTACT
GTAGTGA
AGCCGTC
TCTAGAG
TAATGCA
AGGTCTC
TTTCATG
ATCATGT

GAAACGT
CAATTGC
TATGGAA
ATGCTGG
GATGTGT
TTTTCAAT
TATGTTA
AAGGTAA
TTATGAT
GCAACTA
TGATGGT
GATACGA
TATGGAT
TACAATT
ATCATAT
GAATATT
AAGGCTA
TACGCAC
TACATGT
TACACAC
ATAGACA
CTCAAAC
ATGCAAA
TTGGCAG
AGTTATT
ATTTATTT
GGACATC
ACACATT
ATAGTCT
TACTCTT
ATACAGA
CATCATA
CACACTC
TCATTAT
ATCATAT
CCAATAT
CATACAC
ATCAACC
TACTCAG
ATTTGCT

				ACACACATAATATCTTGTCTCAATTTTCTAACATCTGTGGCATTGGCTCACAGGCAAACAGGCAAGGGAATAAAACAAATATTGCTAGAACCTATTTCTTGAATTCTAAATATCAGACATTTGAAAGCTCTAGTTTTGACCTTCATAATACCTCTGATGGCAGTAACTTTACAAGCACTAATTATGGTCAATTTAGACTGAGAATAGTATCTAGTT					
L1	L1-5_SSa	.	Salmo salar	TAAAAAAAATGTATTTTTTATTATTATTATTTTTTTTGATTGGGGGGTACGTTCAATTTAKCAAAA	TGAATTTGCACTTTTTAACGCACTCTCAATTGTMATATGGTAATCTACCTTTGGGCAGCATGTGCTGGCCCCGCCACCC	WFGALSYLFSDLHAMQATIGTTNERRAGLTFTSWNVKGLNEPIKRGKVLAHLKALSSDIIFLQETHLKNNSHNRLKCKWVGQVYHSNFSAKTRGTAILVRKGIPFLHKTTIVDKEGRYVIVIGEIHSTSVTLLNIYGPNIDNPSFFKRVLALIPDISHTNLVIGGDLNCVLDQYLDRSSTRRTPTSYSSEFLNTYIKNSNLFDIWRISNPTGREYSFYSHVHNVYTRIDYFLVDAKLLPYTCNVRYHDIIISDHSPLTFSLRLGDIVPNERVWRLNPQLLTEPTFCEHLKDQITFFFDTNDNTETSPALLWETLKAYLRGCIISFQTARSRQNRGELVELEGQIHLLDRENASHP	cd01650	535	793

TCCTTGTT	GAGCAGTT	SMEKHKKITSLKFEYNQILSAKIAKSFLYAKQKYFEFGD			
CCSATCTC	GGGAGGGG	KPHKLLARQLRKNVSDRMIHKVKSASGELLSSPKDIND			
CTAACCC	AATAGGGT	RFRQFYETLYTSKADPNPLIMQNFLEDCNLPALNQEDSN			
ACAATAT	GGAATAAG	FLNKEISLEEIRETIKSLKSGKTPGPDGYPGEFYKTFSNM			
GTAATTG	GGGGGGAT	LSPYLHKMLVQANEDGALPSTLDEAFITVIHKKGKDPEE			
TACTGCT	AAAGGGGG	VGSYRPISLLNTDQKILAKTLANRLSTLIGKLVHSDQTGF			
GTCTATG	GTATAAGT	IPNRNSFFNLRRLFNIMYSQRLPNVVDLAVISLDAEKAFD			
TCGGTTG	GGGGGGGT	QVEWPYLFKVLQKCNIGDGFIKWIQLLYRNPCARILTNQ			
CTGATGG	TACATCGAC	SLSPRFNLYRGTRQGCALSPMLFALIIEPLAQAIRSHAAI			
TTTATGTT	CCTTTTTTC	HGYNTKDTLNKISLYADDILLYVTEPQDSIPAILXVINLF			
TAGACCG	TTTCTACCT	GTFSGYRINWNKSELMPIRLQNTSWLEHLPVKLSSEKFT			
GCATTGC	GTCCTTGTT	YLGIVVTKQYSLLFKENFPSLMQKLKANIQFWRTLPISL			
TTAGCAA	CTGTATGTC	LGRINAIKMVFLPQLLYLYQNIPVFIPKSFHKQLDSIINPFI			
TGTAATC	TTGTTTTGT	WDYKTHRIGKKHLCKSKMEGGLSLPNFIFYYWAANLR			
TTGTGAT	AATATTTGT	AVTFLLDDAPPPSSWLSMEREECHPFSIGAVILSPVNLEM			
GCTATAT	TTTGTGCCC	SLYRNNPIIHSTVRIWKQIKVHFELRPISFMLPVVARNPSFA			
CTTGCTT	AGAACTCT	PSNLDNTFERWGELGISTIGDLYIGGTFASFESLRETYNL			
ATCTCAG	GGGTCTTTT	PRSNFFRYLQIRDYVRKHLPTFGNAKPSMFDGCIKICPTS			
GGATTGA	GGTTAATGT	DKLISRIYDAFQSVSAPSTDAIKAKWEEELGTDISVADW			
CCCAAGA	CTGTTCTCT	EDSLEYIHTCSINSRHRLIQFKVLHRLHYSKTKLHRIFPG			
TGAMGCT	TATGTTTAG	TSPMCDKCQAEQGTLLHCFALCSNLYGYWCGIFGVLSE			
TAAGTGC	ATAAGAAC	VLETSIDPDPLLIILGVSDSLNGLTNPQKQLISYSLISAKK			
GCAATAT	TGTAAACTT	LILLFWKKREAPSTKLWLSELANTVHLERIRYILNNKLS			
GTTTAAG	GTCTTTTAT	TFDQIWQPFLSYLLDQAAL			
TTGCAAC	ACCTATACA				
TTATTCTG	CCATTCTTG				
TTTAGGT	TGTATGTGT				
TTAATAG	TCAATAAA				
ATGCTAA	AAATATTTG				
TTGAACA	AACAAAAA				
CTTTAAT	A				
CGCGGGA					
GCCTCCT					
CAGTTCA					
GATGTTA					
GTAGAAG					
TTCTAGG					
ATAGTGA					
GAGGTGA					
ACGTACA					

				GTTGGGA						
				TTTTATTT						
				TTGTTTTA						
				CCTCTTGT						
				TCGAGGT						
				CGCGCCG						
				TGCTTTTT						
				TGGCATC						
				GGCCAGA						
				CAATGTG						
				TTTATTAT						
				TTTTATTT						
				TCATGTTT						
				TTTTCTCT						
				CCTATTT						
				GTGTATG						
				CCTGTTA						
				AATGTGG						
				GTTGATG						
				GGGGGGG						
				TAAATGT						
				TGGGGAA						
				AGTAGGG						
				GGACAGG						
				GTGGAGA						
				GTAAAGG						
				TGCTTGC						
				AGGGGGG						
				GAGGGGT						
				GGGTTGG						
				ATACTGC						
				TCGGGTG						
				TAGGTGC						
				TACATAT						
				GCTGATC						
				GTGA						
	Crack			ATGCTTTT	TAGCTGTGT	MSLRXXXYFAMIVSYVAALLIISVRPCWDDTINLLPENI				
	-			TTTTGCTA	ATTGCCGCT	YTKHIGSPIHRSGLYTAYHNVLNNGQVYLAPYSTEPLRR				
Crac	10_S	.	Salmo	TCGTACC	TTCCATGTT	DKHKSNIYRKLLNYLFTTLLLSGDVQLNPGPNITEPAILT	cd016			
k	Sa		salar	TAACATT	GTCTGTTGT	GVESRGWPRPLIVAAVEVGECYXPMVSLDSPGSRIDFDS	50	727		983

Rex1	Rex1-4_SSa	.	Salmo salar	TTAACGG TTTACAM CGACAGG AGCTTTG TATTTCAT TCATTTTT TACSAGG AACG	CTGTAGCTT GTGAGGTG TGGAAACA CTTTGTTGT TTTTATGGA TTTTGTCTT GTTGCTTTT TGTTCTATG TTGCTCTGT CTGTATGCT ACATCTTGC TTGTCCTAT GTTGCTCTG TCTGTATGC TATGTCTTG CTTGTCCTA TGTTACTCT GCGTGTGCT CACTGTTCA ATGATTGTC TATATTGTA ATTGTTTTT AATAACCT GCCCAGGG ACTGCGGTT GAAAATTA GCCGGCTG GCTAAAAC CGGCACTTT TACTGAAA CGTTGATTA ATGTGCACT GTCCCTGTA AAAATAAA ATAAACTC AAACTCAA	AGATGGC GCCGACA GACATGG CAGCTCT	TAGCCGGCT ACCACCCG GTTACTCAA CCCTGCACC	SNIPKPLYAIPDFASGTQAVLISSPHPVQAGGIVNSATELP LIKTKQNGLNPAVRKHRKCNFFQCVNHSRVIWDPRAKP KGLLGGHLNIRSVIPKSDQIQHLLTDSNLDFLCLSETWL HKNSPCAALIVPGYNVFRRDRIEGRGGLMIYIKEHIRC KQIEWSCDNELECIGLNVTLSPQMSFTLIGMYRPPSTKS VFFDQFNNMLRECDFGKEVILMGDFNINYEDKSCRKTL KRITNTFDLTQLVKGPTRVTCCSKTQIDLVFSNKPERVT KSFNMVTGLSDHNLTLIARKLSKSRFKLSTVRKPDQLRI PKSELNYFENAINGINWNDLLSYTDVEADSEVFLSTIQA TINGFLKKIKSKPGQKSTLPWLNGEIWKLMKERDYALK TALKSKLEHDRRRFTMLRNKVMKEIRQAKANFFINIIGE AKGNSKLIWENLKKLTGKDHSNTAKRLEIMVNNNLTQ DAVEIAIAFNSYFIDSVRVLTQNPSTGLLGSVLVNDAQP VFIIREVSESKVNKVISSLKNSKSKDVFGLDSTFLKNYKE SLIGPITKVTNTSIGLGVFPRVWKSAIITAIFKSGDPADVS NYRPISILPVVSKVVEKCVAEQLISHLNNSPFTLHSMQFG FRAKHSTETANCFLLENVKSKMDKGGVVGAVFLDLRK AFDTVNHEILITKLSKFNFSPDALRWMKSYLEGRTQCVR VSNELSPTLSYDVGVPQGSILGPLLFSLYINDLPSVCTGS EVQMYADDTVIYVHAKSKQQAAQELTTVMVQVTKWL SDSCLHLNVKKTVCMFFTKRATDATEPDVYVSGEKLQV VSDFKYLGIILDSNLSFKKHVKKVIQITKFNLANFRFIRN CLTTEVAKLYFKSMILPHLTYCLTSWAQACYTTLKPIQS VYKQALKVLDRKPNSHHHCYILRKHELLSWENLVQYT DACLVFKILNGLAPPPLSIFVKQKTQTYGSRSTRSAMRG DCIVPLRKSTFSKSAFSVRASHVWNTLPSDTHNCTTYHT FTKCMKTWLKVVNQTCEHNP	LLDIRAAVTHQHYDQEYDFPELDPLFVPPRAIELIPEAAP RRRRRRRGIRSGLLVRLRRRAHHPPLPSILLANVQSLDN KVDELRARISFQRDIRDCNILCFTESWLSPDILSPSIQPAG FSVHRADRNKELSGKKKGGGVCFMINYSWCDCDNVQK	cd01650	539	848

				GCTTCTA GCTCCCTA AGCAACT TTGCAGT ATTTGTTT TTTTGTGT GTTATTTC TTACATT ATTAGCC CAGAACG TTTTTTGT GTTATTA CATACAG CCGGAAA TAA	TTAGAGGCT GCTGCCCTA TATACATAG ACTGGAAT CACTGGCC ACTTTAATA ATGGAACA CTAGTCACT TTAATAATG TTTACATAC TGCTTTACT CATCTCATA TGTATATAC TGTATTCTA TTCTACTGT ATTTTAGTC AATGCCACT CCGACATTG CTCGTCCTA ATATTTATA TATTTCTTA ATTCCATTC TTTTACTTT TAGATTTGT GTGTATTGT TGTGAATTG TTAGATACT ACTGCACTG TTGGAGCTA GGAACACA AGCATTTCG CTACACCCG CAATAACA TCTGCTAAA TATGTGTAT GTGACCAA TAAAATTTG ATTTGATTT GAG	LKSFCSPDLEYLTIKCRPYYLPREFSSVIVTAVYIPPQADT TTALKELHWTLCKLETTYPEAAFIVAGDFNKANLRKTL PKFYQHIDCSTRAAKTLDHCYSPFRDAYKALPRPPFGKS DHDSILLLPSYRQKLKQEVPVLRSIQRWSDQSESMLQDC FDHADWDMFRVASENNIDVYTDTVTEFIRKCIGDVVPT VTIKTYPNQKPWIDGSIRAKLKARTTAFNHGKVTGNMV EYKQCSYSLRKAIKQAKRQYRDKVESQFNGSDTRRMW QGLQTITDYKGKTSHVADTDVLLPDKLNTFFARFEDNT VPPTRPAPKDCGLSFSVADVSKTFKRVNPRKAAGPDGIP SRVLRACADQLAGVFTDIFNLSLSQSAVPTCFKMSTIVP VPKKAKVTELNDYRPVALTSVIMKCFERLVKDHITSTLP DTLDPLQFAYRPNRSTDDAIAIALHTALSHLDKRNTYVR MLFIDYSSAFNTIVPSKLIIKLGALGLNPALCNWVLDFLT GRPQVVKVGNNTSTSLILNTGAPQGCVLSPLLYSLFTHD CVATHASNSIIKFADDTTVVGLITNNDETAYREEVRALA EWCQENNLSLQRQQNEGADRGLQETAEGARPYPHRRD RSGEGGKLQVPRRTHHGPWRSGARKITSPSNVNKTKELI VDFRKQQREHAPIHIDGTAVEKVESFKFLGVHITDDLK WSTHTDSVVKKAQQRLFNLRRLKKFGLAPKTLTNFYRC TIESILSGCITAWYGNCTARNRRALQRVVRSAQRITGGT LPALQDTYSTRCHRKAKKIIKDINHPSHGLFTPLSSRRRG QYRCIKAGTERLKNSFYLKAIRLLNSHH			

Rex 1	Rex1-3B_SSa	.	Salmo salar	GCGCCGGAGGAGATGGCTGCCGTTTTACGGKCTCCTAACCAATTGTGCTATTATGTGGGGTTTTTTTCGCGTTATTTGTAACTTATTTTGTACATAA	TAGTCACTTTAACCATACCTACATGTACATACTACCTCAATCAGCCTGACTAACCGGTGTCTGTATATAGCCTCGCTACTCTTATTTTCAATGTCTTTTTACTGTTGTTTTATTTCTTTACTTACCTATTATTACACACATACTTACCTTTTTTTCCGCACTATTGGTTAGAGCCTGTAAGTAAGCATTTCACTGTAAGGTCTACACCTGTTGTATTCGGCGCACGTGACAAATAAACTTTGATTTGATT	CFCHRILRQKRASGYQDSDHSPRIRQRFFLQQQQAGLTGYSPNTXQGRHPRYSQEEATRRGRRAGCLVRTRRRRVGKLPLPSVLLANVQSLDNKLDEVRSRISYQRDIKNCNILCFTESWLNDDMDIQLAGYTLHRQDRTAHSGKTRGGGLCIFVNNSWCTKSKEVSRFCSPEVEYIVINCRPHYLPREFSAILFVAVYLPPQTDAGTKTALSQLYKDISKQENAHPEAALLVAGDFNAGKLKSVLPNFYQHVKCATRGKKILDHLYSTHRDAYKALPRPPFGKSDHNSILLIPAYKQKLKQEAPVTRSIRKWSDEADAKLQDCFAXTDWNMFRDSSDGIEEYTTSVTGFINKCIEDVVPTVTVRTYPNQKPWITGNIRTELKGRAAAFKERDSNPEAYKKSRYALRRTIKQAKRQYRAKIESYYTGSDARRMWQGLQTITDYKGKHSRELPSDTSLPDELNHFYARFEASNTEACMRAPAVPDDCVITLSVADVSKTFKQVNIHKAAGPDGLPGRVLRACADQLAGVFTDIFNMSLTESVIPTCFKQTTIVPVPKNTKATCLNDYRPVALTSVAMKCFERLVMAHINTIIPETLDPLQFAYRPNRSTDDAISIALHTALSHLDKRNTYVRMLFIDYSSAFNTIVPSKLITKLRTLGLNTSLCNWILDFLTGRPQVVRVGSNTSATLILNTGAPQGCVLSPLLYSLFTHDCMARHDSNTIIKFADDTTVVGLITDNDETAYREEVRDLAVWCQDNNLSLNVTKTKEMIVDYRKRRAEHAPILIDGAVVEQVESFKFLGVHITNKLSWSKHTKTVVKRARQSLFPLRRLKRFGMGPEILKRFYSCTIESILTGCITAWYGNCSASDRKALQRVVRTAQYITGAKLPAIQDLYTRRCQRKALKIVKDPSHPSHRLFSLLPHGKRYRSAKSRTKRLLNSFYPQAIRLLNRDQKAPQQLLPPSHKTPEQVIKWLPGLFALCAPPPSSTLCLSYMHSHFNYTFMYILPQLARPTSAPAHWLTGLSALCPTTRQPLFLRCCYSLFIIYA	cd01650	549	801
Crack	Crack-2B_SSa	.	Salmo salar	GCCCACAGGACTGGAAAACCCACMACCKGCCCAGGTGATAGGCCCAGGCCGAGGTWGCTKTTCTGGAWAGAGCCAAG	TGACCTAGATAGTTTGTGTGTATGCATTGATATGTAGGCTACGTGTGCCTTTTTAAAAATGTATGTAGTTCGTCGAGCTGTTCTTGTCTAWTGAT	MFFSLALFALFHIMSISGKLPRKGLKIAHINICSLRNKVHEINNLLTSDNIHILAISETHLDNSFDDTSVAIQGYNIYRRDRNAYGGGVAVYIQSHIPVMLREDLMSSVIEVLWLQVHLAHLKPFLLGCCYRPPSANSQYLNNMCEMLDSVCDVNREVYFLGDLNIDWFSSSCPLKRKLLTVTSACNLVQVINQPTRVFTNTTGTRSSTCIDHIFTNAVELCSKAVSVPIGCSDHNIVAISRKAKVPTAGPKIVYKRSYKRFCCDSYVDDVKNICWSDVINEEHPDAALDEFMKLLLPIIDKHAPVKKLTVRTVKAPWIDEELKNCMVERDGAKGVANKSGCTSDWLTYCKLRNYVTKLNKKKKKLYYEAKINDIKNDGKKTLEYFKXNYGQKDKFNSIFHRIRWLIHHKTIRMMEKKLWSTLNX	cd01650	557	816

				AACTTGW	GTTCTGTAT	IMGRKTNSTPSFIESDGLFITKPFDVANYFNDYFIGKVGK				
				GAGGAAC	TATGTCATT	LRQEMPTTNSEQLYSCIKKQIMKEKQCKFEFCKVSVGE				
				GTATATC	CTGTATTAT	VEKLLLSINNDKPPGIDNLDGKLLRMVADSIATPICHIFN				
				TTCCTCA	GTTTCATGT	LSLEESLCPQAWREAKVIPLPKSGKAAFTGSNSRPISLLP				
				ACGAGGA	TTTGTGTGG	ALSKLLEKIVFDQIQCYFSVNKLTTDFQHAYREGHSTCT				
				CTGTCCT	ACCCCAGG	ALTQMTDDWLKEIDNKKIVGAVLLDFSAAFDIIDHNLLL				
				GAAGCTG	AAGAGTAG	KKRMCYGFSTSAISWIQSYLSNRTQRVFFNGSFSNVKHV				
				TGCGCCA	CTGCTGCTT	KCGVPQGSSLGPLLFSIFTNDLPLALNKACVSMYADDST				
				GAAGAGG	TTGCAACA	IYASATTANEVTETLNKELQSVLEWVASNKLVLNISKTK				
				AAAGAGC	GCTAATGG	SIVFGTNHSLSSRPQLNLVMNGVAVEQVEETKLLGVTL				
				TGATCCC	GGATCCTA	DCKLSWSKHIDSMVVKMGRGLAVIKRCSAFLTPHSKKQ				
				AGCCATG	ATAAAATA	VLQALVLSNLDYCPVVWSSAARKDLVKLQLAQNRAAR				
				AAAGCTG	CCAAATAC	FALNCNQRADINTMHASLSWLRVEERLTASLLLFIRNIN				
				CCAGAGT	CAAA	VLKIPNCLHSQLTHSSDTHTYPTRHATRGLFTVPKSRTN				
				GCGTGGG		SRKRTVLYRALIAWNFLPSHIAQINSKPGFKKQIKQHLA				
				GACATTG		AQRLSPI				
				CKTACAT						
				CCGCTAC						
				GACAGGC						
				TCATTGT						
				CCACCCT						
				CCCTCCC						
				AGAAGCC						
				TGGAAGG						
				GATGAGA						
				GAGCCAA						
				GCCTATG						
				GGTTCGT						
				AGCCTCA						
				ACCCCAC						
				ACACACA						
				CACACAC						
				ACACACC						
				AATTGAT						
				TAATGGA						
				CTGCTGA						
				ATGT						
Rex 1	Rex1-3D_Sa	.	Salmo salar	TGGCGCC	TAGTCAAAT	MISCRPHYLPREFSSVFFVAVYIPPQSEAGTKTALNELYS	cd01650	372		624
				GGAGAAG	GGCTACCC	AISKQENAHPEAALLVAGDFNAGKLKSVLPNFYQHVKC				
				AAGGCAG	AGACTATTT	ATRGKKTLDHLYSTHRDAYKALPRPPFGKSDHNSILLIP				

				ACGTTTT ACGTGTC CCCAACC GATTGTG TTTTTTTG TTCGTTTA TTTGCGTT GTTTGTA ACTTATTT TTTTACTT ATTTTGT ACATAAT GTTGCCG CTACCGT CTCTTAT GACCGAA AATAACT TCTGGAC ATCAGGA CTGCGAT TACTCAC CACGGAC TGGCAGA ATCCTTTT TTTCCTTT AACGAGT CTGACGA GCCCGAC GCGAACG ATATACT GCTTTCTC GGGAACA GGCCCAG ATCCCCG TGATTTG CGTGAAG AGGAGGC GGAGAAA AAGGGGC CAGAGGG	GCATTGCCC CCCTCCTCC ACACCACT GCCACTCTC TGTTGTCAT CTATGCATA GTCACTTTA ATTAACTCT ACCTACATG TACATACTA CCTCAACTA ACCGGTGC CCCCGCAC ATTGACTCT GTACCGGC ACCCCCCTG TATATAGTC TCGCTATTG TTATTTTAC TGCTGCTCT TTAATTACT TGTTACTTT TATCTCTTA TTCTTATCC GTATTTTTT TAAACTGC ACTGTTGGT TAGGGGCT CGTAAGTA AGCATTTCA CTGTAAGGT CTACACCTG TTGTATTCG GCGCATGT GACTAATA CAATTTGAT TTTTGATTT GATTTG	AYKQKLKQEAPVTRSIKKWSDEADAKLQDCFASTDWN MFRDSSDGIEEYTTSVIGFINKCIDDVVPTVTVRTYPNQK PWITGNIRTELKARAAAFKERDSNPEAYKKSRYALRRTI KQAKRQYRTKIESYYTGSDARRMWQGLQTITDYKGKH SRELPSDTSLPDELNYFYARFEANNTETCMRAPAVPEDC VITLSAADVSKTFKQVNIHKAAGPDGLPGRVLRACADQ LASVFTDIFNLSLSESVIPTCFKQTTIVPVPKNTKVTCLND YRPVALTSVAMKCFERLVMAHINTIIPGTLDPLQFAYRP NRSTDDAISIALHTALSHLDKRNTYVRMLFIDYSSAFNTI VPSKLITKLRTLGLNTSLCNWILDFLTGRPQVVRVGNNT SATLTLNTGAPQGCVLSPLLYSLFTHDCMARHDSNTIIK FADDTTVVGLITDNDETAYREEVRDLAVWCQDNNLSL NVIKTKEMIVDYRKKRTEHAPILIDGAAVEQVESFKFLG VHITNKLTWSKHTKTVVKRARQNLFPLRRLKRFGMGPQ ILKRFYSCTIESILTGCITAWYGNCSASDRKALQRVVRTA QYITGAKLPAIQDLYTRRCQRKALKIVKDSSHPSHRLFS LLPHGKRYRSAKSRSKRLLNSFYPQAIRLLNI			

CGGGCTG
CCTTCTG
AGAATTC
GTAGGCG
ATCGAAT
AAACCCC
CACTTCC
TTCCATTC
TGCTAGC
AAACGTG
CAATCTT
TGGASAA
TAAAATA
GATGACC
TACGCGG
AAGATTA
AACTACC
AACGGGA
CATTCAA
AACTGTA
ATATCTT
ATGCTTC
ACGGAGT
CGTGGCT
GAACGAC
GACACTA
TCAACAT
ACAGCTG
GCTGGTT
ATACGCT
GTACCGG
CAGGATA
GAACAGC
GGCGTCT
GGTAAGA
CAAGGGG
CGGCGGA
CTATGTA
TTTTTGTA
AATAACA

			GCTGGTG						
			CACGATA						
			TCTAAGG						
			AAGTCTC						
			GAGCTAT						
			TGCTCGC						
			CTGAGGT						
			AGAGTAT						
			CTC						
Crack	Crack-13_Sa	.	Salmo salar	ATCCAAG		MYLLLCLLFTQCHPGLNRLNDQWVSELSSPYALQFANT	cd01650	635	891
				ATGGTAG		PTRPSCQYHQELGLSLQLHLYHSQTQLDLHLMAYFQNL			
				CCAATCA		KLLNYFFSVSKDSVRFIDNGAGSTKSSRIGQLRYIVMLLL			
				GTCTGTC		LISGNVRPNPGPVDTMQSLPTPYDFKNRAGFGLLHVNV			
				CGTTTGT		RSLFPKIDMIRIWAKTTNADVMVLSETWLKKSVSNNSIS			
				ATGTGTG		IDGYTVFRTDRMSKGGGVAIYVKSEFNTSEILSITKAKHF			
				ATCCTTG		ELLAVKVNVYKDSHITVVGCYRPPSASGDALNSLSDVL			
				AGATCAG		HKLNDSEFIILGDLNWDMLTSVSDSFKELCDSLNLAQLI			
				AGTTTTA		NAPTRPNPRAPNKSTLLDIILTNTPHKYTSTGIFCNDVSD			
				AGAGTTT		HCAIACVRNTKMTKAKPRYIFKRHFRQFDEQAFLHDLY			
				TTATGTTT		HNIDRVSLIPDVDTAWDYFYMKFVSICDKHAPVKKFRIS			
				TAATGAC		GRDNPWFSDNLAELIRKRNISWAQARHTNAPVDWASFR			
				CCTGATC		ALRNKCTGLIRKSKSDYYLNAVTENLNNPTKFWKLIKS			
				CTGTTTA	TAGGCTGG	VSGSNVSSGLPDHLMMDSNEVKDKAHIVGVFNKHFISA			
				CAGAAGC	ATCATGTTG	GSVFDNGGAQASNVSSVTVNYDIGPHVVNHFNFEPVSYT			
				AAATACT	TTGTATTGT	EVYKALKAIDTKKSAGPDNLDPYLLKIAAGIITEPVAHIF			
				TTTTATGT	ATTGTTGTA	NLSLLTNSIPSIWKSAYVLPLLKGGDPSDANNYRPISKLP			
				GAAGGTG	TGTTTTAAT	ILAKVYESLVNSQLKNFFIEKNILSGVQSGFRSGHSTTTA			
				GATGCTG	TATGTAATG	AMAVANDIINALDKKQHCAALFVDLSKAFDSVDHELLL			
				TATATAA	TATTGATTG	ARLRNIGLSEGAVNWFRNYLSDRTQCVYTDNHKSSFLEI			
				CAACTAT	TTGCTGCCT	NRGVPQGSILAPVLFSIFINDLGNGMQPAKLHLYADDTV			
				TATTTTTA	TCTTGGCCA	IYSCAPSLVQAVEELQTAFQSLQASLYGLKLVLNVQKT			
				GATCTCA	GGTCTCCCT	KFMTFTRARTLPENVSIVTSGGLSIEKVSSYKYLGIWLD			
				GTCTCTA	TGAAAAAG	DKLSFKVHVDNLVTKLKLKLGFYFRNKACFPLMARKK			
				TTTTACG	AGACTCTG	LVQATFLSVIDYGDLLYMHAASSVLQRLDSVYHASLRFI			
				AAATTTA	GGTCTCAAT	TNAKSLTHHCTLYQMVGWTSLYMRRKIHLYVFIYKALL			
				CTTGTAC	GGGCTTTTC	GKLPLYLCSLVSFTTSSYHTRSARWLLLKVPRTFTVLGK			
				AAAGATT	CTGGTTAAA	TAFSSCAPDAWNSLQSMLHLDMLVPLNEFKILMGDSVT			
				TCGGATT	TAAAGGTT	EECKCFF			
				GAGAAGG	AAATAAAA				
				CCTTTATT	A				

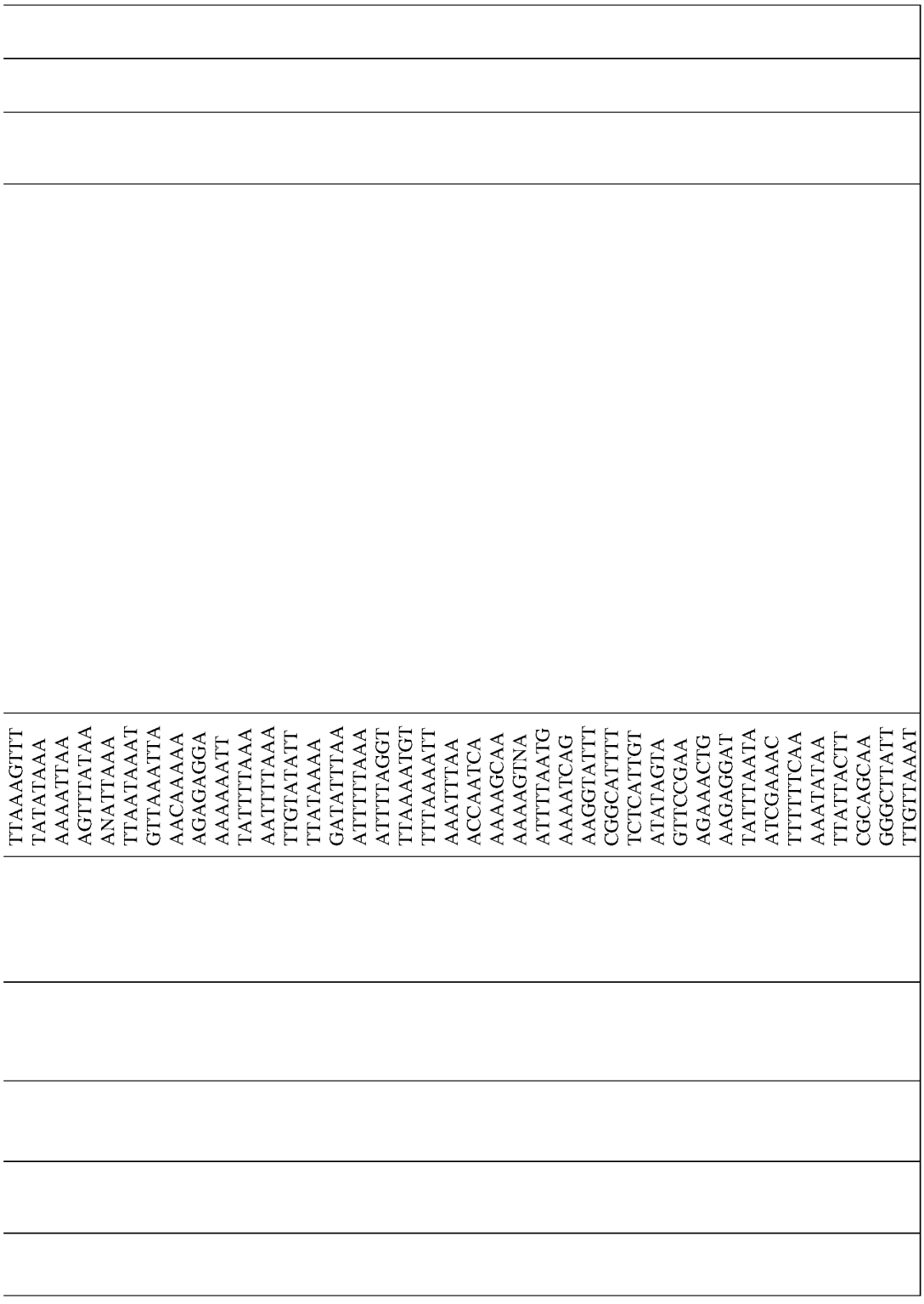
				TTTTTCTACACCTTTCTGGAGTTGGAGCTCGCCGAGTCAGGAGCGTCTGTCTGCGCGCAAGGCCTGCTACGCGAGCGCAGGTGCGCACTGAAGGCTAGGAGCGGAGAAGCGGTTATCGGATTGGAAAACCTGATGCACCGTGGGCGGTGGGAAAATCTACGTTGTTCTACACAACGTCTACTGGTCGTTGCACACCTGTGCTGAAATACCTCTAAAAACCGGCATT					
R2	PERERE-9	BN000800	Schistosoma mansoni	ATCTCACGTTTTAATTTATTTTTGAACTACTGCAGTCTGAGTGCTTCTAA	TAACGGCTGAACGAATAGCCCCCTTCACTCTTAGACATTCCCCCACTGTTGTTGCTTATCT	MPVSTGAETDITSSLPIPASSIVSPNYTLPDSSSTCLICFAIFPTHNILLSHATAIHHISCPPTPVQDGSQQMSCVLCAAAFSSNRGLTQHIRHRHISEYNELIRQRIAVQPTSRIWSPFDDASLLSIANHEAHRFPTKNDLCQHISTILTRRTAEAVKRRLLHLQWSRSPTAITTSSNNHTITDIPNTEARYIFPVDLDEHPPLSDATTPNASTHPLPELLVILTPLPSPTRLQNISESQTSHESNKNSMHTPPTYACDPDETLGATPSSTIPSCFHSYQDPL	cd01650	492	750

CGACCCG	TCATGCTCT	AEQRGKLLRASASLLQSSCTRIRSSSLLAFLQNESTLMDE	
AAGGCTC	TGTGTTAAT	EHVSTFLNSHAEFVFPRTWTPSRPKHPSHAPANVSRKKR	
AGAAACT	TGACTGCTC	RKIEYAHIQRLFHHRPKDASNTVLDGRWRNPYVANHS	
ACCCACT	TCTTCTGGG	MIPDFDCFWTTVFTKTNSPDSREITPIIPMTPSLIDPILPSD	
TCTTGAA	TTGACGTCT	VTWALKEMHGTAGGIDRLTSYDLMRFGKNGLAGYLN	
CTGCTAC	GATTGTCTC	MLLALAYLPTNLSTARVTFVPKSSSPVSPEDFRPISVAPV	
TTTTTGCT	TCTCTCTTT	ATRCLHKILAKRWMPLFPQERLQFAFLNRDGCFEAVNL	
GTTTATC	CCATATTGC	LHSVIRHVHTRHTGASFALLDISRAFDTVSHDSIIRAAKR	
CACAACA	TTGCTCTGC	YGAPELLCRYLNNYYRRSTSCVNRTELHPTCGVKQGDP	
ACAGTTG	CCGCTTACT	LSPLLFIMVLDEVLEGLDPMTHLTVDGESLNYIAYADDL	
TGATTCT	TCCAATAGT	VVFAPNAELLQRKLDRISILLHEAGWSVNPEKSRTLDLIS	
ATTCTCC	TGTCATATT	GGHSKITALSQTEFTIAGMRIPPLSAADTFDYLGIKFNFK	
ANATATT	ATGTCTTTG	GRCPVAHIDLLNNYLTEISCAPLKPQQRMKILKDNLLPR	
CCTTGTG	TTTACTTGC	LLYPLTLGIVHLKTLKSMDRNIHTAIRKWLRLPSDTPLA	
CTTTTGTC	CATGTCTAA	YFHSPVAAGGLGILHLSSSVPFHRRKRLETLLSSPNRLLH	
AACATTA	CGACAATT	KLPTSPTLASYSHLSQLPVRIGHETVTSREEASNSWVRR	
TTCTATA	ACTTTATCT	LHSSCDGKGLLLAPLSTESHAWLRYPQSIFPSVYINAVK	
CCAACTG	ACCTTAGTT	LRGGLLSTKVRRSRGGRVTNGLNCRGGCAHHETIHHIL	
TACCACC	TGTCCTCTT	QHCALTHDIRCKRHNELCNLVAKKLRRQKIHFLQEPCIP	
TACTTCTT	GGTTTCGAT	LEKTYCKPDFIIIRDSIAYVLDVTVSDDGNTHASRLLKIS	
CATCTCA	TGCCTTCAT	KYGNERTVASIKRFLTSSGYIITSVRQTPVVLTFRGILDR	
CGTTTTA	ATGTTCATG	ASSQSLRRLCFSSRDLGDLCLSAIQGSIKIYNTYMRGT	
ATTCTGG	GCGGAATC		
TCTTATTT	TGATGTTTA		
TCTCATC	TAATGACTA		
ATTAGTC	TTCCTATTA		
ACGGAGA	CCACCACTA		
GGGCCTA	CAACTACTA		
TGAACGG	TTATTATTT		
TCCGTGA	TCATTACTA		
CGCGAAA	TTAACATTA		
TTCTATCC	TTATAAACA		
GCGATTT	TTATTACTA		
CGACCTC	TTATTATTA		
TCCTGCT	TTACTATTA		
AGTGGTC	TTACTTCTA		
CCCGAAG	CAATTAATA		
TACGGTT	TTATGGCTA		
CCTCTGG	CTCCTCTCA		
CCTGTCA	GCACACCA		

				GTTGTGTTAAAACTATATAATAACG	ATAAAATATCAATCAAACATCTCAATTATATCCACCTATTAAACTCTCTCTATTTCCCCTGAGTTATAAACTTACAATTCAGTCTAACCGAATATCTCTCTTTTACAAATCTTAAGTATGTAATTTTGTGCCAAACCCATTTGGGTCTGTACAATTTGATACTTAAAAATAAATGTTATTAGCC				
R2	R2Sm-A	.	Schistosoma mansoni	ATGTTTTAATTTATTTTTGAACTACTACTGTCTGAGTGCTTCTTACAACCTGAAGGCTCAGAAACTACCCACTTTTTGCTGTTTATCCACAACAACAGTTGTGAATCTATTCTCCAAATATTCCTTGTG	TAGCCCCCTTCACTCTTAGACATTCCCCCACTGTTGTTGCTTATCTTCATGTTTTTGTGTTAATTGACTGCTCTCTTCTGGGTTGATGTCTGATTGTCTCTCTCTCTTTCCATATTGCTTGCTCTCCCCGCTTACTTCCAATAGTTGTCATATTATGTCTTT	MPVSTGAETDITSSLPIPASSIVSPNYTLPDSSSTCLICFAIFPTHNILLSHATAIHHISCPPTPVQDGSQQMSCVLCAAAFSSNRGLTQHIRHRHISEYNELIRQRIAVQPTSRIWSPFDDASLLSIANHEAHRFPTKNDLYQHISTVLTRRTAEAVKRRLLHLQWSRSPTAITTSSNNHTTTDIPNTEARYIFPVDLDEHPPLSDATTPDASTHPLPELLVILTPLPSPTRLQNISESQTSHESNRNSMHTPPTYACDSDESLGVTPSSTIPSCFHSYRDPLAEQRSKLLRASASLLQSSCTRIRSSSLLAFLQNASTLMDEEHVSTFLNSHGEFVFPRTWTPSRPKHPSHAPANVSRKKRRKIEYAHIQTLFHHRPKDAANTVLDGRWRNPYVANHSMIPDFDCFWTTVFTKTNSPDSREITPIIPMTPSLIDPILPSDVTWALKEMHGTAGGIDRLTSYDLMRFGKNGLAGYLNMLLALAYLPTNLSTARVTFVPKSSSPVSPEDFRPISVAPVATRCLHKILAKRWMPLFPQERLQFAFLNRDGCFEAVNLLHSVIRHVHTRHAGASFALLDISRAFDTVSHDSIIRAAKRYGAPELLCRYLNNYYRRSTSCVNRTELHPTCGVKQGDPLSPLLFIMVLDELLEGLDPMTHLTVDGESLNYIAYADDLVVFAPNAELLQRKLDRISLLLHEAGWSINPEKSRTLDLI	cd01650	492	749

				CTTTTGTC AACATTA TTCTATA CCAACTG TACCACC TACTTCTT CATCTCA CGTTTTA ATTCTGG TCTAATTT TCTCATC ATTAGTC ACGGAGA GGGCCTA TGAACGG TCCGTGA CGCGAAA TTCAATC CACGAAT TCGTCCT CTTCTGCT AGTGGTC CCCGAAA TACGGTT CCTCTGG CCTGTCA GTTGTGT TAAAACT ATATAAT AACG	GTTTACTTG CCATGTCTA ACGACAAT TACTTTATC TACCTTAGT TGGTCCTCT TGGTTTGGT TGCCTTCAT GTGTTCATG GCGGAATC TGATGTTTA TAATGACTA TTCCTACTA CCACCATTA CAACTATTA TTATTATCA CTATTATTA ACATTATTA TTACTTCTA CAATTAGTA TTATGGCTA CTCCTTTCA GCACACCA ATAAAATCT CAATCAAA CATCTCACT TATTAAACT CTCTATTTC CCCTTCGTT ATAAACTTA CAATTCAGT TTAACCGA ATATCTCTC TTTTACAAA TCTTAAGTA TGTAATTTT GTGCCAAG CCCATTTGG GTCTGTACA ATTTGATAC	SGGHSKITALSQTEFTIAGMRIPPLSAADTFDYLGIKSNF KGRCPVAHIDLLNNYLTEISCAPLKPQQRMKILKDNLLP RLLYPLTLGIVHLKTLKSMDRNIHTAIRKWLRLPSDTPL AYFHSPVAAGGLGILHLSSSVPFHRRKRLETLLSSPNRLL HKLPTSPTLASYSHLSQLPVRIGHETVTSREEASNSWVR RLHSSCDGKGLLLAPLSTESHAWLRYPQSIFPSVYINAV KLRGGLLSTKVRRSRGGRVTNGLNCRGGCAHHETIHHI LQHCALTHDIRCKRHNELCNLVAKKLRRQKIHFLQEPCI PLEKTYCKPDFIIIRDSIAYVLDVTVSDDGNTHASRLLKIS KYGNERTVASIKRFLTSSGYIITSVRQTPVLTFRGILERAS SQSLRRLCFSSRDLGDLCLSAIQGSIKIYNTYMRGTQRLN E			

Penelope	Penelope-1_SM	.	Schmidtea mediterranea	ACTTTCG AAGGTTT GTCGAAA GATT	TATACAA ACACCGA TAAACTA TACCTTC CCAATCC CGATCCA GAATTAA ATTCAAC	TTAAAAAT AAATGTTAT TAACTCTGA TACTTTTAC TTAAAGATT AAAACAAT TTTATAGCA AAAAAGGA AAACCTTCA GGAATGAA AATCAGAG TAGAATTGT ATTTAGGTC TTGAAGTTG TTTGTGGAG GGTTGTTTG AAGGCTTTT TTGTTGCGT TTTTAAGAG GAAAGCTTT TTCATATAC ATCTTTCAT AAGACTTTA AAAACTCA GGAACTATT TATTAATTT GCTCTGAAT TTTATAAAA AACTTTCTT TAGTTTTAA TTTTATTAA GAAATTTG AAAGACAA AAAAAAAT GACGTTAG AAGCAGTT ATTTAACAT AAGATATA GAAAACAT TTTAAAATT	MDVVEKEGKRKPYKSNLSNNEQGVLKALKTKPYVYLP SDKGGEFCVIEKSRYCEAAYKHLSDSNTYKKISRMSPNT IETKINKVWKDIAAKMNLKKRLTKSFVTKNSDLPNFYH LIKTHKDGPSIKIRPIVSNTNGPTTKTSWLLSKILKPLLNN VPYHLQSSDCLMARLNKTGKTTHSEFKYPFSLDVVALY TSIDPSEAIECLRRRLENNLLIKEMNTENIINLLTVILENT YFTFDQNIYKQIKGLPMGSSISGLLAIIFMDYLERQTLRTI QSIGFYARYVDDIFITTKSKKEAEEIFHIFNEAHPDIKFEIE YPDENGSLNLLDLRIKIDVDTPTYGFYKKSARKEIFMHQ RSALPYLQKVNVIKNEKQRIANRCTEGKENYKQQLXFD KVLRRNGYHKEDIERMNSKPKTPMDRTQKKKKYFYLNI PFISDTLDFKIRKELNHTGLPIRISHKSWTLRQALANNTK KENICTMENCQTPKKICFRKNVVYCITCNACKEIYIGSTL RNLHQRIKEHLDPAGKSSVYRHLQQCQPRKERKSISIKIL AGDNIQKNLRFKEAILILDKHPSINTKEEISEIASLICC	SSF5 6672	144	314


TTAAAGTTT
TATATAAA
AAAATTAA
AGTTTATAA
ANATTAAA
TTAATAAAT
GTTAAATTA
AACAAAAA
AGAGAGGA
AAAAAATT
TATTTTAAA
AATTTTAAA
TTGTATATT
TTATAAAA
GATATTTAA
ATTTTTAAA
ATTTTAGGT
TTAAAATGT
TTTAAAATT
AAATTTAA
ACCAATCA
AAAAGCAA
AAAAGTNA
ATTTTAATG
AAAATCAG
AAGGTATTT
CGGCATTTT
TCTCATTGT
ATATAGTA
GTTCCGAA
AGAAACTG
AAGAGGAT
TATTTAATA
ATCGAAAC
TTTTTTCAA
AAATATAA
TTATTACTT
CGCAGCAA
GGGCTTATT
TTGTTAAAT

Non-LTR Retrotransposon	LIN10_SM		Schmidtea mediterranea	TTTACAC ATTGACA TTCGATA GCACCGG GCCGAAA TCCGGAT CTGATAA GTTATTTT ATTCAAA ACAACCA TTACTTTA TTTTCCA GATAATA ATGAATT AAATGCT GATCCTA TCTTTGA GAAAAGT TTGCAAG AATTACG TCTT	AATAATAA TAAT TAACTTCTA	MSKNGTDPHLLELCYKAHVFLNGSKEIPLSTAVGHWRE IDSTGIFDSDFSELMEQIIAKPFKTEGLQLKIENAIKTFDK FIADAVNPDKLINHKHLNFFYNQLYYNPLKNFNFSIPIKS ILKNNNIPSHRNFFNNEESTPLNKKSISFKNISHSKPFSYL TALTNNLKYENIDNILISIRNCTDSFFINLNICFFELKHLIK NNFINYFNDYFYIFYLKIQKFNNKKLQKSNYYQNNIPLL FKISSLTNDFFINIENYFSKIKTKRNINLFLNDFLXILKIKI NILLFKYSXFKKLNPKNKHEKISTRSSXASGRLALAVTPS SPLDLTINTTKQTTTDINTDTTNTSHKFSNNRYSLFLDQN EPLDLSIVKHPHKNTNIPSTVKNTYNPLNKQKPFHKTTT KTIFSNAQNNIKKHNIDTSKRPHFSXPIPKSCNTDIKYEIP PIEKNTPKLNDIIXSNITFNPKTPNIQQNISLPKFILPSNLH NNTTNVDLDAESTLVNPHFQLNFNLKPCPNLFIPKXDLP XNLNMEIDQKLLLTNTKASDETISYSPITQTQTDESVLTH KLIELISINTQANRIIEQINSDIDLFSQSPHSYEIPNTETKEI FTVYNNLDIAKKRVEIIEEDILKTKTQTNDDIPMINFNN NYINNNSLFKNILLSNESVENLFRDISDNDSIISKSPNKKI NNQDFLKKTVKINKTMFNLRNDGYLICIGGKCEKKKIGI HNIDNIIEHRKTNHSDTIDKNTTFLCLLCEGSVKRKTILL NKIIQHTETFHSEKHFPVTREELNNKNKILELXMNTEGT LNCRYTNIKKHCQNDLTYKSTYAQIEKHIAKKHKRKTC NINEIICHCGETLEHHNIIEHHNKHIGVLPAEDKLRMIPD NIPVSVKDIGNDISILSTFFCMHLLPQSQQLLYNHTLLKP SYKQNFKTFFETFPFSQYPNWNEIILPFKTSTVVNWAVFYF NKLKAFATIIDPLSPNSLQKHRKIADTILNTIKNIQNIIYPS NINLTELNYPMCANTNESGFFICHIIQCISINHNYITIPIISEI RKRLDNLINKVIIDRAPIDNILKYTNILNNIKILQNNNSEN PDEGLAEINRIHECIFPIKHKIKKXLSLIHHLLLLLKIIKLNK KXXFNPPSTIIKNNKIELTYKLRFNFNINYKSTINKIFNNSI NTLSPHPDRIILNFERETIIINDIKFKYKLPYYPENFIHKPTT IEETINTFKIVDLSSSPGKDQISYLDWKKIDPDNSILCSLF NNILFKSITPLKWKTFKTKLIIKPGKENSSHDLSSWRPLAI LDTSYRLFATLINRRLLEWIKLGNLININQKAIGCPDGCA EHNSVLLLLNERAVRMGTNINIVWLDLADAFGSISHDLI WNVLTNMGINNNTIKLFKNMYNNCNTYYECENNHYNF LKICIIIVTLIMNVKTTTTKTIPINNGVKQGCPMSMTLFCL AINFIITNILKDHPFIINNYNVSILAYADDIVLISDNRIKIAK ILTDIIKYTSDVKLKFRPTKCGFYENNRNHDEKPLKIYNE	cd01650	1274	1556

						IIPIIDENNIYKYLGVGFGGKSNHNIDNILEKALTDAEATF				
						LSDLHPSQKLFAYKTSIHSRLIFPFRTCIVKHMILDCDRN				
						RIVDNRTKQLGFDQKIKRLIKNSIGDIFQNISNNFIYAHCS				
						MGGLGIVPSIDEYLIQSTIQLTRLLMSSDVTMRDFIIGELI				
						ECVRARFNYLENEIEMAMKWLNGDIDDEKVGKPCLIFL				
						KFKKSVYRLKIKFSIICKLTYIDNIFSLEISKNDFKIVINNE				
						NIDNASKIMHILTGEWYALEWYKMTHQGHIAETVGKSQ				
						FSKHLIKSTTLGDNQYFFLIKARNNSLNLNSNPSNHKDK				
						KDVNCRLCRIQWETQAHILNHCTLSQNARRNKHNHVLE				
						IVHNFFASKNYDIDIEKSPPNVDTTLRPDLIIKTTDKNKIK				
						LSLVDIKVPYDNETNFINARKQNLDKYRNLAIETGKDN				
						NCKAAVYTIIVGSLGSWDKENNKDLQKLGLGVNEIGEM				
						ARMCIKAAIEASYMIFMNHVTNNK				
						MLSQHSRSQALVKEEGCDRLGEPTLNLAILMEMKPKRN				
						PLGRNPLSDAPYRNPGMVLNGQGPGRLPLGDAPCRDLG				
						MVSSEQGSGRHFLSGALHRRPGVVKNEQGSSHLSRRVR				
						RARKLVEPTRIRVGSWNVGSLTGKLRELVDVATRRRVN				
						ILCVQETKWKGQKAKEVDNTGFKLWYTGTAANRNGV				
						GVLIDKSLKDGVVDVRRQGDRIILVRLVVGDMVLNVIS				
						AYAPQVGHDESAKRLFWEDLDSLVRTVPSSEKLFIGGD				
						LNGHVGTTSAGFEAVHGGFGYGSRNQEGEEVLDFAVA				
						FDLMIANTFFRKRRSHLVTFSSGQHYSQIDFILTRREDKR				
						ACLDCKVIPGECVVSQHKLVVADFRFQMRARRDKQAKI				
						ARTKWWKLKEETSEVFKERVIKEGPWKEGEDANNMW				
						EKMANCIRKVASEVLGVTKGGRGVDKDTWWWNEEVQ				
						RTIKEKKECYRRLFHDRSVDNIEKYKEAKKIAKRAVSV				
						AKGRAYEDLYQRLGTKEGEKDIYRMAKVRERKTRDFN				
						QVKCIKDEMENILVKEDEIKHRWQEYFDKLFNGENENT				
						TLQLDDSFDDTNRHFVRKIQESEVREALKRMKSGKAMG				
						PDGIPIKVWRCLGDIAIVWLTKLFNHIFRSNKMPEEWRR				
						SILVPIYKNKGDIQSCTNYRGIKLMSHTMKLWERVIEHR				
					TGATTTGTG	LRRITRVSMNQFGFMPGRSTMEAIFLIRQVMERYREKK				
					GCTCTTGTT	KDLHMVFIDLEKAYDKIPRNVVMWWALDKHKVPTKYV				
					GGGTTTCAA	GLIKDMYNNVVTRVRTSDGDTDDFPIRIGLHQGSALSPY				
					CTCTAGCCT	LFALVMDEVTRDIQGDIPWCMLFADDVVLVDESRAGV				
					ACCCCAACT	NRKLELWRETLESKGFRLSRTKTEYMRCDFGTTTHEEG				
					TGCTTGGGA	DVSLESQVVPRKDTFRYLGSMLQRDGDIDEDVSHRIKA				
					CTGAAAGG	GWMKWRQASGILYDKRVPLKLKGKFYRTAIRPAMLYG				
	RTE-1_Slt	.	Setaria italica	AACTCCG CCT	CTATGTTGT	AECWPTKRRHVQQISVAEMRMLRWICGHTRRDRVRNE				
RTE					TGTTGTT	DIRDRLGVAPIEEKLIQHRLRWFGHVQRRPPEAPVRSGT	cd01650	641	890	

						LRRDSNVKRGRGRPKLTWEEAIKGDLKGWNIPKDLALD RSAWKTAIHVPEP			
Penelope	Penelope-5_SIn	.	Solenopsis invicta	TAAATAA	C	KLDPSKHSVITNHRIFFDHSFNWKDIKILDTEENYQKRLI	SSF5 6672	215	444
				CGACCAG		SEMIHIKQQTNGLNSQNDTDLLHDSYFCLLDSLKDNTNI			
				AAAACGA		DCSSNVVYKIFCNNCDATYVGQTKRQLRTRAKEHINNI			
				ATTTTAT		TIYNMVDRALLLSHPTFQQKNLELIIGLLLKNGYPLEFIF			
				GGCATAT		KYIKIRIKNLFNTKMAPIITNDQDQTKKDENTKKFIVVPY			
				TGGGTGG		IRNISEKTTSLLDKNTFTFGYRCINKINNFIKVHKDKTER			
				TGAGAAA		FLDLSLKITDNTIIIDWFHKKTFSGRYLSFFSHHPICHKIG			
				TGTTCCA		VDDIVMAAPPNSLNFILETFNSFHNRLQFTIELEQDRSLS			
				TATAGAT		QTFGTPMGSPLSPIIANIVLHDLEEKALKSINLEIPFYYRY			
				ACCATGT		GIRKRWPSIENNTKIPFDEFIFALKFILSSTFFTFNNVFYK			
				CTAGAGG		NSFDIAKSLSGMKIREEYCLLSLDAKSLFTNIPLDLAMD			
				ATAGCGC		HKENFPFRIIVSSVNTALYPIAKFLHKIINNSLPFDSRQVY			
				TCTATCC		DLNTTIKRWYDXDFISKQSYFSLHSSDSTIPKAYGLPKV			
				GATAGCA		GNITVAMDRLDYINNMEQLLHDKETYSIIKNNPIKKVEK			
				AGGATGA		FLHNNSSLDEVEKELLRMKRATSEFCKSNPNVIFTRADK			
				GGAAAGT		NFCLPPNKKTSITEFIKDIESNRCIQKERGEIRNSVIPLLHN			
				TTTTTGTT		KTLLKVETNPXTSTNNKWFVNISKTEIPPSVSNLLQLGG			
				CCAAATT					
				ATTAATT					
				TCCAATA					
				GAAGTAG					
				TTTTATA					
				ATATCCG					
				CTAGAGG					
				ATAAATT					
				TTTGAAA					
				TAATATA					
				TCTTATTT					
				TTTTTTAT					
				GAATAGA					
				TTGTATT					
				TGTCATTT					
				TGATGGG					
				TTTGTGG					
				TTGATCA					
				TCTGGTC					
				TTTTTTTG					
				TTCTCATC					
				TTTTTGTG					

AGAAAGT
TTTTTTAT
GGAACCA
ATCAATG
ATGATTG
TATTGTC
GGTTATT
TTTAATG
ACAGGTC
CAAGAAA
CTTAGGC
TACGGTC
CTGTTCC
AATTCAA
TGGTGAA
TTGTAAT
CTATTAT
GAAAACT
GTTAAAT
GTCTCGA
GGATAAA
ATTGAGT
GAGTTCG
GAGGTGC
TGCCATA
ACAATGT
CGTCGAC
ATAACGA
TAATAGA
ACGGTAT
TTCTAGA
TTTATTG
ACTTTAA
CGCTTTTT
CTTCTAA
ACCGTGC
AATACGA
TATTCGC
TATAATG
GGTGATA

			AGGGCGA TCCCATT GGTGTAC CAAAAGT CTGTTTGT AA						
R1	R1-2_Sln.	Solenopsis invicta	GGACTTT GCGGCGG GGCTCGG CCTGGTC CTTCTGA ACCGGGG CCGGGTC AGCACCT ACGTGGG GGCCGGG GGGGAGT CCATCGT CGACGTG ACGTGGG CCTCCCC CGGCGCC CTTAACA GGGTAAA GGGATGG CAAGTGG AGACCGG CTCCTTG GAAGAAC TGTCTGA TCACAGA CTTATCTC CATGGAG CTGGTCT CCCCCCC CCCCCCC CGCGGAG GTGCGTG AGCAACG CCGCCGC	TAGGCCCTC AAGCCCTG CCGGGGGC GCCCATCGC TAGGGCGC CCCTGGCG AGCCGAGT CGGCCGCC AAGGCCGA CCGGGTCC GGGGGGCT CCGGAAGT ATGCTGCAC GCGTTCCCG GAGCCCCC CAGTCATG GACCAGGG CAGGACAC CCGGAGAG GTTTTAGTG GGTAGGCC CTCGCCGGC ACGTCGCC GGCGGGGGG AGAGCCCC ACATAACC ACCAGGCC TCCCCCGGGG GCCTGGGGG TATGCGGTA ACGCATTTC CTCTCCGTG AACAAAAA AAA	MMSRACDASMPRSRPMARRAAYWWSAELAELRRATV AARRAHTRAKRRGMEDALNNAAAVLREARKALRTAIA RAKAAAWDELLSELDRDPWGRPYKIVMKRLRPWAPPV SETLDPQFLGEVVDKLFPTRGRDIPEKTGPPPGWTEELE VSRHEMVAAFARITGRKAKAPGPDGIHAKVWARASSII GEALRGPYTKCLREGTFPRRWKRARLVLLKKEGKPAES PSAYRPLCMLDDAGKIFERIIANRIIRHMSRNGPDLSPGQ YGFREARSTVDAIRQVRSLSESITGDGKVALAISLDIANA FNSIPWGEVVAALSEHFRLPPYLVAIVRDYFRDRVLEFR DKTGLQQRRDMSCGVPQGSVLGPLLWNATYDKVLRGA LPPGCHVVGYADDTFVVAGGASWVIAIAMGNWAVNC VVGSIKSLGLRVAPEKTEAIFFHDGSSRVPPRAFVRVDN TRVRVGATLKYLGLLLDGRWAFVEHFDKLAQRLGRRT DALLGLMPNLRGPRVGARRAYAYAAAMAGALYGAPVW FREALASRRIRDVLHAVQRRLARRIVRAYRTAPSAATM VLAGLPPAEFTADALAWSYARAKAVRLQGGVVTPRVA ALLRERARRRAMRAWRESLANDPPAAGSRILEAVLPHL DEWVGRPWGGLSYRTTQVLTGHGCFGEYLCRIRKEPTT QCHHCEADRDSAQHTLEDCPAWEELRGVLREEIGDDLS LRAIISQMVSRERAWVAVSSFCEQVMRQKEEAETVRER RPGGRMPPGEGNDGRGGGDGGHDPPWLPPRPPRRRRP APRPPGPAARLXKGGGGPADQPRSPPPSPQLRRRGATAP IWRGGGDPPPLRLPAPPSGHAAAGGGPLTPASPARQTRP DGPGGATLRHNAESLPPLPRRGGGRGKGLLSPPQGRKX GRAAVSAARRGPGLRRGPDTPGLYPQQHQGKRRGRDG VPLPRPRAD	cd01650	210	472

				AANCGGG CCGGAAA NCGATGG GTCCTGA AGAAACT CGACCCC GAAGCCC TCGAGGT CAGTCTC CTCGCCT CCACTTG GCCGGAG GGAGGAG CTGACCT CGGAGCC GAGGAGG AGGCGGA GCGCCTA TGCGAG					
RTE	RTE1 _Sar	.	Sorex araneus	ATCCACT ACCCAGC CACCGCC ACCCTCC AGGTCGC TCTCCGG ACACTTG GGCCGAA CCTCACG CACGAGT GAAGCCC CGCAAAG CCAGGTA AGCAGAA CTTTGCG ACCTTGA GCTCTAG GTTCAAC GAGACCT GGAAACA AGATCCT	TA	MVTPLRAIGGRPKSLRPSRRARQATESIPPARQSLASYPS RIRYANNSNNNWPHSPVTERAPNTAALRKMSKERLKKS RGRTRLPKRRKTKMIVCTFNARTLASEASIEDLMVQAQ KIKYDIIGLTETRRHQSHHAIFDTGEELFLGTCDNRGVG GVGVLVNTNLAMSIDSFECLTTRIGRLRLNRCGSLPAVSI FVVYAPTSNYDEEEIERFYMELEKFYKEDHTFYKIIVGD FNAKIGPRRSPEELHIGTHGLEWNDQGERLSEFIMSTKTI HGNSQFQKAESKRWTWESPGGQFHNEIDHIIFNRRFCLT DVAVVPKFQTGSDHRLLRAKFYFTERGEKSAKFKSKKA TPRMTTNWELFGTIAATWEDAVLDNIDEEYDRLVQHLH VCAKNAESEKATNRRLSSETLELIRQRGLARASGNHKLT SELTKLCRDAIKEDLKERRAAVLASAAEAGKSIRNARLS FANYKTKMTALRRPDGSITSSRKAMEKIIHDFYSDLFDS HVHLPTYQIPQDGYVVPNVLPSEIRHAISLVKTRTAPGP DKVRPEHLKNLPPVLINTLARLFTRYLSECKVPSQWKTS RTVLLYKKGDIHDIGNYRPICLLSVVYKLFTRVILNRIGR TLDEGQPCEQAGFRRGFSTIDHIHTVTKLIEVSREFKMPL CLTFIDLKKAFDSVETEAVIEALAKQGVQTQYIKILRELY CGFTTRISPFYKEVIIDVKRGVRQGDTISPKLFSAALENI MRRLEWEGMGVKIDGRQLHHLRFADDIVLITPNISQAA QMLADFDHECGKVGLQLNLNKTMFMKNELVPEAPFAL	cd016 50	583	837

				CTGTCTG CCTCCTC CAATCTC CGGCATC TGAGGGA GGGGTCC AACCCAG ACGCCCC ACCCTAC CCAAGAT AAATACC TCACTGG CCGACCC CCACTAC ATCCACT ACCCAGC CACCGCC ACCCTCC AGGTCGC TCTCCGG ACCGCGG ACACACC ATAAAAC ACTTAAT ACCCATT ATTCCCG CACCATA TTTGATA GGGTTCA AATTTGG TGTGAAC C	TGAAGTGG CCGGAGTG GCAATGCCT AAATCGAA TGCGGACG AGAATGGG GCGGTGTA AGGTCAAT	NGTNISECSSYVYLGREINMRNDLVPELRRRRKRAAWNA FKSVEEVVKRTKNLRLRAHLFDSTVLPALTYASETWAL RKQDENAIRVSQRGIERAMLGISRLTQVREGIRSSDLRR RSKIRDAVSFAKASKIRWAGHVMRFRDDRWTRAVTDW IPRDVRRPRGRPPTRWSD			
Ving i	Vingi - 4_SP ur	.	Strongyl ocentrotu s purpurat us	GGGGCAC GCAACAA GAGTCCT ATCTCGC ACCCCGG GGCAACG TGTCTGA CACGGCC	MKMPNRAASQSNSAEGPLSRPALTVLSWNAEGISDAKE EVLAHMCKQMNCDVLCLQETHRGSDNPRPKIKGMKLA IERPHRQYGSAIFMNEKIRPDSTSYTHNNDIEVLNVTIKG LTISSIYKPPNISFSYPSTTTNRHIVIGDYNSHSTQWGYDT TGDDGALVETWAQNNDLKLIHDPKLPKSFASARWKRR YNPDLVFCSRLLAPTSRKFVAEPIPRSQHRPIGIQLTAVIT PNKTPFKRRFNFAKADWPAFTEHLDHALTHLPADPKNY ESFIDIVKKVSRTHIPRGCRTTYIPGIDEESSQLLANYTTA	cd016 50	476	731

				GGTTGTA GCAACTA CTTCGAA GATCAAG ATTAAGA TTAAGAT TGAGAAT AATAATA ATTTAGA CTTAGAC TTAGAGT AATATTA AGATTAC GATTAAG ATCAAGA CCAACAT AACATA	ATGGCAAG ATGGGGCT ATGCCCATG AAGGAGAT ACTCTCTGC GACTGCGG AGAGACTC AGACGATG GACCACCTC TTGCAGTGC GGCGGAAC GGGTTGCA CAGATGAA GACCTGTCC CTTGTTAAT GAAACGGC CATTGCAGC AATCGAGC GATGGAAG AACTTAATT TAATTTAAT TTACTTCTA AGTTAAGG AACTTTACA TTACTATCT TGATCTAAG GAGACACG AACGAACG AACG	YDTDAFSEDTINLGEELISRISQRHRQMWQHMIETTDLR HSSRKAWATIRKLGIDSTRAQQHYNVTADEVANQLLKN GKAPKTTNSHRHKLPGVDQNMSLFTRPISEEELQKGIST LHVNKAAGIDDIRSEQIKHMGSVAQKWLLDMFNNCIEN TCVPKQWLRTKVVAILKPGKDSNEAKNFRPISLLCQTY KLLERLILNRIGPHVDSQLNCEQAGFRPGKSTTGQLLNL TQHIEDGFQNNLITGAVFVDLSAAYDTVNHRRLQTKLH DITNDLKLTRFIQKMMQNRRFFVELDGQKSRWRKQKN GLPQGGVLAPTLYNIYTADIPSTENTRNFIYADDTCITAQ HSTFEAVELHLNTALASLSEYYRANYLRANPSKTQVCV FHLKNREAKRELKLKWNNVQLNHCEHPVYLGVKLDRT LSFKKHIEKTIGKVESRNAIIGKLASSHYGADPKTVQSAA VALCMSSAEYACPVWSRSAHVKKLDTTLNNTYRKITGC LKPTKVEEIYHLAGIGSPNQRRTNTCLKERMKQATDDR HMLHGYLQPHKRLISRHSFLDSVPPPEQNPPPSAKPVLIK GLT			
Rex 1	Rex1-1_SP	.	Strongyl ocentrotu s purpurat us	TGAATGC GCCTAGC AGGCAGC CGTCTCT GTTTTTGC TCTGAAT TTTACCN TGTTTTAC CACTGAT AATCTGA GTGTATG	TGATTGTAT TTAATATAC TCTGAAAC AAATTAAA TTGTCAAAT TACAATTTC ATATTCCTT AGTATATA ATTAATTCA TAAATTTGA GGTCTGATT	MDTYKVYVLFYIALCIIIGLRGHSSYSKHETAFVPVRAPS RAHKQLHRSGYHKTPIKYYANASATFTLPIICCGDVESN PGPFELTSCNTNKRYKYTRGELVNINSNSCGKSLRLPVQ TWKTISELEIGKEFKTRRGCRAGVRKLKHFKHPIPIILPT ARTKLPKKYPRFPRRRCLTSVPVAPQSGCDLPSFLLSNV RSACNKFDELSAVISYLHVDIAAITETWFRNDCPLDNFSI PGYYLCTKSRTNRIGGGVAIYVRELLNPSPADVQVPINL EVTWVKIRPPRLPRSISCIYCAVVYYPQPDVELETELIDH LETTIDVILTNHTDAGFVILGDFNRMNIDPVLLTHSFRQV VDCPTRGAHILDKIITNIHQFYDKPQICSPIGMSDHNSIM WFPLKIKPKLPNTTRYVFNRPMTDSAIRAFGQWISSYDW	cd016 50	658	909

				CTGTTTTA TTGTTGA TCAACTG TCATCTT ATGTACT TGTGACA CTGCATG ACATTTTT GAGGTGA TGTATTT ATATATT AGATACT GGAAACT GTTTTTCT AAATTTT GATTCAA CGGATGT GTACACA GTACATG CAAGTAA AGCCATT TTACACT GGGCTGT TACACAT GTACTAT ACTGCGT GGATGTG TGAACGT CGTTCGT GTGTGAC TGATTTG CCTATCC GTGGTGT GACATTC ATCTACA TACAATG AATAGTC TTGATAC GCACTG	TTTTACCAA GTTAGTCAA TGTATTTTT TATGACATG AAGAATGC GATGCATAT CTATATTTA TATTTTAAT ATCTCTAAT GTTCGCCTT ATAATATTC ATATTTGTC TTGCAATGT ATCTTTGTA TTTTATCAG CATGTACAT AATTGTTTT CCCTCTCTG TCCATCTGT CCCCTTTCT CCCCCCTCT CTCTCTCTC TCTCTCTCT CTCTCTTCT CTCTCTCTC TCTTTTCTC TCTTTCTTT CTCTCTCTC CTGCTTTCC CATCTCTCT TTCTTTTTT AACTCAAA TTTTTAAAT TGAAATATT GTATAAGA TTAATGGTA TATAGTATA TACACGCTC TTCATAATT ATGTAGATT	PEVTEDVDVNVRCGSFYSTLNNAINIYFPVRKVLLHFRD KPWMTGEIKSLIHKRQIAFHQGKGILWKRLRNKVNSSIQ KAKVKYYRSRIQKLKKSNPSRWYQEIRILTKGSRQSLNI PVDGVDPSERDKIANAINEHLISVSSDLQPLNLAALPAYL PSRSLPEVYPWQVNKVLSKIKCSTANGPDNISARLIRLFS VELSGPLTYIINASFSQGKVPASWKCAIVVPIPKVNPARI DKLRPISLTDHFAKVIEGFIAKWMYQDIAQHIDPCQFGN QKRLSTSHCLINILHYVFLNAEVPKTSSTILLTDFQKAFD RIDHTVAIKKLIEIGVRPAVVYWISDFLSERKQRVRYHS DLSEWCQINAGVPQGTKLGPLIFLVMINDFKSSESIHMF KYVDDLTLVECRKTNQISNMQEAVECLSLWADINKMQ LNPTKCIKMDICFSRVPVNQSSIILENCVLSTVSQVKLLG VTIQNDLKWEAHVRDIEKRANAKLYMLRSLGKFGLSHE DLITIFIGYIRPILEYATQVWSGGLTSAQKQRIERLQKRA LRIIFGKDYHSYANALTKSKLQTLESRRNDLCVTFFKKT LSCTDQFKQFIPQRAHNRTLRRTNKIPEPRCKTKRMQNS PIPDLIRLYNKSFL			

RTE	RTE-6_SP	.	Strongylocentrotus purpuratus	GGTATGC GGCTCTT TATGTAG GTAACCC CTTTTAAT TAGCTCA GGGAAAA TCCACCT CCCCTGC CTTGCAT AGATGGC ATTGGAA GGTGTAG CGAGCCT CCATTGT AGCTATA	GTAGTGTG ACATCTTGA AATGTATAC TTGCTTTTT TCCAGCAAT TTACTCATG TGCAATTCA GTTACTTTT TACTGTGAT ATCTCTTGA AATGTATA ATTGTCTTC ACCATCAAT GTTTTATTC ATGCGCAA TTCAGTTAC TTGTAACTG CGATACAT GGAAATGA ATCAATAA ATACCATTT ATCTATCTA TCTATCTAT CTATCA	MMRQTNKNGYRPGCRPGEQEPRPPVTNQAEVGKLNTG FKSKSRIKIIDCKAATLISSFNVRSLKSKEKMGELTALAE EKRIDILCIQEHHIYHENVDIQMKDMGKGWVLAMSSGT KNTVNSTVGGVGILMSPFAYKSLEYLEKVSSRILLAKFS GNPAPTVITCYSPTNCSPDNEAVNFYSLLSDVIKKIPQHN MKIICGDMNAQVSSADSKGFTFGNRTNRNGQYLLNMTF ECDLIACNTRFQKAKEKLWTFTYPNKSQAQLDYILVNK KWRNSVKDCQAYNSFGPLNSDHRIVSTKMQLSLRSNKK VKAKSNLPHDWQLFKSDCSIKEAYTVEVNNRFEALYQD IENPTANEIYSALIEAHEGAATKYVPKRKGKKKRVPWE NEKIREKRKCLKVAYNKKNGAPTAEHAETLDKAKKDL QDAYDEEQSNYIAQKCTEIEIGAVQQQTRKVWDIVKEIT GKXSTPVDQLRGKTXQERLNQWKDHFSNLLGQPPPDN NNPVSKFTIAELKVAVTSLSNNKCPGVDGIPAEVWKSG ALDQQLLEVCNRTLIHGEKPDLWSRSIIIPVPKKGDLSTP QNYRGIALTPIAAKVFNKMLLNRIRPXLEPLLRKNQNGF	cd01650	560	825

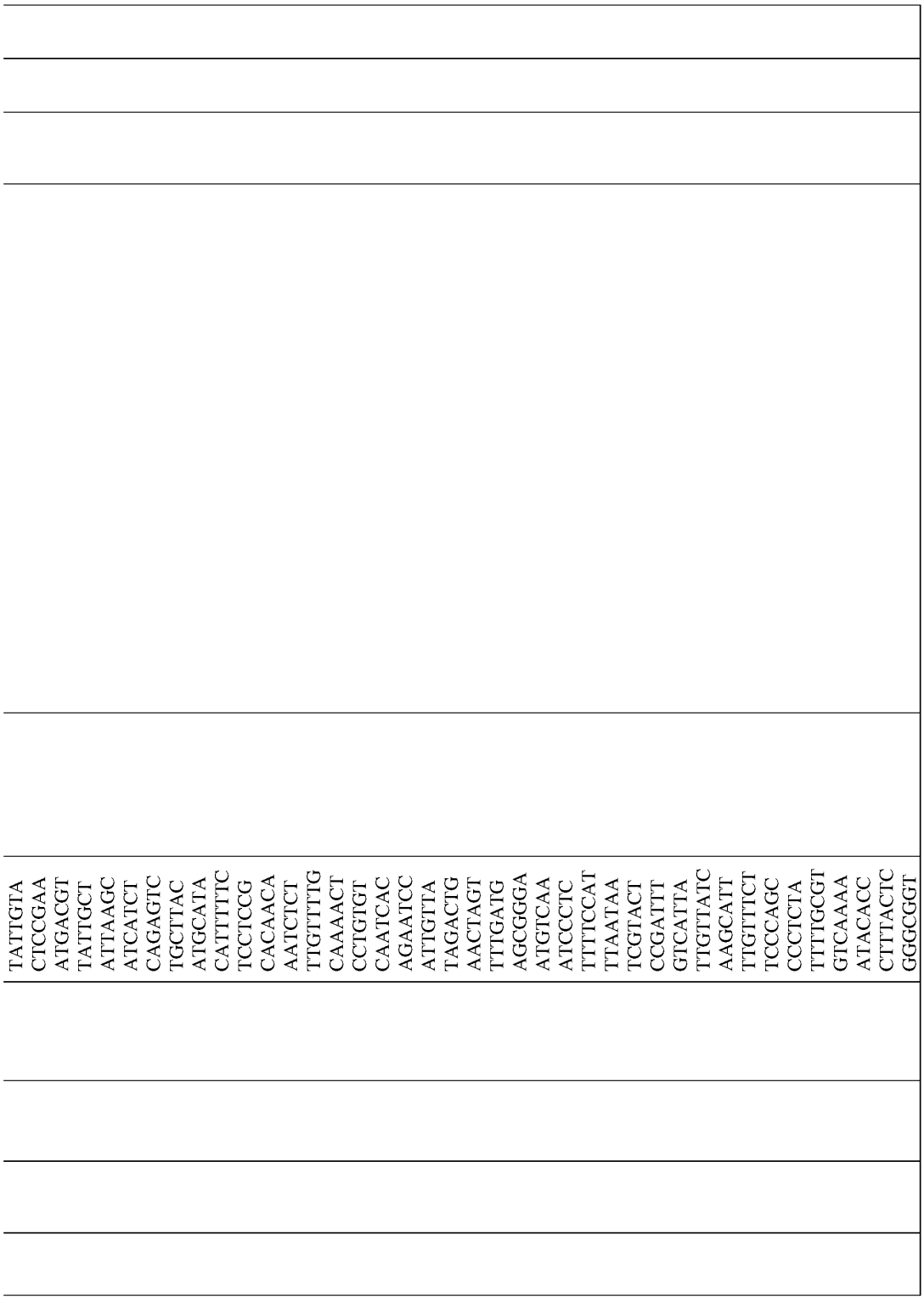
						XPGRSTVXHILTLRRLVEGIKAKHLTAVLTFVDFXKAFD			
						SINRKKMLKILXAYGIPHTIVTAVGVMYKDITAQVRSP			
						DGETDSFTILAGVLQGDTLAPYLFVIALDYALRTAXEGF			
						EDLGFTLQERKSSRYPAVMITDTDFADDIALISDDLEKA			
						QELLXQVESAASQVGLQINSTKTEFMLRNLREGEICTAD			
						XTKSKQVQNFQYLGSWVISKQDLNIRKAKAWTALNKL			
						TAIWKSNLGRKLKVRFFTVETVLLYGSEAWTLTSSLEK			
						QLDGCYTRLLRAALNVNWRDHMTNIELYRDITXISARL			
						RDRRLRFSGHCWRSKSE			
					GGAATAC				
					CATCAGA				
					GTTGGAA				
					CATGGAA				
					TGTGAGA				
					ACACTGT				
					ATCAGAG				
					CGGCAAG				
					CTGGAAA				
					ATGTTAA				
					ACAAGAG				
					ATGATGA	MVITNTWFKEHPRRVYTWKSPGDIVRNQIDYIAINSRFK			
					GGCTAAA	RAVKQSKTYPGADCGSDHVPVICTLQCKLKKLKRPKVV			
					TGTCAAC	KILDFKQLSKPDIRQEFTVQVKNRFELLIDEGEENTWET			
					ATATTGG	MRNILVETAEEYVPTKERKRRNKWMTEEILLMMKDRQ			
					GACTAAA	KIGNRQCKKYRDLDKQIRQKCNEAKESWLNEQCVEIEN			
					TGAAACA	QFDKNSNVYQRINEISGKKSGCSNPGCIKGKDGSILFEM			
					AGATGGC	KDKMSRWTQYIGELFEDSREAMPSFADEVKGEKILKSE			
					TGAATAA	VRWAMSKMKKNKAAGPDGVVTEMFEALEEYGVDRLT			
					CGGCGAG	EIINKIYDDGTFPDELGRSIFITTLPKKPGAVDCEQYRTISLI			
					TTTGTGA	CHITKIILRILLLRARSRITPMVGREQFGFVKDAGTRNAIF			
					TAGATGA	VVRNITERAVEMQKDVYYMCFIDYSKAFDKVKHSELFEE			
					GTACAGA	LKKLDIYRKDLRLLQSLYWNQFACIRVEGECSEYINIRR			
					TTGATTT	GVRQGCVMSPDLFNMYSEIILRELEDENGIQVGGHSITNI			
					ATGCAGG	RYADDTVLFSESEKDLQRLLNVVVKESERMGLEINYTK			
					AGGTGAA	TKCLVVSKKKSSPKCNLKLKEQIIKQVSSFNYLGSMIAED			
					AAGCATG	ARCETEIKRRIALAKSAFSKLGKILTNRTMSMRTRIRVLN			
					AAAGAGG	CYIYPVLMYGSETWTLASDTKKRLESCEMWFLRRMMK			
					TGTGGGG	IPWTDKVSNEEVLTRAGVQRKMIREMLIRQLKFMGHIT			
	RTE-3_SP	048B05	Strongylocentrotus purpuratus	TTACTGC	AGAGAGAG A	RKEGLENLALTGKIVGRRSRGRRRILWMDSLQKLLEER	cd01650	321	578
				TGGATCC		GVKELGVVLIQKARSRGLWNSMIAKVKRYGT			
RTE									

AGATATC
TCAAAGA
GTGTGTT
GGGACAC
TGGTGTG
TAAGTGA
TAGGGTG
CTCCTAG
TGAAATT
ACAGGGA
CACCCTT
TTAACAT
CTCAATC
ATAGTGG
TTTATGC
ACCAACG
GCAGATA
GTGCTGA
CGAAGTT
ATTGACA
CCTTTTAT
GATTCCT
TAGATAA
AGCTAAA
TCTCAGT
GCAAATC
GAATGAC
ATTGTCA
TTGTCCT
GGGAGAC
CTGAACG
CTAAAGT
CGGTCAT
GGTACAG
AGGGCAA
GCAGTAG
GTCCGCA
TGGAATT
GGTGAAA
GGAATGA

			ACGAGGG						
			GACAGAT						
			GGGTTCA						
			GTGGTGT						
			GAGGGGG						
			AGAAC						
RTE	RTE-8_SP	.	TTCTCTAC			MRNSMKLHTGSVVFRVNKDRAVRSGLPSTVPSMLGXK			
			TATAATC			QQQQKDDTSGITTMTFGQWNVRTLLDREGTGRPERRTA			
			AGTGTTT			LVAMELEKYRIDIAALSETRFSESGRLDDMGYTFFWSG			
			TTATCAC			KPDGVRREAGVGFAIRKEILAKLTEMPKAVNDRIMTMT			
			AGCACTG			VPLTKKGNATLISVYAPTMTSPDETKENFYNQLRRTLRD			
			AATAGAC			TPRTDKLILMGDFNARVGRDYEKWQSVIGRHGIGKCNT			
			ATAATCA			NGELLLALCSEFELLLTNTVFKQREEHKTTWMHPRSKH			
			ATAAAGT			WHLIDYIITRQRDRMDVHSTRTMRGANCLTDHQLLRSK			
			AAAGTAA			IAFALRSKRRKQGATKLAKLNTDRLHVNTQRLNLEQD			
			TATTTGG			MEKALTNWGVQEGMTVDQTWASLSKVVYETASNTLG			
			TTGGTTC			KPERKHQDWFDPEDSELLILMKKRDETHQRVLQTRSTR			
			AATTAGT			STTTAYKEACRHLQRYTRKMKTTWWEKKAEDLQRSA			
			ACTCTTA			DNNDMKSFYTGLKEVWGPQTRSTVHLKSLDGNTTFSD			
			TCATAAG			NKRVLERWSQHFETLLNQPGDIEPAARDRINQRPVVTA			
			AGTAATT			LDDIPTRKELNLAIASLGDGKAPGVDGIPSEIWKHGGAT			
			GTAATCG			LTNSLLMLIQQAWAEGSVPQEWKDANIVTIFKKGDRTQ			
			CCTCATT			CGNYRGISLLSIAGKVFARILLNRLNAHITPDILPETQCGF			
			CTTTTATT			RNNRSTVDMIFCLRQLQEKCIEHNRSLYVVFVDFTKAFD			
			AAATGTT			TVGRTGLWQLLRKYGCPEKFTGMIESLHTGMMAKVKE			
			AATTGTT			AGETSDSFPVSNGVKQGCVLAPTLFSIFLSAMLEEAFAD			
			TTTGCCA			FEEGVYIQSRQDADLFNVAHFKAKTKSTQILVRELLFAD			
			TGCTTAG			DSALVAHTPEQMQHVIDVFSSASKKFGLQINIKKTEVLF			
			TTGTAAC			QPGTDHREEEEILVDGTALNQVDDFTYLGSTISKDGRID			
			CACTGCC			SELVKRMSKASSAFGRLRTRLWNNHHVSTRVKCKIYRA			
			TTGTGGT			IVLSTLLYGAESWTLYKSQVKKLHAFMMRHLRAIMRIS			
			TCAGGTT			WKDKVTNKEVLERANLPSMEDLLIRKNLRWTGHVIRM			
			CTCTTTA			PSERLPKQVLFSQLPAGERGIGRPRLRYKDTIKRNLKRR			
			GAATCTA			QIETKTWTTAAGQRAVWRTAVK			
			ACTAAGG	AAATGAAG	AGCTACTAC				
		Strongylocentrotus purpuratus	CTAAGGG	GCGGTCATT	TACTA				
			AGCAAAC	GTCGCTTCG					
			CCTAACA	CGACAGAC					
			GAAAACC	AGCTACTAC			cd01650	589	869
			ATGCCAC	TACTA					

				CGTGACAGCATTGAGGGCGATTGACGATACGGAACTCCCGACATCCTTCGACCAGTCACTGCAGGGTCTAGTGGTTCACACCCGTCCTGGCTTGTCCTCGCCGGTGTGAAACTGCCACCAGTTCCCCGAGAGAGTGCCGTTAGTAACTTGCAAGAGCTACCGGTACTTATCCACCTTTGCAAGCAGCCTTCATT					
Daphne	Daphne-11_SPu	.	Strongylocentrotus purpuratus	TAGTATGTGACAATTATCATGATGGAAAATTCGATTGTCCAAATGGGTAAAGCATTTGAATCTTGAAAGAATTGTACGGTTGTC	TAGTATAACAAAACCGTGACTTATTTATTCTTTATCATGCAACTGCTTACTGCACCTGCACTTGCACCGCACCCTCTTGATCACTCAGTTTTTCGAAGAGAGTTA	MNSHAHTPDILVDDINESKYYTPQEFKNSFSYKPDNISLLHINSRSLNKNFEYVENLLHSLNNFAFSVIGISETWLHQNSPDMFNLPNYKLIRADRKGRRGGGVAFYITQNLNFKIRSDIKLTQAESLFIEINNSNVKNVIIGLIYRPPDSNCELFCDELDFYLHKIGDENKHNFILGDFNINFSPSSDNSNSTNCMQLMHSYGFFSIINKPTRINPHSSTQIDNIFSNVHNNKTIGGILCSEVSDHLPIFLTCECKLSCSKSFDELYRKESKRNIELLKQDLFFEGWAEVYNVNDVNTAYKHFSNKLQYYYEKNIPLGKMSNKRNKPRNPWITKGLLKSIQTRNRLYKLHLRNPTDTNLNTYKIYRNKLTKLIRTSRRLHFSNKIKKASSNTTATWKIIKEVMGKKTNPPPTDNITLNGTKIHDPSLHANSFNSFFTNIGPELASKINRTHAHFKDYLSNQNQESLFLNPTNP	cd01650	531	787

				ACTAATC CTAAATT GTCATGA TTGTATTC CTTCCAG GGGACAT GCAGTAG AACTTTG TATCATG CAGGAAT GTATTCTT TACCCTC AAATGGG CTTGAGA CAAGTAT TCGTTAT CCTTTGA TCTTGAC GACTGGG ATTATCA AGACCGA AGTCCTC ACTCTTT ATCAATC ATTATTA TCTGTAT AGCATTC CAAAAAT TTGAGTC TATCTGTT ATATTGT AATTGAC TGAACTT TTTTTTGT TTTTTGA GCGGGAA TGTCAAA ACCCTCT TTGCCAT TTAATAT	TGTCGCCAA TTCTTTTTA CGAGGCCG CCATCCATC TCCAGCCTA ACCGCTCAC GATGGTGG TCTCCTGCG TTCTAAATT ACCTCTCTA ATTTATTGA AAATATTTC ATATTATGT TATTGTATG CTTCGTACA TGTTTATCG TATGATATA ATTGTACAT GATTTTTTA TTTAATTAC TTTTGTAAA CATTTTATT CATTGTATT TGTTATTGT TCGAATTTT TGTACTTTG AATTTGAAT TTGTATATT TTTGAACGC AGAAATAA AATCAATC AAACAAAC AAACAAA	SEIINITQSLNNSKSHGFDRISMSLLKEIMYPLAKPLTHIF NHSLSQGVFPDLLKIAKVNPILKKDDPHEISNYRPISLLPS ISKVLEKIVHNRLHKFITKHKILNSNQYGFRKNYSTDLA LIQIYDKITSAIANKEHVIGIFCDLSKAFDTLNHAILLSKL SHYGIRGQPLLWFKDYLTNRRQYVTFNSYDSDPLLIQCG VPQGSILGPLLFLLYVNDIINTSPLLSFILFADDTNIFYSHK DLNSLNNTLNYEINKVSDWFKSNKLSLNTKKTHFMYFK HHSHNTETPIYIKIDGMLLEKKTNSKFLGVLIDETLTWN DHLHQVTMCISKSIGIISKLKFLLPHATLFLLYNSLVLPYI TYCNSVWANCGSTKINSIFKLQKKAIRICTGSHFLAHTD PLFYKLKTLKIFDLNTTQVAVIMFKYINNLLPPSFDNMF RFNNSVHSYSTRISGNFHLTNPKLSITHKSVRHSGPDIWN DFPDNIKSCSTIYSLKATLKRQIIQSYAPHSLN			


TATTGTA
CTCCGAA
ATGACGT
TATTGCT
ATTAAGC
ATCATCT
CAGAGTC
TGCTTAC
ATGCATA
CATTTTTC
TCCTCCG
CACAACA
AATCTCT
TTGTTTTG
CAAAACT
CCTGTGT
CAATCAC
AGAATCC
ATTGTTA
TAGACTG
AACTAGT
TTTGATG
AGCGGGA
ATGTCAA
ATCCCTC
TTTTCCAT
TTAATAA
TCGTACT
CCGATTT
GTCATTA
TTGTTATC
AAGCATT
TTGTTTCT
TCCCAGC
CCCTCTA
TTTTGCGT
GTCAAAA
ATACACC
CTTTACTC
GGGCGGT

				CGGGCTCCGCTCTGTTGCTCCTAAGCAATTATTGTAACTGTGACCAAGATGGGTCTTATTGTATTTTCATTTGTATTATATTTTGAAGTTACCCGCACACTTGCACTTCGTACACACACAAACGCACCAGCATCACACATGCAGTTGTTCATCCTCCAAATTTTCGGTACTTTACATAATACGTATCGTTAA					
CR1	CR1-31_SP	.	Strongylocentrotus purpuratus	GCTGTAGTAGTCTACATGTGTACTGACNCTACTGATCTGTTCCTGCCAGCAAGTAATGTATGTATTATTAGTAGTAGTTCTTCAT	TAAATTCCGGCAACGTTTTTTTTTTACTGGCACTATGTATATATACACCTCTCACCTGCACACCTGGACATCCCCATGTCGCCACACACATGCACTCGCACACATCT	MGRKPKYPCGVCSRAVTTRTTGVQCEGTCLKWFHAECIGLTSTEYVAFDSSSIVWICNHCGAQNLAHSCLSSLQLXPDTNRFGVLDPDGQNVISEENSNSCRPTXDSSNPSSPGSPAYSSSPQNIKQRRSKHKPRPRVLRFLNVNCQSIRAKMQPFSILVHQHNPDIVIGTESWLSDGINSNEVFLDEQYSILRRDRSNDPHGGVFIATKTDLIVTREEELETDCEILWCRIEMAGSKTLRVAAYYRPKEDDEVSLAELQRSLSRLDRRHLVVLAGDFNLPGWKWDGPYVGPCAYPALHHQFGEIINDHGLTQLVKEPTRQANTLDLILISNPTLVNSVKVVPGISDHNCPVADIDMSPPRRYQPKRKVPMYKKADWEGFAELMQHVADYICSKSSSASVHDLWQLFKSGIEGGIQTFVPHKMLKQKQDLPWITSGIRKLIYKRDRLHKQVTRLRKQNMLVPFPL	cd01650	632	884

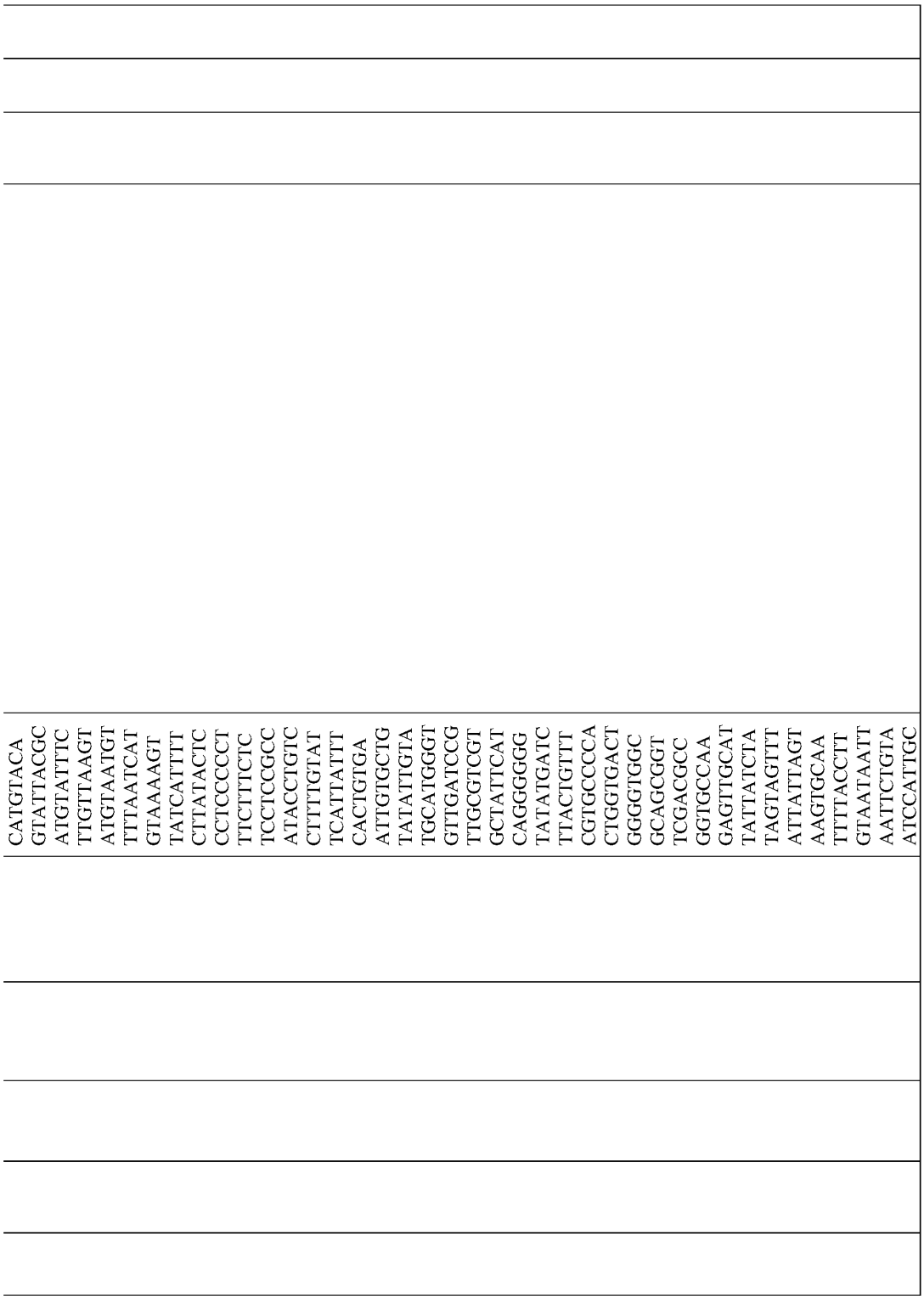
			AATTAGA TATTGTG TA	GCAAGAAT CTTCAGCAG TGAAGGAG GCAGATTAT TTTGAAGA AGAAGAAG AAGAAGAA ATATA	ETELQHLKRLIQAESRKAYWSHLESIFTADGPNEYEGM KRFWKFVKNNRTDRCGIHTLKTDEHTITSPQDKAEALN KQFKMAFTNETDVPADLLPETSPYDSMPDIHISTEGIVK MLKKLKVRKAPGPDGIPALVLRKLSRVVVAPSLRAIFDVS YRTSEVPDDWRLANVVPIYKKGDKTDPENYRPISLTSIC KMMEHIFASNIMGHARRNSILRPLQHGFQAKKSCELQL TGFVTDLLNNMEDNNQTDIVITDFSKAFDKVGHQRLLK KMEYYGVRGHNIDWISSFLSNRKQQVVLDGHTSSTVEV QSGVPQGSVLGPCLFLFYINDLPNQLSTKVRLFADDTVL YHTIQTPEDAELLQQDLDQLCGWSSKWMMALNPKKCQ VITVSRKKTKYVHRYSLNGEILESVSAAKYLGITLTADL NWTQHVSATCKKANNTLAFLRRNLRVNSTRLKTTAYQ ALVRPLLEYGSAVWDPHTIRCVNQLEMVQRRAARYVL NRYHHTSSVSDMLSQLGWTSLQKRREHQRLTLLYKMH YSLVDFDIHHYIQPATVVRTTRSSRPYCYRVMQSRTEQH IHSFFPRTVRQWNALPLEVVLAPSVEAFQRQLKTGGK				
					MMADTSTCVLQLNAEGPYPLKARSDYPHIAKENWVNV ICLQETHCETEEQLHIEGFDLICFIPHKKHGIASYVKQGM KASTTGRSPVNSALQWSTFEAMGTHFCNLYKPPPVDVE IILFPKMPRNSFVCGDFNSRSTAWGYPXTNKNGKLIFEW LESQELVILYDSKDNPTFYSGSXKSWTCPDLSLVSSNIRH TCSRTVLPKFPRSGHCPIIISSDIKLQLSNLPKKRWNFRKA DWPKFQELTNNKLTSRLPDPSTSNVVNDSYSAFCSMLHQ AAKLSIPRGRRHPYIPCWDEECDRLYESCAKSKTIEERE AVGNDLLISLGEKRRIRWEEEVSEINFTHSSRKAWTTIN KLTGRSSAPKQCPVTANSIASVLVNNGKWQDKTTEAKA HSRSVNSEIKNLLKSTPESTSMSEAFTLDEVETALKSIKS GKAPGNDNIHTEFXXNMSRKATSWLQAFLSTCMTTSTI PSKWKTAKVIAILKPKKPADEPKSYRPISLLSHIYKLLER LLLARIVDTVEEKLPTTQAGFRKGKSTTDQVVRLVNNIE TAFQKKQKFGAVLVDLTAAFDTVWHRGLYLKTLQTIPD VRIVKFIMLLVQDRSFVLETSNGECSRKRKLRNGVPQGS VLAPILFNIYVSDMPKGECMQYSYADDTALGFDNHSFE AIKTSLEKDLKMLAEFFNRWHLKMSIPKTVSSVFHLAN RCANRELNISLDGSRLNFVSTPKYLGVTLDRSLTFHEHA KATAQKTQARVSLLKKLAGTGWGANFTTLRTSTLALAF STAEYAAPAWSHSTHVHKVDVVLNDAMRLISGTLTATP VSNLPILSGIPPAQLRRDAQYTKLAQKHDRDKCLVPAPS AFEGQRLPRRHFATQAAELPLEKPPSTLATWIAKRWSDL QWNSSTTNLKEFIPKPCAKPSGSDLPRKAWVNLNRIRTG				
								cd016 50	473	728
Ving i	Vingi - 3_SP ur	.	Strongyl ocentrotu s purpurat us	TAGATGAT GTACTCCCT GATATTCTT TGCACGAA CGAACGAA						

						VGRTQYFLHKIGAXSSPNCACGEIQTMDHVIDDCPIYKS				
						PHGASGLRELDDETICWLNKDLPI				
						TDTDRGKQSKFKRKLRDRKLQIFYTNCDSLLNKRNELL				
						LLLQSHDVDLVLLTEVFPKNHQNPLTHRDFILEGYDLHC				
						NLEDGNRGVAIYIARDLYHLVGECDLQHSFHESVWILIT				
						LRGSDKLLIGCMYRSPSLDNHEELKRAFTQVTSINSSHV				
						LIAGDFNYPNIDWKLGRARSHTEQEFIETVADCYLHQHI				
						GEHTRYREGQRSSMLDLVFTNEDGMINHIDYVPGIGRS				
						DHIGILFGLTVYTPDTHTPTPHLQFYRGQYDDDMCKELD				
					TAGAATAC	QIEWESELENKSVEDAWEFFLVTLTEKMEKFIPKSKPKH				
					CCGACGCA	RRKKQRWMTREALRKQSQKHRAWKEAQRTKHPVDIM				
					GAACTGAG	RARRLAKQLQTLTKQLCRNFEKDLAMNVKSNPKAFWR				
					CCAGACAC	YTNSKLKTSCRLGELRTPDGERISTDKGKADALNTFFAT				
					TTACCAAG	TFTDEDKGHIPKLPPRHKGETLTSISISEDDVYKKLVHLS				
					AAGACACC	KYKCAGPDGIHPCVLKETAASVSEPLSIIYNASLKEGKLP				
					CAACTACG	SAWKVSHVVPVHKKASKMDVKNYRPISLTSVPCKILES				
					AGAAGGAA	LIRDEIVAHLMENDLLCDEQHGFVPGRSCVTQLLTVME				
					GAAAATGA	AWTEMLDKGTPIDVVYLDFKKAFDSVPHLRLLSKLETY				
					AATTGAACT	GICGQIKEWIADFLQGRCQRVRVNDKLSDWTKVRSGVP				
					GGAAGCCC	QGSVLGPTMFVLCINDLPDGLQGSVKIYADDTKLYGKA				
					GACTGCTG	STKDDLCSIQSDLDALMEWSNTWQMHFNTQKCKVLHI				
					AGAACACA	GPGNVKHIYALGDDVLEATSEEKDLGILIDDELKFHNHV				
					GCATTCTTC	AYAVKKANGVLSCIRRSFECRDRSTIPKLYKGLVRPLID				
					CAGAATCTC	YGDSIWYPRFKTDSKTIERVQKRATKMIYDIRNRPYIER				
					GGGTAAAG	LRALNLPSLEYRRRRGDMIQVFKIIHGLERIPVDTFFEHS				
					GTAAAGGT	TSNTRGHSLKIVKQRCSTKIRQDAFSQRVINDWNTLPEE	cd01650	509	757	
					AA	VVTATTVNAFKARLDRVWKDKRFCSPYE				
					TACAGTCTT					
				GACCATA	TCTTTAGGT	MVDRVSLWHKLLRSGVDGKVFTVIKNMYSQAKSCVIV				
				TTTTTGTT	CAAGTATA	DGQMSEAFQCDVGVRQGENLSPLLFAIYLNDFEDFLSK				
				TTGAAAT	GTTTTGATG	HFKGLKFLVENVCADILDKEVNKDLFLQLYTLLYADDT				
				CCCTGCT	GTTTGTTTG	LILADTAEELQVALNALSQYCKIWKLSVNAMKTKVLIFS				
				TGACATT	TTCTTTTGT	RGKVRKFPVFRFGEDEIDVVDDFVYLGVQVNYNGSFKK				
				TACTTGT	TCATTTTTT	AISKQVSQARKALFSMSLKVQRLKLPLDIQCELFDQLVL				
				CTAAACG	TTTTTTTTG	PILIYGCEVWGYEDLSQIEIFHRKFLRSALCVNKFTPDCM				
				GAAGAGA	TTAGTATAC	IYGETGRAWLANSIKHRMVSFWLRIVKGSPSKYSVKFY				
				ATTTACT	TTTTTGTTT	FLQRLISQNVDMSTSFKWLSCIEKILNHAGLGIVWLKQG				
				GCGCTTT	TGTTTGTCT	EGFSNSWILSTLKLRLNDISKQDVMATIWTNSVCLNYRI				
				TATAGAC	TATTACATT	FKEDCLFERYLVLLSKKDRITLSRFRCRSHRLPINNNRFK				
				TATAGAA	ATATTCCAT	EAVNEISIECPCCDMHESGDEFHYVFVCPFFRKERELYL				
							NA	NA	NA	
CR1	CR1-24_SP	.	Strongylocentrotus purpuratus							
RTE	RTE-16_SP	.	Strongylocentrotus purpuratus							

				AAGCTTTTGAT	TATGTTTTTGTTTGTCCAATGGACGATTATGGCGTATTTTGTTAGTCTGCATATATCTTAACAATCCAATCACAACTCACTTACATGTACAAAAATGAAACTTATAATTTCTTTTTTAGTTAAGACAATAATGTATTTAAAAAAAAAAAATTGTAAAATTATTTTTTGTTTTTATCTTTCTTTTTCTTCGTGTTTGTATTTAGATTGTTCATGTGGTATTTTCCACTTTTCATGTATGCTTGCCATATTGTATAATATTGTATTTATGTTTAGCATAGGCTAGGTTTAGATCATTCCTATTTATCGTTATACTTTTATATTGTTGTTTTGTTGTATTCTGTATTCTTTGTAT	GRKVIVKRPCTSHMLNLFSCSGIGQLKKLVLFLRIIMSYFARLYTCKSEQSNHMSPIKRNVTTRSGRLIKQPTILDL			

				TATTAAGCT	PSQRVNIELCCEITVSADKLAPNMLDNFGIIVVLSVIASLP				885
				ATGTTTGTA	HFIAESRPALLAGVTMEFQLVQNSIDTERYSDQYRYTKD				
				TTCTTTTATT	ELLSLRNTSHSRLPDNVFARIKELGIDKRRGCRAGSHKW				
				ATGTATCAT	RKIAVVEGRRRRTPKPPGVRAPQLIPIKTITTVSTDRQSQ				
				TATCATGAC	GNLPTIYMCNPRSLNNKLDEFSVVVHELCCDVCVVTET				
				ATGTACATG	WFRADQPESDFAVDDYLVFAKSRDSRRGGGVAIYAKPE				
				TGTTGTACT	SMPKTVNFLPPHPHHEVIWLQLRPKRLPRDISCLYVAAV				
				CTCATACCC	YSPPDNDCVNDLIQYLVNAIDDIKTRHGDAGVVILGDFN				
				CAGAAGGG	RADTSELTSTCNLRQVVRKATRGTAILDLILTDIHQWYS				
				GGTCGAAT	EPEILSPLGLGGDHNAVLWKPKLNKTNSKNKAVRKFVR				
				GAGTCAAT	PMKDSAIREFGQWISQHTWDDVMASSSAQEKCDKLYA				
				AAATTCTAT	SLQNAMDEYLPMKSVKMHPRDKPWVTPYLKSLIGQRQ				
				TCTATTCTA	KAFSNDRTEEWKRLRNKVQNKKTELKKNFYKDRVLKL				
				TTC	KRENPSSWYKHLKIMTSSKTASNIIIPESPELDDLAIADC				
				TGAACTTTG	VNDHFANVSQDIPELDLQSLPAYLPTPKPQLTVQPHEVY				
				TATAAACTC	KELNRIKPGKSSGPPDGIPARIVKEFAVELASPLCEVLNSS				
				TAATAGAA	YHSGSVPSQWKGAIIAPIPKEKAASSIDKQRPISLTDHFA				
				GGGGCAAA	KIAEVFIYRKLLADVRHQLDPKQFGARAGRSTTHCLVEI				
				CAGACTTTA	YNFLCKRAELPSNVTSIVSTDFSRAFDRVDHTILIRKLITL				
				GTATGACTG	EVHPWVIVWICSFLTDRQNQVRYHGELSSTRSMTAGVP				
				ATTTTAAAT	QGTRLGPLLFLVLIDDALRESEVERWKYVDDMSLASAE				
				GATATTTGC	PSNQHGQFMQNVLVNLEEWCNQNNVRLNPKKCNIMRV				
				ATGTTTTTA	TFSKTPLPDRDFSLCNDVLQEVSAVKLLGVILQRDLKW				
				ATTACTCAC	NDHVSVTISRASSRLYMLRRWKEFHLPAADLVTVYVGF				
				TTTGAAATT	VRPLLEYACPVWHSSLTCAQSAALERIQKRALRLIFPGT				
				AAAATAAT	SYNTSLDLARIVDLETRRKTLCINFATKASRSEVGSDWF				
				GTTTTTTAA					
				AACATATA					
				GTTTGGGAT					
				TATTACTAT					
				TGCCTATTT					
				CCAATGCAT					
				GAAACCTTC					
				GGGCGTAG					
				ACGCGGAG					
				GAGTGGTCT					
				ATTGTTAGT					
				GAGGAGGA					
				TGTCTGTTG					
				TTGTTTTCT					
CR1	CR1-5_SP	.	Strongylocentrotus purpuratus				cd01650	633	

					TTGTTCATCCCCTCACTTCAATAGATAACCTACCGTGCCTGCGCCTGCCAACATGGTTATGTTTTAGGATCTTGTTTATTACATCCCAGCCTTTTTAAATAGATTTAATTCACCTTAGCTAAACTGGATCATGTTTTTTTTTACTCTTTACTAATTAACTGGATAAATCACTCGAGCATTATAATACATGTACCATATTTAAAGCATGACCTATAGTATTTATTCATTCACAAAAAATGTGATTTCATAGAAAATTGAACGATTAATGCTAAACATAGATTTATATTTTTGTAATTTTCCACACGCAACACTACTATTTTCAGAGTTGAAATGTTCTAAACCAGTTAATTGC	TENNPTRNLRKNKRYREPRCRTNRYQNSPIPFMTKLLNFN			

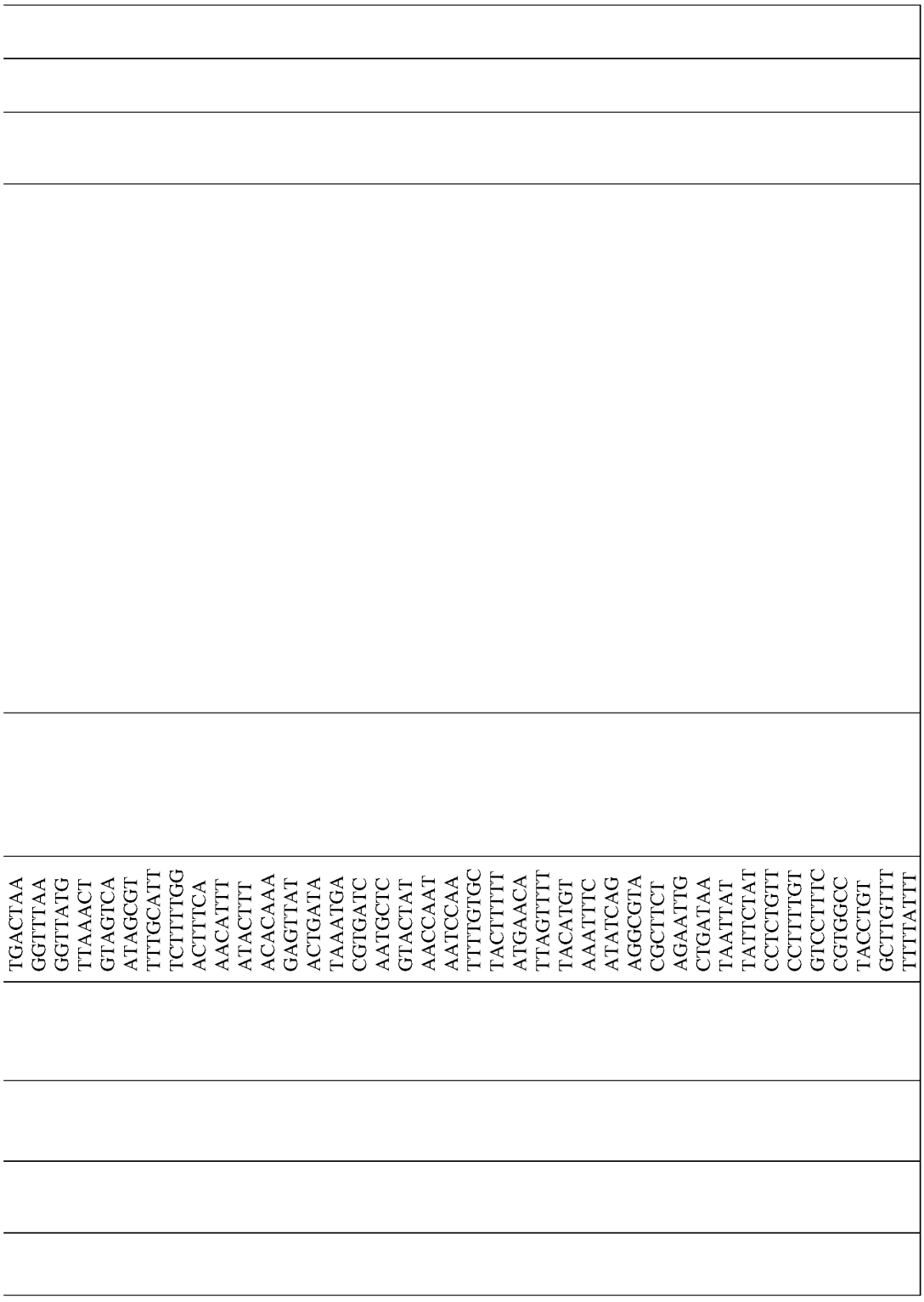

CATGTACA
GTATTACGC
ATGTATTTC
TTGTTAAGT
ATGTAATGT
TTTAATCAT
GTAAAAGT
TATCATTTT
CTTATACTC
CCTCCCCCT
TTCTTTCTC
TCCTCCGCC
ATACCTGTC
CTTTTGTAT
TCATTATTT
CACTGTGA
ATTGTGCTG
TATATTGTA
TGCATGGGT
GTTGATCCG
TTGCGTCGT
GCTATTCAT
CAGGGGGG
TATATGATC
TTACTGTTT
CGTGCCCCA
CTGGTGACT
GGGGTGGC
GCAGCGGT
TCGACGCC
GGTGCCAA
GAGTTGCAT
TATTATCTA
TAGTAGTTT
ATTATTAGT
AAGTGCAA
TTTTACCTT
GTAATAATT
AATTCTGTA
ATCCATTGC

					GCAATTTAG					
					TTTTTAGCT					
					ACGAAGCA					
					TGGAATTA					
					ATGAACATT					
					TATTGTTTA					
					CATGTTTGA					
					TTTGATTTG					
					AATAAACC					
					ATTTATCTA					
					TCTATCTAT					
					CTATCTA					
						LTKTEKDLLSRGLNSAVTPEELPVVDLITATETAIRNNKL				
						PETEAEQIRLKVSAALANAKAPASNITAQEKRALASLAK				
						DKDITILPADKGRCTVVLNTTDYDSKILSLLGDSNTYEK				
						LKRDPTSTYKKKVIDLLQKLEKEQAIDKPQYYRLYPGE				
					TGATGAAG	ATPCFYGLPKIHKEGTPLRPIVSSINSVTYNISKFLASILAP				
					CCTTTTGGA	MVGNTPHHIKNSQDFAEKVSNLQLESGETMVSFDVTSL				
					TAGAAGGC	FTCIPTSEATEAIHKRLLLDKNLQERTTLSPAQICTMLDL				
					GAAACGTTT	CLNTTYFQFRESFYRQKHGCAMGSPVSPIVANLYMEEV				
					TCCAAAAA	EKRALLSFTGAAPSRWFRYVDDTWVKIQTKEVEKFTTH				
					GAAAATTG	LNQTDTFVKFTREDVKGDSLAFLDCEIRIEEDRNLSIEIY				
					AAAAGTCC	RKPTHTDQYLLFDSHHPLEHKLGVIKTLQHRATKVPTTS				
					AGATGAAC	HGRKKEQDHLTTALKTCGYPDWAFTKTRKKRDPSNEE				
					ACAGAATC	EKDKRHSVSIPYMAGVSEKFRRILQKHNIPVQFKPSNTL				
					TTTTCTCGG	RQKLVHPKDKTPRHKQSNVVYAVQCQQECRDLYIGET				
					ATAACTATG	KQPLHKRMAQHRRATTSGQDSAVHLHLKESGHSFEDN				
					ACCTGGAT	QVRVLAREDRWFERGVKEAIHVKMEKPSLNRGGGLRH				
					GACTGAGA	HLSPTYNAVLHSFHKQKPKHTHSFTRAGVTSPGDPAVK				
					ACCTACAC	RKTSQTETRSTTLPTTLPTTPETSIQTTGILMQRLSVDISM	cd00304	296	379	
					AGACA	PMVQKGYIIFKLCTEPDQN				
Non-LTR Retrotransposon	BRIDGE1_TN	.	Tetraodon nigroviridis	CTGGGGA	TAGGGGGC	MGSWIVNFVSVATQTGEFPVDTARRAPVPVTSYPESEC				
				CCGTGGT	TTGGCATTT	HXPLPLTFCNSDVTIWGGVRPEPVDCLGDLPEXYDALP				
				TACAACC	CTCATTGCC	GVAGPREXVGGSPPGEGVRSPGIASPSGTAVQHDFGSPI				
				CGGGCTT	TGCTCCTGA	LVPGAEAAEVSTPVVKVPQDHPACPCCGTRVVKVTALS				
				AGCTGCA	AAGGATAT	EHLRRAHGRKRVLFQCSRCGRMNEKHHSIACHFPKCRG				
				GAGACAG	GGGTCCTGC	PPVEEGPLGAPEWCCEECGQKFNTKSGLSQHKRSVHPL				
R2	R2-1_TGut	.	Tinamus guttatus	TACCTCC	GTCGCGTG	TRNVERIEAARPKGKGKRGAHKGCWTEAEVAQLIELEG				
				CCGTGGT	GTAGGCAG	RFKNQRFINKLIAEHLPSKSAKQISDKRRQLAAATKTSSP	cd01650	542	806	
				TCCCGCC	ACCCATTCG	EKRVTSSTSGESSPEVEKVEGIKREYRRRVGEWLCAGSL				

GGACCCCC	TCCGAGTA	XDQTSFQKILEDVESGSEIVTGPLEELASFARGKLAAAR		
GTAACAT	GGGGGCTT	VRHHRKHPAEAVPAREEQRWMKRRVGRRGLYLRFQRL		
CGGGTGA	GGCAGTNT	FALDRRKLAGIILDDVESIKCPLPMEEVADVFRRRWEEV		
CTGAATC	CCATTGCCT	APFTGSGSFRSLGKADNGAFKPMISAKEVMKNVXEMSR		
TGTCTCT	GTGCCCGA	RSAXGPDGLSLRDLMKIDPQGSRMAELFNLWLLAGRVP		
GCCCCGG	AAGGACGT	DQVKAGRTVLIPKSADPGKIGNIDNWRPITIGSVXLRMF		
GAGTAGT	GGGTCATCT	SRILSARLRRACPINRRQRGFIAAPGCSENLKLLQALIKS		
TCCTCCTT	GGTCTGTCT	AKRDHRTLGVVFVDLAKAFDSVNHQHIFQVLVQKGVD		
GCCCTAT	GCCTACACC	GHIIDILRDLYTNAGTYLESGSQRSGFIKILRGVKQGDPL		
TGACCAG	TCTCTAGAC	SPILFNLALDPLLCRLEDRGLGYKYGDQQIXSLAFADDL		
CGGTCGC	TTGTAACAT	ALLSDSWEGMQQSIRVVEEFCQRTGLRVQAPKCHGFLI		
CGGCTGC	CTAGTCTGT	RPTKESYTINDCDPWTIADMQLDMIDPGSSEKYLGLGID		
TCAATAG	CAACAAGA	PWIGLSRPELSEVLTRWVKNIGGAPLKPLQKVDILRSYA		
TATTCTA	TCAAAATTC	LPRLLFIADHAGLSATCLHSLDLSIRSAVKGWLHLPPSTC		
GGCGTGA	TTCACACAG	DAIIYVSYKDGGLGLPRLASLIPNVQARRLVRIAQSEDD		
AATATAG	ACGACCGA	VIRSVVLQEGIQEEIRKVWISAGGRPEKVPSVTGEFPVM		
CGATAGT	GCTTGCTCA	EAQAADEALSEWERRAPRTIYPIPCKWRKREMENWTNL		
CCTAGTG	GTCTTCCTG	KSQGHGIRNFENDRISNDWLLHYGRIPHRKLITAIQLRA		
GTTGTCTT	TACCCGCA	NVYPTREFLARGLGEGAPRGCRHCPAEWESCSHIIGYCP		
ACTGGGC	GAATTTTGC	AVQEARIKRHNDICGVLAEEARKLGWVIFIEPHLRDNTN		
CATAGCC	TCTTGCTCT	ELFKPDLVLVKGSCAKVVDVTIRYESGLTTLSDAAAEK		
CCTTGCTT	CCTTTGGCT	ARKYQHLAGEVRALTSATTVDFLGFPIGARGKWYVGN		
CAGGGGT	GTGTCCTGG	NGLLSDLGFSTSRVVRIARALSKKALLSSVDIIHIFASRA		
CATTCGC	ACGTGGGA	RQAQTSE		
GAAGTCT	CTATTCCAT			
CTCAGGA	CTCGTCCCA			
GAACTGG	AATGCCGC			
GGGTGGT	GTCCAATTA			
GTTCTTCT	TACCGGATT			
GGGTATA	TGACAAAG			
GCTAAAC	CGGACGGC			
CCCCTAG	CCGCTTTAT			
ACTGTGT	AAGCCGGA			
CCGATCC	AAAGGTGC			
	CTTGTAAAA			
	TTGCAAGGT			
	TCATTAAAT			
	AG			

CR1	CR1-3_TCa	.	Triboliumcastaneum	TATTTGC	AA	GLQSLQFRRICIIIKFLYNLIN	cd01650	915	1166	
				AAATAGT		YKIHIDQLENIQRRFLKFLVFIIDGNYPIRGYDQNLLLNRF				
				TAAAATT		YGFIYRNCRDFKNLSVMKTLFFSLVRSKLEYGALIWHPI				
				TCCTACA		YDISSVTLHRTFIIKDLGVIFDTELAFSEHIRDVTARAIKS				
				TTTTTACT		SLIKSYLANREQRVKYRNYISDSYVATSGVPQGSNLGPL				
				AAAATTA		LFLLFINDICGSLSTCAKLLFADDLKIYTEIKSIEDCLTLQ				
				AAGCTGA		NNINAVVKWCNENRLYLNPSKCNVMSYTKKREFLEFV				
				TTTTTATC		RKGQVDVIYMDFKKAFDQIDHHLLLLKLDQYGFSGSLL				
				CTATCAA		DRIYSSVKRFISPSQHGFVEKRSTVTNLACFTQFISDALD				
				GTTACTC		PNIWKIAHVCPVFKTGDITQIQNYRPISILCNFSKVFESIIY				
				GAAAACA		SKDTLTAGCDGIPSFLLRDCAHIFSKPLVYIFNLILQTSTF				
				ATATCTC		KSVFSDMTLTSDSPSMMTNLNSFITVNEINEHEIIRVLKR				
				TCTTGCTT		FWSFVQAKKRNSRIPGVLYHNGNAISNPNDIVNSFAKFF				
				TTTTGGTT		MHFYEEFKKLRSLSKSLITTAYSQYIGHMENTISKDPKK				
				AACAGTT		PKLFCRKNYRTFPTWFNTDIINCLRRKSNILKKYRRTKN				
				TGCTTAT		FHNFNQFNNIRNNRNNILDLIFSNMDCAVSKSSFSFVPED				
				CTAAATT		LNNFDEINNDNVLFLGDFNTPLFSGFSKNNFSISLKHFAS				
				TTTAGTA		RRDRDLRHLGIHRGGGVLIAVNNKYNCRQLDISNISRRS				
				AGTCGTT		TFFNSFLSTSDEFDLIVITESWLDDTVLDGELFSDNYYVF				
				ACCAAAG		NRFVCLQPSHSKISISINKKTVNSFHGIYQNVRGLKSKSV				
				AAAAGTC		CQDHTNRSAVEQNPVATFLDTSLPSVDKSDSQSFLSVFL				
				AAGACAT		NIYYLDSNDKIFDQFKTDNFVINKAVKRCIIKNLGDLITE				
				AAAGGAG		FLNTNIILIGCPYFVNRPILNNFVYELNCNFYNVAKNNK				
				AGAGGAA		NNEQLGNNDFVVIYAGNHEAQKGKIPDENTLNELINHK				
				QKGRHCIDLLETYIPVSKYVIQSVVKPYSSLQELLKTEN						
				DNIKTANSSDLTLKSTNSKNVSTQSVSKTNRYKVRIFAD						
				CTTTAGT		DLESKLISLLKDTLNREIEKLKNSIKNEVIAEISTNLNRTN				
				TCTAACA		KQETDLIKKTLDDLKTFITESPTTVCTFDKIDEYVRKSFD				
				AATTGTG		GQDLTSISACPCCDKLKSLQNPTSPTLNISTNTIFEDLGIL				
				CTAAATT		MSVCYKCKSSVVNSIACLNCKKAYHPSCLRAFKNYSAI				
				AATTAAA						
				AACCCGA						
				GTTTGAA						
				ATAAAAT						
				CTTTAAA						
				AAATTTG						
				TTAACAA						
				ATAATCT						

ATTTTTCA
TGTAGGA
CTAACCA
TAAGCGA
GATATAA
GTTTGAA
AATAATT
AATATTT
AAAAAAA
GTGCTTT
AACAGAA
AATGCCT
TGTGATT
TAAAATA
CACTTAA
TTTATTG
GGTCCAA
TACTTTAT
TATTTACT
TTATTTTT
TTGGTGC
TATGCTA
GGCTTGT
GCTAATG
TGCTATA
AATAAGG
TGTTACA
ATTTTGT
GCTGTAA
ATAAGGT
GTTGTGC
ATTTTGT
GCTATAA
ACGTCAG
TTTTTTAA
TTTTTCGT
TAATTAT
TTGTGGT
AATAACT
TAATAAG


TGACTAA
GGTTTAA
GGTTATG
TTAAACT
GTAGTCA
ATAGCGT
TTTGCATT
TCTTTTGG
ACTTTCA
AACATTT
ATACTTT
ACACAAA
GAGTTAT
ACTGATA
TAAATGA
CGTGATC
AATGCTC
GTACTAT
AACCAAT
AATCCAA
TTTTGTGC
TACTTTTT
ATGAACA
TTAGTTTT
TACATGT
AAATTTC
ATATCAG
AGGCGTA
CGCTTCT
AGAATTG
CTGATAA
TAATTAT
TATTCTAT
CCTCTGTT
CCTTTTGT
GTCCTTTC
CGTGGCC
TACCTGT
GCTTGTTT
TTTTATTT

CR1	CR1-1_TCa	.	Triboliu m castaneu m	TGAGTCA CTACGAA CGGCCGC GTAGTGT GAAGCGT GTGTCCT TGTCTTTA TTCCCTTG CCTGATT GTATTTA TCAACTC ACTCATA	ATTCTAA GCATTGT AAAAGTA TTTTTTAT CATATAA TATGTAT GAAATGG TTCAAGA AAACACA AGAATAA CCGCACG CAGTAAA ACAAGAA TAGACAA TGTTTTTA CAAATAT GACCTTT AATAACT TTAAAAT AGAAGTT CTTGAAC CACACAT ATCGGAT CACATGG GTATACA GCTAACT ACGAGCA TAAAAAA	CTTATGTCT TAAAATTAT TTTTATGTT ATTGTTATT TCTGTGTAT AATACTGA CGTTTTTCG TATCAATTT TGTTACTAC ATCTGTATA AATTGATA GAAATAAA TTATTATTA TTATTA	MALRAARKNYYGQRIQNADNKSKCAWNIVNKICGRSK KCCEIQIEGDPKQIVNDINQYFINSAPFVRSKSNNNTISSN AKTMYXYPVTSQEIITIIQNLKNKKSAGFDEISNNIIKYC NEEISAPLSFVIQNSLDYGIFPEQLKRAQVLPIYKKGDPT LISNYRPISLLSSFSKIFETVISSRITNFLMKFQLLNSSQHG YIKGKSVDSAIFSFTNALLDKLEEKKVVLGIFLDFSKAFD CLDHNILISKLKKYGIRGIMLKLITSYLNNRYQQVTITKN STVYTSNKLLIKQGVPQGSILGPLFFVIYINDLLNILTVSD NNSAVNFADDTNILIWGENMTDTMKEAKICMNKVINW TKENKLVLNFEKTYGIHFAICNHKNTEKPNKIVISSKEIE LIEKTNFLGICIDHKLSWTNHIDNLLKKLNSVHFTIRVLK KYINIEQLRIIYFANFQSLLSYGIIFWGQSGKISKVFVAQK QVMRTMFNLKYNETCRNIFKSNNFLTVTGLYIYRLLLFF FKQNLLFNVTNDHKYDTRKVDLEYPRHKLTLTEKNSYY AGIKLYNSLPVHLRTPCNLNLFKKKIFDYILDLEPYTLNE FYNRQVF	cd01650	148	408

				TAATACTAATAATACTGATGCAGTAAAAAACGTTAGAATTTTAAATGTATATGCATGTAATCGCAATGATATCAATGCTCAGTGGAGCAATTTTATAAATATTTTCAATGATAAATTTAACGAATGTTTTCCTCTAAAACAAATAAAAGAAAAAAAAACAAGAAAATATTTATAAAAATGATCCTAAAGTAAAATATTATAAATATTATCTAGATTTGTTATCTGTGATAAGTAAGTGTGACGATAAATATGTTATTCTTTATAAGTATTTGAAAA					

NeSL	NeSL-1_TV	.	Trichomonas vaginalis	GGGTGAG TAGTCTA GTGGT	AAAATAC GAC	TAAGAAGA GATAAGAC GAGTGAGA AGAACAGA AGCATAGT AGGATTGG CAGAGCTT AAGCGATG TCACTCGGT ACGAAACG TGTACCAA ACACCGGA TTCCGTGCT AGGAATCA CAAGCCAA AATAAAAG AGACACCA CGAAAATT ACTCACCCT CCCTCAAAC AGATAATA ATATTAACC TCCCATCCA TCAGTCCGT ATGGTCTGA TAACAGAC TAGCACCA CATCCATGA TACACTCAT TGGAGTGA AAACCACC AACAACAA ATCCACCTA GACCAAAT CCTGCCCCA CCTCCACCC AAGTAGCT CGCTTCGCT	MIPVLGTGGPEKLPLQSYVYCGNTAITDSFTPTAKTILKP EEQNLDIVLKNIAALNPENYSDLIRSLSKMEFRLDYPKEI ENYWISEKLFSQSIASLPISLLVASMFSPEDRDLSTEPFHC NADGCNFHCDNCERMVEHIREHHNTDPMINTFETTEDT FRRITAIKIDKTGIEELNPLKYRCSYCDELFTEAEDHAIH MISHLTEKLSPDISFFFNDILRLYKTIDKPTVQNLFPETQV AIFDTLEETNRFRLIVGREAIETIEEAFPPSPPGTDRKPSIII TDTCQLRFVPCMDEPPKGDLGILTLLLRDFSAHNIPIKSL NNKELIADKDIDYSPDFVEGALANAEEHDTTNSQNNNG RYINSAEKLTEFLIQCEDYLTNIKTLEDLERFYTTIKDYR VNKEVIAEDTPIFVYFLVEEGKLPKPGLRCPLESYEGHE DKAFESLRKLCDHFKGEIAKTSFDPKVHTIDIWVEFLAQ AYGTGTFVYKDENGNIDLDTHVFKCPYADCSYTNNDRS KLMDHMKTKKHAKNVYIERYGFFWGIVIEGVNRPKGIV YPTLKDIKEHACRKCPEAGCNTYVTELSDIKEHLKKKH KSTTAGVDGEIAHTDATYCWTIKEELDALHAERARERA EQVDNTPVQQIINADNNEENNENQEDNGNNEEADALDP PNNTTETEDEAVHAVIINPPATEEEEVAIIAEARRNIPELQ QAEERGCVTPKMTSLVRLKLLLKGGGELFNKKLTPLATR YAATGNTEADKIKVDYLTLKCNAALREMIYTNNHSESK FMTAENGEDTAPPPRISEDTRDRIQKAANEIKGTLIKVV KHISHARCLKDSTRDDEHNKFVEMIAKIKNDLRDNKFE QYNIEEIFQGPISDQSILNIVNTTEDNNEFIKKMDYINRILG TPQDASPYARKKLQACFADNPTKTLRNIILADKVPQQSL KPSEYLDYYGPQWANEAEGYENFLHHDYALPERYGQV FANDFLDFMTNESKIIEVIRNKNHLSAHGLDGIPNSVYM LFPVSAAKFLSILFRSIISGHIPDCWKLSKTVMLFKKDDP SLAKNWRPIGITSCTYRIFMTLVNKALQMIPMFHAMQK GFVRGATLSEHIAVANEVLCQSTRTQSEMFQTAIDFTNA FGTVPHQLIFDSLEAKKVPDSIINLLKDLYKGARTAIYTR HAHSEIVPVRRGVIQGGCPLSPILFNCCLDPLLYAVQRRHF EDGYRFQDKAGQYSIAIQAYADDVLVISPTHEGMQRILN TVDEFQKIAKLKVAPQKCVTLAKTSTAIQPFRIGPDEIPI KTSMDNITYLGIPISGTKTSRFAAATGILEKVKAQIRVVVF ASHLALSQKIIALRVFILPQLDFYMFHNVFRVNDLKATD QMIRGLIDKEAPTSNIPVSFFYMPKNKGGFGLVKLELRQ PQLVLTKFARLWLSQQAETKAFFHTMAQEEKSFRKVVE DQENGFLGIKMENGKIVQKNERSKRTNCFITQAAKAAD	cd01650	1049	1310

Ingi	Ingi	.	Trypanosoma brucei	CCCTGGCG	CGCTCACCT	KLEVRFKEWDKGGIQVRGVGENATDWYRSKHIGQISPL			
					AAAACTTTG	IGRVIQQRQYEEFKKDETHSHTFCEPAALAESHDIMKRP			
					CTCGCTCGC	QAVPNNLYSAAIALRTNTAPTPANMHFHNPEVLANCPL			
					TTCGCTCGC	CGCQSCTLFHTLNMCRNRFSLYKWRHNIICDDIYQFIHD			
					TCGTCTTAA	HYPGVTIKCSARITSDGYQTTGPELDDTVKDLLPDLVVY			
					CCCTTTCCG	DEANKMIKIIEVTCPYGTDNNVGNSLDAAYDKKVNKY			
					AATAAACA	KSLAEQTERLFNWTTTLSIIVVSSLGVIPLRTKLDALRISP			
					CTTACAATT	ADHIQLLKRLSMHAIAASACIVFEKVPEFFGMRCRPLPG			
					CCCGGCTCG	RVTAPNAAIPPNNNENNNDTDHGQENQQATSEEQPTNN			
					CCCCATTTT TT	GNAQEDNGQGEQINNSTEQTISVDQIIEEDAENNAIEQA LDQPDEDEFLN			
					TAATAAAC	MPATSTWCQGPVPRIIGGSQEPAAFLSWGTLLCSGYGII			
					GAATAACA	QHRDQQRLAGTPFFICRSLGTCQRAISSIIRTKMLLSGDV			
					AAAATCAA	EENPGPSLRGMQWNCAGLSQGKRLALHKTLVDERIAFC			
					AACAAAAG	LLSETRMTPGEAACFSVAGYQHHGIARNCKGGGVSILV			
					GATTGCTAA	REDLPVETGMAVVGRIEQVHATIHLARGTALTVTSAYIP			
					TTGACATCT	PKHTFTATDLDTLLTTDGAQLIGADANAHALAWDRASP			
					TTTGGAGA	PNTKGETLTQWCIDNQFLVCNTGECTRYARHHGESTPD			
					GTCCGGGG	VTLSRNCTVYTWTSLYSPDSDHHHIFFDVIVGDDTDALS			
					TGGGGGGC	CPRLRKPMYAWLKADWRNFRLKVDELCRKIGREKNVN			
					TTCTCGCCC	TLEQKLSSAIRIATKVSVPRGCRATPPHWTPELAKLDEEI			
					CATCTGCTG	AGCGPSHRREKLVATRKQILDRTTKKRWSTLCSRLAVS			
					TATTCCGTT	DRCSWHIVKKVYAPRPLTTPAVLVDNAAITDYRQAERF			
					CAACTGCG	SKLYSSRARRHPDSHPPAPIKTIASEFSPITMAELRRSIKL			
					GAGCTACA	LPSGSAAGPPDCLYNEALQHLGRTALNVVLRLFNESLRT			
					ACAAAAAT	GVVPPAWKTGVIIPILKAGKKAEDLDSYRPVTLTSCLCK			
					TATAGAGG	VMERIIAARLRDTVESQLTPQQSGFRPGCSTLEQLLHVR			
					GTGTGTTAG	AALCRPTHQYRTGAVFVDYEKAFDTVDHDKIAREMHR			
					GCTGAATA	MKVSPHIVKWCVSFLSNRTGRVRFKEKLSRSRTFERGVP			
					AAAAAGGG	QGTVLGPIMFIVMNSLSQRLAEVPLLQHGFFADDLTLL			
					AGACTCTGC	ARHTERDVINHTLQCGLNVVLQWSKEYFMSVNVAKTK			
					CACAGTCG	CTLFGCIERHPLTLQLDGERIGADRTPKLLGVTFQCLQG			
					CCAGACCG	MATHAAETRRKMDFRLLQIAAISASTWGPRRQVLRAFY			
					ATAGCATCT	LALVQAHTMYGIEVWYWDASERSRDLLAAAQHKASRII			
					CAGGGCTCT	AGIPHGTRKEDSLLEANLLPLKTTTLVRSMKFMLMCES			
					ACGGTGAT	RGGCLRRSAEEVYHSKHPVRALHSRIMRSYPHLRIEPRE			
					GGCTGATG	HPLETSTLRHSCRPLFHTQIKPVCADDPDDVKREASEKW			
					GCCGCGCC	IARHFARRGKEPPRREHYELWTDGSVSLGEKSGAAALL			
					AGTGGGGG	YRNNTLICAPKTGAGELSCSYRAECVALEIGLQRLLKWL		cd01650	550	805
					GAAACTCTC	PAYRSTPSRLSIFSDSLSMLTALQTGPLAVTDPILRRLWR				

Rex 1	Rex1-1_XL	.	Xenopus laevis	CACCCTC	ACGAAGGC	TGACCTTAC	LLLQVQRRKIRIRLQFVFGHCGVKRNEVCDEMAKKAAD				
				CCCCAGG	ACGAAGAA	TATATACCA	LPQLRDTWIPDIIAYAKRVLRSEEVHENTHRFGITGNHFP				
				GAAGACA	AATTCCAA	CAACGTTAC	TKHKEELTREEETALARFRVGSSRHYGWMLRKINPSVP				
				CATCTGC	AAAAAAAA	ACACCCTTG	PQCRWCNPQHAAIGPTIQTAPTVATRTLQRTSEPTKCTE				
				AAGGGCT	A	TATAAATA	CDATYQCRSSAVTHMVNKHGFVRADALRRIKYGDATP				
				CCCAGGG		AATAGCCA	AVDIPPEPPPVVAIVPLPSSTRVPMRPQVLHCTLCASKFA				
				CACTGTT		TTGGATACA	VPGRLLHHLRTIHGIGSGSCRVKRGRENEDSXQGDGRA				
				AAAATCT		ATTGTTCTC	PAAPASQDTRKLPFQCDLCEASFGTRSSLSLHKKFKHKS				
				GCACCCA		TATGAGCA	IVTEDGTVVVVQFPRKRAREGNVDVPGKGEVQCGVCQ				
				TCCGTGC		CATCTGCAC	KVLSCRDSLIRHCKAFHKGEGVELKCSKSTKLCATDSPH				
				TTTACAA		TTTATCATG	TNTSMLVCPTCGRQCASKTGLTLHQKKMHGMKVERAV				
				GTCCCCT		TTCTTTATC	TSQRGDCEETSLHLMSCPGLKELRVRFAVEGVCVQDVC				
				TCTGCCA		ATGTTCTTA	FSKRLAQFLIAVERSRPQVKSSPKVTVIQPTLPSATSPLIA				
				ACATCTC		CTACTATCA	ECGASRKSTGRRNSPDTLRDRGGMQSTSNRNRKKRGRE				
				CCACACC		TATCACAAT		MDVGTTKVQVYNRQTLLRYSNHAYLTIKDGLLHTLRD			
				AGCAGCC		GATACACC		LGLLREPDLLTAVSPGANGGRRSTRKRCTRTRKRGKRA			
				TAAAAGG		CATCATGTC		GVRARLRTNPSRPALPSILLSNVCSIENKMDYIRLQQTTQ			
				GACATCC		TGACGTTTA		REFRDCCMFVFTETWLSERVPDAAIQLDGLVAVRADRV			
				GCTGCCT		GCATTATTT		AALSGKNRGGGLCVYINTKWCRNSVLVSSHCSPLVEFV			
				AAAGGAG		ACTTAACAT		VVKCRPFYLPREIASIHIIGVYIPPSANTKDALWELYSAIT			
				GAAAAAC		ATTACATCT		DIQNTHPDGLIIIAGDFNHANLRSVLPKFHQHVEFATRGP			
				TGGCTCT		GCTGAGTGT		NTLDQVYTNISSAYRAEPRPHFGYSDHISVMLIPAYRPL			
				CTAGCAC		GCACTGTCT		AENLTWSHNTSYLAKKAQQRLYFLRKLRKSHLPPSILKT			
				ACACAGA		TGTCCTGTC		FYRGTIESILSSCITVWAGNCTVMERKTLQRIVKTAEKII			
				GCCGGAG		CCTGCACTG		GVSLPSITDIYTTRCIRKATSIVADPTHPSHSLFTLLPSGK	cd01650	563	816
				AACAAAA		TGCTGTCCT		RYRSIRALTTRLCNSFFPQAIRLLNTQGPDRSLSHSHSI			

				CGGGGCC TACCTCC TCGACAG AAGCCGG GGACTCG AGTAGTA GAGGTAC CCGGGAC CGAGCTG GGGCCTG AACTGCA GGGCAGA CTGCTGG CCTGGTG ACGTGGG GCACCTC GGGGGAG ACTGTGG GCCTAGT GGCGGAC AAGATGG CGGCGCG ACAGTAG ATTGCAG CGGCCTC TCCGGAT CACAAGG TGCTTTCT TTTTTGTG TTTTTTGT GCTTTTTT ACTTTGTT TTTTGTG AAGCT	GTCTTGCAG GAGTTGCAT ATTTTTGCA CTTGCACTT TTTGTCTGT TGTCCGGTA CATCTATGT TGTGTGTAG CACCATGGT CCTAGAGG AACGTTGTT TCGTTTCAC TGTGTACTG TACAAACG TATATGGAT GAAATGAC AATAAACT CACTCTTGA CTCTTGACT TG				
L2	L2- 3_XL	.	Xenopus laevis	CCCCAGA CCTGTGT GCAGCAG CTCCCAG CTTTTACT CTCCCTA	TGAACTGG CACTTATAT AACAGTGT AACAAACT GTAACCCA CAGCACCTT	MAKLIIFAVFACMVLPKVVTSFNRPIVCQYSIYISSSLLPS TYSSSYTLYNYLQHLCPPVTNKTRICRHFKSSSHIFSLFY LLLLAAGDISPNPGPCSIPTYIQPRFPANLRKSSRSSTFSN LIQIPTLPSSCLPFSCALLNTRSLTNKSTLIHDLFALIHLVF LHQKPGSHRMIPPLRLHSLMAVSLSLTPPEPLDVVGELE FSSRCRFKPIPIPPAFSFNSFEVHAIHIFHPLSVRVAVIYRP	cd016 50	664	922

TTTATTA	AATACCCCCT	PSASSSAAFLSDFDTWLSFFLCTDSPAIILGDFNFHIDDAS		
AGACCTC	CTGAATTGT	QPWPSRFLRLTSSFELQQWTNTPTHKDGHFLDLVFTKK		
TTTAGGG	GTCTGTATG	LSLSGFNSEPFPLSDHNLVTFSISHSPCPPASSPTVLVRDT		
GAGAGCA	TTACCCTCC	RDIDLPALAQSFRSSLISLKEMSDPDTLVSEYNHILSSTIG		
TTTGCTG	CATTTAGAC	SFAPLQAKRTRALNPRPWLNSQTRFLRSCTRSAERMWK		
GGGTGAG	TGTAAGCCC	KSRTQADFIHYKFLLSCLNSALSKANKSTTVPLIHKPDSY		
CATTTGC	TACGGGGT	APVRDLLNVCGRNPVRKQTLFTINFYCPASILPCQKPTR		
AAGGGTG	AGGGTCCTC	VLQYPYKQSQIQPTTPVFSIQYFLHPSSTVSVNQTHSPQD		
CTATTCG	TAGCCTTTT	FADFFKDKVESIRNQIPTSTNVNQLLLPKPPSACLNSFPP		
AAGTGAT	GTTTCCTTG	VTMSEVSKLLLSSPLTTCPLDPMPSSLLKHCVAELTPLLT		
TCATTTA	ACACTGAG	HIFNSSLASGTFPSPFKQACVKPILKKATLDPACLSNYRP		
CAAGTGA	CACTTAATC	VSLLPLASKLLERIVFSRITNFLHTHDLLDPLQSGFRPGH		
TCTGGAA	TGTATTGTA	STETALCRVTNDLQVAKAKGHFSILLLLLDLSSAFDTVDH		
GAGTTTA	ATTATATTT	SLLMQILYSLGIRNQAASWLSSYLSNRSFTVSYDNKTSS		
ACAAGGA	TATATTTAT	PVPLNVGVPQGSVLGPLLFSLYTLSLGDLIHSFGFNYHL		
GTTATGA	GTGAATTGT	YADDTQIYLSTPSLTAETEAQISNCLLAISNWMNQRHLK		
ATTGACA	ATTTCTAAT	LNLTKTELIIFPPKPGPTPLFTISIDGMLINPVNSARCLGVI		
TCTCTCTT	AATGCACTT	FDSSLSFSDHINTTVKTCHFFLRNIAKIRPFLSPDTSKTLM		
CCTCTTCT	ATTGTTACT	HALILSRLDYCNLLLTGLPNSHLSPLQSVLNTAARIILLSS		
GTTTATT	TTTTATTCC	KRVQAPPLLKSLSWLPIKQRITYKLLLITFKALHSSAPNY		
ACAGGTA	AATGACCC	ISSLVSLYVPARNLRSSQSNRLVVPPTTTAVSRIKPFSLAT		
AATCTGC	CTTGTTTGT	PYIWNAIPESLRRESSFNVFKTKLKDYLWEHLDNT		
TAAAACT	TACTACTAA			
GCTATTC	TTTATTGTT			
CCATCTC	TTGCTGTAC			
TGCTGTG	AGTGCTTTG			
GCCCTGT	CCCTCAAG			
TGTCTGC	GAGCGCTA			
ATTTTTGC	TACAAATA			
ACCTTTA	AAAATATA			
TTGGTAA	CATACATAC			
ATCTGCT	ATACATAC			
AAAACTG				
CTATGCA				
TTCCCTTC				
CATCTCT				
GCTGTGG				
CCCTGTT				
GTCTGCA				
TTTTTGCA				

				CCATTTA TTGGTAA ATCTGCT AAAACTG CTATTGC ATTCCTTC CATCTCT GCTGTGG CCCTGTT GTCTGCA TTTTTGCA CCATTTA TTGGTAA ATCTGCT AAAACTG CTATTTCC ATCTGCA TTACACC TACACTG TTTAGTG AGTTATA TTTATGC ATTGTTC ATTTGCC ACTTCTA AACTAGT CACATAA TATTGAT TATAAGT TCTCCCT GTAGTCT TGTAACT TGTACAT TTTACGA CCCCCTA TAAACTC CTCTCCA AACAAAC TTTGTAA GTCAGGT					

TATCTCT
GTATTCA
TCCTGAG
TCTTGCTC
TGCTCTC
ACTCTCT
GTCACAG
CTGGGTG
ATCAGTT
TGGTGTT
TGATCAA
TTAGAAT
CAGTTGT
GATTCCC
CAGACCT
GTGTGCA
GCAGCTC
CCAGCTT
CTACTCT
CCCTATTT
ATTAAGA
CCTCTTTA
GGGGAGA
GCATTTG
CTGGGGT
GAGCATT
TGCAAGG
GTGATTC
GTTTCGA
AGTGATT
CATTTAC
AAGTGAT
CTGGAAG
AGTTTAA
CAAGGAA
AGCTACA
AGGAGAG
ACAACAC
AGAGGAC
TCCATTA

				AGGACATAACACTGCTACCGGCTAGAAACTGCA					
Penelope	Penelope-5_XT	.	Xenopus tropicalis	TTGTTTTAGACTTAATTCTACTAG	TAACACCTTCATCAATGAGAATCTTGGTACCATTGTGATTGGATCATACCTTCTTACACCTGTCAGTTTCAGCTAACACCTAACTGAACAAGTTCAATGGAACCATTGTGATTGGATGTCTGTGACTCTACTACCATCAAGAGTTTAAATACCGGGGAATTCCCTACCAGTCATTTGAACTGAAGAAGCCACTCGGATGAGTGGTGAAACGTTTTCAAGAAAAACTCAGAAAAGTCCAGTTGTTTTAGACTTAATTCTACTAGATACTGTATATCATGACCTGGATGAATGAGAATCTTC	ILYIMTWMNENLHRHIATHFGENTLKLVREYERTARKLADYRNHLRFNLRCRYQGITPCSLRLGSSVKGHRAKVILQKAQKQLLNERVRQTNFSIEVLVHKFDNAKQRLTAKLPVLTLQRVVEFTERAQLAQHAKGKERQINKFTSLLSRTNNTQRTDKSTGRKEDENKTCQNIKDLWVKNLSDRELTEPEKDVLAKGLNFAVAPRYVPVVDFITATESSIHNNKIPVDEAENIRLKVSAALANAKAPPSNLSLQERRALTSLAKDSSVTILPADKGRCTVVLNTSDYHAKVSTLLNDSDTYEQLKRDPTSNYKKKVIDCLQHLEKEKVISPALYRRLYPGEATPCLYGLPKIHKEGAPLRPIVSSINCVTYNIAKYVANILAPLVGNTVHHIQNSMDFVKKVKELKLEIDDTMVSYDVTSLFTCIPTAEAIDTVRKRLKQDTTLSSRTKLTPEQVCSLLDLCLNTTYFKHKEVFYRQKHGCAMGSPVSPIVANLYMEEVEKKALDTFKGTTPSHWFRYVDDTWVKIKEREVPAFTKHINSVDKNITFTREDVKDSKLAFLDCAIVIKEGRDLDIEVYRKPTHTDQYLLFDSHHPLEHKLGVIRTLHHRAETVASNAEAKEKEYKHLKGALKTCGYPDWAFIKTKSKTNRKTRPTNRREEHSRRNNIVIPYVAGTSEKLRRIFNKHRIPVFFKPSNTLRQKLVHPKDPTPKHMKSNVVYAVQCSEECSDLYIGETKQPLCKRMAQHRRANSSGQDSAVYLHLKEKGHSFEDSNVHILDREDRWFERGVKEAIYASLEKPSLNRGGGLRHRLSPTYNGALTSLPRQFHNSSHFQSCTLKD	cd00304	482	563

				ATAGACATA					
				TAACACCTG	KGHRASNILQKAQKHLLNERVRQINFTIEALKQKIEQLK				496
				CTTTGAAGA	QSLTKQLPDVTLERVVQFIEHAQMAQHIKSKERQISKFT				
				GTTGCATGA	SLLSRSSGLKRTEELTWRQKDETSKNTKDTWVKNLSDR				
				TACTTGGGT	VLTQPEKDVLAKGLNYAVTPQHIPVVELITATESAINNN				
				AATCCTTTC	HIKVEEAEQLRLKVSAALSSARAPPSNLTTQERRALTSL				
				AAAGTAAC	SKDPNTILPADKGRCTVVLNTTDYHCKMTSLLSDSNTY				
				TCCACAGC	EPLRRDPSSSYKKKVIDCLQQLEKEEAIDRALYHRLYPR				
				ATTCACACC	EATPCLYGLPKIHKDGAPLRPIVSSINSVTYSIAKYLANIL				
				TCTGTGAGT	APLVGKTVHHIQNAKEFVTKIQGVKLEAEETMVSYDVT				
				TTCAGATGT	SLFTCIPTTEATETVRNRLQKDNTLSSRTKLSPNQVCLLL				
				CACCTTTCT	EPLRRDPSSSYKKKVIDCLQQLEKEEAIDRALYHRLYPR				
				GAAGAGTT	DLCLNTTYFKYKDQFYRQKHGCAMGSPVSPIVANLYM				
				ACATAGTG	EEVERKALLTFNGTTPSHWFRYVDDTWVKIRSNEVAAF				
				CCATTGTGA	SEHINSVDNNIKFTREDVHENKLAFLDDCLISIQEGGILKT				
				TTGGATTAG	LYIGETKQLLSKRMAQHRRANTTGQDSAVFLHLKDKG				
				TGGAAGAG	EVYRKPTHTDQYLLFDSHHPLEHKLGVIRTLHHRAESVT				
				TATATATGC	HSFENSNVQILDKEDRWFERGVKEAIHVKVEKPSLNRG	cd00304	413		
				TGAGGACTT	TDTESKDKECKHLRGALKACGYPDWAFVVKTRATKPNR				
				CCCATACCA	GGLRHHLSATYNAVLTSVPRQFQNSSHIHSCNSNK				
				GTCAGTTGA	NTKRNNRPEAERRRNIVIPYVAGVSEKLRRIFNKHHIPVF				
				ACTGAAGA	FKPSNTLRQKLVHPKDPTPKEKQSNVVYAVQCSEECTD				
				AGCTGCTCG	MSLLLLVLSFMFVLPKVTSSFSPPADCPYSITISSSLLHST				
				GATGAGTA	HCCSYNLYTFLKNLAPPTNTSQKHNNFKSFSHLLSLSLL				
				GTGAAACG	LLLAAGDISPNPGPCRIPISYRPRNASLLVKPSQTANLSN				
				TCTTCAATG	LIQIPTLSSSYLPFSCALLNTRSVTNKAALIHDLFCSKSFH				
				ATTACTAAA	LFAITETWLSQNDNAIEAALSYGGLSFSHTPRITGRGGG				
				CAAGTCCA	VGVLLSPHCRFKLIPIPPTFRFPSFEVHAVQLFHPLSVRV	cd01650	616		
				GTTGCTTTA	AVVYRPPSSKSLATFLSDFESWLSFFLCSDTPAIILGDFN				
				AATTTATTT					
				ATACTAGAT					
				AT					
Pene lope	Penel ope-6_XT	.	Xenopus tropicalis	A	TAAAACATT				874
				AGTCCTG	ATTTTGCCT				
				TGTTCAG	AGTCCTGCG				
				GAGTGGC	CTTAAGGG				
				CCAGCTT	CAAATGCC				
				GCTCTCC	CATACCTGG				
				CTATATC	TGCACTCTT				
L2	CR1-L2-1_XT	.	Xenopus tropicalis	TTAAGAA					

				CCTCATTT GGGAAAG CAGCGCA AGGGTAG CAGCGCA CAAGTGA ATCCTTC CAAAGTG AATCATT CACAAGT GAACTGG AAAAGGA GTTTAAC AAGGAGT TAGGAAT TGACACC CTCTCTCT TCTGCTT ATTCTAC AGGTAAT TGCTGCA CTATTTCC ATCTCCA CTGTGGC CCTGAAT ACTGCAT TTTTGCTA GTTTATT GGGTAAT TGCTGCA CTATTTCC ATCTTCA CTGTGGC CCTGAAT ACTGCAT TTTGCTA GTTTATT GGGTAAT TTAGCTA GTTCTATT	ACCTTCCAA TTTGTGCCT GTATGTTAC CCAACCACT TAGATTGTA AGCTCTACG GGGCAGGG ACCTCCTTC CTACTGTGT CTCATACCA CATGGCACT TATTCCCTG TGCATTTAT ATATATTTA TTGTATTTA TTTATTATA ACACTTGTC CTCCCTGTG TGTAATTTT GTATTTTGT AAGATTGT ACAGCGCT GCGTACCCT TGTGGCGCT TTATAAATA AAGTTATAC ATACATAC ATACATAC ATACATAC A	CHIDDPSQPWPSRFLHLTSSFDLRQWTNTATHKDGHFL DLVFSKNLDLSAFNSEPFPLSDHHLISFSISHVSPSPSTNA QPKTLIRNTQRVDITALSNSLSSSLSSFCASSDPELLVNTY NSILSSAINTHAPLQPQHSRAHNPRPWLNPQTRFLRSCA RSAERIWRKSRVKADFLHYKFLMVCLNTALSQAKQVY YNTLVNNNKSNPRRLFSIFNALLHPSPNESIISEHTPQDFA DFFKSKVDSIRKQFSQPSDTGNLNLPKSPLSVFSSFNLVT ESEVSQLLRSSPPTTCSLDPMPSSLLKQCAPALTPALTHIF NSSLASGTFPSLFKQACVKPILKKATLDPSCLSNYRPVSL LPLASKLLERLVFSRVTNFLCTHNLLDPMQSGFRPAHST ETALCRVANDLQTAKAKGRYSVLILLDLSSAFDTVDHS VLMQILYSLGIRDQAASWFSSYLSNRSFSVALANKSSTP VPLSVGVPQGSVLGPLLFSLYTLSLGDLISSFGLKYHLYA DDIQIYLDTPALTTDVQTQIGNCLLAISSWMNQRHLKLN LAKTELMVFPPKPGPSPPFTITIDGMTINPVNSARCLGVIF DQSLSFSNHINNTAKTCRFFLRNIAKIRPFLSQVTAKTLIH ALILSRLDYCNLLLTGLPDSHLSPLQSILNSAARILLLSPK REPVQPHLSSLAWLPVKQRIAYKILLLTFKALHSSAPHYI SSLVSLHVPGRLLRSSQSLRLFTPSTPTALSRLKPFYLAA PYLWNSIPESLRREHSPTLFKKKLSCYLLEH			

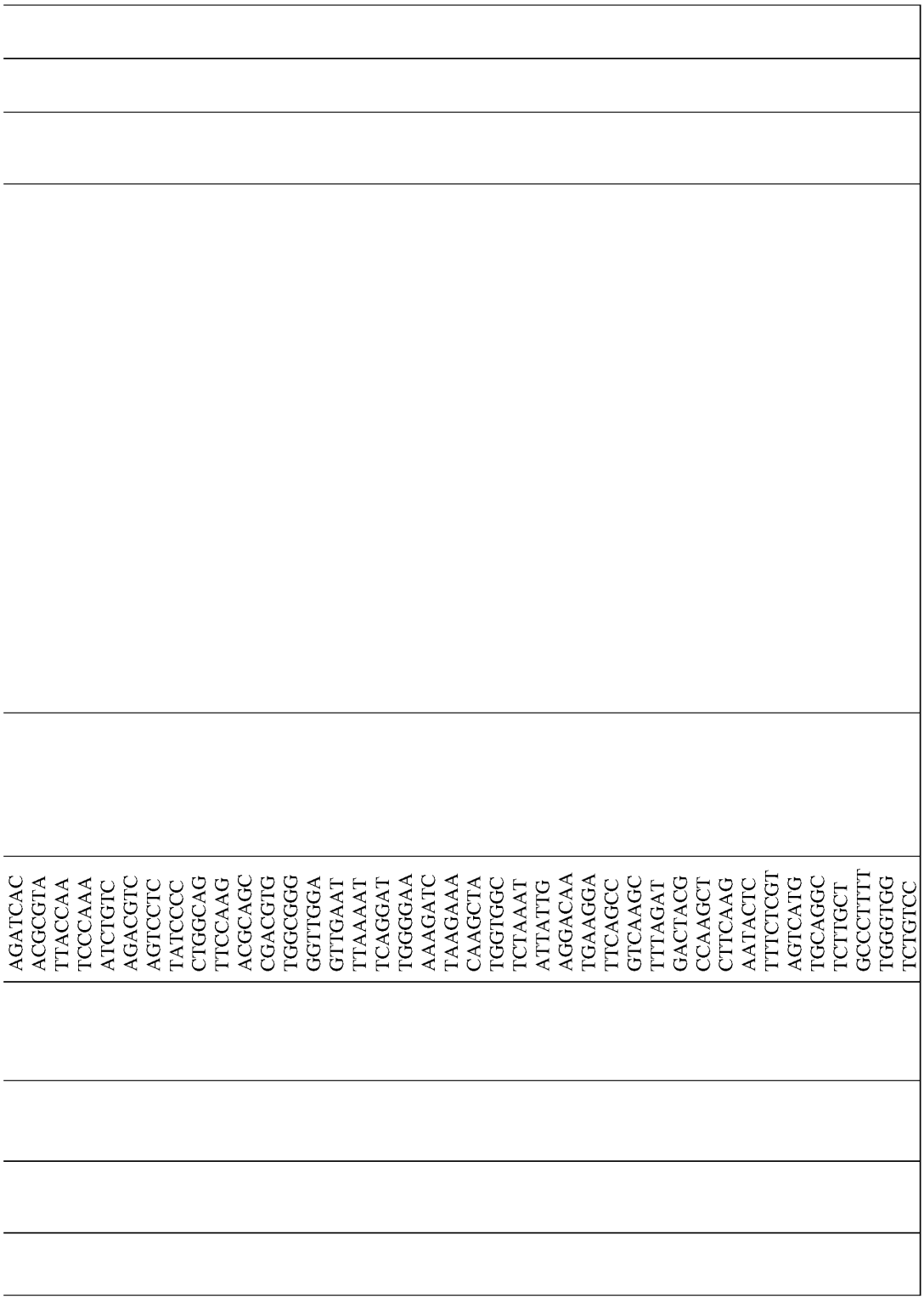
				TCCATTCTCAATTCCCAATATTGCATGCCCCCTATTCCTGTGCTCCACCAATTTGTTGATTTTATGACCCCCATCGTAAACTCCTCTAAACAAACTTTGTAGTTCAGTGTATCTTTGGGCTGATCTTCACTTGTGCACTTCTCTCATCCTGGGACTCAGCTGATTGATTTAATTTGTTATTTAACTAGCTGCAATCAGCTCTGATTCCCCAGTCCTGTGTTCAGGAGTGGCCCAGCTTGCTCTCCCTATATCTTAAGAACCTCATTTGGGAAAGCAGCGCGG					

				GTAGCAGCGCACAAGTGAATCCTTCCAAAGTGAATCATTCACAAGTGAACTGAAAAAGGAGTTTAACAAGGAGTTCAACAAGAGGAGCAACAGGAGGAGCAACAGCAGAAGACCGGCCAGCCTCTGTCCATTTGAATCGGGACCCAGTTTACTGGATATACCTGCATA					
L1	LINE 1-3_ZM	.	Zea mays	GGGAAACCATAACTGTGGAACAATTGCTTAATCCCTTTGCTAGTGTTCACCAGGTTCATCCTGACACCCTGGAGCTAAAGGATGTGGCTGTTTACCGATGTTCTGCTTGGTG	TAGAGTTTTTATAGATGAGTCAGTCTATGCTAGTTTCGTTCAGGCCGCCGTAAGGCGTCAGTTGTACCCCGGTCTATTTTAGACCTTGTTTATTTCTTCTTTTAATATAATGATACGCAGCTCTCCTGCGTGTTCGAGA	MDPEKIIVWNVRGLNSLARQNSVRELLNCYGADIICLQETKMAQPSSRVLLSMLGSDFTGSLALPSNGSRGGILIAWRRHLGITGATRTDIHSISVQFCSVEGHQWWLTCVYGPQSNEAKIEFLQELRHIRAACQGPWMITGDFNLILHAEDKNNSHINRAMMGRFKKLVDDLALKEIPLHGRKFTWSNQQNQPTLVKLDRVLCSIDWEELFPGAVLQSAASEDSDHCPLLLGFKDCKGVKRRFHFESFWTKIEGFHEVVHNAWNSVPVGHCPFSTFDLKLKGTAKALQGWSSKKVGHIETQLHLARDLIHQLEIAQDGRILSVLERWLLSKLKKLALALASTKRTIARLRSRISWLKEGDANSKLFHSFARYRKKKNSVSKLTVGNQVLTAHSDIAAAVDQFYFDLIGASKDRNISINLEALETAGTPFPNLIELDSPITEQEVCNAINSLPSDKSPGPDGFTGRFYKICWSIIKEDLMAAVIAIWNRKFDNFHKLNSAFITLIPKKDNADQVKDFRPISLVHSFAKLITKILASRLASRLNDLVSPNQNAFIKGRFIQDNFMLVQQTAKLLHQQRHPCLMLKLDITKAFDSVSWPFLIEVMQNLGFGQIWRDIISGLLLTS	cd01650	500	762

				TTTAGAC CCGTCCT TGGTGCC TGCTTCA AGGGAGC TTTGGAT CGTCGAA CCTCCCTT TGCCGTT GATGGGA TGTCTTCT GGCAGGA GAACGCT GATGTAT CCAATCC GCATCAC ATTCTCG TCTGCCC CACGGAG GGATGGA TCCGATC CTAAGCT TCATGCT ATTACTT CGGGCGG GGATGGT TCTGGTC ACGTGCA GCAGGGA TCCCCTC CTTTTTCA AGACCTC CCCATTC GAGTGGG AGAAGGG GAGAACA TGGTCAT TCGGGAC GTCTCTC AGTTCAT	AAAAAAAA A	STRVLVNGHPGSIISHRRGLRQGDPLSPMLFVLVMDVLG QLVSKADEEGLLQPLSTRQLQHRVSLYADDVVLFINPV AEDMTMISDILQLFGEASGLYNNNQKSSVYPIRCHEDNL EIIQQFWPCEISSFPCKYLGLPLSLHKLRREQVQPIVDKIA NQLPGWKADLLTKAGRSVLVQSVLTGVLIYSVMAVDL PSTTLKDIDKIRRGFLWRGRKNAMGGHCAVAWGSVCR PRELGGLGIFSITELSWALKMRWAWLQKTEPGRPWASL PLHTPKKIREFLQVALYSEIGNGASTLFWSDRWLDGQRI GDIAPRLIETIPMKLINKRTVQEAIEDNGWINDIPGSLSVG ALADFLRIWDLVAHVDLDPHKEDKHIFRLAANGKYSTK AAYEGLFIGSVEFEPFERIWKTWAPPKCRFFLWLVAHK KCWTADRLEKRGLDHPEKCPLCEQERETIDHLLVNCVF SRECWFLLLRQFGLQGLAPQPTDINFMEWWHQANEAIQ GSFRDGLNSLIMLGAWILWNHRNRCIFDGLSPNVANFLI HVGDERRLWETAGAKGLTSLVASLSHVA			

GAGCGTT
TGGGCCC
AAGAGAT
GAGAACA
TGCAGCA
GCGGAGC
ATTAGCT
CCGGGGT
GGCAGTT
TCAGATT
TGAGGAG
AACTTTG
GAATCTG
ATGGGGG
GACCTTG
TCTCCTA
GGCTGGT
AGAGATG
ACTGATG
ACATGGT
TCTTGTTT
CACCGGT
GAAGGCG
AAGCAGC
TGAAAGT
TTATTTTC
GTCAAAG
ATTCAAG
AATGTCC
TGAACCA
GGAGGAA
GCTGGAA
CTCCCAA
TTCTGTC
ACTGTGC
TTCCTCC
ATCCCCC
ACAGACG
TGGCAGC
TGCTGAC

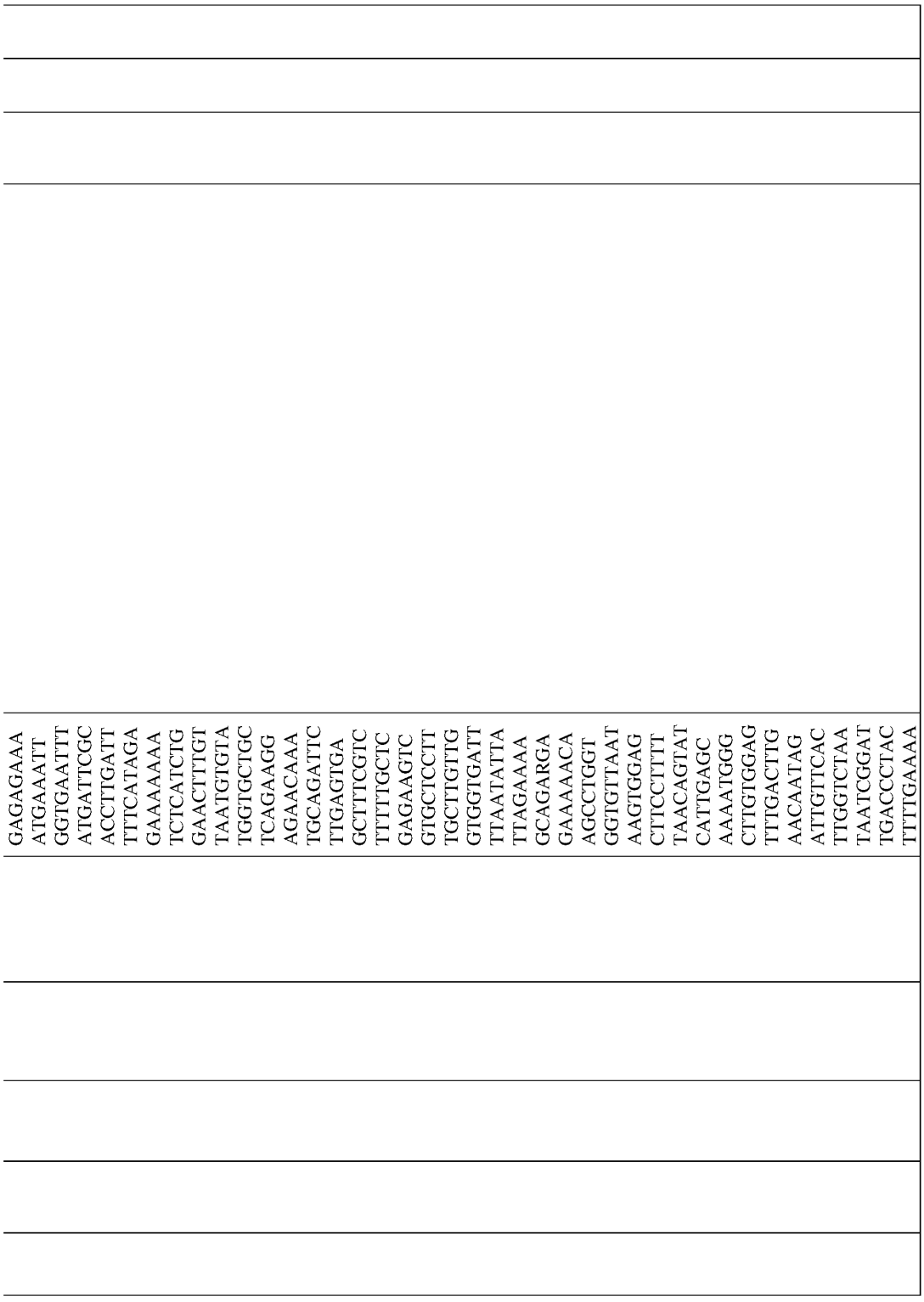
				ACAGATC CATCTGA GTATCCT GTTTCAG ATGTTGA CGCTGTG TTGGCAA CTCCGTC TAGAACA TTACCTC CCTCCCT GTCTGAT GTGGCAG CAGCACA CGCGGCT CCGTCGG AGGCTAA TGCCTTA GACGTTG ATGCTGA GTTGGCT GCTCCAT CTGAGGA GGACATT TCAGCAA CAACACA TGCTGCG TTGGCTG CTCCATC TGAGGAC CCTCCAG CAACAAC ACAAGAT TCTGTGG AGATTTG TGCTCAA GAGCAAT TCATCAG CAAATTG TCAAACC						

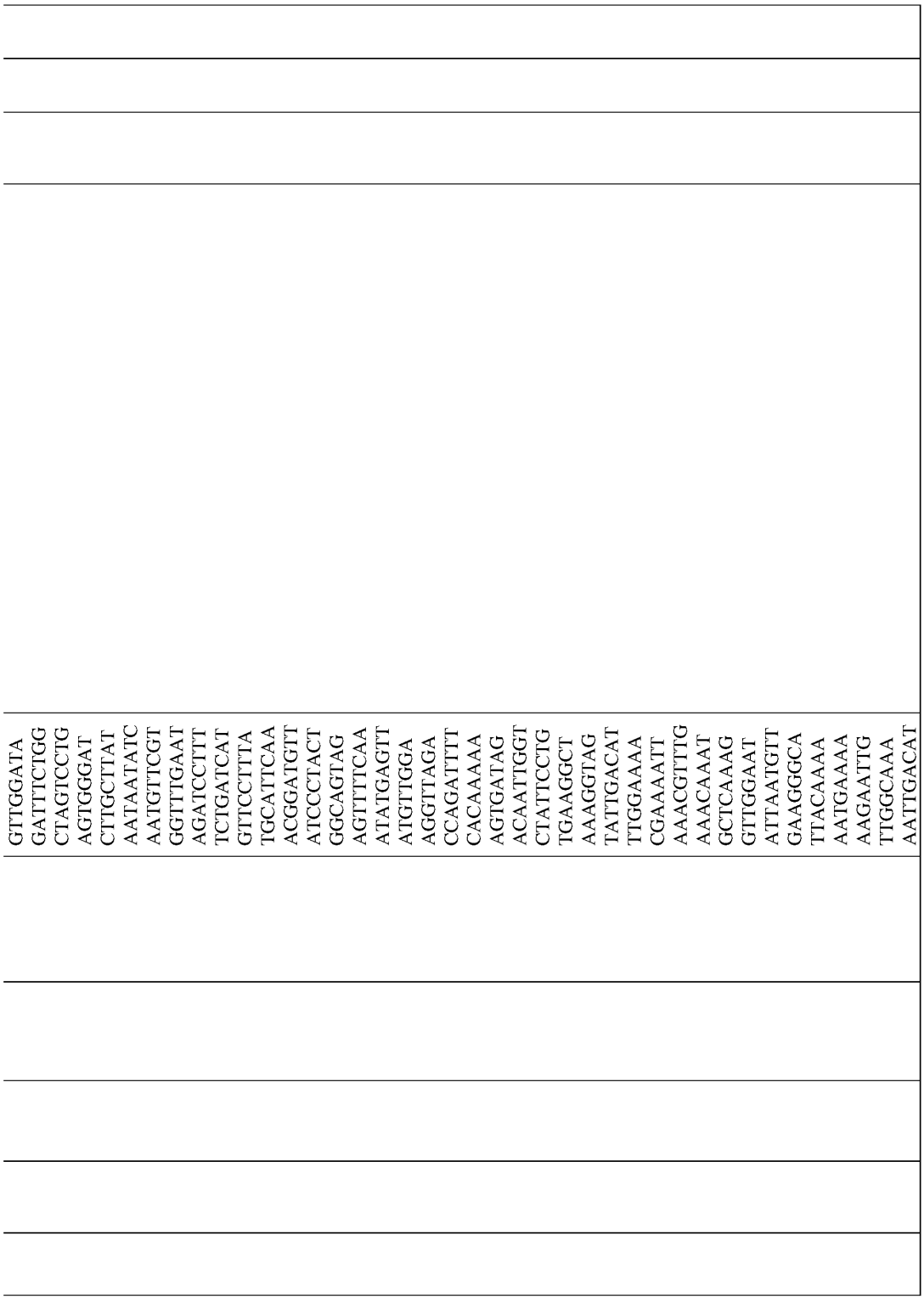

AGATCAC
ACGCGTA
TTACCAA
TCCCAAA
ATCTGTC
AGACGTC
AGTCCTC
TATCCCC
CTGGCAG
TTCCAAG
ACGCAGC
CGACGTG
TGGCGGG
GGTTGGA
GTTGAAT
TTAAAAT
TCAGGAT
TGGGGAA
AAAGATC
TAAGAAA
CAAGCTA
TGGTGGC
TCTAAAT
ATTATTG
AGGACAA
TGAAGGA
TTCAGCC
GTCAAGC
TTTAGAT
GACTACG
CCAAGCT
CTTCAAG
AATACTC
TTTCTCGT
AGTCATG
TGCAGGC
TCTTGCT
GCCCTTTT
TGGGTGG
TCTGTCC

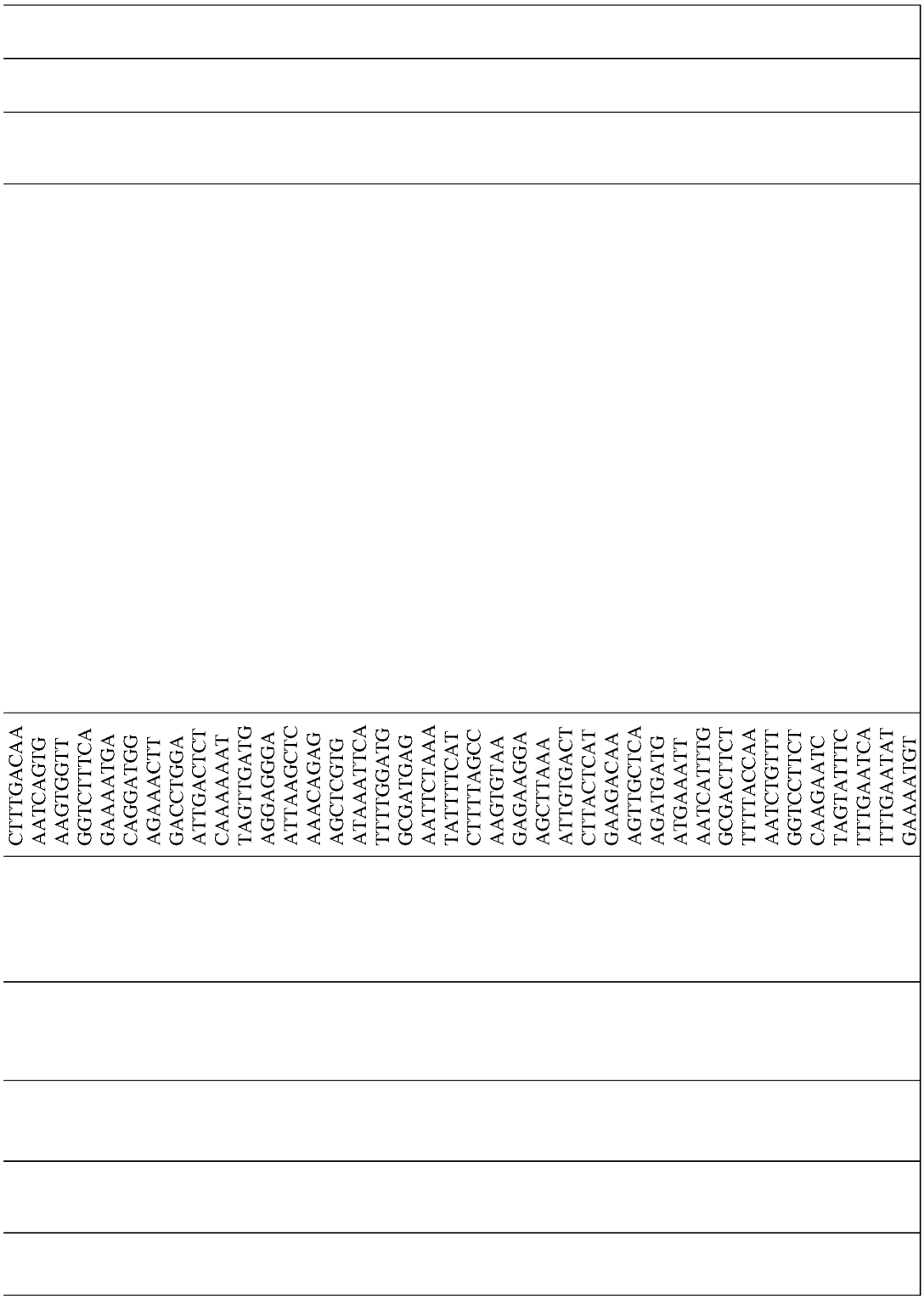
									658
								409	
							cd01650		
RTE	RTE1_ZM	.	Zea mays	AAGGGGC	GTT	AIHVPEP	KRGRGRPNLTWEEAIKRDLKEWNIPMELCLDRSAWKE	AAAATGG	GATGTTGAT
				TATCGAC	GTTGATGTT	VAPIEEKLIQHRLRWFGHVHRRPPEAPVHRGIIRRDNNV		GTGTTAG	GTTGATGTT
				AGCGAGC	AACTTGCTT	VVPKKDTFRYLGSMLQRDGDIDEDVSHRIKAGWMKWR		AGGAGGC	GGGACTAA
				TCGATGC	GCCTACCCC	WRQTLESKGFRISRTKTEYMRCDFGTTISEDGDVSLGGQ		GAGTTAG	AAGGCTAT
				AAGCTAA	TGTTGGGTT	YDKVVTSVRTTDGDTNVFPINIGLHQGSALSPYLFALVI		TAGTTGG	TCAACTCTA
				AACGTAG	TCTCTCTTT	KNKGDIQSCTNYRGIKLMSHTMKLWERVIEHRLRGMT		GGTCCCT	CTCTTTTTG
				TACAGGT	GTGACATCT	VFIDLEKAYDKIPRNLMWWALDKHKVPTKYVTLIKDM		AGGATCC	CTTTTCCCC
				GCCGACT	TTTAGTGCT	DDLYRRLSTKEGEKDVYYKMARIRERKTRDLNQVKCIKD		GTGTAGG	TCCCCTTTC
				AGGGTCT	CTCAAAAG	EMDQLLVKGQDIKQRWQRYFDNLFNGENETMDTQLDD		CGAGGGG	CTTTTCCCC
				TCGCATC	TTTAGTGCT	SFDDLNRCFVRRIQESEVKEALKRMKGGKAMGPDGIPIE		ACGGGTG	TTAGTGCT
				TGTCTCA	CTTTTCCCT	VWKCLGDIAIVWLTKLFNHIFRSNKMPDEWRRSTLVPIF		TAGTTGA	CTTTTCCCT
				CTGGGAC	TGATTAGGC	CIRKVASEVFGVTKGSIGEPKDTWWWTEDVQKAIKEKK		TGGAAGC	TGATTAGGC
				TGGGGAT		ECYRSLFHDRSAVNIERYKVAKKTAKRAVSEAKGRAY		CGAGGGG	
				ATTACCG		WKLKGETSEVFRERVFVEGAWSEEEDANNMWVKMAT		CGATGTG	
				CAACAAG		KVIPGESVVPQHNLVVADFRFRIRTHRDKQAKIARTKW		ACGGGTG	
				CTTATTTT		NTFFRKRDSHLVTFSSGHRSSQIDFVLTRREDKQACLDC		AGCTGGC	
				AGCTGGC		VGSTNVGYELAHGGFGYGSRNQGEDILDFAVAYNLVIA		GATGTTG	
				GATGTTG		MPLSESEKRKFWEDLDGMVRAVPTNEKLFIGGDLNGH		CTGATAT	
				CTGATAT				TGTTTGTC	
				TGTTTGTC				GATTTAC	
				GATTTAC				CTCTCATT	
				CTCTCATT				GTTTCTA	
				GTTTCTA				C	
				C					

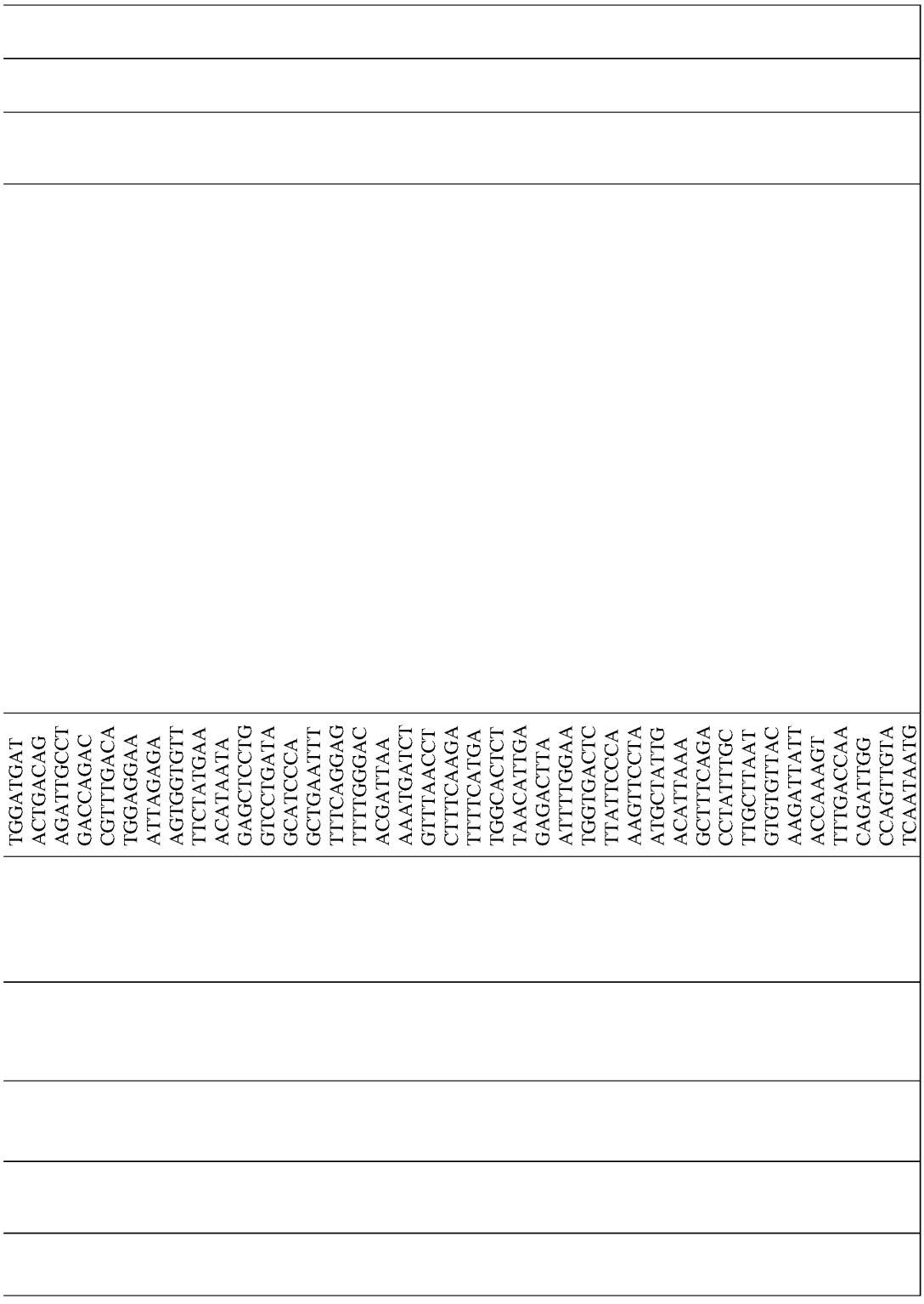
				AAAAAGC GAAGGAG GTGGAGA ACACTGG TTTACTTT GGTACTC AGGAACA GTGGCAA ACAAAAA TGAGTAG GTATCCT GATTGAT AAGACCC TCAAGGA TGGAGTG GTGGACG TCAAAAG GCAGGGA GATAGGA TTATCCT AGTCAAA CAGTTTG GGGATGT AGTCTTG AACATTA TAAGTGC AT					
L1	LINE 1- 43_Z M	.	Zea mays	GTCTAGT CAGCAGA ATCCCCC AACCCTA GTGAAAC TTGACAG AGTATTT ATGCACA CTGGACT GGGTGGA TTTGTTTC CAAATTC CTTGCTTC	TAGCTAGC ATTGTGGTA ATTCGTGTT AGTTTCTTC TTTAGNCAT GGATGGGT TTCTGTATG CCTGTTACT GGCCGCCG TAGGCGCCT CTTTTGTTT TCTTCTGTA ACTGTCCTT	MTIVWQEGAGFILSPFGLSLRDFWRLFRWPGILFQLCSV LLQRLNSKLKATARGLQAWSAKKVGHISASLALARVLL HQLEIAQDSRGLSPAELWFLLCLKKHSLALSSLQRTIAR LRSRIGWLREGDANTKLFHMHSRFRKKKNFVPKLVSNG AVLTSHDAKADLVNEFYGNLLGQCSDREQTISLESLGIP CHDLSALDAPFTEKEVWDTILHLPSDKAPGPDGFTGKFY RVCWEIIKADVMAVFSAVGSRRFANFQILNTAYITLIPKF EGADQVKNFRPISLVHSVAKLLTKLLANRLAGRLHDMV SPNQSAFIKGRFIQDNYMLVQQTTRYLHQQKQPRMLLK LDISKAFDSVSWPFLLEVLQHLGFGQVWRDIISGLLWTS SSQVLLNGLPGQPILHRRGLRQGDPLSPFLFILVMDTLGF LINKAAQDGLLQPLAPRALQHRVSLYADDVVIFLRPAA MDINVILDCLQLFGDASGLRTNVQKSSVYPIRCGDQDLA	cd016 50	263	526

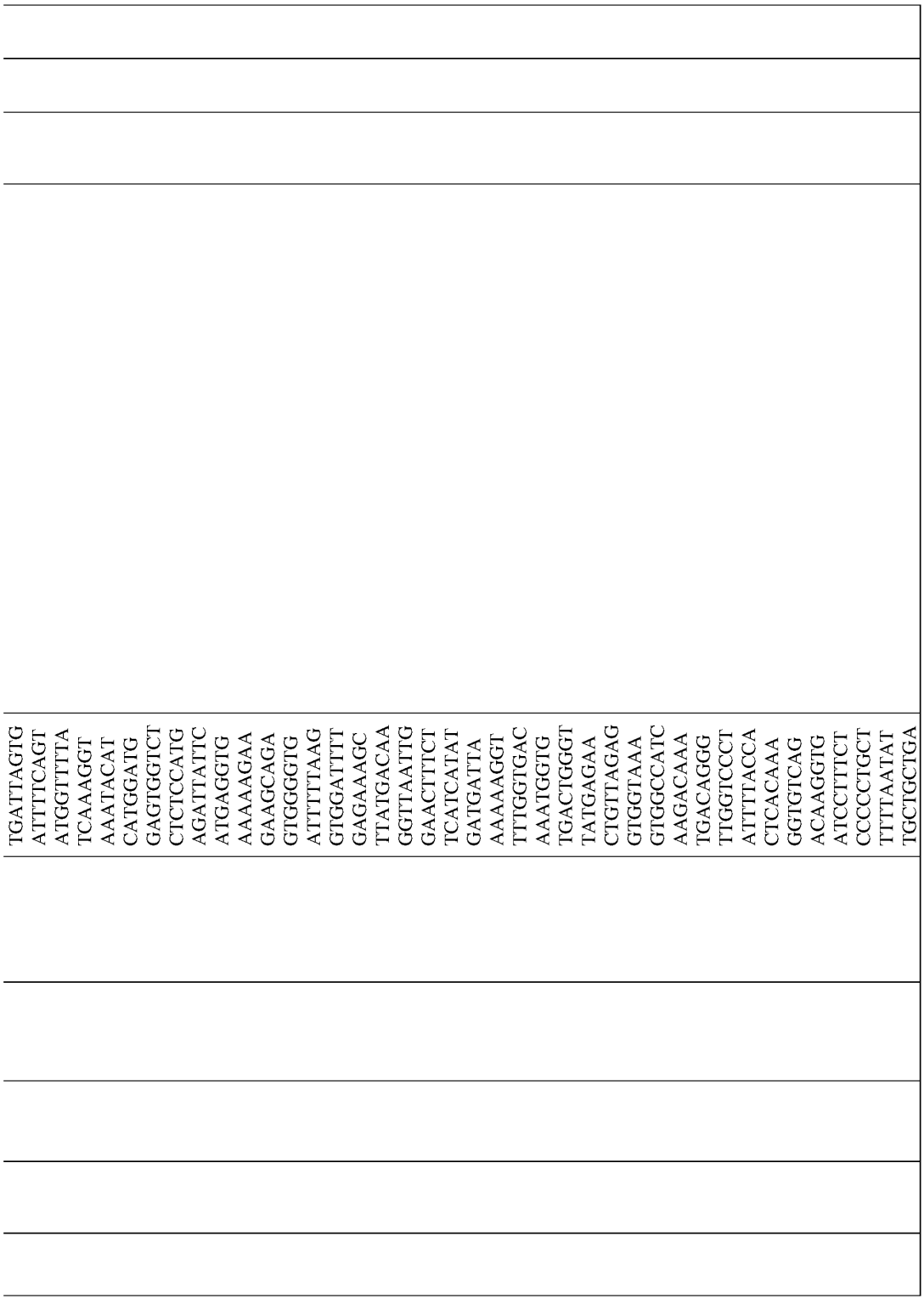
			AGAGTTC TGCCTCN AATGACT CAGACCA CTGCCCC CTCCTTTT GGGCCTG A	CTTCTTTAA TATAATATC GCGCAGTTC TCCTGCGCT TTCGAGAA AAAAAAAA	LLQNMLPCELAEFPCRYLGLPLAIKKLSKIQVQPIIDKIA NQLSGWKAELLTKAGRKVHVQHVLTGMVIYLAMAVD LPAWALKAIDKIRRGFLWRGRRDVKGGHCHVAWGIVC RPRELGGLGISSMKELAWSLRMRWLWLAKTEPSRPWA SLAIQVPKKVKDFFSMAMQTELGNGATTFFWTDRWLS GHRIAELVPRLFGIIPKRRVSKRTVVDAISNQSWISDIQG ALSVGIINEFLILWDLIDSIVLRPQVEDKHFFRLAANGKF SSKEAYRGFFLGSSEFEPFHRIWKSWAPPKTKFFMWLV AHRKVWTADRLQKRGMDHPERCPLCDQDQETLDHMLI GCVFAREFWFKLLLQVNLQFLAPQSEDSAFLEWWRTLC TKVSGIARDGLNSLVILGVWTLWKQRNGCVFDNKSPSI ADAIRRVELEVHLWEMAGAKKLSLLTAPIPGIPAS			
				TGATGATA GGCCTCATT TGGAATTGT CAGGGTTTA GGGAATAG TTCTAAGGT TCAGTTTCT GAAAGAAC TCATTAGAG ATGAGAAA GTTGACTTT ATTGGTTTG CAAGAAAC CCAAAAAA GTCATTTTA GTGATTCTT GGTTGACTT CGCTGGCTG GTAGTAAA CCTTTTGCC TGGTTTTCT GCCCCGCCT AATGGTAG ATCTGGTGG TTTATTGGT GGGTTTTAA TACTGATGT GTTTGATGT	MAKKKRRKAPWAEAPSSPRAPAPVAHGPGGVRSGLVG LPSTVGFAAQGKCAPAPFASSVSWTGWSQGISPLAPTLA HSSRVLLPAATSDAAVSKSRLGFPPDFVGSHSFPHCAAF LSRLWSLPVSSRSDFHGGWCVTSLRPPVPEEEWPRLGSG VVRSSKNLGEMDPKNKNIPNPSWDRWGQKDSAQGSQS WNRNRNLSWRLKPQIGKSDPQEGLMSSSSGDGIQSGNPI LSPSSLKKEIDLARPAFCQKCGVEGHHARNCFNALWCDI CRKDTHVTARCVLPKQSKPNMPIVGMAADGLGFYLSHF AKPISKKPKRVFIGLVKVIEGLVSAEDLAKDFGFHFPWG KTWKATKCHSGFLMQFPSQERLDEMISFPELKMKMSGA KIVVVPWSSRAKPKSRLHTAWIVAENVPEELQNYQSICE IGSMIGAVEEVDLMSLLDSEDIVRFKVHIKSVAMIPPVVE VAVKPFLYDIYFRIESISDEGWNDETINLGKRASVDIHGI NDPVFDKSGKKPRNDDEVGDEDPXPNFKVGESSSQGKE SFKKSEVFMDSSKDKSGDGMRLEDLSDEKVDFDPNEDD LLSSQELEEFAKDMEEMHTDFQAKVGEVLSIPLTKDEK QDXVGKKSSEILMSTSEGIRKSSRLEKNEDIKVADKAISR AEAKDAFLNKGMNNNPFSLLNASNEDLFDISNKLGVCL GSSDSSAMANLDLIRDLEMSRKILAIQSCGSKKNEEPIVD DINEVDSSVHADSEKDIDYDLLDVMVLRKGRKIKHRKK QQKKASPKIRYTPNKKREAENFXLPPDPF			
L1	LINE 1- 15_Z M	.	Zea mays			NA	NA	NA	NA

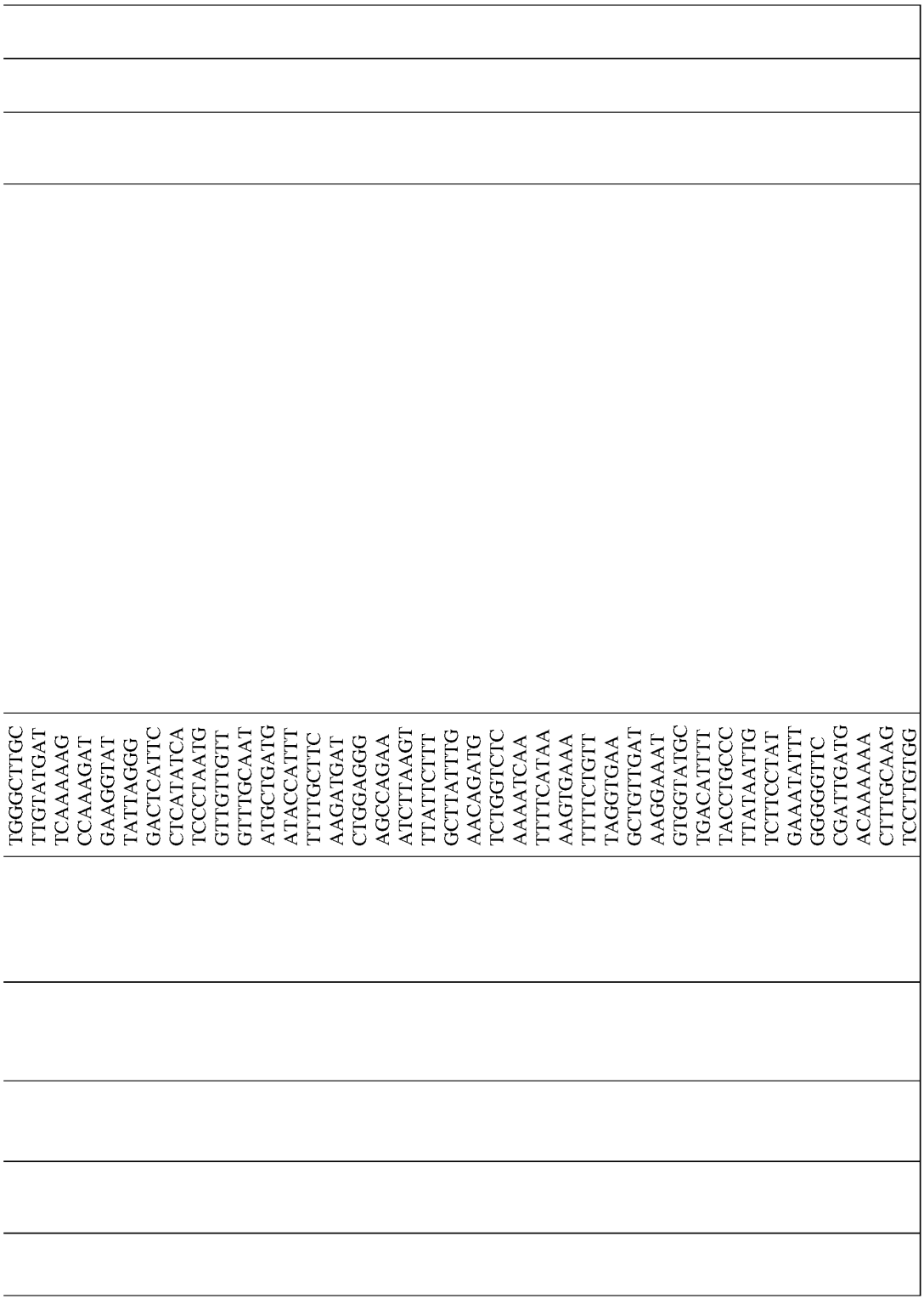

GAGAGAAA
ATGAAATT
GGTGAATTT
ATGATTCGC
ACCTTGATT
TTTCATAGA
GAAAAAAA
TCTCATCTG
GAACTTTGT
TAATGTGTA
TGGTGCTGC
TCAGAAGG
AGAACAAA
TGCAGATTC
TTGAGTGA
GCTTTCGTC
TTTTTGCTC
GAGAAGTC
GTGCTCCTT
TGCTTGTTG
GTGGTGATT
TTAATATTA
TTAGAAAA
GCAGARGA
GAAAAACA
AGCCTGGT
GGTGTTAAT
AAGTGGAG
CTTCCTTTT
TAACAGTAT
CATTGAGC
AAAATGGG
CTTGTGGAG
TTTGACTTG
AACAATAG
ATTGTTCAC
TTGGTCTAA
TAATCGGAT
TGACCCTAC
TTTTGAAAA

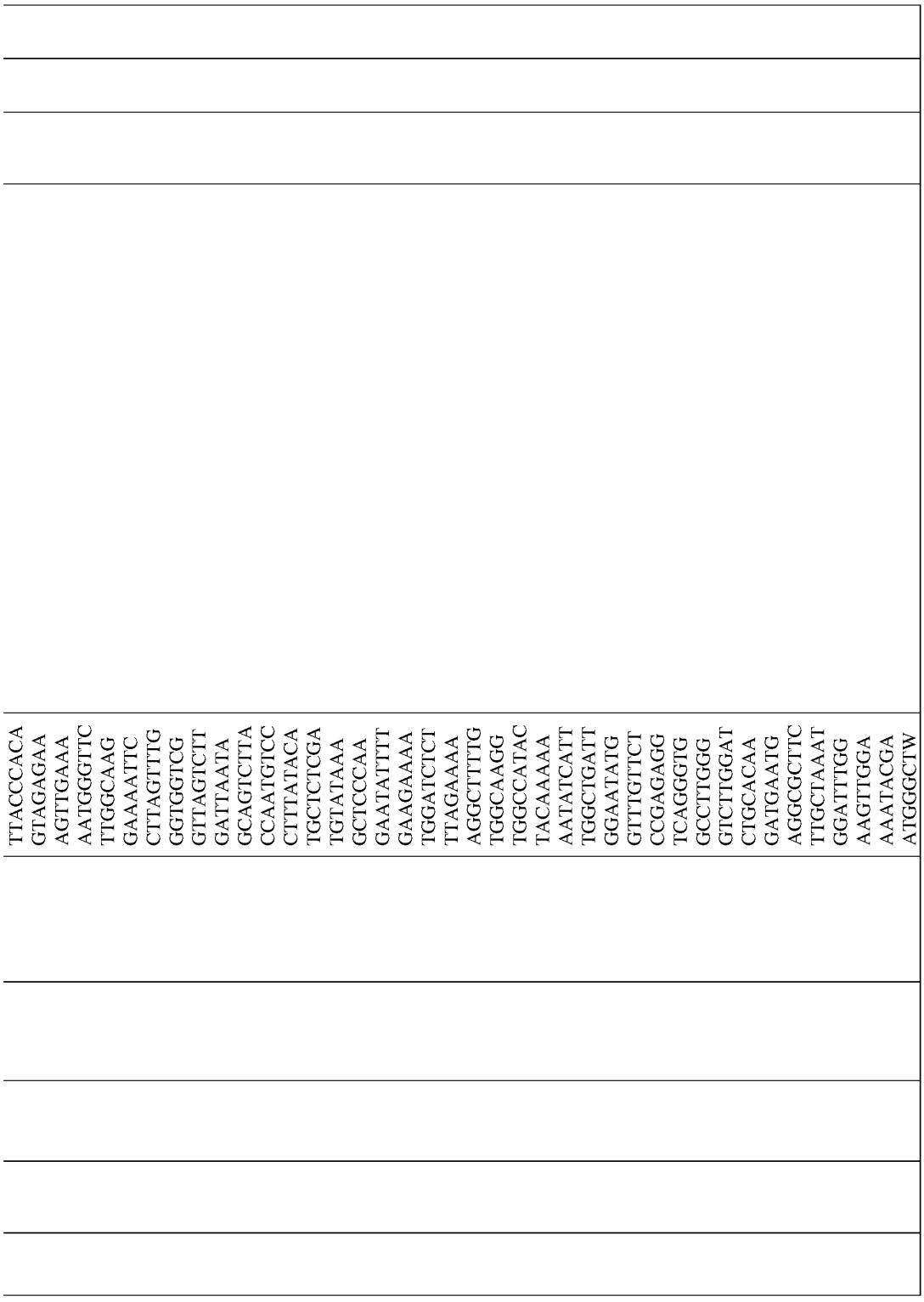

GTTGGATA
GATTTCTGG
CTAGTCCTG
AGTGGGAT
CTTGCTTAT
AATAATATC
AATGTTCGT
GGTTTGAAT
AGATCCTTT
TCTGATCAT
GTTCCTTTA
TGCATTCAA
ACGGATGTT
ATCCCTACT
GGCAGTAG
AGTTTTCAA
ATATGAGTT
ATGTTGGA
AGGTTAGA
CCAGATTTT
CACAAAAA
AGTGATAG
ACAATTGGT
CTATTCCTG
TGAAGGCT
AAAGGTAG
TATTGACAT
TTGGAAAA
CGAAAATT
AAACGTTTG
AAACAAAT
GCTCAAAG
GTTGGAAT
ATTAATGTT
GAAGGGCA
TTACAAAA
AATGAAAA
AAGAATTG
TTGGCAAA
AATTGACAT

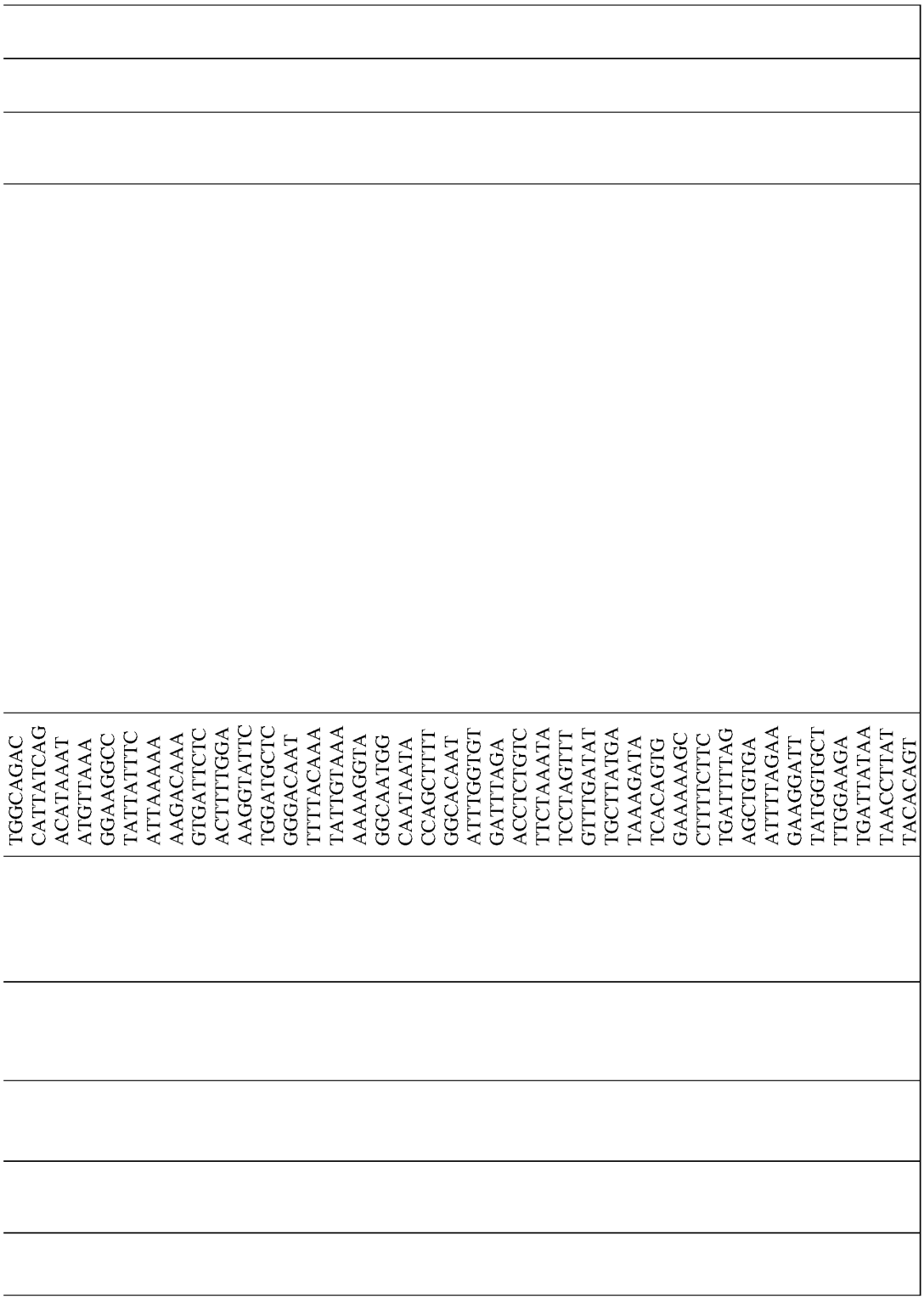

CTTTGACAA
AATCAGTG
AAGTGGTT
GGTCTTTCA
GAAAATGA
CAGGATGG
AGAAACTT
GACCTGGA
ATTGACTCT
CAAAAAAT
TAGTTGATG
AGGAGGGA
ATTAAGCTC
AAACAGAG
AGCTCGTG
ATAAATTCA
TTTTGGATG
GCGATGAG
AATTCTAAA
TATTTTCAT
CTTTTAGCC
AAGTGTAA
GAGAAGGA
AGCTTAAA
ATTGTGACT
CTTACTCAT
GAAGACAA
AGTTGCTCA
AGATGATG
ATGAAATT
AATCATTTG
GCGACTTCT
TTTTACCAA
AATCTGTTT
GGTCCTTCT
CAAGAATC
TAGTATTTC
TTTGAATCA
TTTGAATAT
GAAAATGT


TGGATGAT
ACTGACAG
AGATTGCCT
GACCAGAC
CGTTTGACA
TGGAGGAA
ATTAGAGA
AGTGGTGTT
TTCTATGAA
ACATAATA
GAGCTCCTG
GTCCTGATA
GCATCCCA
GCTGAATTT
TTTCAGGAG
TTTTGGGAC
ACGATTAA
AAATGATCT
GTTTAACCT
CTTTCAAGA
TTTTCATGA
TGGCACTCT
TAACATTGA
GAGACTTA
ATTTTGGAA
TGGTGACTC
TTATTCCCA
AAGTTCCTA
ATGCTATTG
ACATTAAA
GCTTTCAGA
CCTATTTGC
TTGCTTAAT
GTGTGTTAC
AAGATTATT
ACCAAAGT
TTTGACCAA
CAGATTGG
CCAGTTGTA
TCAATAATG

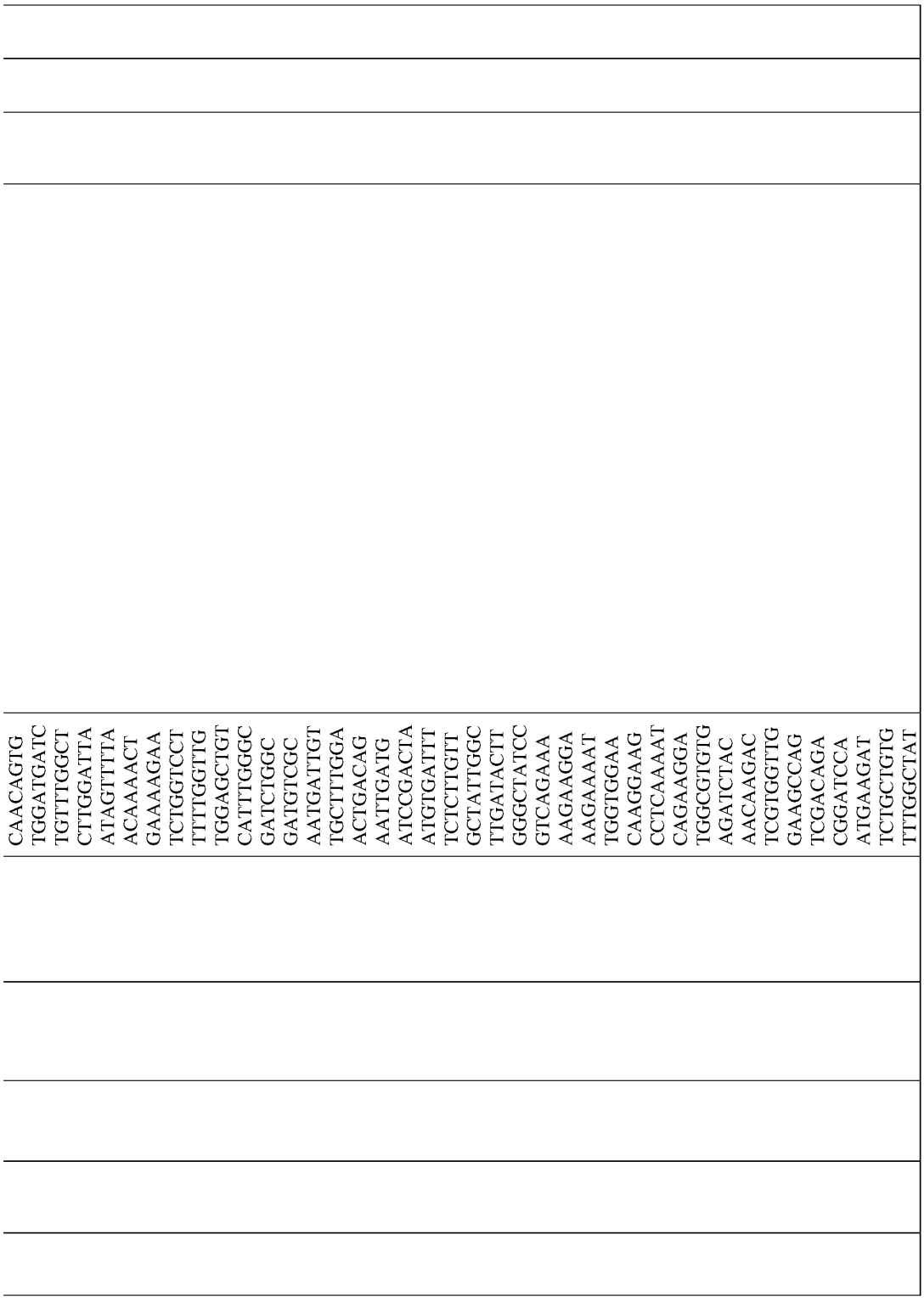

TGATTAGTG
ATTTTCAGT
ATGGTTTTA
TCAAAGGT
AAATACAT
CATGGATG
GAGTGGTCT
CTCTCCATG
AGATTATTC
ATGAGGTG
AAAAAGAA
GAAGCAGA
GTGGGGTG
ATTTTTAAG
GTGGATTTT
GAGAAAGC
TTATGACAA
GGTTAATTG
GAACTTTCT
TCATCATAT
GATGATTA
AAAAAGGT
TTTGGTGAC
AAATGGTG
TGACTGGGT
TATGAGAA
CTGTTAGAG
GTGGTAAA
GTGGCCATC
AAGACAAA
TGACAGGG
TTGGTCCCT
ATTTTACCA
CTCACAAA
GGTGTCAG
ACAAGGTG
ATCCTTTCT
CCCCCTGCT
TTTTAATAT
TGCTGCTGA


TGGGCTTGC
TTGTATGAT
TCAAAAAG
CCAAAGAT
GAAGGTAT
TATTAGGG
GACTCATTC
CTCATATCA
TCCCTAATG
GTTGTTGTT
GTTTGCAAT
ATGCTGATG
ATACCATTT
TTTTGCTTC
AAGATGAT
CTGGAGGG
AGCCAGAA
ATCTTAAGT
TTATTCTTT
GCTTATTTG
AACAGATG
TCTGGTCTC
AAAATCAA
TTTTCATAA
AAGTGAAA
TTTTCTGTT
TAGGTGAA
GCTGTTGAT
AAGGAAAT
GTGGTATGC
TGACATTTT
TACCTGCCC
TTATAATTG
TCTTCCTAT
GAAATATTT
GGGGGTTC
CGATTGATG
ACAAAAAA
CTTTGCAAG
TCCTTGTGG


TTACCCACA
GTAGAGAA
AGTTGAAA
AATGGGTTC
TTGGCAAG
GAAAATTC
CTTAGTTTG
GGTGGTCG
GTTAGTCTT
GATTAATA
GCAGTCTTA
CCAATGTCC
CTTTATACA
TGCTCTCGA
TGTATAAA
GCTCCCAA
GAATATTTT
GAAGAAAA
TGGATCTCT
TTAGAAAA
AGGCTTTTG
TGGCAAGG
TGGCCATAC
TACAAAAA
AATATCATT
TGGCTGATT
GGAATATG
GTTTGTTCT
CCGAGAGG
TCAGGGTG
GCCTTGGG
GTCTTGGAT
CTGCACAA
GATGAATG
AGGCGCTTC
TTGCTAAAT
GGATTTGG
AAGTTGGA
AAATACGA
ATGGGCTW


TGGCAGAC
CATTATCAG
ACATAAAT
ATGTTAAA
GGAAGGCC
TATTATTTC
ATTAAAAA
AAGACAAA
GTGATTCTC
ACTTTTGGA
AAGGTATTC
TGGATGCTC
GGGACAAT
TTTTACAAA
TATTGTAAA
AAAAGGTA
GGCAATGG
CAATAATA
CCAGCTTTT
GGCACAAT
ATTTGGTGT
GATTTAGA
ACCTCTGTC
TTCTAAATA
TCCTAGTTT
GTTTGATAT
TGCTTATGA
TAAAGATA
TCACAGTG
GAAAAAGC
CTTTTCTTC
TGATTTTAG
AGCTGTGA
ATTTTAGAA
GAAGGATT
TATGGTGCT
TTGGAAGA
TGATTATAA
TAACCTTAT
TACACAGT

GTAATAAT
ATTACTATT
TCTGAGGA
GGATGATA
AATCAGTTT
GGCTTCTGG
GTAATAAA
GGTTTTTCT
GTTAATTCT
TTTTACAAA
GAAATCAA
ATGTGCTCA
GATTCCTTT
GCCTTCTAG
TTTCTTGTG
GAAGACTA
GACTCCCAC
ATAAAATC
AAAGCTTTT
TTGTGGCTT
GTTATGCAT
AAAAAGAT
TTTAACCAA
GGATAATTT
GTTCAAAA
AAAATTGG
AAGGGCAA
TCTGGATTG
TACTTTTTG
TGGATTTTTC
AGAATCCA
TAGATCATT
TGTTTTTCA
GTGCTCGGT
TGCTAGATT
TACTTGGAG
AATAGTCTA
CTCGGCTCT
TGGCTTAAT
TTCTATTCC


CAACAGTG
TGGATGATC
TGTTTGGCT
CTTGGATTA
ATAGTTTTA
ACAAAACT
GAAAAGAA
TCTGGTCCT
TTTTGGTTG
TGGAGCTGT
CATTTGGGC
GATCTGGC
GATGTCGC
AATGATTGT
TGCTTTGGA
ACTGACAG
AATTGATG
ATCCGACTA
ATGTGATTT
TCTCTTGTT
GCTATTGGC
TTGATACTT
GGGCTATCC
GTCAGAAA
AAGAAGGA
AAGAAAAT
TGGTGGAA
CAAGGAAG
CCTCAAAAT
CAGAAGGA
TGGCGTGTG
AGATCTAC
AACAAGAC
TCGTGGTTG
GAAGCCAG
TCGACAGA
CGGATCCA
ATGAAGAT
TCTGCTGTG
TTTGGCTAT

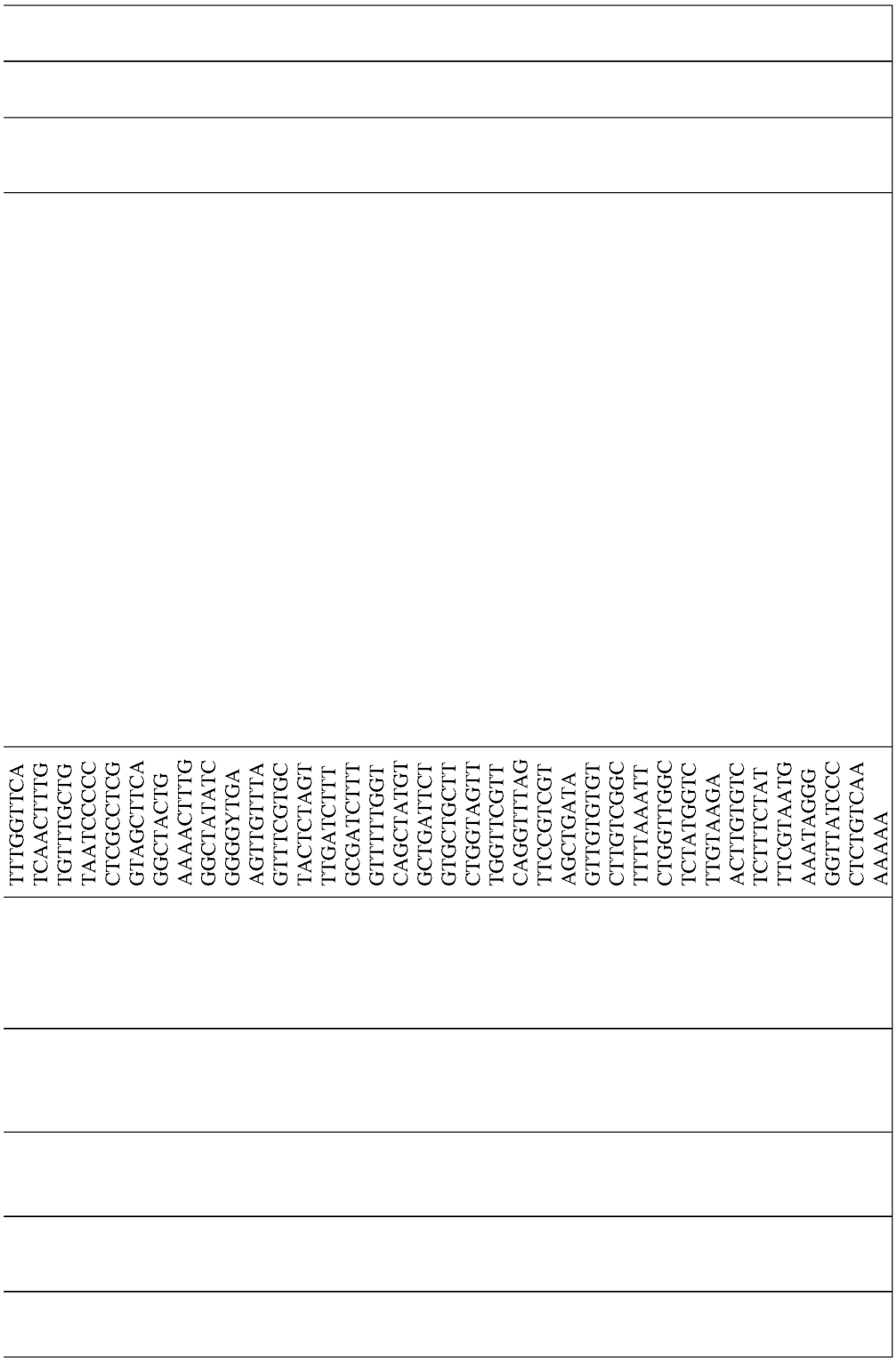

TTTGGTTCA
TCAACTTTG
TGTTTGCTG
TAATCCCCC
CTCGCCTCG
GTAGCTTCA
GGCTACTG
AAAACTTTG
GGCTATATC
GGGGYTGA
AGTTGTTTA
GTTTCGTGC
TACTCTAGT
TTGATCTTT
GCGATCTTT
GTTTTTGGT
CAGCTATGT
GCTGATTCT
GTGCTGCTT
CTGGTAGTT
TGGTTCGTT
CAGGTTTAG
TTCCGTCGT
AGCTGATA
GTTGTGTGT
CTTGTCGGC
TTTTAAATT
CTGGTTGGC
TCTATGGTC
TTGTAAGA
ACTTGTGTC
TCTTTCTAT
TTCGTAATG
AAATAGGG
GGTTATCCC
CTCTGTCAA
AAAAA

Table 11 provides Gene Writers comprising novel retrotransposon elements from Table 10 above, altered for improved efficiency of integration into the human genome. Retrotransposase polypeptides were improved through consensus mapping to rederive the optimal amino acid sequence, listed here as the "Consensus Optimized Protein Sequence", as described in Example 5. As described in Example 6, Template molecules for use with cognate retrotransposase enzymes were mapped back to their host genomes and flanking genomic DNA used to elucidate target site motifs. When detectable, conserved sequence motifs from the flanking genomic DNA of endogenous occurrences of an element were aligned to the human genome, and new sequences were derived from the human genome as 5' or 3' "Human Homology Arms." In some embodiments, a template RNA described herein comprises one or both of a first homology domain comprising a sequence of a 5' Human Homology Arm of Table 11 (or a sequence having at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto) and a second homology domain comprising a sequence of a 3' Human Homology Arm of Table 11 (or a sequence having at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto).

Table 11. Retrotransposase systems with improved integration activity

Element	Predicted Target motif	Consensus Optimized Protein Sequence	5' Human Homology Arm	3' Human Homology Arm
L2-18_ACar	CCACAA GGCCCT GGCAGG GCTtaaa		AGTAGACTGCAGCCC CAGGCCCTTTCCATCT GGGTGTGACTCAGCA CTTCCTAGGAAGGAC CATCTATAAAGCCTG AGTCCACAAGGCCCT GGCAGGGCT	taaa
CR1_AC_1	GTGTGA GTTTCTG CCAGAG CCattctatga		ccacggcacctggcccagtgttttttt aatacaaattatagcctatggatccc actgaggacctctgtttgtgagact gggtgtgagttgctgtcagagcc	attctgcaaacctccagtgaggggaag gttgagggcagtggttggtgctcacggt ggtgctgctcaaaactgacaggcagac acccggcctccagcttgc
RTE-1_Aip	Tacaaggcttt gttgttgttgttg ttgttttttt		TA	caaggctttgttgttgttgttgttttttt
Dong-1_MMa	GGTTCA ACGGGC CGGGCC CGGtaataat aataataaaa	MKANPAMETRSMRNRRMRSPEEGAPTGAGPGTGDRASG QRLEEDNVQERPSSQRAQPVPRTRNRNGSSRNHQGHAAP TDVAVADRRQSLTLAGGRRQRIMWTREMNMYVIRCYYV CTRMETDMSGRSGMLEMFNERFPRFARQLDLNKLYTRQR AIMSHNMLTAAEVEYIKLEVQREIGEEGTRSSDTSRRSSVG LDAPITSESRDSPAPTAPGPTADMQRQQLRDELALHMGTA VTQFRGTDPMSRHRIPKLRYSYRLTSAVSILNQDILPQYLE	AAGGATGGGGTTCAA CGGGCCGGGCCCGG	taataataataataaaa

		AVENLEDLQFIVYSAAIAVVKTLGLRTRPQGEDGTRPNAQ KPVWMRRLESRIATLRSKVGRLTQYKQGNRSTRLVRHVA EIVRPAELRDLREVDITEILDTHVQRLSALAKRLRRYGECS KRKEQNRMFNINEREFYNRIRNDKVDFGEGLPEIGDVTQF WANLWENPTEHDGDGMWLAEEERQCDGIGDMTAVVVT AQDIREATRYTRNWAAPGPDFVHNFWYKKLTTIHGRMAE CFNTVLGDPTQLPEFITRGVTFLLPKDQHTADPAKYRPTTC LSSLYKVLSSVIARRVQAHCDANNVMTEEQKGCRRNTQG CKDQVIVDAVIVGQATQKQRNLSMAYIDYKKAYDSVPHS YLLKVLQLYKVDGNVIRLMQHAMGMWSTSLHITDGTAV LRSRTLSIRRGIFQGDTFSPLWFCLAMNPLSKALNQCNYG YQLKSGERSTRVTHTFYMDDLKLFAESVQRLHQLLQLVT TFSNDIRMEFGIDKCRSIHLRRGQVMDASCFRVNEQEEIRN MVEGETYKYLGFLQLRGIRHTMIKKELQEKFLSRVNCILK SFLSAGNKVKAINTFAVPLLTYSFGVVKWTKTDLEAIERA VRVAFTKHRMRHPKSSIERVTLPRAAGGRGVTDIQALCVS QIQQLRAYFVESQNRHEIYRTVCEADHGFSALHLAQEDYQ LNCDIKTVDEMIAMWKQKELHGTHPHQLELEHIDKVASN TWLVRGDLFSETEGFMVAIQDRVIATKNYRHYILHEDVED RCRKCNSVGETIEHVVAGCSVLAGSAYLDRHNEVAKIVH QQLALKHNLVDRFVPYYKYLPDPVLENSCIKLYWDREIIT DVLIRANRPDIVVYDKRMKRVTLIDIAVPLDHNVQSTFSN KIAKYHDLAEELKQMWHLEDVRIVPVVLSATGIVPKSLLR SLDELELKKDLHSIQKAVILGTCSIVRRFLNHHN	GTTTCGCGTCGTCCG CCTCAGTTTCGGCCCT TGTTTTGACTCGGGA GGCGTGTCCAAACGC CAGGTAACCACTATA TAGTGAGCACGGTAG CATTAGTCG	a
L2-2_DRe	TGAGCA CGGTAG CATTAGT CGa	MCFLIPVVTNTRKTREVRCRRNPHNLRSIHVSTISQLSLSV GLWNCQSAVNKADFITSIATYSDYNLMALTETWLRPEDT ATHATLSANFSFSHTPRQTGRGGGTGLLISKEWKFTLIPSL PTISSFEFHAVTIIHPFYINVVVIYRPPGKLGHFLDELDVLLS SFSNFATPLLVLGDFNIYVDKPQAADFQTLLASFDLKRAPT SATHKSGNQLDLIYTRHCFTDQTIVTPLQISDHFLLSLNIHI TPEPPHTPTLVTFRRNLRSLSPNRLSTIVSDSLPPSRKLTAL DSNSATNTLCSTLASCLDRLCPLASRPARASPPAPWLSDA LREHRSKLRAAERIWRKTKNPAHLLTYQTLLSSFSAEVTS AKQTYYRLKINNATNPRLLFKTFSSLLYPPPPPASSTLTTD DFATFFCTKTAKISAQFAAPTTNTQDTTPTPHTLTSFSQLSE SEVSKLVLSSHATTCPLDPIPSHLLQAISPAVIPTLTHIINTSL DSGLFPTTFKQARVTPLLKKPNLDHTLLENYRPVSLLPFM AKILEKVVFNQVLDFLTQNNLMDNKQSGFKKGHSTETAL LSVVEDLRLAKADSKSSVLILLDLSAAFDTVNHQILLSTLE SLGVAGTVIQWFRSYLSDRSFRVSWRGEVSNLQHLNTGV		

		PQGSVLGPLLFSIYTSSLGPVIQRHGFSYHCYADDTQLYLS FHPDDPSVPARISACLLDISHWMKDHHLQLNLAKTEMLV VSANPTLHHNFSIQMDGATITASKMVKSLGVTIDDQLNFS DHISRTARSCRFALYNIRKIRPFLSEHAAQLLVQALVLSKL DYCNSLLAGLPANSIKPLQLLQNAAARVVFNEPKRAHVTP LLVRLHWLPVAARIKFKALMFAYKVTSGLAPSYLLSLLQI YVPSRNLRSVNERRLVVPSQRGKKSLSRTLTLNLPSWWN ELPNCIRTAESLAIFKKRLKTQLFSLHFTS	caataaaaataataataatattattatt ataa	'AATAATAATAATAATAATAA TAATcaataaaaataataataata'
R4-1_AC	AAATAA TAATAAT AATAAT AATAAT AATAAT AATAAT AATAAT AATcaataa aaataataata atattattattat aa	MENQTQTKSILARSSPRIKRTAVDAADSGHPSTSGLRNHQ NPNEQSQKRQKYTMPENRTIMKCYYKSEPQRRGYQKRM HQLWKQEYPDSQITEPRLADQRRFIIRNKVFSEVELEEIQKI CKVDYHQTTAQTAAETPATLGMVEQIEPEESVALLQEFEE PALETSVEPPGTLTARQQELKDKIMAHAAAANAIRQRLPTL KTVPKRHLAPLMKDVNAALSTVQITSIEQTNQFAYSAAVI VTEELGLLQPRQPQRKSTGKPKWKVRLELKIKKLRSDASN LKNMKERKLKNDKIKQYLIRKYWLNTRKIEEALEIVKEQI TATARKIERYEARIIQYRQNQLFQSDQRRIFYQSLNQTTDT VTIKPEKTATTKFWKELWENNKNYNKNAGWIKEFEGKFS QNKMELMEITTEMISKRVQKVKNWTSPGSDQLHGFWLK HLISLHGKMAQQFNEMLQKGSISEWLTTGRTYLIQKDPAK GAAPGNYRPITCLPTMFKLLTGIIADRIQDYLEEKNILPDEQ KGNKRKSRGTKDQLLIDKMILENCKSRKANLHMTWIDYK KAFDSLPHSWIIKCLDAIGICKTVGTFIENMMEHWKTELFV GNESYGLVNIRRGIFQGDSLSPLLFIIAMIPLSTILQKTNLGY QISKNSHKISHLMYMDDLKLYGKTETEIQSLTNTVRIFSTD INMEFGLDKCSTVALKKGKIIESEGINMPNGQTIKCHQPEA YKYLGILQLDNIKHEHVKTVVSKEYTQRVRKILKSKLNGG NTIKAINTWAIPVIRYTAGIINWTQVELDNLDRKTRKLMTI HHSLHPRSDVDRLYLPRRSGGRGLLQVKQAVKEEEHALA EYVKQSEEPALIEVKNQKLLKTQQTKNQYKKTALQTRAD SWHNKTLHGKFLDKIEGKADKEKTWLWLTNGTLKKETE GLILAAQEQAIRTNAIKAKIEKSADDPKCRLCKETDETIDHI LSCCKKIAQTDYKQRHNYVAQMIHWNLCLKYHLPAAKN WWDHKPAKVLENEHAKILWDFRIQTDKVLEHNTPDITVV EKNKVWIIDVAIPGDSRIDEKQQEKLSRYQDLKIELQRLW QKPVQVVPVVMGTLGAVPKDLSRHLETIDIDKITICQLQK ATLLGSAHIIRKYTTQS	TAATAATAATAATAA TA	taatataat
R4-1_PH	TAATAAT AATAAT	MKMSHNRDTPSNGVKGTSVRLGTSLVRSPVGEAGAVRER GTHPSESVSQDSDASVNATGEGSVREQAPLSPPGAEEATV PTQRRTRHKWSREDRVVLWECFVASKREGPGYLKRLKQL		

AATAtaata ataat		WDERGIPGNFPQASLSGQIRQICSKNLLSEEERLQIAARME AQVASPSADEPARQVPTRPVTPPRSPPVEPARRPSIPSEETP DLGAVPSEIDSADPNRSPSRGPRHLPAHNMSQSESEDDVT DPDVSDQQRSDSLEPRDLLRNSSVESTPGHPPNQELSDTLLS NYVPSEIDSDDPNQSPRRGPRHLPAHDMSLSDSMDEETEP DLSDQQRSDLLELRDLLRNSSVETTPKGHPSLRHLPEPKIR AAAFRVNSVIGKIHTNNITETNALIKAGADLAVRILEVQPR PQRTQRKKDPPWKHRLEKNIAEIRKHLSWISEWRRGNLH DEEKKTLLESRYRCLEVGLTNLEDTLKQRLSAKRSKVRRF EARVAGFHQNQLFNTNQKRLYQTLRGEETSSDSPNAEESI RFWSDIWSKEVRHNNTAEWLHDVKEKNVAADPDLTITSQ QLKKQLSKTKNWKAPGPDMVQGYWIKTFTSLHSRIAAQL NHCLQRGTVPTWMTTGKTVLIQKDKAKGTEVSNYRPITC LPLMWKVLTGIIYERVYQHLDSKKLLPDEQKGCRRNTRG TKDHLLVDKLLTKDARSKKKNLSMAWVDYKKAFDMVP HSWILECLDIYGIAGNIRNLIATTMPNWKTQLTSANKHLG EVSIKRGIFQGDSLSPLLFVLTMIPLSETLNKAGQGYNYSR TMKLNHLLYMDDLKLYAKSKDQVEQLLNIVHQYSQDIK MQFGVSKCGVLNIERGEVTASEGITIEEGTIKDIEEAGYKY LGVMEYNTILHRTMKDSIRKEYLTRLRLILKSHLNGGNTI KAINTWAVPVVRYSAGIINWTKKDCTDMDIKTRKLMTIY RALHPRSCVDRLYINRREGGRGLISVEDCVEAEKRALSQH FRESDDPWARCLVEAKLLKETETADQFKERRRLDRTNKW KSMKMSGQYLEAVQDKIVPDSWNWLLRGELKRETEGTIL AAQEQALRTRYIQNKIDKRNVPSTCRICRSSDETINHVISEC GVLAQKEYKRRHDKVARHLHWTLLRIHNFPVSERWYEH EPAPVVENEAVQIYWDKRMETDRVLHANRPDIVVKDKQE KSAKLIDISIPFDSRIVDKEAEKKEKYRDLAIELQRLWQMK VDVVPVVIGALGAMSKNLKTALRELKCGHLHPGTLQKSA LLGTAHIIRKVL	ATGCGGGAATCAGGGA TGACCATGGATTCTTT GGCTGTGGGTAGAGT CCTTTCTAACCCAGC GAAGGATAAGGGTGC GGTACCGTAATGGAA CGACGCGGG	
Tad1-65B_BG	ATGCGG AATCAG GGATGA CCATGG ATTCTTT GGCTGT GGGTAG AGTCCTT TCTAACC CAGCGA	-		tcaaatcaaatcaaatcaaatcaaatcaa atcaaatcaaatcaaatcaaatcaaatca aatcaaatcaaatcaaatcaaatcaaatc aaatcaaatcaaa

	AGGATA AGGGTG CGGTAC CGTAAT GGAACG ACGCCG Gtcaaatcaa atcaaatcaaa tcaaatcaaat caaatcaaatc aaatcaaatca aatcaaatcaa atcaaatcaaa tcaaatcaaat caaatcaaatc aaa			
L2-24_Acar	CTAACCT CCGCAG GGGTGG TGGAataa a	MPIKMVRPRRLMDPNGFLMALGDVPASSVGDSVDALVTL YNGEMARAIDTIAPERPLAMRRAKLAPWFSGELAAMKRE TRRLERVWRKRRSESNLERVKAQRKTYCKAMAAVKKTY YSSRIEAAKNRPAELFRTVRGLLNNTPQTAPTDDVAEKCE AFSQFFSKKIALIRSELDATVDAISEDVARAPACPILMDSFQ SVQPEDVDKVLGEARATTCILDPCPSWLVKEARGGLVEW VRVVVNASLREGRLPAGLKQAVIKPLLKKPSLDPTNWNN FRPVSNLPFLGKVVERVVASQLQGFLEATDYLDPAQSGFR PGHGTETALVSLVDDLRRELDRGSVSLLVLLDLSAAFDTV DHGILLGRLVGMGLGGTALQWLQSFLEDRTQKVLIGDSC SAPQPLSCGVPQGSVLSPLLFNIYMKPLGEIIRSFGVRCHLY ADDVQLYHSFSPASKEAVQILNRCLAAVSDWMRANKLKL NPDKTEVLLVSRKAEQGIGLQPVLDGVALPLKAQVRSLG VLLDSSLSLEPQVSAVARGAFAQLKLVRQLRPYLGKPDM ATVVHALVTSRLDYCNALYVGLPSKTVRKLELVQRSAAR LLTGAPYREHTTPLLKQLHWLPVSFRAQFKVLALAYKTL NGSGPTYLSERISPYEPPRRLRSSSEALLSVPPPSQVHLAGT RDRAFSVVAPRLWNSLPREIREAPSLMSFRKQLKTWMWD QAFGPS	CTAACCTCCGCAGGG GTGGTGGA	ataaa
			-	-
LINE1-2_ZM	CTGAGG CCTTATT ATTGTCT CCTCCGG TTGAGG	MNLSNILVWNARGLNNKARRDYVRDTILSSKADIICLQET KVASFSSHLLLSVCGSDFDKFLTLPAVGTRGGILIAWKGA VCQVISSRIDNFSVSVQFAGMDGMNWWFTGVYGPQEDD RKIQFLQELLDFRTLCSGPWLLSGDFNMIYQAADKNNANL NRALMGRFRRFLDDCLLKEIPLHGRKFTWSNERSSPTLVR		

	CTGTCAT	LDRVFCCNDWEDFFPNSLLQSASSGVSDHCPLILSLNVQT			
	GTCTACT	HKKRRFHFESYWPKLPGFLEAVSHSWNAPVSIACPVEGIF			
	ACTGCTA	LKLQRLSKGLQKWSHHKVGNVKLQLATAKEILHRLEIAR			
	TGTTGGG	DNRVLSPGEEWLRKKLKLKLHCLGLASLERTIARLRSRILYL			
	GGCCAT	QEGDANTSFFHQQARYRKKKNFIAKFQVEERIVTSQEEKQ			
	GGCTCC	QAALDFYDNLLGTAEHRDYTLDLHSLGIQQHDLGDLELV			
	GGGCCC	FSMEEVWSVVKDLPLDKAPGPDGFTGRFYKVCWDIIKED			
	TGAGCA	MMGALLAVQCGHVSKLKLLNTTFITLLPKKVDALLVKDY			
	TATCCAG	RPISLVHSFAKLVTKVLAARLAPHLPSMVSINQSAFIKGRSI			
	GGTCCCCT	QDNFLLVQQLTRSLHHSKEPHILLKLDISKAFDSISWPFLL			
	Taagtg	DMLQHLGFGRSWCNLISLLLCTSSTRILVNGEPGEYILHHR			
		GLRQGDPLSPMLFILVMEALNAMVSYAFREHMLQPIANQ			
		HAKHCISFYADDVVLFMRPSHQDISAIISILDIFGHATGLKT			
		NISKSSVTPIQCGANELTTISALLPCATKNFPCTYLGIPLSIR			
		RLANSDLLPLVDKVADKLPGWKANLLSRAGRLVMVKTV			
		LSSVPIYLMLALELPKWVFKAIDKRRRGFLWKGQGDANG			
		GNCLVSWEAVQRPLKYGGLGILNLEMFGWALRARWLWL			
		QKTDASRPWAGLPIRVHRNAKALVDVAITSVVGSGESVK			
		FWSDRWLLGKTVAEHCPTLIQVISRRALKRRTVAEGLTNR			
		QWVSDIRGGLSVTVLVEYLQLWNLVDGVVLQPDIADQHI			
		WRLSAHGTYCSKSAYDALFVGSIPFGPWRRVWKTWAPLR			
		CKFFVWLAIKNRVWTADRLAKRGLPHPVACPLCDQAEET			
		IQHILVSCVFARQIWTSMLHNLGLLAIVPQHGCTRFSNWW			
		CQSIKKVEKSLRKGLNSLIILVAWEIWKHRNACVFEGVVP			
		CTQRVQSAVIEEGNVWCLAGASALQDLLLRQLPRGP			

Gene Writers, e.g. Thermostable Gene Writers

While not wishing to be bound by theory, in some embodments, retrotransposases that evolved in cold environments may not function as well at human body temperature. This application provides a number of thermostable Gene Writers, including proteins derived from avian retrotransposases.

Thermostability may be measured, e.g., by testing the ability of a Gene Writer to polymerize DNA *in vitro* at a high temperature (e.g., 37°C) and a low temberature (e.g., 25°C). Suitable conditions for assaying *in vitro* DNA polymerization activity (e.g., processivity) are described, e.g., in Bibillo and Eickbush, "High Processivity of the Reverse Transcriptase from a Non-long Terminal Repeat Retrotransposon" (2002) JBC 277, 34836-34845. In some embodiments, the thermostable Gene Writer polypeptide has an activity, e.g., a DNA polymerization activity, at 37°C that is no less than 70%, 75%, 80%, 85%, 90%, or 95% of its activity at 25°C under otherwise similar conditions.

In some embodiments, a GeneWriter polypeptide (e.g., a sequence of Table 1, 2, 3A, or 3B or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto) is stable in a subject chosen from a mammal (e.g., human) or a bird. In some embodiments, a GeneWriter polypeptide described herein is functional at 37°C. In some embodiments, a GeneWriter polypeptide described herein has greater activity at 37°C than it does at a lower temperature, e.g., at 30°C, 25°C, or 20°C. In some embodiments, a GeneWriter polypeptide described herein has greater activity in a human cell than in a zebrafish cell.

In some embodiments, a GeneWriter polypeptide is active in a human cell cultured at 37°C, e.g., using an assay of Example 6 or Example 7 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety.

In some embodiments, the assay comprises steps of: (1) introducing HEK293T cells into one or more wells of 6.4 mm diameter, at 10,000 cells/well, (2) incubating the cells at 37°C for 24 hr, (3) providing a transfection mixture comprising 0.5µl if FuGENE® HD transfection reagent and 80ng DNA (wherein the DNA is a plasmid comprising, in order, (a) CMV promoter, (b) 100 bp of sequence homologous to the 100 bp upstream of the target site, (c) sequence encoding a 5' untranslated region that binds the GeneWriter protein, (d) sequence encoding the GeneWriter protein, (e) sequence encoding a 3' untranslated region that binds the GeneWriter protein (f) 100 bp of sequence homologous to the 100 bp downstream of the target site, and (g)

BGH polyadenylation sequence) and 10µl Opti-MEM and incubating for 15 min at room temperature, (4) adding the transfection mixture to the cells, (5) incubating the cells for 3 days, and (6) assaying integration of the exogenous sequence into a target locus (e.g., rDNA) in the cell genome, e.g., wherein one or more of the preceding steps are performed as described in Example 6 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety.

In some embodiments, a Gene Writer system is capable of producing a substitution into the target site of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 or more nucleotides. In some embodiments, the substitution is a transition mutation. In some embodiments, the substitution is a transversion mutation. In some embodiments, the substitution converts an adenine to a thymine, an adenine to a guanine, an adenine to a cytosine, a guanine to a thymine, a guanine to a cytosine, a guanine to an adenine, a thymine to a cytosine, a thymine to an adenine, a thymine to a guanine, a cytosine to an adenine, a cytosine to a guanine, or a cytosine to a thymine.

In some embodiments, a Gene Writer system is capable of producing an insertion into the target site of at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides (and optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a Gene Writer system is capable of producing an insertion into the target site of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides (and optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a Gene Writer system is capable of producing an insertion into the target site of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some

embodiments, a Gene Writer system is capable of producing a deletion of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases).

In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a gene. In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a gene by altering, adding, or deleting sequences in a promoter or enhancer, e.g. sequences that bind transcription factors. In some embodiments, an insertion, deletion, substitution, or combination thereof alters translation of a gene (e.g. alters an amino acid sequence), inserts or deletes a start or stop codon, alters or fixes the translation frame of a gene. In some embodiments, an insertion, deletion, substitution, or combination thereof alters splicing of a gene, e.g. by inserting, deleting, or altering a splice acceptor or donor site. In some embodiments, an insertion, deletion, substitution, or combination thereof alters transcript or protein half-life. In some embodiments, an insertion, deletion, substitution, or combination thereof alters protein localization in the cell (e.g. from the cytoplasm to a mitochondria, from the cytoplasm into the extracellular space (e.g. adds a secretion tag)). In some embodiments, an insertion, deletion, substitution, or combination thereof alters (e.g. improves) protein folding (e.g. to prevent accumulation of misfolded proteins). In some embodiments, an insertion, deletion, substitution, or combination thereof, alters, increases, decreases the activity of a gene, e.g. a protein encoded by the gene.

In some embodiments, the GeneWriter polypeptide results in insertion of the heterologous object sequence (e.g., the GFP gene) into the target locus (e.g., rDNA) at an average copy number of at least 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.5, 3, 4, or 5 copies per genome. In some embodiments, a cell described herein (e.g., a cell comprising a heterologous sequence at a target insertion site) comprises the heterologous object sequence at an average copy number of at least 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.5, 3, 4, or 5 copies per genome.

In some embodiments, a GeneWriter causes integration of a sequence in a target RNA with relatively few truncation events at the terminus. For instance, in some embodiments, a Gene Writer protein (e.g., of SEQ ID NO: 1016) results in about 25-100%, 50-100%, 60-100%, 70-100%, 75-95%, 80%-90%, or 86.17% of integrants into the target site being non-truncated, as

measured by an assay described herein, e.g., an assay of Example 6 and Figure 8 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety. In some embodiments, a Gene Writer protein (e.g., of SEQ ID NO: 1016) results in at least about 30%, 40%, 50%, 60%, 70%, 80%, or 90% of integrants into the target site being non-truncated, as measured by an assay described herein. In some embodiments, an integrant is classified as truncated versus non-truncated using an assay comprising amplification with a forward primer situated 565bp from the end of the element (e.g., a wild-type retrotransposon sequence) and a reverse primer situated in the genomic DNA of the target insertion site, e.g., rDNA. In some embodiments, the number of full-length integrants in the target insertion site is greater than the number of integrants truncated by 300-565 nucleotides in the target insertion site, e.g., the number of full-length integrants is at least 1.1x, 1.2x, 1.5x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, or 10x the number of the truncated integrants, or the number of full-length integrants is at least 1.1x-10x, 2x-10x, 3x-10x, or 5x-10x the number of the truncated integrants.

In some embodiments, a system or method described herein results in insertion of the heterologous object sequence only at one target site in the genome of the target cell. Insertion can be measured, e.g., using a threshold of above 1%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, e.g., as described in Example 8 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety. In some embodiments, a system or method described herein results in insertion of the heterologous object sequence wherein less than 1%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, 10%, 20%, 30%, 40%, or 50% of insertions are at a site other than the target site, e.g., using an assay described herein, e.g., an assay of Example 8 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety.

In some embodiments, a system or method described herein results in "scarless" insertion of the heterologous object sequence, while in some embodiments, the target site can show deletions or duplications of endogenous DNA as a result of insertion of the heterologous sequence. The mechanisms of different retrotransposons could result in different patterns of duplications or deletions in the host genome occurring during retrotransposition at the target site. In some embodiments, the system results in a scarless insertion, with no duplications or deletions in the surrounding genomic DNA. In some embodiments, the system results in a deletion of less than 1, 2, 3, 4, 5, 10, 50, or 100 bp of genomic DNA upstream of the insertion. In some embodiments, the system results in a deletion of less than 1, 2, 3, 4, 5, 10, 50, or 100 bp of

genomic DNA downstream of the insertion. In some embodiments, the system results in a duplication of less than 1, 2, 3, 4, 5, 10, 50, or 100 bp of genomic DNA upstream of the insertion. In some embodiments, the system results in a duplication of less than 1, 2, 3, 4, 5, 10, 50, or 100 bp of genomic DNA downstream of the insertion.

In some embodiments, a GeneWriter described herein, or a DNA-binding domain thereof, binds to its target site specifically, e.g., as measured using an assay of Example 21 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety. In some embodiments, the GeneWriter or DNA-binding domain thereof binds to its target site more strongly than to any other binding site in the human genome. For example, in some embodiments, in an assay of Example 21 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety, the target site represents more than 50%, 60%, 70%, 80%, 90%, or 95% of binding events of the GeneWriter or DNA-binding domain thereof to human genomic DNA. In some embodiments, the DNA binding domain of the GeneWriter is heterologous to the remainder of the GeneWriter, e.g., such that the GeneWriter targets a different target site that the endogenous DNA binding domain associated with the remainder of the GeneWriter.

Genetically engineered, e.g., dimerized GeneWriters

Some non-LTR retrotransposons utilize two subunits to complete retrotransposition (Christensen et al PNAS 2006). In some embodiments, a retrotransposase described herein comprises two connected subunits as a single polypeptide. For instance, two wild-type retrotransposases could be joined with a linker to form a covalently "dimerized" protein. In some embodiments, the nucleic acid coding for the retrotransposase codes for two retrotransposase subunits to be expressed as a single polypeptide. In some embodiments, the subunits are connected by a peptide linker, such as has been described herein in the section entitled "Linker" and, e.g., in Chen et al Adv Drug Deliv Rev 2013. In some embodiments, the two subunits in the polypeptide are connected by a rigid linker. In some embodiments, the rigid linker consists of the motif $(EAAAK)_n$ (SEQ ID NO: 1534). In other embodiments, the two subunits in the polypeptide are connected by a flexible linker. In some embodiments, the flexible linker consists of the motif $(Gly)_n$. In some embodiments, the flexible linker consists of the motif $(GGGGS)_n$ (SEQ ID NO: 1535). In some embodiments, the rigid or flexible linker consists of 1,

2, 3, 4, 5, 10, 15, or more amino acids in length to enable retrotransposition. In some embodiments, the linker consists of a combination of rigid and flexible linker motifs.

Based on mechanism, not all functions are required from both retrotransposase subunits. In some embodiments, the fusion protein may consist of a fully functional subunit and a second subunit lacking one or more functional domains. In some embodiments, one subunit may lack reverse transcriptase functionality. In some embodiments, one subunit may lack the reverse transcriptase domain. In some embodiments, one subunit may possess only endonuclease activity. In some embodiments, one subunit may possess only an endonuclease domain. In some embodiments, the two subunits comprising the single polypeptide may provide complimentary functions.

In some embodiments, one subunit may lack endonuclease functionality. In some embodiments, one subunit may lack the endonuclease domain. In some embodiments, one subunit may possess only reverse transcriptase activity. In some embodiments, one subunit may possess only a reverse transcriptase domain. In some embodiments, one subunit may possess only DNA-dependent DNA synthesis functionality.

Linkers:

In some embodiments, domains of the compositions and systems described herein (e.g., the endonuclease and reverse transcriptase domains of a polypeptide or the DNA binding domain and reverse transcriptase domains of a polypeptide) may be joined by a linker. A composition described herein comprising a linker element has the general form S1-L-S2, wherein S1 and S2 may be the same or different and represent two domain moieties (e.g., each a polypeptide or nucleic acid domain) associated with one another by the linker. In some embodiments, a linker may connect two polypeptides. In some embodiments, a linker may connect two nucleic acid molecules. In some embodiments, a linker may connect a polypeptide and a nucleic acid molecule. A linker may be a chemical bond, e.g., one or more covalent bonds or non-covalent bonds. A linker may be flexible, rigid, and/or cleavable. In some embodiments, the linker is a peptide linker. Generally, a peptide linker is at least 2, 3, 4, 5, 6, 7, 8, 9, 10 or more amino acids in length, e.g., 2-50 amino acids in length, 2-30 amino acids in length.

The most commonly used flexible linkers have sequences consisting primarily of stretches of Gly and Ser residues ("GS" linker). Flexible linkers may be useful for joining

domains that require a certain degree of movement or interaction and may include small, non-polar (e.g. Gly) or polar (e.g. Ser or Thr) amino acids. Incorporation of Ser or Thr can also maintain the stability of the linker in aqueous solutions by forming hydrogen bonds with the water molecules, and therefore reduce unfavorable interactions between the linker and the other moieties. Examples of such linkers include those having the structure $[GGS]^{\geq 1}$ or $[GGGS]^{\geq 1}$ (SEQ ID NO: 1536). Rigid linkers are useful to keep a fixed distance between domains and to maintain their independent functions. Rigid linkers may also be useful when a spatial separation of the domains is critical to preserve the stability or bioactivity of one or more components in the agent. Rigid linkers may have an alpha helix-structure or Pro-rich sequence, (XP)n, with X designating any amino acid, preferably Ala, Lys, or Glu. Cleavable linkers may release free functional domains in vivo. In some embodiments, linkers may be cleaved under specific conditions, such as the presence of reducing reagents or proteases. In vivo cleavable linkers may utilize the reversible nature of a disulfide bond. One example includes a thrombin-sensitive sequence (e.g., PRS) between the two Cys residues. In vitro thrombin treatment of CPRSC (SEQ ID NO: 1537) results in the cleavage of the thrombin-sensitive sequence, while the reversible disulfide linkage remains intact. Such linkers are known and described, e.g., in Chen et al. 2013. Fusion Protein Linkers: Property, Design and Functionality. Adv Drug Deliv Rev. 65(10): 1357–1369. In vivo cleavage of linkers in compositions described herein may also be carried out by proteases that are expressed in vivo under pathological conditions (e.g. cancer or inflammation), in specific cells or tissues, or constrained within certain cellular compartments. The specificity of many proteases offers slower cleavage of the linker in constrained compartments.

In some embodiments the amino acid linkers are (or are homologous to) the endogenous amino acids that exist between such domains in a native polypeptide. In some embodiments the endogenous amino acids that exist between such domains are substituted but the length is unchanged from the natural length. In some embodiments, additional amino acid residues are added to the naturally existing amino acid residues between domains.

In some embodiments, the amino acid linkers are designed computationally or screened to maximize protein function (Anad et al., FEBS Letters, 587:19, 2013).

In addition to being fully encoded on a single transcript, a polypeptide can be generated by separately expressing two or more polypeptide fragments that reconstitute the holoenzyme. In

some embodiments, the Gene Writer polypeptide is generated by expressing as separate subunits that reassemble the holoenzyme through engineered protein-protein interactions. In some embodiments, reconstitution of the holoenzyme does not involve covalent binding between subunits. Peptides may also fuse together through *trans*-splicing of inteins (Tornabene et al. Sci Transl Med 11, eaav4523 (2019)). In some embodiments, the Gene Writer holoenzyme is expressed as separate subunits that are designed to create a fusion protein through the presence of split inteins (e.g., as described herein) in the subunits. In some embodiments, the Gene Writer holoenzyme is reconstituted through the formation of covalent linkages between subunits. In some embodiments, protein subunits reassemble through engineered protein-protein binding partners, e.g., SpyTag and SpyCatcher (Zakeri et al. PNAS 109, E690-E697 (2012)). In some embodiments, an additional domain described herein, e.g., a Cas9 nickase, is expressed as a separate polypeptide that associates with the Gene Writer polypeptide through covalent or non-covalent interactions as described above. In some embodiments, the breaking up of a Gene Writer polypeptide into subunits may aid in delivery of the protein by keeping the nucleic acid encoding each part within optimal packaging limits of a viral delivery vector, e.g., AAV (Tornabene et al. Sci Transl Med 11, eaav4523 (2019)). In some embodiments, the Gene Writer polypeptide is designed to be dimerized through the use of covalent or non-covalent interactions as described above.

Inteins

In some embodiments, the Gene Writer system comprises an intein. Generally, an intein comprises a polypeptide that has the capacity to join two polypeptides or polypepide fragments together via a peptide bond. In some embodiments, the intein is a trans-splicing intein that can join two polypeptide fragments, e.g., to form the polypeptide component of a system as described herein. In some embodiments, an intein may be encoded on the same nucleic acid molecule encoding the two polypeptide fragments. In certain embodiments, the intein may be translated as part of a larger polypeptide comprising, e.g., in order, the first polypeptide fragment, the intein, and the second polypeptide fragment. In embodiments, the translated intein may be capable of excising itself from the larger polypeptide, e.g., resulting in separation of the attached polypeptide fragments. In embodiments, the excised intein may be capable of joining

the two polypeptide fragments to each other directly via a peptide bond. Exemplary inteins are described herein, e.g., in Table X.

Template RNA component of Gene Writer™ gene editor system

The Gene Writer systems described herein can transcribe an RNA sequence template into host target DNA sites by target-primed reverse transcription. By writing DNA sequence(s) via reverse transcription of the RNA sequence template directly into the host genome, the Gene Writer system can insert an object sequence into a target genome without the need for exogenous DNA sequences to be introduced into the host cell (unlike, for example, CRISPR systems), as well as eliminate an exogenous DNA insertion step. Therefore, the Gene Writer system provides a platform for the use of customized RNA sequence templates containing object sequences, e.g., sequences comprising heterologous gene coding and/or function information.

In some embodiments the template RNA encodes a Gene Writer protein in *cis* with a heterologous object sequence. Various cis constructs were described, for example, in Kuroki-Kami et al (2019) *Mobile DNA* 10:23 (incorporated by reference herein in its entirety), and can be used in combination with any of the embodiments described herein. For instance, in some embodiments, the template RNA comprises a heterologous object sequence, a sequence encoding a Gene Writer protein (e.g., a protein comprising (i) a reverse transcriptase domain and (ii) an endonuclease domain, e.g., as described herein), a 5' untranslated region, and a 3' untranslated region. The components may be included in various orders. In some embodiments, the Gene Writer protein and heterologous object sequence are encoded in different directions (sense vs. anti-sense), e.g., using an arrangement shown in Figure 3A of Kuroki-Kami et al, *Id.* In some embodiments the Gene Writer protein and heterologous object sequence are encoded in the same direction. In some embodiments, the nucleic acid encoding the polypeptide and the template RNA or the nucleic acid encoding the template RNA are covalently linked, e.g., are part of a fusion nucleic acid and/or are part of the same transcript. In some embodiments, the fusion nucleic acid comprises RNA or DNA.

The nucleic acid encoding the Gene Writer polypeptide may, in some instances, be 5' of the heterologous object sequence. For example, in some embodiments, the template RNA comprises, from 5' to 3', a 5' untranslated region, a sense-encoded Gene Writer polypeptide, a sense-encoded heterologous object sequence, and 3' untranslated region. In some embodiments,

the template RNA comprises, from 5' to 3', a 5' untranslated region, a sense-encoded Gene Writer polypeptide, anti-sense-encoded heterologous object sequence, and 3' untranslated region.

In some embodiments, the RNA further comprises homology to the DNA target site.

It is understood that, when a template RNA is described as comprising an open reading frame or the reverse complement thereof, in some embodiments the template RNA must be converted into double stranded DNA (e.g., through reverse transcription) before the open reading frame can be transcribed and translated.

In certain embodiments, customized RNA sequence template can be identified, designed, engineered and constructed to contain sequences altering or specifying host genome function, for example by introducing a heterologous coding region into a genome; affecting or causing exon structure/alternative splicing; causing disruption of an endogenous gene; causing transcriptional activation of an endogenous gene; causing epigenetic regulation of an endogenous DNA; causing up- or down-regulation of operably liked genes, etc. In certain embodiments, a customized RNA sequence template can be engineered to contain sequences coding for exons and/or transgenes, provide for binding sites to transcription factor activators, repressors, enhancers, etc., and combinations of thereof. In other embodiments, the coding sequence can be further customized with splice acceptor sites, poly-A tails. In certain embodiments the RNA sequence can contain sequences coding for an RNA sequence template homologous to the retrotransposase, be engineered to contain heterologous coding sequences, or combinations thereof.

The template RNA may have some homology to the target DNA. In some embodiments the template RNA has at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 175, 200 or more bases of exact homology to the target DNA at the 3' end of the RNA. In some embodiments the template RNA has at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 175, 180, or 200 or more bases of at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% homology to the target DNA, e.g., at the 5' end of the template RNA. In some embodiments the template RNA has a 3' untranslated region derived from a retrotransposon, e.g. a retrotransposons described herein. In some embodiments the template RNA has a 3' region of at least 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200 or more bases of at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% homology to the 3' sequence of a retrotransposon, e.g., a retrotransposon described herein, e.g. a

retrotransposon in Table X, Z1, Z2, 3A, or 3B. In some embodiments the template RNA has a 5' untranslated region derived from a retrotransposon, e.g. a retrotransposons described herein. In some embodiments the template RNA has a 5' region of at least 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, or 200 or more bases of at least 40%, 50%, 60%, 70%, 80%, 90%, 95% or greater homology to the 5' sequence of a retrotransposon, e.g., a retrotransposon described herein, e.g. a retrotransposon described in Table X, Z1, Z2, 3A, or 3B.

The template RNA component of a Gene Writer genome editing system described herein typically is able to bind the Gene Writer genome editing protein of the system. In some embodiments the template RNA has a 3' region that is capable of binding a Gene Writer genome editing protein. The binding region, e.g., 3' region, may be a structured RNA region, e.g., having at least 1, 2 or 3 hairpin loops, capable of binding the Gene Writer genome editing protein of the system.

The template RNA component of a Gene Writer genome editing system described herein typically is able to bind the Gene Writer genome editing protein of the system. In some embodiments the template RNA has a 5' region that is capable of binding a Gene Writer genome editing protein. The binding region, e.g., 5' region, may be a structured RNA region, e.g., having at least 1, 2 or 3 hairpin loops, capable of binding the Gene Writer genome editing protein of the system. In some embodiments, the 5' untranslated region comprises a pseudoknot, e.g., a pseudoknot that is capable of binding to the Gene Writer protein.

In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 5' untranslated region) comprises a stem-loop sequence. In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 5' untranslated region) comprises a hairpin. In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 5' untranslated region) comprises a helix. In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 5' untranslated region) comprises a psuedoknot. In some embodiments the template RNA comprises a ribozyme. In some embodiments the ribozyme is similar to an hepatitis delta virus (HDV) ribozyme, e.g., has a secondary structure like that of the HDV ribozyme and/or has one or more activities of the HDV ribozyme, e.g., a self-cleavage activity. See, e.g., Eickbush et al., Molecular and Cellular Biology, 2010, 3142-3150.

In some embodiments, the template RNA (e.g., an untranslated region of the hairpin RNA, e.g., a 3' untranslated region) comprises one or more stem-loops or helices. Exemplary structures of R2 3' UTRs are shown, for example, in Ruschak et al. "Secondary structure models of the 3′ untranslated regions of diverse R2 RNAs" RNA. 2004 Jun; 10(6): 978–987, e.g., at Figure 3, therein, and in Eikbush and Eikbush, "R2 and R2/R1 hybrid non-autonomous retrotransposons derived by internal deletions of full-length elements" Mobile DNA (2012) 3:10; e.g., at Figure 3 therein, which articles are hereby incorporated by reference in their entirety.

In some embodiments, a template RNA described herein comprises a sequence that is capable of binding to a GeneWriter protein described herein. For instance, in some embodiments, the template RNA comprises an MS2 RNA sequence capable of binding to an MS2 coat protein sequence in the GeneWriter protein. In some embodiments, the template RNA comprises an RNA sequence capable of binding to a B-box sequence. In some embodiments, the template RNA comprises an RNA sequence (e.g., a crRNA sequence and/or tracrRNA sequence, a gRNA sequence) capable of binding to a Cas9 domain in the GeneWriter protein. In some embodiments, in addition to or in place of a UTR, the template RNA is linked (e.g., covalently) to a non-RNA UTR, e.g., a protein or small molecule.

In some embodiments the template RNA has a poly-A tail at the 3' end. In some embodiments the template RNA does not have a poly-A tail at the 3' end.

In some embodiments the template RNA has a 5' region of at least 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200 or more bases of at least 40%, 50%, 60%, 70%, 80%, 90%, 95% or greater homology to the 5' sequence of a retrotransposon, e.g., a retrotransposon described herein.

The template RNA of the system typically comprises an object sequence for insertion into a target DNA. The object sequence may be coding or non-coding.

In some embodiments a system or method described herein comprises a single template RNA. In some embodiments a system or method described herein comprises a plurality of template RNAs.

In some embodiments, the object sequence may contain an open reading frame. In some embodiments, the template RNA has a Kozak sequence. In some embodiments, the template RNA has an internal ribosome entry site. In some embodiments, the template RNA has a self-cleaving peptide such as a T2A or P2A site. In some embodiments, the template RNA has a start

codon. In some embodiments, the template RNA has a splice acceptor site. In some embodiments, the template RNA has a splice donor site. Exemplary splice acceptor and splice donor sites are described in WO2016044416, incorporated herein by reference in its entirety. Exemplary splice acceptor site sequences are known to those of skill in the art and include, by way of example only, CTGACCCTTCTCTCTCTCCCCCAGAG (from human HBB gene) and TTTCTCTCCCACAAG (from human immunoglobulin-gamma gene). In some embodiments the template RNA, has a microRNA binding site downstream of the stop codon. In some embodiments, the template RNA has a polyA tail downstream of the stop codon of an open reading frame. In some embodiments, the template RNA comprises one or more exons. In some embodiments, the template RNA comprises one or more introns. In some embodiments, the template RNA comprises a eukaryotic transcriptional terminator. In some embodiments, the template RNA comprises an enhanced translation element or a translation enhancing element. In some embodiments, the RNA comprises the human T-cell leukemia virus (HTLV-1) R region. In some embodiments, the RNA comprises a posttranscriptional regulatory element that enhances nuclear export, such as that of Hepatitis B Virus (HPRE) or Woodchuck Hepatitis Virus (WPRE). In some embodiments, in the template RNA, the heterologous object sequence encodes a polypeptide and is coded in an antisense direction with respect to the 5' and 3' UTR. In some embodiments, in the template RNA, the heterologous object sequence encodes a polypeptide and is coded in a sense direction with respect to the 5' and 3' UTR.

In some embodiments, a nucleic acid described herein (e.g., a template RNA or a DNA encoding a template RNA) comprises a microRNA binding site. In some embodiments, the microRNA binding site is used to increase the target-cell specificity of a GeneWriter system. For instance, the microRNA binding site can be chosen on the basis that is is recognized by a miRNA that is present in a non-target cell type, but that is not present (or is present at a reduced level relative to the non-target cell) in a target cell type. Thus, when the template RNA is present in a non-target cell, it would be bound by the miRNA, and when the template RNA is present in a target cell, it would not be bound by the miRNA (or bound but at reduced levels relative to the non-target cell). While not wishing to be bound by theory, binding of the miRNA to the template RNA may interfere with insertion of the heterologous object sequence into the genome. Accordingly, the heterologous object sequence would be inserted into the genome of target cells more efficiently than into the genome of non-target cells. A system having a microRNA binding

site in the template RNA (or DNA encoding it) may also be used in combination with a nucleic acid encoding a GeneWriter polypeptide, wherein expression of the GeneWriter polypeptide is regulated by a second microRNA binding site, e.g., as described herein, e.g., in the section entitled "Polypeptide component of Gene Writer gene editor system".

In some embodiments, the object sequence may contain a non-coding sequence. For example, the template RNA may comprise a promoter or enhancer sequence. In some embodiments, the template RNA comprises a tissue specific promoter or enhancer, each of which may be unidirectional or bidirectional. In some embodiments, the promoter is an RNA polymerase I promoter, RNA polymerase II promoter, or RNA polymerase III promoter. In some embodiments, the promoter comprises a TATA element. In some embodiments, the promoter comprises a B recognition element. In some embodiments, the promoter has one or more binding sites for transcription factors. In some embodiments, the non-coding sequence is transcribed in an antisense-direction with respect to the 5' and 3' UTR. In some embodiments, the non-coding sequence is transcribed in a sense direction with respect to the 5' and 3' UTR.

In some embodiments, a nucleic acid described herein (e.g., a template RNA or a DNA encoding a template RNA) comprises a promoter sequence, e.g., a tissue specific promoter sequence. In some embodiments, the tissue-specific promoter is used to increase the target-cell specificity of a GeneWriter system. For instance, the promoter can be chosen on the basis that it is active in a target cell type but not active in (or active at a lower level in) a non-target cell type. Thus, even if the promoter integrated into the genome of a non-target cell, it would not drive expression (or only drive low level expression) of an integrated gene. A system having a tissue-specific promoter sequence in the template RNA may also be used in combination with a microRNA binding site, e.g., in the template RNA or a nucleic acid encoding a GeneWriter protein, e.g., as described herein. A system having a tissue-specific promoter sequence in the template RNA may also be used in combination with a DNA encoding a GeneWriter polypeptide, driven by a tissue-specific promoter, e.g., to achieve higher levels of GeneWriter protein in target cells than in non-target cells.

In some embodiments, a heterologous object sequence comprised by a template RNA (or DNA encoding the template RNA) is operably linked to at least one regulatory sequence. In some embodiments, the heterologous object sequence is operably linked to a tissue-specific promoter, such that expression of the heterologous object sequence, e.g., a therapeutic protein, is

upregulated in target cells, as above. In some embodiments, the heterologous object sequence is operably linked to a miRNA binding site, such that expression of the heterologous object sequence, e.g., a therapeutic protein, is downregulated in cells with higher levels of the corresponding miRNA, e.g., non-target cells, as above.

In some embodiments, the template RNA comprises a microRNA sequence, a siRNA sequence, a guide RNA sequence, a piwi RNA sequence.

In some embodiments, the template RNA comprises a non-coding heterologous object sequence, e.g., a regulatory sequence. In some embodiments, integration of the heterologous object sequence thus alters the expression of an endogenous gene. In some embodiments, integration of the heterologous object sequence upregulates expression of an endogenous gene. In some embodiments, integration of the heterologous object sequence downregulated expression of an endogenous gene.

In some embodiments, the template RNA comprises a site that coordinates epigenetic modification. In some embodiments, the template RNA comprises an element that inhibits, e.g., prevents, epigenetic silencing. In some embodiments, the template RNA comprises a chromatin insulator. For example, the template RNA comprises a CTCF site or a site targeted for DNA methylation.

In order to promote higher level or more stable gene expression, the template RNA may include features that prevent or inhibit gene silencing. In some embodiments, these features prevent or inhibit DNA methylation. In some embodiments, these features promote DNA demethylation. In some embodiments, these features prevent or inhibit histone deacetylation. In some embodiments, these features prevent or inhibit histone methylation. In some embodiments, these features promote histone acetylation. In some embodiments, these features promote histone demethylation. In some embodiments, multiple features may be incorporated into the template RNA to promote one or more of these modifications. CpG dinculeotides are subject to methylation by host methyl transferases. In some embodiments, the template RNA is depleted of CpG dinucleotides, e.g., does not comprise CpG nucleotides or comprises a reduced number of CpG dinucleotides compared to a corresponding unaltered sequence. In some embodiments, the promoter driving transgene expression from integrated DNA is depleted of CpG dinucleotides.

In some embodiments, the template RNA comprises a gene expression unit composed of at least one regulatory region operably linked to an effector sequence. The effector sequence

may be a sequence that is transcribed into RNA (e.g., a coding sequence or a non-coding sequence such as a sequence encoding a micro RNA).

In some embodiments, the object sequence of the template RNA is inserted into a target genome in an endogenous intron. In some embodiments, the object sequence of the template RNA is inserted into a target genome and thereby acts as a new exon. In some embodiments, the insertion of the object sequence into the target genome results in replacement of a natural exon or the skipping of a natural exon.

In some embodiments, the object sequence of the template RNA is inserted into the target genome in a genomic safe harbor site, such as AAVS1, CCR5, or ROSA26. In some embodiments, the object sequence of the template RNA is inserted into the albumin locus. In some embodiments, the object sequence of the template RNA is inserted into the TRAC locus. In some embodiments, the object sequence of the template RNA is added to the genome in an intergenic or intragenic region. In some embodiments, the object sequence of the template RNA is added to the genome 5' or 3' within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb of an endogenous active gene. In some embodiments, the object sequence of the template RNA is added to the genome 5' or 3' within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb of an endogenous promoter or enhancer. In some embodiments, the object sequence of the template RNA can be, e.g., 50-50,000 base pairs (e.g., between 50-40,000 bp, between 500-30,000 bp between 500-20,000 bp, between 100-15,000 bp, between 500-10,000 bp, between 50-10,000 bp, between 50-5,000 bp. In some embodiments, the heterologous object sequence is less than 1,000, 1,300, 1500, 2,000, 3,000, 4,000, 5,000, or 7,500 nucleotides in length.

In some embodiments, the genomic safe harbor site is a Natural Harbor™ site. In some embodiments, the Natural Harbor™ site is ribosomal DNA (rDNA). In some embodiments, the Natural Harbor™ site is 5S rDNA, 18S rDNA, 5.8S rDNA, or 28S rDNA. In some embodiments, the Natural Harbor™ site is the Mutsu site in 5S rDNA. In some embodiments, the Natural Harbor™ site is the R2 site, the R5 site, the R6 site, the R4 site, the R1 site, the R9 site, or the RT site in 28S rDNA. In some embodiments, the Natural Harbor™ site is the R8 site or the R7 site in 18S rDNA. In some embodiments, the Natural Harbor™ site is DNA encoding transfer RNA (tRNA). In some embodiments, the Natural Harbor™ site is DNA encoding tRNA-

Asp or tRNA-Glu. In some embodiments, the Natural Harbor™ site is DNA encoding spliceosomal RNA. In some embodiments, the Natural Harbor™ site is DNA encoding small nuclear RNA (snRNA) such as U2 snRNA.

Thus, in some aspects, the present disclosure provides a method of inserting a heterologous object sequence into a Natural Harbor™ site. In some embodimetns, the method comprises using a GeneWriter system described herein, e.g., using a polypeptide of any of Table X, Z1, Z2, 3A, or 3B or a polypeptide having sequence similarity thereto, e.g., at least 80%, 85%, 90%, or 95% identity thereto. In some embodiments, the method comprises using an enzyme, e.g., a retrotransposase, to insert the heterologous object sequence into the Natural Harbor™ site. In some aspects, the present disclosure provides a host human cell comprising a heterologous object sequence (e.g., a sequence encoding a therapeutic polypeptide) situated at a Natural Harbor™ site in the genome of the cell. In some embodiments, the Natural Harbor™ site is a site described in Table 4 below. In some embodiments, the heterologous object sequence is inserted within 20, 50, 100, 150, 200, 250, 500, or 1000 base pairs of a sequence shown in Table 4. In some embodiments, the heterologous object sequence is inserted within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb of a sequence shown in Table 4. In some embodiments, the heterologous object sequence is inserted into a site having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to a sequence shown in Table 4. In some embodiments, the heterologous object sequence is inserted within 20, 50, 100, 150, 200, 250, 500, or 1000 base pairs, or within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb, of a site having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to a sequence shown in Table 4. In some embodiments, the heterologous object sequence is inserted within a gene indicated in Column 5 of Table 4, or within 20, 50, 100, 150, 200, 250, 500, or 1000 base pairs, or within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb, of the gene.

Table 4. Natural Harbor™ sites. Column 1 indicates a retrotransposon that inserts into the Natural Harbor™ site. Column 2 indicates the gene at the Natural Harbor™ site. Columns 3 and 4 show exemplary human genome sequence 5' and 3' of the insertion site (for example, 250 bp). Columns 5 and 6 list the example gene symbol and corresponding Gene ID.

Target Site	Target Gene	5' flanking sequence	3' flanking sequence	Example Gene Symbol	Example Gene ID
R2	28S rDNA	CCGGTCCCCCCCGCCGGGTCC GCCCCCGGGGCCGCGGTTCC GCGCGGCGCCTCGCCTCGGC CGGCGCCTAGCAGCCGACTT AGAACTGGTGCGGACCAGGG GAATCCGACTGTTTAATTAAA ACAAAGCATCGCGAAGGCCC GCGGCGGGTGTTGACGCGAT GTGATTTCTGCCCAGTGCTCT GAATGTCAAAGTGAAGAAAT TCAATGAAGCGCGGGTAAAC GGCGGGAGTAACTATGACTC TCTTAAG (SEQ ID NO: 1508)	GTAGCCAAATGCCTCGTCATC TAATTAGTGACGCGCATGAAT GGATGAACGAGATTCCCACT GTCCCTACCTACTATCCAGCG AAACCACAGCCAAGGGAACG GGCTTGGCGGAATCAGCGGG GAAAGAAGACCCTGTTGAGC TTGACTCTAGTCTGGCACGGT GAAGAGACATGAGAGGTGTA GAATAAGTGGGAGGCCCCCG GCGCCCCCCCCGGTGTCCCCGC GAGGGGCCCGGGGCGGGGT CCGCCG (SEQ ID NO: 1513)	RNA28SN1	106632264
R4	28S rDNA	GCGGTTCCGCGCGGCGCCTC GCCTCGGCCGGCGCCTAGCA GCCGACTTAGAACTGGTGCG GACCAGGGGAATCCGACTGT TTAATTAAAACAAAGCATCGC GAAGGCCCGCGGCGGGTGTT GACGCGATGTGATTTCTGCCC AGTGCTCTGAATGTCAAAGT GAAGAAATTCAATGAAGCGC GGGTAAACGGCGGGAGTAAC TATGACTCTCTTAAGGTAGCC AAATGCCTCGTCATCTAATTA GTGACG (SEQ ID NO: 1509)	CGCATGAATGGATGAACGAG ATTCCCACTGTCCCTACCTACT ATCCAGCGAAACCACAGCCA AGGGAACGGGCTTGGCGGA ATCAGCGGGGAAAGAAGACC CTGTTGAGCTTGACTCTAGTC TGGCACGGTGAAGAGACATG AGAGGTGTAGAATAAGTGGG AGGCCCCCGGCGCCCCCCCG GTGTCCCCGCGAGGGGCCCG GGGCGGGGTCCGCCGGCCCT GCGGGCCGCCGGTGAAATAC CACTACTC (SEQ ID NO: 1514)	RNA28SN1	106632264
R5	28S rDNA	TCCCCCCCGCCGGGTCCGCCC CCGGGGCCGCGGTTCCGCGC GGCGCCTCGCCTCGGCCGGC GCCTAGCAGCCGACTTAGAA CTGGTGCGGACCAGGGGAAT CCGACTGTTTAATTAAAACAA AGCATCGCGAAGGCCCGCGG CGGGTGTTGACGCGATGTGA TTTCTGCCCAGTGCTCTGAAT GTCAAAGTGAAGAAATTCAA TGAAGCGCGGGTAAACGGCG GGAGTAACTATGACTCTCTTA AGGTAG (SEQ ID NO: 1510)	CCAAATGCCTCGTCATCTAAT TAGTGACGCGCATGAATGGA TGAACGAGATTCCCACTGTCC CTACCTACTATCCAGCGAAAC CACAGCCAAGGGAACGGGCT TGGCGGAATCAGCGGGGAAA GAAGACCCTGTTGAGCTTGA CTCTAGTCTGGCACGGTGAA GAGACATGAGAGGTGTAGAA TAAGTGGGAGGCCCCCGGCG CCCCCCCGGTGTCCCCGCGAG GGGCCCGGGGCGGGGTCCG CCGGCCC (SEQ ID NO: 1515)	RNA28SN1	106632264

R9	28S rDNA	CGGCGCGCTCGCCGGCCGAG GTGGGATCCCGAGGCCTCTC CAGTCCGCCGAGGGCGCACC ACCGGCCCGTCTCGCCCGCCG CGCCGGGGAGGTGGAGCAC GAGCGCACGTGTTAGGACCC GAAAGATGGTGAACTATGCC TGGGCAGGGCGAAGCCAGA GGAAACTCTGGTGGAGGTCC GTAGCGGTCCTGACGTGCAA ATCGGTCGTCCGACCTGGGT ATAGGGGCGAAAGACTAATC GAACCATCTAG (SEQ ID NO: 1511)	TAGCTGGTTCCCTCCGAAGTT TCCCTCAGGATAGCTGGCGCT CTCGCAGACCCGACGCACCCC CGCCACGCAGTTTTATCCGGT AAAGCGAATGATTAGAGGTC TTGGGGCCGAAACGATCTCA ACCTATTCTCAAACTTTAAAT GGGTAAGAAGCCCGGCTCGC TGGCGTGGAGCCGGGCGTGG AATGCGAGTGCCTAGTGGGC CACTTTTGGTAAGCAGAACTG GCGCTGCGGGATGAACCGAA CGCC (SEQ ID NO: 1516)	RNA28SN1	106632264
R8	18S rDNA	GCATTCGTATTGCGCCGCTAG AGGTGAAATTCTTGGACCGG CGCAAGACGGACCAGAGCGA AAGCATTTGCCAAGAATGTTT TCATTAATCAAGAACGAAAGT CGGAGGTTCGAAGACGATCA GATACCGTCGTAGTTCCGACC ATAAACGATGCCGACCGGCG ATGCGGCGGCGTTATTCCCAT GACCCGCCGGGCAGCTTCCG GGAAACCAAAGTCTTTGGGT TCCGGGGGGAGTATGGTTGC AAAGC (SEQ ID NO: 1512)	TGAAACTTAAAGGAATTGAC GGAAGGGCACCACCAGGAGT GGAGCCTGCGGCTTAATTTG ACTCAACACGGGAAACCTCA CCCGGCCCGGACACGGACAG GATTGACAGATTGATAGCTCT TTCTCGATTCCGTGGGTGGTG GTGCATGGCCGTTCTTAGTTG GTGGAGCGATTTGTCTGGTT AATTCCGATAACGAACGAGA CTCTGGCATGCTAACTAGTTA CGCGACCCCCGAGCGGTCGG CGTCCC (SEQ ID NO: 1517)	RNA18SN1	106631781
R4-2_SRa	tRNA-Asp			TRD-GTC1-1	100189207
LIN25_SM	tRNA-Glu			TRE-CTC1-1	100189384
R1	28S rDNA	TAGCAGCCGACTTAGAACTG GTGCGGACCAGGGGAATCCG ACTGTTTAATTAAAACAAAGC ATCGCGAAGGCCCGCGGCGG GTGTTGACGCGATGTGATTTC TGCCCAGTGCTCTGAATGTCA AAGTGAAGAAATTCAATGAA GCGCGGGTAAACGGCGGGA GTAACTATGACTCTCTTAAGG TAGCCAAATGCCTCGTCATCT AATTAGTGACGCGCATGAAT	ACCTACTATCCAGCGAAACCA CAGCCAAGGGAACGGGCTTG GCGGAATCAGCGGGGAAAG AAGACCCTGTTGAGCTTGACT CTAGTCTGGCACGGTGAAGA GACATGAGAGGTGTAGAATA AGTGGGAGGCCCCCGGCGCC CCCCCGGTGTCCCCGCGAGG GGCCCGGGGCGGGGTCCGCC GGCCCTGCGGGCCGCCGGTG AAATACCACTACTCTGATCGT TTTTTCACTGACCCGGTGAGG	RNA28SN1	106632264

		GGATGAACGAGATTCCCACT GTCCCT (SEQ ID NO: 1518)	CGGGGGG (SEQ ID NO: 1524)		
R6	28S rDNA	CCCCCCGCCGGGTCCGCCCCC GGGGCCGCGGTTCCGCGCGG CGCCTCGCCTCGGCCGGCGC CTAGCAGCCGACTTAGAACT GGTGCGGACCAGGGGAATCC GACTGTTTAATTAAAACAAAG CATCGCGAAGGCCCGCGGCG GGTGTTGACGCGATGTGATT TCTGCCCAGTGCTCTGAATGT CAAAGTGAAGAAATTCAATG AAGCGCGGGTAAACGGCGG GAGTAACTATGACTCTCTTAA GGTAGCC (SEQ ID NO: 1519)	AAATGCCTCGTCATCTAATTA GTGACGCGCATGAATGGATG AACGAGATTCCCACTGTCCCT ACCTACTATCCAGCGAAACCA CAGCCAAGGGAACGGGCTTG GCGGAATCAGCGGGGAAAG AAGACCCTGTTGAGCTTGACT CTAGTCTGGCACGGTGAAGA GACATGAGAGGTGTAGAATA AGTGGGAGGCCCCCGGCGCC CCCCCGGTGTCCCCGCGAGG GGCCCGGGGCGGGGTCCGCC GGCCCTG (SEQ ID NO: 1525)	RNA28SN1	106632264
R7	18S rDNA	GCGCAAGACGGACCAGAGCG AAAGCATTTGCCAAGAATGTT TTCATTAATCAAGAACGAAAG TCGGAGGTTCGAAGACGATC AGATACCGTCGTAGTTCCGAC CATAAACGATGCCGACCGGC GATGCGGCGGCGTTATTCCC ATGACCCGCCGGGCAGCTTC CGGGAAACCAAAGTCTTTGG GTTCCGGGGGGAGTATGGTT GCAAAGCTGAAACTTAAAGG AATTGACGGAAGGGCACCAC CAGGAGT (SEQ ID NO: 1520)	GGAGCCTGCGGCTTAATTTG ACTCAACACGGGAAACCTCA CCCGGCCCGGACACGGACAG GATTGACAGATTGATAGCTCT TTCTCGATTCCGTGGGTGGTG GTGCATGGCCGTTCTTAGTTG GTGGAGCGATTTGTCTGGTT AATTCCGATAACGAACGAGA CTCTGGCATGCTAACTAGTTA CGCGACCCCCGAGCGGTCGG CGTCCCCCAACTTCTTAGAGG GACAAGTGGCGTTCAGCCAC CCGAG (SEQ ID NO: 1526)	RNA18SN1	106631781
RT	28S rDNA	GGCCGGGCGCGACCCGCTCC GGGGACAGTGCCAGGTGGG GAGTTTGACTGGGGCGGTAC ACCTGTCAAACGGTAACGCA GGTGTCCTAAGGCGAGCTCA GGGAGGACAGAAACCTCCCG TGGAGCAGAAGGGCAAAAG CTCGCTTGATCTTGATTTTCA GTACGAATACAGACCGTGAA AGCGGGGCCTCACGATCCTTC TGACCTTTTGGGTTTTAAGCA GGAGGTGTCAGAAAAGTTAC CACAGGGAT (SEQ ID NO: 1521)	AACTGGCTTGTGGCGGCCAA GCGTTCATAGCGACGTCGCTT TTTGATCCTTCGATGTCGGCT CTTCCTATCATTGTGAAGCAG AATTCACCAAGCGTTGGATTG TTCACCCACTAATAGGGAACG TGAGCTGGGTTTAGACCGTC GTGAGACAGGTTAGTTTTACC CTACTGATGATGTGTTGTTGC CATGGTAATCCTGCTCAGTAC GAGAGGAACCGCAGGTTCAG ACATTTGGTGTATGTGCTTGG C (SEQ ID NO: 1527)	RNA28SN1	106632264

Mutsu	5S rDNA	GTCTACGGCCATACCACCC (SEQ ID NO: 1522)	TGAACGCGCCCGATCTCGTCT GATCTCGGAAGCTAAGCAGG GTCGGGCCTGGTTAGTACTT GGATGGGAGACCGCCTGGGA ATACCGGGTGCTGTAGGCTTT (SEQ ID NO: 1528)	RNA5S1	100169751
Utopia /Keno	U2 snRNA	ATCGCTTCTCGGCCTTTTGGC TAAGATCAAGTGTAGTA (SEQ ID NO: 1523)	TCTGTTCTTATCAGTTTAATAT CTGATACGTCCTCTATCCGAG GACAATATATTAAATGGATTT TTGGAGCAGGGAGATGGAAT AGGAGCTTGCTCCGTCCACTC CACGCATCGACCTGGTATTGC AGTACCTCCAGGAACGGTGC ACCC (SEQ ID NO: 1529)	RNU2-1	6066

In some embodiments, a system or method described herein results in insertion of a heterologous sequence into a target site in the human genome. In some embodiments, the target site in the human genome has sequence similarity to the corresponding target site of the corresponding wild-type retrotransposase (e.g., the retrotransposase from which the GeneWriter was derived) in the genome of the organism to which it is native. For instance, in some embodiments, the identity between the 40 nucleotides of human genome sequence centered at the insertion site and the 40 nucleotides of native organism genome sequence centered at the insertion site is less than 99.5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60%, or 50%, or is between 50-60%, 60-70%, 70-80%, 80-90%, or 90-100%. In some embodiments, the identity between the 100 nucleotides of human genome sequence centered at the insertion site and the 100 nucleotides of native organism genome sequence centered at the insertion site is less than 99.5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60%, or 50%, or is between 50-60%, 60-70%, 70-80%, 80-90%, or 90-100%. In some embodiments, the identity between the 500 nucleotides of human genome sequence centered at the insertion site and the 500 nucleotides of native organism genome sequence centered at the insertion site is less than 99.5%, 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 60%, or 50%, or is between 50-60%, 60-70%, 70-80%, 80-90%, or 90-100%.

The template nucleic acid (e.g., template RNA) component of a Gene Writer genome editing system described herein typically is able to bind the Gene Writer genome editing protein

of the system. In some embodiments, the template nucleic acid (e.g., template RNA) has a 3' region that is capable of binding a Gene Writer genome editing protein. The binding region, e.g., 3' region, may be a structured RNA region, e.g., having at least 1, 2 or 3 hairpin loops, capable of binding the Gene Writer genome editing protein of the system. The binding region may associate the template nucleic acid (e.g., template RNA) with any of the polypeptide modules. In some embodiments, the binding region of the template nucleic acid (e.g., template RNA) may associate with an RNA-binding domain in the polypeptide. In some embodiments, the binding region of the template nucleic acid (e.g., template RNA) may associate with the reverse transcription domain of the polypeptide (e.g., specifically bind to the RT domain). For example, where the reverse transcription domain is derived from a non-LTR retrotransposon, the template nucleic acid (e.g., template RNA) may contain a binding region derived from a non-LTR retrotransposon, e.g., a 3' UTR from a non-LTR retrotransposon. In some embodiments, the template nucleic acid (e.g., template RNA) may associate with the DNA binding domain of the polypeptide, e.g., a gRNA associating with a Cas9-derived DNA binding domain. In some embodiments, the binding region may also provide DNA target recognition, e.g., a gRNA hybridizing to the target DNA sequence and binding the polypeptide, e.g., a Cas9 domain. In some embodiments, the template nucleic acid (e.g., template RNA) may associate with multiple components of the polypeptide, e.g., DNA binding domain and reverse transcription domain. For example, the template nucleic acid (e.g., template RNA) may comprise a gRNA region that associates with a Cas9-derived DNA binding domain and a 3' UTR from a non-LTR retrotransposon that associated with a non-LTR retrotransposon-derived reverse transcription domain.

In some embodiments a system or method described herein comprises a single template nucleic acid (e.g., template RNA). In some embodiments a system or method described herein comprises a plurality of template nucleic acids (e.g., template RNAs). For example, a system described herein comprises a first RNA comprising (e.g., from 5' to 3') a sequence that binds the Gene Writer polypeptide (e.g., the DNA-binding domain and/or the endonuclease domain, e.g., a gRNA) and a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), and a second RNA (e.g., a template RNA) comprising (e.g., from 5' to 3') optionally a sequence that binds the Gene Writer polypeptide (e.g., that specifically binds the RT domain), a heterologous object sequence, and a 3' homology domain. In some embodiments, when the

system comprises a plurality of nucleic acids, each nucleic acid comprises a conjugating domain. In some embodiments, a conjugating domain enables association of nucleic acid molecules, e.g., by hybridization of complementary sequences.

In some embodiments, the template nucleic acid may further comprise a gRNA. In some embodiments, the template nucleic acid may bind to the Gene Writer polypeptide by interaction of a gRNA portion of the template nucleic acid with a heterologous RNA binding domain. In some embodiments, the heterologous RNA binding domain is Cas9.

In some embodiments, the template nucleic acid may comprise one or more UTRs (e.g., a 5' UTR or a 3' UTR, e.g., from an R2-type retrotransposon) and a gRNA. In some embodiments, the UTR facilitates interaction of the template with the reverse transcriptase domain of the polypeptide. In some embodiments, the gRNA facilitates interaction with the binding domain of the polypeptide. In some embodiments, the gRNA directs the polypeptide to the matching target sequence. In some embodiments, the template may contain only the reverse transcriptase binding motif (e.g., a 3' UTR from R2) and the gRNA may be provided as a second RNA molecule for target site recognition. In some embodiments, the template containing the RT-binding motif may exist on the same molecule as the gRNA, but be processed into two RNA molecules by cleavage activity (e.g., a ribozyme).

In some embodiments, the template possesses one or more sequences aiding in association of the template with the Gene Writer polypeptide. In some embodiments, these sequences may be derived from retrotransposon UTRs. In some embodiments, the UTRs may be located flanking the desired insertion sequence. In some embodiments, a sequence with target site homology may be located outside of one or both UTRs. In some embodiments, the sequence with target site homology can anneal to the target sequence to prime reverse transcription. In some embodiments, the 5' and/or 3' UTR may be located terminal to the target site homology sequence, e.g., such that target primed reverse transcription excludes reverse transcription of the 5' and/or 3' UTR. In some embodiments, the Gene Writer system may result in the insertion of a desired payload without any additional sequence (e.g., a gene expression unit without UTRs used to bind the Gene Writer protein).

The template nucleic acid (e.g., template RNA) can be designed to result in insertions, mutations, or deletions at the target DNA locus. In some embodiments, the template nucleic acid (e.g., template RNA) may be designed to cause an insertion in the target DNA. For example, the

template nucleic acid (e.g., template RNA) may contain a heterologous sequence, wherein the reverse transcription will result in insertion of the heterologous sequence into the target DNA. In other embodiments, the RNA template may be designed to write a deletion into the target DNA. For example, the template nucleic acid (e.g., template RNA) may match the target DNA upstream and downstream of the desired deletion, wherein the reverse transcription will result in the copying of the upstream and downstream sequences from the template nucleic acid (e.g., template RNA) without the intervening sequence, e.g., causing deletion of the intervening sequence. In other embodiments, the template nucleic acid (e.g., template RNA) may be designed to write an edit into the target DNA. For example, the template RNA may match the target DNA sequence with the exception of one or more nucleotides, wherein the reverse transcription will result in the copying of these edits into the target DNA, e.g., resulting in mutations, e.g., transition or transversion mutations.

In some embodiments, a Gene Writer system is capable of producing an insertion into the target site of at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides (and optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a Gene Writer system is capable of producing an insertion into the target site of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides (and optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a Gene Writer system is capable of producing an insertion into the target site of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a Gene Writer system is capable of producing a deletion of at least 0.2, 0.3, 0.4,

0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases).

Additional Template Features

In some embodiments, the template (e.g., template RNA) comprises certain structural features, e.g., determined *in silico*. In embodiments, the template RNA is predicted to have minimal energy structures between -280 and -480 kcal/mol (e.g., between -280 to -300, -300 to -350, -350 to -400, -400 to -450, or -450 to -480 kcal/mol), e.g., as measured by RNAstructure, e.g., as described in Turner and Mathews *Nucleic Acids Res* 38:D280-282 (2009) (incorporated herein by reference in its entirety).

In some embodiments, the template (e.g., template RNA) comprises certain structural features, e.g., determined *in vitro*. In embodiments, the template RNA is sequence optimized, e.g., to reduce secondary structure as determined *in vitro*, for example, by SHAPE-MaP (e.g., as described in Siegfried et al. Nat Methods 11:959-965 (2014); incorporated herein by reference in its entirety).In some embodiments, the template (e.g., template RNA) comprises certain structural features, e.g., determined in cells. In embodiments, the template RNA is sequence optimized, e.g., to reduce secondary structure as measured in cells, for example, by DMS-MaPseq (e.g., as described in Zubradt et al. Nat Methods 14:75-82 (2017); incorporated by reference herein in its entirety).

Additional Functional Characteristics for Gene Writers™

A Gene Writer as described herein may, in some instances, be characterized by one or more functional measurements or characteristics. In some embodiments, the DNA binding domain has one or more of the functional characteristics described below. In some embodiments, the RNA binding domain has one or more of the functional characteristics described below. In some embodiments, the endonuclease domain has one or more of the functional characteristics described below. In some embodiments, the reverse transcriptase domain has one or more of the functional characteristics described below. In some embodiments, the template (e.g., template RNA) has one or more of the functional characteristics

described below. In some embodiments, the target site bound by the Gene Writer has one or more of the functional characteristics described below.

In embodiments, the disclosure provides a nucleic acid molecule or a system for retargeting, e.g., of a Gene Writer polypeptide or nucleic acid molecule, or of a system as described herein. Retargeting (e.g., of a Gene Writer polypeptide or nucleic acid molecule, or of a system as described herein) generally comprises : (i) directing the polypeptide to bind and cleave at the target site; and/or (ii) designing the template RNA to have complementarity to the target sequence. In some embodiments, the template RNA has complementarity to the target sequence 5' of the first-strand nick, e.g., such that the 3' end of the template RNA anneals and the 5' end of the target site serves as the primer, e.g., for target-primed reverse transcription (TPRT). In some embodiments, the endonuclease domain of the polypeptide and the 5' end of the RNA template are also modified as described.

Gene Writer Polypeptide

DNA Binding Domain

In some embodiments, the DNA binding domain is capable of binding to a target sequence (e.g., a dsDNA target sequence) with greater affinity than a reference DNA binding domain. In some embodiments, the reference DNA binding domain is a DNA binding domain from R2_BM of B. mori. In some embodiments, the DNA binding domain is capable of binding to a target sequence (e.g., a dsDNA target sequence) with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM).

In some embodiments, the affinity of a DNA binding domain for its target sequence (e.g., dsDNA target sequence) is measured in vitro, e.g., by thermophoresis, e.g., as described in Asmari et al. Methods 146:107-119 (2018) (incorporated by reference herein in its entirety).

In embodiments, the DNA binding domain is capable of binding to its target sequence (e.g., dsDNA target sequence), e.g, with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM) in the presence of a molar excess of scrambled sequence competitor dsDNA, e.g., of about 100-fold molar excess.

In some embodiments, the DNA binding domain is found associated with its target sequence (e.g., dsDNA target sequence) more frequently than any other sequence in the genome of a target cell, e.g., human target cell, e.g., as measured by ChIP-seq (e.g., in HEK293T cells), e.g., as described in He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (incorporated herein by

reference in its entirety). In some embodiments, the DNA binding domain is found associated with its target sequence (e.g., dsDNA target sequence) at least about 5-fold or 10-fold, more frequently than any other sequence in the genome of a target cell, e.g., as measured by ChIP-seq (e.g., in HEK293T cells), e.g., as described in He and Pu (2010), *supra*.

In some embodiments, a Gene Writer polypeptide comprises a modification to a DNA-binding domain, e.g., relative to the wild-type polypeptide. In some embodiments, the DNA-binding domain comprises an addition, deletion, replacement, or modification to the amino acid sequence of the original DNA-binding domain. In some embodiments, the DNA-binding domain is modified to include a heterologous functional domain that binds specifically to a target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the functional domain replaces at least a portion (e.g., the entirety of) the prior DNA-binding domain of the polypeptide. In some embodiments, the functional domain comprises a zinc finger (e.g., a zinc finger that specifically binds to the target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the functional domain comprises a Cas domain (e.g., a Cas domain that specifically binds to the target nucleic acid (e.g., DNA) sequence of interest. In embodiments, the Cas domain comprises a Cas9 or a mutant or variant thereof (e.g., as described herein). In embodiments, the Cas domain is associated with a guide RNA (gRNA), e.g., as described herein. In embodiments, the Cas domain is directed to a target nucleic acid (e.g., DNA) sequence of interest by the gRNA. In embodiments, the Cas domain is encoded in the same nucleic acid (e.g., RNA) molecule as the gRNA. In embodiments, the Cas domain is encoded in a different nucleic acid (e.g., RNA) molecule from the gRNA.

In some embodiments, a Gene Writer polypeptide comprises a modification to an endonuclease domain, e.g., relative to the wild-type polypeptide. In some embodiments, the endonuclease domain comprises an addition, deletion, replacement, or modification to the amino acid sequence of the original endonuclease domain. In some embodiments, the endonuclease domain is modified to include a heterologous functional domain that binds specifically to and/or induces endonuclease cleavage of a target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the endonuclease domain comprises a zinc finger. In some embodiments, the endonuclease domain comprises a Cas domain (e.g., a Cas9 or a mutant or variant thereof). In embodiments, the endonuclease domain comprising the Cas domain is associated with a guide RNA (gRNA), e.g., as described herein. In some embodiments, the endonuclease domain is

modified to include a functional domain that does not target a specific target nucleic acid (e.g., DNA) sequence. In embodiments, the endonuclease domain comprises a Fok1 domain.

In some embodiments, the reverse transcriptase (RT) domain exhibits enhanced stringency of target-primed reverse transcription (TPRT) initiation, e.g., relative to an endogenous RT domain. In some embodiments, the RT domain initiates TPRT when the 3 nt in the target site immediately upstream of the first strand nick, e.g., the genomic DNA priming the RNA template, have at least 66% or 100% complementarity to the 3 nt of homology in the RNA template. In some embodiments, the RT domain initiates TPRT when there are less than 5 nt mismatched (e.g., less than 1, 2, 3, 4, or 5 nt mismatched) between the template RNA homology and the target DNA priming reverse transcription. In some embodiments, the RT domain is modified such that the stringency for mismatches in priming the TPRT reaction is increased, e.g., wherein the RT domain does not tolerate any mismatches or tolerates fewer mismatches in the priming region relative to a wild-type (e.g., unmodified) RT domain. In some embodiments, the RT domain comprises a HIV-1 RT domain. In embodiments, the HIV-1 RT domain initiates lower levels of synthesis even with three nucleotide mismatches relative to an alternative RT domain (e.g., as described by Jamburuthugoda and Eickbush J Mol Biol 407(5):661-672 (2011); incorporated herein by reference in its entirety).

RNA Binding Domain

In some embodiments, the RNA binding domain is capable of binding to a template RNA with greater affinity than a reference RNA binding domain. In some embodiments, the reference RNA binding domain is an RNA binding domain from R2_BM of B. mori. In some embodiments, the RNA binding domain is capable of binding to a template RNA with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM). In some embodiments, the affinity of a RNA binding domain for its template RNA is measured in vitro, e.g., by thermophoresis, e.g., as described in Asmari et al. Methods 146:107-119 (2018) (incorporated by reference herein in its entirety). In some embodiments, the affinity of a RNA binding domain for its template RNA is measured in cells (e.g., by FRET or CLIP-Seq).

In some embodiments, the RNA binding domain is associated with the template RNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled RNA. In some embodiments, the frequency of association between the RNA binding domain and the template RNA or scrambled RNA is measured by CLIP-seq, e.g., as described in Lin and Miles (2019)

Nucleic Acids Res 47(11):5490-5501 (incorporated by reference herein in its entirety). In some embodiments, the RNA binding domain is associated with the template RNA in cells (e.g., in HEK293T cells) at a frequency at least about 5-fold or 10-fold higher than with a scrambled RNA. In some embodiments, the frequency of association between the RNA binding domain and the template RNA or scrambled RNA is measured by CLIP-seq, e.g., as described in Lin and Miles (2019), *supra.*

Endonuclease Domain

In some embodiments, the endonuclease domain is associated with the target dsDNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled dsDNA. In some embodiments, the endonuclease domain is associated with the target dsDNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled dsDNA, e.g., in a cell (e.g., a HEK293T cell). In some embodiments, the frequency of association between the endonuclease domain and the target DNA or scrambled DNA is measured by ChIP-seq, e.g., as described in He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (incorporated by reference herein in its entirety).

In some embodiments, the endonuclease domain can catalyze the formation of a nick at a target sequence, e.g., to an increase of at least about 5-fold or 10-fold relative to a non-target sequence (e.g., relative to any other genomic sequence in the genome of the target cell). In some embodiments, the level of nick formation is determined using NickSeq, e.g., as described in Elacqua et al. (2019) *bioRxiv* doi.org/10.1101/867937 (incorporated herein by reference in its entirety).

In some embodiments, the endonuclease domain is capable of nicking DNA in vitro. In embodiments, the nick results in an exposed base. In embodiments, the exposed base can be detected using a nuclease sensitivity assay, e.g., as described in Chaudhry and Weinfeld (1995) *Nucleic Acids Res* 23(19):3805-3809 (incorporated by reference herein in its entirety). In embodiments, the level of exposed bases (e.g., detected by the nuclease sensitivity assay) is increased by at least 10%, 50%, or more relative to a reference endonuclease domain. In some embodiments, the reference endonuclease domain is an endonuclease domain from R2_BM of B. mori.

In some embodiments, the endonuclease domain is capable of nicking DNA in a cell. In embodiments, the endonuclease domain is capable of nicking DNA in a HEK293T cell. In embodiments, an unrepaired nick that undergoes replication in the absence of Rad51 results in increased NHEJ rates at the site of the nick, which can be detected, e.g., by using a Rad51 inhibition assay, e.g., as described in Bothmer et al. (2017) *Nat Commun* 8:13905 (incorporated by reference herein in its entirety). In embodiments, NHEJ rates are increased above 0-5%. In embodiments, NHEJ rates are increased to 20-70% (e.g., between 30%-60% or 40-50%), e.g., upon Rad51 inhibition.

In some embodiments, the endonuclease domain releases the target after cleavage. In some embodiments, release of the target is indicated indirectly by assessing for multiple turnovers by the enzyme, e.g., as described in Yourik at al. RNA 25(1):35-44 (2019) (incorporated herein by reference in its entirety) and shown in Figure 2. In some embodiments, the k_{exp} of an endonuclease domain is 1 x 10^{-3} – 1 x 10^{-}5 min-1 as measured by such methods.

In some embodiments, the endonuclease domain has a catalytic efficiency (k_{cat}/K_m) greater than about 1 x 10^8 s^{-1} M^{-1} in vitro. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1 x 10^5, 1 x 10^6, 1 x 10^7, or 1 x 10^8, s^{-1} M^{-1} in vitro. In embodiments, catalytic efficiency is determined as described in Chen et al. (2018) *Science* 360(6387):436-439 (incorporated herein by reference in its entirety). In some embodiments, the endonuclease domain has a catalytic efficiency (k_{cat}/K_m) greater than about 1 x 10^8 s^{-1} M^{-1} in cells. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1 x 10^5, 1 x 10^6, 1 x 10^7, or 1 x 10^8 s^{-1} M^{-1} in cells.

Reverse Transcriptase Domain

In some embodiments, the reverse transcriptase domain has a lower probability of premature termination rate (P_{off}) *in vitro* relative to a reference reverse transcriptase domain. In some embodiments, the reference reverse transcriptase domain is a reverse transcriptase domain from R2_BM of B. mori or a viral reverse transcriptase domain, e.g., the RT domain from M-MLV.

In some embodiments, the reverse transcriptase domain has a lower probability of premature termination rate (P_{off}) *in vitro* of less than about 5 x 10^{-3}/nt, 5 x 10^{-4}/nt, or 5 x 10^{-6}/nt, e.g., as measured on a 1094 nt RNA. In embodiments, the *in vitro* premature termination rate is

determined as described in Bibillo and Eickbush (2002) J Biol Chem 277(38):34836-34845 (incorporated by reference herein its entirety).

In some embodiments, the reverse transcriptase domain is able to complete at least about 30% or 50% of integrations in cells. The percent of complete integrations can be measured by dividing the number of substantially full-length integration events (e.g., genomic sites that comprise at least 98% of the expected integrated sequence) by the number of total (including substantially full-length and partial) integration events in a population of cells. In embodiments, the integrations in cells is determined (e.g., across the integration site) using long-read amplicon sequencing, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein its in entirety).

In embodiments, quantifying integrations in cells comprises counting the fraction of integrations that contain at least about 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the DNA sequence corresponding to the template RNA (e.g., a template RNA having a length of at least 0.05, 0.1, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 3, 4, or 5 kb, e.g., a length between 0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-0.9, 1.0-1.2, 1.2-1.4, 1.4-1.6, 1.6-1.8, 1.8-2.0, 2-3, 3-4, or 4-5 kb).

In some embodiments, the reverse transcriptase domain is capable of polymerizing dNTPs *in vitro*. In embodiments, the reverse transcriptase domain is capable of polymerizing dNTPs *in vitro* at a rate between 0.1 – 50 nt/sec (e.g., between 0.1-1, 1-10, or 10-50 nt/sec). In embodiments, polymerization of dNTPs by the reverse transcriptase domain is measured by a single-molecule assay, e.g., as described in Schwartz and Quake (2009) *PNAS* 106(48):20294-20299 (incorporated by reference in its entirety).

In some embodiments, the reverse transcriptase domain has an *in vitro* error rate (e.g., misincorporation of nucleotides) of between $1 \times 10^{-3} – 1 \times 10^{-4}$ or $1 \times 10^{-4} – 1 \times 10^{-5}$ substitutions/nt , e.g., as described in Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (incorporated herein by reference in its entirety). In some embodiments, the reverse transcriptase domain has an error rate (e.g., misincorporation of nucleotides) in cells (e.g., HEK293T cells) of between $1 \times 10^{-3} – 1 \times 10^{-4}$ or $1 \times 10^{-4} – 1 \times 10^{-5}$ substitutions/nt, e.g., by long-read amplicon sequencing, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein in its entirety).

In some embodiments, the reverse transcriptase domain is capable of performing reverse transcription of a target RNA in vitro. In some embodiments, the reverse transcriptase requires a

primer of at least 3 nt to initiate reverse transcription of a template. In some embodiments, reverse transcription of the target RNA is determined by detection of cDNA from the target RNA (e.g., when provided with a ssDNA primer, e.g., which anneals to the target with at least 3, 4, 5, 6, 7, 8, 9, or 10 nt at the 3' end), e.g., as described in Bibillo and Eickbush (2002) *J Biol Chem* 277(38):34836-34845 (incorporated herein by reference in its entirety).

In some embodiments, the reverse transcriptase domain performs reverse transcription at least 5 or 10 times more efficiently (e.g., by cDNA production), e.g., when converting its RNA template to cDNA, for example, as compared to an RNA template lacking the protein binding motif (e.g., a 3' UTR). In embodiments, efficiency of reverse transcription is measured as described in Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (incorporated by reference herein in its entirety).

In some embodiments, the reverse transcriptase domain specifically binds a specific RNA template with higher frequency (e.g., about 5 or 10-fold higher frequency) than any endogenous cellular RNA, e.g., when expressed in cells (e.g., HEK293T cells). In embodiments, frequency of specific binding between the reverse transcriptase domain and the template RNA are measured by CLIP-seq, e.g., as described in Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (incorporated herein by reference in its entirety).

Target Site

In some embodiments, after Gene Writing, the target site surrounding the integrated sequence contains a limited number of insertions or deletions, for example, in less than about 50% or10% of integration events, e.g., as determined by long-read amplicon sequencing of the target site, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein in its entirety). In some embodiments, the target site does not show multiple insertion events, e.g., head-to-tail or head-to-head duplications, e.g., as determined by long-read amplicon sequencing of the target site, e.g., as described in Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (incorporated herein by reference in its entirety). In some embodiments, the target site contains an integrated sequence corresponding to the template RNA. In some embodiments, the target site does not contain insertions resulting from endogenous RNA in more than about 1% or10% of events, e.g., as determined by long-read amplicon sequencing of the target site, e.g., as described in Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020)

(incorporated herein by reference in its entirety). In some embodiments, the target site contains the integrated sequence corresponding to the template RNA.

In some embodiments, the target site contains an integrated sequence corresponding to the template RNA. In embodiments, the target site does not comprise sequence outside of the RT template (e.g., gRNA scaffold, vector backbone, and/or ITRs), e.g., as determined by long-read amplicon sequencing of the target site (for example, as described in Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020); incorporated herein by reference in its entirety).

Evolved Variants of Gene Writers

In some embodiments, the invention provides evolved variants of Gene Writers. Evolved variants can, in some embodiments, be produced by mutagenizing a reference Gene Writer, or one of the fragments or domains comprised therein. In some embodiments, one or more of the domains (e.g., the reverse transcriptase, DNA binding (including, for example, sequence-guided DNA binding elements), RNA-binding, or endonuclease domain) is evolved. One or more of such evolved variant domains can, in some embodiments, be evolved alone or together with other domains. An evolved variant domain or domains may, in some embodiments, be combined with unevolved cognate component(s) or evolved variants of the cognate component(s), e.g., which may have been evolved in either a parallel or serial manner.

In some embodiments, the process of mutagenizing a reference Gene Writer, or fragment or domain thereof, comprises mutagenizing the reference Gene Writer or fragment or domain thereof. In embodiments, the mutagenesis comprises a continuous evolution method (e.g., PACE) or non-contious evolution method (e.g., PANCE), e.g., as described herein. In some embodiments, the evolved Gene Writer, or a fragment or domain thereof, comprises one or more amino acid variations introduced into its amino acid sequence relative to the amino acid sequence of the reference Gene Writer, or fragment or domain thereof. In embodiments, amino acid sequence variations may include one or more mutated residues (e.g., conservative substitutions, non-conservative substitutions, or a combination thereof) within the amino acid sequence of a reference Gene Writer, e.g., as a result of a change in the nucleotide sequence encoding the gene writer that results in, e.g., a change in the codon at any particular position in the coding sequence, the deletion of one or more amino acids (e.g., a truncated protein), the insertion of one or more amino acids, or any combination of the foregoing. The evolved variant

Gene Writer may include variants in one or more components or domains of the Gene Writer (e.g., variants introduced into a reverse transcriptase domain, endonuclease domain, DNA binding domain, RNA binding domain, or combinations thereof).

In some aspects, the invention provides Gene Writers, systems, kits, and methods using or comprising an evolved variant of a Gene Writer, e.g., employs an evolved variant of a Gene Writer or a Gene Writer produced or produceable by PACE or PANCE. In embodiments, the unevolved reference Gene Writer is a Gene Writer as disclosed herein.

The term "phage-assisted continuous evolution (PACE)."as used herein, generally refers to continuous evolution that employs phage as viral vectors. Examples of PACE technology have been described, for example, in International PCT Application No. PCT/US 2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; U.S. Patent No. 9,023,594, issued May 5, 2015; U.S. Patent No. 9,771,574, issued September 26, 2017; U.S. Patent No. 9,394,537, issued July 19, 2016; International PCT Application, PCT/US2015/012022, filed January 20, 2015, published as WO 2015/134121 on September 11, 2015; U.S. Patent No. 10,179,911, issued January 15, 2019; and International PCT Application, PCT/US2016/027795, filed April 15, 2016, published as WO 2016/168631 on October 20, 2016, the entire contents of each of which are incorporated herein by reference.

The term"phage-assisted non-continuous evolution (PANCE)," as used herein, generally refers to non-continuous evolution that employs phage as viral vectors. Examples of PANCE technology have been described, for example, in Suzuki T. et al, Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase, Nat Chem Biol. 13(12): 1261-1266 (2017), incorporated herein by reference in its entirety. Briefly, PANCE is a technique for rapid in vivo directed evolution using serial flask transfers of evolving selection phage (SP), which contain a gene of interest to be evolved, across fresh host cells (e.g., E. coli cells). Genes inside the host cell may be held constant while genes contained in the SP continuously evolve. Following phage growth, an aliquot of infected cells may be used to transfect a subsequent flask containing host E. coli. This process can be repeated and/or continued until the desired phenotype is evolved, e.g., for as many transfers as desired.

Methods of applying PACE and PANCE to Gene Writers may be readily appreciated by the skilled artisan by reference to, *inter alia*, the foregoing references. Additional exemplary

methods for directing continuous evolution of genome-modifying proteins or systems, e.g., in a population of host cells, e.g., using phage particles, can be applied to generate evolved variants of Gene Writers, or fragments or subdomains thereof. Non-limiting examples of such methods are described in International PCT Application, PCT/US2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; U.S. Patent No. 9,023,594, issued May 5, 2015; U.S. Patent No. 9,771,574, issued September 26, 2017; U.S. Patent No. 9,394,537, issued July 19, 2016; International PCT Application, PCT/US2015/012022, filed January 20, 2015, published as WO 2015/134121 on September 11, 2015; U.S. Patent No. 10,179,911, issued January 15, 2019; International Application No. PCT/US2019/37216, filed June 14, 2019, International Patent Publication WO 2019/023680, published January 31, 2019, International PCT Application, PCT/US2016/027795, filed April 15, 2016, published as WO 2016/168631 on October 20, 2016, and International Patent Publication No. PCT/US2019/47996, filed August 23, 2019, each of which is incorporated herein by reference in its entirety.

In some non-limiting illustrative embodiments, a method of evolution of a evolved variant Gene Writer, of a fragment or domain thereof, comprises: (a) contacting a population of host cells with a population of viral vectors comprising the gene of interest (the starting Gene Writer or fragment or domain thereof), wherein: (1) the host cell is amenable to infection by the viral vector; (2) the host cell expresses viral genes required for the generation of viral particles; (3) the expression of at least one viral gene required for the production of an infectious viral particle is dependent on a function of the gene of interest; and/or (4) the viral vector allows for expression of the protein in the host cell, and can be replicated and packaged into a viral particle by the host cell. In some embodiments, the method comprises (b) contacting the host cells with a mutagen, using host cells with mutations that elevate mutation rate (e.g., either by carrying a mutation plasmid or some genome modification—e.g., proofing-impaired DNA polymerase, SOS genes, such as UmuC, UmuD', and/or RecA, which mutations, if plasmid-bound, may be under control of an inducible promoter), or a combination thereof. In some embodiments, the method comprises (c) incubating the population of host cells under conditions allowing for viral replication and the production of viral particles, wherein host cells are removed from the host cell population, and fresh, uninfected host cells are introduced into the population of host cells,

thus replenishing the population of host cells and creating a flow of host cells. In some embodiments, the cells are incubated under conditions allowing for the gene of interest to acquire a mutation. In some embodiments, the method further comprises (d) isolating a mutated version of the viral vector, encoding an evolved gene product (e.g., an evolved variant Gene Writer, or fragment or domain thereof), from the population of host cells.

The skilled artisan will appreciate a variety of features employable within the above-described framework. For example, in some embodiments, the viral vector or the phage is a filamentous phage, for example, an M13 phage, e.g., an M13 selection phage. In certain embodiments, the gene required for the production of infectious viral particles is the M13 gene III (gIII). In embodiments, the phage may lack a functional gIII, but otherwise comprise gI, gII, gIV, gV, gVI, gVII, gVIII, gIX, and a gX. In some embodiments, the generation of infectious VSV particles involves the envelope protein VSV-G. Various embodiments can use different retroviral vectors, for example, Murine Leukemia Virus vectors, or Lentiviral vectors. In embodiments, the retroviral vectors can efficiently be packaged with VSV-G envelope protein, e.g., as a substitute for the native envelope protein of the virus.

In some embodiments, host cells are incubated according to a suitable number of viral life cycles, e.g., at least 10, at least 20, at least 30, at least 40, at least 50, at least 100, at least 200, at least 300, at least 400, at least, 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1250, at least 1500, at least 1750, at least 2000, at least 2500, at least 3000, at least 4000, at least 5000, at least 7500, at least 10000, or more consecutive viral life cycles, which in on illustrative and non-limiting examples of M13 phage is 10-20 minutes per virus life cycle. Similarly, conditions can be modulated to adjust the time a host cell remains in a population of host cells, e.g., about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 70, about 80, about 90, about 100, about 120, about 150, or about 180 minutes. Host cell populations can be controlled in part by density of the host cells, or, in some embodiments, the host cell density in an inflow, e.g., 10^3 cells/ml, about 10^4 cells/ml, about 10^5 cells/ml, about 5- 10^5 cells/ml, about 10^6 cells/ml, about 5- 10^6 cells/ml, about 10^7 cells/ml, about 5- 10^7 cells/ml, about 10^8 cells/ml, about 5- 10^8 cells/ml, about 10^9 cells/ml, about 5- 10^9 cells/ml, about 10^{10} cells/ml, or about 5- 10^{10} cells/ml.

Promoters

In some embodiments, one or more promoter or enhancer elements are operably linked to a nucleic acid encoding a Gene Writer protein or a template nucleic acid, e.g., that controls expression of the heterologous object sequence. In certain embodiments, the one or more promoter or enhancer elements comprise cell-type or tissue specific elements. In some embodiments, the promoter or enhancer is the same or derived from the promoter or enhancer that naturally controls expression of the heterologous object sequence. For example, the ornithine transcarbomylase promoter and enhancer may be used to control expression of the ornithine transcarbomylase gene in a system or method provided by the invention for correcting ornithine transcarbomylase deficiencies. In some embodiments, a promoter for use in the invention is for a gene described in any one of Tables **9-22**, e.g., which may be used with an allele of the reference gene, or, in other embodiments, with a heterologous gene. In some embodiments, the promoter is a promoter of Table 33 or a functional fragment or variant thereof.

Exemplary tissue specific promoters that are commercially available can be found, for example, at a uniform resource locator (e.g., https://www.invivogen.com/tissue-specific-promoters). In some embodiments, a promoter is a **native promoter** or a minimal promoter, e.g., which consists of a single fragment from the 5' region of a given gene. In some embodiments, a native promoter comprises a core promoter and its natural 5' UTR. In some embodiments, the 5' UTR comprises an intron. In other embodiments, these include **composite promoters, which** combine promoter elements of different origins or were generated by assembling a distal enhancer with a minimal promoter of the same origin.

Exemplary cell or tissue specific promoters are provided in the tables, below, and exemplary nucleic acid sequences encoding them are known in the art and can be readily accessed using a variety of resources, such as the NCBI database, including RefSeq, as well as the Eukaryotic Promoter Database (//epd.epfl.ch//index.php).

Table 33. Exemplary cell or tissue-specific promoters

Promoter	**Target cells**
B29 Promoter	B cells
CD14 Promoter	Monocytic Cells

CD43 Promoter	Leukocytes and platelets
CD45 Promoter	Hematopoeitic cells
CD68 promoter	macrophages
Desmin promoter	muscle cells
Elastase-1 promoter	pancreatic acinar cells
Endoglin promoter	endothelial cells
fibronectin promoter	differentiating cells, healing tissue
Flt-1 promoter	endothelial cells
GFAP promoter	Astrocytes
GPIIB promoter	megakaryocytes
ICAM-2 Promoter	Endothelial cells
INF-Beta promoter	Hematopoeitic cells
Mb promoter	muscle cells
Nphs1 promoter	podocytes
OG-2 promoter	Osteoblasts, Odonblasts
SP-B promoter	Lung
Syn1 promoter	Neurons
WASP promoter	Hematopoeitic cells
SV40/bAlb promoter	Liver
SV40/bAlb promoter	Liver
SV40/Cd3 promoter	Leukocytes and platelets
SV40/CD45 promoter	hematopoeitic cells
NSE/RU5' promoter	Mature Neurons

Table 34. Additional exemplary cell or tissue-specific promoters

Promoter	**Gene Description**	**Gene Specificity**
APOA2	Apolipoprotein A-II	Hepatocytes (from hepatocyte progenitors)

SERPINA 1 (hAAT)	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1 (also named alpha 1 anti-tryps in)	Hepatocytes (from definitive endoderm stage)
CYP3A	Cytochrome P450, family 3, subfamily A, polypeptide	Mature Hepatocytes
MIR122	MicroRNA 122	Hepatocytes (from early stage embryonic liver cells) and endoderm
Pancreatic specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
INS	Insulin	Pancreatic beta cells (from definitive endoderm stage)
IRS2	Insulin receptor substrate 2	Pancreatic beta cells
Pdx1	Pancreatic and duodenal homeobox 1	Pancreas (from definitive endoderm stage)

Alx3	Aristaless-like homeobox 3	Pancreatic beta cells (from definitive endoderm stage)
Ppy	Pancreatic polypeptide	PP pancreatic cells (gamma cells)
Cardiac specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
Myh6 (aMHC)	Myosin, heavy chain 6, cardiac muscle, alpha	Late differentiation marker of cardiac muscle cells (atrial specificity)
MYL2 (MLC-2v)	Myosin, light chain 2, regulatory, cardiac, slow	Late differentiation marker of cardiac muscle cells (ventricular specificity)
ITNNl3 (cTnl)	Troponin I type 3 (cardiac)	Cardiomyocytes (from immature state)
ITNNl3 (cTnl)	Troponin I type 3 (cardiac)	Cardiomyocytes (from immature state)
NPPA (ANF)	Natriuretic peptide precursor A (also named Atrial Natriuretic Factor)	Atrial specificity in adult cells
Slc8a1 (Ncx1)	Solute carrier family 8 (sodium/calcium exchanger), member 1	Cardiomyocytes from early developmental stages
CNS specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
SYN1 (hSyn)	Synapsin I	Neurons
GFAP	Glial fibrillary acidic protein	Astrocytes

INA	lnternexin neuronal intermediate filament protein, alpha (a-internexin)	Neuroprogenitors
NES	Nestin	Neuroprogenitors and ectoderm
MOBP	Myelin-associated oligodendrocyte basic protein	Oligodendrocytes
MBP	Myelin basic protein	Oligodendrocytes
TH	Tyrosine hydroxylase	Dopaminergic neurons
FOXA2 (HNF3 beta)	Forkhead box A2	Dopaminergic neurons (also used as a marker of endoderm)
Skin specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
FLG	Filaggrin	Keratinocytes from granular layer
K14	Keratin 14	Keratinocytes from granular and basal layers
TGM3	Transglutaminase 3	Keratinocytes from granular layer
Immune cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
ITGAM (CD11B)	lntegrin, alpha M (complement component 3 receptor 3 subunit)	Monocytes, macrophages , granulocytes , natural killer cells
Urogential cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
Pbsn	Probasin	Prostatic epithelium
Upk2	Uroplakin 2	Bladder
Sbp	Spermine binding protein	Prostate
Fer1l4	Fer-1-like 4	Bladder
Endothelial cell specific promoters		

Promoter	Gene Description	Gene Specificity
ENG	Endoglin	Endothelial cells
Pluripotent and embryonic cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
Oct4 (POU5F1)	POU class 5 homeobox 1	Pluripotent cells (germ cells, ES cells, iPS cells)
NANOG	Nanog homeobox	Pluripotent cells (ES cells, iPS cells)
Synthetic Oct4	Synthetic promoter based on a Oct-4 core enhancer element	Pluripotent cells (ES cells, iPS cells)
T brachyury	Brachyury	Mesoderm
NES	Nestin	Neuroprogenitors and Ectoderm
SOX17	SRY (sex determining region Y)-box 17	Endoderm
FOXA2 (HNFJ beta)	Forkhead box A2	Endoderm (also used as a marker of dopaminergic neurons)
MIR122	MicroRNA 122	Endoderm and hepatocytes (from early stage embryonic liver cells~

Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g.,

Bitter et al. (1987) *Methods in Enzymology*, 153:516-544; incorporated herein by reference in its entirety).

In some embodiments, a nucleic acid encoding a Gene Writer or template nucleic acid is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may, in some embodiment, be functional in either a eukaryotic cell, e.g., a mammalian cell; or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a polypeptide is operably linked to multiple control elements, e.g., that allow expression of the nucleotide sequence encoding the polypeptide in both prokaryotic and eukaryotic cells.

For illustration purposes, examples of spatially restricted promoters include, but are not limited to, neuron-specific promoters, adipocyte-specific promoters, cardiomyocyte-specific promoters, smooth muscle-specific promoters, photoreceptor-specific promoters, etc. Neuron-specific spatially restricted promoters include, but are not limited to, a neuron-specific enolase (NSE) promoter (see, e.g., EMBL HSENO2, X51956); an aromatic amino acid decarboxylase (AADC) promoter, a neurofilament promoter (see, e.g., GenBank HUMNFL, L04147); a synapsin promoter (see, e.g., GenBank HUMSYNIB, M55301); a thy-1 promoter (see, e.g., Chen et al. (1987) Cell 51:7-19; and Llewellyn, et al. (2010) Nat. Med. 16(10):1161-1166); a serotonin receptor promoter (see, e.g., GenBank S62283); a tyrosine hydroxylase promoter (TH) (see, e.g., Oh et al. (2009) Gene Ther 16:437; Sasaoka et al. (1992) Mol. Brain Res. 16:274; Boundy et al. (1998) J. Neurosci. 18:9989; and Kaneda et al. (1991) Neuron 6:583-594); a GnRH promoter (see, e.g., Radovick et al. (1991) Proc. Natl. Acad. Sci. USA 88:3402-3406); an L7 promoter (see, e.g., Oberdick et al. (1990) Science 248:223-226); a DNMT promoter (see, e.g., Bartge et al. (1988) Proc. Natl. Acad. Sci. USA 85:3648-3652); an enkephalin promoter (see, e.g., Comb et al. (1988) EMBO J. 17:3793-3805); a myelin basic protein (MBP) promoter; a Ca2+-calmodulin-dependent protein kinase II-alpha (CamKIIα) promoter (see, e.g., Mayford et al. (1996) Proc. Natl. Acad. Sci. USA 93:13250; and Casanova et al. (2001) Genesis 31:37); a CMV enhancer/platelet-derived growth factor-β promoter (see, e.g., Liu et al. (2004) Gene Therapy 11:52-60); and the like.

Adipocyte-specific spatially restricted promoters include, but are not limited to, the aP2 gene promoter/enhancer, e.g., a region from −5.4 kb to +21 bp of a human aP2 gene (see, e.g., Tozzo et al. (1997) Endocrinol. 138:1604; Ross et al. (1990) Proc. Natl. Acad. Sci. USA

87:9590; and Pavjani et al. (2005) Nat. Med. 11:797); a glucose transporter-4 (GLUT4) promoter (see, e.g., Knight et al. (2003) Proc. Natl. Acad. Sci. USA 100:14725); a fatty acid translocase (FAT/CD36) promoter (see, e.g., Kuriki et al. (2002) Biol. Pharm. Bull. 25:1476; and Sato et al. (2002) J. Biol. Chem. 277:15703); a stearoyl-CoA desaturase-1 (SCD1) promoter (Tabor et al. (1999) J. Biol. Chem. 274:20603); a leptin promoter (see, e.g., Mason et al. (1998) Endocrinol. 139:1013; and Chen et al. (1999) Biochem. Biophys. Res. Comm. 262:187); an adiponectin promoter (see, e.g., Kita et al. (2005) Biochem. Biophys. Res. Comm. 331:484; and Chakrabarti (2010) Endocrinol. 151:2408); an adipsin promoter (see, e.g., Platt et al. (1989) Proc. Natl. Acad. Sci. USA 86:7490); a resistin promoter (see, e.g., Seo et al. (2003) Molec. Endocrinol. 17:1522); and the like.

Cardiomyocyte-specific spatially restricted promoters include, but are not limited to, control sequences derived from the following genes: myosin light chain-2, α-myosin heavy chain, AE3, cardiac troponin C, cardiac actin, and the like. Franz et al. (1997) Cardiovasc. Res. 35:560-566; Robbins et al. (1995) Ann. N.Y. Acad. Sci. 752:492-505; Linn et al. (1995) Circ. Res. 76:584-591; Parmacek et al. (1994) Mol. Cell. Biol. 14:1870-1885; Hunter et al. (1993) Hypertension 22:608-617; and Sartorelli et al. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051.

Smooth muscle-specific spatially restricted promoters include, but are not limited to, an SM22α promoter (see, e.g., Akyürek et al. (2000) Mol. Med. 6:983; and U.S. Pat. No. 7,169,874); a smoothelin promoter (see, e.g., WO 2001/018048); an α-smooth muscle actin promoter; and the like. For example, a 0.4 kb region of the SM22α promoter, within which lie two CArG elements, has been shown to mediate vascular smooth muscle cell-specific expression (see, e.g., Kim, et al. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, et al., (1996) J. Cell Biol. 132, 849-859; and Moessler, et al. (1996) Development 122, 2415-2425).

Photoreceptor-specific spatially restricted promoters include, but are not limited to, a rhodopsin promoter; a rhodopsin kinase promoter (Young et al. (2003) Ophthalmol. Vis. Sci. 44:4076); a beta phosphodiesterase gene promoter (Nicoud et al. (2007) J. Gene Med. 9:1015); a retinitis pigmentosa gene promoter (Nicoud et al. (2007) supra); an interphotoreceptor retinoid-binding protein (IRBP) gene enhancer (Nicoud et al. (2007) supra); an IRBP gene promoter (Yokoyama et al. (1992) Exp Eye Res. 55:225); and the like.

Nonlimiting Exemplary Cells-Specific Promoters

Cell-specific promoters known in the art may be used to direct expression of a Gene Writer protein, e.g., as described herein. Nonlimiting exemplary mammalian cell-specific promoters have been characterized and used in mice expressing Cre recombinase in a cell-specific manner. Certain nonlimiting exemplary mammalian cell-specific promoters are listed in Table 1 of US9845481, incorporated herein by reference.

In some embodiments, a cell-specific promoters is a promoter that is active in plants. Many exemplary cell-specific plant promoters are known in the art. See, e.g., U.S. Pat. Nos. 5,097,025; 5,783,393; 5,880,330; 5,981,727; 7,557,264; 6,291,666; 7,132,526; and 7,323,622; and U.S. Publication Nos. 2010/0269226; 2007/0180580; 2005/0034192; and 2005/0086712, which are incorporated by reference herein in their entireties for any purpose.

In some embodiments, a vector as described herein comprises an expression cassette. The term "expression cassette", as used herein, refers to a nucleic acid construct comprising nucleic acid elements sufficient for the expression of the nucleic acid molecule of the instant invention. Typically, an expression cassette comprises the nucleic acid molecule of the instant invention operatively linked to a promoter sequence. The term"operatively linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operatively linked with a coding sequence when it is capable of affecting the expression of that coding sequence (e.g., the coding sequence is under the transcriptional control of the promoter). Encoding sequences can be operatively linked to regulatory sequences in sense or antisense orientation. In certain embodiments, the promoter is a heterologous promoter. The term"heterologous promoter", as used herein, refers to a promoter that is not found to be operatively linked to a given encoding sequence in nature. In certain embodiments, an expression cassette may comprise additional elements, for example, an intron, an enhancer, a polyadenylation site, a woodchuck response element (WRE), and/or other elements known to affect expression levels of the encoding sequenceA"promoter" typically controls the expression of a coding sequence or functional RNA. In certain embodiments, a promoter sequence comprises proximal and more distal upstream elements and can further comprise an enhancer element. An "enhancer" can typically stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. In certain embodiments, the promoter is derived in its entirety from a native gene. In certain embodiments,

the promoter is composed of different elements derived from different naturally occurring promoters. In certain embodiments, the promoter comprises a synthetic nucleotide sequence. It will be understood by those skilled in the art that different promoters will direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions or to the presence or the absence of a drug or transcriptional co-factor. Ubiquitous, cell-type-specific, tissue-specific, developmental stage-specific, and conditional promoters, for example, drug-responsive promoters (e.g ., tetracycline-responsive promoters) are well known to those of skill in the art. Examples of promoter include, but are not limited to, the phosphoglycerate kinase (PKG) promoter, CAG (composite of the CMV enhancer the chicken beta actin promoter (CBA) and the rabbit beta globin intron.), NSE (neuronal specific enolase), synapsin or NeuN promoters, the SV40 early promoter, mouse mammary tumor virus LTR promoter; adenovirus major late promoter (Ad MLP); a herpes simplex virus (HSV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE), SFFV promoter, rous sarcoma virus (RSV) promoter, synthetic promoters, hybrid promoters, and the like. Other promoters can be of human origin or from other species, including from mice. Common promoters include, e.g., the human cytomegalovirus (CMV) immediate early gene promoter, the SV40 early promoter, the Rous sarcoma virus long terminal repeat, [beta]- actin, rat insulin promoter, the phosphoglycerate kinase promoter, the human alpha- 1 antitrypsin (hAAT) promoter, the transthyretin promoter, the TBG promoter and other liver-specific promoters, the desmin promoter and similar muscle-specific promoters, the EF1 -alpha promoter, the CAG promoter and other constitutive promoters, hybrid promoters with multi-tissue specificity, promoters specific for neurons like synapsin and glyceraldehyde-3 - phosphate dehydrogenase promoter, all of which are promoters well known and readily available to those of skill in the art, can be used to obtain high-level expression of the coding sequence of interest. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, will also find use herein. Such promoter sequences are commercially available from, e.g., Stratagene (San Diego, CA). Additional exemplary promoter sequences are described, for example, in WO2018213786A1 (incorporated by reference herein in its entirety).

In some embodiments, the apolipoprotein E enhancer (ApoE) or a functional fragment thereof is used, e.g., to drive expression in the liver. In some embodiments, two copies of the ApoE enhancer or a functional fragment thereof is used. In some embodiments, the ApoE

enhancer or functional fragment thereof is used in combination with a promoter, e.g., the human alpha-1 antitrypsin (hAAT) promoter.

In some embodiments, the regulatory sequences impart tissue-specific gene expression capabilities. In some cases, the tissue-specific regulatory sequences bind tissue-specific transcription factors that induce transcription in a tissue specific manner. Various tissue-specific regulatory sequences (e.g., promoters, enhancers, etc.) are known in the art. Exemplary tissue-specific regulatory sequences include, but are not limited to, the following tissue-specific promoters: a liver-specific thyroxin binding globulin (TBG) promoter, a insulin promoter, a glucagon promoter, a somatostatin promoter, a pancreatic polypeptide (PPY) promoter, a synapsin-1 (Syn) promoter, a creatine kinase (MCK) promoter, a mammalian desmin (DES) promoter, a α-myosin heavy chain (a-MHC) promoter, or a cardiac Troponin T (cTnT) promoter. Other exemplary promoters include Beta-actin promoter, hepatitis B virus core promoter, Sandig et al., Gene Ther., 3:1002-9 (1996); alpha-fetoprotein (AFP) promoter, Arbuthnot et al., Hum. Gene Ther., 7:1503-14 (1996)), bone osteocalcin promoter (Stein et al., Mol. Biol. Rep., 24:185-96 (1997)); bone sialoprotein promoter (Chen et al., J. Bone Miner. Res., 11:654-64 (1996)), CD2 promoter (Hansal et al., J. Immunol., 161:1063-8 (1998); immunoglobulin heavy chain promoter; T cell receptor α-chain promoter, neuronal such as neuron-specific enolase (NSE) promoter (Andersen et al., Cell. Mol. Neurobiol., 13:503-15 (1993)), neurofilament light-chain gene promoter (Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991)), and the neuron-specific vgf gene promoter (Piccioli et al., Neuron, 15:373-84 (1995)), and others. Additional exemplary promoter sequences are described, for example, in U.S. Patent No. 10300146 (incorporated herein by reference in its entirety). In some embodiments, a tissue-specific regulatory element, e.g., a tissue-specific promoter, is selected from one known to be operably linked to a gene that is highly expressed in a given tissue, e.g., as measured by RNA-seq or protein expression data, or a combination thereof. Methods for analyzing tissue specificity by expression are taught in Fagerberg et al. *Mol Cell Proteomics* 13(2):397-406 (2014), which is incorporated herein by reference in its entirety.

In some embodiments, a vector described herein is a multicistronic expression construct. Multicistronic expression constructs include, for example, constructs harboring a first expression cassette, e.g. comprising a first promoter and a first encoding nucleic acid sequence, and a second expression cassette, e.g. comprising a second promoter and a second encoding nucleic

acid sequence. Such multicistronic expression constructs may, in some instances, be particularly useful in the delivery of non-translated gene products, such as hairpin RNAs, together with a polypeptide, for example, a gene writer and gene writer template. In some embodiments, multicistronic expression constructs may exhibit reduced expression levels of one or more of the included transgenes, for example, because of promoter interference or the presence of incompatible nucleic acid elements in close proximity. If a multicistronic expression construct is part of a viral vector, the presence of a self-complementary nucleic acid sequence may, in some instances, interfere with the formation of structures necessary for viral reproduction or packaging.

In some embodiments, the sequence encodes an RNA with a hairpin. In some embodiments, the hairpin RNA is an a guide RNA, a template RNA, shRNA, or a microRNA. In some embodiments, the first promoter is an RNA polymerase I promoter. In some embodiments, the first promoter is an RNA polymerase II promoter. In some embodiments, the second promoter is an RNA polymerase III promoter. In some embodiments, the second promoter is a U6 or H1 promoter. In some embodiments, the nucleic acid construct comprises the structure of AAV construct B1 or B2.

Without wishing to be bound by theory, multicistronic expression constructs may not achieve optimal expression levels as compared to expression systems containing only one cistron. One of the suggested causes of lower expression levels achieved with multicistronic expression constructs comprising two ore more promoter elements is the phenomenon of promoter interference (see, e.g., Curtin J A, Dane A P, Swanson A, Alexander I E, Ginn S L. *Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct*. Gene Ther. 2008 March; 15(5):384-90; and Martin-Duque P, Jezzard S, Kaftansis L, Vassaux G. *Direct comparison of the insulating properties of two genetic elements in an adenoviral vector containing two different expression cassettes*. Hum Gene Ther. 2004 October; 15(10):995-1002; both references incorporated herein by reference for disclosure of promoter interference phenomenon). In some embodiments, the problem of promoter interference may be overcome, e.g., by producing multicistronic expression constructs comprising only one promoter driving transcription of multiple encoding nucleic acid sequences separated by internal ribosomal entry sites, or by separating cistrons comprising their own promoter with transcriptional insulator elements. In some embodiments, single-promoter driven

expression of multiple cistrons may result in uneven expression levels of the cistrons. In some embodiments, a promoter cannot efficiently be isolated and isolation elements may not be compatible with some gene transfer vectors, for example, some retroviral vectors.

MicroRNAs

miRNAs and other small interfering nucleic acids generally regulate gene expression via target RNA transcript cleavage/degradation or translational repression of the target messenger RNA (mRNA). miRNAs may, in some instances, be natively expressed, typically as final 19-25 non-translated RNA products. miRNAs generally exhibit their activity through sequence-specific interactions with the 3′ untranslated regions (UTR) of target mRNAs. These endogenously expressed miRNAs may form hairpin precursors that are subsequently processed into an miRNA duplex, and further into a mature single stranded miRNA molecule. This mature miRNA generally guides a multiprotein complex, miRISC, which identifies target 3′ UTR regions of target mRNAs based upon their complementarity to the mature miRNA. Useful transgene products may include, for example, miRNAs or miRNA binding sites that regulate the expression of a linked polypeptide. A non-limiting list of miRNA genes; the products of these genes and their homologues are useful as transgenes or as targets for small interfering nucleic acids (e.g., miRNA sponges, antisense oligonucleotides), e.g., in methods such as those listed in US10300146, 22:25-25:48, incorporated by reference. In some embodiments, one or more binding sites for one or more of the foregoing miRNAs are incorporated in a transgene, e.g., a transgene delivered by a rAAV vector, e.g., to inhibit the expression of the transgene in one or more tissues of an animal harboring the transgene. In some embodiments, a binding site may be selected to control the expression of a trangene in a tissue specific manner. For example, binding sites for the liver-specific miR-122 may be incorporated into a transgene to inhibit expression of that transgene in the liver. Additional exemplary miRNA sequences are described, for example, in U.S. Patent No. 10300146 (incorporated herein by reference in its entirety). For liver-specific Gene Writing, however, overexpression of miR-122 may be utilized instead of using binding sites to effect miR-122-specific degradation. This miRNA is positively associated with hepatic differentiation and maturation, as well as enhanced expression of liver specific genes. Thus, in some embodiments, the coding sequence for miR-122 may be added to a component of a Gene Writing system to enhance a liver-directed therapy.

A miR inhibitor or miRNA inhibitor is generally an agent that blocks miRNA expression and/or processing. Examples of such agents include, but are not limited to, microRNA antagonists, microRNA specific antisense, microRNA sponges, and microRNA oligonucleotides (double-stranded, hairpin, short oligonucleotides) that inhibit miRNA interaction with a Drosha complex. MicroRNA inhibitors, e.g., miRNA sponges, can be expressed in cells from transgenes (e.g., as described in Ebert, M. S. Nature Methods, Epub Aug. 12, 2007; incorporated by reference herein in its entirety). In some embodiments, microRNA sponges, or other miR inhibitors, are used with the AAVs. microRNA sponges generally specifically inhibit miRNAs through a complementary heptameric seed sequence. In some embodiments, an entire family of miRNAs can be silenced using a single sponge sequence. Other methods for silencing miRNA function (derepression of miRNA targets) in cells will be apparent to one of ordinary skill in the art.

In some embodiments, a miRNA as described herein comprises a sequence listed in Table 4 of PCT Publication No. WO2020014209, incorporated herein by reference. Also incorporated herein by reference are the listing of exemplary miRNA sequences from WO2020014209.

In some embodiments, it is advantageous to silence one or more components of a Gene Writing system (e.g., mRNA encoding a Gene Writer polypeptide, a Gene Writer Template RNA, or a heterologous object sequence expressed from the genome after successful Gene Writing) in a portion of cells. In some embodiments, it is advantageous to restrict expression of a component of a Gene Writing system to select cell types within a tissue of interest.

For example, it is known that in a given tissue, e.g., liver, macrophages and immune cells, e.g., Kupffer cells in the liver, may engage in uptake of a delivery vehicle for one or more components of a Gene Writing system. In some embodiments, at least one binding site for at least one miRNA highly expressed in macrophages and immune cells, e.g., Kupffer cells, is included in at least one component of a Gene Writing system, e.g., nucleic acid encoding a Gene Writing polypeptide or a transgene. In some embodiments, a miRNA that targets the one or more binding sites is listed in a table referenced herein, e.g., miR-142, e.g., mature miRNA hsa-miR-142-5p or hsa-miR-142-3p.

In some embodiments, there may be a benefit to decreasing Gene Writer levels and/or Gene Writer activity in cells in which Gene Writer expression or overexpression of a transgene

may have a toxic effect. For example, it has been shown that delivery of a transgene overexpression cassette to dorsal root ganglion neurons may result in toxicity of a gene therapy (see Hordeaux et al Sci Transl Med 12(569):eaba9188 (2020), incorporated herein by reference in its entirety). In some embodiments, at least one miRNA binding site may be incorporated into a nucleic acid component of a Gene Writing system to reduce expression of a system component in a neuron, e.g., a dorsal root ganglion neuron. In some embodiments, the at least one miRNA binding site incorporated into a nucleic acid component of a Gene Writing system to reduce expression of a system component in a neuron is a binding site of miR-182, e.g., mature miRNA hsa-miR-182-5p or hsa-miR-182-3p. In some embodiments, the at least one miRNA binding site incorporated into a nucleic acid component of a Gene Writing system to reduce expression of a system component in a neuron is a binding site of miR-183, e.g., mature miRNA hsa-miR-183-5p or hsa-miR-183-3p. In some embodiments, combinations of miRNA binding sites may be used to enhance the restriction of expression of one or more components of a Gene Writing system to a tissue or cell type of interest.

Table A5 below provides exemplary miRNAs and corresponding expressing cells, e.g., a miRNA for which one can, in some embodiments, incorporate binding sites (complementary sequences) in the transgene or polypeptide nucleic acid, e.g., to decrease expression in that off-target cell.

Table A5: Exemplary miRNA from off-target cells and tissues

Silenced cell type	miRNA name	Mature miRNA	miRNA sequence
Kupffer cells	miR-142	hsa-miR-142-5p	cauaaaguagaaagcacuacu
Kupffer cells	miR-142	hsa-miR-142-3p	uguaguguuuccuacuuuaugga
Dorsal root ganglion neurons	miR-182	hsa-miR-182-5p	uuuggcaaugguagaacucacacu
Dorsal root ganglion neurons	miR-182	hsa-miR-182-3p	ugguucuagacuugccaacua
Dorsal root ganglion neurons	miR-183	hsa-miR-183-5p	uauggcacugguagaauucacu
Table XDorsal root ganglion neurons	miR-183	hsa-miR-183-3p	gugaauuaccgaagggccauaa
Hepatocytes	miR-122	hsa-miR-122-5p	uggagugugacaaugguguuug
Hepatocytes	miR-122	hsa-miR-122-3p	aacgccauuaucacacuaaaua

Anticrispr systems for regulating GeneWriter activity

Various approaches for modulating Cas molecule activity may be used in conjunction with the systems and methods described herein. For instance, in some embodiments, a polypeptide described herein (e.g., a Cas molecule or a GeneWriter comprising a Cas domain) can be regulated using an anticrispr agent (e.g., an anticrispr protein or anticrispr small molecule). In some embodiments, the Cas molecule or Cas domain comprises a responsive intein such as, for example, a 4-hydroxytamoxifen (4-HT)-responsive intein, an iCas molecule (e.g., iCas9); a 4-HT-responsive Cas (e.g., allosterically regulated Cas9 (arC9) or dead Cas9 (dC9)). The systems and methods described herein can also utilize a chemically-induced dimerization system of split protein fragments (e.g., rapamycin-mediated dimerization of FK506 binding protein 12 (FKBP) and FKBP rapamycin binding domain (FRB), an abscisic acid-inducible ABI-PYL1 and gibberellin-inducible GID1-GAI heterodimerization domains); a dimer of BCL-xL peptide and BH3 peptides, a A385358 (A3) small molecule, a degron system (e.g., a FKBP-Cas9 destabilized system, an auxin-inducible degron (AID) or an E.coliDHFR degron system), an aptamer or aptazyme fused with gRNA (e.g., tetracycline- and theophylline-responsive bioswitches), AcrIIA2 and AcrIIA4 proteins, and BRD0539.

In some embodiments, a small molecule-responsive intein (e.g., 4-hydroxytamoxifen (4-HT)-responsive intein) is inserted at specific sites within a Cas molecule (e.g., Cas9). In some embodiments, the insertion of a 4HT-responsive intein disrupts Cas9 enzymatic activity. In some embodiments, a Cas molecule (e.g., iCas9) is fused to the hormone binding domain of the estrogen receptor (ERT2). In some embodiments, the ligand binding domain of the human estrogen receptor-α can be inserted into a Cas molecule (e.g., Cas9 or dead Cas9 (dC9)), e.g., at position 231, yielding a 4HT-responsive anticrispr Cas9 (e.g., arC9 or dC9). In some embodiments, dCas9 can provide 4-HT dose-dependent repression of Cas9 function. In some embodiments, arC9 can provide 4-HT dose-dependent control of Cas9 function. In some embodiments, a Cas molecule (e.g., Cas9) is fused to split protein fragments. In some embodiments, chemically-induced dimerization of split protein fragments (e.g., rapamycin-mediated dimerization of FK506 binding protein 12 (FKBP) and FKBP rapamycin binding

domain (FRB)) can induce low levels of Cas9 molecule activity. In some embodiments, a chemically-induced dimerization system (e.g., abscisic acid-inducible ABI-PYL1 and gibberellin-inducible GID1-GAI heterodimerization domains) can induce a dose-dependent and reversible transcriptional activation/repression of Cas9. In some embodiments, a Cas9 inducible system (ciCas9) comprises the replacement of a Cas molecule (e.g., Cas9) REC2 domain with a BCL-xl peptide and attachment of a BH3 peptide to the N- and C-termini of the modified Cas9.BCL. In some embodiments, the interaction between BCL-xL and BH3 peptides can keep Cas9 in an inactive state. In some embodiments, a small molecule (e.g., A-385358 (A3)) can disrupt the interaction between BLC-xl and BH3 peptides to activate Cas9. In some embodiments, a Cas9 inducible system can exhibit dose-dependent control of nuclease activity. In some embodiments, a degron system can induce degradation of a Cas molecule (e.g., Cas9) upon activation or deactivation by an external factor (e.g., small-molecule ligand, light, temperature, or a protein). In some embodiments, a small molecule BRD0539 inhibits a Cas molecule (e.g., Cas9) reversibly. Additional information on anticrispr proteins or anticrispr small molecules can be found, for example, in Gangopadhyay, S.A. et al. Precision control of CRISPR-Cas9 using small molecules and light, Biochemistry, 2019, Maji, B. et al. A high-throughput platform to identify small molecule inhibitors of CRISPR-Cas9, and Pawluk Anti-CRISPR: discovery, mechanism and function Nature Reviews Microbiology volume 16, pages12–17(2018), each of which is incorporated by reference in its entirety.

Self-inactivating modules for regulating GeneWriter activity

In some embodiments the Gene Writer systems described herein includes a self-inactivating module. The self-inactivating module leads to a decrease of expression of the Gene Writer polypeptide, the Gene Writer template, or both. Without wishing to be bound by the theory, the self-inactivating module provides for a temporary period of Gene Writer expression prior to inactivation. Without wishing to be bound by theory, the activity of the Gene Writer polypeptide at a target site introduces a mutation (e.g. a substitution, insertion, or deletion) into the DNA encoding the Gene Writer polypeptide or Gene Writer template which results in a decrease of Gene Writer polypeptide or template expression. In some embodiments of the self-inactivating module, a target site for the Gene Writer polypeptide is included in the DNA encoding the Gene Writer polypeptide or Gene Writer template. In some embodiments, one, two,

three, four, five, or more copies of the target site are included in the DNA encoding the Gene Writer polypeptide or Gene Writer template. In some embodiments, the target site in the DNA encoding the Gene Writer polypeptide or Gene Writer template is the same target site as the target site on the genome. In some embodiments, the target site is a different target site than the target site on the genome. In some embodiments, the self-inactivation module target site uses the same or a different template RNA or guide RNA as the genome target site. In some embodiments, the target site is modified via target primed reverse transcription based upon a template RNA. In some embodiments the target side is nicked. The target site may be incorporated into an enhancer, a promoter, an untranslated region, an exon, an intron, an open reading frame, or a stuffer sequence.

In some embodiments, upon inactivation, the decrease of expression is 25%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 99.9%, or more lower than a Gene Writing system that does not contain the self-inactivating module. In some embodiments, a Gene Writer system that contains the self-inactivating module has a 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% 99%, or higher rate of integrations in target sites than off-target sites compared to a Gene Writing system that does not contain the self-inactivation module. a Gene Writer system that contains the self-inactivating module has a 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% 99%, or higher efficiency of target site modification compared to a Gene Writing system that does not contain the self-inactivation module. In some embodiments, the self-inactivating module is included when the Gene Writer polypeptide is delivered as DNA, e.g. via a viral vector.

Self-inactivating modules have been described for nucleases. See, e.g. in Li et al A Self-Deleting AAV-CRISPR System for In Vivo Genome Editing, Mol Ther Methods Clin Dev. 2019 Mar 15; 12: 111–122, P. Singhal, Self-Inactivating Cas9: a method for reducing exposure while maintaining efficacy in virally delivered Cas9 applications (available at www.editasmedicine.com/wp-content/uploads/2019/10/aef_asgct_poster_2017_final_-_present_5-11-17_515pm1_1494537387_1494558495_1497467403.pdf), and Epstein and Schaffer Engineering a Self-Inactivating CRISPR System for AAV Vectors Targeted Genome Editing IlVolume 24, SUPPLEMENT 1, S50, May 01, 2016, *and* WO2018106693A1.

Small Molecules

In some embodiments a polypeptide described herein (e.g., a Gene Writer polypeptide) is controllable via a small molecule. In some embodiments the polypeptide is dimerized via a small molecule.

In some embodiment, the polypeptide is controllable via Chemical Induction of Dimerization (CID) with small molecules. CID is generally used to generate switches of protein function to alter cell physiology. An exemplary high specificity, efficient dimerizer is rimiducid (AP1903), which has two identical, protein-binding surfaces arranged tail-to-tail, each with high affinity and specificity for a mutant of FKBP12: FKBP12(F36V) (FKBP12v36, F_{V36} or F_V), Attachment of one or more Fv domains onto one or more cell signaling molecules that normally rely on homodimerization can convert that protein to rimiducid control. Homodimerization with rimiducid is used in the context of an inducible caspase safety switch. This molecular switch that is controlled by a distinct dimerizer ligand, based on the heterodimerizing small molecule, rapamycin, or rapamycin analogs ("rapalogs"). Rapamycin binds to FKBP12, and its variants, and can induce heterodimerization of signaling domains that are fused to FKBP12 by binding to both FKBP12 and to polypeptides that contain the FKBP-rapamycin-binding (FRB) domain of mTOR. Provided in some embodiments of the present application are molecular switches that greatly augment the use of rapamycin, rapalogs and rimiducid as agents for therapeutic applications.

In some embodiments of the dual switch technology, a homodimerizer, such as AP1903 (rimiducid), directly induces dimerization or multimerization of polypeptides comprising an FKBP12 multimerizing region. In other embodiments, a polypeptide comprising an FKBP12 multimerization is multimerized, or aggregated by binding to a heterodimerizer, such as rapamycin or a rapalog, which also binds to an FRB or FRB variant multimerizing region on a chimeric polypeptide, also expressed in the modified cell, such as, for example, a chimeric antigen receptor. Rapamycin is a natural product macrolide that binds with high affinity (<1 nM) to FKBP12 and together initiates the high-affinity, inhibitory interaction with the FKBP-Rapamycin-Binding (FRB) domain of mTOR. FRB is small (89 amino acids) and can thereby be used as a protein "tag" or "handle" when appended to many proteins. Coexpression of a FRB-fused protein with a FKBP12-fused protein renders their approximation rapamycin-inducible (12-16). This can serve as the basis for a cell safety switch regulated by the orally available ligand, rapamycin, or derivatives of rapamycin (rapalogs) that do not inhibit mTOR at a low,

therapeutic dose but instead bind with selected, Caspase-9-fused mutant FRB domains. (see Sabatini D M, et al., Cell. 1994; 78(1):35-43; Brown E J, et al., Nature. 1994; 369(6483):756-8; Chen J, et al., Proc Natl Acad Sci USA. 1995; 92(11):4947-51; and Choi J, Science. 1996; 273(5272):239-42).

In some embodiments, two levels of control are provided in the therapeutic cells. In embodiments, the first level of control may be tunable, i.e., the level of removal of the therapeutic cells may be controlled so that it results in partial removal of the therapeutic cells. In some embodiments, the chimeric antigen polypeptide comprises a binding site for rapamycin, or a rapamycin analog. In embodiments, also present in the therapeutic cell is a suicide gene, such as, for example, one encoding a caspase polypeptide. Using this controllable first level, the need for continued therapy may, in some embodiments, be balanced with the need to eliminate or reduce the level of negative side effects. In some embodiments, a rapamycin analog, a rapalog is administered to the patient, which then binds to both the caspase polypeptide and the chimeric antigen receptor, thus recruiting the caspase polypeptide to the location of the CAR, and aggregating the caspase polypeptide. Upon aggregation, the caspase polypeptide induces apoptosis. The amount of rapamycin or rapamycin analog administered to the patient may vary; if the removal of a lower level of cells by apoptosis is desired in order to reduce side effects and continue CAR therapy, a lower level of rapamycin or rapamycin may be administered to the patient. In some embodiments, the second level of control may be designed to achieve the maximum level of cell elimination. This second level may be based, for example, on the use of rimiducid, or AP1903. If there is a need to rapidly eliminate up to 100% of the therapeutic cells, the AP1903 may be administered to the patient. The multimeric AP1903 binds to the caspase polypeptide, leading to multimerization of the caspase polypeptide and apoptosis. In certain examples, second level may also be tunable, or controlled, by the level of AP1903 administered to the subject.

In certain embodiments, small molecules can be used to control genes, as described in for example, US10584351 at 47:53-56:47 (incorporated by reference herein in its entirety), together suitable ligands for the control features, e.g., in US10584351 at 56:48, et seq. as well as U10046049 at 43:27-52:20, incorporated by reference as well as the description of ligands for such control systems at 52:21, et seq.

Chemically modified nucleic acids and nucleic acid end features

A nucleic acid described herein (e.g., a template nucleic acid, e.g., a template RNA; or a nucleic acid (e.g., mRNA) encoding a GeneWriter) can comprise unmodified or modified nucleobases. Naturally occurring RNAs are synthesized from four basic ribonucleotides: ATP, CTP, UTP and GTP, but may contain post-transcriptionally modified nucleotides. Further, approximately one hundred different nucleoside modifications have been identified in RNA (Rozenski, J, Crain, P, and McCloskey, J. (1999). The RNA Modification Database: 1999 update. Nucl Acids Res 27: 196-197). An RNA can also comprise wholly synthetic nucleotides that do not occur in nature.

In some embodiments, the chemically modification is one provided in PCT/US2016/032454, US Pat. Pub. No. 20090286852, of International Application No. WO/2012/019168, WO/2012/045075, WO/2012/135805, WO/2012/158736, WO/2013/039857, WO/2013/039861, WO/2013/052523, WO/2013/090648, WO/2013/096709, WO/2013/101690, WO/2013/106496, WO/2013/130161, WO/2013/151669, WO/2013/151736, WO/2013/151672, WO/2013/151664, WO/2013/151665, WO/2013/151668, WO/2013/151671, WO/2013/151667, WO/2013/151670, WO/2013/151666, WO/2013/151663, WO/2014/028429, WO/2014/081507, WO/2014/093924, WO/2014/093574, WO/2014/113089, WO/2014/144711, WO/2014/144767, WO/2014/144039, WO/2014/152540, WO/2014/152030, WO/2014/152031, WO/2014/152027, WO/2014/152211, WO/2014/158795, WO/2014/159813, WO/2014/164253, WO/2015/006747, WO/2015/034928, WO/2015/034925, WO/2015/038892, WO/2015/048744, WO/2015/051214, WO/2015/051173, WO/2015/051169, WO/2015/058069, WO/2015/085318, WO/2015/089511, WO/2015/105926, WO/2015/164674, WO/2015/196130, WO/2015/196128, WO/2015/196118, WO/2016/011226, WO/2016/011222, WO/2016/011306, WO/2016/014846, WO/2016/022914, WO/2016/036902, WO/2016/077125, or WO/2016/077123, each of which is herein incorporated by reference in its entirety. It is understood that incorporation of a chemically modified nucleotide into a polynucleotide can result in the modification being incorporated into a nucleobase, the backbone, or both, depending on the location of the modification in the nucleotide. In some embodiments, the backbone modification is one provided in EP 2813570, which is herein incorporated by reference in its entirety. In some embodiments, the modified cap is one provided in US Pat. Pub. No. 20050287539, which is herein incorporated by reference in its entirety.

In some embodiments, the chemically modified nucleic acid (e.g., RNA, e.g., mRNA) comprises one or more of ARCA: anti-reverse cap analog (m27.3'-OGP3G), GP3G (Unmethylated Cap Analog), m7GP3G (Monomethylated Cap Analog), m32.2.7GP3G (Trimethylated Cap Analog), m5CTP (5'-methyl-cytidine triphosphate), m6ATP (N6-methyl-adenosine-5'-triphosphate), s2UTP (2-thio-uridine triphosphate), and Ψ (pseudouridine triphosphate).

In some embodiments, the chemically modified nucleic acid comprises a 5' cap, e.g.: a 7-methylguanosine cap (e.g., a O-Me-m7G cap); a hypermethylated cap analog; an NAD+-derived cap analog (e.g., as described in Kiledjian, Trends in Cell Biology 28, 454-464 (2018)); or a modified, e.g., biotinylated, cap analog (e.g., as described in Bednarek et al., Phil Trans R Soc B 373, 20180167 (2018)).

In some embodiments, the chemically modified nucleic acid comprises a 3' feature selected from one or more of: a polyA tail; a 16-nucleotide long stem-loop structure flanked by unpaired 5 nucleotides (e.g., as described by Mannironi et al., Nucleic Acid Research 17, 9113-9126 (1989)); a triple-helical structure (e.g., as described by Brown et al., PNAS 109, 19202-19207 (2012)); a tRNA, Y RNA, or vault RNA structure (e.g., as described by Labno et al., Biochemica et Biophysica Acta 1863, 3125-3147 (2016)); incorporation of one or more deoxyribonucleotide triphosphates (dNTPs), 2'O-Methylated NTPs, or phosphorothioate-NTPs; a single nucleotide chemical modification (e.g., oxidation of the 3' terminal ribose to a reactive aldehyde followed by conjugation of the aldehyde-reactive modified nucleotide); or chemical ligation to another nucleic acid molecule.

In some embodiments, the the nucleic acid (e.g., template nucleic acid) comprises one or more modified nucleotides, e.g., selected from dihydrouridine, inosine, 7-methylguanosine, 5-methylcytidine (5mC), 5′ Phosphate ribothymidine, 2′-O-methyl ribothymidine, 2′-O-ethyl ribothymidine, 2′-fluoro ribothymidine, C-5 propynyl-deoxycytidine (pdC), C-5 propynyl-deoxyuridine (pdU), C-5 propynyl-cytidine (pC), C-5 propynyl-uridine (pU), 5-methyl cytidine, 5-methyl uridine, 5-methyl deoxycytidine, 5-methyl deoxyuridine methoxy, 2,6-diaminopurine, 5′-Dimethoxytrityl-N4-ethyl-2′-deoxycytidine, C-5 propynyl-f-cytidine (pfC), C-5 propynyl-f-uridine (pfU), 5-methyl f-cytidine, 5-methyl f-uridine, C-5 propynyl-m-cytidine (pmC), C-5 propynyl-f-uridine (pmU), 5-methyl m-cytidine, 5-methyl m-uridine, LNA (locked nucleic acid),

MGB (minor groove binder) pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), or 5-methoxyuridine (5-MO-U).

In some embodiments, the nucleic acid comprises a backbone modification, e.g., a modification to a sugar or phosphate group in the backbone. In some embodiments, the nucleic acid comprises a nucleobase modification.

In some embodiments, the nucleic acid comprises one or more chemically modified nucleotides of Table M1, one or more chemical backbone modifications of Table M2, one or more chemically modified caps of Table M3. For instance, in some embodiments, the nucleic acid comprises two or more (e.g., 3, 4, 5, 6, 7, 8, 9, or 10 or more) different types of chemical modifications. As an example, the nucleic acid may comprise two or more (e.g., 3, 4, 5, 6, 7, 8, 9, or 10 or more) different types of modified nucleobases, e.g., as described herein, e.g., in Table M1. Alternatively or in combination, the nucleic acid may comprise two or more (e.g., 3, 4, 5, 6, 7, 8, 9, or 10 or more) different types of backbone modifications, e.g., as described herein, e.g., in Table M2. Alternatively or in combination, the nucleic acid may comprise one or more modified cap, e.g., as described herein, e.g., in Table M3. For instance, in some embodiments, the nucleic acid comprises one or more type of modified nucleobase and one or more type of backbone modification; one or more type of modified nucleobase and one or more modified cap; one or more type of modified cap and one or more type of backbone modification; or one or more type of modified nucleobase, one or more type of backbone modification, and one or more type of modified cap.

In some embodiments, the nucleic acid comprises one or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, or more) modified nucleobases. In some embodiments, all nucleobases of the nucleic acid are modified. In some embodiments, the nucleic acid is modified at one or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, or more) positions in the backbone. In some embodiments, all backbone positions of the nucleic acid are modified.

Table M1. Modified nucleotides

5-aza-uridine	N2-methyl-6-thio-guanosine
2-thio-5-aza-midine	N2,N2-dimethyl-6-thio-guanosine
2-thiouridine	pyridin-4-one ribonucleoside

4-thio-pseudouridine	2-thio-5-aza-uridine
2-thio-pseudouridine	2-thiomidine
5-hydroxyuridine	4-thio-pseudomidine
3-methyluridine	2-thio-pseudowidine
5-carboxymethyl-uridine	3-methylmidine
1-carboxymethyl-pseudouridine	1-propynyl-pseudomidine
5-propynyl-uridine	1-methyl-1-deaza-pseudomidine
1-propynyl-pseudouridine	2-thio-1-methyl-1-deaza-pseudouridine
5-taurinomethyluridine	4-methoxy-pseudomidine
1-taurinomethyl-pseudouridine	5'-O-(1-Thiophosphate)-Adenosine
5-taurinomethyl-2-thio-uridine	5'-O-(1-Thiophosphate)-Cytidine
1-taurinomethyl-4-thio-uridine	5'-O-(1-thiophosphate)-Guanosine
5-methyl-uridine	5'-O-(1-Thiophophate)-Uridine
1-methyl-pseudouridine	5'-O-(1-Thiophosphate)-Pseudouridine
4-thio-1-methyl-pseudouridine	2'-O-methyl-Adenosine
2-thio-1-methyl-pseudouridine	2'-O-methyl-Cytidine
1-methyl-1-deaza-pseudouridine	2'-O-methyl-Guanosine
2-thio-1-methyl-1-deaza-pseudomidine	2'-O-methyl-Uridine
dihydrouridine	2'-O-methyl-Pseudouridine
dihydropseudouridine	2'-O-methyl-Inosine
2-thio-dihydromidine	2-methyladenosine
2-thio-dihydropseudouridine	2-methylthio-N6-methyladenosine
2-methoxyuridine	2-methylthio-N6 isopentenyladenosine
2-methoxy-4-thio-uridine	2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine
4-methoxy-pseudouridine	N6-methyl-N6-threonylcarbamoyladenosine
4-methoxy-2-thio-pseudouridine	N6-hydroxynorvalylcarbamoyladenosine
5-aza-cytidine	2-methylthio-N6-hydroxynorvalyl carbamoyladenosine
pseudoisocytidine	2'-O-ribosyladenosine (phosphate)
3-methyl-cytidine	1,2'-O-dimethylinosine
N4-acetylcytidine	5,2'-O-dimethylcytidine
5-formylcytidine	N4-acetyl-2'-O-methylcytidine
N4-methylcytidine	Lysidine
5-hydroxymethylcytidine	7-methylguanosine
1-methyl-pseudoisocytidine	N2,2'-O-dimethylguanosine
pyrrolo-cytidine	N2,N2,2'-O-trimethylguanosine
pyrrolo-pseudoisocytidine	2'-O-ribosylguanosine (phosphate)
2-thio-cytidine	Wybutosine
2-thio-5-methyl-cytidine	Peroxywybutosine
4-thio-pseudoisocytidine	Hydroxywybutosine
4-thio-1-methyl-pseudoisocytidine	undermodified hydroxywybutosine
4-thio-1-methyl-1-deaza-pseudoisocytidine	methylwyosine
1-methyl-1-deaza-pseudoisocytidine	queuosine
zebularine	epoxyqueuosine
5-aza-zebularine	galactosyl-queuosine
5-methyl-zebularine	
5-aza-2-thio-zebularine	

2-thio-zebularine	mannosyl-queuosine
2-methoxy-cytidine	7-cyano-7-deazaguanosine
2-methoxy-5-methyl-cytidine	7-aminomethyl-7-deazaguanosine
4-methoxy-pseudoisocytidine	archaeosine
4-methoxy-1-methyl-pseudoisocytidine	5,2'-O-dimethyluridine
2-aminopurine	4-thiouridine
2,6-diaminopurine	5-methyl-2-thiouridine
7-deaza-adenine	2-thio-2'-O-methyluridine
7-deaza-8-aza-adenine	3-(3-amino-3-carboxypropyl)uridine
7-deaza-2-aminopurine	5-methoxyuridine
7-deaza-8-aza-2-aminopurine	uridine 5-oxyacetic acid
7-deaza-2,6- diaminopurine	uridine 5-oxyacetic acid methyl ester
7-deaza-8-aza-2,6-diarninopurine	5-(carboxyhydroxymethyl)uridine)
1-methyladenosine	5-(carboxyhydroxymethyl)uridine methyl ester
N6-isopentenyladenosine	5-methoxycarbonylmethyluridine
N6-(cis-hydroxyisopentenyl)adenosine	5-methoxycarbonylmethyl-2'-O-methyluridine
2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine	5-methoxycarbonylmethyl-2-thiouridine
	5-aminomethyl-2-thiouridine
N6-glycinylcarbamoyladenosine	5-methylaminomethyluridine
N6-threonylcarbamoyladenosine	5-methylaminomethyl-2-thiouridine
2-methylthio-N6-threonyl carbamoyladenosine	5-methylaminomethyl-2-selenouridine
	5-carbamoylmethyluridine
N6,N6-dimethyladenosine	5-carbamoylmethyl-2'-O-methyluridine
7-methyladenine	5-carboxymethylaminomethyluridine
2-methylthio-adenine	5-carboxymethylaminomethyl-2'-O-methyluridine
2-methoxy-adenine	
inosine	5-carboxymethylaminomethyl-2-thiouridine
1-methyl-inosine	N4,2'-O-dimethylcytidine
wyosine	5-carboxymethyluridine
wybutosine	N6,2'-O-dimethyladenosine
7-deaza-guanosine	N,N6,O-2'-trimethyladenosine
7-deaza-8-aza-guanosine	N2,7-dimethylguanosine
6-thio-guanosine	N2,N2,7-trimethylguanosine
6-thio-7-deaza-guanosine	3,2'-O-dimethyluridine
6-thio-7-deaza-8-aza-guanosine	5-methyldihydrouridine
7-methyl-guanosine	5-formyl-2'-O-methylcytidine
6-thio-7-methyl-guanosine	1,2'-O-dimethylguanosine
7-methylinosine	4-demethylwyosine
6-methoxy-guanosine	Isowyosine
1-methylguanosine	N6-acetyladenosine
N2-methylguanosine	
N2,N2-dimethylguanosine	
8-oxo-guanosine	
7-methyl-8-oxo-guanosine	
1-methyl-6-thio-guanosine	

Table M2. Backbone modifications

2’-O-Methyl backbone
Peptide Nucleic Acid (PNA) backbone
phosphorothioate backbone
morpholino backbone
carbamate backbone
siloxane backbone
sulfide backbone
sulfoxide backbone
sulfone backbone
formacetyl backbone
thioformacetyl backbone
methyleneformacetyl backbone
riboacetyl backbone
alkene containing backbone
sulfamate backbone
sulfonate backbone
sulfonamide backbone
methyleneimino backbone
methylenehydrazino backbone
amide backbone

Table M3. Modified caps

m7GpppA
m7GpppC
m2,7GpppG
m2,2,7GpppG
m7Gpppm7G
m7,2'OmeGpppG
m72'dGpppG
m7,3'OmeGpppG
m7,3'dGpppG
GppppG
m7GppppG
m7GppppA
m7GppppC
m2,7GppppG
m2,2,7GppppG
m7Gppppm7G
m7,2'OmeGppppG
m72'dGppppG
m7,3'OmeGppppG
m7,3'dGppppG

Production of Compositions and Systems

Methods of designing and constructing nucleic acid constructs and proteins or polypeptides (such as the systems, constructs and polypeptides described herein) are known. Generally, recombinant methods may be used. See, in general, Smales & James (Eds.), *Therapeutic Proteins: Methods and Protocols* (Methods in Molecular Biology), Humana Press (2005); and Crommelin, Sindelar & Meibohm (Eds.), *Pharmaceutical Biotechnology: Fundamentals and Applications*, Springer (2013). Methods of designing, preparing, evaluating, purifying and manipulating nucleic acid compositions are described in Green and Sambrook (Eds.), *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012).

The disclosure provides, in part, a nucleic acid, e.g., vector, encoding a Gene Writer polypeptide described herein, a template nucleic acid described herein, or both. In some embodiments, a vector comprises a selective marker, e.g., an antibiotic resistance marker. In some embodiments, the antibiotic resistance marker is a kanamycin resistance marker. In some embodiments, the antibiotic resistance marker does not confer resistance to beta-lactam antibiotics. In some embodiments, the vector does not comprise an ampicillin resistance marker. In some embodiments, the vector comprises a kanamycin resistance marker and does not comprise an ampicillin resistance marker. In some embodiments, a vector encoding a Gene Writer polypeptide is integrated into a target cell genome (e.g., upon administration to a target cell, tissue, organ, or subject). In some embodiments, a vector encoding a Gene Writer polypeptide is not integrated into a target cell genome (e.g., upon administration to a target cell, tissue, organ, or subject). In some embodiments, a vector encoding a template nucleic acid (e.g., template RNA) is not integrated into a target cell genome (e.g., upon administration to a target cell, tissue, organ, or subject). In some embodiments, if a vector is integrated into a target site in a target cell genome, the selective marker is not integrated into the genome. In some embodiments, if a vector is integrated into a target site in a target cell genome, genes or sequences involved in vector maintenance (e.g., plasmid maintenance genes) are not integrated into the genome. In some embodiments, if a vector is integrated into a target site in a target cell genome, transfer regulating sequences (e.g., inverted terminal repeats, e.g., from an AAV) are

not integrated into the genome. In some embodiments, administration of a vector (e.g., encoding a Gene Writer polypeptide described herein, a template nucleic acid described herein, or both) to a target cell, tissue, organ, or subject results in integration of a portion of the vector into one or more target sites in the genome(s) of said target cell, tissue, organ, or subject. In some embodiments, less than 99, 95, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5, 4, 3, 2, or 1% of target sites (e.g., no target sites) comprising integrated material comprise a selective marker (e.g., an antibiotic resistance gene), a transfer regulating sequence (e.g., an inverted terminal repeat, e.g., from an AAV), or both from the vector.

Exemplary methods for producing a therapeutic pharmaceutical protein or polypeptide described herein involve expression in mammalian cells, although recombinant proteins can also be produced using insect cells, yeast, bacteria, or other cells under control of appropriate promoters. Mammalian expression vectors may comprise non-transcribed elements such as an origin of replication, a suitable promoter, and other 5' or 3' flanking non-transcribed sequences, and 5' or 3' non-translated sequences such as necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, and termination sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, splice, and polyadenylation sites may be used to provide other genetic elements required for expression of a heterologous DNA sequence. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, and mammalian cellular hosts are described in Green & Sambrook, *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012).

Various mammalian cell culture systems can be employed to express and manufacture recombinant protein. Examples of mammalian expression systems include CHO, COS, HEK293, HeLA, and BHK cell lines. Processes of host cell culture for production of protein therapeutics are described in Zhou and Kantardjieff (Eds.), *Mammalian Cell Cultures for Biologics Manufacturing (Advances in Biochemical Engineering/Biotechnology)*, Springer (2014). Compositions described herein may include a vector, such as a viral vector, e.g., a lentiviral vector, encoding a recombinant protein. In some embodiments, a vector, e.g., a viral vector, may comprise a nucleic acid encoding a recombinant protein.

Purification of protein therapeutics is described in Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); and in Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

In some embodiments, quality standards include, but are not limited to:

(i) the length of mRNA encoding the GeneWriter polypeptide, e.g., whether the mRNA has a length that is above a reference length or within a reference length range, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA present is greater than 3000, 4000, or 5000 nucleotides long;

(ii) the presence, absence, and/or length of a polyA tail on the mRNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA present contains a polyA tail (e.g., a polyA tail that is at least 5, 10, 20, 30, 50, 70, 100 nucleotides in length);

(iii) the presence, absence, and/or type of a 5' cap on the mRNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA present contains a 5' cap, e.g., whether that cap is a 7-methylguanosine cap, e.g., a O-Me-m7G cap;

(iv) the presence, absence, and/or type of one or more modified nucleotides (e.g., selected from pseudouridine, dihydrouridine, inosine, 7-methylguanosine, 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U), 5-methylcytidine (5mC), or a locked nucleotide) in the mRNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA present contains one or more modified nucleotides;

(v) the stability of the mRNA (e.g., over time and/or under a pre-selected condition), e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the mRNA remains intact (e.g., greater than 100, 125, 150, 175, or 200 nucleotides long) after a stability test; or

(vi) the potency of the mRNA in a system for modifying DNA, e.g., whether at least 1% of target sites are modified after a system comprising the mRNA is assayed for potency.

Kits, Articles of Manufacture, and Pharmaceutical Compositions

In an aspect the disclosure provides a kit comprising a Gene Writer or a Gene Writing system, e.g., as described herein. In some embodiments, the kit comprises a Gene Writer polypeptide (or a nucleic acid encoding the polypeptide) and a template RNA (or DNA encoding the template RNA). In some embodiments, the kit further comprises a reagent for introducing the system into a cell, e.g., transfection reagent, LNP, and the like. In some embodiments, the kit is suitable for any of the methods described herein. In some embodiments, the kit comprises one or more elements, compositions (e.g., pharmaceutical compositions), Gene Writers, and/or Gene

Writer systems, or a functional fragment or component thereof, e.g., disposed in an article of manufacture. In some embodiments, the kit comprises instructions for use thereof.

In an aspect, the disclosure provides an article of manufacture, e.g., in which a kit as described herein, or a component thereof, is disposed.

In an aspect, the disclosure provides a pharmaceutical composition comprising a Gene Writer or a Gene Writing system, e.g., as described herein. In some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier or excipient. In some embodiments, the pharmaceutical composition comprises a template RNA and/or an RNA encoding the polypeptide. In embodiments, the pharmaceutical composition has one or more (e.g., 1, 2, 3, or 4) of the following characteristics:

(a) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) DNA template relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(b) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) uncapped RNA relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(c) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) partial length RNAs relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(d) substantially lacks unreacted cap dinucleotides.

Chemistry, Manufacturing, and Controls (CMC)

Purification of protein therapeutics is described, for example, in Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); and in Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

In some embodiments, a Gene Writer™ system, polypeptide, and/or template nucleic acid (e.g., template RNA) conforms to certain quality standards. In some embodiments, a Gene Writer™ system, polypeptide, and/or template nucleic acid (e.g., template RNA) produced by a method described herein conforms to certain quality standards. Accordingly, the disclosure is directed, in some aspects, to methods of manufacturing a Gene Writer™ system, polypeptide, and/or template nucleic acid (e.g., template RNA) that conforms to certain quality standards, e.g., in which said quality standards are assayed. The disclosure is also directed, in some aspects, to methods of assaying said quality standards in a Gene Writer™ system, polypeptide, and/or

template nucleic acid (e.g., template RNA). In some embodiments, quality standards include, but are not limited to, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12) of the following:

(i) the length of the template RNA, e.g., whether the template RNA has a length that is above a reference length or within a reference length range, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present is greater than 100, 125, 150, 175, or 200 nucleotides long;

(ii) the presence, absence, and/or length of a polyA tail on the template RNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains a polyA tail (e.g., a polyA tail that is at least 5, 10, 20, 30, 50, 70, 100 nucleotides in length);

(iii) the presence, absence, and/or type of a 5' cap on the template RNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains a 5' cap, e.g., whether that cap is a 7-methylguanosine cap, e.g., a O-Me-m7G cap;

(iv) the presence, absence, and/or type of one or more modified nucleotides (e.g., selected from pseudouridine, dihydrouridine, inosine, 7-methylguanosine, 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U), 5-methylcytidine (5mC), or a locked nucleotide) in the template RNA, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains one or more modified nucleotides;

(v) the stability of the template RNA (e.g., over time and/or under a pre-selected condition), e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA remains intact (e.g., greater than 100, 125, 150, 175, or 200 nucleotides long) after a stability test;

(vi) the potency of the template RNA in a system for modifying DNA, e.g., whether at least 1% of target sites are modified after a system comprising the template RNA is assayed for potency;

(vii) the length of the polypeptide, first polypeptide, or second polypeptide, e.g., whether the polypeptide, first polypeptide, or second polypeptide has a length that is above a reference length or within a reference length range, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the polypeptide, first polypeptide, or second polypeptide present is greater than 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, or 2000 amino acids long (and optionally, no larger than 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, or 600 amino acids long);

(viii) the presence, absence, and/or type of post-translational modification on the polypeptide, first polypeptide, or second polypeptide, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the polypeptide, first polypeptide, or second polypeptide contains phosphorylation, methylation, acetylation, myristoylation, palmitoylation, isoprenylation, glipyatyon, or lipoylation, or any combination thereof;

(ix) the presence, absence, and/or type of one or more artificial, synthetic, or non-canonical amino acids (e.g., selected from ornithine, β-alanine, GABA, δ-Aminolevulinic acid, PABA, a D-amino acid (e.g., D-alanine or D-glutamate), aminoisobutyric acid, dehydroalanine, cystathionine, lanthionine, Djenkolic acid, Diaminopimelic acid, Homoalanine, Norvaline, Norleucine, Homonorleucine, homoserine, O-methyl-homoserine and O-ethyl-homoserine, ethionine, selenocysteine, selenohomocysteine, selenomethionine, selenoethionine, tellurocysteine, or telluromethionine) in the polypeptide, first polypeptide, or second polypeptide, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the polypeptide, first polypeptide, or second polypeptide present contains one or more artificial, synthetic, or non-canonical amino acids;

(x) the stability of the polypeptide, first polypeptide, or second polypeptide (e.g., over time and/or under a pre-selected condition), e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the polypeptide, first polypeptide, or second polypeptide remains intact (e.g., greater than 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, or 2000 amino acids long (and optionally, no larger than 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, or 600 amino acids long)) after a stability test;

(xi) the potency of the polypeptide, first polypeptide, or second polypeptide in a system for modifying DNA, e.g., whether at least 1 % of target sites are modified after a system comprising the polypeptide, first polypeptide, or second polypeptide is assayed for potency; or

(xii) the presence, absence, and/or level of one or more of a pyrogen, virus, fungus, bacterial pathogen, or host cell protein, e.g., whether the system is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, or host cell protein contamination.

In some embodiments, a system or pharmaceutical composition described herein is endotoxin free.

In some embodiments, the presence, absence, and/or level of one or more of a pyrogen, virus, fungus, bacterial pathogen, and/or host cell protein is determined. In embodiments, whether the system is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, and/or host cell protein contamination is determined.

In some embodiments, a pharmaceutical composition or system as described herein has one or more (e.g., 1, 2, 3, or 4) of the following characteristics:

(a) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) DNA template relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(b) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) uncapped RNA relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(c) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) partial length RNAs relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(d) substantially lacks unreacted cap dinucleotides.

Applications

By integrating coding genes into a RNA sequence template, the Gene Writer system can address therapeutic needs, for example, by providing expression of a therapeutic transgene in individuals with loss-of-function mutations, by replacing gain-of-function mutations with normal transgenes, by providing regulatory sequences to eliminate gain-of-function mutation expression, and/or by controlling the expression of operably linked genes, transgenes and systems thereof. In certain embodiments, the RNA sequence template encodes a promotor region specific to the therapeutic needs of the host cell, for example a tissue specific promotor or enhancer. In still other embodiments, a promotor can be operably linked to a coding sequence.

In embodiments, the Gene Writer™ gene editor system can provide therapeutic transgenes expressing, e.g., replacement blood factors or replacement enzymes, e.g., lysosomal enzymes. For example, the compositions, systems and methods described herein are useful to express, in a target human genome, agalsidase alpha or beta for treatment of Fabry Disease; imiglucerase, taliglucerase alfa, velaglucerase alfa, or alglucerase for Gaucher Disease; sebelipase alpha for lysosomal acid lipase deficiency (Wolman disease/CESD); laronidase, idursulfase, elosulfase alpha, or galsulfase for mucopolysaccharidoses; alglucosidase alpha for Pompe disease. For example, the compositions, systems and methods described herein are useful

to express, in a target human genome factor I, II, V, VII, X, XI, XII or XIII for blood factor deficiencies.

In some embodiments, the heterologous object sequence encodes an intracellular protein (e.g., a cytoplasmic protein, a nuclear protein, an organellar protein such as a mitochondrial protein or lysosomal protein, or a membrane protein). In some embodiments, the heterologous object sequence encodes a membrane protein, e.g., a membrane protein other than a CAR, and/or an endogenous human membrane protein. In some embodiments, the heterologous object sequence encodes an extracellular protein. In some embodiments, the heterologous object sequence encodes an enzyme, a structural protein, a signaling protein, a regulatory protein, a transport protein, a sensory protein, a motor protein, a defense protein, or a storage protein. Other exemplary proteins that may be encoded by a heterologous object sequence include, without limitation, a immune receptor protein, e.g. a synthetic immune receptor protein such as a chimeric antigen receptor protein (CAR), a T cell receptor, a B cell receptor, or an antibody.

Administration

The composition and systems described herein may be used in vitro or in vivo. In some embodiments the system or components of the system are delivered to cells (e.g., mammalian cells, e.g., human cells), e.g., in vitro or in vivo. In some embodiments, the cells are eukaryotic cells, e.g., cells of a multicellular organism, e.g., an animal, e.g., a mammal (e.g., human, swine, bovine) a bird (e.g., poultry, such as chicken, turkey, or duck), or a fish. In some embodiments, the cells are non-human animal cells (e.g., a laboratory animal, a livestock animal, or a companion animal). In some embodiments, the cell is a stem cell (e.g., a hematopoietic stem cell), a fibroblast, or a T cell. In some embodiments, the cell is a non-dividing cell, e.g., a non-dividing fibroblast or non-dividing T cell. In some embodiments, the cell is an HSC and p53 is not upregulated or is upregulated by less than 10%, 5%, 2%, or 1%, e.g., as determined according to the method described in Example 30 of PCT Application No. PCT/US2019/048607, incorporated herein by reference in its entirety. The components of the Gene Writer system may, in some instances, be delivered in the form of polypeptide, nucleic acid (e.g., DNA, RNA), and combinations thereof.

For instance, delivery can use any of the following combinations for delivering the retrotransposase (e.g., as DNA encoding the retrotransposase protein, as RNA encoding the

retrotransposase protein, or as the protein itself) and the template RNA (e.g., as DNA encoding the RNA, or as RNA):

1. Retrotransposase DNA + template DNA
2. Retrotransposase RNA + template DNA
3. Retrotransposase DNA + template RNA
4. Retrotransposase RNA + template RNA
5. Retrotransposase protein + template DNA
6. Retrotransposase protein + template RNA
7. Retrotransposase virus + template virus
8. Retrotransposase virus + template DNA
9. Retrotransposase virus + template RNA
10. Retrotransposase DNA + template virus
11. Retrotransposase RNA + template virus
12. Retrotransposase protein + template virus

As indicated above, in some embodiments, the DNA or RNA that encodes the retrotransposase protein is delivered using a virus, and in some embodiments, the template RNA (or the DNA encoding the template RNA) is delivered using a virus.

In one embodiments the system and/or components of the system are delivered as nucleic acid. For example, the Gene Writer polypeptide may be delivered in the form of a DNA or RNA encoding the polypeptide, and the template RNA may be delivered in the form of RNA or its complementary DNA to be transcribed into RNA. In some embodiments the system or components of the system are delivered on 1, 2, 3, 4, or more distinct nucleic acid molecules. In some embodiments the system or components of the system are delivered as a combination of DNA and RNA. In some embodiments the system or components of the system are delivered as a combination of DNA and protein. In some embodiments the system or components of the system are delivered as a combination of RNA and protein. In some embodiments the Gene Writer genome editor polypeptide is delivered as a protein.

In some embodiments the system or components of the system are delivered to cells, e.g. mammalian cells or human cells, using a vector. The vector may be, e.g., a plasmid or a virus. In some embodiments delivery is in vivo, in vitro, ex vivo, or in situ. In some embodiments the virus is an adeno associated virus (AAV), a lentivirus, an adenovirus. In some embodiments the

system or components of the system are delivered to cells with a viral-like particle or a virosome. In some embodiments the delivery uses more than one virus, viral-like particle or virosome.

In some embodiments, nucleic acid (e.g., encoding a polypeptide, or a template DNA, or both) delivered to cells is covalently closed linear DNA, or so-called "doggybone" DNA. During its lifecycle, the bacteriophage N15 employs protelomerase to convert its genome from circular plasmid DNA to a linear plasmid DNA (Ravin et al. J Mol Biol 2001). This process has been adapted for the production of covalently closed linear DNA in vitro (see, for example, WO2010086626A1). In some embodiments, a protelomerase is contacted with a DNA containing one or more protelomerase recognition sites, wherein protelomerase results in a cut at the one or more sites and subsequent ligation of the complementary strands of DNA, resulting in the covalent linkage between the complementary strands. In some embodiments, nucleic acid (e.g., encoding a transposase, or a template DNA, or both) is first generated as circular plasmid DNA containing a single protelomerase recognition site that is then contacted with protelomerase to yield a covalently closed linear DNA. In some embodiments, nucleic acid (e.g., encoding a transposase, or a template DNA, or both) flanked by protelomerase recognition sites on plasmid or linear DNA is contacted with protelomerase to generate a covalently closed linear DNA containing only the DNA contained between the protelomerase recognition sites. In some embodiments, the approach of flanking the desired nucleic acid sequence by protelomerase recognition sites results in covalently closed circular DNA lacking plasmid elements used for bacterial cloning and maintenance. In some embodiments, the plasmid or linear DNA containing the nucleic acid and one or more protelomerase recognition sites is optionally amplified prior to the protelomerase reaction, e.g., by rolling circle amplification or PCR.

In one embodiment, the compositions and systems described herein can be formulated in liposomes or other similar vesicles. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes may be anionic, neutral or cationic. Liposomes are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

Vesicles can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Methods for preparation of multilamellar vesicle lipids are known in the art (see for example U.S. Pat. No. 6,693,086, the teachings of which relating to multilamellar vesicle lipid preparation are incorporated herein by reference). Although vesicle formation can be spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review). Extruded lipids can be prepared by extruding through filters of decreasing size, as described in Templeton et al., Nature Biotech, 15:647-652, 1997, the teachings of which relating to extruded lipid preparation are incorporated herein by reference.

Lipid nanoparticles are another example of a carrier that provides a biocompatible and biodegradable delivery system for the pharmaceutical compositions described herein. Nanostructured lipid carriers (NLCs) are modified solid lipid nanoparticles (SLNs) that retain the characteristics of the SLN, improve drug stability and loading capacity, and prevent drug leakage. Polymer nanoparticles (PNPs) are an important component of drug delivery. These nanoparticles can effectively direct drug delivery to specific targets and improve drug stability and controlled drug release. Lipid–polymer nanoparticles (PLNs), a new type of carrier that combines liposomes and polymers, may also be employed. These nanoparticles possess the complementary advantages of PNPs and liposomes. A PLN is composed of a core–shell structure; the polymer core provides a stable structure, and the phospholipid shell offers good biocompatibility. As such, the two components increase the drug encapsulation efficiency rate, facilitate surface modification, and prevent leakage of water-soluble drugs. For a review, see, e.g., Li et al. 2017, Nanomaterials 7, 122; doi:10.3390/nano7060122.

Exosomes can also be used as drug delivery vehicles for the compositions and systems described herein. For a review, see Ha et al. July 2016. Acta Pharmaceutica Sinica B. Volume 6, Issue 4, Pages 287-296; https://doi.org/10.1016/j.apsb.2016.02.001.

Fusosomes interact and fuse with target cells, and thus can be used as delivery vehicles for a variety of molecules. They generally consist of a bilayer of amphipathic lipids enclosing a lumen or cavity and a fusogen that interacts with the amphipathic lipid bilayer. The fusogen component has been shown to be engineerable in order to confer target cell specificity for the

fusion and payload delivery, allowing the creation of delivery vehicles with programmable cell specificity (see, for example, the relating to fusosome design, preparation, and usage in PCT Publication No. WO/2020014209, incorporated herein by reference in its entirety).

A Gene Writer system can be introduced into cells, tissues and multicellular organisms. In some embodiments the system or components of the system are delivered to the cells via mechanical means or physical means.

Formulation of protein therapeutics is described in Meyer (Ed.), *Therapeutic Protein Drug Products: Practical Approaches to formulation in the Laboratory, Manufacturing, and the Clinic*, Woodhead Publishing Series (2012).

Tissue Specific Activity/Administration

In some embodiments, a system, template RNA, or polypeptide described herein is administered to or is active in (e.g., is more active in) a target tissue, e.g., a first tissue. In some embodiments, the system, template RNA, or polypeptide is not administered to or is less active in (e.g., not active in) a non-target tissue. In some embodiments, a system, template RNA, or polypeptide described herein is useful for modifying DNA in a target tissue, e.g., a first tissue, (e.g., and not modifying DNA in a non-target tissue).

In some embodiments, a system comprises (a) a polypeptide described herein or a nucleic acid encoding the same, (b) a template nucleic acid (e.g., template RNA) described herein, and (c) one or more first tissue-specific expression-control sequences specific to the target tissue, wherein the one or more first tissue-specific expression-control sequences specific to the target tissue are in operative association with (a), (b), or (a) and (b), wherein, when associated with (a), (a) comprises a nucleic acid encoding the polypeptide.

In some embodiments, the nucleic acid in (b) comprises RNA.

In some embodiments, the nucleic acid in (b) comprises DNA.

In some embodiments, the nucleic acid in (b): (i) is single-stranded or comprises a single-stranded segment, e.g., is single-stranded DNA or comprises a single-stranded segment and one or more double stranded segments; (ii) has inverted terminal repeats; or (iii) both (i) and (ii).

In some embodiments, the nucleic acid in (b) is double-stranded or comprises a double-stranded segment.

In some embodiments, (a) comprises a nucleic acid encoding the polypeptide.

In some embodiments, the nucleic acid in (a) comprises RNA.

In some embodiments, the nucleic acid in (a) comprises DNA.

In some embodiments, the nucleic acid in (a): (i) is single-stranded or comprises a single-stranded segment, e.g., is single-stranded DNA or comprises a single-stranded segment and one or more double stranded segments; (ii) has inverted terminal repeats; or (iii) both (i) and (ii).

In some embodiments, the nucleic acid in (a) is double-stranded or comprises a double-stranded segment.

In some embodiments, the nucleic acid in (a), (b), or (a) and (b) is linear.

In some embodiments, the nucleic acid in (a), (b), or (a) and (b) is circular, e.g., a plasmid or minicircle.

In some embodiments, the heterologous object sequence is in operative association with a first promoter.

In some embodiments, the one or more first tissue-specific expression-control sequences comprises a tissue specific promoter.

In some embodiments, the tissue-specific promoter comprises a first promoter in operative association with: i. the heterologous object sequence, ii. a nucleic acid encoding the transposase, or iii. (i) and (ii).

In some embodiments, the one or more first tissue-specific expression-control sequences comprises a tissue-specific microRNA recognition sequence in operative association with: i. the heterologous object sequence, ii. a nucleic acid encoding the transposase, or iii. (i) and (ii).

In some embodiments, a system comprises a tissue-specific promoter, and the system further comprises one or more tissue-specific microRNA recognition sequences, wherein: i. the tissue specific promoter is in operative association with: I. the heterologous object sequence, II. a nucleic acid encoding the transposase, or III. (I) and (II); and/or ii. the one or more tissue-specific microRNA recognition sequences are in operative association with: I. the heterologous object sequence, II. a nucleic acid encoding the transposase, or III. (I) and (II).

In some embodiments, wherein (a) comprises a nucleic acid encoding the polypeptide, the nucleic acid comprises a promoter in operative association with the nucleic acid encoding the polypeptide.

In some embodiments, the nucleic acid encoding the polypeptide comprises one or more second tissue-specific expression-control sequences specific to the target tissue in operative association with the polypeptide coding sequence.

In some embodiments, the one or more second tissue-specific expression-control sequences comprises a tissue specific promoter.

In some embodiments, the tissue-specific promoter is the promoter in operative association with the nucleic acid encoding the polypeptide.

In some embodiments, the one or more second tissue-specific expression-control sequences comprises a tissue-specific microRNA recognition sequence.

In some embodiments, the promoter in operative association with the nucleic acid encoding the polypeptide is a tissue-specific promoter, the system further comprising one or more tissue-specific microRNA recognition sequences.

In some embodiments, a Gene Writer™ system described herein is delivered to a tissue or cell from the cerebrum, cerebellum, adrenal gland, ovary, pancreas, parathyroid gland, hypophysis, testis, thyroid gland, breast, spleen, tonsil, thymus, lymph node, bone marrow, lung, cardiac muscle, esophagus, stomach, small intestine, colon, liver, salivary gland, kidney, prostate, blood, or other cell or tissue type. In some embodiments, a Gene Writer™ system described herein is used to treat a disease, such as a cancer, inflammatory disease, infectious disease, genetic defect, or other disease. A cancer can be cancer of the cerebrum, cerebellum, adrenal gland, ovary, pancreas, parathyroid gland, hypophysis, testis, thyroid gland, breast, spleen, tonsil, thymus, lymph node, bone marrow, lung, cardiac muscle, esophagus, stomach, small intestine, colon, liver, salivary gland, kidney, prostate, blood, or other cell or tissue type, and can include multiple cancers.

In some embodiments, a Gene Writer™ system described herein described herein is administered by enteral administration (e.g. oral, rectal, gastrointestinal, sublingual, sublabial, or buccal administration). In some embodiments, a Gene Writer™ system described herein is administered by parenteral administration (e.g., intravenous, intramuscular, subcutaneous, intradermal, epidural, intracerebral, intracerebroventricular, epicutaneous, nasal, intra-arterial, intra-articular, intracavernous, intraocular, intraosseous infusion, intraperitoneal, intrathecal, intrauterine, intravaginal, intravesical, perivascular, or transmucosal administration). In some

embodiments, a Gene Writer™ system described herein is administered by topical administration (e.g., transdermal administration).

In some embodiments, a Gene Writer™ system as described herein can be used to modify an animal cell, plant cell, or fungal cell. In some embodiments, a Gene Writer™ system as described herein can be used to modify a mammalian cell (e.g., a human cell). In some embodiments, a Gene Writer™ system as described herein can be used to modify a cell from a livestock animal (e.g., a cow, horse, sheep, goat, pig, llama, alpaca, camel, yak, chicken, duck, goose, or ostrich). In some embodiments, a Gene Writer™ system as described herein can be used as a laboratory tool or a research tool, or used in a laboratory method or research method, e.g., to modify an animal cell, e.g., a mammalian cell (e.g., a human cell), a plant cell, or a fungal cell.

In some embodiments, a Gene Writer™ system as described herein can be used to express a protein, template, or heterologous object sequence (e.g., in an animal cell, e.g., a mammalian cell (e.g., a human cell), a plant cell, or a fungal cell). In some embodiments, a Gene Writer™ system as described herein can be used to express a protein, template, or heterologous object sequence under the control of an inducible promoter (e.g., a small molecule inducible promoter). In some embodiments, a Gene Writing system or payload thereof is designed for tunable control, e.g., by the use of an inducible promoter. For example, a promoter, e.g., Tet, driving a gene of interest may be silent at integration, but may, in some instances, activated upon exposure to a small molecule inducer, e.g., doxycycline. In some embodiments, the tunable expression allows post-treatment control of a gene (e.g., a therapeutic gene), e.g., permitting a small molecule-dependent dosing effect. In embodiments, the small molecule-dependent dosing effect comprises altering levels of the gene product temporally and/or spatially, e.g., by local administration. In some embodiments, a promoter used in a system described herein may be inducible, e.g., responsive to an endogenous molecule of the host and/or an exogenous small molecule administered thereto.

In some embodiments, a Gene Writing system is used to make changes to non-coding and/or regulatory control regions, e.g., to tune the expression of endogenous genes. In some embodiments, a Gene Writing system is used to induce upregulation or downregulation of gene expression. In some embodiments, a regulatory control region comprises one or more of a promoter, enhancer, UTR, CTCF site, and/or a gene expression control region.

In some embodiments, a Gene Writing system may be used to treat or prevent a repeat expansion disease (e.g., a disease of Table 26), or to reduce the severity or a symptom thereof. In some embodiments, the repeat expansion disease comprises expansion of a trinucleotide repeat. In some embodiments, the subject has at least 10, 20, 30, 40, or 50 copies of the repeat. In embodiments, the repeat expansion disease is an inherited disease. Non-limiting examples of repeat expansion diseases include Huntington's disease (HD) and myotonic dystrophy. For example, healthy individuals may possess between 10 and 35 tandem copies of the CAG trinucleotide repeat, while Huntington's patients frequently possess >40 copies, which can result, e.g., in an elongated and dysfunctional Huntingtin protein. In some embodiments, a Gene Writer corrects a repeat expansion, e.g., by recognizing DNA at the terminus of the repeat region and nicking one strand (Figure 30). In some embodiments, the template RNA component of the Gene Writer comprises a region with a number of repeats characterstic of a healthy subject, e.g., about 20 repeats (e.g., between 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, or 35-40 repeats). In some embodiments, the template RNA component of the Gene Writer is copied by TPRT into the target site. In some embodiments, a second strand nick and second strand synthesis then results in the integration of the newly copied DNA comprising a correct number of repeats (e.g., as described herein). In some embodiments, the system recognizes DNA at the terminus of the repeat region and the template carries the information for the new number of repeats. In embodiments, a Gene Writer can be used in this way regardless of the number of repeats present in an individual and/or in an individual cell. Owing to the presence of multiple repeats, an alternative non-GeneWriter therapeutic (e.g., a CRISPR-based homologous recombination therapeutic) might, in some embodiments, result in unpredictable repair behavior. Further non-limiting examples of repeat expansion diseases and the causative repeats can be found, for example, in La Spada and Taylor Nat Rev Genet 11(4):247-258 (2010), which is incorporated herein by reference in its entirety.

In some embodiments, a Gene Writing system may be used to treat a healthy individual, e.g., as a preventative therapy. Gene Writing systems can, in some embodiments, be targeted to generate mutations, e.g., that have been shown to be protective towards a disease of interest. An exemplary list of such diseases and protective mutation targets can be found in Table 22.

In some embodiments, a nucleic acid component of a system provided by the invention a sequence (e.g., encoding the polypeptide or comprising a heterologous object sequence) is

flanked by untranslated regions (UTRs) that modify protein expression levels. Various 5' and 3' UTRs can affect protein expression. For example, in some embodiments, the coding sequence may be preceded by a 5' UTR that modifies RNA stability or protein translation. In some embodiments, the sequence may be followed by a 3' UTR that modifies RNA stability or translation. In some embodiments, the sequence may be preceded by a 5' UTR and followed by a 3' UTR that modify RNA stability or translation. In some embodiments, the 5' and/or 3' UTR may be selected from the 5' and 3' UTRs of complement factor 3 (C3) (cactcctccccatcctctccctctgtccctctgtccctctgaccctgcactgtcccagcacc) or orosomucoid 1 (ORM1) (caggacacagccttggatcaggacagagacttgggggccatcctgcccctccaacccgacatgtgtacctcagctttttccctcacttgcat caataaagcttctgtgtttggaacagctaa) (Asrani et al. RNA Biology 2018). In certain embodiments, the 5' UTR is the 5' UTR from C3 and the 3' UTR is the 3' UTR from ORM1.

In certain embodiments, a 5' UTR and 3' UTR for protein expression, e.g., mRNA (or DNA encoding the RNA) for a Gene Writer polypeptide or heterologous object sequence, comprise optimized expression sequences. In some embodiments, the 5' UTR comprises GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC and/or the 3' UTR comprising UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCCC AGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA, e.g., as described in Richner et al. *Cell* 168(6): P1114-1125 (2017), the sequences of which are incorporated herein by reference.

In some embodiments, a 5' and/or 3' UTR may be selected to enhance protein expression. In some embodiments, a 5' and/or 3' UTR may be selected to modify protein expression such that overproduction inhibition is minimized. In some embodiments, UTRs are around a coding sequence, e.g., outside the coding sequence and in other embodiments proximal to the coding sequence, In some embodiments additional regulatory elements (e.g., miRNA binding sites, cis-regulatory sites) are included in the UTRs.

In some embodiments, an open reading frame (ORF) of a Gene Writer system, e.g., an ORF of an mRNA (or DNA encoding an mRNA) encoding a Gene Writer polypeptide or one or more ORFs of an mRNA (or DNA encoding an mRNA) of a heterologous object sequence, is flanked by a 5' and/or 3' untranslated region (UTR) that enhances the expression thereof. In some embodiments, the 5' UTR of an mRNA component (or transcript produced from a DNA

component) of the system comprises the sequence 5'-GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC-3'. In some embodiments, the 3' UTR of an mRNA component (or transcript produced from a DNA component) of the system comprises the sequence 5'-UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCCCAGCCCCUCCUCCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA-3'. This combination of 5' UTR and 3' UTR has been shown to result in desirable expression of an operably linked ORF by Richner et al. *Cell* 168(6): P1114-1125 (2017), the teachings and sequences of which are incorporated herein by reference. In some embodiments, a system described herein comprises a DNA encoding a transcript, wherein the DNA comprises the corresponding 5' UTR and 3' UTR sequences, with T substituting for U in the above-listed sequence). In some embodiments, a DNA vector used to produce an RNA component of the system further comprises a promoter upstream of the 5' UTR for initiating in vitro transcription, e.g, a T7, T3, or SP6 promoter. The 5' UTR above begins with GGG, which is a suitable start for optimizing transcription using T7 RNA polymerase. For tuning transcription levels and altering the transcription start site nucleotides to fit alternative 5' UTRs, the teachings of Davidson et al. *Pac Symp Biocomput* 433-443 (2010) describe T7 promoter variants, and the methods of discovery thereof, that fulfill both of these traits.

Viral vectors and components thereof

Viruses are a useful source of delivery vehicles for the systems described herein, in addition to a source of relevant enzymes or domains as described herein, e.g., as sources of polymerases and polymerase functions used herein, e.g., DNA-dependent DNA polymerase, RNA-dependent RNA polymerase, RNA-dependent DNA polymerase, DNA-dependent RNA polymerase, reverse transcriptase. Some enzymes, e.g., reverse transcriptases, may have multiple activities, e.g., be capable of both RNA-dependent DNA polymerization and DNA-dependent DNA polymerization, e.g., first and second strand synthesis. In some embodiments, the virus used as a Gene Writer delivery system or a source of components thereof may be selected from a group as described by Baltimore *Bacteriol Rev* 35(3):235-241 (1971).

In some embodiments, the virus is selected from a Group I virus, e.g., is a DNA virus and packages dsDNA into virions. In some embodiments, the Group I virus is selected from, e.g., Adenoviruses, Herpesviruses, Poxviruses.

In some embodiments, the virus is selected from a Group II virus, e.g., is a DNA virus and packages ssDNA into virions. In some embodiments, the Group II virus is selected from, e.g., Parvoviruses. In some embodiments, the parvovirus is a dependoparvovirus, e.g., an adeno-associated virus (AAV).

In some embodiments, the virus is selected from a Group III virus, e.g., is an RNA virus and packages dsRNA into virions. In some embodiments, the Group III virus is selected from, e.g., Reoviruses. In some embodiments, one or both strands of the dsRNA contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps.

In some embodiments, the virus is selected from a Group IV virus, e.g., is an RNA virus and packages ssRNA(+) into virions. In some embodiments, the Group IV virus is selected from, e.g., Coronaviruses, Picornaviruses, Togaviruses. In some embodiments, the ssRNA(+) contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps.

In some embodiments, the virus is selected from a Group V virus, e.g., is an RNA virus and packages ssRNA(-) into virions. In some embodiments, the Group V virus is selected from, e.g., Orthomyxoviruses, Rhabdoviruses. In some embodiments, an RNA virus with an ssRNA(-) genome also carries an enzyme inside the virion that is transduced to host cells with the viral genome, e.g., an RNA-dependent RNA polymerase, capable of copying the ssRNA(-) into ssRNA(+) that can be translated directly by the host.

In some embodiments, the virus is selected from a Group VI virus, e.g., is a retrovirus and packages ssRNA(+) into virions. In some embodiments, the Group VI virus is selected from, e.g., Retroviruses. In some embodiments, the retrovirus is a lentivirus, e.g., HIV-1, HIV-2, SIV, BIV. In some embodiments, the retrovirus is a spumavirus, e.g., a foamy virus, e.g., HFV, SFV, BFV. In some embodiments, the ssRNA(+) contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into

protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps. In some embodiments, the ssRNA(+) is first reverse transcribed and copied to generate a dsDNA genome intermediate from which mRNA can be transcribed in the host cell. In some embodiments, an RNA virus with an ssRNA(+) genome also carries an enzyme inside the virion that is transduced to host cells with the viral genome, e.g., an RNA-dependent DNA polymerase, capable of copying the ssRNA(+) into dsDNA that can be transcribed into mRNA and translated by the host. In some embodiments, the reverse transcriptase from a Group VI retrovirus is incorporated as the reverse transcriptase domain of a Gene Writer polypeptide.

In some embodiments, the virus is selected from a Group VII virus, e.g., is a retrovirus and packages dsRNA into virions. In some embodiments, the Group VII virus is selected from, e.g., Hepadnaviruses. In some embodiments, one or both strands of the dsRNA contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps. In some embodiments, one or both strands of the dsRNA contained in such virions is first reverse transcribed and copied to generate a dsDNA genome intermediate from which mRNA can be transcribed in the host cell. In some embodiments, an RNA virus with a dsRNA genome also carries an enzyme inside the virion that is transduced to host cells with the viral genome, e.g., an RNA-dependent DNA polymerase, capable of copying the dsRNA into dsDNA that can be transcribed into mRNA and translated by the host. In some embodiments, the reverse transcriptase from a Group VII retrovirus is incorporated as the reverse transcriptase domain of a Gene Writer polypeptide.

In some embodiments, virions used to deliver nucleic acid in this invention may also carry enzymes involved in the process of Gene Writing. For example, a retroviral virion may contain a reverse transcriptase domain that is delivered into a host cell along with the nucleic acid. In some embodiments, an RNA template may be associated with a Gene Writer polypeptide within a virion, such that both are co-delivered to a target cell upon transduction of the nucleic acid from the viral particle. In some embodiments, the nucleic acid in a virion may comprise DNA, e.g., linear ssDNA, linear dsDNA, circular ssDNA, circular dsDNA, minicircle DNA, dbDNA, ceDNA. In some embodiments, the nucleic acid in a virion may comprise RNA, e.g., linear ssRNA, linear dsRNA, circular ssRNA, circular dsRNA. In some embodiments, a viral

genome may circularize upon transduction into a host cell, e.g., a linear ssRNA molecule may undergo a covalent linkage to form a circular ssRNA, a linear dsRNA molecule may undergo a covalent linkage to form a circular dsRNA or one or more circular ssRNA. In some embodiments, a viral genome may replicate by rolling circle replication in a host cell. In some embodiments, a viral genome may comprise a single nucleic acid molecule, e.g., comprise a non-segmented genome. In some embodiments, a viral genome may comprise two or more nucleic acid molecules, e.g., comprise a segmented genome. In some embodiments, a nucleic acid in a virion may be associated with one or proteins. In some embodiments, one or more proteins in a virion may be delivered to a host cell upon transduction. In some embodiments, a natural virus may be adapted for nucleic acid delivery by the addition of virion packaging signals to the target nucleic acid, wherein a host cell is used to package the target nucleic acid containing the packaging signals.

In some embodiments, a virion used as a delivery vehicle may comprise a commensal human virus. In some embodiments, a virion used as a delivery vehicle may comprise an anellovirus, the use of which is described in WO2018232017A1, which is incorporated herein by reference in its entirety.

Adeno-associated viruses

In some embodiments, the virus is an adeno-associated virus (AAV). In some embodiments, the AAV genome comprises two genes that encode four replication proteins and three capsid proteins, respectively. In some embodiments, the genes are flanked on either side by 145-bp inverted terminal repeats (ITRs). In some embodiments, the virion comprises up to three capsid proteins (Vp1, Vp2, and/or Vp3), e.g., produced in a 1:1:10 ratio. In some embodiments, the capsid proteins are produced from the same open reading frame and/or from differential splicing (Vp1) and alternative translational start sites (Vp2 and Vp3, respectively). Generally, Vp3 is the most abundant subunit in the virion and participates in receptor recognition at the cell surface defining the tropism of the virus. In some embodiments, Vp1 comprises a phospholipase domain, e.g., which functions in viral infectivity, in the N-terminus of Vp1.

In some embodiments, packaging capacity of the viral vectors limits the size of the base editor that can be packaged into the vector. For example, the packaging capacity of the AAVs

can be about 4.5 kb (e.g., about 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, or 6.0 kb), e.g., including one or two inverted terminal repeats (ITRs), e.g., 145 base ITRs.

In some embodiments, recombinant AAV (rAAV) comprises cis-acting 145-bp ITRs flanking vector transgene cassettes, e.g., providing up to 4.5 kb for packaging of foreign DNA. Subsequent to infection, rAAV can, in some instances, express a fusion protein of the invention and persist without integration into the host genome by existing episomally in circular head-to-tail concatemers. rAAV can be used, for example, in vitro and in vivo. In some embodiments, AAV-mediated gene delivery requires that the length of the coding sequence of the gene is equal or greater in size than the wild-type AAV genome.

AAV delivery of genes that exceed this size and/or the use of large physiological regulatory elements can be accomplished, for example, by dividing the protein(s) to be delivered into two or more fragments. In some embodiments, the N-terminal fragment is fused to a split intein-N. In some embodiments, the C- terminal fragment is fused to a split intein-C. In embodiments, the fragments are packaged into two or more AAV vectors.

In some embodiments, dual AAV vectors are generated by splitting a large transgene expression cassette in two separate halves (5 and 3 ends, or head and tail), e.g., wherein each half of the cassette is packaged in a single AAV vector (of <5 kb). The re-assembly of the full-length transgene expression cassette can, in some embodiments, then be achieved upon co-infection of the same cell by both dual AAV vectors. In some embodiments, co-infection is followed by one or more of: (1) homologous recombination (HR) between 5 and 3 genomes (dual AAV overlapping vectors); (2) ITR-mediated tail-to-head concatemerization of 5 and 3 genomes (dual AAV trans-splicing vectors); and/or (3) a combination of these two mechanisms (dual AAV hybrid vectors). In some embodiments, the use of dual AAV vectors in vivo results in the expression of full-length proteins. In some embodiments, the use of the dual AAV vector platform represents an efficient and viable gene transfer strategy for transgenes of greater than about 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, or 5.0 kb in size.In some embodiments, AAV vectors can also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides. In some embodiments, AAV vectors can be used for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest.94:1351 (1994); each of which is incorporated herein by reference in

their entirety). The construction of recombinant AAV vectors is described in a number of publications, including U.S. Patent No.5,173,414; Tratschin et al., Mol. Cell. Biol.5:3251- 3260 (1985); Tratschin, et al., Mol. Cell. Biol.4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol.63:03822-3828 (1989) (incorporated by reference herein in their entirety).

In some embodiments, a Gene Writer described herein (e.g., with or without one or more guide nucleic acids) can be delivered using AAV, lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Patent No. 8,454,972 (formulations, doses for adenovirus), U.S. Patent No.8,404,658 (formulations, doses for AAV) and U.S. Patent No.5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For example, for AAV, the route of administration, formulation and dose can be as described in U.S. Patent No.8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as described in U.S. Patent No.8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as described in U.S. Patent No.5,846,946 and as in clinical studies involving plasmids. Doses can be based on or extrapolated to an average 70 kg individual (e.g. a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. In some embodiments, the viral vectors can be injected into the tissue of interest. For cell-type specific Gene Writing, the expression of the Gene Writer and optional guide nucleic acid can, in some embodiments, be driven by a cell-type specific promoter.

In some embodiments, AAV allows for low toxicity, for example, due to the purification method not requiring ultracentrifugation of cell particles that can activate the immune response. In some embodiments, AAV allows low probability of causing insertional mutagenesis, for example, because it does not substantially integrate into the host genome.

In some embodiments, AAV has a packaging limit of about 4.4, 4.5, 4.6, 4.7, or 4.75 kb. In some embodiments, a Gene Writer, promoter, and transcription terminator can fit into a single

viral vector. SpCas9 (4.1 kb) may, in some instances, be difficult to package into AAV. Therefore, in some embodiments, a Gene Writer is used that is shorter in length than other Gene Writers or base editors. In some embodiments, the Gene Writers are less than about 4.5 kb, 4.4 kb, 4.3 kb, 4.2 kb, 4.1 kb, 4 kb, 3.9 kb, 3.8 kb, 3.7 kb, 3.6 kb, 3.5 kb, 3.4 kb, 3.3 kb, 3.2 kb, 3.1 kb, 3 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.6 kb, 2.5 kb, 2 kb, or 1.5 kb.

An AAV can be AAV1, AAV2, AAV5 or any combination thereof. In some embodiments, the type of AAV is selected with respect to the cells to be targeted; e.g., AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof can be selected for targeting brain or neuronal cells; or AAV4 can be selected for targeting cardiac tissue. In some embodiments, AAV8 is selected for delivery to the liver. Exemplary AAV serotypes as to these cells are described, for example, in Grimm, D. et al, J. Virol.82: 5887-5911 (2008) (incorporated herein by reference in its entirety). In some embodiments, AAV refers all serotypes, subtypes, and naturally-occurring AAV as well as recombinant AAV. AAV may be used to refer to the virus itself or a derivative thereof. In some embodiments, AAV includes AAV1, AAV2, AAV3, AAV3B, AAV4, AAV5, AAV6, AAV6.2, AAV7, AAVrh.64Rl, AAVhu.37, AAVrh.8, AAVrh.32.33, AAV8, AAV9, AAV-DJ, AAV2/8, AAVrhlO, AAVLK03, AV10, AAV11, AAV 12, rhlO, and hybrids thereof, avian AAV, bovine AAV, canine AAV, equine AAV, primate AAV, nonprimate AAV, and ovine AAV. The genomic sequences of various serotypes of AAV, as well as the sequences of the native terminal repeats (TRs), Rep proteins, and capsid subunits are known in the art. Such sequences may be found in the literature or in public databases such as GenBank. Additional exemplary AAV serotypes are listed in Table 36.

Table 36. Exemplary AAV serotypes.

Target Tissue	Vehicle	Reference
Liver	AAV (AAV8[1], AAVrh.8[1], AAVhu.37[1], AAV2/8, AAV2/rh10[2], AAV9, AAV2, NP40[3], NP59[2,3], AAV3B[5], AAV-DJ[4], AAV-LK01[4], AAV-LK02[4], AAV-LK03[4], AAV-LK19[4] Adenovirus (Ad5, HC-AdV[6])	1. Wang *et al.*, *Mol. Ther.* **18**, 118–25 (2010) 2. Ginn *et al.*, *JHEP Reports*, 100065 (2019) 3. Paulk *et al.*, *Mol. Ther.* **26**, 289–303 (2018). 4. L. Lisowski *et al.*, *Nature.* **506**, 382–6 (2014).

		5. L. Wang *et al.*, *Mol. Ther.* **23**, 1877–87 (2015). 6. Hausl *Mol Ther* (2010)
Lung	AAV (AAV4, AAV5, AAV6[1], AAV9, H22[2]) Adenovirus (Ad5, Ad3, Ad21, Ad14)[3]	1. Duncan *et al.*, *Mol Ther Methods Clin Dev* (2018) 2. Cooney *et al.*, Am J Respir Cell Mol Biol (2019) 3. Li *et al.*, *Mol Ther Methods Clin Dev* (2019)
Skin	AAV (AAV6[1], AAV-LK19[2])	1. Petek et al., *Mol. Ther.* (2010) 2. L. Lisowski *et al.*, *Nature.* **506**, 382–6 (2014).
HSCs	Adenovirus (HDAd5/35^{++})	Wang et al. *Blood Adv* (2019)

In some embodiments, a pharmaceutical composition (e.g., comprising an AAV as dscribed herein) has less than 10% empty capsids, less than 8% empty capsids, less than 7% empty capsids, less than 5% empty capsids, less than 3% empty capsids, or less than 1 % empty capsids. In some embodiments, the pharmaceutical composition has less than about 5% empty capsids. In some embodiments, the number of empty capsids is below the limit of detection. In some embodiments, it is advantageous for the pharmaceutical composition to have low amounts of empty capsids, e.g., because empty capsids may generate an adverse response (e.g., immune response, inflammatory response, liver response, and/or cardiac response), e.g., with little or no substantial therapeutic benefit.

In some embodiments, the residual host cell protein (rHCP) in the pharmaceutical composition is less than or equal to 100 ng/ml rHCP per 1 x 10^{13} vg/ml, e.g., less than or equal to 40 ng/ml rHCP per 1 x 10^{13} vg/ml or 1-50 ng/ml rHCP per 1 x 10^{13} vg/ml. In some embodiments, the pharmaceutical composition comprises less than 10 ng rHCP per 1.0 x 10^{13} vg, or less than 5 ng rHCP per 1.0 x 10^{13} vg, less than 4 ng rHCP per 1.0 x 10^{13} vg, or less than 3 ng rHCP per 1.0 x 10^{13} vg, or any concentration in between. In some embodiments, the residual host cell DNA (hcDNA) in the pharmaceutical composition is less than or equal to 5 x 10^{6} pg/ml hcDNA per 1 x 10^{13} vg/ml, less than or equal to 1.2 x 10^{6} pg/ml hcDNA per 1 x 10^{13} vg/ml, or 1 x 10^{5} pg/ml hcDNA per 1 x 10^{13} vg/ml. In some embodiments, the residual host cell DNA in said pharmaceutical composition is less than 5.0 x 10^{5} pg per 1 x 10^{13} vg, less than 2.0 x 10^{5} pg per 1.0 x 10^{13} vg, less than 1.1 x 10^{5} pg per 1.0 x 10^{13} vg, less than 1.0 x 10^{5} pg hcDNA per 1.0 x

10^{13} vg, less than 0.9 x 10^5 pg hcDNA per 1.0 x 10^{13} vg, less than 0.8 x 10^5 pg hcDNA per 1.0 x 10^{13} vg, or any concentration in between.

In some embodiments, the residual plasmid DNA in the pharmaceutical composition is less than or equal to 1.7 x 10^5 pg/ml per 1.0 x 10^{13} vg/ml, or 1 x 10^5 pg/ml per 1 x 1.0 x 10^{13} vg/ml, or 1.7 x 10^6 pg/ml per 1.0 x 10^{13} vg/ml. In some embodiments, the residual DNA plasmid in the pharmaceutical composition is less than 10.0 x 10^5 pg by 1.0 x 10^{13} vg, less than 8.0 x 10^5 pg by 1.0 x 10^{13} vg or less than 6.8 x 10^5 pg by 1.0 x 10^{13} vg. In embodiments, the pharmaceutical composition comprises less than 0.5 ng per 1.0 x 10^{13} vg, less than 0.3 ng per 1.0 x 10^{13} vg, less than 0.22 ng per 1.0 x 10^{13} vg or less than 0.2 ng per 1.0 x 10^{13} vg or any intermediate concentration of bovine serum albumin (BSA). In embodiments, the benzonase in the pharmaceutical composition is less than 0.2 ng by 1.0 x 10^{13} vg, less than 0.1 ng by 1.0 x 10^{13} vg, less than 0.09 ng by 1.0 x 10^{13} vg, less than 0.08 ng by 1.0 x 10^{13} vg or any intermediate concentration. In embodiments, Poloxamer 188 in the pharmaceutical composition is about 10 to 150 ppm, about 15 to 100 ppm or about 20 to 80 ppm. In embodiments, the cesium in the pharmaceutical composition is less than 50 pg / g (ppm), less than 30 pg / g (ppm) or less than 20 pg / g (ppm) or any intermediate concentration.

In embodiments, the pharmaceutical composition comprises total impurities, e.g., as determined by SDS-PAGE, of less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or any percentage in between. In embodiments, the total purity, e.g., as determined by SDS-PAGE, is greater than 90%, greater than 92%, greater than 93%, greater than 94%, greater than 95%, greater than 96%, greater than 97%, greater than 98%, or any percentage in between. In embodiments, no single unnamed related impurity, e.g., as measured by SDS-PAGE, is greater than 5%, greater than 4%, greater than 3% or greater than 2%, or any percentage in between. In embodiments, the pharmaceutical composition comprises a percentage of filled capsids relative to total capsids (e.g., peak 1 + peak 2 as measured by analytical ultracentrifugation) of greater than 85%, greater than 86%, greater than 87%, greater than 88%, greater than 89%, greater than 90%, greater than 91%, greater than 91.9%, greater than 92%, greater than 93%, or any percentage in between. In embodiments of the pharmaceutical composition, the percentage of filled capsids measured in peak 1 by analytical ultracentrifugation is 20-80%, 25-75%, 30-75%, 35-75%, or 37.4-70.3%. In embodiments of the

pharmaceutical composition, the percentage of filled capsids measured in peak 2 by analytical ultracentrifugation is 20-80%, 20-70%, 22-65%, 24-62%, or 24.9-60.1%.

In one embodiment, the pharmaceutical composition comprises a genomic titer of 1.0 to 5.0 x 10^{13} vg / mL, 1.2 to 3.0 x 10^{13} vg / mL or 1.7 to 2.3 x 10^{13} vg / ml. In one embodiment, the pharmaceutical composition exhibits a biological load of less than 5 CFU / mL, less than 4 CFU / mL, less than 3 CFU / mL, less than 2 CFU / mL or less than 1 CFU / mL or any intermediate contraction. In embodiments, the amount of endotoxin according to USP, for example, USP <85> (incorporated by reference in its entirety) is less than 1.0 EU / mL, less than 0.8 EU / mL or less than 0.75 EU / mL. In embodiments, the osmolarity of a pharmaceutical composition according to USP, for example, USP <785> (incorporated by reference in its entirety) is 350 to 450 mOsm / kg, 370 to 440 mOsm / kg or 390 to 430 mOsm / kg. In embodiments, the pharmaceutical composition contains less than 1200 particles that are greater than 25 μm per container, less than 1000 particles that are greater than 25 μm per container, less than 500 particles that are greater than 25 μm per container or any intermediate value. In embodiments, the pharmaceutical composition contains less than 10,000 particles that are greater than 10 μm per container, less than 8000 particles that are greater than 10 μm per container or less than 600 particles that are greater than 10 pm per container.

In one embodiment, the pharmaceutical composition has a genomic titer of 0.5 to 5.0 x 10^{13} vg / mL, 1.0 to 4.0 x 10^{13} vg / mL, 1.5 to 3.0 x 10^{13} vg / ml or 1.7 to 2.3 x 10^{13} vg / ml. In one embodiment, the pharmaceutical composition described herein comprises one or more of the following: less than about 0.09 ng benzonase per 1.0 x 10^{13} vg, less than about 30 pg / g (ppm) of cesium, about 20 to 80 ppm Poloxamer 188, less than about 0.22 ng BSA per 1.0 x 10^{13} vg, less than about 6.8 x 10^{5} pg of residual DNA plasmid per 1.0 x 10^{13} vg, less than about 1.1 x 10^{5} pg of residual hcDNA per 1.0 x 10^{13} vg, less than about 4 ng of rHCP per 1.0 x 10^{13} vg, pH 7.7 to 8.3, about 390 to 430 mOsm / kg, less than about 600 particles that are > 25 μm in size per container, less than about 6000 particles that are > 10 μm in size per container, about 1.7 x 10^{13} - 2.3 x 10^{13} vg / mL genomic titer, infectious titer of about 3.9 x 10^{8} to 8.4 x 10^{10} IU per 1.0 x 10^{13} vg, total protein of about 100-300 pg per 1.0 x 10^{13} vg, mean survival of >24 days in A7SMA mice with about 7.5 x 10^{13} vg / kg dose of viral vector, about 70 to 130% relative potency based on an in vitro cell based assay and / or less than about 5% empty capsid. In various embodiments, the pharmaceutical compositions described herein comprise any of the

viral particles discussed here, retain a potency of between ± 20%, between ± 15%, between ± 10% or within ± 5% of a reference standard. In some embodiments, potency is measured using a suitable in vitro cell assay or in vivo animal model.

Additional methods of preparation, characterization, and dosing AAV particles are taught in WO2019094253, which is incorporated herein by reference in its entirety.

Additional rAAV constructs that can be employed consonant with the invention include those described in Wang et al 2019, available at: //doi.org/10.1038/s41573-019-0012-9, including Table 1 thereof, which is incorporated by reference in its entirety.

AAV Administration

In some embodiments, an adeno-associated virus (AAV) is used in conjunction with the system, template nucleic acid, and/or polypeptide described herein. In some embodiments, an AAV is used to deliver, administer, or package the system, template nucleic acid, and/or polypeptide described herein. In some embodiments, the AAV is a recombinant AAV (rAAV).

In some embodiments, a system comprises (a) a polypeptide described herein or a nucleic acid encoding the same, (b) a template nucleic acid (e.g., template RNA) described herein, and (c) one or more first tissue-specific expression-control sequences specific to the target tissue, wherein the one or more first tissue-specific expression-control sequences specific to the target tissue are in operative association with (a), (b), or (a) and (b), wherein, when associated with (a), (a) comprises a nucleic acid encoding the polypeptide.

In some embodiments, a system described herein further comprises a first recombinant adeno-associated virus (rAAV) capsid protein; wherein the at least one of (a) or (b) is associated with the first rAAV capsid protein, wherein at least one of (a) or (b) is flanked by AAV inverted terminal repeats (ITRs) .

In some embodiments, (a) and (b) are associated with the first rAAV capsid protein.

In some embodiments, (a) and (b) are on a single nucleic acid.

In some embodiments, the system further comprises a second rAAV capsid protein, wherein at least one of (a) or (b) is associated with the second rAAV capsid protein, and wherein the at least one of (a) or (b) associated with the second rAAV capsid protein is different from the at least one of (a) or (b) is associated with the first rAAV capsid protein.

In some embodiments, the at least one of (a) or (b) is associated with the first or second rAAV capsid protein is dispersed in the interior of the first or second rAAV capsid protein, which first or second rAAV capsid protein is in the form of an AAV capsid particle.

In some embodiments, the system further comprises a nanoparticle, wherein the nanoparticle is associated with at least one of (a) or (b).

In some embodiments, (a) and (b), respectively are associated with: a) a first rAAV capsid protein and a second rAAV capsid protein; b) a nanoparticle and a first rAAV capsid protein; c) a first rAAV capsid protein; d) a first adenovirus capsid protein; e) a first nanoparticle and a second nanoparticle; or f) a first nanoparticle.

Viral vectors are useful for delivering all or part of a system provided by the invention, e.g., for use in methods provided by the invention. Systems derived from different viruses have been employed for the delivery of polypeptides, nucleic acids, or transposons; for example: integrase-deficient lentivirus, adenovirus, adeno-associated virus (AAV), herpes simplex virus, and baculovirus (reviewed in Hodge et al. Hum Gene Ther 2017; Narayanavari et al. Crit Rev Biochem Mol Biol 2017; Boehme et al. Curr Gene Ther 2015).

Adenoviruses are common viruses that have been used as gene delivery vehicles given well-defined biology, genetic stability, high transduction efficiency, and ease of large-scale production (see, for example, review by Lee et al. Genes & Diseases 2017). They possess linear dsDNA genomes and come in a variety of serotypes that differ in tissue and cell tropisms. In order to prevent replication of infectious virus in recipient cells, adenovirus genomes used for packaging are deleted of some or all endogenous viral proteins, which are provided in trans in viral production cells. This renders the genomes helper-dependent, meaning they can only be replicated and packaged into viral particles in the presence of the missing components provided by so-called helper functions. A helper-dependent adenovirus system with all viral ORFs removed may be compatible with packaging foreign DNA of up to ~37 kb (Parks et al. J Virol 1997). In some embodiments, an adenoviral vector is used to deliver DNA corresponding to the polypeptide or template component of the Gene Writing™ system, or both are contained on separate or the same adenoviral vector. In some embodiments, the adenovirus is a helper-dependent adenovirus (HD-AdV) that is incapable of self-packaging. In some embodiments, the adenovirus is a high-capacity adenovirus (HC-AdV) that has had all or a substantial portion of endogenous viral ORFs deleted, while retaining the necessary sequence components for

packaging into adenoviral particles. For this type of vector, the only adenoviral sequences required for genome packaging are noncoding sequences: the inverted terminal repeats (ITRs) at both ends and the packaging signal at the 5'-end (Jager et al. Nat Protoc 2009). In some embodiments, the adenoviral genome also comprises stuffer DNA to meet a minimal genome size for optimal production and stability (see, for example, Hausl et al. Mol Ther 2010). Adenoviruses have been used in the art for the delivery of transposons to various tissues. In some embodiments, an adenovirus is used to deliver a Gene Writing™ system to the liver.

In some embodiments, an adenovirus is used to deliver a Gene Writing™ system to HSCs, e.g., HDAd5/35++. HDAd5/35++ is an adenovirus with modified serotype 35 fibers that de-target the vector from the liver (Wang et al. Blood Adv 2019). In some embodiments, the adenovirus that delivers a Gene Writing™ system to HSCs utilizes a receptor that is expressed specifically on primitive HSCs, e.g., CD46.

Adeno-associated viruses (AAV) belong to the parvoviridae family and more specifically constitute the dependoparvovirus genus. The AAV genome is composed of a linear single-stranded DNA molecule which contains approximately 4.7 kilobases (kb) and consists of two major open reading frames (ORFs) encoding the non-structural Rep (replication) and structural Cap (capsid) proteins. A second ORF within the cap gene was identified that encodes the assembly-activating protein (AAP). The DNAs flanking the AAV coding regions are two cis-acting inverted terminal repeat (ITR) sequences, approximately 145 nucleotides in length, with interrupted palindromic sequences that can be folded into energetically stable hairpin structures that function as primers of DNA replication. In addition to their role in DNA replication, the ITR sequences have been shown to be involved in viral DNA integration into the cellular genome, rescue from the host genome or plasmid, and encapsidation of viral nucleic acid into mature virions (Muzyczka, (1992) Curr. Top. Micro. Immunol. 158:97-129). In some embodiments, one or more Gene Writing™ nucleic acid components is flanked by ITRs derived from AAV for viral packaging. See, e.g., WO2019113310.

In some embodiments, one or more components of the Gene Writing™ system are carried via at least one AAV vector. In some embodiments, the at least one AAV vector is selected for tropism to a particular cell, tissue, organism. In some embodiments, the AAV vector is pseudotyped, e.g., AAV2/8, wherein AAV2 describes the design of the construct but the capsid protein is replaced by that from AAV8. It is understood that any of the described vectors

could be pseudotype derivatives, wherein the capsid protein used to package the AAV genome is derived from that of a different AAV serotype. In some embodiments, an AAV to be employed for Gene Writing™ may be evolved for novel cell or tissue tropism as has been demonstrated in the literature (e.g., Davidsson et al. Proc Natl Acad Sci U S A 2019).

In some embodiments, the AAV delivery vector is a vector which has two AAV inverted terminal repeats (ITRs) and a nucleotide sequence of interest (for example, a sequence coding for a Gene Writer™ polypeptide or a DNA template, or both), each of said ITRs having an interrupted (or noncontiguous) palindromic sequence, i.e., a sequence composed of three segments: a first segment and a last segment that are identical when read 5'→ 3' but hybridize when placed against each other, and a segment that is different that separates the identical segments. Such sequences, notably the ITRs, form hairpin structures. See, for example, WO2012123430.

Conventionally, AAV virions with capsids are produced by introducing a plasmid or plasmids encoding the rAAV or scAAV genome, Rep proteins, and Cap proteins (Grimm et al, 1998). Upon introduction of these helper plasmids in trans, the AAV genome is "rescued" (i.e., released and subsequently recovered) from the host genome, and is further encapsidated to produce infectious AAV. In some embodiments, one or more Gene Writing™ nucleic acids are packaged into AAV particles by introducing the ITR-flanked nucleic acids into a packaging cell in conjunction with the helper functions.

In some embodiments, the AAV genome is a so called self-complementary genome (referred to as scAAV), such that the sequence located between the ITRs contains both the desired nucleic acid sequence (e.g., DNA encoding the Gene Writer™ polypeptide or template, or both) in addition to the reverse complement of the desired nucleic acid sequence, such that these two components can fold over and self-hybridize. In some embodiments, the self-complementary modules are separated by an intervening sequence that permits the DNA to fold back on itself, e.g., forms a stem-loop. An scAAV has the advantage of being poised for transcription upon entering the nucleus, rather than being first dependent on ITR priming and second-strand synthesis to form dsDNA. In some embodiments, one or more Gene Writing™ components is designed as an scAAV, wherein the sequence between the AAV ITRs contains two reverse complementing modules that can self-hybridize to create dsDNA.

In some embodiments, nucleic acid (e.g., encoding a polypeptide, or a template, or both) delivered to cells is closed-ended, linear duplex DNA (CELiD DNA or ceDNA). In some embodiments, ceDNA is derived from the replicative form of the AAV genome (Li et al. PLoS One 2013). In some embodiments, the nucleic acid (e.g., encoding a polypeptide, or a template DNA, or both) is flanked by ITRs, e.g., AAV ITRs, wherein at least one of the ITRs comprises a terminal resolution site and a replication protein binding site (sometimes referred to as a replicative protein binding site). In some embodiments, the ITRs are derived from an adeno-associated virus, e.g., AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, or a combination thereof. In some embodiments, the ITRs are symmetric. In some embodiments, the ITRs are asymmetric. In some embodiments, at least one Rep protein is provided to enable replication of the construct. In some embodiments, the at least one Rep protein is derived from an adeno-associated virus, e.g., AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, or a combination thereof. In some embodiments, ceDNA is generated by providing a production cell with (i) DNA flanked by ITRs, e.g., AAV ITRs, and (ii) components required for ITR-dependent replication, e.g., AAV proteins Rep78 and Rep52 (or nucleic acid encoding the proteins). In some embodiments, ceDNA is free of any capsid protein, e.g., is not packaged into an infectious AAV particle. In some embodiments, ceDNA is formulated into LNPs (see, for example, WO2019051289A1).

In some embodiments, the ceDNA vector consists of two self complementary sequences, e.g., asymmetrical or symmetrical or substantially symmetrical ITRs as defined herein, flanking said expression cassette, wherein the ceDNA vector is not associated with a capsid protein. In some embodiments, the ceDNA vector comprises two self-complementary sequences found in an AAV genome, where at least one ITR comprises an operative Rep-binding element (RBE) (also sometimes referred to herein as "RBS") and a terminal resolution site (trs) of AAV or a functional variant of the RBE. See, for example, WO2019113310.

Inteins

In some embodiments, as described in more detail below, Intein-N may be fused to the N-terminal portion of a first domain described herein, and and intein-C may be fused to the C-terminal portion of a second domain described herein for the joining of the N-terminal portion to the C-terminal portion, thereby joining the first and second domains. In some embodiments, the

first and second domains are each independent chosen from a DNA binding domain, an RNA binding domain, an RT domain, and an endonuclease domain.

As used herein, "intein" refers to a self-splicing protein intron (e.g., peptide), e.g., which ligates flanking N-terminal and C-terminal exteins (e.g., fragments to be joined). An intein may, in some instances, comprise a fragment of a protein that is able to excise itself and join the remaining fragments (the exteins) with a peptide bond in a process known as protein splicing. Inteins are also referred to as "protein introns." The process of an intein excising itself and joining the remaining portions of the protein is herein termed "protein splicing" or "intein-mediated protein splicing." In some embodiments, an intein of a precursor protein (an intein containing protein prior to intein-mediated protein splicing) comes from two genes. Such intein is referred to herein as a split intein (e.g., split intein-N and split intein-C). For example, in cyanobacteria, DnaE, the catalytic subunit a of DNA polymerase III, is encoded by two separate genes, dnaE-n and dnaE-c. The intein encoded by the dnaE-n gene may be herein referred as "intein-N." The intein encoded by the dnaE-c gene may be herein referred as "intein-C."

Use of inteins for joining heterologous protein fragments is described, for example, in Wood et al., J. Biol. Chem.289(21); 14512-9 (2014) (incorporated herein by reference in its entirety). For example, when fused to separate protein fragments, the inteins IntN and IntC may recognize each other, splice themselves out, and/or simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused, thereby reconstituting a full-length protein from the two protein fragments.

In some embodiments, a synthetic intein based on the dnaE intein, the Cfa-N (e.g., split intein-N) and Cfa-C (e.g., split intein-C) intein pair, is used. Examples of such inteins have been described, e.g., in Stevens et al., J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5 (incorporated herein by reference in its entirety). Non-limiting examples of intein pairs that may be used in accordance with the present disclosure include: Cfa DnaE intein, Ssp GyrB intein, Ssp DnaX intein, Ter DnaE3 intein, Ter ThyX intein, Rma DnaB intein and Cne Prp8 intein (e.g., as described in U.S. Pat. No. 8,394,604, incorporated herein by reference.

In some embodiments, Intein-N and intein-C may be fused to the N-terminal portion of the split Cas9 and the C-terminal portion of a split Cas9, respectively, for the joining of the N-terminal portion of the split Cas9 and the C-terminal portion of the split Cas9. For example, in some embodiments, an intein-N is fused to the C-terminus of the N-terminal portion of the split

Cas9, i.e., to form a structure of N— [N-terminal portion of the split Cas9]-[intein-N]~ C. In some embodiments, an intein-C is fused to the N-terminus of the C-terminal portion of the split Cas9, i.e., to form a structure of N-[intein-C]~ [C-terminal portion of the split Cas9]-C. The mechanism of intein-mediated protein splicing for joining the proteins the inteins are fused to (e.g., split Cas9) is described in Shah et al., Chem Sci. 2014; 5(l):446-46l, incorporated herein by reference. Methods for designing and using inteins are known in the art and described, for example by WO2020051561, W02014004336, WO2017132580, US20150344549, and US20180127780, each of which is incorporated herein by reference in their entirety.

In some embodiments, a split refers to a division into two or more fragments. In some embodiments, a split Cas9 protein or split Cas9 comprises a Cas9 protein that is provided as an N-terminal fragment and a C-terminal fragment encoded by two separate nucleotide sequences. The polypeptides corresponding to the N-terminal portion and the C-terminal portion of the Cas9 protein may be spliced to form a reconstituted Cas9 protein. In embodiments, the Cas9 protein is divided into two fragments within a disordered region of the protein, e.g., as described in Nishimasu et al., Cell, Volume 156, Issue 5, pp. 935-949, 2014, or as described in Jiang et al. (2016) Science 351: 867-871 and PDB file: 5F9R (each of which is incorporated herein by reference in its entirety). A disordered region may be determined by one or more protein structure determination techniques known in the art, including, without limitation, X-ray crystallography, NMR spectroscopy, electron microscopy (e.g., cryoEM), and/or in silico protein modeling. In some embodiments, the protein is divided into two fragments at any C, T, A, or S, e.g., within a region of SpCas9 between amino acids A292- G364, F445-K483, or E565-T637, or at corresponding positions in any other Cas9, Cas9 variant (e.g., nCas9, dCas9), or other napDNAbp. In some embodiments, protein is divided into two fragments at SpCas9 T310, T313, A456, S469, or C574. In some embodiments, the process of dividing the protein into two fragments is referred to as splitting the protein.

In some embodiments, a protein fragment ranges from about 2-1000 amino acids (e.g., between 2-10, 10-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 amino acids) in length. In some embodiments, a protein fragment ranges from about 5-500 amino acids (e.g., between 5-10, 10-50, 50-100, 100-200, 200-300, 300-400, or 400-500 amino acids) in length. In some embodiments, a protein fragment ranges from about 20-200 amino acids (e.g., between 20-30, 30-40, 40-50, 50-100, or 100-200 amino acids) in length.

In some embodiments, a portion or fragment of a Gene Writer (e.g., Cas9-R2Tg) is fused to an intein. The nuclease can be fused to the N-terminus or the C-terminus of the intein. In some embodiments, a portion or fragment of a fusion protein is fused to an intein and fused to an AAV capsid protein. The intein, nuclease and capsid protein can be fused together in any arrangement (e.g., nuclease-intein-capsid, intein-nuclease-capsid, capsid-intein-nuclease, etc.). In some embodiments, the N-terminus of an intein is fused to the C-terminus of a fusion protein and the C-terminus of the intein is fused to the N-terminus of an AAV capsid protein.

In some embodiments, an endonuclease domain (e.g., a nickase Cas9 domain) is fused to intein-N and a polypeptide comprising an RT domain is fused to an intein-C.

Exemplary nucleotide and amino acid sequences of interns are provided below:

DnaE Intein-N DNA:
TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGG
AAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAA
CATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCG
AATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATG
ACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTC
ATGCGAGTTGACAACCTTCCTAAT

DnaE Intein-N Protein:
CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL
EDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

DnaE Intein-C DNA:
ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGG
AGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT

Intein-C:
MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

Cfa-N DNA:
TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAA
AGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTC
GTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGA
GTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGA
CCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTC
AAACAAGTGGATGGATTG CCA

Cfa-N Protein:
CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCL
EDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

Cfa-C DNA:
ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGT
AAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGA
GAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC

Cfa-C Protein:
MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

Lipid Nanoparticles

The methods and systems provided by the invention, may employ any suitable carrier or delivery modality, including, in certain embodiments, lipid nanoparticles (LNPs). Lipid nanoparticles, in some embodiments, comprise one or more ionic lipids, such as non-cationic lipids (e.g., neutral or anionic, or zwitterionic lipids); one or more conjugated lipids (such as PEG-conjugated lipids or lipids conjugated to polymers described in Table 5 of WO2019217941; incorporated herein by reference in its entirety); one or more sterols (e.g., cholesterol); and, optionally, one or more targeting molecules (e.g., conjugated receptors, receptor ligands, antibodies); or combinations of the foregoing.

Lipids that can be used in nanoparticle formations (e.g., lipid nanoparticles) include, for example those described in Table 4 of WO2019217941, which is incorporated by reference—e.g., a lipid-containing nanoparticle can comprise one or more of the lipids in table 4 of WO2019217941. Lipid nanoparticles can include additional elements, such as polymers, such as the polymers described in table 5 of WO2019217941, incorporated by reference.

In some embodiments, conjugated lipids, when present, can include one or more of PEG-diacylglycerol (DAG) (such as l-(monomethoxy-polyethyleneglycol)-2,3- dimyristoylglycerol (PEG-DMG)), PEG-dialkyloxypropyl (DAA), PEG-phospholipid, PEG- ceramide (Cer), a pegylated phosphatidylethanoloamine (PEG-PE), PEG succinate diacylglycerol (PEGS-DAG) (such as 4-0-(2',3'-di(tetradecanoyloxy)propyl-l-0-(w- methoxy(polyethoxy)ethyl) butanedioate (PEG-S-DMG)), PEG dialkoxypropylcarbam, N- (carbonyl-methoxypoly ethylene glycol 2000)-1 ,2-distearoyl-sn-glycero-3-phosphoethanolamine sodium salt, and those described in Table 2 of WO2019051289 (incorporated by reference), and combinations of the foregoing.

In some embodiments, sterols that can be incorporated into lipid nanoparticles include one or more of cholesterol or cholesterol derivatives, such as those in W02009/127060 or US2010/0130588, which are incorporated by reference. Additional exemplary sterols include phytosterols, including those described in Eygeris et al (2020), dx.doi.org/10.1021/acs.nanolett.0c01386, incorporated herein by reference.

In some embodiments, the lipid particle comprises an ionizable lipid, a non-cationic lipid, a conjugated lipid that inhibits aggregation of particles, and a sterol. The amounts of these components can be varied independently and to achieve desired properties. For example, in some embodiments, the lipid nanoparticle comprises an ionizable lipid is in an amount from about 20 mol % to about 90 mol % of the total lipids (in other embodiments it may be 20-70% (mol), 30-60% (mol) or 40-50% (mol); about 50 mol % to about 90 mol % of the total lipid present in the lipid nanoparticle), a non-cationic lipid in an amount from about 5 mol % to about 30 mol % of the total lipids, a conjugated lipid in an amount from about 0.5 mol % to about 20 mol % of the total lipids, and a sterol in an amount from about 20 mol % to about 50 mol % of the total lipids. The ratio of total lipid to nucleic acid (e.g., encoding the Gene Writer or template nucleic acid) can be varied as desired. For example, the total lipid to nucleic acid (mass or weight) ratio can be from about 10: 1 to about 30: 1.

In some embodiments, an ionizable lipid may be a cationic lipid, an ionizable cationic lipid, e.g., a cationic lipid that can exist in a positively charged or neutral form depending on pH, or an amine-containing lipid that can be readily protonated. In some embodiments, the cationic lipid is a lipid capable of being positively charged, e.g., under physiological conditions. Exemplary cationic lipids include one or more amine group(s) which bear the positive charge. In some embodiments, the lipid particle comprises a cationic lipid in formulation with one or more of neutral lipids, ionizable amine-containing lipids, biodegradable alkyn lipids, steroids, phospholipids including polyunsaturated lipids, structural lipids (e.g., sterols), PEG, cholesterol and polymer conjugated lipids. In some embodiments, the cationic lipid may be an ionizable cationic lipid. An exemplary cationic lipid as disclosed herein may have an effective pKa over 6.0. In embodiments, a lipid nanoparticle may comprise a second cationic lipid having a different effective pKa (e.g., greater than the first effective pKa), than the first cationic lipid. A lipid nanoparticle may comprise between 40 and 60 mol percent of a cationic lipid, a neutral lipid, a steroid, a polymer conjugated lipid, and a therapeutic agent, e.g., a nucleic acid (e.g., RNA)

described herein (e.g., a template nucleic acid or a nucleic acid encoding a GeneWriter), encapsulated within or associated with the lipid nanoparticle. In some embodiments, the nucleic acid is co-formulated with the cationic lipid. The nucleic acid may be adsorbed to the surface of an LNP, e.g., an LNP comprising a cationic lipid. In some embodiments, the nucleic acid may be encapsulated in an LNP, e.g., an LNP comprising a cationic lipid. In some embodiments, the lipid nanoparticle may comprise a targeting moiety, e.g., coated with a targeting agent. In embodiments, the LNP formulation is biodegradable. In some embodiments, a lipid nanoparticle comprising one or more lipid described herein, e.g., Formula (i), (ii), (ii), (vii) and/or (ix) encapsulates at least 1%, at least 5%, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98% or 100% of an RNA molecule, e.g., template RNA and/or a mRNA encoding the Gene Writer polypeptide.

In some embodiments, the lipid to nucleic acid ratio (mass/mass ratio; w/w ratio) can be in the range of from about 1 : 1 to about 25: 1, from about 10: 1 to about 14: 1, from about 3 : 1 to about 15: 1, from about 4: 1 to about 10: 1, from about 5: 1 to about 9: 1, or about 6: 1 to about 9: 1. The amounts of lipids and nucleic acid can be adjusted to provide a desired N/P ratio, for example, N/P ratio of 3, 4, 5, 6, 7, 8, 9, 10 or higher. Generally, the lipid nanoparticle formulation's overall lipid content can range from about 5 mg/ml to about 30 mg/mL.

Exemplary ionizable lipids that can be used in lipid nanoparticle formulations include, without limitation, those listed in Table 1 of WO2019051289, incorporated herein by reference. Additional exemplary lipids include, without limitation, one or more of the following formulae: X of US2016/0311759; I of US20150376115 or in US2016/0376224; I, II or III of US20160151284; I, IA, II, or IIA of US20170210967; I-c of US20150140070; A of US2013/0178541; I of US2013/0303587 or US2013/0123338; I of US2015/0141678; II, III, IV, or V of US2015/0239926; I of US2017/0119904; I or II of WO2017/117528; A of US2012/0149894; A of US2015/0057373; A of WO2013/116126; A of US2013/0090372; A of US2013/0274523; A of US2013/0274504; A of US2013/0053572; A of W02013/016058; A of W02012/162210; I of US2008/042973; I, II, III, or IV of US2012/01287670; I or II of US2014/0200257; I, II, or III of US2015/0203446; I or III of US2015/0005363; I, IA, IB, IC, ID, II, IIA, IIB, IIC, IID, or III-XXIV of US2014/0308304; of US2013/0338210; I, II, III, or IV of

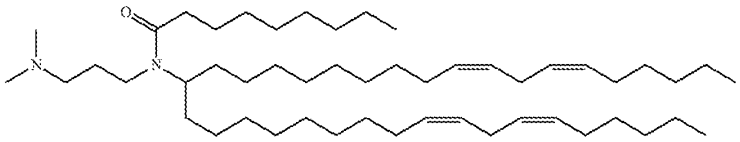
W02009/132131; A of US2012/01011478; I or XXXV of US2012/0027796; XIV or XVII of US2012/0058144; of US2013/0323269; I of US2011/0117125; I, II, or III of US2011/0256175; I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII of US2012/0202871; I, II, III, IV, V, VI, VII, VIII, X, XII, XIII, XIV, XV, or XVI of US2011/0076335; I or II of US2006/008378; I of US2013/0123338; I or X-A-Y-Z of US2015/0064242; XVI, XVII, or XVIII of US2013/0022649; I, II, or III of US2013/0116307; I, II, or III of US2013/0116307; I or II of US2010/0062967; I-X of US2013/0189351; I of US2014/0039032; V of US2018/0028664; I of US2016/0317458; I of US2013/0195920; 5, 6, or 10 of US10,221,127; III-3 of WO2018/081480; I-5 or I-8 of WO2020/081938; 18 or 25 of US9,867,888; A of US2019/0136231; II of WO2020/219876; 1 of US2012/0027803; OF-02 of US2019/0240349; 23 of US10,086,013; cKK-E12/A6 of Miao et al (2020); C12-200 of WO2010/053572; 7C1 of Dahlman et al (2017); 304-O13 or 503-O13 of Whitehead et al; TS-P4C2 of US9,708,628; I of WO2020/106946; I of WO2020/106946.

.

In some embodiments, the ionizable lipid is MC3 (6Z,9Z,28Z,3 lZ)-heptatriaconta-6,9,28,3 l-tetraen-l9-yl-4-(dimethylamino) butanoate (DLin-MC3-DMA or MC3), e.g., as described in Example 9 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is the lipid ATX-002, e.g., as described in Example 10 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is (l3Z,l6Z)-A,A-dimethyl-3- nonyldocosa-l3, l6-dien-l-amine (Compound 32), e.g., as described in Example 11 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is Compound 6 or Compound 22, e.g., as described in Example 12 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is heptadecan-9-yl 8-((2-hydroxyethyl)(6-oxo-6-(undecyloxy)hexyl)amino)octanoate (SM-102); e.g., as described in Example 1 of US9,867,888(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is 9Z,12Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9,12-dienoate (LP01) e.g., as synthesized in Example 13 of WO2015/095340(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is Di((Z)-non-2-en-1-yl) 9-((4-dimethylamino)butanoyl)oxy)heptadecanedioate (L319), , e.g. as synthesized in Example 7, 8, or

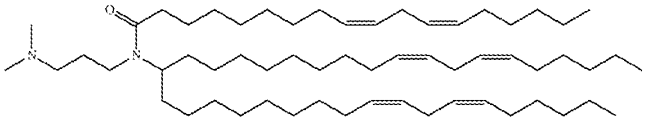
9 of US2012/0027803(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is 1,1'-((2-(4-(2-((2-(Bis(2-hydroxydodecyl)amino)ethyl)(2-hydroxydodecyl) amino)ethyl)piperazin-1-yl)ethyl)azanediyl)bis(dodecan-2-ol) (C12-200), e.g., as synthesized in Examples 14 and 16 of WO2010/053572(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is; Imidazole cholesterol ester (ICE) lipid (3S, 10R, 13R, 17R)-10, 13-dimethyl-17- ((R)-6-methylheptan-2-yl)-2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17-tetradecahydro-lH- cyclopenta[a]phenanthren-3-yl 3-(1H-imidazol-4-yl)propanoate, e.g., Structure (I) from WO2020/106946 (incorporated by reference herein in its entirety).

Some non-limiting example of lipid compounds that may be used (e.g., in combination with other lipid components) to form lipid nanoparticles for the delivery of compositions described herein, e.g., nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a GeneWriter) includes,



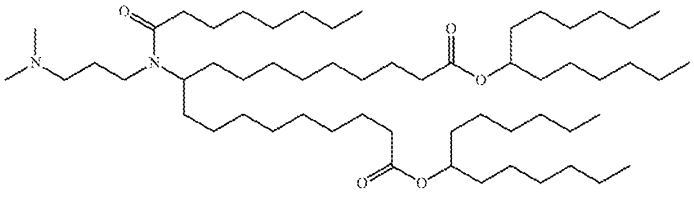
(i)

In some embodiments an LNP comprising Formula (i) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



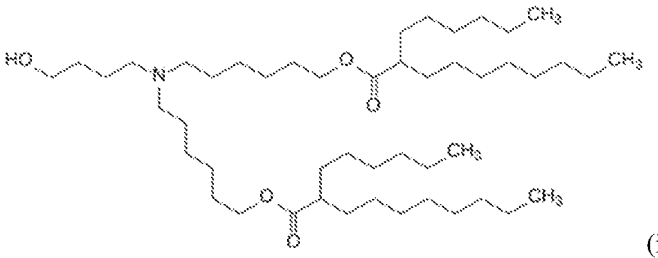
(ii)

In some embodiments an LNP comprising Formula (ii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

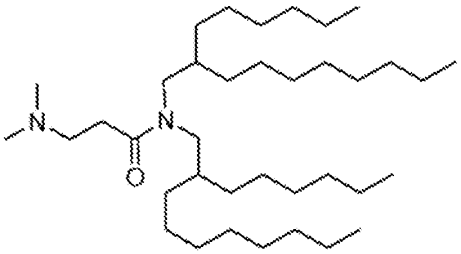


(iii)

In some embodiments an LNP comprising Formula (iii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

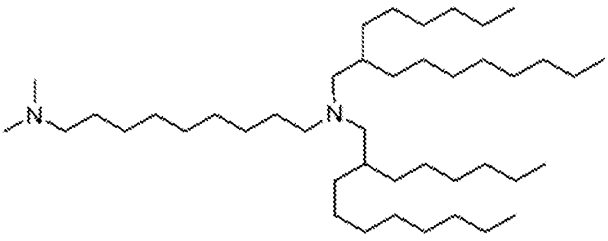


(iv)



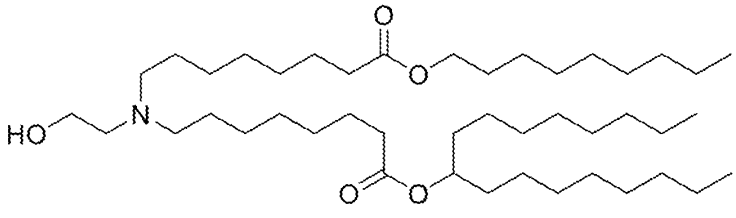
(v)

In some embodiments an LNP comprising Formula (v) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

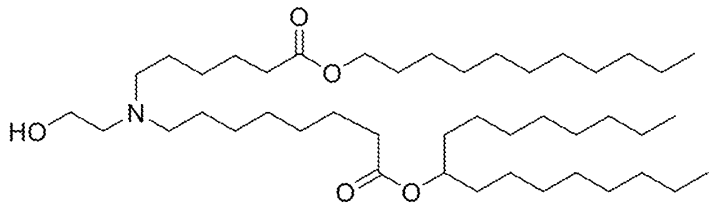


(vi)

In some embodiments an LNP comprising Formula (vi) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

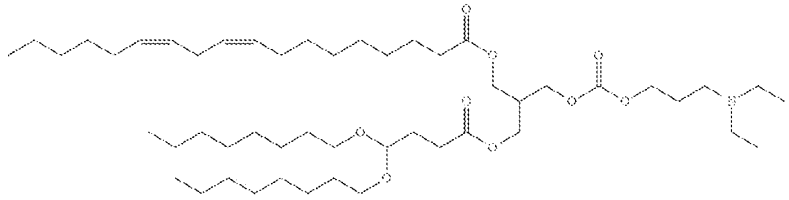


(vii)



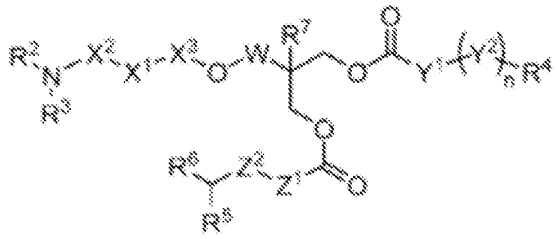
(viii)

In some embodiments an LNP comprising Formula (viii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



(ix)

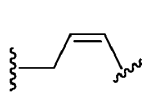
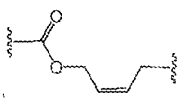
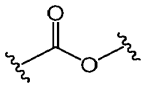
In some embodiments an LNP comprising Formula (ix) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



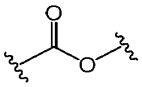
(x)

wherein

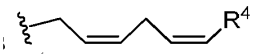
X^1 is O, NR^1, or a direct bond, X^2 is C2-5 alkylene, X^3 is C(=O) or a direct bond, R^1 is H or Me, R^3 is Ci-3 alkyl, R^2 is Ci-3 alkyl, or R^2 taken together with the nitrogen atom to which it is attached and 1-3 carbon atoms of X^2 form a 4-, 5-, or 6-membered ring, or X^1 is NR^1, R^1 and R^2 taken together with the nitrogen atoms to which they are attached form a 5- or 6-membered ring, or R^2 taken together with R^3 and the nitrogen atom to which they are attached form a 5-, 6-, or 7-membered ring, Y^1 is C2-12 alkylene, Y^2 is selected from


(in either orientation),

(in either orientation),

(in either orientation),

n is 0 to 3, R^4 is Ci-15 alkyl, Z^1 is Ci-6 alkylene or a direct bond,

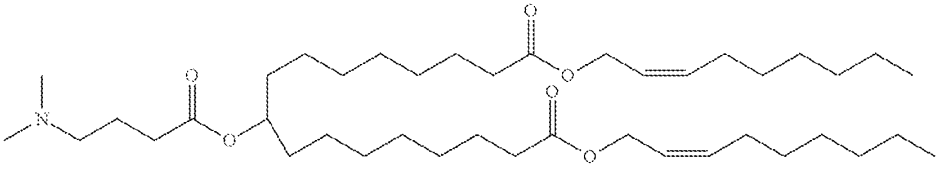
Z^2 is 
(in either orientation) or absent, provided that if Z^1 is a direct bond, Z^2 is absent;

R^5 is C5-9 alkyl or C6-10 alkoxy, R^6 is C5-9 alkyl or C6-10 alkoxy, W is methylene or a direct bond, and R^7 is H or Me, or a salt thereof, provided that if R^3 and R^2 are C2 alkyls, X^1 is O, X^2 is linear C3 alkylene, X^3 is C(=O), Y^1 is linear Ce alkylene, (Y^2)n-R^4 is



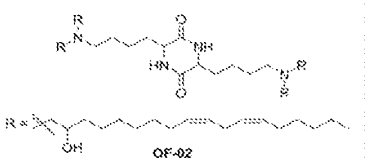
, R^4 is linear C5 alkyl, Z^1 is C2 alkylene, Z^2 is absent, W is methylene, and R^7 is H, then R^5 and R^6 are not Cx alkoxy.

In some embodiments an LNP comprising Formula (xii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

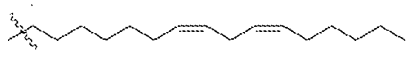


(xi)

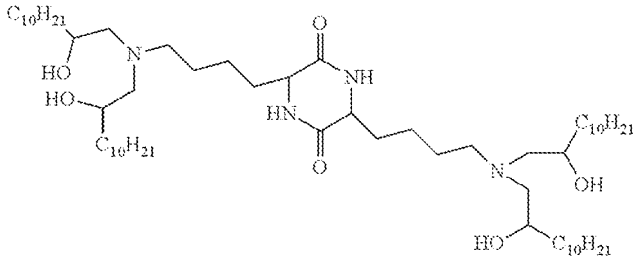
In some embodiments an LNP comprising Formula (xi) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



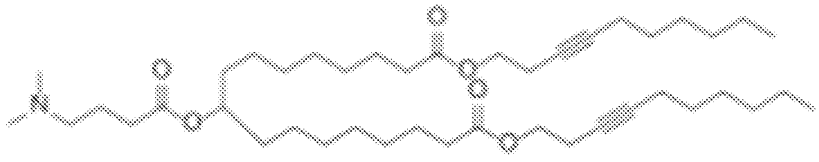
where R=



(xii)

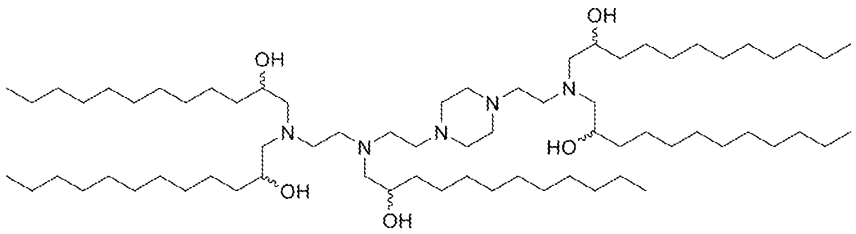


(xiii)



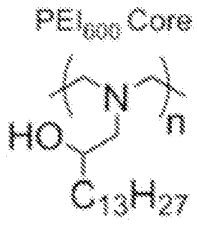
(xiv)

In some embodiments an LNP comprises a compound of Formula (xiii) and a compound of Formula (xiv).



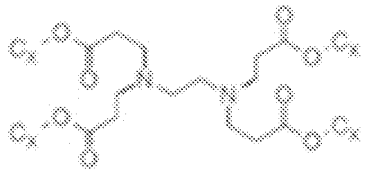
(xv)

In some embodiments an LNP comprising Formula (xv) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

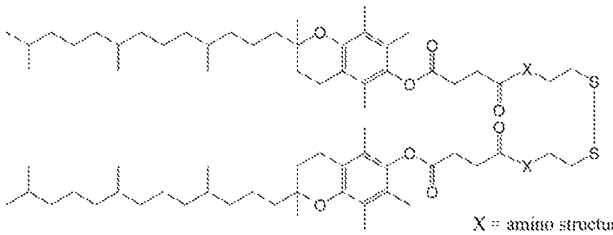


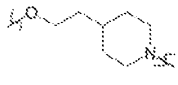
(xvi)

In some embodiments an LNP comprising a formulation of Formula (xvi) is used to deliver a GeneWriter composition described herein to the lung endothelial cells.



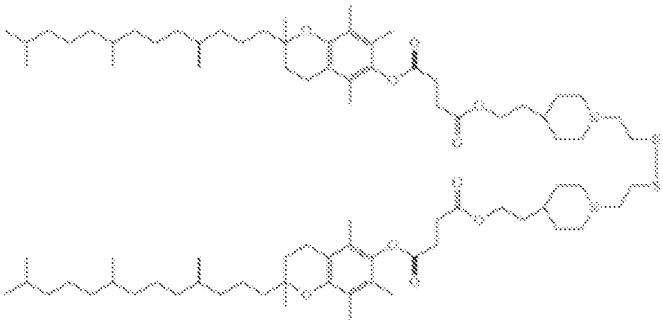
(xvii)



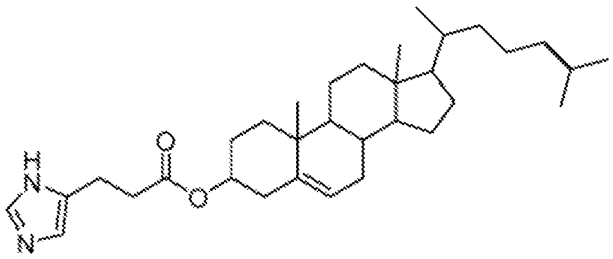
where X= 

(xviii)

(a)

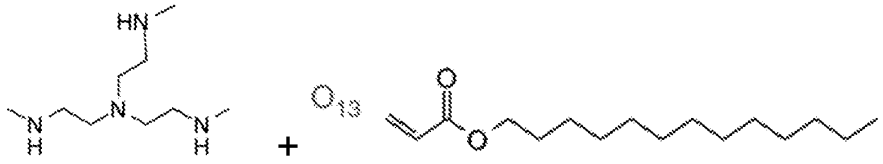


(xviii)(b)

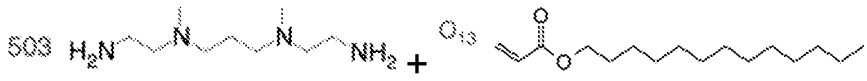


(xix)

In some embodiments, a lipid compound used to form lipid nanoparticles for the delivery of compositions described herein, e.g., nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a GeneWriter) is made by one of the following reactions:



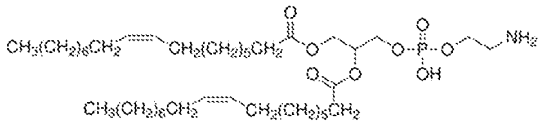
(xx) (a)



(xx)(b)

Exemplary non-cationic lipids include, but are not limited to, distearoyl-sn-glycero-phosphoethanolamine, distearoylphosphatidylcholine (DSPC), dioleoylphosphatidylcholine (DOPC), dipalmitoylphosphatidylcholine (DPPC), dioleoylphosphatidylglycerol (DOPG), 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), dipalmitoylphosphatidylglycerol (DPPG), dioleoyl-phosphatidylethanolamine (DOPE), palmitoyloleoylphosphatidylcholine (POPC),

palmitoyloleoylphosphatidylethanolamine (POPE), dioleoyl-phosphatidylethanolamine 4-(N-maleimidomethyl)-cyclohexane- 1 - carboxylate (DOPE-mal), dipalmitoyl phosphatidyl ethanolamine (DPPE), dimyristoylphosphoethanolamine (DMPE), distearoyl-phosphatidyl-ethanolamine (DSPE), monomethyl-phosphatidylethanolamine (such as 16-O-monomethyl PE), dimethyl- phosphatidylethanolamine (such as 16-O-dimethyl PE), l8-l-trans PE, l-stearoyl-2-oleoyl- phosphatidyethanolamine (SOPE), hydrogenated soy phosphatidylcholine (HSPC), egg phosphatidylcholine (EPC), dioleoylphosphatidylserine (DOPS), sphingomyelin (SM), dimyristoyl phosphatidylcholine (DMPC), dimyristoyl phosphatidylglycerol (DMPG), distearoylphosphatidylglycerol (DSPG), dierucoylphosphatidylcholine (DEPC), palmitoyloleyolphosphatidylglycerol (POPG), dielaidoyl-phosphatidylethanolamine (DEPE), lecithin, phosphatidylethanolamine, lysolecithin, lysophosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, sphingomyelin, egg sphingomyelin (ESM), cephalin, cardiolipin, phosphatidicacid,cerebrosides, dicetylphosphate, lysophosphatidylcholine, dilinoleoylphosphatidylcholine, or mixtures thereof. It is understood that other diacylphosphatidylcholine and diacylphosphatidylethanolamine phospholipids can also be used. The acyl groups in these lipids are preferably acyl groups derived from fatty acids having C10-C24 carbon chains, e.g., lauroyl, myristoyl, paimitoyl, stearoyl, or oleoyl. Additional exemplary lipids, in certain embodiments, include, without limitation, those described in Kim et al. (2020) dx.doi.org/10.1021/acs.nanolett.0c01386, incorporated herein by reference. Such lipids include, in some embodiments, plant lipids found to improve liver transfection with mRNA (e.g., DGTS In some embodiments, the non-cationic lipid may have the following structure,



(xxi)

Other examples of non-cationic lipids suitable for use in the lipid nanopartieles include, without limitation, nonphosphorous lipids such as, e.g., stearylamine, dodeeylamine, hexadecylamine, acetyl palmitate, glycerol ricinoleate, hexadecyl stereate, isopropyl myristate, amphoteric acrylic polymers, triethanolamine-lauryl sulfate, alkyl-aryl sulfate polyethyloxylated fatty acid amides, dioctadecyl dimethyl ammonium bromide, ceramide, sphingomyelin, and the

like. Other non-cationic lipids are described in WO2017/099823 or US patent publication US2018/0028664, the contents of which is incorporated herein by reference in their entirety.

In some embodiments, the non-cationic lipid is oleic acid or a compound of Formula I, II, or IV of US2018/0028664, incorporated herein by reference in its entirety. The non-cationic lipid can comprise, for example, 0-30% (mol) of the total lipid present in the lipid nanoparticle. In some embodiments, the non-cationic lipid content is 5-20% (mol) or 10-15% (mol) of the total lipid present in the lipid nanoparticle. In embodiments, the molar ratio of ionizable lipid to the neutral lipid ranges from about 2:1 to about 8:1 (e.g., about 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, or 8:1).

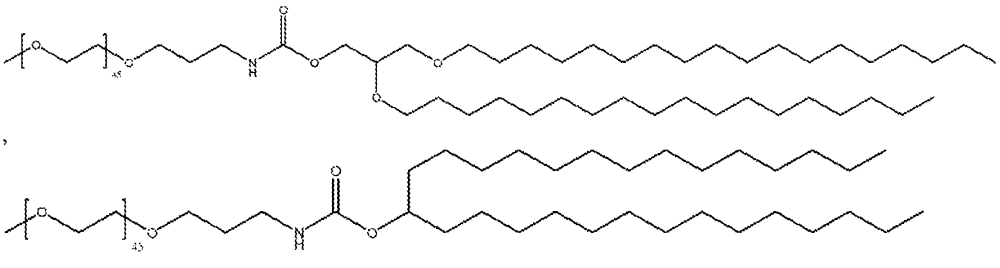
In some embodiments, the lipid nanoparticles do not comprise any phospholipids.

In some aspects, the lipid nanoparticle can further comprise a component, such as a sterol, to provide membrane integrity. One exemplary sterol that can be used in the lipid nanoparticle is cholesterol and derivatives thereof. Non-limiting examples of cholesterol derivatives include polar analogues such as 5a-choiestanol, 53-coprostanol, choiesteryl-(2'-hydroxy)-ethyl ether, choiesteryl-(4'- hydroxy)-butyl ether, and 6-ketocholestanol; non-polar analogues such as 5a-cholestane, cholestenone, 5a-cholestanone, 5p-cholestanone, and cholesteryl decanoate; and mixtures thereof. In some embodiments, the cholesterol derivative is a polar analogue, e.g., choiesteryl-(4 '-hydroxy)-buty1 ether. Exemplary cholesterol derivatives are described in PCT publication W02009/127060 and US patent publication US2010/0130588, each of which is incorporated herein by reference in its entirety.

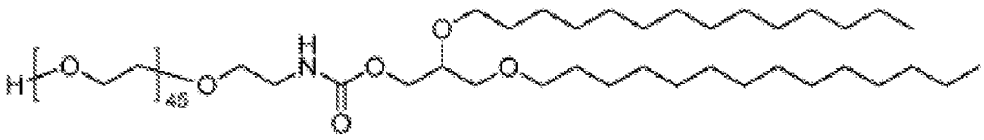
In some embodiments, the component providing membrane integrity, such as a sterol, can comprise 0-50% (mol) (e.g., 0-10%, 10-20%, 20-30%, 30-40%, or 40-50%) of the total lipid present in the lipid nanoparticle. In some embodiments, such a component is 20-50% (mol) 30-40% (mol) of the total lipid content of the lipid nanoparticle.

In some embodiments, the lipid nanoparticle can comprise a polyethylene glycol (PEG) or a conjugated lipid molecule. Generally, these are used to inhibit aggregation of lipid nanoparticles and/or provide steric stabilization. Exemplary conjugated lipids include, but are not limited to, PEG-lipid conjugates, polyoxazoline (POZ)-lipid conjugates, polyamide-lipid conjugates (such as ATTA-lipid conjugates), cationic-polymer lipid (CPL) conjugates, and mixtures thereof. In some embodiments, the conjugated lipid molecule is a PEG-lipid conjugate, for example, a (methoxy polyethylene glycol)-conjugated lipid.

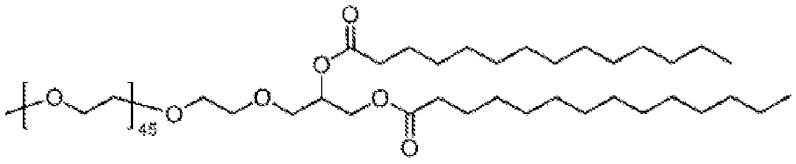
Exemplary PEG-lipid conjugates include, but are not limited to, PEG-diacylglycerol (DAG) (such as l-(monomethoxy-polyethyleneglycol)-2,3-dimyristoylglycerol (PEG-DMG)), PEG-dialkyloxypropyl (DAA), PEG-phospholipid, PEG-ceramide (Cer), a pegylated phosphatidylethanoloamine (PEG-PE), 1,2-dimyristoyl-sn-glycerol, methoxypoly ethylene glycol (DMG-PEG-2K), PEG succinate diacylglycerol (PEGS-DAG) (such as 4-0-(2',3'-di(tetradecanoyloxy)propyl-l-0-(w-methoxy(polyethoxy)ethyl) butanedioate (PEG-S-DMG)), PEG dialkoxypropylcarbam, N-(carbonyl-methoxypolyethylene glycol 2000)-l,2-distearoyl-sn-glycero-3-phosphoethanolamine sodium salt, or a mixture thereof. Additional exemplary PEG-lipid conjugates are described, for example, in US5,885,6l3, US6,287,59l, US2003/0077829, US2003/0077829, US2005/0175682, US2008/0020058, US2011/0117125, US2010/0130588, US2016/0376224, US2017/0119904, and US/099823, the contents of all of which are incorporated herein by reference in their entirety. In some embodiments, a PEG-lipid is a compound of Formula III, III-a-I, III-a-2, III-b-1, III-b-2, or V of US2018/0028664, the content of which is incorporated herein by reference in its entirety. In some embodiments, a PEG-lipid is of Formula II of US20150376115 or US2016/0376224, the content of both of which is incorporated herein by reference in its entirety. In some embodiments, the PEG-DAA conjugate can be, for example, PEG-dilauryloxypropyl, PEG- dimyristyloxypropyl, PEG-dipalmityloxypropyl, or PEG-distearyloxypropyl. The PEG-lipid can be one or more of PEG-DMG, PEG-dilaurylglycerol, PEG-dipalmitoylglycerol, PEG- disterylglycerol, PEG-dilaurylglycamide, PEG-dimyristylglycamide, PEG- dipalmitoylglycamide, PEG-disterylglycamide, PEG-cholesterol (l-[8'-(Cholest-5-en-3[beta]- oxy)carboxamido-3',6'-dioxaoctanyl] carbamoyl-[omega]-methyl-poly(ethylene glycol), PEG- DMB (3,4-Ditetradecoxylbenzyl- [omega]-methyl-poly(ethylene glycol) ether), and 1,2- dimyristoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000]. In some embodiments, the PEG-lipid comprises PEG-DMG, 1,2- dimyristoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000]. In some embodiments, the PEG-lipid comprises a structure selected from:



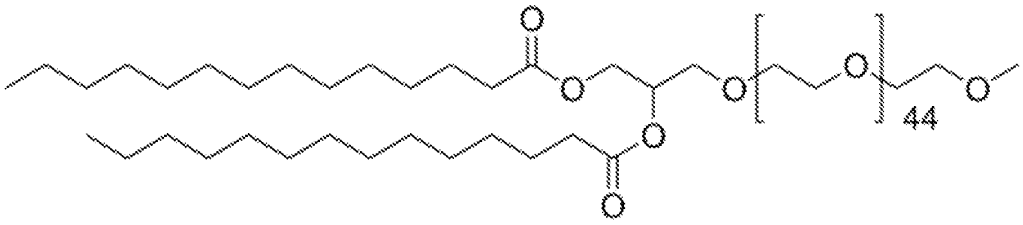
xxii),



(xxiii),



(xxiv), and



(xxv).

In some embodiments, lipids conjugated with a molecule other than a PEG can also be used in place of PEG-lipid. For example, polyoxazoline (POZ)-lipid conjugates, polyamide-lipid conjugates (such as ATTA-lipid conjugates), and cationic-polymer lipid (GPL) conjugates can be used in place of or in addition to the PEG-lipid.

Exemplary conjugated lipids, i.e., PEG-lipids, (POZ)-lipid conjugates, ATTA-lipid conjugates and cationic polymer-lipids are described in the PCT and LIS patent applications listed in Table 2 of WO2019051289A9 and in WO2020106946A1, the contents of all of which are incorporated herein by reference in their entirety.

In some embodiments an LNP comprises a compound of Formula (xix), a compound of Formula (xxi) and a compound of Formula (xxv). In some embodiments a LNP comprising a

formulation of Formula (xix), Formula (xxi) and Formula (xxv)is used to deliver a GeneWriter composition described herein to the lung or pulmonary cells.

In some embodiments, the PEG or the conjugated lipid can comprise 0-20% (mol) of the total lipid present in the lipid nanoparticle. In some embodiments, PEG or the conjugated lipid content is 0.5- 10% or 2-5% (mol) of the total lipid present in the lipid nanoparticle. Molar ratios of the ionizable lipid, non-cationic-lipid, sterol, and PEG/conjugated lipid can be varied as needed. For example, the lipid particle can comprise 30-70% ionizable lipid by mole or by total weight of the composition, 0-60% cholesterol by mole or by total weight of the composition, 0-30% non-cationic-lipid by mole or by total weight of the composition and 1-10% conjugated lipid by mole or by total weight of the composition. Preferably, the composition comprises 30-40% ionizable lipid by mole or by total weight of the composition, 40-50% cholesterol by mole or by total weight of the composition, and 10- 20% non-cationic-lipid by mole or by total weight of the composition. In some other embodiments, the composition is 50-75% ionizable lipid by mole or by total weight of the composition, 20-40% cholesterol by mole or by total weight of the composition, and 5 to 10% non-cationic-lipid, by mole or by total weight of the composition and 1-10% conjugated lipid by mole or by total weight of the composition. The composition may contain 60-70% ionizable lipid by mole or by total weight of the composition, 25-35% cholesterol by mole or by total weight of the composition, and 5-10% non-cationic-lipid by mole or by total weight of the composition. The composition may also contain up to 90% ionizable lipid by mole or by total weight of the composition and 2 to 15% non-cationic lipid by mole or by total weight of the composition. The formulation may also be a lipid nanoparticle formulation, for example comprising 8-30% ionizable lipid by mole or by total weight of the composition, 5-30% non- cationic lipid by mole or by total weight of the composition, and 0-20% cholesterol by mole or by total weight of the composition; 4-25% ionizable lipid by mole or by total weight of the composition, 4-25% non-cationic lipid by mole or by total weight of the composition, 2 to 25% cholesterol by mole or by total weight of the composition, 10 to 35% conjugate lipid by mole or by total weight of the composition, and 5% cholesterol by mole or by total weight of the composition; or 2-30% ionizable lipid by mole or by total weight of the composition, 2-30% non-cationic lipid by mole or by total weight of the composition, 1 to 15% cholesterol by mole or by total weight of the composition, 2 to 35% conjugate lipid by mole or by total weight of the

composition, and 1-20% cholesterol by mole or by total weight of the composition; or even up to 90% ionizable lipid by mole or by total weight of the composition and 2-10% non-cationic lipids by mole or by total weight of the composition, or even 100% cationic lipid by mole or by total weight of the composition. In some embodiments, the lipid particle formulation comprises ionizable lipid, phospholipid, cholesterol and a PEG-ylated lipid in a molar ratio of 50: 10:38.5: 1.5. In some other embodiments, the lipid particle formulation comprises ionizable lipid, cholesterol and a PEG-ylated lipid in a molar ratio of 60:38.5: 1.5.

In some embodiments, the lipid particle comprises ionizable lipid, non-cationic lipid (e.g. phospholipid), a sterol (e.g., cholesterol) and a PEG-ylated lipid, where the molar ratio of lipids ranges from 20 to 70 mole percent for the ionizable lipid, with a target of 40-60, the mole percent of non-cationic lipid ranges from 0 to 30, with a target of 0 to 15, the mole percent of sterol ranges from 20 to 70, with a target of 30 to 50, and the mole percent of PEG-ylated lipid ranges from 1 to 6, with a target of 2 to 5.

In some embodiments, the lipid particle comprises ionizable lipid / non-cationic- lipid / sterol / conjugated lipid at a molar ratio of 50: 10:38.5: 1.5.

In an aspect, the disclosure provides a lipid nanoparticle formulation comprising phospholipids, lecithin, phosphatidylcholine and phosphatidylethanolamine.

In some embodiments, one or more additional compounds can also be included. Those compounds can be administered separately or the additional compounds can be included in the lipid nanoparticles of the invention. In other words, the lipid nanoparticles can contain other compounds in addition to the nucleic acid or at least a second nucleic acid, different than the first. Without limitations, other additional compounds can be selected from the group consisting of small or large organic or inorganic molecules, monosaccharides, disaccharides, trisaccharides, oligosaccharides, polysaccharides, peptides, proteins, peptide analogs and derivatives thereof, peptidomimetics, nucleic acids, nucleic acid analogs and derivatives, an extract made from biological materials, or any combinations thereof.

In some embodiments, a lipid nanoparticle (or a formulation comprising lipid nanoparticles) lacks reactive impurities (e.g., aldehydes or ketones), or comprises less than a preselected level of reactive impurities (e.g., aldehydes or ketones). While not wishing to be bound by theory, in some embodiments, a lipid reagent is used to make a lipid nanoparticle formulation, and the lipid reagent may comprise a contaminating reactive impurity (e.g., an

aldehyde or ketone). A lipid regent may be selected for manufacturing based on having less than a preselected level of reactive impurities (e.g., aldehydes or ketones). Without wishing to be bound by theory, in some embodiments, aldehydes can cause modification and damage of RNA, e.g., cross-linking between bases and/or covalently conjugating lipid to RNA (e.g., forming lipid-RNA adducts). This may, in some instances, lead to failure of a reverse transcriptase reaction and/or incorporation of inappropriate bases, e.g., at the site(s) of lesion(s), e.g., a mutation in a newly synthesized target DNA.

In some embodiments, a lipid nanoparticle formulation is produced using a lipid reagent comprising less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content. In some embodiments, a lipid nanoparticle formulation is produced using a lipid reagent comprising less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, a lipid nanoparticle formulation is produced using a lipid reagent comprising: (i) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content; and (ii) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, the lipid nanoparticle formulation is produced using a plurality of lipid reagents, and each lipid reagent of the plurality independently meets one or more criterion described in this paragraph. In some embodiments, each lipid reagent of the plurality meets the same criterion, e.g., a criterion of this paragraph.

In some embodiments, the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content. In some embodiments, the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, the lipid nanoparticle formulation comprises: (i) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content; and (ii) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

In some embodiments, one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%,

1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content. In some embodiments, one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise: (i) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content; and (ii) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

In some embodiments, total aldehyde content and/or quantity of any single reactive impurity (e.g., aldehyde) species is determined by liquid chromatography (LC), e.g., coupled with tandem mass spectrometry (MS/MS), e.g., according to the method described in Example 34. In some embodiments, reactive impurity (e.g., aldehyde) content and/or quantity of reactive impurity (e.g., aldehyde) species is determined by detecting one or more chemical modifications of a nucleic acid molecule (e.g., an RNA molecule, e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents. In some embodiments, reactive impurity (e.g., aldehyde) content and/or quantity of reactive impurity (e.g., aldehyde) species is determined by detecting one or more chemical modifications of a nucleotide or nucleoside (e.g., a ribonucleotide or ribonucleoside, e.g., comprised in or isolated from a template nucleic acid, e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents, e.g., as described in Example 7. In embodiments, chemical modifications of a nucleic acid molecule, nucleotide, or nucleoside are detected by determining the presence of one or more modified nucleotides or nucleosides, e.g., using LC-MS/MS analysis, e.g., as described in Example 7.

In some embodiments, a nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a GeneWriter) does not comprise an aldehyde modification, or comprises less than a preselected amount of aldehyde modifications. In some embodiments, on average, a nucleic acid has less than 50, 20, 10, 5, 2, or 1 aldehyde modifications per 1000 nucleotides, e.g., wherein a single cross-linking of two nucleotides is a

single aldehyde modification. In some embodiments, the aldehyde modification is an RNA adduct (e.g., a lipid-RNA adduct). In some embodiments, the aldehyde-modified nucleotide is cross-linking between bases. In some embodiments, a nucleic acid (e.g., RNA) described herein comprises less than 50, 20, 10, 5, 2, or 1 cross-links between nucleotide.

In some embodiments, LNPs are directed to specific tissues by the addition of targeting domains. For example, biological ligands may be displayed on the surface of LNPs to enhance interaction with cells displaying cognate receptors, thus driving association with and cargo delivery to tissues wherein cells express the receptor. In some embodiments, the biological ligand may be a ligand that drives delivery to the liver, e.g., LNPs that display GalNAc result in delivery of nucleic acid cargo to hepatocytes that display asialoglycoprotein receptor (ASGPR). The work of Akinc et al. *Mol Ther* 18(7):1357-1364 (2010) teaches the conjugation of a trivalent GalNAc ligand to a PEG-lipid (GalNAc-PEG-DSG) to yield LNPs dependent on ASGPR for observable LNP cargo effect (see, e.g., Figure 6). Other ligand-displaying LNP formulations, e.g., incorporating folate, transferrin, or antibodies, are discussed in WO2017223135, which is incorporated herein by reference in its entirety, in addition to the references used therein, namely Kolhatkar et al., Curr Drug Discov Technol. 2011 8:197-206; Musacchio and Torchilin, Front Biosci. 2011 16:1388-1412; Yu et al., Mol Membr Biol. 2010 27:286-298; Patil et al., Crit Rev Ther Drug Carrier Syst. 2008 25:1-61 ; Benoit et al., Biomacromolecules. 2011 12:2708-2714; Zhao et al., Expert Opin Drug Deliv. 2008 5:309-319; Akinc et al., Mol Ther. 2010 18:1357-1364; Srinivasan et al., Methods Mol Biol. 2012 820:105-116; Ben-Arie et al., Methods Mol Biol. 2012 757:497-507; Peer 2010 J Control Release. 20:63-68; Peer et al., Proc Natl Acad Sci U S A. 2007 104:4095-4100; Kim et al., Methods Mol Biol. 2011 721:339-353; Subramanya et al., Mol Ther. 2010 18:2028-2037; Song et al., Nat Biotechnol. 2005 23:709-717; Peer et al., Science. 2008 319:627-630; and Peer and Lieberman, Gene Ther. 2011 18:1127-1133.

In some embodiments, LNPs are selected for tissue-specific activity by the addition of a Selective ORgan Targeting (SORT) molecule to a formulation comprising traditional components, such as ionizable cationic lipids, amphipathic phospholipids, cholesterol and poly(ethylene glycol) (PEG) lipids. The teachings of Cheng et al. Nat Nanotechnol 15(4):313-320 (2020) demonstrate that the addition of a supplemental "SORT" component precisely alters the in vivo RNA delivery profile and mediates tissue-specific (e.g., lungs, liver, spleen) gene

delivery and editing as a function of the percentage and biophysical property of the SORT molecule.

In some embodiments, the LNPs comprise biodegradable, ionizable lipids. In some embodiments, the LNPs comprise (9Z,l2Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9,l2-dienoate, also called 3- ((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl (9Z,l2Z)-octadeca-9,l2-dienoate) or another ionizable lipid. See, e.g, lipids of WO2019/067992, WO/2017/173054, WO2015/095340, and WO2014/136086, as well as references provided therein. In some embodiments, the term cationic and ionizable in the context of LNP lipids is interchangeable, e.g., wherein ionizable lipids are cationic depending on the pH.

In some embodiments, multiple components of a Gene Writer system may be prepared as a single LNP formulation, e.g., an LNP formulation comprises mRNA encoding for the Gene Writer polypeptide and an RNA template. Ratios of nucleic acid components may be varied in order to maximize the properties of a therapeutic. In some embodiments, the ratio of RNA template to mRNA encoding a Gene Writer polypeptide is about 1:1 to 100:1, e.g., about 1:1 to 20:1, about 20:1 to 40:1, about 40:1 to 60:1, about 60:1 to 80:1, or about 80:1 to 100:1, by molar ratio. In other embodiments, a system of multiple nucleic acids may be prepared by separate formulations, e.g., one LNP formulation comprising a template RNA and a second LNP formulation comprising an mRNA encoding a Gene Writer polypeptide. In some embodiments, the system may comprise more than two nucleic acid components formulated into LNPs. In some embodiments, the system may comprise a protein, e.g., a Gene Writer polypeptide, and a template RNA formulated into at least one LNP formulation.

In some embodiments, the average LNP diameter of the LNP formulation may be between 10s of nm and 100s of nm, e.g., measured by dynamic light scattering (DLS). In some embodiments, the average LNP diameter of the LNP formulation may be from about 40 nm to about 150 nm, such as about 40 nm, 45 nm, 50 nm, 55 nm, 60 nm, 65 nm, 70 nm, 75 nm, 80 nm, 85 nm, 90 nm, 95 nm, 100 nm, 105 nm, 110 nm, 115 nm, 120 nm, 125 nm, 130 nm, 135 nm, 140 nm, 145 nm, or 150 nm. In some embodiments, the average LNP diameter of the LNP formulation may be from about 50 nm to about 100 nm, from about 50 nm to about 90 nm, from about 50 nm to about 80 nm, from about 50 nm to about 70 nm, from about 50 nm to about 60 nm, from about 60 nm to about 100 nm, from about 60 nm to about 90 nm, from about 60 nm to

about 80 nm, from about 60 nm to about 70 nm, from about 70 nm to about 100 nm, from about 70 nm to about 90 nm, from about 70 nm to about 80 nm, from about 80 nm to about 100 nm, from about 80 nm to about 90 nm, or from about 90 nm to about 100 nm. In some embodiments, the average LNP diameter of the LNP formulation may be from about 70 nm to about 100 nm. In a particular embodiment, the average LNP diameter of the LNP formulation may be about 80 nm. In some embodiments, the average LNP diameter of the LNP formulation may be about 100 nm. In some embodiments, the average LNP diameter of the LNP formulation ranges from about lmm to about 500 mm, from about 5 mm to about 200 mm, from about 10 mm to about 100 mm, from about 20 mm to about 80 mm, from about 25 mm to about 60 mm, from about 30 mm to about 55 mm, from about 35 mm to about 50 mm, or from about 38 mm to about 42 mm.

A LNP may, in some instances, be relatively homogenous. A polydispersity index may be used to indicate the homogeneity of a LNP, e.g., the particle size distribution of the lipid nanoparticles. A small (e.g., less than 0.3) polydispersity index generally indicates a narrow particle size distribution. A LNP may have a polydispersity index from about 0 to about 0.25, such as 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, or 0.25. In some embodiments, the polydispersity index of a LNP may be from about 0.10 to about 0.20.

The zeta potential of a LNP may be used to indicate the electrokinetic potential of the composition. In some embodiments, the zeta potential may describe the surface charge of a LNP. Lipid nanoparticles with relatively low charges, positive or negative, are generally desirable, as more highly charged species may interact undesirably with cells, tissues, and other elements in the body. In some embodiments, the zeta potential of a LNP may be from about -10 mV to about +20 mV, from about -10 mV to about +15 mV, from about -10 mV to about +10 mV, from about -10 mV to about +5 mV, from about -10 mV to about 0 mV, from about -10 mV to about -5 mV, from about -5 mV to about +20 mV, from about -5 mV to about +15 mV, from about -5 mV to about +10 mV, from about -5 mV to about +5 mV, from about -5 mV to about 0 mV, from about 0 mV to about +20 mV, from about 0 mV to about +15 mV, from about 0 mV to about +10 mV, from about 0 mV to about +5 mV, from about +5 mV to about +20 mV, from about +5 mV to about +15 mV, or from about +5 mV to about +10 mV.

The efficiency of encapsulation of a protein and/or nucleic acid, e.g., Gene Writer polypeptide or mRNA encoding the polypeptide, describes the amount of protein and/or nucleic

acid that is encapsulated or otherwise associated with a LNP after preparation, relative to the initial amount provided. The encapsulation efficiency is desirably high (e.g., close to 100%). The encapsulation efficiency may be measured, for example, by comparing the amount of protein or nucleic acid in a solution containing the lipid nanoparticle before and after breaking up the lipid nanoparticle with one or more organic solvents or detergents. An anion exchange resin may be used to measure the amount of free protein or nucleic acid (e.g., RNA) in a solution. Fluorescence may be used to measure the amount of free protein and/or nucleic acid (e.g., RNA) in a solution. For the lipid nanoparticles described herein, the encapsulation efficiency of a protein and/or nucleic acid may be at least 50%, for example 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%. In some embodiments, the encapsulation efficiency may be at least 80%. In some embodiments, the encapsulation efficiency may be at least 90%. In some embodiments, the encapsulation efficiency may be at least 95%.

A LNP may optionally comprise one or more coatings. In some embodiments, a LNP may be formulated in a capsule, film, or table having a coating. A capsule, film, or tablet including a composition described herein may have any useful size, tensile strength, hardness or density.

Additional exemplary lipids, formulations, methods, and characterization of LNPs are taught by WO2020061457, which is incorporated herein by reference in its entirety.

In some embodiments, in vitro or ex vivo cell lipofections are performed using Lipofectamine MessengerMax (Thermo Fisher) or TransIT-mRNA Transfection Reagent (Mirus Bio). In certain embodiments, LNPs are formulated using the GenVoy_ILM ionizable lipid mix (Precision NanoSystems). In certain embodiments, LNPs are formulated using 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) or dilinoleylmethyl-4-dimethylaminobutyrate (DLin-MC3-DMA or MC3), the formulation and in vivo use of which are taught in Jayaraman et al. Angew Chem Int Ed Engl 51(34):8529-8533 (2012), incorporated herein by reference in its entirety.

LNP formulations optimized for the delivery of CRISPR-Cas systems, e.g., Cas9-gRNA RNP, gRNA, Cas9 mRNA, are described in WO2019067992 and WO2019067910, both incorporated by reference.

Additional specific LNP formulations useful for delivery of nucleic acids are described in US8158601 and US8168775, both incorporated by reference, which include formulations used in patisiran, sold under the name ONPATTRO.

Exemplary dosing of Gene Writer LNP may include about 0.1, 0.25, 0.3, 0.5, 1, 2, 3, 4, 5, 6, 8, 10, or 100 mg/kg (RNA). Exemplary dosing of AAV comprising a nucleic acid encoding one or more components of the system may include an MOI of about 10^{11}, 10^{12}, 10^{13}, and 10^{14} vg/kg.

Suitable Indications

Exemplary suitable diseases and disorders that can be treated by the systems or methods provided herein, for example, those comprising Gene Writers, include, without limitation: Baraitser-Winter syndromes 1 and 2; Diabetes mellitus and insipidus with optic atrophy and deafness; Alpha- 1- antitrypsin deficiency; Heparin cofactor II deficiency; Adrenoleukodystrophy; Keppen-Lubinsky syndrome; Treacher collins syndrome 1; Mitochondrial complex I, II, III, III (nuclear type 2, 4, or 8) deficiency; Hypermanganesemia with dystonia, polycythemia and cirrhosis; Carcinoid tumor of intestine; Rhabdoid tumor predisposition syndrome 2; Wilson disease; Hyperphenylalaninemia, bh4-deficient, a, due to partial pts deficiency, BH4-deficient, D, and non-pku; Hyperinsulinemic hypoglycemia familial 3, 4, and 5; Keratosis follicularis; Oral-facial-digital syndrome; SeSAME syndrome; Deafness, nonsyndromic sensorineural, mitochondrial; Proteinuria; Insulin-dependent diabetes mellitus secretory diarrhea syndrome; Moyamoya disease 5; Diamond-Blackfan anemia 1, 5, 8, and 10; Pseudoachondroplastic spondyloepiphyseal dysplasia syndrome; Brittle cornea syndrome 2; Methylmalonic acidemia with homocystinuria, ; Adams-Oliver syndrome 5 and 6; autosomal recessive Agammaglobulinemia 2; Cortical malformations, occipital; Febrile seizures, familial, 11; Mucopolysaccharidosis type VI, type VI (severe), and type VII; Marden Walker like syndrome; Pseudoneonatal adrenoleukodystrophy; Spheroid body myopathy; Cleidocranial dysostosis; Multiple Cutaneous and Mucosal Venous Malformations; Liver failure acute infantile; Neonatal intrahepatic cholestasis caused by citrin deficiency; Ventricular septal defect 1; Oculodentodigital dysplasia; Wilms tumor 1; Weill- Marchesani-like syndrome; Renal adysplasia; Cataract 1, 4, autosomal dominant, autosomal dominant, multiple types, with microcornea, coppock-like, juvenile, with microcornea and glucosuria, and nuclear diffuse

nonprogressive; Odontohypophosphatasia; Cerebro-oculo-facio- skeletal syndrome; Schizophrenia 15; Cerebral amyloid angiopathy, APP-related; Hemophagocytic lymphohistiocytosis, familial, 3; Porphobilinogen synthase deficiency; Episodic ataxia type 2; Trichorhinophalangeal syndrome type 3; Progressive familial heart block type IB; Glioma susceptibility 1; Lichtenstein-Knorr Syndrome; Hypohidrotic X-linked ectodermal dysplasia; Bartter syndrome types 3, 3 with hypocalciuria , and 4; Carbonic anhydrase VA deficiency, hyperammonemia due to; Cardiomyopathy; Poikiloderma, hereditary fibrosing, with tendon contractures, myopathy, and pulmonary fibrosis; Combined d-2- and l-2- hydroxyglutaric aciduria; Arginase deficiency; Cone-rod dystrophy 2 and 6; Smith-Lemli-Opitz syndrome; Mucolipidosis III Gamma; Blau syndrome; Werner syndrome; Meningioma; Iodotyrosyl coupling defect; Dubin-Johnson syndrome; 3-Oxo-5 alpha-steroid delta 4-dehydrogenase deficiency; Boucher Neuhauser syndrome; Iron accumulation in brain; Mental Retardation, X-Linked 102 and syndromic 13; familial, Pituitary adenoma predisposition; Hypoplasia of the corpus callosum; Hyperalphalipoproteinemia 2; Deficiency of ferroxidase; Growth hormone insensitivity with immunodeficiency; Marinesco-Sj\xc3\xb6gren syndrome; Martsolf syndrome; Gaze palsy, familial horizontal, with progressive scoliosis; Mitchell-Riley syndrome; Hypocalciuric hypercalcemia, familial, types 1 and 3; Rubinstein-Taybi syndrome; Epstein syndrome; Juvenile retinoschisis; Becker muscular dystrophy; Loeys-Dietz syndrome 1, 2, 3; Congenital muscular hypertrophy-cerebral syndrome; Familial juvenile gout; Spermatogenic failure 11, 3, and 8; Orofacial cleft 11 and 7, Cleft lip/palate- ectodermal dysplasia syndrome; Mental retardation, X-linked, nonspecific, syndromic, Hedera type, and syndromic, wu type; Combined oxidative phosphorylation deficiencies 1, 3, 4, 12, 15, and 25; Frontotemporal dementia; Kniest dysplasia; Familial cardiomyopathy; Benign familial hematuria; Pheochromocytoma; Aminoglycoside-induced deafness; Gamma-aminobutyric acid transaminase deficiency; Oculocutaneous albinism type IB, type 3, and type 4; Renal coloboma syndrome; CNS hypomyelination; Hennekam lymphangiectasia-lymphedema syndrome 2; Migraine, familial basilar; Distal spinal muscular atrophy, X-linked 3; X-linked periventricular heterotopia; Microcephaly; Mucopolysaccharidosis, MPS-I-H/S, MPS-II, MPS-III-A, MPS-III-B, MPS-III-C, MPS-IV-A, MPS-IV-B; Infantile Parkinsonism-dystonia; Frontotemporal dementia with TDP43 inclusions, TARDBP-related; Hereditary diffuse gastric cancer; Sialidosis type I and II; Microcephaly-capillary malformation syndrome; Hereditary breast and ovarian

cancer syndrome; Brain small vessel disease with hemorrhage; Non-ketotic hyperglycinemia; Navajo neurohepatopathy; Auriculocondylar syndrome 2; Spastic paraplegia 15, 2, 3, 35, 39, 4, autosomal dominant, 55, autosomal recessive, and 5A; Autosomal recessive cutis laxa type IA and IB; Hemolytic anemia, nonspherocytic, due to glucose phosphate isomerase deficiency; Hutchinson-Gilford syndrome; Familial amyloid nephropathy with urticaria and deafness; Supravalvar aortic stenosis; Diffuse palmoplantar keratoderma, Bothnian type; Holt-Oram syndrome; Coffin Siris/Intellectual Disability; Left-right axis malformations; Rapadilino syndrome; Nanophthalmos 2; Craniosynostosis and dental anomalies; Paragangliomas 1; Snyder Robinson syndrome; Ventricular fibrillation; Activated PI3K-delta syndrome; Howel-Evans syndrome; Larsen syndrome, dominant type; Van Maldergem syndrome 2; MYH-associated polyposis; 6-pymvoyl-tetrahydropterin synthase deficiency; Alagille syndromes 1 and 2; Lymphangiomyomatosis; Muscle eye brain disease; WFS1-Related Disorders; Primary hypertrophic osteoarthropathy, autosomal recessive 2; Infertility; Nestor- Guillermo progeria syndrome; Mitochondrial trifunctional protein deficiency; Hypoplastic left heart syndrome 2; Primary dilated cardiomyopathy; Retinitis pigmentosa; Hirschsprung disease 3; Upshaw-Schulman syndrome; Desbuquois dysplasia 2; Diarrhea 3 (secretory sodium, congenital, syndromic) and 5 (with tufting enteropathy, congenital); Pachyonychia congenita 4 and type 2; Cerebral autosomal dominant and recessive arteriopathy with subcortical infarcts and leukoencephalopathy; Vi tel 1i form dystrophy ; type II, type IV, IV (combined hepatic and myopathic), type V, and type VI; Atypical Rett syndrome; Atrioventricular septal defect 4; Papillon-Lef\xc3\xa8vre syndrome; Leber amaurosis; X-linked hereditary motor and sensory neuropathy; Progressive sclerosing poliodystrophy; Goldmann-Favre syndrome; Renal-hepatic-pancreatic dysplasia; Pallister-Hall syndrome; Amyloidogenic transthyretin amyloidosis; Melnick-Needles syndrome; Hyperimmunoglobulin E syndrome; Posterior column ataxia with retinitis pigmentosa; Chondrodysplasia punctata 1, X-linked recessive and 2 X-linked dominant; Ectopia lentis, isolated autosomal recessive and dominant; Familial cold urticarial; Familial adenomatous polyposis 1 and 3; Porokeratosis 8, disseminated superficial actinic type; PIK3CA Related Overgrowth Spectrum; Cerebral cavernous malformations 2; Exudative vitreoretinopathy 6; Megalencephaly cutis marmorata telangiectatica congenital; TARP syndrome; Diabetes mellitus, permanent neonatal, with neurologic features; Short-rib thoracic dysplasia 11 or 3 with or without polydactyly; Hypertrichotic osteochondrodysplasia; beta

Thalassemia; Niemann-Pick disease type C1, C2, type A, and type C1, adult form; Charcot-Marie-Tooth disease types 1B, 2B2, 2C, 2F, 21, 2U (axonal), 1C (demyelinating), dominant intermediate C, recessive intermediate A, 2A2, 4C, 4D, 4H, 1F, IVF, and X; Tyrosinemia type I; Paroxysmal atrial fibrillation; UV- sensitive syndrome; Tooth agenesis, selective, 3 and 4; Merosin deficient congenital muscular dystrophy; Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; Congenital aniridia; Left ventricular noncompaction 5; Deficiency of aromatic-L-amino-acid decarboxylase; Coronary heart disease; Leukonychia totalis; Distal arthrogryposis type 2B; Retinitis pigmentosa 10, 11, 12, 14, 15, 17, and 19; Robinow Sorauf syndrome; Tenorio Syndrome; Prolactinoma; Neurofibromatosis, type 1and type 2; Congenital muscular dystrophy-dystroglycanopathy with brain and eye anomalies, types A2, A7, A8, A11, and A14; Heterotaxy, visceral, 2, 4, and 6, autosomal; Jankovic Rivera syndrome; Lipodystrophy, familial partial, type 2 and 3; Hemoglobin H disease, nondeletional; Multicentric osteolysis, nodulosis and arthropathy; Thyroid agenesis; deficiency of Acyl-CoA dehydrogenase family, member 9; Alexander disease; Phytanic acid storage disease; Breast-ovarian cancer, familial 1, 2, and 4; Proline dehydrogenase deficiency; Childhood hypophosphatasia; Pancreatic agenesis and congenital heart disease; Vitamin D-dependent rickets, types 1and 2; Iridogoniodysgenesis dominant type and type 1; Autosomal recessive hypohidrotic ectodermal dysplasia syndrome; Mental retardation, X-linked, 3, 21, 30, and 72; Hereditary hemorrhagic telangiectasia type 2; Blepharophimosis, ptosis, and epicanthus inversus; Adenine phosphoribosyltransferase deficiency; Seizures, benign familial infantile, 2; Acrodysostosis 2, with or without hormone resistance; Tetralogy of Fallot; Retinitis pigmentosa 2, 20, 25, 35, 36, 38, 39, 4, 40, 43, 45, 48, 66, 7, 70, 72; Lysosomal acid lipase deficiency; Eichsfeld type congenital muscular dystrophy; Walker-Warburg congenital muscular dystrophy; TNF receptor-associated periodic fever syndrome (TRAPS); Progressive myoclonus epilepsy with ataxia; Epilepsy, childhood absence 2, 12 (idiopathic generalized, susceptibility to) 5 (nocturnal frontal lobe), nocturnal frontal lobe type 1, partial, with variable foci, progressive myoclonic 3, and X-linked, with variable learning disabilities and behavior disorders; Long QT syndrome; Dicarboxylic aminoaciduria; Brachydactyly types A1 and A2; Pseudoxanthoma elasticum-like disorder with multiple coagulation factor deficiency; Multisystemic smooth muscle dysfunction syndrome; Syndactyly Cenani Lenz type; Joubert syndrome 1, 6, 7, 9/15 (digenic), 14, 16, and 17, and Orofaciodigital syndrome xiv; Digitorenocerebral syndrome; Retinoblastoma;

Dyskinesia, familial, with facial myokymia; Hereditary sensory and autonomic neuropathy type IIB amd IIA; familial hyperinsulinism; Megalencephalic leukoencephalopathy with subcortical cysts land 2a; Aase syndrome; Wiedemann- Steiner syndrome; Ichthyosis exfoliativa; Myotonia congenital; Granulomatous disease, chronic, X-linked, variant; Deficiency of 2-methylbutyryl-CoA dehydrogenase; Sarcoidosis, early-onset; Glaucoma, congenital and Glaucoma, congenital, Coloboma; Breast cancer, susceptibility to; Ceroid lipofuscinosis neuronal 2, 6, 7, and 10; Congenital generalized lipodystrophy type 2; Fructose-biphosphatase deficiency; Congenital contractural arachnodactyly; Lynch syndrome I and II; Phosphoglycerate dehydrogenase deficiency; Burn- Mckeown syndrome; Myocardial infarction 1; Achromatopsia 2 and 7; Retinitis Pigmentosa 73; Protan defect; Polymicrogyria, asymmetric, bilateral frontoparietal; Spinal muscular atrophy, distal, autosomal recessive, 5; Methylmalonic aciduria due to methylmalonyl-CoA mutase deficiency; Familial porencephaly; Hurler syndrome; Oto-palato-digital syndrome, types I and II; Sotos syndrome 1 or 2; Cardioencephalomyopathy, fatal infantile, due to cytochrome c oxidase deficiency; Parastremmatic dwarfism; Thyrotropin releasing hormone resistance, generalized; Diabetes mellitus, type 2, and insulin-dependent, 20; Thoracic aortic aneurysms and aortic dissections; Estrogen resistance; Maple syrup urine disease type 1A and type 3; Hypospadias 1 and 2, X-linked; Metachromatic leukodystrophy juvenile, late infantile, and adult types; Early T cell progenitor acute lymphoblastic leukemia; Neuropathy, Hereditary Sensory, Type IC; Mental retardation, autosomal dominant 31; Retinitis pigmentosa 39; Breast cancer, early-onset; May-Hegglin anomaly; Gaucher disease type 1 and Subacute neuronopathic; Temtamy syndrome; Spinal muscular atrophy, lower extremity predominant 2, autosomal dominant; Fanconi anemia, complementation group E, I, N, and O; Alkaptonuria; Hirschsprung disease; Combined malonic and methylmalonic aciduria; Arrhythmogenic right ventricular cardiomyopathy types 5, 8, and 10; Congenital lipomatous overgrowth, vascular malformations, and epidermal nevi; Timothy syndrome; Deficiency of guanidinoacetate methyltransferase; Myoclonic dystonia; Kanzaki disease; Neutral 1 amino acid transport defect; Neurohypophyseal diabetes insipidus; Thyroid hormone metabolism, abnormal; Benign scapuloperoneal muscular dystrophy with cardiomyopathy; Hypoglycemia with deficiency of glycogen synthetase in the liver; Hypertrophic cardiomyopathy; Myasthenic Syndrome, Congenital, 11, associated with acetylcholine receptor deficiency; Mental retardation X-linked syndromic 5; Stormorken syndrome; Aplastic anemia; Intellectual disability; Normokalemic

periodic paralysis, potassium-sensitive; Danon disease; Nephronophthisis 13, 15 and 4; Thyrotoxic periodic paralysis and Thyrotoxic periodic paralysis 2; Infertility associated with multi-tailed spermatozoa and excessive DNA; Glaucoma, primary open angle, juvenile-onset; Afibrinogenemia and congenital Afibrinogenemia; Polycystic kidney disease 2, adult type, and infantile type; Familial porphyria cutanea tarda; Cerebello-oculo-renal syndrome (nephronophthisis, oculomotor apraxia and cerebellar abnormalities); Frontotemporal Dementia Chromosome 3- Linked and Frontotemporal dementia ubiquitin-positive; Metatrophic dysplasia; Immunodeficiency-centromeric instability-facial anomalies syndrome 2; Anemia, nonspherocytic hemolytic, due to G6PD deficiency; Bronchiectasis with or without elevated sweat chloride 3; Congenital myopathy with fiber type disproportion; Carney complex, type 1; Cryptorchidism, unilateral or bilateral; Ichthyosis bullosa of Siemens; Isolated lutropin deficiency; DFNA 2 Nonsyndromic Hearing Loss; Klein-Waardenberg syndrome; Gray platelet syndrome; Bile acid synthesis defect, congenital, 2; 46, XY sex reversal, type 1, 3, and 5; Acute intermittent porphyria; Cornelia de Fange syndromes 1 and 5; Hyperglycinuria; Cone-rod dystrophy 3; Dysfibrinogenemia; Karak syndrome; Congenital muscular dystrophy-dystroglycanopathy without mental retardation, type B5; Infantile nystagmus, X-linked; Dyskeratosis congenita, autosomal recessive, 1, 3, 4, and 5; Microcephaly with or without chorioretinopathy, lymphedema, or mental retardation; Hyperlysinemia; Bardet-Biedl syndromes 1, 11, 16, and 19; Autosomal recessive centronuclear myopathy; Frasier syndrome; Caudal regression syndrome; Fibrosis of extraocular muscles, congenital, 1, 2, 3a (with or without extraocular involvement), 3b; Prader-Willi-like syndrome; Malignant melanoma; Bloom syndrome; Darier disease, segmental; Multicentric osteolysis nephropathy; Hemochromatosis type 1, 2B, and 3; Cerebellar ataxia infantile with progressive external ophthalmoplegi and Cerebellar ataxia, mental retardation, and dysequilibrium syndrome 2; Hypoplastic left heart syndrome; Epilepsy, Hearing Loss, And Mental Retardation Syndrome; Transferrin serum level quantitative trait locus 2; Ocular albinism, type I; Marfan syndrome; Congenital muscular dystrophy- dystroglycanopathy with brain and eye anomalies, type A14 and B14; Hyperammonemia, type III; Cryptophthalmos syndrome; Alopecia universalis congenital; Adult hypophosphatasia; Mannose-binding protein deficiency; Bull eye macular dystrophy; Autosomal dominant torsion dystonia 4; Nephrotic syndrome, type 3, type 5, with or without ocular abnormalities, type 7, and type 9; Seizures, Early infantile epileptic encephalopathy 7; Persistent

hyperinsulinemic hypoglycemia of infancy; Thrombocytopenia, X-linked; Neonatal hypotonia; Orstavik Lindemann Solberg syndrome; Pulmonary hypertension, primary, 1, with hereditary hemorrhagic telangiectasia; Pituitary dependent hypercortisolism; Epidermodysplasia verruciformis; Epidermolysis bullosa, junctional, localisata variant; Cytochrome c oxidase i deficiency; Kindler syndrome; Myosclerosis, autosomal recessive; Truncus arteriosus; Duane syndrome type 2; ADULT syndrome; Zellweger syndrome spectrum; Leukoencephalopathy with ataxia, with Brainstem and Spinal Cord Involvement and Lactate Elevation, with vanishing white matter, and progressive, with ovarian failure; Antithrombin III deficiency; Holoprosencephaly 7; Roberts-SC phocomelia syndrome; Mitochondrial DNA-depletion syndrome 3 and 7, hepatocerebral types, and 13 (encephalomyopathic type); Porencephaly 2; Microcephaly, normal intelligence and immunodeficiency; Giant axonal neuropathy; Sturge-Weber syndrome, Capillary malformations, congenital, 1; Fabry disease and Fabry disease, cardiac variant; Glutamate formiminotransferase deficiency; Fanconi-Bickel syndrome; Acromicric dysplasia; Epilepsy, idiopathic generalized, susceptibility to, 12; Basal ganglia calcification, idiopathic, 4; Polyglucosan body myopathy 1 with or without immunodeficiency; Malignant tumor of prostate; Congenital ectodermal dysplasia of face; Congenital heart disease; Age-related macular degeneration 3, 6, 11, and 12; Congenital myotonia, autosomal dominant and recessive forms; Hypomagnesemia 1, intestinal; Sulfite oxidase deficiency, isolated; Pick disease; Plasminogen deficiency, type I; Syndactyly type 3; Cone-rod dystrophy amelogenesis imperfecta; Pseudoprimary hyperaldosteronism; Terminal osseous dysplasia; Bartter syndrome antenatal type 2; Congenital muscular dystrophy- dystroglycanopathy with mental retardation, types B2, B3, B5, and B15; Familial infantile myasthenia; Lymphoproliferative syndrome 1, 1 (X-linked), and 2; Hypercholesterolaemia and Hypercholesterolemia, autosomal recessive; Neoplasm of ovary; Infantile GM1 gangliosidosis; Syndromic X-linked mental retardation 16; Deficiency of ribose-5-phosphate isomerase; Alzheimer disease, types, 1, 3, and 4; Andersen Tawil syndrome; Multiple synostoses syndrome 3; Chilbain lupus 1; Hemophagocytic lymphohistiocytosis, familial, 2; Axenfeld-Rieger syndrome type 3; Myopathy, congenital with cores; Osteoarthritis with mild chondrodysplasia; Peroxisome biogenesis disorders; Severe congenital neutropenia; Hereditary neuralgic amyotrophy; Palmoplantar keratoderma, nonepidermolytic, focal or diffuse; Dysplasminogenemia; Familial colorectal cancer; Spastic ataxia 5, autosomal recessive, Charlevoix-Saguenay type, 1,10, or 11, autosomal recessive; Frontometaphyseal dysplasia land

3; Hereditary factors II, IX, VIII deficiency disease; Spondylocheirodysplasia, Ehlers-Danlos syndrome-like, with immune dysregulation, Aggrecan type, with congenital joint dislocations, short limb-hand type, Sedaghatian type, with cone-rod dystrophy, and Kozlowski type; Ichthyosis prematurity syndrome; Stickler syndrome type 1; Focal segmental glomerulosclerosis 5; 5-Oxoprolinase deficiency; Microphthalmia syndromic 5, 7, and 9; Juvenile polyposis/hereditary hemorrhagic telangiectasia syndrome; Deficiency of butyryl-CoA dehydrogenase; Maturity-onset diabetes of the young, type 2; Mental retardation, syndromic, Claes-Jensen type, X-linked; Deafness, cochlear, with myopia and intellectual impairment, without vestibular involvement, autosomal dominant, X-linked 2; Spondylocarpotarsal synostosis syndrome; Sting-associated vasculopathy, infantile-onset; Neutral lipid storage disease with myopathy; Immune dysfunction with T-cell inactivation due to calcium entry defect 2; Cardiofaciocutaneous syndrome; Corticosterone methyloxidase type 2 deficiency; Hereditary myopathy with early respiratory failure; Interstitial nephritis, karyomegalic; Trimethylaminuria; Hyperimmunoglobulin D with periodic fever; Malignant hyperthermia susceptibility type 1; Trichomegaly with mental retardation, dwarfism and pigmentary degeneration of retina; Breast adenocarcinoma; Complement factor B deficiency; Ullrich congenital muscular dystrophy; Left ventricular noncompaction cardiomyopathy; Fish-eye disease; Finnish congenital nephrotic syndrome; Limb-girdle muscular dystrophy, type IB, 2A, 2B, 2D, Cl, C5, C9, C14; Idiopathic fibrosing alveolitis, chronic form; Primary familial hypertrophic cardiomyopathy; Angiotensin i-converting enzyme, benign serum increase; Cd8 deficiency, familial; Proteus syndrome; Glucose-6-phosphate transport defect; Borjeson-Forssman-Lehmann syndrome; Zellweger syndrome; Spinal muscular atrophy, type II; Prostate cancer, hereditary, 2; Thrombocytopenia, platelet dysfunction, hemolysis, and imbalanced globin synthesis; Congenital disorder of glycosylation types IB, ID, 1G, 1H, 1 J, IK, IN, IP, 2C, 2J, 2K, Ilm; Junctional epidermolysis bullosa gravis of Herlitz; Generalized epilepsy with febrile seizures plus 3, type 1, type 2; Schizophrenia 4; Coronary artery disease, autosomal dominant 2; Dyskeratosis congenita, autosomal dominant, 2 and 5; Subcortical laminar heterotopia, X-linked; Adenylate kinase deficiency; X- linked severe combined immunodeficiency; Coproporphyria; Amyloid Cardiomyopathy, Transthyretin-related; Hypocalcemia, autosomal dominant 1; Brugada syndrome; Congenital myasthenic syndrome, acetazolamide- responsive; Primary hypomagnesemia; Sclerosteosis; Frontotemporal dementia and/or amyotrophic lateral sclerosis 3

and 4; Mevalonic aciduria; Schwannomatosis 2; Hereditary motor and sensory neuropathy with optic atrophy; Porphyria cutanea tarda; Osteochondritis dissecans; Seizures, benign familial neonatal, 1, and/or myokymia; Long QT syndrome, LQT1 subtype; Mental retardation, anterior maxillary protrusion, and strabismus; Idiopathic hypercalcemia of infancy; Hypogonadotropic hypogonadism 11 with or without anosmia; Polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy; Primary autosomal recessive microcephaly 10, 2, 3, and 5; Interrupted aortic arch; Congenital amegakaryocytic thrombocytopenia; Hermansky-Pudlak syndrome 1, 3, 4, and 6; Long QT syndrome 1, 2, 2/9, 2/5, (digenic), 3, 5 and 5, acquired, susceptibility to; Andermann syndrome; Retinal cone dystrophy 3B; Erythropoietic protoporphyria; Sepiapterin reductase deficiency; Very long chain acyl-CoA dehydrogenase deficiency; Hyperferritinemia cataract syndrome; Silver spastic paraplegia syndrome; Charcot-Marie-Tooth disease; Atrial septal defect 2; Carnevale syndrome; Hereditary insensitivity to pain with anhidrosis; Catecholaminergic polymorphic ventricular tachycardia; Hypokalemic periodic paralysis 1 and 2; Sudden infant death syndrome; Hypochromic microcytic anemia with iron overload; GLUT1 deficiency syndrome 2; Leukodystrophy, Hypomyelinating, 11 and 6; Cone monochromatism; Osteopetrosis autosomal dominant type 1 and 2, recessive 4, recessive 1, recessive 6; Severe congenital neutropenia 3, autosomal recessive or dominant; Methionine adenosyltransferase deficiency, autosomal dominant; Paroxysmal familial ventricular fibrillation; Pyruvate kinase deficiency of red cells; Schneckenbecken dysplasia; Torsades de pointes; Distal myopathy Markesbery-Griggs type; Deficiency of UDPglucose-hexose-1- phosphate uridylyltransferase; Sudden cardiac death; Neu-Laxova syndrome 1; Atransferrinemia; Hyperparathyroidism 1 and 2; Cutaneous malignant melanoma 1; Symphalangism, proximal, Ib; Progressive pseudorheumatoid dysplasia; Werdnig-Hoffmann disease; Achondrogenesis type 2; Holoprosencephaly 2, 3,7, and 9; Schindler disease, type 1; Cerebroretinal microangiopathy with calcifications and cysts; Heterotaxy, visceral, X-linked; Tuberous sclerosis syndrome; Kartagener syndrome; Thyroid hormone resistance, generalized, autosomal dominant; Bestrophinopathy, autosomal recessive; Nail disorder, nonsyndromic congenital, 8; Mohr-Tranebjaerg syndrome; Cone-rod dystrophy 12; Hearing impairment; Ovarioleukodystrophy; Renal tubular acidosis, proximal, with ocular abnormalities and mental retardation; Dihydropteridine reductase deficiency; Focal epilepsy with speech disorder with or without mental retardation; Ataxia- telangiectasia syndrome; Brown- Vialetto- Van laere syndrome and

Brown- Vialetto-Van Laere syndrome 2; Cardiomyopathy; Peripheral demyelinating neuropathy, central dysmyelination; Corneal dystrophy, Fuchs endothelial, 4; Cowden syndrome 3; Dystonia 2 (torsion, autosomal recessive), 3 (torsion, X-linked), 5 (Dopa-responsive type), 10, 12, 16, 25, 26 (Myoclonic); Epiphyseal dysplasia, multiple, with myopia and conductive deafness; Cardiac conduction defect, nonspecific; Branchiootic syndromes 2 and 3; Peroxisome biogenesis disorder 14B, 2A, 4A, 5B, 6A, 7A, and 7B; Familial renal glucosuria; Candidiasis, familial, 2, 5, 6, and 8; Autoimmune disease, multisystem, infantile-onset; Early infantile epileptic encephalopathy 2, 4, 7, 9, 10, 11, 13, and 14; Segawa syndrome, autosomal recessive; Deafness, autosomal dominant 3a, 4, 12, 13, 15, autosomal dominant nonsyndromic sensorineural 17, 20, and 65; Congenital dyserythropoietic anemia, type I and II; Enhanced s-cone syndrome; Adult neuronal ceroid lipofuscinosis; Atrial fibrillation, familial, 11, 12, 13, and 16; Norum disease; Osteosarcoma; Partial albinism; Biotinidase deficiency; Combined cellular and humoral immune defects with granulomas; Alpers encephalopathy; Holocarboxylase synthetase deficiency; Maturity-onset diabetes of the young, type 1, type 2, type 11, type 3, and type 9; Variegate porphyria; Infantile cortical hyperostosis; Testosterone 17-beta- dehydrogenase deficiency; L-2-hydroxyglutaric aciduria; Tyrosinase-negative oculocutaneous albinism; Primary ciliary dyskinesia 24; Pontocerebellar hypoplasia type 4; Ciliary dyskinesia, primary, 7, 11, 15, 20 and 22; Idiopathic basal ganglia calcification 5; Brain atrophy; Craniosynostosis 1 and 4; Keratoconus 1; Rasopathy; Congenital adrenal hyperplasia and Congenital adrenal hypoplasia, X-linked; Mitochondrial DNA depletion syndrome 11, 12 (cardiomyopathic type), 2, 4B (MNGIE type), 8B (MNGIE type); Brachydactyly with hypertension; Cornea plana 2; Aarskog syndrome; Multiple epiphyseal dysplasia 5 or Dominant; Corneal endothelial dystrophy type 2; Aminoacylase 1 deficiency; Delayed speech and language development; Nicolaides-Baraitser syndrome; Enterokinase deficiency; Ectrodactyly, ectodermal dysplasia, and cleft lip/palate syndrome 3; Arthrogryposis multiplex congenita, distal, X-linked; Perrault syndrome 4; Jervell and Lange-Nielsen syndrome 2; Hereditary Nonpolyposis Colorectal Neoplasms; Robinow syndrome, autosomal recessive, autosomal recessive, with brachy-syn-polydactyly; Neurofibrosarcoma; Cytochrome-c oxidase deficiency ; Vesicoureteral reflux 8; Dopamine beta hydroxylase deficiency; Carbohydrate-deficient glycoprotein syndrome type I and II; Progressive familial intrahepatic cholestasis 3; Benign familial neonatal-infantile seizures; Pancreatitis, chronic, susceptibility to; Rhizomelic chondrodysplasia punctata type 2 and type 3; Disordered

steroidogenesis due to cytochrome p450 oxidoreductase deficiency; Deafness with labyrinthine aplasia microtia and microdontia (FAMM); Rothmund-Thomson syndrome; Cortical dysplasia, complex, with other brain malformations 5 and 6; Myasthenia, familial infantile, 1; Trichorhinophalangeal dysplasia type I; Worth disease; Splenic hypoplasia; Molybdenum cofactor deficiency, complementation group A; Sebastian syndrome; Progressive familial intrahepatic cholestasis 2 and 3; Weill-Marchesani syndrome 1 and 3; Microcephalic osteodysplastic primordial dwarfism type 2; Surfactant metabolism dysfunction, pulmonary, 2 and 3; Severe X-linked myotubular myopathy; Pancreatic cancer 3; Platelet- type bleeding disorder 15 and 8; Tyrosinase-positive oculocutaneous albinism; Borrone Di Rocco Crovato syndrome; ATR-X syndrome; Sucrase-isomaltase deficiency; Complement component 4, partial deficiency of, due to dysfunctional c1 inhibitor; Congenital central hypoventilation; Infantile hypophosphatasia; Plasminogen activator inhibitor type 1 deficiency; Malignant lymphoma, non-Hodgkin; Hyperornithinemia-hyperammonemia-homocitrullinuria syndrome; Schwartz Jampel syndrome type 1; Fetal hemoglobin quantitative trait locus 1; Myopathy, distal, with anterior tibial onset; Noonan syndrome 1 and 4, LEOPARD syndrome 1; Glaucoma 1, open angle, e, F, and G; Kenny-Caffey syndrome type 2; PTEN hamartoma tumor syndrome; Duchenne muscular dystrophy; Insulin-resistant diabetes mellitus and acanthosis nigricans; Microphthalmia, isolated 3, 5, 6, 8, and with coloboma 6; Raine syndrome; Premature ovarian failure 4, 5, 7, and 9; Allan-Herndon-Dudley syndrome; Citrullinemia type I; Alzheimer disease, familial, 3, with spastic paraparesis and apraxia; Familial hemiplegic migraine types 1 and 2; Ventriculomegaly with cystic kidney disease; Pseudoxanthoma elasticum; Homocysteinemia due to MTHFR deficiency, CBS deficiency, and Homocystinuria, pyridoxine-responsive; Dilated cardiomyopathy 1A, 1AA, 1C, 1G, IBB, 1DD, IFF, 1HH, II, IKK, IN, IS, 1Y, and 3B; Muscle AMP guanine oxidase deficiency; Familial cancer of breast; Hereditary sideroblastic anemia; Myoglobinuria, acute recurrent, autosomal recessive; Neuroferritinopathy; Cardiac arrhythmia; Glucose transporter type 1 deficiency syndrome; Holoprosencephaly sequence; Angiopathy, hereditary, with nephropathy, aneurysms, and muscle cramps; Isovaleryl-CoA dehydrogenase deficiency; Kallmann syndrome 1, 2, and 6; Permanent neonatal diabetes mellitus; Acrocallosal syndrome, Schinzel type; Gordon syndrome; MYH9 related disorders; Donnai Barrow syndrome; Severe congenital neutropenia and 6, autosomal recessive; Charcot-Marie-Tooth disease, types ID and IVF; Coffin-Lowry syndrome; mitochondrial 3-hydroxy-3- methylglutaryl-CoA synthase

deficiency; Hypomagnesemia, seizures, and mental retardation; Ischiopatellar dysplasia; Multiple congenital anomalies -hypotonia- seizures syndrome 3; Spastic paraplegia 50, autosomal recessive; Short stature with nonspecific skeletal abnormalities; Severe myoclonic epilepsy in infancy; Propionic academia; Adolescent nephronophthisis; Macrocephaly, macrosomia, facial dysmorphism syndrome; Stargardt disease 4; Ehlers-Danlos syndrome type 7 (autosomal recessive), classic type, type 2 (progeroid), hydroxylysine-deficient, type 4, type 4 variant, and due to tenascin-X deficiency; Myopia 6; Coxa plana; Familial cold autoinflammatory syndrome 2; Malformation of the heart and great vessels; von Willebrand disease type 2M and type 3; Deficiency of galactokinase; Brugada syndrome 1; X-linked ichthyosis with steryl-sulfatase deficiency; Congenital ocular coloboma; Histiocytosis-lymphadenopathy plus syndrome; Aniridia, cerebellar ataxia, and mental retardation; Left ventricular noncompaction 3; Amyotrophic lateral sclerosis types 1, 6, 15 (with or without frontotemporal dementia), 22 (with or without frontotemporal dementia), and 10; Osteogenesis imperfecta type 12, type 5, type 7, type 8, type I, type III, with normal sclerae, dominant form, recessive perinatal lethal; Hematologic neoplasm; Favism, susceptibility to; Pulmonary Fibrosis And/Or Bone Marrow Failure, Telomere-Related, 1 and 3; Dominant hereditary optic atrophy; Dominant dystrophic epidermolysis bullosa with absence of skin; Muscular dystrophy, congenital, megaconial type; Multiple gastrointestinal atresias; McCune- Albright syndrome; Nail-patella syndrome; McLeod neuroacanthocytosis syndrome; Common variable immunodeficiency 9; Partial hypoxanthine-guanine phosphoribosyltransferase deficiency; Pseudohypoaldosteronism type 1 autosomal dominant and recessive and type 2; Urocanate hydratase deficiency; Heterotopia; Meckel syndrome type 7; Ch\xc3\xa9diak-Higashi syndrome , Chediak-Higashi syndrome, adult type; Severe combined immunodeficiency due to ADA deficiency, with microcephaly, growth retardation, and sensitivity to ionizing radiation, atypical, autosomal recessive, T cell-negative, B cell-positive, NK cell-negative of NK- positive; Insulin resistance; Deficiency of steroid 11 -beta-monooxygenase; Popliteal pterygium syndrome; Pulmonary arterial hypertension related to hereditary hemorrhagic telangiectasia; Deafness, autosomal recessive 1A, 2, 3, 6, 8, 9, 12, 15, 16, 18b, 22, 28, 31, 44, 49, 63, 77, 86, and 89; Primary hyperoxaluria, type I, type, and type III; Paramyotonia congenita of von Eulenburg; Desbuquois syndrome; Carnitine palmitoyltransferase I , II, II (late onset), and II (infantile) deficiency; Secondary hypothyroidism; Mandibulofacial dysostosis, Treacher Collins type,

autosomal recessive; Cowden syndrome 1; Li-Fraumeni syndrome 1; Asparagine synthetase deficiency; Malattia leventinese; Optic atrophy 9; Infantile convulsions and paroxysmal choreoathetosis, familial; Ataxia with vitamin E deficiency; Islet cell hyperplasia; Miyoshi muscular dystrophy 1; Thrombophilia, hereditary, due to protein C deficiency, autosomal dominant and recessive; Fechtner syndrome; Properdin deficiency, X-linked; Mental retardation, stereotypic movements, epilepsy, and/or cerebral malformations; Creatine deficiency, X-linked; Pilomatrixoma; Cyanosis, transient neonatal and atypical nephropathic; Adult onset ataxia with oculomotor apraxia; Hemangioma, capillary infantile; PC-K6a; Generalized dominant dystrophic epidermolysis bullosa; Pelizaeus-Merzbacher disease; Myopathy, centronuclear, 1, congenital, with excess of muscle spindles, distal, 1, lactic acidosis, and sideroblastic anemia 1, mitochondrial progressive with congenital cataract, hearing loss, and developmental delay, and tubular aggregate, 2; Benign familial neonatal seizures 1 and 2; Primary pulmonary hypertension; Lymphedema, primary, with myelodysplasia; Congenital long QT syndrome; Familial exudative vitreoretinopathy, X-linked; Autosomal dominant hypohidrotic ectodermal dysplasia; Primordial dwarfism; Familial pulmonary capillary hemangiomatosis; Carnitine acylcarnitine translocase deficiency; Visceral myopathy; Familial Mediterranean fever and Familial mediterranean fever, autosomal dominant; Combined partial and complete 17-alpha-hydroxylase/ 17, 20-lyase deficiency; Oto-palato-digital syndrome, type I; Nephrolithiasis/osteoporosis, hypophosphatemic, 2; Familial type 1 and 3 hyperlipoproteinemia; Phenotypes; CHARGE association; Fuhrmann syndrome; Hypotrichosis-lymphedema-telangiectasia syndrome; Chondrodysplasia Blomstrand type; Acroerythrokeratoderma; Slowed nerve conduction velocity, autosomal dominant; Hereditary cancer-predisposing syndrome; Craniodiaphyseal dysplasia, autosomal dominant; Spinocerebellar ataxia autosomal recessive 1 and 16; Proprotein convertase 1/3 deficiency; D-2-hydroxyglutaric aciduria 2; Hyperekplexia 2 and Hyperekplexia hereditary; Central core disease; Opitz G/BBB syndrome; Cystic fibrosis; Thiel-Behnke corneal dystrophy; Deficiency of bisphosphoglycerate mutase; Mitochondrial short-chain Enoyl-CoA Hydratase 1 deficiency; Ectodermal dysplasia skin fragility syndrome; Wolfram-like syndrome, autosomal dominant; Microcytic anemia; Pyruvate carboxylase deficiency; Leukocyte adhesion deficiency type I and III; Multiple endocrine neoplasia, types land 4; Transient bullous dermolysis of the newborn; Primrose syndrome; Non-small cell lung cancer; Congenital muscular dystrophy; Lipase deficiency combined; COLE-CARPENTER

SYNDROME 2; Atrioventricular septal defect and common atrioventricular junction; Deficiency of xanthine oxidase; Waardenburg syndrome type 1, 4C, and 2E (with neurologic involvement); Stickler syndrome, types I(nonsyndromic ocular) and 4; Comeal fragility keratoglobus, blue sclerae and joint hypermobility; Microspherophakia; Chudley-McCullough syndrome; Epidermolysa bullosa simplex and limb girdle muscular dystrophy, simplex with mottled pigmentation, simplex with pyloric atresia, simplex, autosomal recessive, and with pyloric atresia; Rett disorder; Abnormality of neuronal migration; Growth hormone deficiency with pituitary anomalies; Leigh disease; Keratosis palmoplantaris striata 1; Weissenbacher-Zweymuller syndrome; Medium-chain acyl-coenzyme A dehydrogenase deficiency; UDPglucose-4- epimerase deficiency; susceptibility to Autism, X-linked 3; Rhegmatogenous retinal detachment, autosomal dominant; Familial febrile seizures 8; Ulna and fibula absence of with severe limb deficiency; Left ventricular noncompaction 6; Centromeric instability of chromosomes 1,9 and 16 and immunodeficiency; Hereditary diffuse leukoencephalopathy with spheroids; Cushing syndrome; Dopamine receptor d2, reduced brain density of; C-like syndrome; Renal dysplasia, retinal pigmentary dystrophy, cerebellar ataxia and skeletal dysplasia; Ovarian dysgenesis 1; Pierson syndrome; Polyneuropathy, hearing loss, ataxia, retinitis pigmentosa, and cataract; Progressive intrahepatic cholestasis; autosomal dominant, autosomal recessive, and X-linked recessive Alport syndromes; Angelman syndrome; Amish infantile epilepsy syndrome; Autoimmune lymphoproliferative syndrome, type Ia; Hydrocephalus; Marfanoid habitus; Bare lymphocyte syndrome type 2, complementation group E; Recessive dystrophic epidermolysis bullosa; Factor H, VII, X, v and factor viii, combined deficiency of 2, xiii, a subunit, deficiency; Zonular pulverulent cataract 3; Warts, hypogammaglobulinemia, infections, and myelokathexis; Benign hereditary chorea; Deficiency of hyaluronoglucosaminidase; Microcephaly, hiatal hernia and nephrotic syndrome; Growth and mental retardation, mandibulofacial dysostosis, microcephaly, and cleft palate; Lymphedema, hereditary, id; Delayed puberty; Apparent mineralocorticoid excess; Generalized arterial calcification of infancy 2; METHYLMALONIC ACIDURIA, mut(0) TYPE; Congenital heart disease, multiple types, 2; Familial hypoplastic, glomerulocystic kidney; Cerebrooculofacioskeletal syndrome 2; Stargardt disease 1; Mental retardation, autosomal recessive 15, 44, 46, and 5; Prolidase deficiency; Methylmalonic aciduria cblB type, ; Oguchi disease; Endocrine-cerebroosteodysplasia; Lissencephaly 1, 2 (X-linked), 3, 6 (with

microcephaly), X-linked; Somatotroph adenoma; Gamstorp- Wohlfart syndrome; Lipid proteinosis; Inclusion body myopathy 2 and 3; Enlarged vestibular aqueduct syndrome; Osteoporosis with pseudoglioma; Acquired long QT syandrome; Phenylketonuria; CHOPS syndrome; Global developmental delay; Bietti crystalline corneoretinal dystrophy; Noonan syndrome-like disorder with or without juvenile myelomonocytic leukemia; Congenital erythropoietic porphyria; Atrophia bulborum hereditaria; Paragangliomas 3; Van der Woude syndrome; Aromatase deficiency; Birk Barel mental retardation dysmorphism syndrome; Amyotrophic lateral sclerosis type 5; Methemoglobinemia types I 1 and 2; Congenital stationary night blindness, type 1A, IB, 1C, IE, IF, and 2A; Seizures; Thyroid cancer, follicular; Lethal congenital contracture syndrome 6; Distal hereditary motor neuronopathy type 2B; Sex cord-stromal tumor; Epileptic encephalopathy, childhood-onset, early infantile, 1, 19, 23, 25, 30, and 32; Myofibrillar myopathy 1 and ZASP-related; Cerebellar ataxia infantile with progressive external ophthalmoplegia; Purine-nucleoside phosphorylase deficiency; Forebrain defects; Epileptic encephalopathy Lennox-Gastaut type; Obesity; 4, Left ventricular noncompaction 10; Verheij syndrome; Mowat-Wilson syndrome; Odontotrichomelic syndrome; Patterned dystrophy of retinal pigment epithelium; Lig4 syndrome; Barakat syndrome; IRAK4 deficiency; Somatotroph adenoma; Branched-chain ketoacid dehydrogenase kinase deficiency; Cystinuria; Familial aplasia of the vermis; Succinyl-CoA acetoacetate transferase deficiency; Scapuloperoneal spinal muscular atrophy; Pigmentary retinal dystrophy; Glanzmann thrombasthenia; Primary open angle glaucoma juvenile onset 1; Aicardi Goutieres syndromes 1, 4, and 5; Renal dysplasia; Intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, and genital anomalies; Beaded hair; Short stature, onychodysplasia, facial dysmorphism, and hypotrichosis; Metachromatic leukodystrophy; Cholestanol storage disease; Three M syndrome 2; Leber congenital amaurosis 11, 12, 13, 16, 4, 7, and 9; Mandibuloacral dysplasia with type A or B lipodystrophy, atypical; Meier-Gorlin syndromes land 4; Hypotrichosis 8 and 12; Short QT syndrome 3; Ectodermal dysplasia l ib; Anonychia; Pseudohypoparathyroidism type 1A, Pseudopseudohypoparathyroidism; Leber optic atrophy; Bainbridge- Ropers syndrome; Weaver syndrome; Short stature, auditory canal atresia, mandibular hypoplasia, skeletal abnormalities; Deficiency of alpha-mannosidase; Macular dystrophy, vitelliform, adult-onset; Glutaric aciduria, type 1; Gangliosidosis GM1 typel (with cardiac involvenment) 3; Mandibuloacral dysostosis; Hereditary lymphedema type I; Atrial

standstill 2; Kabuki make-up syndrome; Bethlem myopathy and Bethlem myopathy 2; Myeloperoxidase deficiency; Fleck corneal dystrophy; Hereditary acrodermatitis enteropathica; Hypobetalipoproteinemia, familial, associated with apob32; Cockayne syndrome type A, ; Hyperparathyroidism, neonatal severe; Ataxia-telangiectasia-like disorder; Pendred syndrome; I blood group system; Familial benign pemphigus; Visceral heterotaxy 5, autosomal; Nephrogenic diabetes insipidus, Nephrogenic diabetes insipidus, X-linked; Minicore myopathy with external ophthalmoplegia; Perry syndrome; hypohidrotic/hair/tooth type, autosomal recessive; Hereditary pancreatitis; Mental retardation and microcephaly with pontine and cerebellar hypoplasia; Glycogen storage disease 0 (muscle), II (adult form), IXa2, IXc, type 1A; Osteopathia striata with cranial sclerosis; Gluthathione synthetase deficiency; Brugada syndrome and Brugada syndrome 4; Endometrial carcinoma; Hypohidrotic ectodermal dysplasia with immune deficiency; Cholestasis, intrahepatic, of pregnancy 3; Bernard-Soulier syndrome, types A1 and A2 (autosomal dominant); Salla disease; Ornithine aminotransferase deficiency; PTEN hamartoma tumor syndrome; Distichiasis- lymphedema syndrome; Corticosteroid-binding globulin deficiency; Adult neuronal ceroid lipofuscinosis; Dejerine-Sottas disease; Tetraamelia, autosomal recessive; Senior-Loken syndrome 4 and 5, ; Glutaric acidemia IIA and IIB; Aortic aneurysm, familial thoracic 4, 6, and 9; Hyperphosphatasia with mental retardation syndrome 2, 3, and 4; Dyskeratosis congenita X-linked; Arthrogryposis, renal dysfunction, and cholestasis 2; Bannayan-Riley-Ruvalcaba syndrome; 3- Methylglutaconic aciduria; Isolated 17,20-lyase deficiency; Gorlin syndrome; Hand foot uterus syndrome; Tay-Sachs disease, B1 variant, Gm2-gangliosidosis (adult), Gm2-gangliosidosis (adult-onset); Dowling-degos disease 4; Parkinson disease 14, 15, 19 (juvenile-onset), 2, 20 (early-onset), 6, (autosomal recessive early-onset, and 9; Ataxia, sensory, autosomal dominant; Congenital microvillous atrophy; Myoclonic- Atonic Epilepsy; Tangier disease;2- methyl-3-hydroxybutyric aciduria; Familial renal hypouricemia; Schizencephaly; Mitochondrial DNA depletion syndrome 4B, MNGIE type; Feingold syndrome 1; Renal carnitine transport defect; Familial hypercholesterolemia; Townes-Brocks-branchiootorenal-like syndrome; Griscelli syndrome type 3; Meckel-Gruber syndrome; Bullous ichthyosiform erythroderma; Neutrophil immunodeficiency syndrome; Myasthenic Syndrome, Congenital, 17, 2A (slow-channel), 4B (fast-channel), and without tubular aggregates; Microvascular complications of diabetes 7; McKusick Kaufman syndrome; Chronic granulomatous disease, autosomal recessive cytochrome b-positive, types 1 and 2; Arginino

succinate lyase deficiency; Mitochondrial phosphate carrier and pyruvate carrier deficiency; Lattice corneal dystrophy Type III; Ectodermal dysplasia-syndactyly syndrome 1; Hypomyelinating leukodystrophy 7; Mental retardation, autosomal dominant 12, 13, 15, 24, 3, 30, 4, 5, 6, and 9; Generalized epilepsy with febrile seizures plus, types 1 and 2; Psoriasis susceptibility 2; Frank Ter Haar syndrome; Thoracic aortic aneurysms and aortic dissections; Crouzon syndrome; Granulosa cell tumor of the ovary; Epidermolytic palmoplantar keratoderma; Leri Weill dyschondrosteosis; 3 beta-Hydroxysteroid dehydrogenase deficiency; Familial restrictive cardiomyopathy 1; Autosomal dominant progressive external ophthalmoplegia with mitochondrial DNA deletions 1 and 3; Antley-Bixler syndrome with genital anomalies and disordered steroidogenesis; Hajdu-Cheney syndrome; Pigmented nodular adrenocortical disease, primary, 1; Episodic pain syndrome, familial, 3; Dejerine-Sottas syndrome, autosomal dominant; FG syndrome and FG syndrome 4; Dendritic cell, monocyte, B lymphocyte, and natural killer lymphocyte deficiency; Hypothyroidism, congenital, nongoitrous, 1; Miller syndrome; Nemaline myopathy 3 and 9; Oligodontia- colorectal cancer syndrome; Cold-induced sweating syndrome 1; Van Buchem disease type 2; Glaucoma 3, primary congenital, d; Citrullinemia type I and II; Nonaka myopathy; Congenital muscular dystrophy due to partial LAMA2 deficiency; Myoneural gastrointestinal encephalopathy syndrome; Leigh syndrome due to mitochondrial complex I deficiency; Medulloblastoma; Pyruvate dehydrogenase E1 -alpha deficiency; Carcinoma of colon; Nance-Horan syndrome; Sandhoff disease, adult and infantil types; Arthrogryposis renal dysfunction cholestasis syndrome; Autosomal recessive hypophosphatemic bone disease; Doyne honeycomb retinal dystrophy; Spinocerebellar ataxia 14, 21, 35, 40, and 6; Lewy body dementia; RRM2B -related mitochondrial disease; Brody myopathy; Megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome 2; Usher syndrome, types 1, IB, ID, 1G, 2A, 2C, and 2D; hypocalcification type and hypomaturation type, IIA1 Amelogenesis imperfecta; Pituitary hormone deficiency, combined 1, 2, 3, and 4; Cushing symphalangism; Renal tubular acidosis, distal, autosomal recessive, with late-onset sensorineural hearing loss, or with hemolytic anemia; Infantile nephronophthisis; Juvenile polyposis syndrome; Sensory ataxic neuropathy, dysarthria, and ophthalmoparesis; Deficiency of 3-hydroxyacyl-CoA dehydrogenase; Parathyroid carcinoma; X-linked agammaglobulinemia; Megaloblastic anemia, thiamine-responsive, with diabetes mellitus and sensorineural deafness; Multiple sulfatase deficiency; Neurodegeneration with brain iron accumulation 4 and 6; Cholesterol monooxygenase (side-chain cleaving)

deficiency; hemolytic anemia due to Adenylosuccinate lyase deficiency; Myoclonus with epilepsy with ragged red fibers; Pitt- Hopkins syndrome; Multiple pterygium syndrome Escobar type; Homocystinuria-Megaloblastic anemia due to defect in cobalamin metabolism, cblE complementation type; Cholecystitis; Spherocytosis types 4 and 5; Multiple congenital anomalies; Xeroderma pigmentosum, complementation group b, group D, group E, and group G; Leiner disease; Groenouw corneal dystrophy type I; Coenzyme Q10 deficiency, primary 1, 4, and 7; Distal spinal muscular atrophy, congenital nonprogressive; Warburg micro syndrome 2 and 4; Bile acid synthesis defect, congenital, 3; Acth-independent macronodular adrenal hyperplasia 2; Acrocapitofemoral dysplasia; Paget disease of bone, familial; Severe neonatal-onset encephalopathy with microcephaly; Zimmermann-Laband syndrome and Zimmermann-Laband syndrome 2; Reifenstein syndrome; Familial hypokalemia-hypomagnesemia; Photosensitive trichothiodystrophy; Adult junctional epidermolysis bullosa; Lung cancer; Freeman-Sheldon syndrome; Hyperinsulinism-hyperammonemia syndrome; Posterior polar cataract type 2; Sclerocornea, autosomal recessive; Juvenile GM>1< gangliosidosis; Cohen syndrome, ; Hereditary Paraganglioma- Pheochromocytoma Syndromes; Neonatal insulin-dependent diabetes mellitus; Hypochondrogenesis; Floating-Harbor syndrome; Cutis laxa with osteodystrophy and with severe pulmonary, gastrointestinal, and urinary abnormalities; Congenital contractures of the limbs and face, hypotonia, and developmental delay; Dyskeratosis congenita autosomal dominant and autosomal dominant, 3; Histiocytic medullary reticulosis; Costello syndrome; Immunodeficiency 15, 16, 19, 30, 31C, 38, 40, 8, due to defect in cd3-zeta, with hyper IgM type 1 and 2, and X-Linked, with magnesium defect, Epstein-Barr vims infection, and neoplasia; Atrial septal defects 2, 4, and 7 (with or without atrioventricular conduction defects); GTP cyclohydrolase I deficiency; Talipes equinovarus; Phosphoglycerate kinase 1 deficiency; Tuberous sclerosis 1 and 2; Autosomal recessive congenital ichthyosis 1, 2, 3, 4A, and 4B; and Familial hypertrophic cardiomyopathy 1, 2, 3, 4, 7, 10, 23 and 24.

Indications by tissue

Additional suitable diseases and disorders that can be treated by the systems and methods provided herein include, without limitation, diseases of the central nervous system (CNS) (see exemplary diseases and affected genes in **Table 13**), diseases of the eye (see exemplary diseases and affected genes in **Table 14**), diseases of the heart (see exemplary diseases and affected genes

in **Table 15**), diseases of the hematopoietic stem cells (HSC) (see exemplary diseases and affected genes in **Table 16**), diseases of the kidney (see exemplary diseases and affected genes in **Table 17**), diseases of the liver (see exemplary diseases and affected genes in **Table 18**), diseases of the lung (see exemplary diseases and affected genes in **Table 19**), diseases of the skeletal muscle (see exemplary diseases and affected genes in **Table 20**), and diseases of the skin (see exemplary diseases and affected genes in **Table 21**). **Table 22** provides exemplary protective mutations that reduce risks of the indicated diseases. In some embodiments, a Gene Writer system described herein is used to treat an indication of any of Tables 13-21. In some embodiments, the GeneWriter system modifies a target site in genomic DNA in a cell, wherein the target site is in a gene of any of Tables 13-21, e.g., in a subject having the corresponding indication listed in any of Tables 13-21. In some embodiments, the GeneWriter corrects a mutation in the gene. In some embodiments, the GeneWriter inserts a sequence that had been deleted from the gene (e.g., through a disease-causing mutation). In some embodiments, the GeneWriter deletes a sequence that had been duplicated in the gene (e.g., through a disease-causing mutation). In some embodiments, the GeneWriter replaces a mutation (e.g., a disease-causing mutation) with the corresponding wild-type sequence. In some embodiments, the mutation is a substitution, insertion, deletion, or inversion.

Table 13. CNS diseases and genes affected.

Disease	**Gene Affected**
Alpha-mannosidosis	MAN2B1
Ataxia-telangiectasia	ATM
CADASIL	NOTCH3
Canavan disease	ASPA
Carbamoyl-phosphate synthetase 1 deficiency	CPS1
CLN1 disease	PPT1
CLN2 Disease	TPP1
CLN3 Disease (Juvenile neuronal ceroid lipofuscinosis, Batten Disease)	CLN3
Coffin-Lowry syndrome	RPS6KA3
Congenital myasthenic syndrome 5	COLQ

Cornelia de Lange syndrome (NIPBL)	NIPBL
Cornelia de Lange syndrome (SMC1A)	SMC1A
Dravet syndrome (SCN1A)	SCN1A
Glycine encephalopathy (GLDC)	GLDC
GM1 gangliosidosis	GLB1
Huntington's Disease	HTT
Hydrocephalus with stenosis of the aqueduct of Sylvius	L1CAM
Leigh Syndrome	SURF1
Metachromatic leukodystrophy (ARSA)	ARSA
MPS type 2	IDS
MPS type 3	Type 3a: SGSH Type 3b: NAGLU
Mucolipidosis IV	MCOLN1
Neurofibromatosis Type 1	NF1
Neurofibromatosis type 2	NF2
Pantothenate kinase-associated neurodegeneration	PANK2
Pyridoxine-dependent epilepsy	ALDH7A1
Rett syndrome (MECP2)	MECP2
Sandhoff disease	HEXB
Semantic dementia (Frontotemporal dementia)	MAPT
Spinocerebellar ataxia with axonal neuropathy (Ataxia with Oculomotor Apraxia)	SETX
Tay-Sachs disease	HEXA
X-linked Adrenoleukodystrophy	ABCD1

Table 14. Eye diseases and genes affected.

Disease	**Gene Affected**
Achromatopsia	CNGB3
Amaurosis Congenita (LCA1)	GUCY2D
Amaurosis Congenita (LCA10)	CEP290

Disease	Gene Affected
Amaurosis Congenita (LCA2)	RPE65
Amaurosis Congenita (LCA8)	CRB1
Choroideremia	CHM
Cone Rod Dystrophy (ABCA4)	ABCA4
Cone Rod Dystrophy (CRX)	CRX
Cone Rod Dystrophy (GUCY2D)	GUCY2D
Cystinosis, Ocular Nonnephropathic	CTNS
Lattice corneal dystrophy type I	TGFBI
Macular Corneal Dystrophy (MCD)	CHST6
Optic Atrophy	OPA1
Retinitis Pigmentosa (AR)	USH2A
Retinitis Rigmentosa (AD)	RHO
Stargardt Disease	ABCA4
Vitelliform Macular Dystrophy	BEST1; PRPH2

Table 15. Heart diseases and genes affected.

Disease	Gene Affected
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	PKP2
Barth syndrome	TAZ
Becker muscular dystrophy	DMD
Brugada syndrome	SCN5A
Catecholaminergic polymorphic ventricular tachycardia (RYR2)	RYR2
Dilated cardiomyopathy (LMNA)	LMNA
Dilated cardiomyopathy (TTN)	TTN
Duchenne muscular dystrophy	DMD
Emery-Dreifuss Muscular Dystrophy Type I	EMD
Familial hypertrophic cardiomyopathy	MYH7
Familial hypertrophic cardiomyopathy	MYBPC3
Jervell Lange-Nielsen syndrome	KCNQ1
LCHAD deficiency	HADHA
Limb-girdle muscular dystrophy type 1B (Emery-Dreifuss EDMD2)	LMNA

Limb-girdle muscular dystrophy, type 2D	SGCA
Long QT syndrome 1 (Romano Ward)	KCNQ1

Table 16. HSC diseases and genes affected.

Disease	Gene Affected
ADA-SCID	ADA
Adrenoleukodystrophy (CALD)	ABCD1
Alpha-mannosidosis	MAN2B1
Chronic granulomatous disease	CYBB; CYBA; NCF1; NCF2; NCF4
Common variable immunodeficiency	TNFRSF13B
Fanconi anemia	FANCA; FANCC; FANCG
Gaucher disease	GBA
Globoid cell leukodystrophy (Krabbe disease)	GALC
Hemophagocytic lymphohistiocytosis	PRF1; STX11; STXBP2; UNC13D
IL-7R SCID	IL7R
JAK-3 SCID	JAK3
Malignant infantile osteopetrosis- autosomal recessive osteopetrosis	TCIRG1; Many genes implicated
Metachromatic leukodystrophy	ARSA; PSAP
MPS 1S (Scheie syndrome)	IDUA
MPS2	IDS
MPS7	GUSB
Mucolipidosis II	GNPTAB
Niemann-Pick disease A and B	SMPD1
Niemann-Pick disease C	NPC1
Paroxysmal Nocturnal Hemoglobinuria	PIGA
Pompe disease	GAA

Pyruvate kinase deficiency (PKD)	PKLR
RAG 1/2 Deficiency (SCID with granulomas)	RAG1/RAG2
Severe Congenital Neutropenia	ELANE; HAX1
Sickle cell disease (SCD)	HBB
Tay Sachs	HEXA
Thalassemia	HBB
Wiskott-Aldrich Syndrome	WAS
X-linked agammaglobulinemia	BTK
X-linked SCID	IL2RG

Table 17. Kidney diseases and genes affected.

Disease	**Gene Affected**
Alport syndrome	COL4A5
Autosomal dominant polycystic kidney disease (PKD1)	PKD1
Autosomal dominant polycystic kidney disease (PKD2)	PDK2
Autosomal dominant tubulointerstitial kidney disease (MUC1)	MUC1
Autosomal dominant tubulointerstitial kidney disease (UMOD)	UMOD
Autosomal recessive polycystic kidney disease	PKHD1
Congenital nephrotic syndrome	NPHS2
Cystinosis	CTNS

Table 18. Liver diseases and genes affected.

Disease	**Gene Affected**
Acute intermittent porphyria	HMBS

Alagille syndrome	JAG1
Alpha-1-antitrypsin deficiency	SERPINA1
Carbamoyl phosphate synthetase I deficiency	CPS1
Citrullinemia I	ASS1
Crigler-Najjar	UGT1A1
Fabry	LPL
Familial chylomicronemia syndrome	GLA
Gaucher	GBE1
GSD IV	GBA
Heme A	F8
Heme B	F9
Hereditary amyloidosis (hTTR)	TTR
Hereditary angioedema	SERPING1 (KLKB1 for CRISPR)
HoFH	LDLRAP1
Hypercholesterolemia	PCSK9
Methylmalonic acidemia	MMUT
MPS II	IDS
MPS III	Type IIIa: SGSH Type IIIb: NAGLU Type IIIc: HGSNAT Type IIId: GNS
MPS IV	Type IVA: GALNS Type IVB: GLB1
MPS VI	ARSB
MSUD	Type Ia: BCKDHA Type Ib: BCKDHB Type II: DBT
OTC Deficiency	OTC
Polycystic Liver Disease	PRKCSH
Pompe	GAA

Primary Hyperoxaluria 1	AGXT (HAO1 or LDHA for CRISPR)
Progressive familial intrahepatic cholestasis type 1	ATP8B1
Progressive familial intrahepatic cholestasis type 2	ABCB11
Progressive familial intrahepatic cholestasis type 3	ABCB4
Propionic acidemia	PCCB; PCCA
Wilson's Disease	ATP7B
Glycogen storage disease, Type 1a	G6PC
Glycogen storage disease, Type IIIb	AGL
Isovaleric acidemia	IVD
Wolman disease	LIPA

Table 19. Lung diseases and genes affected.

Disease	**Gene Affected**
Alpha-1 antitrypsin deficiency	SERPINA1
Cystic fibrosis	CFTR
Primary ciliary dyskinesia	DNAI1
Primary ciliary dyskinesia	DNAH5
Primary pulmonary hypertension I	BMPR2
Surfactant Protein B (SP-B) Deficiency (pulmonary surfactant metabolism dysfunction 1)	SFTPB

Table 20. Skeletal muscle diseases and genes affected.

Disease	**Gene Affected**
Becker muscular dystrophy	DMD
Becker myotonia	CLCN1
Bethlem myopathy	COL6A2
Centronuclear myopathy, X-linked (myotubular)	MTM1
Congenital myasthenic syndrome	CHRNE
Duchenne muscular dystrophy	DMD

Emery-Dreifuss muscular dystrophy, AD	LMNA
Facioscapulohumeral Muscular Dystrophy	DUX4 - D4Z4 chromosomal region
Hyperkalemic periodic paralysis	SCN4A
Hypokalemic periodic paralysis	CACNA1S
Limb-girdle muscular dystrophy 2A	CAPN3
Limb-girdle muscular dystrophy 2B	DYSF
Limb-girdle muscular dystrophy, type 2D	SGCA
Miyoshi muscular dystrophy 1	DYSF
Paramyotonia congenita	SCN4A
Thomsen myotonia	CLCN1
VCP myopathy (IBMPFD) 1	VCP

Table 21. Skin diseases and genes affected.

Disease	**Gene Affected**
Epidermolysis Bullosa Dystrophica Dominant	COL7A1
Epidermolysis Bullosa Dystrophica Recessive (Hallopeau-Siemens)	COL7A1
Epidermolysis Bullosa Junctional	LAMB3
Epidermolysis Bullosa Simplex	KRT5; KRT14
Epidermolytic Ichthyosis	KRT1; KRT10
Hailey-Hailey Disease	ATP2C1
Lamellar Ichthyosis/Nonbullous Congenital Ichthyosiform Erythroderma (ARCI)	TGM1
Netherton Syndrome	SPINK5

Table 22. Exemplary protective mutations that reduce disease risk.

Disease	**Gene**	**Exemplary Protective Mutation**
Alzheimer's	APP	A673T
Parkinson's	SGK1	
Diabetes (Type II)	SLC30A8	p.Arg138X; p.Lys34SerfsX50
Cardiovascular Disease	PCSK9	R46L

Cardiovascular Disease	ASGR1	NM_001671.4, c.284-36_283+33delCTGGGGCTGGGG; NP_001662.1, p.W158X
Cardiovascular Disease	NPC1L1	p.Arg406X
Cardiovascular Disease	APOC3	R19X; IVS2+1G→A; A43T
Cardiovascular Disease	LPA	
Cardiovascular Disease	ANGPTL4	E40K
Cardiovascular Disease	ANGPTL3	p.Ser17Ter; p.Asn121fs; p.Asn147fs; c.495+6T→C
HIV infection	CCR5	CCR5-delta32

Pathogenic mutations

In some embodiments, the systems or methods provided herein can be used to correct a pathogenic mutation. The pathogenic mutation can be a genetic mutation that increases an individual's susceptibility or predisposition to a certain disease or disorder. In some embodiments, the pathogenic mutation is a disease-causing mutation in a gene associated with a disease or disorder. In some embodiments, the systems or methods provided herein can be used to revert the pathogenic mutation to its wild-type counterpart. In some embodiments, the systems or methods provided herein can be used to change the pathogenic mutation to a sequence not causing the disease or disorder.

Table 23 provides exemplary indications (column 1), underlying genes (column 2), and pathogenic mutations that can be corrected using the systems or methods described herein (column 3).

Table 23. Indications, genes, and causitive pathogenic mutations.

Disease	**Gene**	**Pathogenic Mutation**[#]
Achromatopsia	CNGB3	1148delC
Alpha-1 Antitrypsin Deficiency	SERPINAI	E342K
Alpha-1 Antitrypsin Deficiency	SERPINAI	E342K
Alpha-1 Antitrypsin Deficiency	SERPINAI	R48C (R79C)
Amaurosis Congenita (LCA10)	CEP290	2991+1655A>G
Andersen– Tawil syndrome	KCNJ2	R218W

Disease	Gene	Pathogenic Mutation[#]
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	PKP2	c.235C>T
associated with congenital factor XI deficiency	F11	E117*
associated with congenital factor XI deficiency	F11	F283L
ATTR amyloidosis	TTR	V50M/N30M
autosomal dominant deafness	COCH	G88E
autosomal dominant deafness	TECTA	Y1870C
autosomal dominant Parkinson's disease	SNCA	A53T
autosomal dominant Parkinson's disease	SNCA	A30P
Autosomal dominant rickets	FGF23	R176Q
autosomal recessive deafness	CX30	T5M
autosomal recessive deafness	DFNB59	R183W
autosomal recessive deafness	TMC1	Y182C
autosomal recessive hypercholesterolemia	ARH	Q136*
Blackfan-Diamond anemia	RPS19	R62Q
blue-cone monochromatism	OPN1LW	C203R
Brugada syndrome	SCN5A	E1784K
CADASIL syndrome	NOTCH3 gene	R90C
CADASIL syndrome	NOTCH3 gene	R141C
Canavan disease	ASPA	E285A
Canavan disease	ASPA	Y231X
Canavan disease	ASPA	A305E
carnitine palmitoyltransferase II deficiency	CPT2	S113L
choroideremia	CHM	R293*
choroideremia	CHM	R270*
choroideremia	CHM	A117A
Citrullinemia Type I	ASS	G390R
classic galactosemia	GALT	Q188R
classic homocysteinuria	CBS	T191M
classic homocysteinuria	CBS	G307S
CLN2 Disease	TPP1	c.509–1 G>C
CLN2 Disease	TPP1	c.622 C>T
CLN2 Disease	TPP1	c.851 G>T
cone-rod dystrophy	GUCY2D	R838C
congenital factor V deficiency	F5	R506Q
congenital factor V deficiency	F5	R534Q
congenital factor VII deficiency	F7	A294V
congenital factor VII deficiency	F7	C310F
congenital factor VII deficiency	F7	R304Q
congenital factor VII deficiency	F7	QI00R
Creutzfeldt- Jakob disease (CJD)	PRNP	E200K
Creutzfeldt- Jakob disease (CJD)	PRNP	M129V
Creutzfeldt- Jakob disease (CJD)	PRNP	P102L
Creutzfeldt- Jakob disease (CJD)	PRNP	D178N

Disease	Gene	Pathogenic Mutation[#]
cystic fibrosis	CFTR	G551D
cystic fibrosis	CFTR	W1282*
cystic fibrosis	CFTR	R553*
cystic fibrosis	CFTR	R117H
cystic fibrosis	CFTR	delta F508
cystinosis	CTNS	W138*
Darier disease	ATP2A2	N767S
Darier disease	ATP2A2	N767S
Darier disease	ATP2A2	N767S
Epidermolysis Bullosa Junctional	LAMB3	R42X
Epidermolysis Bullosa Junctional	LAMB3	R635X
familial amyotrophic lateral sclerosis (ALS)	SOD1	A4V
familial amyotrophic lateral sclerosis (ALS)	SOD1	H46R
familial amyotrophic lateral sclerosis (ALS)	SOD1	G37R
Gaucher disease	GBA	N370S
Gaucher disease	GBA	N370S
Gaucher disease	GBA	L444P
Gaucher disease	GBA	L444P
Gaucher disease	GBA	L483P
glutaryl-CoA dehydrogenase deficiency	GCDH	R138G
glutaryl-CoA dehydrogenase deficiency	GCDH	M263V
glutaryl-CoA dehydrogenase deficiency	GCDH	R402W
glycine encephalopathy	GLDC	A389V
glycine encephalopathy	GLDC	G771R
glycine encephalopathy	GLDC	T269M
hemophilia A	F8	R2178C
hemophilia A	F8	R550C
hemophilia A	F8	R2169H
hemophilia A	F8	R1985Q
hemophilia B	F9	T342M
hemophilia B	F9	R294Q
hemophilia B	F9	R43Q
hemophilia B	F9	R191H
hemophilia B	F9	G106S
hemophilia B	F9	A279T
hemophilia B	F9	R75*
hemophilia B	F9	R294*
hemophilia B	F9	R379Q
Hereditary antithrombin deficiency type I	SERPINCI	R48C (R79C)
hereditary chronic pancreatitis	PRSS1	R122H
Hunter syndrome	IDS	R88C
Hunter syndrome	IDS	G374G
Hurler syndrome (MPS1)	IDUA	Q70*
Hurler syndrome (MPS1)	IDUA	W402*

Disease	Gene	Pathogenic Mutation[#]
Hyperkalemic periodic paralysis	SCN4A	T704M
Hyperkalemic periodic paralysis	SCN4A	M1592V
Hyperkalemic periodic paralysis	CACNA1S	p.Arg528X
Hyperkalemic periodic paralysis	CACNA1S	p.Arg1239
intermittent porphyria	HMBS	Rl73W
isolated agammaglobulinemia	E47	E555K
Lattice corneal dystrophy type I	TGFBI	Arg124Cys
LCHAD deficiency	HADHA	Glu474Gln
Leber congenital amaurosis 2	RPE65	R44*
Leber congenital amaurosis 2	RPE65	IVS1
Leber congenital amaurosis 2	RPE65	G-A, +5
Lesch-Nyhan syndrome	HPRTI	R51*
Lesch-Nyhan syndrome	HPRTI	R170*
Limb-girdle muscular dystrophy, type 2D	SGCA	Arg77Cys
Marteauz- Lamy Syndrome (MSPVI)	ARSB	Y210C
Mediterranean G6PD deficiency	G6PD	S188D
medium-chain acyl-CoA dehydrogenase deficiency	ACADM	K329E
medium-chain acyl-CoA dehydrogenase deficiency	ACADM	K329E
medium-chain acyl-CoA dehydrogenase deficiency	ACADM	K329E
Meesmann epithelial corneal dystrophy	KRT12	L132P
metachromatic leukodystrophy	ARSA	P426L
metachromatic leukodystrophy	ARSA	c.459+1G>A
Morquio Syndrome (MPSIVA)	GALNS	R386C
Mucolipidosis IV	MCOLN1	406-2A>G
Mucolipidosis IV	MCOLN1	511_6943del
Neimann-Pick disease type A	SMPDI	L302P
Neuronal ceroid lipofuscinosis (NCL)	CLN2	R208*
neuronal ceroid lipofuscinosis 1	PPT1	R151*
Parkinson's disease	LRRK2	G2019S
Pendred syndrome	PDS	T461P
Pendred syndrome	PDS	L236P
Pendred syndrome	PDS	c.1001+1G>A
Pendred syndrome	PDS	IVS8, +1 G>A,
phenylketonuria	PAH	R408W
phenylketonuria	PAH	I65T
phenylketonuria	PAH	R261Q
phenylketonuria	PAH	IVS10-11G>A
phenylketonuria	PCDH15	R245*
phenylketonuria	PCDH15	R245*
Pompe disease	GAA	c.-32-13T>G
Primary ciliary dyskinesia	DNAI1	IVS1+2_3insT

Disease	Gene	Pathogenic Mutation[#]
Primary ciliary dyskinesia	DNAH5	10815delT
primary hypoxaluria	AGXT	G170R
Progressive familial intrahepatic cholestasis type 2	ABCB11	D482G (c.1445A>G)
Progressive familial intrahepatic cholestasis type 2	ABCB11	E297G
Propionic acidemia	PCCB; PCCA	c.1218_1231del14ins12
pseudoxanthoma elasticum	ABCC6	R1141*
Pyruvate kinase deficiency (PKD)	PKLR	c.1456c->T
retinitis pigmentos	USH2a	C759F
retinitis pigmentosa	IMPDHI	D226N
retinitis pigmentosa	PDE6A	V685M
retinitis pigmentosa	PDE6A	D670G
retinitis pigmentosa	PRPF3	T494M
retinitis pigmentosa	PRPF8	H2309R
retinitis pigmentosa	RHO	P23H
retinitis pigmentosa	RHO	P347L
retinitis pigmentosa	RHO	P347L
retinitis pigmentosa	RHO	D190N
retinitis pigmentosa	RPI	R667*
retinitis pigmentosa/Usher syndrome type 1C	USH1C	V72V
Rett syndrome	MECP2	R106W
Rett syndrome	MECP2	R133C
Rett syndrome	MECP2	R306C
Rett syndrome	MECP2	R168*
Rett syndrome	MECP2	R255*
Sanfilippo syndrome A (MPSIIIA)	SGSH	R74C
Sanfilippo syndrome A (MPSIIIA)	SGSH	R245H
Sanfilippo syndrome B (MPSIIIB)	NAGLU	R297*
Sanfilippo syndrome B (MPSIIIB)	NAGLU	Y140C
severe combined immunodeficiency	ADA	G216R
severe combined immunodeficiency	ADA	G216R
severe combined immunodeficiency	ADA	Q3*
sickle cell disease	HBB	E6V
sickle cell disease	HBB	E6V
sickle cell disease	HBB	E6V
sickle cell disease	HBB	E26K
sickle cell disease	HBB	E26K
sickle cell disease	HBB	E7K
sickle cell disease	HBB	c.-138C>T
sickle cell disease	HBB	IVS2
sickle cell disease	HBB	654 C>T
Sly Syndrome (MPSVII)	GUSB	L175F
Stargardt disease	ABCA4	A1038V

Disease	Gene	Pathogenic Mutation[#]
Stargardt disease	ABCA4	A1038V
Stargardt disease	ABCA4	L541P
Stargardt disease	ABCA4	G1961E
Stargardt disease	ABCA4	G1961E
Stargardt disease	ABCA4	G1961E
Stargardt disease	ABCA4	G1961E
Stargardt disease	ABCA4	c.2588G>C
Stargardt disease	ABCA4	c.5461-10 T>C
Stargardt disease	ABCA4	c.5714+5G>A
Tay Sachs	HEXA	InsTATC1278
tyrosinemia type 1	FAH	P261L
Usher syndrome type 1F	PCDH15	R245*
variegate porphyria	PPOX	R59W
VCP myopathy (IBMPFD) 1	VCP	R1555X
von Gierke disease	G6PC	Q347*
von Gierke disease	G6PC	Q347*
von Gierke disease	G6PC	Q347*
von Gierke disease	G6PC	R83C
Wilson's Disease	ATP7B	E297G
X-linked myotubular myopathy	MTMI	c.1261-10A>G
X-linked retinoschisis	RS1	R102W
X-linked retinoschisis	RS1	R141C

[#]: See J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, herein incorporated by reference in its entirety, for details of the nomenclatures of gene mutations. * means a stop codon.

Compensatory edits

In some embodiments, the systems or methods provided herein can be used to introduce a compensatory edit. In some embodiments, the compensatory edit is at a position of a gene associated with a disease or disorder, which is different from the position of a disease-causing mutation. In some embodiments, the compensatory mutation is not in the gene containing the causitive mutation. In some embodiments, the compensatory edit can negate or compensate for a disease-causing mutation. In some embodiments, the compensatory edit can be introduced by the systems or methods provided herein to suppress or reverse the mutant effect of a disease-causing mutation.

Table 24 provides exemplary indications (column 1), genes (column 2), and compensatory edits that can be introduced using the systems or methods described herein

(column 3). In some embodiments, the compensatory edits provided in Table 24 can be introduced to suppress or reverse the mutant effect of a disease-causing mutation.

Table 24. Indications, genes, compensatory edits, and exemplary design features.

Disease	**Gene**	**Nucleotide Change#**
Alpha-1 Antitrypsin Deficiency	SERPINAI	F51L
Alpha-1 Antitrypsin Deficiency	SERPINAI	M374I
Alpha-1 Antitrypsin Deficiency	SERPINAI	A348V/A347V
Alpha-1 Antitrypsin Deficiency	SERPINAI	K387R
Alpha-1 Antitrypsin Deficiency	SERPINAI	T59A
Alpha-1 Antitrypsin Deficiency	SERPINAI	T68A
ATTR amyloidosis	TTR	Al08V
ATTR amyloidosis	TTR	Rl04H
ATTR amyloidosis	TTR	T119M
Cystic fibrosis	CFTR	R555K
Cystic fibrosis	CFTR	F409L
Cystic fibrosis	CFTR	F433L
Cystic fibrosis	CFTR	H667R
Cystic fibrosis	CFTR	Rl070W
Cystic fibrosis	CFTR	R29K
Cystic fibrosis	CFTR	R553Q
Cystic fibrosis	CFTR	1539T
Cystic fibrosis	CFTR	G550E
Cystic fibrosis	CFTR	F429S
Cystic fibrosis	CFTR	Q637R
Sickle cell disease	HBB	A70T
Sickle cell disease	HBB	A70V
Sickle cell disease	HBB	L88P
Sickle cell disease	HBB	F85L and/or F85P
Sickle cell disease	HBB	E22G
Sickle cell disease	HBB	G16D and/or G16N

#: See J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, herein incorporated by reference in its entirety, for details of the nomenclatures of gene mutations.

Regulatory edits

In some embodiments, the systems or methods provided herein can be used to introduce a regulatory edit. In some embodiments, the regulatory edit is introduced to a regulatory sequence of a gene, for example, a gene promoter, gene enhancer, gene repressor, or a sequence that regulates gene splicing. In some embodiments, the regulatory edit increases or decreases the expression level of a target gene. In some embodiments, the target gene is the same as the gene containing a disease-causing mutation. In some embodiment, the target gene is different from the gene containing a disease-causing mutation. For example, the systems or methods provided herein can be used to upregulate the expression of fetal hemoglobin by introducing a regulatory edit at the promoter of bcl11a, thereby treating sickle cell disease.

Table 25 provides exemplary indications (column 1), genes (column 2), and regulatory edits that can be introduced using the systems or methods described herein (column 3).

Table 25. Indications, genes, and compensatory regulatory edits.

Disease	**Gene**	**Nucleotide Change#**
homozygous familial hypercholesterolaemia	LDLR	c.81C>T
Porphyrias	ALAS1	c.3G>A
Porphyrias	ALAS1	c.2T>C
Porphyrias	ALAS1	c.46C>T
Porphyrias	ALAS1	c.91C>T
Porphyrias	ALAS1	c.91C>T
Porphyrias	ALAS1	c.226C>T
Porphyrias	ALAS1	c.226C>T
Porphyrias	ALAS1	c.226C>T
Porphyrias	ALAS1	c.229C>T
Porphyrias	ALAS1	c.247C>T
Porphyrias	ALAS1	c.247C>T
Porphyrias	ALAS1	c.250C>T
Porphyrias	ALAS1	c.250C>T
Porphyrias	ALAS1	c.340C>T
Porphyrias	ALAS1	c.340C>T
Porphyrias	ALAS1	c.349C>T
Porphyrias	ALAS1	c.391C>T
Porphyrias	ALAS1	c.391C>T
Porphyrias	ALAS1	c.403C>T

Disease	Gene	Nucleotide Change#
Porphyrias	ALAS1	c.403C>T
Porphyrias	ALAS1	c.199+1G>A
Porphyrias	ALAS1	c.199+1G>A
Porphyrias	ALAS1	c.199+1G>A
Porphyrias	ALAS1	c.199+1G>A
Porphyrias	ALAS1	c.199+2T>C
Porphyrias	ALAS1	c.199+2T>C
Porphyrias	ALAS1	c.199+2T>C
Porphyrias	ALAS1	c.199+2T>C
Porphyrias	ALAS1	c.200-2A>G
Porphyrias	ALAS1	c.427+ 1G>A
Porphyrias	ALAS1	c.427+2T>C
Porphyrias	ALAS1	c.1165+1G >A
Porphyrias	ALAS1	c. l 165+2T>C
Porphyrias	ALAS1	c.1166-lA>G
Porphyrias	ALAS1	c.133 l-2A>G
sickle cell disease	BCL11A	c.386-24278G>A
sickle cell disease	BCL11A	c.386-24983T>C
sickle cell disease	HBG1	c.-167C>T
sickle cell disease	HBG1	c.-170G>A
sickle cell disease	HBG1	c.-249C>T
sickle cell disease	HBG2	c.-211C>T
sickle cell disease	HBG2	c.-228T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -198 T>C
sickle cell disease	HBG1/2	C. -175 T>C
sickle cell disease	HBG1/2	C. -175 T>C
sickle cell disease	HBG1/2	C. -175 T>C
sickle cell disease	HBG1/2	C. -175 T>C
sickle cell disease	HBG1/2	C. -175 T>C

Disease	Gene	Nucleotide Change[#]
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	C. -114~ -102 deletion
sickle cell disease	HBG1/2	c. -90 BCLlIA Binding
sickle cell disease	HBG1/2	c. -90 BCLlIA Binding
sickle cell disease	HBG1/2	C. -202 C>T, -201 C>T, -198 T>C, -197 C>T, -196 C>T, -195 C>G
sickle cell disease	HBG1/2	C. -197 C>T,-196 C>T, -195 C>G

[#]: See J T den Dunnen and S E Antonarakis, Hum Mutat. 2000;15(1):7-12, herein incorporated by reference in its entirety, for details of the nomenclatures of gene mutations.

Repeat expansion diseases

In some embodiments, the systems or methods provided herein can be used to a repeat expansion disease, for example, a repeat expansion disease provided in **Table 26**. Table 26 provides the indication (column 1), the gene (column 2), minimal repeat sequence of the repeat that is expanded in the condition (column 3), and the location of the repeat relative to the listed gene for each indication (column 4). In some embodiments, the systems or methods provided herein, for example, those comprising Gene Writers, can be used to treat repeat expansion diseases by resetting the number of repeats at the locus according to a customized RNA template (see, e.g., Example 24).

Table 26. Exemplary repeat expansion diseases, genes, causal repeats, and repeat locations.

Disease	Gene	Causal repeat	Repeat location
myotonic dystrophy 1	DMPK/DM1	CTG	3' UTR

Disease	Gene	Causal repeat	Repeat location
myotonic dystrophy 2	ZNF9/CNBP	CCTG	Intron 1
dentatorubral-pallidoluysian atrophy	ATN1	CAG	Coding
fragile X mental retardation syndrome	FMR1	CGG	5' UTR
fragile X E mental retardation	FMR2	GCC	5' UTR
Friedreich's ataxia	FXN	GAA	Intron
fragile X tremor ataxia syndrome	FMR1	CGG	5' UTR
Huntington's disease	HTT	CAG	Coding
Huntington's disease-like 2	JPH3	CTG	3' UTR, coding
myoclonic epilepsy of Unverricht and Lundborg	CSTB	CCCCGCCCCGCG	Promoter
oculopharyngeal muscular dystrophy	PABPN1	GCG	Coding
spinal and bulbar muscular atrophy	AR	CAG	Coding
spinocerebellar ataxia 1	ATXN1	CAG	Coding
spinocerebellar ataxia 2	ATXN2	CAG	Coding
spinocerebellar ataxia 3	ATXN3	CAG	Coding
spinocerebellar ataxia 6	CACNA1A	CAG	Coding
spinocerebellar ataxia 7	ATXN7	CAG	Coding
spinocerebellar ataxia 8	ATXN8	CTG/CAG	CTG/CAG (ATXN8)
spinocerebellar ataxia 10	ATXN10	ATTCT	Intron
spinocerebellar ataxia 12	PPP2R2B	CAG	Promoter, 5' UTR?
spinocerebellar ataxia 17	TBP	CAG	Coding
Syndromic/non-syndromic X-linked mental retardation	ARX	GCG	Coding

Exemplary templates

In some embodiments, the systems or methods provided herein use the template sequences listed in **Table 27**. Table 27 provides exemplary template RNA sequences (column 5) and optional second-nick gRNA sequences (column 6) designed to be paired with a Gene Writing polypeptide to correct the indicated pathogenic mutations (column 4). All the templates in Table 27 are meant to exemplify the total sequence of: (1) targeting gRNA for first strand nick, (2) polypeptide binding domain, (3) heterologous object sequence, and (4) target homology domain for setting up TPRT at first strand nick.

Table 27. Exemplary diseases, tissues, genes, pathogenic mutations, template RNA sequences, and second nick gRNA sequences.

Disease	Tissue	Gene	Mutation	Template RNA	Second nick gRNA
Alpha-1 antitrypsin	Liver	SERPINA1	PiZ	TCCCCTCCAGGCCGTGCATAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTcGTCGATGGTCAGCACAGCCTTATGCACGGCCTGGA	TTTGTTGAACTTGACCTCGG
Cystic fibrosis	Lung	CFTR	deltaF508	ACCATTAAAGAAAATATCATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACCAaagATGATATTTTCTTTA	AaagATGATATTTTCTTAA
Sickle cell	HSC	HBB	HbS	GTAACGGCAGACTTCTCCACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGACTCCTGaGGAGAAGTCTGCC	TGGTGAGGCCCTGGGCAGGT
Wilson's Disease	Liver	ATP7B	H1069Q	TTGGTGACTGCCACGCCCAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACAcCCCTTGGGCGTGGCAGTC	GGCCAGCAGTGAACAcCCCT
ARVC	Heart	PKP2	235C>T	ACTCAGGAACACTGCTGGTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTTCACtGAACCAGCAGTGTTCC	TTGGTTGAAAATGATTTTGT
Long QT syndrome 1	Heart	KCNQ1	P343S	CCAGGGAAAACGCACCCACGGTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCcCAGCGGTAGGTGCCCCGTGGGTGCGTTTTC	CTCCTTCTTTGCGCTCcCAG
Mucolipidosis IV	CNS	MCOLN1	406-2A>G	GCCCTCCCCTTCTCTGCCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGT	TCAGGCAACGCCAGGTACtG

Disease	Tissue	Gene	Mutation	Template RNA	Second nick gRNA
				CGGTCCGGTACtGTGGGCAGAGA AGGGG	

In some embodiments, the systems or methods provided herein use the template sequences listed in **Table 35**. Table 35 provides exemplary template RNA sequences (column 5) and optional second-nick gRNA sequences (column 6) designed to be paired with a Gene Writing polypeptide to correct the indicated pathogenic mutations (column 4). All the templates in Table 35 are meant to exemplify the total sequence of: (1) targeting gRNA for first strand nick, (2) polypeptide binding domain, (3) heterologous object sequence, and (4) target homology domain for setting up TPRT at first strand nick.

Table 35. Exemplary Gene Writing templates and second nick gRNA sequences for the correction of exemplary repeat expansion diseases. The region of the template spanning the repeat(s) is indicated in lowercase.

Disease	Gene	Reference Accession	Repeat	Location	Template RNA	Second-nick gRNA
myotonic dystrophy 1	DMPK	NC_000019.10 (45769709..45782490, complement)	CTG	3' UTR	CTCGAAGGGTC CTTGTAGCCGTT TTAGAGCTAGA AATAGCAAGTT AAAATAAGGCT AGTCCGTTATCA ACTTGAAAAAG TGGGACCGAGT CGGTCCGTGAT CCCCCcagcagcagc agcagcagcagcagcag cagcagcagcagcagca gcagcagcagcagcag CATTCCCGGCTA CAAGGACCCT	ATCA CAGG ACTG GAGC TGGG

myotonic dystrophy 2	CNBP	NC_000003.12 (129167827..129183896, complement)	CCTG	Intron 1	ACCACTGCACTCCAGCCTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGcctgcctgcctgcctgcctgcctgcctgcctggctgcctgtctgcctgtctgcctgcctgcctgcctgcctgcctgcctgTCTGTCTCACTTTGTCCCCTAGGCTGGAGTGCA	GCCTCAGCCTCCTGAGTAGC
fragile X mental retardation syndrome	FMR1	NC_000023.11 (147911919..147951127)	CGG	5' UTR	GGGGGCGTGCGGCAGCGCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTGCGGGCGCTCGAGGCCCAGccgccgccgccgccgccgccgccgccgcctccgccgccgccgccgccgccgccgccgccgCGCTGCCGCACG	GCTCAGAGGCGGCCCTCCAC

Friedreich's ataxia	FXN	NC_000009.12 (69035752..69079076)	GAA	Intron	CAGGCGCGCGACACCACGCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCAACCCAGTATCTACTAAAAAATACAAAAAAAAAAAAAAgaagaagaagaagaagaaAATAAAGAAAAGTTAGCCGGGCGTGGTGTCGCGC	CGCTTGAGCCCGGGAGGCAG
Huntington disease	HTT	NC_000004.12 (3074681..3243960)	CAG	Coding	GGCGGCTGAGGAAGCTGAGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCAGTCCCTCAAGTCCTTCcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcaacagccgccaccgccgccgccgccgccgccgcctcctCAGCTTCCTCAG	CGCTGCACCGACCGTGAGTT

spinocerebellar ataxia 1	ATXN1	NC_000006.12 (16299112..16761490, complement)	CAG	Coding	TGAGCCCCGGAGCCCTGCTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCACAAGGCTGAGcagcagcagcagcagcagcagcagcagcagcagcagcatcagcatcagcagcagcagcagcagcagcagcagcagcagcagcagcagCACCTCAGCAGGGCTCCGGG	TCCAGTTCTCCGCAGAACAC

Exemplary heterologous object sequences

In some embodiments, the systems or methods provided herein comprise a heterologous object sequence, wherein the heterologous object sequence or a reverse complementary sequence thereof, encodes a protein (e.g., an antibody) or peptide. In some embodiments, the therapy is one approved by a regulatory agency such as FDA.

In some embodiments, the protein or peptide is a protein or peptide from the THPdb database (Usmani et al. PLoS One 12(7):e0181748 (2017), herein incorporated by reference in its entirety. In some embodiments, the protein or peptide is a protein or peptide disclosed in **Table 28**. In some embodiments, the systems or methods disclosed herein, for example, those comprising Gene Writers, may be used to integrate an expression cassette for a protein or peptide from Table 28 into a host cell to enable the expression of the protein or peptide in the host. In some embodiments, the sequences of the protein or peptide in the first column of Table 28 can be found in the patents or applications provided in the third column of Table 28, incorporated by reference in their entireties.

In some embodiments, the protein or peptide is an antibody disclosed in Table 1 of Lu et al. J Biomed Sci 27(1):1 (2020), herein incorporated by reference in its entirety. In some embodiments, the protein or peptide is an antibody disclosed in **Table 29**. In some embodiments, the systems or methods disclosed herein, for example, those comprising Gene

Writers, may be used to integrate an expression cassette for an antibody from Table 29 into a host cell to enable the expression of the antibody in the host. In some embodiments, a system or method described herein is used to express an agent that binds a target of column 2 of Table 29 (e.g., a monoclonal antibody of column 1 of Table 29) in a subject having an indication of column 3 of Table 29.

Table 28. Exemplary protein and peptide therapeutics.

Therapeutic peptide	**Category**	**Patent Number**
Lepirudin	Antithrombins and Fibrinolytic Agents	CA1339104
Cetuximab	Antineoplastic Agents	CA1340417
Dor se alpha	Enzymes	CA2184581
Denileukin diftitox	Antineoplastic Agents	
Etanercept	Immunosuppressive Agents	CA2476934
Bivalirudin	Antithrombins	US7582727
Leuprolide	Antineoplastic Agents	
Peginterferon alpha-2a	Immunosuppressive Agents	CA2203480
Alteplase	Thrombolytic Agents	
Interferon alpha-n1	Antiviral Agents	
Darbepoetin alpha	Anti-anemic Agents	CA2165694
Reteplase	Fibrinolytic Agents	CA2107476
Epoetin alpha	Hematinics	CA1339047
Salmon Calcitonin	Bone Density Conservation Agents	US6440392
Interferon alpha-n3	Immunosuppressive Agents	
Pegfilgrastim	Immunosuppressive Agents	CA1341537
Sargramostim	Immunosuppressive Agents	CA1341150
Secretin	Diagnostic Agents	
Peginterferon alpha-2b	Immunosuppressive Agents	CA1341567
Asparagi se	Antineoplastic Agents	
Thyrotropin alpha	Diagnostic Agents	US5840566
Antihemophilic Factor	Coagulants and Thrombotic agents	CA2124690
A kinra	Antirheumatic Agents	CA2141953
Gramicidin D	Anti-Bacterial Agents	
Intravenous Immunoglobulin	Immunologic Factors	
Anistreplase	Fibrinolytic Agents	
Insulin Regular	Antidiabetic Agents	
Tenecteplase	Fibrinolytic Agents	CA2129660
Menotropins	Fertility Agents	
Interferon gamma-1b	Immunosuppressive Agents	US6936695
Interferon alpha-2a, Recombi nt		CA2172664

Therapeutic peptide	Category	Patent Number
Coagulation factor VIIa	Coagulants	
Oprelvekin	Antineoplastic Agents	
Palifermin	Anti-Mucositis Agents	
Glucagon recombi nt	Hypoglycemic Agents	
Aldesleukin	Antineoplastic Agents	
Botulinum Toxin Type B	Antidystonic Agents	
Omalizumab	Anti-Allergic Agents	CA2113813
Lutropin alpha	Fertility Agents	US5767251
Insulin Lispro	Hypoglycemic Agents	US5474978
Insulin Glargine	Hypoglycemic Agents	US7476652
Collage se		
Rasburicase	Gout Suppressants	CA2175971
Adalimumab	Antirheumatic Agents	CA2243459
Imiglucerase	Enzyme Replacement Agents	US5549892
Abciximab	Anticoagulants	CA1341357
Alpha-1-protei se inhibitor	Serine Protei se Inhibitors	
Pegaspargase	Antineoplastic Agents	
Interferon beta-1a	Antineoplastic Agents	CA1341604
Pegademase bovine	Enzyme Replacement Agents	
Human Serum Albumin	Serum substitutes	US6723303
Eptifibatide	Platelet Aggregation Inhibitors	US6706681
Serum albumin iodo ted	Diagnostic Agents	
Infliximab	Antirheumatic Agents, Anti-Inflammatory Agents, Non-Steroidal, Dermatologic Agents, Gastrointesti l Agents and Immunosuppressive Agents	CA2106299
Follitropin beta	Fertility Agents	US7741268
Vasopressin	Antidiuretic Agents	
Interferon beta-1b	Adjuvants, Immunologic and Immunosuppressive Agents	CA1340861
Interferon alphacon-1	Antiviral Agents and Immunosuppressive Agents	CA1341567
Hyaluronidase	Adjuvants, Anesthesia and Permeabilizing Agents	
Insulin, porcine	Hypoglycemic Agents	
Trastuzumab	Antineoplastic Agents	CA2103059
Rituximab	Antineoplastic Agents, Immunologic Factors and Antirheumatic Agents	CA2149329
Basiliximab	Immunosuppressive Agents	CA2038279
Muromo b	Immunologic Factors and Immunosuppressive Agents	

Therapeutic peptide	Category	Patent Number
Digoxin Immune Fab (Ovine)	Antidotes	
Ibritumomab		CA2149329
Daptomycin		US6468967
Tositumomab		
Pegvisomant	Hormone Replacement Agents	US5849535
Botulinum Toxin Type A	Neuromuscular Blocking Agents, Anti-Wrinkle Agents and Antidystonic Agents	CA2280565
Pancrelipase	Gastrointesti l Agents and Enzyme Replacement Agents	
Streptoki se	Fibrinolytic Agents and Thrombolytic Agents	
Alemtuzumab		CA1339198
Alglucerase	Enzyme Replacement Agents	
Capromab	Indicators, Reagents and Diagnostic Agents	
Laronidase	Enzyme Replacement Agents	
Urofollitropin	Fertility Agents	US5767067
Efalizumab	Immunosuppressive Agents	
Serum albumin	Serum substitutes	US6723303
Choriogo dotropin alpha	Fertility Agents and Go dotropins	US6706681
Antithymocyte globulin	Immunologic Factors and Immunosuppressive Agents	
Filgrastim	Immunosuppressive Agents, Antineutropenic Agents and Hematopoietic Agents	CA1341537
Coagulation factor ix	Coagulants and Thrombotic Agents	
Becaplermin	Angiogenesis Inducing Agents	CA1340846
Agalsidase beta	Enzyme Replacement Agents	CA2265464
Interferon alpha-2b	Immunosuppressive Agents	CA1341567
Oxytocin	Oxytocics, Anti-tocolytic Agents and Labor Induction Agents	
Enfuvirtide	HIV Fusion Inhibitors	US6475491
Palivizumab	Antiviral Agents	CA2197684
Daclizumab	Immunosuppressive Agents	
Bevacizumab	Angiogenesis Inhibitors	CA2286330
Arcitumomab	Diagnostic Agents	US8420081
Arcitumomab	Diagnostic Agents	US7790142
Eculizumab		CA2189015
Panitumumab		
Ranibizumab	Ophthalmics	CA2286330
Idursulfase	Enzyme Replacement Agents	

Therapeutic peptide	Category	Patent Number
Alglucosidase alpha	Enzyme Replacement Agents	CA2416492
Exe tide	Hypoglycemic Agents	US6872700
Mecasermin		US5681814
Pramlintide		US5686411
Galsulfase	Enzyme Replacement Agents	
Abatacept	Antirheumatic Agents and Immunosuppressive Agents	CA2110518
Cosyntropin	Hormones and Diagnostic Agents	
Corticotropin		
Insulin aspart	Hypoglycemic Agents and Antidiabetic Agents	US5866538
Insulin detemir	Antidiabetic Agents	US5750497
Insulin glulisine	Antidiabetic Agents	US6960561
Pegaptanib	Intended for the prevention of respiratory distress syndrome (RDS) in premature infants at high risk for RDS.	
Nesiritide		
Thymalphasin		
Defibrotide	Antithrombins	
tural alpha interferon OR multiferon		
Glatiramer acetate		
Preotact		
Teicoplanin	Anti-Bacterial Agents	
Ca kinumab	Anti-Inflammatory Agents and Monoclo l antibodies	
Ipilimumab	Antineoplastic Agents and Monoclo l antibodies	CA2381770
Sulodexide	Antithrombins and Fibrinolytic Agents and Hypoglycemic Agents and Anticoagulants and Hypolipidemic Agents	
Tocilizumab		CA2201781
Teriparatide	Bone Density Conservation Agents	US6977077
Pertuzumab	Monoclo l antibodies	CA2376596
Rilo cept	Immunosuppressive Agents	US5844099
Denosumab	Bone Density Conservation Agents and Monoclo l antibodies	CA2257247
Liraglutide		US6268343
Golimumab	Antipsoriatic Agents and Monoclo l antibodies and TNF inhibitor	

Therapeutic peptide	Category	Patent Number
Belatacept	Antirheumatic Agents and Immunosuppressive Agents	
Buserelin		
Velaglucerase alpha	Enzymes	US7138262
Tesamorelin		US5861379
Brentuximab vedotin		
Taliglucerase alpha	Enzymes	
Belimumab	Monoclo l antibodies	
Aflibercept	Antineoplastic Agents and Ophthalmics	US7306799
Asparagi se erwinia chrysanthemi	Enzymes	
Ocriplasmin	Ophthalmics	
Glucarpidase	Enzymes	
Teduglutide		US5789379
Raxibacumab	Anti-Infective Agents and Monoclo l antibodies	
Certolizumab pegol	TNF inhibitor	CA2380298
Insulin,isophane	Hypoglycemic Agents and Antidiabetic Agents	
Epoetin zeta		
Obinutuzumab	Antineoplastic Agents	
Fibrinolysin aka plasmin		US3234106
Follitropin alpha		
Romiplostim	Colony-Stimulating Factors and Thrombopoietic Agents	
Luci ctant	Pulmo ry surfactants	US5407914
talizumab	Immunosuppressive agents	
Aliskiren	Renin inhibitor	
Ragweed Pollen Extract		
Secukinumab	Inhibitor	US20130202610
Somatotropin Recombi nt	Hormone Replacement Agents	CA1326439
Drotrecogin alpha	Antisepsis	CA2036894
Alefacept	Dermatologic and Immunosupressive agents	
OspA lipoprotein	Vaccines	
Uroki se		US4258030
Abarelix	Anti-Testosterone Agents	US5968895
Sermorelin	Hormone Replacement Agents	
Aprotinin		US5198534
Gemtuzumab ozogamicin	Antineoplastic agents and Immunotoxins	US5585089
Satumomab Pendetide	Diagnostic Agents	

Therapeutic peptide	Category	Patent Number
Albiglutide	Drugs used in diabetes; alimentary tract and metabolism; blood glucose lowering drugs, excl. insulins.	
Alirocumab		
Ancestim		
Antithrombin alpha		
Antithrombin III human		
Asfotase alpha	Enzymes Alimentary Tract and Metabolism	
Atezolizumab		
Autologous cultured chondrocytes		
Beractant		
Bli tumomab	Antineoplastic Agents Immunosuppressive Agents Monoclo l antibodies Antineoplastic and Immunomodulating Agents	US20120328618
C1 Esterase Inhibitor (Human)		
Coagulation Factor XIII A-Subunit (Recombi nt)		
Conestat alpha		
Daratumumab	Antineoplastic Agents	
Desirudin		
Dulaglutide	Hypoglycemic Agents; Drugs Used in Diabetes; Alimentary Tract and Metabolism; Blood Glucose Lowering Drugs, Excl. Insulins	
Elosulfase alpha	Enzymes; Alimentary Tract and Metabolism	
Elotuzumab		US2014055370
Evolocumab	Lipid Modifying Agents, Plain; Cardiovascular System	
Fibrinogen Concentrate (Human)		
Filgrastim-sndz		
Gastric intrinsic factor		
Hepatitis B immune globulin		
Human calcitonin		

Therapeutic peptide	Category	Patent Number
Human clostridium tetani toxoid immune globulin		
Human rabies virus immune globulin		
Human Rho(D) immune globulin		
Hyaluronidase (Human Recombi nt)		US7767429
Idarucizumab	Anticoagulant	
Immune Globulin Human	Immunologic Factors; Immunosuppressive Agents; Anti-Infective Agents	
Vedolizumab	Immunosupressive agent, Antineoplastic agent	US2012151248
Ustekinumab	Deramtologic agent, Immunosuppressive agent, antineoplastic agent	
Turoctocog alpha		
Tuberculin Purified Protein Derivative		
Simoctocog alpha	Antihaemorrhagics: blood coagulation factor VIII	
Siltuximab	Antineoplastic and Immunomodulating Agents, Immunosuppressive Agents	US7612182
Sebelipase alpha	Enzymes	
Sacrosidase	Enzymes	
Ramucirumab	Antineoplastic and Immunomodulating Agents	US2013067098
Prothrombin complex concentrate		
Poractant alpha	Pulmo ry Surfactants	
Pembrolizumab	Antineoplastic and Immunomodulating Agents	US2012135408
Peginterferon beta-1a		
Ofatumumab	Antineoplastic and Immunomodulating Agents	US8337847
Obiltoxaximab		
Nivolumab	Antineoplastic and Immunomodulating Agents	US2013173223
Necitumumab		
Metreleptin		US20070099836
Methoxy polyethylene glycol-epoetin beta		

Therapeutic peptide	Category	Patent Number
Mepolizumab	Antineoplastic and Immunomodulating Agents, Immunosuppressive Agents, Interleukin Inhibitors	US2008134721
Ixekizumab		
Insulin Pork	Hypoglycemic Agents, Antidiabetic Agents	
Insulin Degludec		
Insulin Beef		
Thyroglobulin	Hormone therapy	US5099001
Anthrax immune globulin human	Plasma derivative	
Anti-inhibitor coagulant complex	Blood Coagulation Factors, Antihemophilic Agent	
Anti-thymocyte Globulin (Equine)	Antibody	
Anti-thymocyte Globulin (Rabbit)	Antibody	
Brodalumab	Antineoplastic and Immunomodulating Agents	
C1 Esterase Inhibitor (Recombi nt)	Blood and Blood Forming Organs	
Ca kinumab	Antineoplastic and Immunomodulating Agents	
Chorionic Go dotropin (Human)	Hormones	US6706681
Chorionic Go dotropin (Recombi nt)	Hormones	US5767251
Coagulation factor X human	Blood Coagulation Factors	
Dinutuximab	Antibody, Immunosuppresive agent, Antineoplastic agent	US20140170155
Efmoroctocog alpha	Antihemophilic Factor	
Factor IX Complex (Human)	Antihemophilic agent	
Hepatitis A Vaccine	Vaccine	
Human Varicella-Zoster Immune Globulin	Antibody	
Ibritumomab tiuxetan	Antibody, Immunosuppressive Agents	CA2149329
Lenograstim	Antineoplastic and Immunomodulating Agents	
Pegloticase	Enzymes	

Therapeutic peptide	Category	Patent Number
Protamine sulfate	Heparin Antagonists, Hematologic Agents	
Protein S human	Anticoagulant plasma protein	
Sipuleucel-T	Antineoplastic and Immunomodulating Agents	US8153120
Somatropin recombi nt	Hormones, Hormone Substitutes, and Hormone Antagonists	CA1326439,CA2252535 ,US5288703,US5849700,U S5849704,US5898030 ,US6004297,US6152897 ,US6235004 ,US6899699
Susoctocog alpha	Blood coagulation factors, Antihaemorrhagics	
Thrombomodulin alpha	Anticoagulant agent, Antiplatelet agent	

Table 29. Exemplary monoclonal antibody therapies.

mAb	Target	Indication
Muromonab-CD3	CD3	Kidney transplant rejection
Abciximab	GPIIb/IIIa	Prevention of blood clots in angioplasty
Rituximab	CD20	Non-Hodgkin lymphoma
Palivizumab	RSV	Prevention of respiratory syncytial virus infection
Infliximab	TNFα	Crohn's disease
Trastuzumab	HER2	Breast cancer
Alemtuzumab	CD52	Chronic myeloid leukemia
Adalimumab	TNFα	Rheumatoid arthritis
Ibritumomab tiuxetan	CD20	Non-Hodgkin lymphoma
Omalizumab	IgE	Asthma
Cetuximab	EGFR	Colorectal cancer
Bevacizumab	VEGF-A	Colorectal cancer
Natalizumab	ITGA4	Multiple sclerosis
Panitumumab	EGFR	Colorectal cancer
Ranibizumab	VEGF-A	Macular degeneration
Eculizumab	C5	Paroxysmal nocturnal hemoglobinuria
Certolizumab pegol	TNFα	Crohn's disease
Ustekinumab	IL-12/23	Psoriasis
Canakinumab	IL-1β	Muckle-Wells syndrome
Golimumab	TNFα	Rheumatoid and psoriatic arthritis, ankylosing spondylitis
Ofatumumab	CD20	Chronic lymphocytic leukemia
Tocilizumab	IL-6R	Rheumatoid arthritis
Denosumab	RANKL	Bone loss

Belimumab	BLyS	Systemic lupus erythematosus
Ipilimumab	CTLA-4	Metastatic melanoma
Brentuximab vedotin	CD30	Hodgkin lymphoma, systemic anaplastic large cell lymphoma
Pertuzumab	HER2	Breast Cancer
Trastuzumab emtansine	HER2	Breast cancer
Raxibacumab	*B. anthrasis* PA	Anthrax infection
Obinutuzumab	CD20	Chronic lymphocytic leukemia
Siltuximab	IL-6	Castleman disease
Ramucirumab	VEGFR2	Gastric cancer
Vedolizumab	α4β7 integrin	Ulcerative colitis, Crohn disease
Blinatumomab	CD19, CD3	Acute lymphoblastic leukemia
Nivolumab	PD-1	Melanoma, non-small cell lung cancer
Pembrolizumab	PD-1	Melanoma
Idarucizumab	Dabigatran	Reversal of dabigatran-induced anticoagulation
Necitumumab	EGFR	Non-small cell lung cancer
Dinutuximab	GD2	Neuroblastoma
Secukinumab	IL-17α	Psoriasis
Mepolizumab	IL-5	Severe eosinophilic asthma
Alirocumab	PCSK9	High cholesterol
Evolocumab	PCSK9	High cholesterol
Daratumumab	CD38	Multiple myeloma
Elotuzumab	SLAMF7	Multiple myeloma
Ixekizumab	IL-17α	Psoriasis
Reslizumab	IL-5	Asthma
Olaratumab	PDGFRα	Soft tissue sarcoma
Bezlotoxumab	*Clostridium difficile* enterotoxin B	Prevention of *Clostridium difficile* infection recurrence
Atezolizumab	PD-L1	Bladder cancer
Obiltoxaximab	*B. anthrasis* PA	Prevention of inhalational anthrax
Inotuzumab ozogamicin	CD22	Acute lymphoblastic leukemia
Brodalumab	IL-17R	Plaque psoriasis
Guselkumab	IL-23 p19	Plaque psoriasis
Dupilumab	IL-4Rα	Atopic dermatitis
Sarilumab	IL-6R	Rheumatoid arthritis
Avelumab	PD-L1	Merkel cell carcinoma
Ocrelizumab	CD20	Multiple sclerosis
Emicizumab	Factor IXa, X	Hemophilia A
Benralizumab	IL-5Rα	Asthma
Gemtuzumab ozogamicin	CD33	Acute myeloid leukemia
Durvalumab	PD-L1	Bladder cancer
Burosumab	FGF23	X-linked hypophosphatemia

Lanadelumab	Plasma kallikrein	Hereditary angioedema attacks
Mogamulizumab	CCR4	Mycosis fungoides or Sézary syndrome
Erenumab	CGRPR	Migraine prevention
Galcanezumab	CGRP	Migraine prevention
Tildrakizumab	IL-23 p19	Plaque psoriasis
Cemiplimab	PD-1	Cutaneous squamous cell carcinoma
Emapalumab	IFNγ	Primary hemophagocytic lymphohistiocytosis
Fremanezumab	CGRP	Migraine prevention
Ibalizumab	CD4	HIV infection
Moxetumomab pasudodox	CD22	Hairy cell leukemia
Ravulizumab	C5	Paroxysmal nocturnal hemoglobinuria
Caplacizumab	von Willebrand factor	Acquired thrombotic thrombocytopenic purpura
Romosozumab	Sclerostin	Osteoporosis in postmenopausal women at increased risk of fracture
Risankizumab	IL-23 p19	Plaque psoriasis
Polatuzumab vedotin	CD79β	Diffuse large B-cell lymphoma
Brolucizumab	VEGF-A	Macular degeneration
Crizanlizumab	P-selectin	Sickle cell disease

Plant-modification Methods

Gene Writer systems described herein may be used to modify a plant or a plant part (e.g., leaves, roots, flowers, fruits, or seeds), e.g., to increase the fitness of a plant.

A. Delivery to a Plant

Provided herein are methods of delivering a Gene Writer system described herein to a plant. Included are methods for delivering a Gene Writer system to a plant by contacting the plant, or part thereof, with a Gene Writer system. The methods are useful for modifying the plant to, e.g., increase the fitness of a plant.

More specifically, in some embodiments, a nucleic acid described herein (e.g., a nucleic acid encoding a GeneWriter) may be encoded in a vector, e.g., inserted adjacent to a plant promoter, e.g., a maize ubiquitin promoter (ZmUBI) in a plant vector (e.g., pHUC411). In some embodiments, the nucleic acids described herein are introduced into a plant (e.g., japonica rice) or part of a plant (e.g., a callus of a plant) via agrobacteria. In some embodiments, the systems and methods described herein can be used in plants by replacing a plant gene (e.g., hygromycin phosphotransferase (HPT)) with a null allele (e.g., containing a base substitution at the start

codon). Systems and methods for modifying a plant genome are described in Xu et. al. Development of plant prime-editing systems for precise genome editing, 2020, Plant Communications.

In one aspect, provided herein is a method of increasing the fitness of a plant, the method including delivering to the plant the Gene Writer system described herein (e.g., in an effective amount and duration) to increase the fitness of the plant relative to an untreated plant (e.g., a plant that has not been delivered the Gene Writer system).

An increase in the fitness of the plant as a consequence of delivery of a Gene Writer system can manifest in a number of ways, e.g., thereby resulting in a better production of the plant, for example, an improved yield, improved vigor of the plant or quality of the harvested product from the plant, an improvement in pre- or post-harvest traits deemed desirable for agriculture or horticulture (e.g., taste, appearance, shelf life), or for an improvement of traits that otherwise benefit humans (e.g., decreased allergen production). An improved yield of a plant relates to an increase in the yield of a product (e.g., as measured by plant biomass, grain, seed or fruit yield, protein content, carbohydrate or oil content or leaf area) of the plant by a measurable amount over the yield of the same product of the plant produced under the same conditions, but without the application of the instant compositions or compared with application of conventional plant-modifying agents. For example, yield can be increased by at least about 0.5%, about 1%, about 2%, about 3%, about 4%, about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 100%, or more than 100%. In some instances, the method is effective to increase yield by about 2x-fold, 5x-fold, 10x-fold, 25x-fold, 50x-fold, 75x-fold, 100x-fold, or more than 100x-fold relative to an untreated plant. Yield can be expressed in terms of an amount by weight or volume of the plant or a product of the plant on some basis. The basis can be expressed in terms of time, growing area, weight of plants produced, or amount of a raw material used. For example, such methods may increase the yield of plant tissues including, but not limited to: seeds, fruits, kernels, bolls, tubers, roots, and leaves.

An increase in the fitness of a plant as a consequence of delivery of a Gene Writer system can also be measured by other means, such as an increase or improvement of the vigor rating, the stand (the number of plants per unit of area), plant height, stalk circumference, stalk length, leaf number, leaf size, plant canopy, visual appearance (such as greener leaf color), root rating, emergence, protein content, increased tillering, bigger leaves, more leaves, less dead basal leaves, stronger tillers, less fertilizer needed, less seeds needed, more productive tillers, earlier flowering, early grain or seed maturity, less plant verse (lodging), increased shoot growth, earlier germination, or any combination of these factors, by a measurable or noticeable amount over the same factor of the plant produced under the same conditions, but without the administration of the instant compositions or with application of conventional plant-modifying agents (e.g., plant-modifying agents delivered without PMPs).

Accordingly, provided herein is a method of modifying a plant, the method including delivering to the plant an effective amount of any of the Gene Writer systems provided herein, wherein the method modifies the plant and thereby introduces or increases a beneficial trait in the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant. In particular, the method may increase the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In some instances, the increase in plant fitness is an increase (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) in disease resistance, drought tolerance, heat tolerance, cold tolerance, salt tolerance, metal tolerance, herbicide tolerance, chemical tolerance, water use efficiency, nitrogen utilization, resistance to nitrogen stress, nitrogen fixation, pest resistance, herbivore resistance, pathogen resistance, yield, yield under water-limited conditions, vigor, growth, photosynthetic capability, nutrition, protein content, carbohydrate content, oil content, biomass, shoot length, root length, root architecture, seed weight, or amount of harvestable produce.

In some instances, the increase in fitness is an increase (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) in development, growth, yield, resistance to abiotic stressors, or resistance to biotic stressors. An abiotic stress refers to an environmental stress condition that a plant or a plant part is subjected to that includes, e.g., drought stress, salt stress, heat stress, cold stress, and low nutrient stress. A biotic stress refers to an environmental stress condition that a plant or plant part is subjected to that includes, e.g. nematode stress, insect herbivory stress, fungal pathogen stress, bacterial pathogen stress, or viral pathogen stress. The stress may be temporary, e.g. several hours, several days, several months, or permanent, e.g. for the life of the plant.

In some instances, the increase in plant fitness is an increase (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) in quality of products harvested from the plant. For example, the increase in plant fitness may be an improvement in commercially favorable features (e.g., taste or appearance) of a product harvested from the plant. In other instances, the increase in plant fitness is an increase in shelf-life of a product harvested from the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%).

Alternatively, the increase in fitness may be an alteration of a trait that is beneficial to human or animal health, such as a reduction in allergen production. For example, the increase in fitness may be a decrease (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) in production of an allergen (e.g., pollen) that stimulates an immune response in an animal (e.g., human).

The modification of the plant (e.g., increase in fitness) may arise from modification of one or more plant parts. For example, the plant can be modified by contacting leaf, seed, pollen, root, fruit, shoot, flower, cells, protoplasts, or tissue (e.g., meristematic tissue) of the plant. As such, in another aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting pollen of the plant with an effective amount of any of the plant-modifying compositions herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In yet another aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting a seed of the plant with an effective amount of any of the Gene Writer systems disclosed herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In another aspect, provided herein is a method including contacting a protoplast of the plant with an effective amount of any of the Gene Writer systems described herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In a further aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting a plant cell of the plant with an effective amount of any of the Gene Writer system described herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In another aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting meristematic tissue of the plant with an effective amount of any of the plant-modifying compositions herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

In another aspect, provided herein is a method of increasing the fitness of a plant, the method including contacting an embryo of the plant with an effective amount of any of the plant-modifying compositions herein, wherein the method increases the fitness of the plant (e.g., by about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more than 100%) relative to an untreated plant.

B. Application Methods

A plant described herein can be exposed to any of the Gene Writer system compositions described herein in any suitable manner that permits delivering or administering the composition to the plant. The Gene Writer system may be delivered either alone or in combination with other active (e.g., fertilizing agents) or inactive substances and may be applied by, for example, spraying, injection (e.g,. microinjection), through plants, pouring, dipping, in the form of concentrated liquids, gels, solutions, suspensions, sprays, powders, pellets, briquettes, bricks and the like, formulated to deliver an effective concentration of the plant-modifying composition. Amounts and locations for application of the compositions described herein are generally determined by the habitat of the plant, the lifecycle stage at which the plant can be targeted by the plant-modifying composition, the site where the application is to be made, and the physical and functional characteristics of the plant-modifying composition.

In some instances, the composition is sprayed directly onto a plant, e.g., crops, by e.g., backpack spraying, aerial spraying, crop spraying/dusting etc. In instances where the Gene Writer system is delivered to a plant, the plant receiving the Gene Writer system may be at any stage of plant growth. For example, formulated plant-modifying compositions can be applied as a seed-coating or root treatment in early stages of plant growth or as a total plant treatment at later stages of the crop cycle. In some instances, the plant-modifying composition may be applied as a topical agent to a plant.

Further, the Gene Writer system may be applied (e.g., in the soil in which a plant grows, or in the water that is used to water the plant) as a systemic agent that is absorbed and distributed through the tissues of a plant. In some instances, plants or food organisms may be genetically transformed to express the Gene Writer system.

Delayed or continuous release can also be accomplished by coating the Gene Writer system or a composition with the plant-modifying composition(s) with a dissolvable or bioerodable coating layer, such as gelatin, which coating dissolves or erodes in the environment of use, to then make the plant-modifying com Gene Writer system position available, or by dispersing the agent in a dissolvable or erodable matrix. Such continuous release and/or dispensing means devices may be advantageously employed to consistently maintain an effective concentration of one or more of the plant-modifying compositions described herein.

In some instances, the Gene Writer system is delivered to a part of the plant, e.g., a leaf, seed, pollen, root, fruit, shoot, or flower, or a tissue, cell, or protoplast thereof. In some instances, the Gene Writer system is delivered to a cell of the plant. In some instances, the Gene Writer system is delivered to a protoplast of the plant. In some instances, the Gene Writer system is delivered to a tissue of the plant. For example, the composition may be delivered to meristematic tissue of the plant (e.g., apical meristem, lateral meristem, or intercalary meristem). In some instances, the composition is delivered to permanent tissue of the plant (e.g., simple tissues (e.g., parenchyma, collenchyma, or sclerenchyma) or complex permanent tissue (e.g., xylem or phloem)). In some instances, the Gene Writer system is delivered to a plant embryo.

C. Plants

A variety of plants can be delivered to or treated with a Gene Writer system described herein. Plants that can be delivered a Gene Writer system (i.e., "treated") in accordance with the present methods include whole plants and parts thereof, including, but not limited to, shoot vegetative organs/structures (e.g., leaves, stems and tubers), roots, flowers and floral organs/structures (e.g., bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, cotyledons, and seed coat) and fruit (the mature ovary), plant tissue (e.g., vascular tissue, ground tissue, and the like) and cells (e.g., guard cells, egg cells, and the like), and progeny of same. Plant parts can further refer parts of the plant such as the shoot, root, stem, seeds, stipules, leaves, petals, flowers, ovules, bracts, branches, petioles, internodes, bark, pubescence, tillers, rhizomes, fronds, blades, pollen, stamen, and the like.

The class of plants that can be treated in a method disclosed herein includes the class of higher and lower plants, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, horsetails, psilophytes, lycophytes, bryophytes, and algae (e.g., multicellular or unicellular algae). Plants that can be treated in accordance with the present methods further include any vascular plant, for example monocotyledons or dicotyledons or gymnosperms, including, but not limited to alfalfa, apple, Arabidopsis, banana, barley, canola, castor bean, chrysanthemum, clover, cocoa, coffee, cotton, cottonseed, corn, crambe, cranberry, cucumber, dendrobium, dioscorea, eucalyptus, fescue, flax, gladiolus, liliacea, linseed, millet, muskmelon, mustard, oat, oil palm, oilseed rape, papaya, peanut, pineapple, ornamental plants, Phaseolus, potato, rapeseed, rice, rye, ryegrass, safflower, sesame, sorghum, soybean, sugarbeet, sugarcane, sunflower, strawberry, tobacco, tomato, turfgrass, wheat and vegetable crops such as

lettuce, celery, broccoli, cauliflower, cucurbits; fruit and nut trees, such as apple, pear, peach, orange, grapefruit, lemon, lime, almond, pecan, walnut, hazel; vines, such as grapes (e.g., a vineyard), kiwi, hops; fruit shrubs and brambles, such as raspberry, blackberry, gooseberry; forest trees, such as ash, pine, fir, maple, oak, chestnut, popular; with alfalfa, canola, castor bean, corn, cotton, crambe, flax, linseed, mustard, oil palm, oilseed rape, peanut, potato, rice, safflower, sesame, soybean, sugarbeet, sunflower, tobacco, tomato, and wheat. Plants that can be treated in accordance with the methods of the present invention include any crop plant, for example, forage crop, oilseed crop, grain crop, fruit crop, vegetable crop, fiber crop, spice crop, nut crop, turf crop, sugar crop, beverage crop, and forest crop. In certain instances, the crop plant that is treated in the method is a soybean plant. In other certain instances, the crop plant is wheat. In certain instances, the crop plant is corn. In certain instances, the crop plant is cotton. In certain instances, the crop plant is alfalfa. In certain instances, the crop plant is sugarbeet. In certain instances, the crop plant is rice. In certain instances, the crop plant is potato. In certain instances, the crop plant is tomato.

In certain instances, the plant is a crop. Examples of such crop plants include, but are not limited to, monocotyledonous and dicotyledonous plants including, but not limited to, fodder or forage legumes, ornamental plants, food crops, trees, or shrubs selected from Acer spp., Allium spp., Amaranthus spp., Ananas comosus, Apium graveolens, Arachis spp, Asparagus officinalis, Beta vulgaris, Brassica spp. (e.g., Brassica napus, Brassica rapa ssp. (canola, oilseed rape, turnip rape), Camellia sinensis, Canna indica, Cannabis saliva, Capsicum spp., Castanea spp., Cichorium endivia, Citrullus lanatus, Citrus spp., Cocos spp., Coffea spp., Coriandrum sativum, Corylus spp., Crataegus spp., Cucurbita spp., Cucumis spp., Daucus carota, Fagus spp., Ficus carica, Fragaria spp., Ginkgo biloba, Glycine spp. (e.g., Glycine max, Soja hispida or Soja max), Gossypium hirsutum, Helianthus spp. (e.g., Helianthus annuus), Hibiscus spp., Hordeum spp. (e.g., Hordeum vulgare), Ipomoea batatas, Juglans spp., Lactuca sativa, Linum usitatissimum, Litchi chinensis, Lotus spp., Luffa acutangula, Lupinus spp., Lycopersicon spp. (e.g., Lycopersicon esculenturn, Lycopersicon lycopersicum, Lycopersicon pyriforme), Malus spp., Medicago sativa, Mentha spp., Miscanthus sinensis, Morus nigra, Musa spp., Nicotiana spp., Olea spp., Oryza spp. (e.g., Oryza sativa, Oryza latifolia), Panicum miliaceum, Panicum virgatum, Passiflora edulis, Petroselinum crispum, Phaseolus spp., Pinus spp., Pistacia vera, Pisum spp., Poa spp., Populus spp., Prunus spp., Pyrus communis, Quercus spp., Raphanus

sativus, Rheum rhabarbarum, Ribes spp., Ricinus communis, Rubus spp., Saccharum spp., Salix sp., Sambucus spp., Secale cereale, Sesamum spp., Sinapis spp., Solanum spp. (e.g., Solanum tuberosum, Solanum integrifolium or Solanum lycopersicum), Sorghum bicolor, Sorghum halepense, Spinacia spp., Tamarindus indica, Theobroma cacao, Trifolium spp., Triticosecale rimpaui, Triticum spp. (e.g., Triticum aestivum, Triticum durum, Triticum turgidum, Triticum hybernum, Triticum macha, Triticum sativum or Triticum vulgare), Vaccinium spp., Vicia spp., Vigna spp., Viola odorata, Vitis spp., and Zea mays. In certain embodiments, the crop plant is rice, oilseed rape, canola, soybean, corn (maize), cotton, sugarcane, alfalfa, sorghum, or wheat.

The plant or plant part for use in the present invention include plants of any stage of plant development. In certain instances, the delivery can occur during the stages of germination, seedling growth, vegetative growth, and reproductive growth. In certain instances, delivery to the plant occurs during vegetative and reproductive growth stages. In some instances, the composition is delivered to pollen of the plant. In some instances, the composition is delivered to a seed of the plant. In some instances, the composition is delivered to a protoplast of the plant. In some instances, the composition is delivered to a tissue of the plant. For example, the composition may be delivered to meristematic tissue of the plant (e.g., apical meristem, lateral meristem, or intercalary meristem). In some instances, the composition is delivered to permanent tissue of the plant (e.g., simple tissues (e.g., parenchyma, collenchyma, or sclerenchyma) or complex permanent tissue (e.g., xylem or phloem)). In some instances, the composition is delivered to a plant embryo. In some instances, the composition is delivered to a plant cell. The stages of vegetative and reproductive growth are also referred to herein as “adult” or “mature” plants.

In instances where the Gene Writer system is delivered to a plant part, the plant part may be modified by the plant-modifying agent. Alternatively, the Gene Writer system may be distributed to other parts of the plant (e.g., by the plant’s circulatory system) that are subsequently modified by the plant-modifying agent.

Table X. Exemplary nucleic acid sequences comprising coding sequences for retrotransposons from different types of sources, and 5' UTR and/or 3' UTR sequences flanking same *(following pages)*

Name	Accession	Organism	Clade	NID
5S-Sauria	.	Anolis carolinensis	SINE3/5S	
AARA8_AG	.	Anopheles gambiae	I	
Aca-Nin-DC-SINE1	.	Aplysia californica	SINE2/tRNA	
ACASINE	.	Anolis carolinensis	SINE2/tRNA	
ACASINE2	.	Anolis carolinensis	SINE2/tRNA	
ACASINE2a	.	Anolis carolinensis	SINE	
ACASINE2b	.	Anolis carolinensis	SINE	
ACASINE2c	.	Anolis carolinensis	SINE	
ACASINE2d	.	Anolis carolinensis	SINE	
ACASINE3	.	Anolis carolinensis	SINE	
ACSINE1	.	Anolis carolinensis	SINE	
AFESINE	.	Azemiops feae	SINE	
AFRO_LA	.	Loxodonta africana	SINE2/tRNA	
AFROA1	.	Orycteropus afer	SINE2/tRNA	
AFROA1B	.	Orycteropus afer	SINE2/tRNA	
AFROSINE	.	Proboscidea	SINE2/tRNA	
AFROSINE-1_LA	.	Loxodonta africana	SINE2/tRNA	
AFROSINE-2_LA	.	Loxodonta africana	SINE2/tRNA	
AFROSINE1B	.	Proboscidea	SINE2/tRNA	
AFROSINE2	.	Proboscidea	SINE2/tRNA	
AFROSINE3	.	Proboscidea	SINE2/tRNA	
Ag-CR1-22	.	Anopheles gambiae	CR1	1,3,11,4,2
Ag-CR1-7	.	Anopheles gambiae	CR1	2,11,56,57,68,1,58,4,3,59
Ag-I-2	.	Anopheles gambiae	I	6,2,9,10,11,13,1,4,15,16,17,3,7
Ag-Jen-1	.	Anopheles gambiae	Jockey	16,1,3,11,4,2,15
Ag-Jen-11	.	Anopheles gambiae	Jockey	16,1,3,11,4,2,15
Ag-Jock-1	.	Anopheles gambiae	Jockey	16,1,3,11,4,2,15
Ag-Jock-12	.	Anopheles gambiae	Jockey	16,1,3,11,4,2,15
Ag-Jock-13	.	Anopheles gambiae	Jockey	16,1,3,82,4,11,2,15
Ag-L1-5	.	Anopheles gambiae	L1	16,1,26,43,3,44,24,11,4,2
Ag-Loner-1	AAAB01008849	Anopheles gambiae	I	
Ag-Outcast-2	.	Anopheles gambiae	Outcast	6,2,9,10,11,80,1,4,15,16,17,3,7
Ag-Outcast-5	.	Anopheles gambiae	Outcast	6,2,9,10,11,80,1,4,15,16,17,3,7
Ag-Outcast-6	.	Anopheles gambiae	Outcast	6,2,9,10,11,80,1,4,15,16,17,3,7
Ag-R1-6	.	Anopheles gambiae	R1	2,8,11,86,12,1,14,4,15,16,3
ALINE	.	Mammalia	RTEX	16,11,15
ALINE	.	Mammalia	RTEX	16,11,15
ALINE	.	Mammalia	RTEX	16,11,15
ALPINE1	.	Vicugna pacos	SINE2/tRNA	
ALPINE1B	.	Vicugna pacos	SINE2/tRNA	
ALPINE2	.	Vicugna pacos	SINE2/tRNA	
ALU	.	Primates	SINE1/7SL	
Alu1_OG	.	Otolemur garnettii	SINE1/7SL	
Alu1_TS	.	Carlito syrichta	SINE1/7SL	
Alu2_OG	.	Otolemur garnettii	SINE1/7SL	
Alu2_TS	.	Carlito syrichta	SINE1/7SL	

Alu3_TS	.	Carlito syrichta	SINE1/7SL
AluJ_Mim	.	Microcebus murinus	SINE1/7SL
AluJr	.	Primates	SINE1/7SL
AluJr4	.	Primates	SINE1/7SL
AluMacYa3	.	Macaca	SINE1/7SL
AluMacYb2	.	Macaca	SINE1/7SL
AluMacYb4	.	Macaca	SINE1/7SL
AluSc5	.	Primates	SINE1/7SL
AluSc8	.	Primates	SINE1/7SL
AluSg1	.	Primates	SINE1/7SL
AluSg4	.	Primates	SINE1/7SL
AluSg7	.	Primates	SINE1/7SL
AluSq10	.	Primates	SINE1/7SL
AluSq2	.	Primates	SINE1/7SL
AluSq4	.	Primates	SINE1/7SL
AluSx1	.	Primates	SINE1/7SL
AluSx3	.	Primates	SINE1/7SL
AluSx4	.	Primates	SINE1/7SL
AluSz6	.	Primates	SINE1/7SL
AluYb10	.	Homo sapiens	SINE1/7SL
AluYb11	.	Homo sapiens	SINE1/7SL
AluYb8a1	.	Homo sapiens	SINE1/7SL
AluYc5	.	Primates	SINE1/7SL
AluYd3a1_gib	.	Nomascus leucogenys	SINE1/7SL
AluYf5	.	Primates	SINE1/7SL
AluYk11	.	Primates	SINE1/7SL
AluYk12	.	Primates	SINE1/7SL
AluYk13	.	Homo sapiens	SINE1/7SL
Ambal-1_FCy	.	Fragilariopsis cylindrus	Ambal 238,10,11,6
Ambal-1_PMul	.	Pseudo-nitzschia multiseries	Ambal 1,17,10,11,6,2
Ambal-1_TP	.	Thalassiosira pseudonana	Ambal 17,11
Ambal-1N1_TP	.	Thalassiosira pseudonana	Ambal
Ambal-2_FCy	.	Fragilariopsis cylindrus	Ambal 17,10,11,6
Ambal-2_TP	.	Thalassiosira pseudonana	Ambal 17,10,11,247,6,7
Ambal-3_FCy	.	Fragilariopsis cylindrus	Ambal 1,17,10,11,6,2
Ambal-3N1_FCy	.	Fragilariopsis cylindrus	Ambal
AmnSINE1_AMi	.	Alligator mississippiensis	SINE3/5S
AmnSINE1_Crp	.	Crocodylus porosus	SINE3/5S
AmnSINE1_GG	.	Gallus gallus	SINE3/5S
AmnSINE1_HS	.	Homo sapiens	SINE3/5S
AMPLICON_AA	AF065437	Aedes albopictus	RTE
AnolisSINE2	.	Anolis carolinensis	SINE2/tRNA
AnolisSINE2a	.	Anolis carolinensis	SINE
AnolisSINE2b	.	Anolis carolinensis	SINE
AnolisSINE2c	.	Anolis carolinensis	SINE
AnolisSINE2s	.	Anolis carolinensis	SINE2/tRNA
ARMER1	.	Hippopotamidae	L1
ATLINE1_1	.	Arabidopsis thaliana	L1
ATLINE1_10	AL161518	Arabidopsis thaliana	L1 78,2,24,11,1,81,77,4,16,75,3,76
ATLINE1_10A	AC006300	Arabidopsis thaliana	L1
43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76			

ATLINE1_11	AL161498		Arabidopsis thaliana	L1	43,78,2,44,24,11,12,1,14,81,77,4,16,75,3,76
ATLINE1_12	AC002387		Arabidopsis thaliana	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
ATLINE1_2	.		Arabidopsis thaliana	L1	
ATLINE1_3A	.		Arabidopsis thaliana	L1	
ATLINE1_4	AC006248		Arabidopsis thaliana	L1	
ATLINE1_5	AC007047		Arabidopsis thaliana	L1	
ATLINE1_6	AC007047,	AL161499	Arabidopsis thaliana	L1	16,1,75,3,76,77,24,11,4,2
ATLINE1_7	AL138652		Arabidopsis thaliana	L1	
43,78,79,2,44,10,11,80,1,81,77,4,17,75,3,76,7					
ATLINE1_8	AB073153		Arabidopsis thaliana	L1	1,43,3,44,4,2
ATLINE1_9	AC005897		Arabidopsis thaliana	L1	1,43,3,44,11,4,2
ATLINE1A	.		Arabidopsis thaliana	L1	
ATLINE2	AC005724		Arabidopsis thaliana	L1	
ATLINEIII	.		Arabidopsis thaliana	L1	
AtSB2	.		Arabidopsis thaliana	SINE2/tRNA	
AtSB3	AY033702		Arabidopsis thaliana	SINE2/tRNA	
AtSB4	.		Arabidopsis thaliana	SINE2/tRNA	
AtSB5	.		Arabidopsis thaliana	SINE2/tRNA	
AtSB6	.		Arabidopsis thaliana	SINE2/tRNA	
ATSINE2A	AC006528		Arabidopsis thaliana	SINE	
Au	.		Aegilops umbellulata	SINE2/tRNA	
AvaIII	.		Salmoniformes	SINE2/tRNA	
AviRTE_AMac	.		Ara macao	RTE	16,1,26,3,4,11,2,15
AviRTE_ApVit	.		Apaloderma vittatum	RTE	16,1,26,3,11,4,2,15
AviRTE_AVi	.		Amazona vittata	RTE	
AviRTE_BRh	.		Buceros rhinoceros	RTE	16,1,26,3,11,4,2,15
AviRTE_CAn	.		Calypte anna	RTE	16,1,26,3,11,4,2,15
AviRTE_GRu	.		Gymnopithys rufigula	RTE	16,1,26,3,11,4,2,15
AviRTE_LLo	.		Loa loa	RTE	16,1,26,3,11,4,2,15
AviRTE_MUn	.		Melopsittacus undulatus	RTE	16,1,26,3,11,4,2,15
AviRTE_MUni	.		Mesitornis unicolor	RTE	16,1,26,3,11,4,2,15
AviRTE_MVi	.		Manacus vitellinus	RTE	16,1,26,3,11,4,2,15
AviRTE_OMe	.		Oreotrochilus melanogaster	RTE	16,1,26,3,24,11,4,2
AviRTE_TiGu	.		Tinamus guttatus	RTE	16,1,26,3,11,4,2,15
AviRTE_WBa	.		Wuchereria bancrofti	RTE	65,11
AviRTE_ZCh	.		Zimmerius chrysops	RTE	16,1,26,3,24,11,4,2
B1	.		Mus musculus	SINE1/7SL	
B1_Mio	.		Microtus ochrogaster	SINE1/7SL	
B1_Mm	.		Mus	SINE1/7SL	
B1_Mur1	.		Muridae	SINE1/7SL	
B1_Mur2	.		Muridae	SINE1/7SL	
B1_Mur3	.		Muridae	SINE1/7SL	
B1_Mur4	.		Muridae	SINE1/7SL	
B1_Mus1	.		Mus	SINE1/7SL	
B1_Mus2	.		Mus	SINE1/7SL	
B1_Rn	.		Rattus	SINE1/7SL	
B1-dID	.		Rodentia	SINE1/7SL	
B1F	.		Mus musculus	SINE1/7SL	
B1F-1_CLan	.		Chinchilla lanigera	SINE1/7SL	
B1F1	.		Muridae	SINE1/7SL	
B1F1A	.		Cavia porcellus	SINE1/7SL	

B1F2	.	Muridae	SINE1/7SL	
B2	.	Mus musculus	SINE2/tRNA	
B2_Mm1a	.	Mus	SINE2/tRNA	
B2_Mm1t	.	Mus	SINE2/tRNA	
B2_Mm2	.	Mus	SINE2/tRNA	
B2_Rat1	.	Rattus	SINE2/tRNA	
B2_Rat2	.	Rattus	SINE2/tRNA	
B2_Rat3	.	Rattus	SINE2/tRNA	
B2_Rat4	.	Rattus	SINE2/tRNA	
B2_Rn	.	Rattus	SINE2/tRNA	
B2_Rn1	.	Rattus	SINE2/tRNA	
B2_Rn2	.	Rattus	SINE2/tRNA	
B2_RnY	.	Rattus	SINE2/tRNA	
B2L_S	.	Dipodidae	SINE2/tRNA	
B3	.	Rodentia	SINE2/tRNA	
B3A	.	Rodentia	SINE2/tRNA	
B4	.	Rodentia	SINE2/tRNA	
B4A	.	Rodentia	SINE2/tRNA	
Baggins-1_NVi	.	Nasonia vitripennis	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
Baggins-2_NVi	.	Nasonia vitripennis	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
Baggins-3_NVi	.	Nasonia vitripennis	Loa	6,2,9,10,11,13,1,4,17,3,7
BAGGINS1	.	Drosophila melanogaster	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
Baggins1_Cis	.	Ciona savignyi	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
BALN1_HV	AB016119	Hordeum vulgare	L1	
BALN2_HV	AB016120	Hordeum vulgare	L1	
BALN3_HV	AB016121	Hordeum vulgare	L1	
BC1	.	Rattus norvegicus	SINE2/tRNA	
BC1_Cp	.	Mus musculus	SINE2/tRNA	
BC1_Cpo	.	Cavia porcellus	SINE2/tRNA	
BC1_Dor	.	Dipodomys ordii	SINE2/tRNA	
BC1_Ma	.	Mesocricetus	SINE2/tRNA	
BC1_Mm	.	Mus	SINE2/tRNA	
BC1_Rn	.	Rattus	SINE2/tRNA	
BCS	X51356	Bos taurus	SINE	
BflSINE1	.	Branchiostoma floridae	SINE3/5S	
BflSINE2	.	Branchiostoma floridae	SINE2/tRNA	
BILBO	U73803	Drosophila subobscura	Loa	
BivaCORE-SINE1_RuDe.	.	Ruditapes decussatus	SINE2/tRNA	
BivaCORE-SINE2_MiYe.	.	Mizuhopecten yessoensis	SINE2/tRNA	
BivaMD-SINE1_CrVi	.	Crassostrea virginica	SINE2/tRNA	
BivaMD-SINE1_TeGr	.	Tegillarca granosa	SINE2/tRNA	
BivaMeta-SINE1_HyCu.	.	Hyriopsis cumingii	SINE2/tRNA	
BivaV-SINE1_BaAz	.	Bathymodiolus azoricus	SINE2/tRNA	
BivaV-SINE1_ChFa	.	Azumapecten farreri	SINE2/tRNA	
BivaV-SINE1_HyCu	.	Hyriopsis cumingii	SINE2/tRNA	
BivaV-SINE1_MiYe	.	Mizuhopecten yessoensis	SINE2/tRNA	
BivaV-SINE1_MyCa	.	Mytilus californianus	SINE2/tRNA	
BivaV-SINE1_RuDe	.	Ruditapes decussatus	SINE2/tRNA	
BivaV-SINE1_SiCo	.	Sinonovacula constricta	SINE2/tRNA	
BivaV-SINE2_CrGi	.	Crassostrea gigas	SINE2/tRNA	
BivaV-SINE2_MiYe	.	Mizuhopecten yessoensis	SINE2/tRNA	

BivaV-SINE2_PiFu	.	Pinctada fucata	SINE2/tRNA		
BivaV-SINE3_CoAn	.	Coelomactra antiquata	SINE2/tRNA		
BivaV-SINE3_SpSo	.	Spisula solidissima	SINE2/tRNA		
BM1	X03547	Bombyx mori	SINE2/tRNA		
BM2A	X70929	Bombyx mori	L1		
BMC1	AB018558	Bombyx mori	Jockey	16,1,87,3,88,11,4,2,15	
BmSE	.	Bombyx mori	SINE2/tRNA		
BNR1	EU564339	Beta vulgaris	L1	1,3,4,2	
BoSB10A	.	Brassica oleracea	SINE		
BoSB10A	.	Brassica oleracea	SINE2/tRNA		
BoSB11	.	Brassica oleracea	SINE		
BoSB11	.	Brassica oleracea	SINE2/tRNA		
BoSB12	.	Brassica oleracea	SINE		
BoSB12	.	Brassica oleracea	SINE2/tRNA		
BoSB13	.	Brassica oleracea	SINE		
BoSB13	.	Brassica oleracea	SINE2/tRNA		
BoSB14A	.	Brassica oleracea	SINE		
BoSB14A	.	Brassica oleracea	SINE2/tRNA		
BoSB14B	.	Brassica oleracea	SINE		
BoSB14B	.	Brassica oleracea	SINE2/tRNA		
BoSB15	.	Brassica oleracea	SINE		
BoSB15	.	Brassica oleracea	SINE2/tRNA		
BoSB1A	.	Brassica oleracea	SINE		
BoSB1A	.	Brassica oleracea	SINE2/tRNA		
BoSB2	.	Brassica oleracea	SINE		
BoSB2	.	Brassica oleracea	SINE2/tRNA		
BoSB3	.	Brassica oleracea	SINE		
BoSB3	.	Brassica oleracea	SINE2/tRNA		
BoSB5A	.	Brassica oleracea	SINE		
BoSB5A	.	Brassica oleracea	SINE2/tRNA		
BoSB5B	.	Brassica oleracea	SINE		
BoSB5B	.	Brassica oleracea	SINE2/tRNA		
BoSB5C	.	Brassica oleracea	SINE		
BoSB5C	.	Brassica oleracea	SINE2/tRNA		
BoSB5D	.	Brassica oleracea	SINE		
BoSB5D	.	Brassica oleracea	SINE2/tRNA		
BoSB6A	.	Brassica oleracea	SINE		
BoSB6A	.	Brassica oleracea	SINE2/tRNA		
BoSB6B	.	Brassica oleracea	SINE		
BoSB6B	.	Brassica oleracea	SINE2/tRNA		
BoSB6C	.	Brassica oleracea	SINE		
BoSB6C	.	Brassica oleracea	SINE2/tRNA		
BoSB6D	.	Brassica oleracea	SINE		
BoSB6D	.	Brassica oleracea	SINE2/tRNA		
BoSB7A	.	Brassica oleracea	SINE		
BoSB7A	.	Brassica oleracea	SINE2/tRNA		
BoSB8A	.	Brassica oleracea	SINE		
BoSB8A	.	Brassica oleracea	SINE2/tRNA		
BoSB8B	.	Brassica oleracea	SINE		
BoSB8B	.	Brassica oleracea	SINE2/tRNA		
BoSB9A	.	Brassica oleracea	SINE		

BoSB9A	.	Brassica oleracea	SINE2/tRNA		
BoSB9B	.	Brassica oleracea	SINE		
BoSB9B	.	Brassica oleracea	SINE2/tRNA		
Bov-tA1	.	Ruminantia	SINE2/tRNA		
Bov-tA2	.	Ruminantia	SINE2/tRNA		
Bov-tA3	.	Ruminantia	SINE2/tRNA		
BOVA2	.	Bos taurus	SINE		
BovB	.	Bos taurus	RTE	16,1,26,3,24,11,4,2	
BovB_Ac	.	Anolis carolinensis	RTE		
BovB_ACo	.	Agkistrodon contortrix	RTE	1,3,4,2	
BovB_Af	.	Antechinus flavipes	RTE		
BovB_ALi	.	Amblyomma limbatum	RTE		
BovB_Ano	.	Amphibolurus norrisi	RTE		
BovB_Bh	.	Bothriocroton hydrosauri	RTE		
BovB_BM	.	Bombyx mori	RTE		
BovB_Bp	.	Burramys parvus	RTE		
BovB_Ca	.	Ctenotus atlas	RTE		
BovB_Cg	.	Christinus	RTE		
BovB_Dr	.	Christinus	RTE		
BovB_Ec	.	Equus caballus	RTE		
BovB_Er	.	Eremiascincus richardsonii	RTE		
BovB_Es	.	Egernia stokesii	RTE		
BovB_Et	.	Christinus	RTE		
BovB_Gd	.	Eremiascincus douglasi	RTE		
BovB_Gl	.	Gehyra lazelli	RTE		
BovB_Gv	.	Gehyra variegata	RTE		
BovB_Hs	.	Hydrophis spiralis	RTE		
BovB_Io	.	Isoodon obesulus	RTE		
BovB_La	.	Loxodonta africana	RTE		
BovB_Ma	.	Macropus antilopinus	RTE		
BovB_Me	.	Notamacropus eugenii	RTE		
BovB_Ml	.	Gehyra variegata	RTE		
BovB_Oa	.	Ovis aries	RTE		
BovB_Pb	.	Petaurus breviceps	RTE		
BovB_Pc	.	Procavia capensis	RTE		
BovB_Pg	.	Perameles gunnii	RTE		
BovB_Plat	.	Ornithorhynchus	RTE		
BovB_PMo	.	Python molurus	RTE	65,16,1,26,3,11,4,2,15	
BovB_Ta	.	Tachyglossus aculeatus	RTE	1,3,11,4,2	
BovB_Tr	.	Tiliqua rugosa	RTE		
BovB_VA	AF332697	Vipera ammodytes	RTE	16,1,26,3,24,11,4,2	
BovBa-1_EF	.	Eptesicus fuscus	RTE		
Bovc-tA2	.	Ruminantia	SINE2/tRNA		
BOVTA	.	Bos taurus	SINE2/tRNA		
BRIDGE1_FR	.	Takifugu rubripes	BRIDGE		
BRIDGE1_TN	.	Tetraodon nigroviridis	BRIDGE	74,1,3,71,103,72,2	
BRIDGE2_FR	.	Takifugu rubripes	Penelope		
BRIDGE2_FR	.	Takifugu rubripes	Penelope		
BS	X77571	Drosophila melanogaster	Jockey	16,1,87,88,3,11,4,2,15	
BS2	.	Drosophila melanogaster	Jockey	16,1,87,14,3,88,11,4,2,15	
BS3_DM	.	Drosophila melanogaster	Jockey	16,1,3,11,4,2,15	

BS4_DM	.	Drosophila melanogaster	Jockey	1,3,2	
BTALUL1	.	Bos taurus	SINE2/tRNA		
BTALUL2	X05090	Bos taurus	RTE		
BuceSINE	.	Buceros rhinoceros	SINE		
BvL1	FM993986	Beta vulgaris	L1	1,3,11,4,2	
BvL1	FM993986	Beta vulgaris	L1	1,3,11,4,2	
BvL1-2	FM993987	Beta vulgaris	L1		
BvL1-2	FM993987	Beta vulgaris	L1		
CAM1_GG	X70342	Gallus gallus	CR1		
CAN	.	Carnivora	SINE2/tRNA		
CASINE	.	Oryza sativa	SINE		
CATS	.	Bombyx mori	R1	6,2,9,10,11,86,13,1,4,15,16,17,3,7	
CAVID	.	Cavia porcellus	SINE2/tRNA		
CAVID1B	.	Cavia porcellus	SINE2/tRNA		
CAVID1C	.	Cavia porcellus	SINE2/tRNA		
CAVID2	.	Cavia porcellus	SINE2/tRNA		
CAVID2A	.	Cavia porcellus	SINE2/tRNA		
CAVID2B	.	Cavia porcellus	SINE2/tRNA		
CAVID2B1	.	Cavia porcellus	SINE2/tRNA		
CAVID2B2	.	Cavia porcellus	SINE2/tRNA		
CAVID2C	.	Cavia porcellus	SINE2/tRNA		
CAVID2C2	.	Cavia porcellus	SINE2/tRNA		
CAVID2D	.	Cavia porcellus	SINE2/tRNA		
CAVID2E	.	Cavia porcellus	SINE2/tRNA		
CELE45	.	Caenorhabditis elegans	SINE2/tRNA		
CERE1	.	Ceratotherium simum simum	SINE2/tRNA		
CGT1	L76172	Colletotrichum gloeosporioides	L1	16,11,15	
ChaetoSINE1	.	Chaetopterus variopedatus	SINE2/tRNA		
CHR-1	AB005033	Physeter catodon	SINE2/tRNA		
CHR-2	AB005034	Physeter catodon	SINE2/tRNA		
CHR-2_BT	.	Bos taurus	SINE2/tRNA		
CHR-2A	.	Ruminantia	SINE2/tRNA		
CHR-2B	.	Ruminantia	SINE2/tRNA		
CHR2A_Ttr	.	Tursiops truncatus	SINE2/tRNA		
CHR2B_Ttr	.	Tursiops truncatus	SINE2/tRNA		
CHR2C_Ttr	.	Tursiops truncatus	SINE2/tRNA		
CHR2C2_Ttr	.	Tursiops truncatus	SINE2/tRNA		
CHR2D_Ttr	.	Tursiops truncatus	SINE2/tRNA		
CHR2E_Ttr	.	Tursiops truncatus	SINE2/tRNA		
CHR2F_Ttr	.	Tursiops truncatus	SINE2/tRNA		
CHRL	.	Cetartiodactyla	SINE2/tRNA		
CHRL1_BT	.	Bos taurus	SINE2/tRNA		
Cics-1	.	Ciona intestinalis	SINE2/tRNA		
CiL2	.	Ciona intestinalis	CR1		
Cili-1	AJ226259	Ciona intestinalis	L1		
Cili-2	AJ227045	Ciona intestinalis	Loa		
CIN4E_ZM	Y00086	Zea mays	L1	1,43,3,44,11,4,2	
CIN4L_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,1,14,4,16,3	
ClySINE1	.	Clytia hemisphaerica	SINE2/tRNA		
Cnl1	.	Cryptococcus neoformans	CRE	1,3,2	
CoeSINE1a	.	Latimeria chalumnae	SINE2/tRNA		

CoeSINE1c	Latimeria chalumnae	SINE2/tRNA	
CoeSINE1d	Latimeria chalumnae	SINE2/tRNA	
CoeSINE1e	Latimeria chalumnae	SINE2/tRNA	
CoeSINE1f	Latimeria chalumnae	SINE2/tRNA	
CoeSINE1g	Latimeria chalumnae	SINE2/tRNA	
CoeSINE1h	Latimeria chalumnae	SINE2/tRNA	
CoeSINE1i	Latimeria chalumnae	SINE2/tRNA	
CoeSINE1j	Latimeria chalumnae	SINE2/tRNA	
CoeSINE1k	Latimeria chalumnae	SINE2/tRNA	
CoeSINE2	Latimeria chalumnae	SINE2/tRNA	
CoeSINE3	Latimeria chalumnae	SINE2/tRNA	
CoeSINE4	Latimeria chalumnae	SINE2/tRNA	
CoeSINE5	Latimeria chalumnae	SINE3/5S	
CoeSINE6	Latimeria chalumnae	SINE2/tRNA	
Copia3_HV_I	Hordeum vulgare	Copia	
Copia3_HV_LTR	Hordeum vulgare	Copia	
Coprina_Cc1	Coprinopsis cinerea	Penelope	
Coprina_Pc1	Phanerochaete chrysosporium	Penelope	3,1
Coprina_Pc2	Phanerochaete chrysosporium	Penelope	
Coprina-1_SLL	Serpula lacrymans	Penelope	
Coprina-2_SLL	Serpula lacrymans	Penelope	
Coprina-3_SLL	Serpula lacrymans	Penelope	
CR1_1a_XT	Xenopus tropicalis	CR1	
CR1_1b_Xt	Xenopus tropicalis	CR1	
CR1_AC_1	Anolis carolinensis	CR1	2,63,11,48,1,64,4,15,16,3
CR1_B	Gallus gallus	CR1	
CR1_E_PPU	Picoides pubescens	CR1	
CR1_E_VSP	Veniliornis spilogaster	CR1	
CR1_Ele1	Aedes aegypti	CR1	1,3,4,11,2
CR1_Ele10	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele11	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele12	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele13	Aedes aegypti	CR1	16,1,3,59,4,11,57,2,15
CR1_Ele14	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele15	Aedes aegypti	CR1	1,3,59,11,4,57,2
CR1_Ele16	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele17	Aedes aegypti	CR1	16,1,3,59,11,4,2,15
CR1_Ele18	Aedes aegypti	CR1	16,1,3,59,11,4,2,15
CR1_Ele19	Aedes aegypti	CR1	54,1,3,59,53,11,4,2
CR1_Ele2	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59
CR1_Ele20	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele21	Aedes aegypti	CR1	54,1,3,59,53,11,4,2
CR1_Ele22	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele23	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele24	Aedes aegypti	CR1	53,92,2,54,11,56,57,68,1,58,4,15,16,3,59,60
CR1_Ele25	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1_Ele26	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele27	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1_Ele28	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1_Ele29	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1_Ele3	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2

CR1_Ele30	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele31	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele32	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele33	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele34	.	Aedes aegypti	CR1	16,1,54,3,58,59,53,11,4,56,2,15
CR1_Ele35	.	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1_Ele36	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele37	.	Aedes aegypti	CR1	54,1,3,58,59,53,11,56,4,2
CR1_Ele38	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1_Ele38B_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele39	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59,60
CR1_Ele4	.	Aedes aegypti	CR1	16,1,3,59,11,4,2,15
CR1_Ele40	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1_Ele41	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_Ele42	.	Aedes aegypti	CR1	53,2,54,11,57,68,1,4,15,16,3,59
CR1_Ele5	.	Aedes aegypti	CR1	42,1,8,14,3,11,4,12,2
CR1_Ele6	.	Aedes aegypti	CR1	53,2,54,8,11,12,1,4,15,16,3,59
CR1_Ele7	.	Aedes aegypti	CR1	16,1,3,59,11,4,2,15
CR1_Ele8	.	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59
CR1_Ele9	.	Aedes aegypti	CR1	1,3,11,4,2
CR1_F	.	Gallus gallus	CR1	16,1,3,11,4,48,2,15
CR1_GD	X14729	Gallus gallus	CR1	
CR1_GG	U88211	Gallus gallus	CR1	
CR1_HS	.	Homo sapiens	CR1	
CR1_Mam	.	Mammalia	CR1	
CR1_Mam	.	Mammalia	CR1	
CR1_Mam	.	Mammalia	CR1	
CR1-1_AC	.	Aplysia californica	CR1	16,1,3,11,4,2,15
CR1-1_ACC	.	Aquila chrysaetos canadensis	CR1	16,1,3,11,4,48,2,15
CR1-1_ACo	.	Agkistrodon contortrix	CR1	1,3,4,2
CR1-1_AG	.	Anopheles gambiae	L2B	54,1,8,3,59,53,11,4,57,12,2
CR1-1_ASi	.	Alligator sinensis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-1_BF	.	Branchiostoma floridae	CR1	53,2,54,11,56,57,1,58,4,15,16,3,59
CR1-1_BM	.	Bombyx mori	CR1	1,3,11,4,2
CR1-1_BTa	.	Bemisia tabaci	CR1	53,2,54,11,57,68,1,4,3,59
CR1-1_CBre	.	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-1_CGi	.	Crassostrea gigas	CR1	16,1,3,11,4,2,15
CR1-1_CJap	.	Caenorhabditis japonica	CR1	16,1,3,4,11,2,15
CR1-1_CM	DQ524336	Callorhinchus milii	CR1	
CR1-1_CPB	.	Chrysemys picta bellii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-1_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-1_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-1_Croc	.	Crocodylidae	CR1	16,1,3,4,11,48,2,15
CR1-1_CTe	.	Capitella teleta	CR1	1,3,4,2
CR1-1_DAn	.	Drosophila ananassae	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-1_DBi	.	Drosophila biarmipes	CR1	
CR1-1_DBp	.	Drosophila bipectinata	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-1_DEl	.	Drosophila elegans	CR1	1,3,11,4,2
CR1-1_DEu	.	Drosophila eugracilis	CR1	1,16,3,11,4,57,2,15
CR1-1_DF	.	Drosophila ficusphila	CR1	1,3,4,2
CR1-1_DK	.	Drosophila kikkawai	CR1	1,3,4,2

CR1-1_DPer	.	Drosophila persimilis	CR1	16,1,3,11,4,57,2,15
CR1-1_DRh	.	Drosophila rhopaloa	CR1	16,1,3,59,60,4,11,57,2,15
CR1-1_DSer	.	Drosophila serrata	CR1	1,3,11,4,2
CR1-1_DT	.	Drosophila takahashii	CR1	16,1,3,11,4,57,2,15
CR1-1_DVi	.	Drosophila virilis	CR1	16,1,3,11,4,2,15
CR1-1_DWil	.	Drosophila willistoni	CR1	1,3,11,4,57,2
CR1-1_Gav	.	Gavialis gangeticus	CR1	16,1,3,11,4,2,15
CR1-1_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-1_Hmel	.	Heliconius melpomene	CR1	1,3,11,4,2
CR1-1_HMM	.	Heliconius melpomene melpomene	CR1	1,3,4,2
CR1-1_IS	.	Ixodes scapularis	CR1	16,1,3,11,4,2,15
CR1-1_LCh	.	Latimeria chalumnae	CR1	2,63,11,48,1,64,4,15,16,3
CR1-1_LG	.	Lottia gigantea	CR1	21,16,1,3,11,15,4,2
CR1-1_Lme	.	Latimeria menadoensis	CR1	16,1,3,11,4,2,15
CR1-1_LMi	.	Locusta migratoria	CR1	16,1,3,11,4,2,15
CR1-1_LSal	.	Lepeophtheirus salmonis	CR1	16,1,3,11,4,2,15
CR1-1_LVa	.	Litopenaeus vannamei	CR1	16,1,3,11,4,2,15
CR1-1_MUn	.	Melopsittacus undulatus	CR1	16,1,3,11,4,48,2,15
CR1-1_NV	.	Nematostella vectensis	CR1	
CR1-1_NVi	.	Nasonia vitripennis	CR1	1,8,3,11,4,12,2
CR1-1_OL	.	Oryzias latipes	CR1	16,1,3,4,11,2,15
CR1-1_PH	.	Parhyale hawaiensis	CR1	16,1,3,11,4,2,15
CR1-1_PMo	.	Python molurus	CR1	
CR1-1_PPac	.	Pristionchus pacificus	CR1	1,3,4,2
CR1-1_PPo	.	Papilio polytes	CR1	
CR1-1_PSi	.	Pelodiscus sinensis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-1_SP	.	Strongylocentrotus purpuratus	CR1	
CR1-1_TCa	.	Tribolium castaneum	CR1	1,3,4,2
CR1-1_XL	.	Xenopus laevis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-10_AG	NT_078268.2	Anopheles gambiae	CR1	16,1,3,11,4,2,15
CR1-10_AMi	.	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-10_BF	.	Branchiostoma floridae	CR1	
CR1-10_CGi	.	Crassostrea gigas	CR1	1,3,11,4,2
CR1-10_CJap	.	Caenorhabditis japonica	CR1	1,3,4,2
CR1-10_CM	.	Callorhinchus milii	CR1	16,1,3,11,15,4,2
CR1-10_CPB	.	Chrysemys picta bellii	CR1	82,2,63,11,48,1,64,4,15,16,3
CR1-10_CQ	.	Culex quinquefasciatus	CR1	1,3,4,2
CR1-10_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-10_Crp	.	Crocodylus porosus	CR1	2,63,11,1,64,4,15,16,3
CR1-10_CTe	.	Capitella teleta	CR1	16,1,3,11,4,2,15
CR1-10_DR	.	Danio rerio	CR1	16,1,3,11,4,2,15
CR1-10_DTa	.	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-10_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-10_NV	.	Nematostella vectensis	CR1	
CR1-10_SP	.	Strongylocentrotus purpuratus	CR1	16,1,3,11,4,2,15
CR1-100_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59,60
CR1-101_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-102_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-103_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,127,3,59
CR1-104_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-105_AAe	.	Aedes aegypti	CR1	1,3,172,173,59,11,4,2

CR1-106_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,59
CR1-107_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1-108_AAe	.	Aedes aegypti	CR1	1,3,4,11,2
CR1-109_AAe	.	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-10B_AMi	.	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-10B_CM	.	Callorhinchus milii	CR1	63,16,1,64,3,11,4,48,2,15
CR1-10B_CPB	.	Chrysemys picta bellii	CR1	82,2,11,48,1,4,15,16,3
CR1-10D_AMi	.	Alligator mississippiensis	CR1	2,63,11,1,64,4,15,16,3
CR1-11_AG	.	Anopheles gambiae	CR1	2,11,56,57,68,1,58,4,127,3,59,60
CR1-11_AMi	.	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-11_BF	.	Branchiostoma floridae	CR1	
CR1-11_CGi	.	Crassostrea gigas	CR1	1,11,4,2
CR1-11_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-11_CM	.	Callorhinchus milii	CR1	
CR1-11_CQ	.	Culex quinquefasciatus	CR1	1,3,4,2
CR1-11_CRem	.	Caenorhabditis remanei	CR1	1,3,4,2
CR1-11_Crp	.	Crocodylus porosus	CR1	
CR1-11_CTe	.	Capitella teleta	CR1	1,3,4,2
CR1-11_DTa	.	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-11_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-11_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-11_NV	.	Nematostella vectensis	CR1	
CR1-110_AAe	.	Aedes aegypti	CR1	1,3,11,4,57,2
CR1-111_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-112_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-113_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59,60
CR1-114_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,57,2
CR1-115_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-116_AAe	.	Aedes aegypti	CR1	2,11,56,57,68,1,58,4,3,59
CR1-117_AAe	.	Aedes aegypti	CR1	16,1,3,60,4,11,2,15
CR1-118_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,69,3,59
CR1-119_AAe	.	Aedes aegypti	CR1	16,54,1,3,59,53,11,4,57,2,15
CR1-12_AG	.	Anopheles gambiae	CR1	16,54,1,3,53,11,4,244,2,15
CR1-12_AMi	.	Alligator mississippiensis	CR1	
CR1-12_BF	.	Branchiostoma floridae	CR1	
CR1-12_CGi	.	Crassostrea gigas	CR1	1,3,4,2
CR1-12_CJap	.	Caenorhabditis japonica	CR1	1,3,4,2
CR1-12_CM	.	Callorhinchus milii	CR1	
CR1-12_CPB	.	Chrysemys picta bellii	CR1	
CR1-12_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-12_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-12_Crp	.	Crocodylus porosus	CR1	16,1,3,11,4,2,15
CR1-12_CTe	.	Capitella teleta	CR1	2,11,57,1,4,15,16,3,59
CR1-12_DR	.	Danio rerio	CR1	1,3,4,2
CR1-12_DTa	.	Drosophila takahashii	CR1	1,3,11,4,57,2
CR1-12_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-12_HRo	.	Helobdella robusta	CR1	16,1,3,60,11,4,2,15
CR1-12_NV	.	Nematostella vectensis	CR1	
CR1-120_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-121_AAe	.	Aedes aegypti	CR1	1,3,4,2
CR1-122_AAe	.	Aedes aegypti	CR1	53,2,54,11,57,1,4,15,16,3,59

CR1-123_AAe	Aedes aegypti	CR1	1,3,59,11,4,2
CR1-124_AAe	Aedes aegypti	CR1	54,1,8,3,59,53,11,4,12,2
CR1-125_AAe	Aedes aegypti	CR1	1,3,59,11,4,2
CR1-126_AAe	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-127_AAe	Aedes aegypti	CR1	54,1,3,59,53,11,4,2
CR1-12B_CPB	Chrysemys picta bellii	CR1	
CR1-13_AMi	Alligator mississippiensis	CR1	
CR1-13_BF	Branchiostoma floridae	CR1	
CR1-13_CGi	Crassostrea gigas	CR1	16,1,3,24,4,11,2
CR1-13_CJap	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-13_CM	Callorhinchus milii	CR1	
CR1-13_CPB	Chrysemys picta bellii	CR1	
CR1-13_CQ	Culex quinquefasciatus	CR1	16,1,3,11,4,2,15
CR1-13_CRem	Caenorhabditis remanei	CR1	1,3,11,4,2
CR1-13_Crp	Crocodylus porosus	CR1	
CR1-13_CTe	Capitella teleta	CR1	1,3,4,2
CR1-13_DTa	Drosophila takahashii	CR1	1,3,11,4,2
CR1-13_HRo	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-13_NV	Nematostella vectensis	CR1	
CR1-13_SP	Strongylocentrotus purpuratus	CR1	
CR1-14_AMi	Alligator mississippiensis	CR1	
CR1-14_BF	Branchiostoma floridae	CR1	
CR1-14_CGi	Crassostrea gigas	CR1	21,1,3,11,4,2
CR1-14_CJap	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-14_CQ	Culex quinquefasciatus	CR1	1,3,59,11,4,57,2
CR1-14_Croc	Crocodylidae	CR1	2,11,48,1,4,15,16,3
CR1-14_CTe	Capitella teleta	CR1	1,3,11,4,2
CR1-14_HRo	Helobdella robusta	CR1	16,1,3,4,11,2,15
CR1-14B_CQ	Culex quinquefasciatus	CR1	16,1,3,11,4,2,15
CR1-15_AMi	Alligator mississippiensis	CR1	
CR1-15_BF	Branchiostoma floridae	CR1	
CR1-15_CJap	Caenorhabditis japonica	CR1	1,3,4,2
CR1-15_CQ	Culex quinquefasciatus	CR1	1,3,4,11,2
CR1-15_Crp	Crocodylus porosus	CR1	
CR1-15_CTe	Capitella teleta	CR1	1,3,4,2
CR1-15_HM	Hydra vulgaris	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-15_NV	Nematostella vectensis	CR1	
CR1-16_AMi	Alligator mississippiensis	CR1	
CR1-16_BF	Branchiostoma floridae	CR1	
CR1-16_CJap	Caenorhabditis japonica	CR1	1,3,4,2
CR1-16_CQ	Culex quinquefasciatus	CR1	54,1,3,59,53,11,4,57,2
CR1-16_Croc	Crocodylidae	CR1	2,63,11,1,64,4,15,16,3
CR1-16_Crp	Crocodylus porosus	CR1	
CR1-16_HRo	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-16B_CQ	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-17_AMi	Alligator mississippiensis	CR1	2,63,24,11,1,64,4,16,3
CR1-17_BF	Branchiostoma floridae	CR1	
CR1-17_CJap	Caenorhabditis japonica	CR1	1,3,4,2
CR1-17_CQ	Culex quinquefasciatus	CR1	1,3,4,2
CR1-17_Croc	Crocodylidae	CR1	
CR1-17_Crp	Crocodylus porosus	CR1	

CR1-17_Gav	.	Gavialis gangeticus	CR1	1,3,4,2
CR1-17_HM	.	Hydra vulgaris	CR1	1,23,3,11,4,2
CR1-18_BF	.	Branchiostoma floridae	CR1	
CR1-18_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-18_Crp	.	Crocodylus porosus	CR1	2,63,11,1,64,4,15,16,3
CR1-18_DR	.	Danio rerio	CR1	16,1,3,11,4,2,15
CR1-18_HRo	.	Helobdella robusta	CR1	16,1,3,11,15,4,2
CR1-18_NV	.	Nematostella vectensis	CR1	1,3,11,4,2
CR1-19_BF	.	Branchiostoma floridae	CR1	
CR1-19_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-19_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-19_HRo	.	Helobdella robusta	CR1	1,3,4,2
CR1-19_NV	.	Nematostella vectensis	CR1	16,1,3,11,4,2,15
CR1-1a_AG	.	Anopheles gambiae	L2B	53,2,54,8,11,57,12,68,1,4,3,59
CR1-1B_CM	.	Callorhinchus milii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-1B_CPB	.	Chrysemys picta bellii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-1b_Lch	.	Latimeria chalumnae	CR1	
CR1-1B_PH	.	Parhyale hawaiensis	CR1	16,1,3,11,15,4,2
CR1-1B_PSi	.	Pelodiscus sinensis	CR1	
CR1-1C_CM	.	Callorhinchus milii	CR1	
CR1-1C_PSi	.	Pelodiscus sinensis	CR1	
CR1-2_ACC	.	Aquila chrysaetos canadensis	CR1	16,1,3,4,11,48,2,15
CR1-2_AG	.	Anopheles gambiae	CR1	54,1,3,59,53,11,4,57,2
CR1-2_AMi	.	Alligator mississippiensis	CR1	2,11,48,1,4,15,16,3
CR1-2_ASi	.	Alligator sinensis	CR1	16,1,3,11,4,2,15
CR1-2_BF	.	Branchiostoma floridae	CR1	53,2,54,11,56,57,1,58,4,15,16,3,59
CR1-2_BTa	.	Bemisia tabaci	CR1	2,21,23,55,11,57,1,4,3,59,61
CR1-2_CBre	.	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-2_CGi	.	Crassostrea gigas	CR1	21,1,23,3,11,4,2
CR1-2_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-2_CPB	.	Chrysemys picta bellii	CR1	2,63,11,1,64,4,15,16,3
CR1-2_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2,126
CR1-2_CRem	.	Caenorhabditis remanei	CR1	1,3,4,2
CR1-2_Crp	.	Crocodylus porosus	CR1	2,63,11,48,1,64,4,15,16,3
CR1-2_DAn	.	Drosophila ananassae	CR1	1,3,11,4,57,2
CR1-2_DBi	.	Drosophila biarmipes	CR1	16,1,3,11,4,57,2,15
CR1-2_DBp	.	Drosophila bipectinata	CR1	1,3,11,4,2
CR1-2_DEl	.	Drosophila elegans	CR1	
CR1-2_DEu	.	Drosophila eugracilis	CR1	16,1,3,11,4,57,2,15
CR1-2_DF	.	Drosophila ficusphila	CR1	16,1,3,11,4,2,15
CR1-2_DK	.	Drosophila kikkawai	CR1	16,1,3,11,4,2,15
CR1-2_DPer	.	Drosophila persimilis	CR1	16,1,161,162,3,11,4,2,15
CR1-2_DRh	.	Drosophila rhopaloa	CR1	16,54,1,3,53,11,4,2,15
CR1-2_DSer	.	Drosophila serrata	CR1	1,3,59,60,11,4,57,2
CR1-2_DT	.	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-2_DWi	.	Drosophila willistoni	CR1	16,1,3,11,4,57,2,15
CR1-2_EL	.	Esox lucius	CR1	16,1,3,11,4,2,15
CR1-2_FR	.	Takifugu rubripes	CR1	1,47,40,3,4,41,2
CR1-2_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-2_Hmel	.	Heliconius melpomene	CR1	1,3,4,2
CR1-2_HMM	.	Heliconius melpomene melpomene	CR1	21,1,3,11,4,57,2

CR1-2_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-2_IS	.	Ixodes scapularis	CR1	1,3,4,2
CR1-2_LCh	.	Latimeria chalumnae	CR1	16,1,3,11,4,2,15
CR1-2_LG	.	Lottia gigantea	CR1	16,1,3,11,4,2,15
CR1-2_LMi	.	Locusta migratoria	CR1	16,1,3,11,4,48,62,2,15
CR1-2_LSal	.	Lepeophtheirus salmonis	CR1	16,1,3,11,4,2,15
CR1-2_LVa	.	Litopenaeus vannamei	CR1	1,3,11,4,2
CR1-2_MUn	.	Melopsittacus undulatus	CR1	16,1,3,11,4,48,2,15
CR1-2_NV	.	Nematostella vectensis	CR1	
CR1-2_NVi	.	Nasonia vitripennis	CR1	195,1,8,14,3,11,4,12,2
CR1-2_OL	.	Oryzias latipes	CR1	1,3,2
CR1-2_PH	.	Parhyale hawaiensis	CR1	145,54,1,16,3,53,146,11,4,2,15
CR1-2_PPac	.	Pristionchus pacificus	CR1	16,1,3,11,4,2,15
CR1-2_PPo	.	Papilio polytes	CR1	
CR1-2_PSi	.	Pelodiscus sinensis	CR1	16,1,3,11,4,2,15
CR1-2_SP	.	Strongylocentrotus purpuratus	CR1	
CR1-2_TCa	.	Tribolium castaneum	CR1	1,3,4,2
CR1-2_XL	.	Xenopus laevis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-2_XT	.	Xenopus tropicalis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-20_AMi	.	Alligator mississippiensis	CR1	16,1,11,4,2,15
CR1-20_BF	.	Branchiostoma floridae	CR1	
CR1-20_CQ	.	Culex quinquefasciatus	CR1	1,54,3,59,53,11,4,57,2
CR1-20_HM	.	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-20_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-20_NV	.	Nematostella vectensis	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,59
CR1-20_SP	081C18	Strongylocentrotus purpuratus	CR1	1,3,11,4,2
CR1-21_BF	.	Branchiostoma floridae	CR1	
CR1-21_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-21_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-21_NV	.	Nematostella vectensis	CR1	53,2,54,11,56,57,1,58,4,15,16,69,3,59
CR1-22_BF	.	Branchiostoma floridae	CR1	
CR1-22_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-22_DR	.	Danio rerio	CR1	1,3,4,2
CR1-22_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-22_HRo	.	Helobdella robusta	CR1	1,3,11,4,2
CR1-22_SP	.	Strongylocentrotus purpuratus	CR1	1,3,4,2
CR1-22-NV	.	Nematostella vectensis	CR1	
CR1-23_BF	.	Branchiostoma floridae	CR1	
CR1-23_CQ	.	Culex quinquefasciatus	CR1	1,3,4,2
CR1-23_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-23_HRo	.	Helobdella robusta	CR1	16,1,3,11,15,4,2
CR1-24_BF	.	Branchiostoma floridae	CR1	
CR1-24_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-24_DR	.	Danio rerio	CR1	1,3,4,2
CR1-24_HRo	.	Helobdella robusta	CR1	16,1,3,4,11,2,15
CR1-24_SP	.	Strongylocentrotus purpuratus	CR1	16,1,3,4,11,2,15
CR1-25_BF	.	Branchiostoma floridae	CR1	
CR1-25_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-25_DR	.	Danio rerio	L2	16,1,3,11,4,2,15
CR1-25_HM	.	Hydra vulgaris	CR1	53,2,54,11,56,57,1,58,4,3,59,60
CR1-25_HRo	.	Helobdella robusta	CR1	16,1,3,4,2,15

CR1-25_SP	.	Strongylocentrotus purpuratus	CR1	1,3,4,2
CR1-26_BF	.	Branchiostoma floridae	CR1	16,21,1,23,3,4,15,11,2
CR1-26_CQ	.	Culex quinquefasciatus	CR1	16,1,3,4,11,2,15
CR1-26_HM	.	Hydra vulgaris	CR1	16,21,1,23,3,11,4,2,15
CR1-27_BF	.	Branchiostoma floridae	CR1	
CR1-27_CQ	.	Culex quinquefasciatus	CR1	1,3,4,2
CR1-27_HM	.	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-28_BF	.	Branchiostoma floridae	CR1	
CR1-28_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-29_BF	.	Branchiostoma floridae	CR1	
CR1-2B_CPB	.	Chrysemys picta bellii	CR1	
CR1-2B_Croc	.	Crocodylidae	CR1	2,11,48,1,4,15,16,3
CR1-2B_Crp	.	Crocodylus porosus	CR1	2,63,24,11,48,1,64,4,16,3
CR1-2b_Lch	.	Latimeria chalumnae	CR1	
CR1-2B_MUn	.	Melopsittacus undulatus	CR1	1,3,4,2
CR1-2B_PPac	.	Pristionchus pacificus	CR1	16,1,3,11,4,2,15
CR1-2C_Croc	.	Crocodylidae	CR1	2,63,11,48,1,64,4,15,16,3
CR1-3_AG	.	Anopheles gambiae	CR1	1,54,3,58,59,53,4,56,57,2
CR1-3_BF	.	Branchiostoma floridae	CR1	
CR1-3_CBre	.	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-3_CGi	.	Crassostrea gigas	CR1	1,3,4,2
CR1-3_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-3_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-3_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-3_Croc	.	Crocodylidae	CR1	16,1,3,11,4,2,15
CR1-3_DAn	.	Drosophila ananassae	CR1	1,3,11,4,57,2
CR1-3_DBi	.	Drosophila biarmipes	CR1	16,1,3,11,4,2,15
CR1-3_DBp	.	Drosophila bipectinata	CR1	1,3,11,4,2
CR1-3_DEl	.	Drosophila elegans	CR1	1,3,4,2
CR1-3_DK	.	Drosophila kikkawai	CR1	1,3,11,4,57,2
CR1-3_DPer	.	Drosophila persimilis	CR1	
CR1-3_DT	.	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-3_DWi	.	Drosophila willistoni	CR1	1,3,11,4,57,2
CR1-3_FR	.	Takifugu rubripes	CR1	1,2
CR1-3_Gav	.	Gavialis gangeticus	CR1	16,1,3,11,15,4,2
CR1-3_HM	.	Hydra vulgaris	CR1	180,2,11,56,57,68,1,58,4,127,3,59
CR1-3_Hmel	.	Heliconius melpomene	CR1	1,3,11,4,2
CR1-3_IS	ABJB010111403	Ixodes scapularis	CR1	1,3,4,11,2
CR1-3_Lch	.	Latimeria chalumnae	CR1	
CR1-3_LMi	.	Locusta migratoria	CR1	16,1,3,11,4,62,2,15
CR1-3_LVa	.	Litopenaeus vannamei	CR1	1,14,3,11,4,2
CR1-3_NV	.	Nematostella vectensis	CR1	
CR1-3_NVi	.	Nasonia vitripennis	CR1	1,3,11,4,2
CR1-3_PH	.	Parhyale hawaiensis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-3_PPac	.	Pristionchus pacificus	CR1	1,3,4,2
CR1-3_PPo	.	Papilio polytes	CR1	
CR1-3_PSi	.	Pelodiscus sinensis	CR1	16,1,3,11,4,48,2,15
CR1-3_SP	.	Strongylocentrotus purpuratus	CR1	
CR1-3_TCa	.	Tribolium castaneum	CR1	1,3,11,4,2
CR1-30_BF	.	Branchiostoma floridae	CR1	
CR1-30_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15

CR1-30_SP	.	Strongylocentrotus purpuratus	CR1	180,2,11,56,57,68,1,58,4,15,16,3,59
CR1-31_BF	.	Branchiostoma floridae	CR1	
CR1-31_HM	.	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-31_SP	.	Strongylocentrotus purpuratus	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,59
CR1-32_BF	.	Branchiostoma floridae	CR1	
CR1-32_HM	.	Hydra vulgaris	CR1	1,23,3,11,4,2
CR1-33_BF	.	Branchiostoma floridae	CR1	
CR1-33_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-34_BF	.	Branchiostoma floridae	CR1	
CR1-34_HM	.	Hydra vulgaris	L2A	16,1,3,11,62,2,15
CR1-35_BF	.	Branchiostoma floridae	CR1	
CR1-35_HM	.	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-36_BF	.	Branchiostoma floridae	CR1	
CR1-36_HM	.	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-37_BF	.	Branchiostoma floridae	CR1	
CR1-37_DR	.	Danio rerio	CR1	16,1,3,11,4,2,15
CR1-37_HM	.	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-38_BF	.	Branchiostoma floridae	CR1	
CR1-39_BF	.	Branchiostoma floridae	CR1	
CR1-39_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-3b_LCh	.	Latimeria chalumnae	CR1	
CR1-3B_LMi	.	Locusta migratoria	CR1	16,1,3,11,4,48,62,2,15
CR1-3B_PSi	.	Pelodiscus sinensis	CR1	
CR1-3c_LCh	.	Latimeria chalumnae	CR1	63,16,1,64,3,11,4,48,2,15
CR1-3N1_LMi	.	Locusta migratoria	CR1	
CR1-4_AG	.	Anopheles gambiae	CR1	1,3,59,11,4,2
CR1-4_BF	.	Branchiostoma floridae	CR1	
CR1-4_CBre	.	Caenorhabditis brenneri	CR1	16,1,3,4,11,2,15
CR1-4_CGi	.	Crassostrea gigas	CR1	21,1,23,3,4,11,2
CR1-4_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-4_CM	.	Callorhinchus milii	CR1	2,11,48,1,4,15,16,109,3,110
CR1-4_CPB	.	Chrysemys picta bellii	CR1	16,1,3,4,11,48,2,15
CR1-4_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-4_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-4_Croc	.	Crocodylidae	CR1	2,63,11,48,1,64,4,15,16,3
CR1-4_CTe	.	Capitella teleta	CR1	1,3,4,2
CR1-4_DAn	.	Drosophila ananassae	CR1	1,3,11,4,2
CR1-4_DBi	.	Drosophila biarmipes	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-4_DEl	.	Drosophila elegans	CR1	16,1,3,4,11,2,15
CR1-4_DK	.	Drosophila kikkawai	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-4_DRh	.	Drosophila rhopaloa	CR1	16,1,3,11,4,2,15
CR1-4_DT	.	Drosophila takahashii	CR1	16,1,3,59,11,4,57,2,15
CR1-4_DWi	.	Drosophila willistoni	CR1	16,1,3,11,4,2,15
CR1-4_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-4_Hmel	.	Heliconius melpomene	CR1	1,3,4,2
CR1-4_LCh	.	Latimeria chalumnae	CR1	16,1,3,11,4,2,15
CR1-4_LMi	.	Locusta migratoria	CR1	16,1,3,11,4,62,2,15
CR1-4_LVa	.	Litopenaeus vannamei	CR1	1,14,3,11,4,2
CR1-4_MUn	.	Melopsittacus undulatus	CR1	1,3,2
CR1-4_NV	.	Nematostella vectensis	CR1	
CR1-4_NVi	.	Nasonia vitripennis	CR1	1,8,3,11,4,12,2

CR1-4_PH	.	Parhyale hawaiensis	CR1	16,1,3,11,4,2,15
CR1-4_PPac	.	Pristionchus pacificus	CR1	1,3,4,2
CR1-4_PSi	.	Pelodiscus sinensis	CR1	2,63,11,48,1,64,4,15,16,3
CR1-4_SP	.	Strongylocentrotus purpuratus	CR1	
CR1-4_TCa	.	Tribolium castaneum	CR1	2,11,57,1,4,15,16,3,59,60
CR1-40_BF	.	Branchiostoma floridae	CR1	
CR1-41_BF	.	Branchiostoma floridae	CR1	
CR1-42_BF	.	Branchiostoma floridae	CR1	
CR1-42_DR	.	Danio rerio	CR1	16,1,3,11,4,2,15
CR1-42_HM	.	Hydra vulgaris	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-43_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-43_BF	.	Branchiostoma floridae	CR1	
CR1-43_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-44_AAe	.	Aedes aegypti	CR1	92,2,11,56,57,1,58,4,15,16,3,59,60
CR1-44_BF	.	Branchiostoma floridae	CR1	
CR1-44_DR	.	Danio rerio	CR1	1,3,2
CR1-44_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-45_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,3,59
CR1-45_BF	.	Branchiostoma floridae	CR1	
CR1-45_HM	.	Hydra vulgaris	CR1	21,16,1,23,3,24,11,4,2
CR1-46_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-46_BF	.	Branchiostoma floridae	CR1	
CR1-46_DR	.	Danio rerio	CR1	16,1,3,4,11,2,15
CR1-46_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-47_AAe	.	Aedes aegypti	CR1	16,1,3,4,11,2,15
CR1-47_BF	.	Branchiostoma floridae	CR1	
CR1-47_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-48_AAe	.	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-48_BF	.	Branchiostoma floridae	CR1	
CR1-48_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-49_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,127,3,59
CR1-49_BF	.	Branchiostoma floridae	CR1	
CR1-49_DR	.	Danio rerio	CR1	1,3,4,2
CR1-49_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-4B_AG	.	Anopheles gambiae	CR1	1,3,58,59,4,56,57,2
CR1-4B_CPB	.	Chrysemys picta bellii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-4B_LMi	.	Locusta migratoria	CR1	
CR1-4C_CPB	.	Chrysemys picta bellii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-5_AG	.	Anopheles gambiae	CR1	54,1,3,59,53,11,4,57,2,68
CR1-5_AMi	.	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-5_BF	.	Branchiostoma floridae	CR1	
CR1-5_CBre	.	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-5_CGi	.	Crassostrea gigas	CR1	2,21,23,24,11,1,4,16,3
CR1-5_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-5_CM	.	Callorhinchus milii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-5_CPB	.	Chrysemys picta bellii	CR1	16,1,3,4,11,2,15
CR1-5_CQ	.	Culex quinquefasciatus	CR1	54,1,3,59,53,11,4,2
CR1-5_CRem	.	Caenorhabditis remanei	CR1	16,1,3,11,4,2,15
CR1-5_Croc	.	Crocodylidae	CR1	2,63,24,11,1,64,4,16,3
CR1-5_CTe	.	Capitella teleta	CR1	1,3,11,4,2
CR1-5_DEl	.	Drosophila elegans	CR1	

CR1-5_DK	.	Drosophila kikkawai	CR1	
CR1-5_DRh	.	Drosophila rhopaloa	CR1	16,1,3,11,4,57,2,15
CR1-5_DT	.	Drosophila takahashii	CR1	16,1,3,11,4,2,15
CR1-5_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-5_Hmel	.	Heliconius melpomene	CR1	
CR1-5_HMM	.	Heliconius melpomene melpomene	CR1	16,1,3,11,4,2,15
CR1-5_LCh	.	Latimeria chalumnae	CR1	16,1,3,11,4,2,15
CR1-5_LVa	.	Litopenaeus vannamei	CR1	2,142,1,14,3,11,4,12
CR1-5_MUn	.	Melopsittacus undulatus	CR1	1,3,4,2
CR1-5_NV	.	Nematostella vectensis	CR1	
CR1-5_NVi	.	Nasonia vitripennis	CR1	1,8,3,11,4,12,2
CR1-5_PH	.	Parhyale hawaiensis	CR1	16,1,3,11,4,2,15
CR1-5_PPac	.	Pristionchus pacificus	CR1	1,3,4,2
CR1-5_PSi	.	Pelodiscus sinensis	CR1	2,63,11,1,64,4,15,16,3
CR1-5_SP	.	Strongylocentrotus purpuratus	CR1	1,3,11,4,2
CR1-50_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-50_BF	.	Branchiostoma floridae	CR1	
CR1-50_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-51_AAe	.	Aedes aegypti	CR1	53,2,54,8,11,57,12,42,1,4,3,59
CR1-51_BF	.	Branchiostoma floridae	CR1	
CR1-51_DR	.	Danio rerio	CR1	1,3,4,2
CR1-52_AAe	.	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1-52_BF	.	Branchiostoma floridae	CR1	
CR1-53_AAe	.	Aedes aegypti	CR1	54,1,3,58,59,53,11,4,56,2
CR1-53_BF	.	Branchiostoma floridae	CR1	16,1,21,3,11,15,4,2
CR1-53_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-54_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-54_BF	.	Branchiostoma floridae	CR1	
CR1-54_DR	.	Danio rerio	CR1	
CR1-54_HM	.	Hydra vulgaris	CR1	21,1,23,3,4,11,2
CR1-55_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-55_BF	.	Branchiostoma floridae	CR1	
CR1-55_DR	.	Danio rerio	CR1	
CR1-55_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-56_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,57,2
CR1-56_BF	.	Branchiostoma floridae	CR1	
CR1-57_AAe	.	Aedes aegypti	CR1	54,1,8,3,59,53,11,4,12,2
CR1-58_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-58_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-59_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,2
CR1-5B_AMi	.	Alligator mississippiensis	CR1	1,3,2
CR1-5B_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-5B_CM	.	Callorhinchus milii	CR1	16,1,3,11,4,2,15
CR1-5B_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-5B_Crp	.	Crocodylus porosus	CR1	16,1,3,4,11,2,15
CR1-6_AG	.	Anopheles gambiae	CR1	53,2,54,11,56,57,1,58,4,15,16,3,60
CR1-6_BF	.	Branchiostoma floridae	CR1	
CR1-6_CBre	.	Caenorhabditis brenneri	CR1	16,1,3,11,4,2,15
CR1-6_CGi	.	Crassostrea gigas	CR1	16,1,3,24,11,4,2
CR1-6_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-6_CQ	.	Culex quinquefasciatus	CR1	1,3,11,4,2

CR1-6_CRem	Caenorhabditis remanei	CR1	1,3,4,2
CR1-6_Croc	Crocodylidae	CR1	2,63,24,11,48,1,64,4,16,3
CR1-6_CTe	Capitella teleta	CR1	16,1,3,11,4,2,15
CR1-6_DT	Drosophila takahashii	CR1	
CR1-6_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-6_Hmel	Heliconius melpomene	CR1	1,3,2
CR1-6_HMM	Heliconius melpomene melpomene	CR1	1,3,4,2
CR1-6_HRo	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-6_IVa	Litopenaeus vannamei	CR1	16,1,3,59,11,4,57,2,15
CR1-6_MUn	Melopsittacus undulatus	CR1	
CR1-6_NV	Nematostella vectensis	CR1	1,3,11,4,2
CR1-6_NVi	Nasonia vitripennis	CR1	1,3,4,2
CR1-6_PPac	Pristionchus pacificus	CR1	1,3,11,4,2
CR1-60_AAe	Aedes aegypti	CR1	16,1,23,3,11,4,2,15
CR1-60_HM	Hydra vulgaris	CR1	1,3,4,2
CR1-61_AAe	Aedes aegypti	CR1	21,1,23,3,11,4,2
CR1-61_HM	Hydra vulgaris	CR1	53,2,54,11,56,57,1,58,4,3,59
CR1-62_AAe	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-62_HM	Hydra vulgaris	CR1	16,1,3,59,11,4,2,15
CR1-63_AAe	Aedes aegypti	CR1	21,1,23,3,4,11,2
CR1-63_HM	Hydra vulgaris	CR1	54,1,3,59,53,11,4,2
CR1-64_AAe	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-64_HM	Hydra vulgaris	CR1	1,3,11,4,2
CR1-64B_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-65_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-66_AAe	Aedes aegypti	Crack	21,1,23,3,11,4,2
CR1-66_HM	Hydra vulgaris	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-67_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-67_HM	Hydra vulgaris	CR1	53,2,54,11,57,1,4,3,59,60
CR1-68_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-69_AAe	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-69_HM	Hydra vulgaris	CR1	
CR1-6B_CM	Callorhinchus milii	CR1	1,11,2
CR1-6B_CQ	Culex quinquefasciatus	CR1	
CR1-6C_AMi	Alligator mississippiensis	CR1	
CR1-6C_CM	Callorhinchus milii	CR1	2,63,11,1,64,4,15,16,3
CR1-6D_AMi	Alligator mississippiensis	CR1	2,63,24,11,48,1,64,4,16,3
CR1-6F_AMi	Anopheles gambiae	CR1	1,3,4,11,2
CR1-7_AG	Branchiostoma floridae	CR1	
CR1-7_BF	Caenorhabditis brenneri	CR1	1,3,2
CR1-7_CBre	Crassostrea gigas	CR1	1,3,4,2
CR1-7_CGi	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-7_CJap	Callorhinchus milii	CR1	
CR1-7_CM	Chrysemys picta bellii	CR1	1,3,11,4,2
CR1-7_CPB	Culex quinquefasciatus	CR1	16,1,3,4,11,2,15
CR1-7_CQ	Caenorhabditis remanei	CR1	
CR1-7_CRem	Crocodylidae	CR1	16,1,3,11,4,2,15
CR1-7_Croc	Crocodylus porosus	CR1	
CR1-7_Crp	Drosophila takahashii	CR1	1,3,4,2
CR1-7_DT	Gavialis gangeticus	CR1	
CR1-7_Gav			

CR1-7_HM	.	Hydra vulgaris	CR1	1,3,4,2
CR1-7_Hmel	.	Heliconius melpomene	CR1	
CR1-7_HMM	.	Heliconius melpomene melpomene	CR1	2,21,11,56,57,1,58,4,15,16,3,59,60
CR1-7_HRo	.	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-7_LVa	.	Litopenaeus vannamei	CR1	16,1,3,11,4,2,15
CR1-7_NV	.	Nematostella vectensis	CR1	
CR1-7_NVi	.	Nasonia vitripennis	CR1	212,1,213,8,3,11,4,12,2
CR1-7_PPac	.	Pristionchus pacificus	CR1	1,3,4,2
CR1-7_SP	.	Strongylocentrotus purpuratus	CR1	1,3,11,4,2
CR1-70_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,1,58,4,15,16,3,59
CR1-70_HM	.	Hydra vulgaris	CR1	16,1,23,3,24,11,4,2
CR1-71_AAe	.	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-71_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-72_AAe	.	Aedes aegypti	CR1	16,1,3,59,60,4,11,57,2,15
CR1-72_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-73_AAe	.	Aedes aegypti	CR1	1,3,59,11,4,57,2
CR1-74_AAe	.	Aedes aegypti	CR1	16,1,3,59,11,4,57,2,68,15
CR1-74_HM	.	Hydra vulgaris	CR1	16,1,23,3,11,4,2,15
CR1-75_AAe	.	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,127,3,59
CR1-75_HM	.	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-76_AAe	.	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1-76_HM	.	Hydra vulgaris	CR1	1,3,11,4,2
CR1-77_AAe	.	Aedes aegypti	CR1	16,1,3,4,11,2,15
CR1-77_HM	.	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-78_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-79_AAe	.	Aedes aegypti	CR1	54,1,8,3,59,53,11,4,12,2
CR1-79_HM	.	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-7B_Gav	.	Gavialis gangeticus	CR1	16,1,3,11,4,2,15
CR1-8_AG	.	Anopheles gambiae	CR1	1,3,4,2
CR1-8_AMi	.	Alligator mississippiensis	CR1	16,1,3,11,4,2,15
CR1-8_BF	.	Branchiostoma floridae	CR1	
CR1-8_CBre	.	Caenorhabditis brenneri	CR1	1,3,4,2
CR1-8_CGi	.	Crassostrea gigas	CR1	21,1,23,3,11,4,2
CR1-8_CJap	.	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-8_CM	.	Callorhinchus milii	CR1	2,63,11,48,1,64,4,15,16,3
CR1-8_CPB	.	Chrysemys picta bellii	CR1	
CR1-8_CQ	.	Culex quinquefasciatus	CR1	16,1,3,11,4,57,2,15
CR1-8_CRem	.	Caenorhabditis remanei	CR1	1,3,4,2
CR1-8_Crp	.	Crocodylus porosus	CR1	
CR1-8_CTe	.	Capitella teleta	CR1	1,3,4,2
CR1-8_DTa	.	Drosophila takahashii	CR1	1,3,11,4,2
CR1-8_Gav	.	Gavialis gangeticus	CR1	16,1,3,4,2,15
CR1-8_Hmel	.	Heliconius melpomene	CR1	1,3,11,4,2
CR1-8_HMM	.	Heliconius melpomene melpomene	CR1	1,3,55,59,11,4,57,61,2
CR1-8_HRo	.	Helobdella robusta	CR1	16,1,23,3,4,11,2,15
CR1-8_NV	.	Nematostella vectensis	CR1	
CR1-8_PPac	.	Pristionchus pacificus	CR1	1,3,4,2
CR1-8_SP	.	Strongylocentrotus purpuratus	CR1	
CR1-80_AAe	.	Aedes aegypti	CR1	1,3,11,4,2
CR1-80_HM	.	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-81_AAe	.	Aedes aegypti	CR1	1,3,11,4,2

CR1-81_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-82_AAe	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-82_HM	Hydra vulgaris	CR1	21,1,23,3,11,4,2
CR1-83_AAe	Aedes aegypti	CR1	16,1,3,11,4,2,15
CR1-83_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-84_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-84_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-85_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-85_HM	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-86_AAe	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1-86_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-87_AAe	Aedes aegypti	CR1	1,8,3,4,11,12,2
CR1-87_HM	Hydra vulgaris	CR1	1,3,4,2
CR1-88_AAe	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,59
CR1-88_HM	Hydra vulgaris	CR1	1,3,11,4,2
CR1-89_AAe	Aedes aegypti	CR1	1,69,3,59,11,4,2
CR1-89_HM	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-9_AG	Anopheles gambiae	CR1	1,54,16,3,53,11,4,57,2,15
CR1-9_AMi	Alligator mississippiensis	CR1	
CR1-9_BF	Branchiostoma floridae	CR1	
CR1-9_CGi	Crassostrea gigas	CR1	1,3,4,2
CR1-9_CJap	Caenorhabditis japonica	CR1	16,1,3,11,4,2,15
CR1-9_CM	Callorhinchus milii	CR1	
CR1-9_CPB	Chrysemys picta bellii	CR1	
CR1-9_CQ	Culex quinquefasciatus	CR1	1,3,11,4,2
CR1-9_CRem	Caenorhabditis remanei	CR1	1,3,4,2
CR1-9_Crp	Crocodylus porosus	CR1	
CR1-9_DTa	Drosophila takahashii	CR1	1,3,11,4,2
CR1-9_HM	Hydra vulgaris	Crack	1,3,4,2
CR1-9_Hmel	Heliconius melpomene	CR1	1,3,4,2
CR1-9_HRo	Helobdella robusta	CR1	16,1,3,11,4,2,15
CR1-9_NV	Nematostella vectensis	CR1	
CR1-9_NVi	Nasonia vitripennis	CR1	1,8,3,11,4,12,2
CR1-9_SP	Strongylocentrotus purpuratus	CR1	1,3,4,2
CR1-90_AAe	Aedes aegypti	CR1	53,2,54,11,56,57,68,1,58,4,3,59
CR1-91_AAe	Aedes aegypti	CR1	53,2,54,11,57,68,1,4,15,16,3,59
CR1-92_AAe	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1-93_AAe	Aedes aegypti	CR1	54,1,8,3,59,53,4,11,12,2
CR1-93_HMa	Hydra vulgaris	CR1	16,1,3,4,11,2,15
CR1-94_AAe	Aedes aegypti	CR1	1,3,4,11,2
CR1-94_HMa	Hydra vulgaris	CR1	16,1,3,11,4,2,15
CR1-95_AAe	Aedes aegypti	CR1	1,3,11,4,57,2
CR1-96_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-96_HMa	Hydra vulgaris	CR1	16,1,3,11,15,4,2
CR1-97_AAe	Aedes aegypti	CR1	54,1,3,59,53,11,4,57,2
CR1-98_AAe	Aedes aegypti	CR1	1,54,3,59,53,11,4,57,2
CR1-99_AAe	Aedes aegypti	CR1	1,3,11,4,2
CR1-9B_AMi	Alligator mississippiensis	CR1	16,1,3,4,11,2,15
CR1-9B_CM	Callorhinchus milii	CR1	
CR1-9C_AMi	Alligator mississippiensis	CR1	2,63,11,1,64,4,15,16,3
CR1-9D_AMi	Alligator mississippiensis	CR1	16,1,3,4,11,2,15

CR1-B2	.	Gallus	CR1	
CR1-C	.	Gallus	CR1	16,1,3,11,4,2,15
CR1-C2	.	Gallus	CR1	
CR1-C3	.	Gallus	CR1	
CR1-C4	.	Gallus	CR1	
CR1-D	.	Gallus	CR1	
CR1-D2	.	Gallus	CR1	
CR1-E_Pass	.	Passeriformes	CR1	
CR1-F2	.	Gallus	CR1	
CR1-G	.	Gallus	CR1	
CR1-H	.	Gallus	CR1	
CR1-H2	.	Gallus	CR1	
CR1-I_Tgu	.	Passeriformes	CR1	
CR1-J1_Pass	.	Passeriformes	CR1	
CR1-J2_Pass	.	Passeriformes	CR1	
CR1-J3_Pass	.	Passeriformes	CR1	
CR1-K1_Tgu	.	Estrildidae	CR1	
CR1-K2_Tgu	.	Estrildidae	CR1	
CR1-K3_Tgu	.	Estrildidae	CR1	
CR1-K4_Tgu	.	Estrildidae	CR1	
CR1-L1_Tgu	.	Estrildidae	CR1	
CR1-L2_Tgu	.	Estrildidae	CR1	
CR1-L2-1_XT	.	Xenopus tropicalis	L2	1,3,11,4,2
CR1-L3A_Croc	.	Crocodylidae	CR1	2,63,11,48,1,64,4,15,16,3
CR1-L3B_Croc	.	Crocodylidae	CR1	16,1,3,4,11,2,15
CR1-X1_3end	.	Gallus	CR1	
CR1-X1_Pass	.	Passeriformes	CR1	
CR1-X2_3end	.	Gallus	CR1	
CR1-X2_Pass	.	Passeriformes	CR1	
CR1-X3_Pass	.	Passeriformes	CR1	
CR1-Y	.	Gallus	CR1	
CR1-Y_Pass	.	Passeriformes	CR1	
CR1-Y1_Aves	.	Aves	CR1	
CR1-Y2	.	Gallus	CR1	
CR1-Y2_Aves	.	Aves	CR1	
CR1-Y3	.	Gallus	CR1	
CR1-Y4	.	Gallus	CR1	
CR1-YB1_Tgu	.	Passeriformes	CR1	
CR1-YB2_Pass	.	Passeriformes	CR1	
CR1-Z1_Pass	.	Passeriformes	CR1	
CR1-Z2_Pass	.	Passeriformes	CR1	
CR1AVI	.	Neognathae	CR1	
CR1L	.	Vertebrata	CR1	
CR1X_SP	.	Strongylocentrotus purpuratus	CR1	16,1,3,11,4,2,15
CR1Y_SP	.	Strongylocentrotus purpuratus	CR1	16,1,3,4,11,2,15
Crack-1_AAe	.	Aedes aegypti	Crack	1,3,60,11,4,57,2
Crack-1_AFC	.	Cichlidae	Crack	1,3,11,4,2
Crack-1_BF	.	Branchiostoma floridae	Crack	21,16,1,23,3,11,4,2,15
Crack-1_CP	.	Culex pipiens	Crack	2,11,57,1,4,15,16,3,59,60
Crack-1_CQ	.	Culex quinquefasciatus	Crack	1,3,11,4,2
Crack-1_CS1	.	Capitella sp. 1	Crack	16,21,1,23,3,11,4,2,15

Crack-1_DR	.	Danio rerio	Crack
Crack-1_EL	.	Esox lucius	Crack
Crack-1_GA	.	Gasterosteus aculeatus	Crack
Crack-1_HM	.	Hydra vulgaris	Daphne
Crack-1_IS	.	Ixodes scapularis	Crack
Crack-1_LG	.	Lottia gigantea	Crack
Crack-1_NV	.	Nematostella vectensis	Crack
Crack-1_PM	.	Petromyzon marinus	Crack
Crack-1_SP	.	Strongylocentrotus purpuratus	Crack
Crack-1_SSa	.	Salmo salar	Crack
Crack-10_AAe	.	Aedes aegypti	Crack
Crack-10_BF	.	Branchiostoma floridae	Crack
Crack-10_CQ	.	Culex quinquefasciatus	Crack
Crack-10_SSa	.	Salmo salar	Crack
Crack-11_AAe	.	Aedes aegypti	Crack
Crack-11_BF	.	Branchiostoma floridae	Crack
Crack-11_SSa	.	Salmo salar	Crack
Crack-12_AAe	.	Aedes aegypti	Crack
Crack-12_BF	.	Branchiostoma floridae	Crack
Crack-12_SSa	.	Salmo salar	Crack
Crack-13_AAe	.	Aedes aegypti	Crack
Crack-13_BF	.	Branchiostoma floridae	Crack
Crack-13_SSa	.	Salmo salar	Crack
Crack-14_AAe	.	Aedes aegypti	Crack
Crack-14_BF	.	Branchiostoma floridae	Crack
Crack-14_SSa	.	Salmo salar	Crack
Crack-15_AAe	.	Aedes aegypti	Crack
Crack-15_BF	.	Branchiostoma floridae	Crack
Crack-15_SSa	.	Salmo salar	Crack
Crack-16_AAe	.	Aedes aegypti	Crack
Crack-16_BF	.	Branchiostoma floridae	Crack
Crack-16_SSa	.	Salmo salar	Crack
Crack-17_AAe	.	Aedes aegypti	Crack
Crack-17_BF	.	Branchiostoma floridae	Crack
Crack-18_AAe	.	Aedes aegypti	Crack
Crack-18_BF	.	Branchiostoma floridae	Crack
Crack-19_AAe	.	Aedes aegypti	Crack
Crack-19_BF	.	Branchiostoma floridae	Crack
Crack-1B_PM	.	Petromyzon marinus	Crack
Crack-1B_SSa	.	Salmo salar	Crack
Crack-2_AAe	.	Aedes aegypti	Crack
Crack-2_AFC	.	Cichlidae	Crack
Crack-2_BF	.	Branchiostoma floridae	Crack
Crack-2_CP	.	Culex pipiens	Crack
Crack-2_CQ	.	Culex quinquefasciatus	Crack
Crack-2_DR	.	Danio rerio	Crack
Crack-2_EL	.	Esox lucius	Crack
Crack-2_HM	.	Hydra vulgaris	Daphne
Crack-2_NV	.	Nematostella vectensis	Crack
Crack-2_PM	.	Petromyzon marinus	Crack
Crack-2_SP	.	Strongylocentrotus purpuratus	Crack

21,1,23,3,4,2
21,1,23,3,11,4,2
21,1,23,3,11,4,2
2,21,55,11,57,1,4,15,16,3,59,61
1,3,11,4,2
53,2,21,54,11,56,57,68,1,58,4,15,16,69,3,59
16,1,3,11,4,2,15
16,1,21,3,11,4,2,15
1,3,4,11,2
1,3,11,4,2
53,2,54,11,56,57,68,1,58,4,15,16,127,3,59
1,3,2
16,1,3,11,4,2,15
16,1,3,55,60,11,15,57,4,2,61
53,2,54,11,56,57,68,1,58,4,15,16,3,59
1,3,11,4,2
2,55,24,11,57,1,4,16,3,59,61
1,3,4,2
1,3,11,4,2
1,3,4,11,2
1,3,4,2
16,1,3,11,4,2,15
1,3,11,4,57,2
1,3,4,2
16,1,3,11,4,2,15
16,1,3,59,60,11,4,57,2,15
16,1,3,11,4,2,15
1,3,4,2
16,1,3,4,11,2,15
1,3,11,4,2
16,1,3,11,4,2,15
1,3,11,4,2
53,2,54,11,56,57,68,1,58,4,15,16,3,59
1,3,11,4,2
16,1,3,11,4,2,15
1,3,55,11,4,2,61
1,3,4,2
1,3,11,4,2
16,1,3,24,11,4,2
16,1,3,11,4,2,15
16,21,1,23,3,11,4,2,15
1,3,59,60,11,4,57,2
1,3,11,4,2
1,3,4,11,2
16,21,1,23,3,11,4,2,15
1,3,11,4,2
2,1
1,3,4,2

Crack-2_SSa	Salmo salar	Crack	2,63,11,1,64,4,15,16,3
Crack-20_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-20_BF	Branchiostoma floridae	Crack	1,3,11,4,2
Crack-21_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-21_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-22_AAe	Aedes aegypti	Crack	16,1,3,11,4,2,15
Crack-22_BF	Branchiostoma floridae	Crack	1,3,11,4,2
Crack-23_AAe	Aedes aegypti	Crack	21,1,3,11,4,2
Crack-23_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-23_SP	Strongylocentrotus purpuratus	Crack	16,1,3,11,4,2,15
Crack-24_AAe	Aedes aegypti	Crack	1,3,60,11,4,57,2
Crack-24_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-25_AAe	Aedes aegypti	Crack	2,55,11,57,1,4,15,16,3,59,60,61
Crack-25_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-26_AAe	Aedes aegypti	Crack	1,3,59,11,4,57,2
Crack-26_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-27_AAe	Aedes aegypti	Crack	1,3,59,60,11,4,57,2
Crack-27_BF	Branchiostoma floridae	Crack	16,1,3,11,4,2,15
Crack-28_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-28_BF	Branchiostoma floridae	Crack	16,1,3,24,11,4,2
Crack-29_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-29_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-2B_SSa	Salmo salar	Crack	1,3,4,11,2
Crack-2C_SSa	Salmo salar	Crack	
Crack-3_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-3_AFC	Cichlidae	Crack	1,3,4,2
Crack-3_BF	Branchiostoma floridae	Daphne	21,1,23,3,11,4,2
Crack-3_CP	Culex pipiens	Crack	16,1,3,60,24,4,11,2
Crack-3_CQ	Culex quinquefasciatus	Crack	1,3,11,4,2
Crack-3_DR	Danio rerio	Crack	16,1,3,11,4,2,15
Crack-3_EL	Esox lucius	Crack	1,3,4,2
Crack-3_HM	Hydra vulgaris	Daphne	21,16,1,23,3,11,4,2,15
Crack-3_NVe	Nematostella vectensis	Crack	1,3,11,4,2
Crack-3_PM	Petromyzon marinus	Crack	
Crack-3_SP	Strongylocentrotus purpuratus	Crack	16,1,3,11,4,2,15
Crack-3_SSa	Salmo salar	Crack	1,3,11,4,2
Crack-30_AAe	Aedes aegypti	Crack	16,1,3,11,4,2,15
Crack-30_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-31_AAe	Aedes aegypti	Crack	1,3,59,11,4,57,2
Crack-31_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-32_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-32_BF	Branchiostoma floridae	Crack	1,3,4,2
Crack-33_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-34_AAe	Aedes aegypti	Crack	16,1,3,11,4,57,2,15
Crack-35_AAe	Aedes aegypti	Crack	2,11,57,1,4,15,16,3,59,60
Crack-36_AAe	Aedes aegypti	Crack	1,3,11,4,2
Crack-37_AAe	Aedes aegypti	Crack	16,1,3,11,4,2,15
Crack-4_AAe	Aedes aegypti	Crack	16,1,3,11,4,2,15
Crack-4_CP	Culex pipiens	Crack	1,3,11,4,2
Crack-4_CQ	Culex quinquefasciatus	Crack	16,1,3,24,11,4,2
Crack-4_DR	Danio rerio	Crack	

Crack-4_EL	.	Esox lucius	Crack	
Crack-4_HM	.	Hydra vulgaris	Daphne	1,3,4,2
Crack-4_PM	.	Petromyzon marinus	Crack	
Crack-4_SP	.	Strongylocentrotus purpuratus	Crack	63,1,64,3,11,4,62,2
Crack-4_SSa	.	Salmo salar	Crack	21,16,1,3,11,4,2,15
Crack-5_AAe	.	Aedes aegypti	Crack	1,3,4,11,2
Crack-5_BF	.	Branchiostoma floridae	Crack	1,3,4,2
Crack-5_CQ	.	Culex quinquefasciatus	Crack	16,1,3,59,11,4,57,2,15
Crack-5_DR	.	Danio rerio	Crack	1,4,2
Crack-5_HM	.	Hydra vulgaris	Daphne	1,3,11,4,2
Crack-5_SP	.	Strongylocentrotus purpuratus	Crack	2,11,12,1,14,4,15,16,3
Crack-5_SSa	.	Salmo salar	Crack	16,1,3,4,11,2,15
Crack-6_AAe	.	Aedes aegypti	Crack	1,3,11,4,2
Crack-6_BF	.	Branchiostoma floridae	Crack	16,1,3,11,4,2,15
Crack-6_CQ	.	Culex quinquefasciatus	Crack	1,3,4,2
Crack-6_DR	.	Danio rerio	Crack	
Crack-6_SP	.	Strongylocentrotus purpuratus	Crack	2,21,11,56,57,1,58,4,15,16,3,59
Crack-6_SSa	.	Salmo salar	Crack	53,2,21,54,23,24,11,57,1,4,16,3,60
Crack-7_AAe	.	Aedes aegypti	Crack	1,3,59,11,4,57,2
Crack-7_BF	.	Branchiostoma floridae	Crack	1,3,11,4,2
Crack-7_CQ	.	Culex quinquefasciatus	Crack	53,2,54,55,24,11,56,57,68,1,58,4,16,127,3,59,61
Crack-7_DR	.	Danio rerio	Crack	
Crack-7_HM	.	Hydra vulgaris	Daphne	2,11,56,57,68,1,58,4,127,3,59
Crack-7_SP	.	Strongylocentrotus purpuratus	Crack	2,21,11,56,57,68,1,58,4,127,3,59
Crack-7_SSa	.	Salmo salar	Crack	63,16,1,64,3,11,4,2,15
Crack-8_AAe	.	Aedes aegypti	Crack	1,3,60,11,4,57,2
Crack-8_BF	.	Branchiostoma floridae	Crack	16,1,3,11,4,2,15
Crack-8_CQ	.	Culex quinquefasciatus	Crack	1,3,11,4,2
Crack-8_HMa	.	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Crack-8_SP	.	Strongylocentrotus purpuratus	Crack	1,3,4,2
Crack-8_SSa	.	Salmo salar	Crack	16,1,3,11,4,2,15
Crack-9_AAe	.	Aedes aegypti	Crack	1,3,59,60,11,4,57,2
Crack-9_BF	.	Branchiostoma floridae	Crack	1,3,11,4,2
Crack-9_CQ	.	Culex quinquefasciatus	Crack	1,3,4,2
Crack-9_SP	.	Strongylocentrotus purpuratus	Crack	16,1,3,11,4,2,15
Crack-9_SSa	.	Salmo salar	Crack	16,1,3,11,4,2,15
CRE-1_ACas	.	Acanthamoeba castellanii	CRE	18,5,1,19,3,2
Cre-1_BM	.	Bombyx mori	CRE	18,5,1,19,3,2
CRE-1_CCri	.	Chondrus crispus	CRE	2,1
Cre-1_FCy	.	Fragilariopsis cylindrus	CRE	
CRE-1_FoCa	.	Folsomia candida	CRE	2,121,56,57,122,1,58,123,124,69,3,59,125
Cre-1_HM	.	Hydra vulgaris	CRE	1,52,2
CRE-1_HRo	.	Helobdella robusta	CRE	1,3,2
CRE-1_LSa	.	Lactuca sativa	CRE	1,3,99,2
Cre-1_MB	.	Monosiga brevicollis	CRE	163,1,3,2
Cre-1_NV	.	Nematostella vectensis	CRE	1,52,2
CRE-1_PXu	.	Papilio xuthus	CRE	
CRE-10_CCri	.	Chondrus crispus	CRE	2,1
CRE-11_CCri	.	Chondrus crispus	CRE	
CRE-12_CCri	.	Chondrus crispus	CRE	2,1

CRE-13_CCri	.	Chondrus crispus	CRE	2,1
CRE-14_CCri	.	Chondrus crispus	CRE	1,2
CRE-15_CCri	.	Chondrus crispus	CRE	
231,232,233,234				
CRE-16_CCri	.	Chondrus crispus	CRE	
CRE-17_CCri	.	Chondrus crispus	CRE	2,1
CRE-1B_FoCa	.	Folsomia candida	CRE	121,1,124,3,58,164,59,56,57,123,2,125
Cre-2_BM	.	Bombyx mori	CRE	3,2,1
CRE-2_CCri	.	Chondrus crispus	CRE	2,1
CRE-2_FoCa	.	Folsomia candida	CRE	121,1,127,124,3,58,59,56,57,2,68,125
CRE-2_HMa	.	Hydra vulgaris	CRE	
52				
CRE-2_HRo	.	Helobdella robusta	CRE	1,3,2
CRE-2_NV	.	Nematostella vectensis	CRE	
52				
CRE-2_PXu	.	Papilio xuthus	CRE	
CRE-3_CCri	.	Chondrus crispus	CRE	2,1
CRE-3_FoCa	.	Folsomia candida	CRE	121,1,127,124,3,58,59,56,57,2,68,125
CRE-3_HRo	.	Helobdella robusta	CRE	1,3,2
CRE-3_NV	.	Nematostella vectensis	CRE	
52				
CRE-4_CCri	.	Chondrus crispus	CRE	1,2
CRE-4_HRo	.	Helobdella robusta	CRE	1,3,2
CRE-5_CCri	.	Chondrus crispus	CRE	2,1
CRE-5_FoCa	.	Folsomia candida	CRE	121,1,124,3,164,2
CRE-5_HRo	.	Helobdella robusta	CRE	1,3,2
CRE-6_CCri	.	Chondrus crispus	CRE	1,3,2
CRE-6_FoCa	.	Folsomia candida	CRE	121,1,124,3,123,2,125
CRE-6_HRo	.	Helobdella robusta	CRE	1,3,2
CRE-7_CCri	.	Chondrus crispus	CRE	1,3,2
CRE-7_FoCa	.	Folsomia candida	CRE	53,2,54,56,57,68,1,58,4,127,3,59,125
CRE-8_CCri	.	Chondrus crispus	CRE	2,1
CRE-9_CCri	.	Chondrus crispus	CRE	2,1
CRE1	M33009	Crithidia fasciculata	CRE	
CRE2	U19151	Crithidia fasciculata	CRE	
CroRTE-1	.	Crocus sativus	RTE	1,26,3,11,4,2
CroSINE-10	.	Crocus sativus	SINE2/tRNA	
CroSINE-11a	.	Crocus sativus	SINE2/tRNA	
CroSINE-11b	.	Crocus sativus	SINE2/tRNA	
CroSINE-12	.	Crocus sativus	SINE2/tRNA	
CroSINE-13	.	Crocus sativus	SINE2/tRNA	
CroSINE-1a	.	Crocus sativus	SINE2/tRNA	
CroSINE-1b	.	Crocus sativus	SINE2/tRNA	
CroSINE-1c	.	Crocus sativus	SINE2/tRNA	
CroSINE-2a	.	Crocus sativus	SINE2/tRNA	
CroSINE-2b	.	Crocus sativus	SINE2/tRNA	
CroSINE-2c	.	Crocus sativus	SINE2/tRNA	
CroSINE-2d	.	Crocus sativus	SINE	
CroSINE-3a	.	Crocus sativus	SINE2/tRNA	
CroSINE-3b	.	Crocus sativus	SINE2/tRNA	
CroSINE-4a	.	Crocus sativus	SINE2/tRNA	

CroSINE-4b	.	Crocus sativus	SINE2/tRNA	
CroSINE-4c	.	Crocus sativus	SINE2/tRNA	
CroSINE-5a	.	Crocus sativus	SINE2/tRNA	
CroSINE-5b	.	Crocus sativus	SINE2/tRNA	
CroSINE-6	.	Crocus sativus	SINE2/tRNA	
CroSINE-7	.	Crocus sativus	SINE2/tRNA	
CroSINE-8	.	Crocus sativus	SINE2/tRNA	
CroSINE-9	.	Crocus sativus	SINE2/tRNA	
CRP1	.	Crocodylus porosus	SINE	
CRYI	.	Testudinoidea	SINE2/tRNA	
CryIIB	AB125437	Macrochelys temminckii	SINE2/tRNA	
CSINE1	.	Oryctolagus cuniculus	SINE	
CSINE2	.	Oryctolagus cuniculus	SINE	
CSINE2_OP	.	Ochotona princeps	SINE	
CSINE2A2_OP	.	Ochotona princeps	SINE	
CSINE2B	.	Oryctolagus cuniculus	SINE	
CSINE2B_OP	.	Ochotona princeps	SINE	
CSINE2C_OP	.	Ochotona princeps	SINE	
CSINE3_OP	.	Ochotona princeps	SINE	
CSINE3A	.	Oryctolagus cuniculus	SINE	
CSP2096	J01188	unidentified sea urchin	CR1	
Cte-Nin-DC-SINE1	.	Capitella teleta	SINE2/tRNA	
CYN-I	.	Cynocephalus	SINE	
CYN-II1	.	Cynocephalus	SINE2/tRNA	
CYN-II2	.	Cynocephalus	SINE2/tRNA	
CYN-II3	.	Cynocephalus	SINE2/tRNA	
CYN-III1	.	Cynocephalus	SINE2/tRNA	
CYN-III2	.	Cynocephalus	SINE2/tRNA	
CYN-III3	.	Cynocephalus	SINE2/tRNA	
CYN-III4	.	Cynocephalus	SINE2/tRNA	
CYN-III5	.	Cynocephalus	SINE2/tRNA	
Cynin-1_PH	.	Parhyale hawaiensis	RTE	53,2,54,11,56,57,68,1,58,4,15,16,127,3,59
Cynin-2_PH	.	Parhyale hawaiensis	RTE	1,3,59,11,4,2
CZAR	M62862	Trypanosoma cruzi	CRE	2,1
D_CH	M57436	Capra hircus	SINE2/tRNA	
DANA	L42295	Danio rerio	SINE	
Daphe-16_SPu	.	Strongylocentrotus purpuratus	Daphne	21,1,3,11,4,2
Daphne_DS	.	Darwinula stevensoni	Daphne	1,3,11,4,2
Daphne-1_AFC	.	Cichlidae	Daphne	
Daphne-1_BTa	.	Bemisia tabaci	Daphne	53,2,21,54,23,11,56,57,68,1,58,4,127,3,59
Daphne-1_CSav	.	Ciona savignyi	Daphne	1,3,4,2
Daphne-1_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-1_GA	.	Gasterosteus aculeatus	Daphne	1,3,11,4,2
Daphne-1_Hmel	.	Heliconius melpomene	Daphne	16,1,3,11,4,2,15
Daphne-1_HMM	.	Heliconius melpomene melpomene	Daphne	1,3,11,4,2
Daphne-1_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-1_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-1_LVa	.	Litopenaeus vannamei	Daphne	21,1,23,3,11,4,2
Daphne-1_NV	.	Nematostella vectensis	Daphne	
Daphne-1_OL	.	Oryzias latipes	Daphne	16,1,3,11,4,2,15
Daphne-1_PH	.	Parhyale hawaiensis	Daphne	1,3,11,4,2

Daphne-1_PPo	.	Papilio polytes	Daphne	1,3,11,4,2
Daphne-1_SK	.	Saccoglossus kowalevskii	Daphne	1,3,11,4,2
Daphne-1_TCa	.	Tribolium castaneum	Crack	16,1,3,11,4,2,15
Daphne-10_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-10_HMa	.	Hydra vulgaris	Daphne	16,1,3,11,4,2,15
Daphne-10_HMM	.	Heliconius melpomene melpomene	Daphne	1,3,4,2
Daphne-10_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-10_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-11_CTe	.	Capitella teleta	Daphne	21,16,1,3,24,11,4,2
Daphne-11_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-11_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-11_SPu	.	Strongylocentrotus purpuratus	Daphne	1,3,11,4,2
Daphne-12_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-12_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-12_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-12_SPu	.	Strongylocentrotus purpuratus	Daphne	21,1,3,11,4,2
Daphne-13_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-13_HMa	.	Hydra vulgaris	Daphne	1,3,4,11,2
Daphne-13_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-13_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,62,2,15
Daphne-13B_HRo	.	Helobdella robusta	Daphne	16,1,3,11,15,4,2
Daphne-14_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-14_HMa	.	Hydra vulgaris	Daphne	1,3,11,4,2
Daphne-14_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-14_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-14_NV	.	Nematostella vectensis	Daphne	
Daphne-14_SPu	.	Strongylocentrotus purpuratus	Daphne	2,21,23,11,1,4,15,16,3
Daphne-15_CTe	.	Capitella teleta	Daphne	1,3,11,4,2
Daphne-15_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-15_LMi	.	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-15_NVe	.	Nematostella vectensis	Daphne	16,1,3,11,4,2,15
Daphne-15_SP	.	Strongylocentrotus purpuratus	Daphne	1,3,11,4,2
Daphne-16_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-16_HMa	.	Hydra vulgaris	Daphne	1,3,11,4,2
Daphne-16_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-16_LMi	.	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-17_CTe	.	Capitella teleta	Daphne	16,1,3,24,11,4,2
Daphne-17_HRo	.	Helobdella robusta	Daphne	16,1,3,24,11,4,2
Daphne-17_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-17_SP	.	Strongylocentrotus purpuratus	Daphne	1,3,11,4,2
Daphne-18_CTe	.	Capitella teleta	Daphne	1,3,4,2
Daphne-18_HM	.	Hydra vulgaris	Daphne	21,16,1,23,3,11,4,2,15
Daphne-18_HRo	.	Helobdella robusta	Daphne	1,3,4,11,2
Daphne-18_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-18_SP	.	Strongylocentrotus purpuratus	Daphne	21,1,3,11,4,2
Daphne-19_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-19_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-19_SP	.	Strongylocentrotus purpuratus	Daphne	1,3,11,4,2
Daphne-1B_LMi	.	Locusta migratoria	Daphne	1,3,11,4,57,48,62,2
Daphne-2_AFC	.	Cichlidae	Daphne	
Daphne-2_BTa	.	Bemisia tabaci	Daphne	16,1,3,11,15,4,2

Daphne-2_CTe	.	Capitella teleta	Daphne	1,3,2
Daphne-2_Hmel	.	Heliconius melpomene	Daphne	16,1,3,11,4,2,15
Daphne-2_HMM	.	Heliconius melpomene melpomene	Daphne	1,3,11,4,2
Daphne-2_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-2_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-2_NV	.	Nematostella vectensis	Daphne	
53,2,21,54,55,11,56,57,1,58,4,15,16,3,59,60,61				
Daphne-2_OL	.	Oryzias latipes	Daphne	1,3,4,2
Daphne-2_PH	.	Parhyale hawaiensis	Daphne	
Daphne-2_PPo	.	Papilio polytes	Daphne	
Daphne-2_SK	.	Saccoglossus kowalevskii	Daphne	21,1,3,82,4,11,2
Daphne-20_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-20_LMi	.	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-21_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-21_HRo	.	Helobdella robusta	Daphne	16,1,3,24,4,11,2
Daphne-21_LMi	.	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-21_SPu	.	Strongylocentrotus purpuratus	Daphne	21,1,23,3,11,4,2
Daphne-22_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-22_LMi	.	Locusta migratoria	Daphne	1,3,4,62,2
Daphne-22_SPur	.	Strongylocentrotus purpuratus	Daphne	2,21,23,11,1,4,15,16,3
Daphne-23_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-23_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-23_SPur	.	Strongylocentrotus purpuratus	Daphne	1,3,4,2
Daphne-24_HMa	.	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-24_HRo	.	Helobdella robusta	Daphne	16,1,3,24,11,4,2
Daphne-24_LMi	.	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-24_SPur	.	Strongylocentrotus purpuratus	Daphne	16,21,1,3,11,4,2,15
Daphne-25_HRo	.	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-25_LMi	.	Locusta migratoria	Daphne	1,3,4,2
Daphne-25_SPur	.	Strongylocentrotus purpuratus	Daphne	21,1,23,3,11,4,2
Daphne-26_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-26_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,62,2,15
Daphne-26_SPur	.	Strongylocentrotus purpuratus	Daphne	1,3,11,4,2
Daphne-26B_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-27_DRe	.	Danio rerio	Daphne	1,3,2
Daphne-27_HRo	.	Helobdella robusta	Daphne	1,3,4,2
Daphne-27_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-28_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-28_LMi	.	Locusta migratoria	Daphne	1,3,4,11,62,2
Daphne-29_HMa	.	Hydra vulgaris	Daphne	1,3,11,4,2
Daphne-29_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-29_LMi	.	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-3_AFC	.	Cichlidae	Daphne	
Daphne-3_BTa	.	Bemisia tabaci	Daphne	2,1,3,11,4,62
Daphne-3_CTe	.	Capitella teleta	Daphne	1,3,2
Daphne-3_HMM	.	Heliconius melpomene melpomene	Daphne	1,3,4,11,2
Daphne-3_HRo	.	Helobdella robusta	Daphne	1,3,11,4,183,2
Daphne-3_LMi	.	Locusta migratoria	Daphne	53,62,2,54,11,57,68,1,4,15,16,3,59
Daphne-3_LVa	.	Litopenaeus vannamei	Daphne	2,21,23,11,1,4,15,16,3
Daphne-3_NV	.	Nematostella vectensis	Daphne	
Daphne-3_OL	.	Oryzias latipes	Daphne	21,1,3,11,4,2

Daphne-3_SK	Saccoglossus kowalevskii	Daphne	1,3,4,2
Daphne-30_HRo	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-30_LMi	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-31_HRo	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-31_LMi	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-32_HRo	Helobdella robusta	Daphne	16,1,3,4,11,2,15
Daphne-32_LMi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-32B_HRo	Helobdella robusta	Daphne	1,3,4,2
Daphne-33_HRo	Helobdella robusta	Daphne	16,1,3,4,11,2,15
Daphne-33_LMi	Locusta migratoria	Daphne	54,1,59,53,11,4,57,2,68
Daphne-34_HRo	Helobdella robusta	Daphne	16,1,3,24,11,4,2
Daphne-34_LMi	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-35_HRo	Helobdella robusta	Daphne	1,3,4,2
Daphne-35_LMi	Locusta migratoria	Daphne	1,3,4,2
Daphne-36_HRo	Helobdella robusta	Daphne	1,3,4,2
Daphne-36_LMi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-37_HRo	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-37_LMi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-38_HMa	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-38_HRo	Helobdella robusta	Daphne	1,3,4,2
Daphne-38_LMi	Locusta migratoria	Daphne	1,3,4,62,2
Daphne-39_HRo	Helobdella robusta	Daphne	16,1,3,59,24,11,4,57,2
Daphne-39_LMi	Locusta migratoria	Daphne	1,3,11,62,2
Daphne-4_AFC	Cichlidae	Daphne	
Daphne-4_BF	Branchiostoma floridae	Daphne	16,1,3,11,4,2,15
Daphne-4_CTe	Capitella teleta	Daphne	1,3,4,2
Daphne-4_HRo	Helobdella robusta	Daphne	2,11,57,68,1,4,15,16,3,59
Daphne-4_LMi	Locusta migratoria	Daphne	53,62,2,54,11,57,48,1,4,15,16,3,59
Daphne-4_LVa	Litopenaeus vannamei	Daphne	1,3,4,11,2
Daphne-4_NV	Nematostella vectensis	Daphne	21,1,23,3,11,4,2
Daphne-4_PH	Parhyale hawaiensis	Daphne	1,3,11,4,2
Daphne-40_HMa	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-40_HRo	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-40_LMi	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-41_HMa	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-41_LMi	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-42_LMi	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-43_LMi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-44_LMi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-45_LMi	Locusta migratoria	Daphne	1,3,4,2
Daphne-46_LMi	Locusta migratoria	Daphne	1,3,4,2
Daphne-47_LMi	Locusta migratoria	Daphne	1,3,4,2
Daphne-48_LMi	Locusta migratoria	Daphne	16,1,3,11,4,62,2,15
Daphne-49_LMi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-5_AFC	Cichlidae	Daphne	
Daphne-5_CTe	Capitella teleta	Daphne	1,3,4,2
Daphne-5_HRo	Helobdella robusta	Daphne	16,1,3,11,4,2,15
Daphne-5_LMi	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-5_LVa	Litopenaeus vannamei	Daphne	2,21,23,11,1,4,15,16,3
Daphne-50_LMi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-51_HMa	Hydra vulgaris	Daphne	21,1,23,3,11,4,30,2

Daphne-51_1Mi	Locusta migratoria	Daphne	1,54,3,60,53,11,4,57,62,2
Daphne-52_HMa	Hydra vulgaris	Daphne	21,1,23,3,11,4,2
Daphne-52_1Mi	Locusta migratoria	Daphne	16,1,3,11,15,4,2
Daphne-53_1Mi	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-54_1Mi	Locusta migratoria	Daphne	16,1,11,62,2,15
Daphne-55_1Mi	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-56_HMa	Hydra vulgaris	Daphne	21,1,23,3,11,4,30,2
Daphne-56_1Mi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-57_HMa	Hydra vulgaris	Daphne	21,16,1,23,3,11,4,2,15
Daphne-57_1Mi	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-58_1Mi	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-59_1Mi	Locusta migratoria	Daphne	1,3,4,2
Daphne-6_CTe	Capitella teleta	Daphne	1,3,4,2
Daphne-6_HMM	Heliconius melpomene melpomene	Daphne	1,3,11,4,2
Daphne-6_HRo	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-6_1Mi	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-6_1Va	Litopenaeus vannamei	Daphne	21,16,1,23,3,4,11,2,15
Daphne-6_SPu	Strongylocentrotus purpuratus	Daphne	
Daphne-60_1Mi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-61_1Mi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-62_1Mi	Locusta migratoria	Daphne	16,1,3,11,4,2,15
Daphne-64_1Mi	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-65_HMa	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-65_1Mi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-66_1Mi	Locusta migratoria	Daphne	1,3,11,4,62,2
Daphne-67_1Mi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-68_1Mi	Locusta migratoria	Daphne	
Daphne-69_1Mi	Locusta migratoria	Daphne	
Daphne-7_CTe	Capitella teleta	Daphne	1,3,4,2
Daphne-7_HMM	Heliconius melpomene melpomene	Daphne	1,3,11,4,2
Daphne-7_HRo	Helobdella robusta	Daphne	16,1,3,4,11,2,15
Daphne-7_1Mi	Locusta migratoria	Daphne	16,1,3,11,4,62,2,15
Daphne-7_1Va	Litopenaeus vannamei	Daphne	21,16,1,23,3,11,4,2,15
Daphne-70_1Mi	Locusta migratoria	Daphne	
Daphne-71_1Mi	Locusta migratoria	Daphne	
Daphne-72_1Mi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-73_HMa	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15
Daphne-73_1Mi	Locusta migratoria	Daphne	
Daphne-74_1Mi	Locusta migratoria	Daphne	1,3,4,11,2
Daphne-75_1Mi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-76_1Mi	Locusta migratoria	Daphne	
Daphne-77_1Mi	Locusta migratoria	Daphne	
Daphne-78_HMa	Hydra vulgaris	Daphne	1,54,127,3,58,59,53,4,56,57,11,2
Daphne-78_1Mi	Locusta migratoria	Daphne	
Daphne-79_1Mi	Locusta migratoria	Daphne	1,3,11,4,2
Daphne-8_CTe	Capitella teleta	Daphne	1,3,11,4,2
Daphne-8_HMa	Hydra vulgaris	Daphne	16,1,3,11,4,2,15
Daphne-8_HRo	Helobdella robusta	Daphne	1,3,11,4,2
Daphne-8_1Mi	Locusta migratoria	Daphne	16,1,3,4,11,2,15
Daphne-80_1Mi	Locusta migratoria	Daphne	
Daphne-81_1Mi	Locusta migratoria	Daphne	1,11,2

Daphne-8B_HRo	.	Helobdella robusta	Daphne	1,3,4,2	
Daphne-9_CTe	.	Capitella teleta	Daphne	1,3,4,2	
Daphne-9_HRo	.	Helobdella robusta	Daphne	1,3,11,4,2	
Daphne-9_LMi	.	Locusta migratoria	Daphne	16,1,3,11,4,48,62,2,15	
Daphne-91_HMa	.	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15	
Daphne-92_HMa	.	Hydra vulgaris	CR1	21,1,23,3,11,4,30,2	
Daphne-95_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,2	
Daphne-96_HMa	.	Hydra vulgaris	Daphne	2,21,23,24,11,1,4,30,16,3	
Daphne-97_HMa	.	Hydra vulgaris	Daphne	16,21,1,23,3,11,4,2,15	
Daphne-98_HMa	.	Hydra vulgaris	Daphne	21,1,23,3,11,4,2	
Daphne-99_HMa	.	Hydra vulgaris	Daphne	2,21,23,11,1,4,15,16,3	
DASIa	.	Dasypus	SINE2/tRNA		
DASIb	.	Dasypus	SINE2/tRNA		
DASIIa	.	Dasypus	SINE2/tRNA		
DASIIb	.	Dasypus	SINE2/tRNA		
DASIIIb1	.	Dasypus	SINE2/tRNA		
DASIIIb1a	.	Dasypus	SINE2/tRNA		
DASIIIb2	.	Dasypus	SINE2/tRNA		
DASIIIb3	.	Dasypus	SINE2/tRNA		
Deceiver-1_CCO	supercontig_2.42	Coprinopsis cinerea	Non-LTRRetrotransposo	16,1,3,24,4,11,2	
Deceiver-1_MLP	.	Melampsora larici-populina	Non-LTRRetrotransposo	1,3,11,4,2	
Deceiver-2_MLP	.	Melampsora larici-populina	Non-LTRRetrotransposo	1,3,11,4,2	
Deceiver-3_MLP	.	Melampsora larici-populina	Non-LTRRetrotransposo	1,3,11,4,2	
DeuSINE	.	Metazoa Eukaryota.	SINE3/5S		
DINOLT1	NW_633168	Dictyostelium discoideum	L1	2,1,3,11,4	
DIP	.	Allactaga	SINE2/tRNA		
DIP0	.	Dipodidae	SINE2/tRNA		
DIPODE1	.	Dipodomys ordii	SINE2/tRNA		
DIPODE2	.	Dipodomys ordii	SINE2/tRNA		
DIPODE3	.	Dipodomys ordii	SINE1/7SL		
DIPODE4	.	Dipodomys ordii	SINE1/7SL		
D1_TRE3-A	LODT01000000	Dictyostelium lacteum	L1	16,1,3,11,4,2,15	
D1_TRE5-A	LODT01000000	Dictyostelium lacteum	Non-LTRRetrotransposo	1,3,11,4,2	
DMCR1A	.	Drosophila melanogaster	CR1	16,1,3,11,4,2,15	
DMRT1A	.	Drosophila melanogaster	R1	2,8,11,86,12,1,14,4,15,16,3	
DMRT1B	.	Drosophila melanogaster	R1	2,8,11,86,12,1,14,4,15,16,3	
DMRT1C	.	Drosophila melanogaster	R1	16,1,3,11,4,2,15	
DOC	X17551	Drosophila melanogaster	Jockey	87,2,11,1,14,4,15,16,3,88	
DOC2_DM	.	Drosophila melanogaster	Jockey	1,16,87,3,88,11,4,2,15	
DOC3_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15	
DOC4_DM	.	Drosophila melanogaster	Jockey	1,87,88,11,2	
DOC5_DM	.	Drosophila melanogaster	Jockey	16,1,87,88,3,11,4,2,15	
DOC6_DM	.	Drosophila melanogaster	Jockey	16,1,3,11,4,2,15	
Dong	.	Bombyx mori	R4	1,3,4,2	
DONG_FR2	.	Takifugu rubripes	R4		
Dong-1_ADi	.	Acropora digitifera	R4	1,3,4,2	
Dong-1_AFC	.	Cichlidae	R4		
Dong-1_HMM	.	Heliconius melpomene melpomene	R4	1,3,4,2	
Dong-1_MMa	.	Mesobuthus martensii	Loa	1,3,4,2	
Dong-1_NVe	.	Nematostella vectensis	R4	1,3,4,2	
Dong-1_PPo	.	Papilio polytes	R4		

Dong-1_PXu	.	Papilio xuthus	R4	
Dong-1_TCas	.	Tribolium castaneum	R4	1,3,4,2
Dong-2_AD1	.	Acropora digitifera	R4	1,3,2
Dong-2_BM	.	Bombyx mori	R4	1,3,4,2
Dong-2_HMM	.	Heliconius melpomene melpomene	R4	1,3,4,2
Dong-2_Lch	.	Latimeria chalumnae	R4	1,3,4,2
Dong-2_PPo	.	Papilio polytes	R4	
Dong-3_AD1	.	Acropora digitifera	R4	1,3,4,2
DongAa	.	Aedes aegypti	R4	1,3,4,2
DongAG	AB097127	Anopheles gambiae	R4	1,3,4,2
Doro-1_PH	.	Parhyale hawaiensis	CR1	1,16,176,3,11,4,177,2,15
Doro-2_PH	.	Parhyale hawaiensis	CR1	1,3,4,11,2
Doro-3_PH	.	Parhyale hawaiensis	CR1	16,1,3,4,11,2,15
Doro-4_PH	.	Parhyale hawaiensis	CR1	16,1,3,4,11,2,15
Doro-4B_PH	.	Parhyale hawaiensis	CR1	
Doro-5_PH	.	Parhyale hawaiensis	CR1	
Dp_NLTR-A	.	Dictyostelium purpureum	Non-LTRRetrotransposo	6,2,10,11,13,1,4,100,17,3,7
Dp_TRE3-A	.	Dictyostelium purpureum	Non-LTRRetrotransposo	1,3,11,4,2
Dp_TRE3-B	.	Dictyostelium purpureum	Non-LTRRetrotransposo	1,3,11,4,2
DRE	X57034	Dictyostelium discoideum	L1	1,3,11,4,2
DualenCr4	.	Chlamydomonas reinhardtii	RandI	34,35,6,2,33,10,11,13,1,32,4,17,3,7
EbuSINE1	.	Eptatretus burgeri	SINE3/5S	
EbuSINE2	.	Eptatretus burgeri	SINE3/5S	
EdSINE1	.	Entamoeba dispar	SINE	
EGLN1_SM	AB016142	Solanum melongena	L1	
EGLN1_SM	AB016142	Solanum melongena	L1	
EGLN2_SM	AB016143	Solanum melongena	L1	
EGLN2_SM	AB016143	Solanum melongena	L1	
EGLN3_SM	AB016144	Solanum melongena	L1	
EGLN3_SM	AB016144	Solanum melongena	L1	
EGRT1	X86077	Blumeria graminis	Non-LTRRetrotransposo	
EHAPT2	.	Entamoeba histolytica	SINE	
EhRLE2	AB097128	Entamoeba histolytica	R4	1,3,4,2
EhRLE3	AB097129	Entamoeba histolytica	R4	
EptSINE1	.	Eptatretus burgeri	SINE2/tRNA	
ERE1	D26566	Equus caballus	SINE2/tRNA	
ERE1B	.	Equus caballus	SINE2/tRNA	
ERE1C	.	Equus caballus	SINE2/tRNA	
ERE2	.	Equus	SINE2/tRNA	
ERE3	.	Equus caballus	SINE2/tRNA	
ERE3B	.	Equus caballus	SINE2/tRNA	
ERE3B2	.	Equus caballus	SINE2/tRNA	
ERE3C	.	Equus caballus	SINE2/tRNA	
ERE3C2	.	Equus caballus	SINE2/tRNA	
ERE3C3	.	Equus caballus	SINE2/tRNA	
ERE3C4	.	Equus caballus	SINE2/tRNA	
ERE3D	.	Equus caballus	SINE2/tRNA	
ERE4	.	Equus caballus	SINE2/tRNA	
ERE4B	.	Equus caballus	SINE2/tRNA	
Eri-1	AF195903AF195904AF19	Mesechinus dauuricus	SINE2/tRNA	
Eri-2	AF195910, AF195911,	Mesechinus dauuricus	SINE2/tRNA	

Eri-2A_EE	.	Erinaceus europaeus	SINE2/tRNA		
Eri-2B_EE	.	Erinaceus europaeus	SINE2/tRNA		
Eri-2B2_EE	.	Erinaceus europaeus	SINE2/tRNA		
Eri-2C_EE	.	Erinaceus europaeus	SINE2/tRNA		
ERI1	.	Erinaceus	SINE2/tRNA		
Erine1	.	Erinaceus europaeus	SINE2/tRNA		
Erine1b	.	Erinaceus europaeus	SINE2/tRNA		
Erine1c	.	Erinaceus europaeus	SINE2/tRNA		
Erine2a	.	Erinaceus europaeus	SINE2/tRNA		
Erine2b	.	Erinaceus europaeus	SINE2/tRNA		
Erine2c	.	Erinaceus europaeus	SINE2/tRNA		
Expander	.	Takifugu rubripes	RTE	1,26,16,3,11,4,2,15	
Expander1_Cis	.	Ciona savignyi	RTE		
EXPANDER1_DR	.	Danio rerio	RTE	1,26,16,3,11,4,2,15	
EXPANDER2	.	Takifugu rubripes	RTE	16,1,26,49,3,11,4,50,2,15	
F524	.	Oryza sativa	SINE2/tRNA		
FAM	.	Primates	SINE1/7SL		
FEILAI_AA	.	Aedes aegypti	SINE2/tRNA		
FEILAI_B	.	Aedes aegypti	SINE2/tRNA		
FEILAI-1B_AAe	.	Aedes aegypti	SINE2/tRNA		
FLAM	.	Primates	SINE1/7SL		
FMLN1_SI	AB016110	Setaria italica	L1		
FMLN2_SI	AB016111	Setaria italica	L1		
FMLN3_SI	AB016112	Setaria italica	L1		
FRAM	.	Primates	SINE1/7SL		
FW	.	Drosophila melanogaster	Jockey	16,1,87,3,88,24,11,4,2	
FW2_DM	.	Drosophila melanogaster	Jockey	16,1,87,14,3,88,11,4,2,15	
FW3_DM	.	Drosophila melanogaster	Jockey	1,87,3,88,11,4,2	
G_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15	
G2_DM	AC006215	Drosophila melanogaster	Jockey	16,1,3,11,4,2,15	
G3_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15	
G4_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15	
G5_DM	.	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15	
G5A_DM	.	Drosophila melanogaster	Jockey	16,1,3,11,4,2,15	
G6_DM	.	Drosophila melanogaster	Jockey	87,88	
G7_DM	AC003788	Drosophila melanogaster	Jockey		
GAG_CA	AJ411828	Cicer arietinum	Non-LTRRetrotransposo		
GAMERA	.	Oryzias latipes	Non-LTRRetrotransposo		
GarnAlu1	.	Otolemur garnettii	SINE2/tRNA		
GarnAlu2	.	Otolemur garnettii	SINE2/tRNA		
GarnAlu3	.	Otolemur garnettii	SINE2/tRNA		
GarnAlu3B	.	Otolemur garnettii	SINE2/tRNA		
Garnel1	.	Otolemur garnettii	SINE2/tRNA		
Gecko	.	Aedes aegypti	SINE2/tRNA		
GENIE1_GI	.	Giardia intestinalis	Genie		
GilD	.	Giardia intestinalis	Non-LTRRetrotransposo		
GilM	.	Giardia intestinalis	Non-LTRRetrotransposo	18,5,1,3,4,2	
GymnSINE	.	Gymnopithys rufigula	SINE3/5S		
HAL1	.	Eutheria	L1		
HAL1	.	Eutheria	L1		
HAL1_SS	.	Sus scrofa	L1	31,25,21,20,23,27,28,30	

HAL1-1_Cja	.	Callithrix jacchus	L1	31,25,21,20,23,27,28,30
HAL1-1_Eeu	.	Erinaceus europaeus	L1	
HAL1-1_NS	.	Nicotiana sylvestris	L1	75,76,77
HAL1-1_Opr	.	Ochotona princeps	L1	
HAL1-1A_Cho	.	Choloepus hoffmanni	L1	25,21,20,23,28,30
HAL1-1A_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1A_ML	.	Myotis lucifugus	L1	25,21,20,23,28,30
HAL1-1A_Opr	.	Ochotona princeps	L1	
HAL1-1B_Cho	.	Choloepus hoffmanni	L1	25,21,20,23,28,30
HAL1-1B_Cja	.	Callithrix jacchus	L1	
HAL1-1B_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1B_ML	.	Myotis lucifugus	L1	
HAL1-1B_Tbel	.	Tupaia belangeri	L1	
HAL1-1C_Cja	.	Callithrix jacchus	L1	
HAL1-1C_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1C_Tbel	.	Tupaia belangeri	L1	
HAL1-1D_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1D_Tbel	.	Tupaia belangeri	L1	
HAL1-1E_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1E_ML	.	Myotis lucifugus	L1	
HAL1-1E_Tbel	.	Tupaia belangeri	L1	
HAL1-1F_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1G_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-1H_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-2_Eeu	.	Erinaceus europaeus	L1	
HAL1-2_ML	.	Myotis lucifugus	L1	
HAL1-2A_Cho	.	Choloepus hoffmanni	L1	
HAL1-2A_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-2B_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-2C_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-3_Cho	.	Choloepus hoffmanni	L1	
HAL1-3_ML	.	Myotis lucifugus	L1	
HAL1-3_PVa	.	Pteropus vampyrus	L1	25,21,20,23,28,30
HAL1-3A_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-3B_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-3C_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-3D_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-4_Cho	.	Choloepus hoffmanni	L1	
HAL1-4_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30
HAL1-4A_Cpo	.	Cavia porcellus	L1	
HAL1-4B_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30,36
HAL1-4C_Cpo	.	Cavia porcellus	L1	25,21,20,23,28,30,36
HAL1-5_Cho	.	Choloepus hoffmanni	L1	25,21,20,23,28,30
HAL1B	.	Homo sapiens	L1	
HAL1B	.	Mammalia	L1	25,37,21,38,23,20,28,30
HAL1M8	.	Mammalia	L1	25,37,21,38,23,20,28,30
HAL1M8	.	Mammalia	L1	25,37,21,38,23,20,28,30
HalFs1	.	Forficula scudderi	R1	2,8,11,132,12,86,42,131,1,14,4,15,16,3
HalSINE1	.	Haliotis discus	SINE2/tRNA	
HaSE1	.	Helicoverpa armigera	SINE2/tRNA	

HaSE2_DP	.	Danaus plexippus	SINE2/tRNA	
HaSE3	.	Helicoverpa armigera	SINE3/5S	
HaTad1-1_AB	.	Agaricus bisporus	Tad1	
HaTad1-1_BG	.	Blumeria graminis	Tad1	
HaTad1-1_BoBo	.	Botryobasidium botryosum	Tad1	
HaTad1-2_AB	.	Agaricus bisporus	Tad1	
HaTad1-2_BG	.	Blumeria graminis	Tad1	
HaTad1-2_BoBo	.	Botryobasidium botryosum	Tad1	
HaTad1-3_BG	.	Blumeria graminis	Tad1	
HaTad1-3_BoBo	.	Botryobasidium botryosum	Tad1	
HaTad1-4_BG	.	Blumeria graminis	Tad1	
HaTad1-4_BoBo	.	Botryobasidium botryosum	Tad1	
HaTad1-5_BG	.	Blumeria graminis	Tad1	
HE1_DR1	.	Danio rerio	SINE2/tRNA	
HE1_HJ	.	Heterodontus japonicus	SINE2/tRNA	
HE1_MM	.	Mustelus manazo	SINE2/tRNA	
HE1_SINE	.	Mustelus manazo	SINE2/tRNA	
HE2_DR	.	Danio rerio	SINE2/tRNA	
Hebe_Av	.	Adineta vaga	Jockey	16,1,3,4,11,2,15
HELENA	.	Drosophila virilis	Jockey	
Helena_Ds	chr3R	Drosophila simulans	Non-LTRRetrotransposo	16,1,87,3,88,11,4,2,15
HELENA_RT	.	Drosophila melanogaster	Jockey	
HER_LINE	.	Scyliorhinus torazame	CR1	1,3,4,2
HER1_HJ	AB027741	Heterodontus japonicus	Non-LTRRetrotransposo	
HER1_ST	.	Scyliorhinus torazame	Non-LTRRetrotransposo	
HERO-1_AC	.	Aplysia californica	Hero	1,3,4,2
HERO-1_AFC	.	Cichlidae	Hero	
HERO-1_BF	.	Branchiostoma floridae	Hero	1,3,4,2
HERO-1_HR	.	Helobdella robusta	Hero	1,3,4,2
HERO-1_PP	.	Physarum polycephalum	Hero	1,3,4,2
HERO-1_SP	AAGJ02121261	Strongylocentrotus purpuratus	Hero	1,3,4,2
HERO-2_BF	.	Branchiostoma floridae	Hero	1,3,4,2
HERO-2_DR	.	Danio rerio	Hero	1,3,4,2
HERO-2_HR	.	Helobdella robusta	Hero	1,3,4,2
Hero-2_SPur	048B05	Strongylocentrotus purpuratus	Hero	1,3,4,2
HERO-3_BF	.	Branchiostoma floridae	Hero	1,3,4,2
HERO-3_DR	.	Danio rerio	Hero	44,43
HERO-3_HR	.	Helobdella robusta	Hero	1,3,4,2
Hero-3_SPur	.	Strongylocentrotus purpuratus	Hero	1,3,4,2
HERO-4_DR	.	Danio rerio	Hero	
HERO-4_HR	.	Helobdella robusta	Hero	1,3,4,2
HERO-5_HR	.	Helobdella robusta	Hero	1,3,4,2
HERO-6_HR	.	Helobdella robusta	Hero	1,3,4,2
HERO-7_HR	.	Helobdella robusta	Hero	1,4,2
HERO-8_HR	.	Helobdella robusta	Hero	1,3,4,2
HERO-9_HR	.	Helobdella robusta	Hero	1,3,4,2
HERODr	.	Danio rerio	Hero	1,3,4,2
HEROFr	.	Takifugu rubripes	Hero	1,3,4,2
HEROTn	.	Tetraodon nigroviridis	Hero	1,3,4,2
HET-A_DV	.	Drosophila virilis	Jockey	
HeT-A2_DVi	.	Drosophila virilis	Jockey	87,88

HeT-A3_DVi	.	Drosophila virilis	Jockey	87,88
HETA	U06920	Drosophila melanogaster	Jockey	87,88
HETA_DSi	.	Drosophila simulans	Jockey	87,88
HidaAg1	AB090822	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3
HOLN1_EA	AB0116148	Equisetum arvense	L1	
HOLN2_EA	AB0116149	Equisetum arvense	L1	
HOLN3_EA	AB0116150	Equisetum arvense	L1	
HOPEBm2	.	Bombyx mori	R1	2,8,11,86,12,1,14,4,15,16,3
HPAI	.	Oncorhynchus nerka	SINE2/tRNA	
HydSINE1	.	Hydractinia symbiolongicarpus	SINE2/tRNA	
I_Av	.	Adineta vaga	Penelope	1,3,48,62
I_DM	.	Drosophila melanogaster	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I_Ele10	.	Aedes aegypti	I	6,2,7,9,10,11,1,4,15,16,17,3
I_Ele11	.	Aedes aegypti	I	6,2,10,11,13,1,4,15,16,17,3,7
I_Ele13	.	Aedes aegypti	I	6,2,7,8,9,10,11,86,12,13,1,186,4,15,16,17,3
I_Ele14	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele15	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele16	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele17	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,1,4,15,16,17,3
I_Ele18	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele19	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele2	.	Aedes aegypti	I	6,2,7,9,8,10,11,12,13,1,14,4,15,16,17,3
I_Ele20	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
I_Ele22	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I_Ele24	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I_Ele25	.	Aedes aegypti	I	6,2,7,9,10,11,1,4,15,16,17,3
I_Ele28	.	Aedes aegypti	I	6,2,7,9,10,11,1,4,15,16,17,3
I_Ele33	.	Aedes aegypti	I	6,2,7,8,10,11,12,1,4,15,16,17,3
I_Ele34	.	Aedes aegypti	I	16,1,8,3,11,4,12,2,15
I_Ele35	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I_Ele36	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I_Ele37	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I_Ele39	.	Aedes aegypti	I	6,2,7,9,8,10,11,12,1,4,15,16,17,3
I_Ele4	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I_Ele42	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele43	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele43B_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele43C_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I_Ele4B_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4C_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4D_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4E_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4F_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele4G_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I_Ele7	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I_Ele8	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
I_Ele9	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-1_AA	.	Aedes aegypti	Nimb	6,2,7,9,10,11,12,13,1,14,4,15,16,17,3
I-1_AN	.	Aspergillus nidulans	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_AO	.	Aspergillus oryzae	I	
I-1_BM	.	Bombyx mori	Nimb	6,2,9,8,10,11,12,13,42,1,14,4,15,16,17,3,7

I-1_BTa	.	Bemisia tabaci	I	6,2,7,9,10,11,1,4,15,16,17,3
I-1_CH	.	Bipolaris maydis	I	
I-1_CI	.	Ciona intestinalis	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
I-1_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,13,1,4,17,3,7
I-1_CTe	.	Capitella teleta	I	16,1,8,14,3,11,4,12,2,15
I-1_DAn	.	Drosophila ananassae	I	6,2,9,10,11,86,1,14,4,15,16,17,3,7
I-1_DBi	.	Drosophila biarmipes	I	
I-1_DBp	.	Drosophila bipectinata	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_DEl	.	Drosophila elegans	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_DEu	.	Drosophila eugracilis	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_DF	.	Drosophila ficusphila	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I-1_DK	.	Drosophila kikkawai	I	6,2,9,10,11,86,1,4,15,16,17,3,7
I-1_DP	.	Daphnia pulex	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_DPe	.	Drosophila persimilis	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I-1_DR	AL672145	Danio rerio	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I-1_DSer	.	Drosophila serrata	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-1_DT	.	Drosophila takahashii	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-1_DVi	.	Drosophila virilis	I	
I-1_DWi	.	Drosophila willistoni	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_DYa	.	Drosophila yakuba	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_Hmel	.	Heliconius melpomene	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_HMM	.	Heliconius melpomene melpomene	I	7,1,17,9,3,10,4,6,2,13
I-1_LHu	.	Linepithema humile	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-1_LMi	.	Locusta migratoria	I	6,2,9,8,10,11,12,1,4,15,16,17,3,7
I-1_NVi	AAZX01003261.1	Nasonia vitripennis	I	6,2,9,10,11,13,1,15,16,17,3,7
I-1_PBa	.	Pogonomyrmex barbatus	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-1_PH	.	Parhyale hawaiensis	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-1_PPo	.	Papilio polytes	I	
I-1_RPr	.	Rhodnius prolixus	I	2,8,11,12,1,14,4,15,16,3
I-1_SP	.	Strongylocentrotus purpuratus	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-10_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-10_DBi	.	Drosophila biarmipes	I	17,9,10,6,7,13
I-10_LMi	.	Locusta migratoria	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-10B_LMi	.	Locusta migratoria	I	
I-11_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-11_LMi	.	Locusta migratoria	I	7,16,1,8,3,11,4,12,2,15
I-12_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-13_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-14_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-15_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-2_AO	.	Aspergillus oryzae	I	
I-2_BF	.	Branchiostoma floridae	I	
53,2,54,8,11,56,57,12,68,1,14,58,4,15,16,69,59				
I-2_BM	.	Bombyx mori	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-2_BTa	.	Bemisia tabaci	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_CH	.	Bipolaris maydis	I	
I-2_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_DAn	.	Drosophila ananassae	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_DBi	.	Drosophila biarmipes	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-2_DBp	.	Drosophila bipectinata	I	6,2,9,10,11,86,1,4,15,16,17,3,7
I-2_DEl	.	Drosophila elegans	I	6,2,9,10,11,13,1,4,15,16,17,3,7

I-2_DEu	.	Drosophila eugracilis	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I-2_DF	.	Drosophila ficusphila	I	
I-2_DK	.	Drosophila kikkawai	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_DVi	.	Drosophila virilis	I	6,2,9,10,11,86,1,4,15,16,17,3,7
I-2_DWi	.	Drosophila willistoni	I	6,2,9,10,11,86,13,1,4,15,16,17,3,7
I-2_DYa	.	Drosophila yakuba	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-2_LMi	.	Locusta migratoria	I	6,2,9,10,11,12,13,1,4,15,16,17,3,7
I-2_NVi	.	Nasonia vitripennis	I	2,8,11,12,1,14,4,15,16,3
I-2_PPo	.	Papilio polytes	I	
I-2_RPr	.	Rhodnius prolixus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-2_SPur	.	Strongylocentrotus purpuratus	I	6,2,9,10,11,13,1,14,4,15,16,17,3,7
I-2B_HMM	.	Heliconius melpomene melpomene	I	1,3,4,2
I-3_AC	.	Aplysia californica	Nimb	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
I-3_AO	.	Aspergillus oryzae	Tad1	
I-3_BTa	.	Bemisia tabaci	I	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
I-3_CQ	.	Culex quinquefasciatus	I	6,2,7,9,10,11,12,13,1,4,15,16,17,3
I-3_DAn	.	Drosophila ananassae	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-3_DBi	.	Drosophila biarmipes	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-3_DBp	.	Drosophila bipectinata	I	7,1,17,9,3,10,6,2,13
I-3_DEu	.	Drosophila eugracilis	I	6,2,7,9,10,11,86,13,1,4,15,16,17,3
I-3_DF	.	Drosophila ficusphila	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-3_DK	.	Drosophila kikkawai	I	7,1,17,9,3,10,4,6,2,13
I-3_DR	.	Danio rerio	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
I-3_DYa	.	Drosophila yakuba	I	7,1,17,16,3,11,4,2,15
I-3_HMM	.	Heliconius melpomene melpomene	I	7,1,17,9,3,10,4,6,2,13
I-3_LMi	.	Locusta migratoria	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-3_PPo	.	Papilio polytes	I	17,10,6,7
I-3_RPr	.	Rhodnius prolixus	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-3_SPur	.	Strongylocentrotus purpuratus	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
I-3B_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,12,13,1,4,15,16,17,3,7
I-4_AC	.	Aplysia californica	Nimb	6,2,7,9,10,11,86,12,1,14,4,15,16,17,3
I-4_AO	.	Aspergillus oryzae	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
I-4_BTa	.	Bemisia tabaci	I	6,2,7,9,10,11,12,13,1,14,4,15,16,17,3
I-4_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-4_DAn	.	Drosophila ananassae	I	7,1,17,9,3,10,4,6,2,13
I-4_DBi	.	Drosophila biarmipes	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-4_DBp	.	Drosophila bipectinata	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-4_DEu	.	Drosophila eugracilis	I	
I-4_DF	.	Drosophila ficusphila	I	
I-4_DK	.	Drosophila kikkawai	I	7,1,17,9,3,10,4,6,2,13
I-4_DR	.	Danio rerio	Nimb	17,9,10,6,7,13
I-4_DYa	.	Drosophila yakuba	I	6,2,9,10,24,11,13,1,4,16,17,3,7
I-4_LMi	.	Locusta migratoria	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-4_RPr	.	Rhodnius prolixus	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-44_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-45_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-48_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,1,4,15,16,17,3
I-49_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-5_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-5_AO	.	Aspergillus oryzae	Tad1	2,8,11,12,42,1,14,4,15,16,17,3
I-5_BTa	.	Bemisia tabaci	I	6,2,7,9,10,11,13,1,4,15,16,17,3

I-5_CQ	.	Culex quinquefasciatus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-5_DBi	.	Drosophila biarmipes	I	6,2,9,10,11,13,1,14,4,15,16,17,3,7
I-5_DBp	.	Drosophila bipectinata	I	7,1,17,9,3,10,4,6,2,13
I-5_DF	.	Drosophila ficusphila	I	
I-5_DK	.	Drosophila kikkawai	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-5_LMi	.	Locusta migratoria	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-5_RPr	.	Rhodnius prolixus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-50_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-51_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-52_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-53_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-54_AAe	.	Aedes aegypti	I	16,1,3,11,2,15
I-55_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,178,13,1,179,4,15,16,17,3
I-56_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-57_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-58_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-59_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-5B_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-6_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-6_AO	.	Aspergillus oryzae	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
I-6_BTa	.	Bemisia tabaci	I	6,2,7,9,10,11,13,1,15,16,17,3
I-6_DBi	.	Drosophila biarmipes	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-6_DBp	.	Drosophila bipectinata	I	7,1,17,9,3,10,4,6,2,13
I-6_DF	.	Drosophila ficusphila	I	
I-6_DY	.	Drosophila yakuba	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-6_LMi	.	Locusta migratoria	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-6_RPr	.	Rhodnius prolixus	I	7,1,17,9,3,10,4,2,6,13
I-60_AAe	.	Aedes aegypti	I	16,11,15
I-61_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-62_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-63_AAe	.	Aedes aegypti	I	6,2,9,10,11,1,4,15,16,17,3,7
I-64_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,1,4,15,16,17,3
I-65_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-66_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-67_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-68_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,1,14,4,15,16,17,3
I-69_AAe	.	Aedes aegypti	I	6,2,7,9,8,10,11,12,13,1,4,15,16,17,3
I-7_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7_BTa	.	Bemisia tabaci	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7_DBi	.	Drosophila biarmipes	I	6,2,7,9,10,11,13,1,14,4,15,16,17,3
I-7_DBp	.	Drosophila bipectinata	I	
I-7_LMi	.	Locusta migratoria	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7_RPr	.	Rhodnius prolixus	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-70_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,1,4,15,16,17,3
I-71_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-72_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-73_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-74_AAe	.	Aedes aegypti	I	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
I-75_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
I-76_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-77_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7

I-78_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-79_AAe	.	Aedes aegypti	I	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
I-7B_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7C_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-7D_AAe	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
I-8_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-8_BTa	.	Bemisia tabaci	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-8_DBi	.	Drosophila biarmipes	I	
I-8_DBp	.	Drosophila bipectinata	I	6,2,9,10,11,1,4,15,16,17,3,7
I-8_LMi	.	Locusta migratoria	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-80_AAe	.	Aedes aegypti	I	6,2,10,11,1,4,15,16,17,3,7
I-80B_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-81_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-9_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-9_LMi	.	Locusta migratoria	I	7,1,17,9,3,10,11,6,2,13
I-9B_AAe	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
I-N1_CQ	.	Culex quinquefasciatus	I	
12,207,208,209,210,200				
I1_As	.	Adineta	Penelope	1,3,62
I2_As	.	Adineta	Penelope	1,3,62
ID_B1	.	Muridae	SINE2/tRNA	
ID_Rn1	.	Rattus	SINE2/tRNA	
ID_Rn2	.	Rattus	SINE2/tRNA	
ID-B1_Cpo	.	Cavia porcellus	SINE2/tRNA	
ID-Spe	.	Spermophilus	SINE2/tRNA	
ID4_	.	Muridae	SINE2/tRNA	
IdioSINE1	.	Idiosepius paradoxus	SINE2/tRNA	
IdioSINE2	.	Idiosepius paradoxus	SINE2/tRNA	
IDL-Geo_Cca	DQ206804DQ206805DQ20	Chaetodipus californicus	SINE2/tRNA	
IDL-Geo_Tbo	AJ294708	Thomomys bottae	SINE2/tRNA	
ILRC_GL	.	Ocyphaps lophotes	CR1	
ILRC2_GL	.	Ocyphaps lophotes	Non-LTRRetrotransposo	
IN25	.	Homo sapiens	L1	
Ingi	.	Trypanosoma brucei	Ingi	6,2,18,7,9,10,11,13,1,4,143,15,5,16,19,17,3
Ingi-1_AC	.	Aplysia californica	Ingi	6,2,7,9,10,11,57,13,1,4,15,16,17,3,59
Ingi-1_CCri	.	Chondrus crispus	Ingi	6,2,10,24,11,1,4,16,17,3,7
Ingi-1_HMi	.	Hymenolepis microstoma	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7
Ingi-1_LMi	.	Locusta migratoria	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7
Ingi-1_LVa	.	Litopenaeus vannamei	Ingi	6,2,9,10,11,13,1,4,17,3,7
Ingi-1_Rpro	ACPB01056162.1	Rhodnius prolixus	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7
Ingi-2_AC	.	Aplysia californica	Ingi	6,2,7,9,10,11,13,1,4,15,16,17,3
Ingi-2_LMi	.	Locusta migratoria	Ingi	6,2,9,10,11,13,1,4,17,3,7
Ingi-2_LVa	.	Litopenaeus vannamei	Ingi	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
Ingi-2_RPr	.	Rhodnius prolixus	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7
Ingi-3_AC	.	Aplysia californica	Ingi	6,2,9,10,24,11,13,1,4,16,17,3,7
Ingi-3_LVa	.	Litopenaeus vannamei	Ingi	
Ingi-4_AC	.	Aplysia californica	Ingi	6,2,7,9,10,11,13,1,4,15,16,17,3
Ingi-4_LVa	.	Litopenaeus vannamei	Ingi	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Ingi-5_AC	.	Aplysia californica	Ingi	6,2,7,9,10,11,13,1,4,15,16,17,3
Ingi-6_AC	.	Aplysia californica	Ingi	16,1,3,4,11,2,15
Ingi-7_AC	.	Aplysia californica	Ingi	6,2,9,10,11,13,1,4,15,16,17,3,7

Ingi-8_AC	.	Aplysia californica	Ingi	
53,6,2,7,54,9,10,11,56,57,68,13,1,58,4,15,16,17,3,59				
Ingi2	.	Trypanosoma brucei	Ingi	
105,6,2,18,7,9,10,11,13,1,4,143,15,5,16,19,17,3				
Inkcap-1_CCi	.	Coprinopsis cinerea	RTEX	3,2,1
Inkcap-2_CCi	AACS02000013	Coprinopsis cinerea	RTEX	240,7,1,17,3,10,4,11,6,2,13
ISABELLE_TM	AF326781	Triticum monococcum	L1	
Isc-Nin-DC-SINE	.	Ixodes scapularis	SINE2/tRNA	
ISEAG1	AB097148	Anopheles gambiae	L1	
IVK_DM	.	Drosophila melanogaster	Jockey	6,2,9,10,11,1,4,15,16,17,3,7
JAM1	.	Aedes aegypti	RTE	1,26,3,11,4,2
JAM1B_AAe	.	Aedes aegypti	RTE	1,26,3,11,4,2
JAM1C_AAe	.	Aedes aegypti	RTE	1,26,3,11,4,2
JN_As	.	Adineta	Penelope	1,3
JN_Av	.	Adineta vaga	Penelope	7,1,17,129,3,130
Jockey	M22874	Drosophila melanogaster	Jockey	16,1,87,88,3,24,4,11,2
Jockey_Ele10	.	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey_Ele2	.	Aedes aegypti	Jockey	16,1,8,14,3,4,11,12,2,15
Jockey_Ele3	.	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey_Ele5	.	Aedes aegypti	Jockey	
Jockey_Ele7	.	Aedes aegypti	Jockey	16,1,3,4,11,2,15
Jockey_Ele8	.	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey-1_BTa	.	Bemisia tabaci	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_CoFl	.	Copidosoma floridanum	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_CQ	.	Culex quinquefasciatus	Jockey	16,1,14,3,11,4,2,15
Jockey-1_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DBi	.	Drosophila biarmipes	Jockey	87,14,88
Jockey-1_DBp	.	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-1_DEl	.	Drosophila elegans	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DEr	.	Drosophila erecta	Jockey	87,2,11,12,1,14,4,15,16,3,88
Jockey-1_DEu	.	Drosophila eugracilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DF	.	Drosophila ficusphila	Jockey	16,1,87,88,3,11,4,2,15
Jockey-1_DGri	.	Drosophila grimshawi	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DK	.	Drosophila kikkawai	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DPer	.	Drosophila persimilis	Jockey	87,2,11,1,4,15,16,3,88,174
Jockey-1_DRh	.	Drosophila rhopaloa	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_Dro	.	Drosophila	Jockey	
Jockey-1_DSe	.	Drosophila sechellia	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DSer	.	Drosophila serrata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DSi	.	Drosophila simulans	Jockey	16,1,3,11,4,2,15
Jockey-1_DT	.	Drosophila takahashii	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DVi	.	Drosophila virilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DWi	.	Drosophila willistoni	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_DYa	.	Drosophila yakuba	Jockey	16,1,87,3,88,11,4,2,15
Jockey-1_Hmel	.	Heliconius melpomene	Jockey	1,3,4,2
Jockey-1_LMi	.	Locusta migratoria	Jockey	1,3,11,4,2
Jockey-1_LVa	.	Litopenaeus vannamei	Jockey	16,1,14,3,11,4,2,15
Jockey-1_PH	.	Parhyale hawaiensis	Jockey	16,1,8,3,11,4,12,2,15
Jockey-1_PPo	.	Papilio polytes	Jockey	1,3,4,2
Jockey-1_PXu	.	Papilio xuthus	Jockey	
Jockey-1_RPr	.	Rhodnius prolixus	Jockey	1,3,4,2

Jockey-1_TCa	Tribolium castaneum	Jockey	16,1,3,11,4,2,15
Jockey-10_BTa	Bemisia tabaci	Jockey	16,1,3,11,4,2,15
Jockey-10_DAn	Drosophila ananassae	Jockey	16,1,87,14,3,88,4,11,2,15
Jockey-10_DBp	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-10_DEu	Drosophila eugracilis	Jockey	
Jockey-10_DF	Drosophila ficusphila	Jockey	
Jockey-10_DRh	Drosophila rhopaloa	Jockey	16,1,87,3,88,11,4,2,15
Jockey-10_DY	Drosophila yakuba	Jockey	87,88
Jockey-11_AAe	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey-11_BTa	Bemisia tabaci	Jockey	16,1,3,11,4,2,15
Jockey-11_DAn	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-11_DBp	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-11_DF	Drosophila ficusphila	Jockey	
Jockey-11B_DBp	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-12_AAe	Aedes aegypti	Jockey	16,1,8,14,3,11,4,12,2,15
Jockey-12_BTa	Bemisia tabaci	Jockey	16,1,3,11,4,2,15
Jockey-12_DAn	Drosophila ananassae	Jockey	16,1,3,11,4,2,15
Jockey-12_DBp	Drosophila bipectinata	Jockey	
Jockey-12_DF	Drosophila ficusphila	Jockey	
Jockey-13_AAe	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey-13_BTa	Bemisia tabaci	Jockey	16,1,3,4,11,2,15
Jockey-13_DAn	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-13_DBp	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-14_AAe	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey-14_DAn	Drosophila ananassae	Jockey	1,3,4,2
Jockey-14_DBp	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-15_AAe	Aedes aegypti	Jockey	16,1,3,11,4,2,15
Jockey-15_DBp	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-16_AAe	Aedes aegypti	Jockey	16,1,14,3,11,4,2,15
Jockey-17_AAe	Aedes aegypti	Jockey	16,1,3,4,11,2,15
Jockey-1B_Hmel	Heliconius melpomene melpomene	Jockey	16,1,3,11,4,2,15
Jockey-1B_LMi	Locusta migratoria	Jockey	
Jockey-2_BTa	Bemisia tabaci	Jockey	16,1,131,3,4,11,132,2,15
Jockey-2_CQ	Culex quinquefasciatus	Jockey	16,1,8,14,3,11,4,12,2,15
Jockey-2_DAn	Drosophila ananassae	Jockey	16,1,3,11,4,2,15
Jockey-2_DBi	Drosophila biarmipes	Jockey	1,3,11,4,2
Jockey-2_DBp	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-2_DEl	Drosophila elegans	Jockey	16,1,87,3,88,11,4,2,15
Jockey-2_DEu	Drosophila eugracilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-2_DF	Drosophila ficusphila	Jockey	16,1,87,88,3,4,11,2,15
Jockey-2_DGri	Drosophila grimshawi	Jockey	16,1,3,24,11,4,2
Jockey-2_DK	Drosophila kikkawai	Jockey	16,1,87,88,3,11,4,2,15
Jockey-2_DPer	Drosophila persimilis	Jockey	87,2,11,1,4,15,16,3,88
Jockey-2_DRh	Drosophila rhopaloa	Jockey	16,1,87,3,88,11,4,2,15
Jockey-2_DSer	Drosophila serrata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-2_DSim	Drosophila simulans	Jockey	16,11,15
Jockey-2_DT	Drosophila takahashii	Jockey	93,16,1,87,3,88,11,4,2,15
Jockey-2_DVi	Drosophila virilis	Jockey	87,2,11,1,4,15,16,3,88
Jockey-2_DYa	Drosophila yakuba	Jockey	16,1,3,4,11,2,15
Jockey-2_Hmel	Heliconius melpomene	Jockey	16,1,87,3,88,11,4,2,15
Jockey-2_HMM	Heliconius melpomene melpomene	Jockey	92,16,1,87,3,88,11,4,2,15

Jockey-2_LMi	.	Locusta migratoria	Jockey	1,3,4,2
Jockey-2_LVa	.	Litopenaeus vannamei	Jockey	16,1,8,3,46,11,4,12,2,15
Jockey-2_PPo	.	Papilio polytes	Jockey	
Jockey-2_RPr	.	Rhodnius prolixus	Jockey	1,87,3,88,4,2
Jockey-2B_LMi	.	Locusta migratoria	Jockey	
Jockey-3_BTa	.	Bemisia tabaci	Jockey	16,1,3,11,4,2,15
Jockey-3_CQ	.	Culex quinquefasciatus	Jockey	16,1,3,11,4,2,15
Jockey-3_DAn	.	Drosophila ananassae	Jockey	16,1,3,11,4,2,15
Jockey-3_DBp	.	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-3_DEl	.	Drosophila elegans	Jockey	16,1,87,3,88,24,11,4,2
Jockey-3_DEu	.	Drosophila eugracilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-3_DF	.	Drosophila ficusphila	Jockey	16,1,87,3,88,11,4,2,15
Jockey-3_DGri	.	Drosophila grimshawi	Jockey	87,88
Jockey-3_DK	.	Drosophila kikkawai	Jockey	16,1,87,3,88,11,4,2,15
Jockey-3_DPer	.	Drosophila persimilis	Jockey	87,2,11,1,4,15,16,3,88,174
Jockey-3_DRh	.	Drosophila rhopaloa	Jockey	16,1,87,3,88,24,11,4,2
Jockey-3_DSim	.	Drosophila simulans	Jockey	16,1,87,3,88,11,4,2,15
Jockey-3_DT	.	Drosophila takahashii	Jockey	1,87,88,3,11,4,2
Jockey-3_DVi	.	Drosophila virilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-3_DYa	.	Drosophila yakuba	Jockey	16,1,3,11,4,2,15
Jockey-3_Hmel	.	Heliconius melpomene	Jockey	16,1,3,11,4,2,15
Jockey-3_HMM	.	Heliconius melpomene melpomene	Jockey	16,1,3,11,4,2,15
Jockey-3_LMi	.	Locusta migratoria	Jockey	1,3,4,2
Jockey-3_RPr	.	Rhodnius prolixus	Jockey	16,1,87,3,88,11,4,2,15
Jockey-3B_LMi	.	Locusta migratoria	Jockey	
Jockey-4_BTa	.	Bemisia tabaci	Jockey	16,1,3,11,4,2,15
Jockey-4_CQ	.	Culex quinquefasciatus	Jockey	16,1,3,11,4,2,15
Jockey-4_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,24,11,4,2
Jockey-4_DBp	.	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DEl	.	Drosophila elegans	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DEu	.	Drosophila eugracilis	Jockey	16,1,87,14,3,88,11,4,2,15
Jockey-4_DF	.	Drosophila ficusphila	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DK	.	Drosophila kikkawai	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DPer	.	Drosophila persimilis	Jockey	16,1,87,88,3,11,4,2,15
Jockey-4_DRh	.	Drosophila rhopaloa	Jockey	16,1,3,11,4,2,15
Jockey-4_DSim	.	Drosophila simulans	Jockey	1,3,4,2
Jockey-4_DT	.	Drosophila takahashii	Jockey	
Jockey-4_DVi	.	Drosophila virilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_DYa	.	Drosophila yakuba	Jockey	1,3,4,2
Jockey-4_Hmel	.	Heliconius melpomene	Jockey	16,1,3,11,4,2,15
Jockey-4_HMM	.	Heliconius melpomene melpomene	Jockey	16,1,87,3,88,11,4,2,15
Jockey-4_LMi	.	Locusta migratoria	Jockey	1,3,4,2
Jockey-4_RPr	.	Rhodnius prolixus	Jockey	16,1,3,11,4,2,15
Jockey-4B_DVi	.	Drosophila virilis	Jockey	1,87,3,88,11,4,2
Jockey-5_BTa	.	Bemisia tabaci	Jockey	16,131,1,3,11,4,132,2,15
Jockey-5_CQ	.	Culex quinquefasciatus	Jockey	16,1,14,3,11,4,2,15
Jockey-5_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-5_DBp	.	Drosophila bipectinata	Jockey	16,1,87,3,88,11,4,2,15
Jockey-5_DEl	.	Drosophila elegans	Jockey	16,1,87,3,88,11,4,2,15
Jockey-5_DEu	.	Drosophila eugracilis	Jockey	1,3,4,2
Jockey-5_DF	.	Drosophila ficusphila	Jockey	16,1,3,11,4,2,15

Jockey-5_DK	.	Drosophila kikkawai	Jockey	1,3,4,2
Jockey-5_DPer	.	Drosophila persimilis	Jockey	16,1,87,88,3,11,4,2,15
Jockey-5_DTa	.	Drosophila takahashii	Jockey	16,1,87,3,88,11,4,2,15
Jockey-5_DVi	.	Drosophila virilis	Jockey	87,2,119,11,1,4,15,16,3,88
Jockey-5_DYa	.	Drosophila yakuba	Jockey	16,1,3,4,11,2,15
Jockey-5_Hmel	.	Heliconius melpomene	Jockey	1,3,4,2
Jockey-5_IMi	.	Locusta migratoria	Jockey	1,3,4,2
Jockey-5_RPr	.	Rhodnius prolixus	Jockey	1,3,4,2
Jockey-5B_HMM	.	Heliconius melpomene melpomene	Jockey	1,3,4,2
Jockey-6_BTa	.	Bemisia tabaci	Jockey	16,131,1,3,4,11,132,2,15
Jockey-6_CQ	.	Culex quinquefasciatus	Jockey	16,1,3,11,4,2,15
Jockey-6_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-6_DBp	.	Drosophila bipectinata	Jockey	16,1,3,11,4,2,15
Jockey-6_DEl	.	Drosophila elegans	Jockey	16,1,87,3,88,11,4,2,15
Jockey-6_DEu	.	Drosophila eugracilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-6_DF	.	Drosophila ficusphila	Jockey	16,1,87,3,88,11,4,2,15
Jockey-6_DK	.	Drosophila kikkawai	Jockey	16,1,3,11,4,2,15
Jockey-6_DPer	.	Drosophila persimilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-6_DRh	.	Drosophila rhopaloa	Jockey	16,1,3,11,4,2,15
Jockey-6_DTa	.	Drosophila takahashii	Jockey	16,1,87,3,88,11,4,2,15
Jockey-6_DVi	.	Drosophila virilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-6_DYa	.	Drosophila yakuba	Jockey	16,1,3,4,11,2,15
Jockey-6_Hmel	.	Heliconius melpomene	Jockey	1,3,4,2
Jockey-6_IMi	.	Locusta migratoria	Jockey	
Jockey-6_RPr	.	Rhodnius prolixus	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_BTa	.	Bemisia tabaci	Jockey	16,1,3,82,11,4,2,15
Jockey-7_CQ	.	Culex quinquefasciatus	Jockey	16,1,3,11,4,2,15
Jockey-7_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DEl	.	Drosophila elegans	Jockey	16,1,3,4,11,2,15
Jockey-7_DEu	.	Drosophila eugracilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DF	.	Drosophila ficusphila	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DK	.	Drosophila kikkawai	Jockey	1,3,4,2
Jockey-7_DPer	.	Drosophila persimilis	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DRh	.	Drosophila rhopaloa	Jockey	16,1,87,3,88,11,4,2,15
Jockey-7_DVi	.	Drosophila virilis	Jockey	1,3,4,2
Jockey-7_DYa	.	Drosophila yakuba	Jockey	16,1,3,11,4,2,15
Jockey-7_Hmel	.	Heliconius melpomene	Jockey	16,1,3,11,4,2,15
Jockey-7_HMM	.	Heliconius melpomene melpomene	Jockey	16,1,3,11,4,2,15
Jockey-7_IMi	.	Locusta migratoria	Jockey	1,3,11,4,2
Jockey-7_RPr	.	Rhodnius prolixus	Jockey	16,87,88,11,15
Jockey-8_BTa	.	Bemisia tabaci	Jockey	16,1,3,11,4,2,15
Jockey-8_CQ	.	Culex quinquefasciatus	Jockey	16,1,3,11,4,2,15
Jockey-8_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15
Jockey-8_DEu	.	Drosophila eugracilis	Jockey	
Jockey-8_DF	.	Drosophila ficusphila	Jockey	16,1,87,3,88,11,4,2,15
Jockey-8_DPer	.	Drosophila persimilis	Jockey	16,1,3,11,4,2,15
Jockey-8_DRh	.	Drosophila rhopaloa	Jockey	16,1,87,3,88,11,4,2,15
Jockey-8_DVi	.	Drosophila virilis	Jockey	87,91,2,11,93,1,4,15,16,3,88
Jockey-8_DYa	.	Drosophila yakuba	Jockey	16,87,88,11,15
Jockey-8_Hmel	.	Heliconius melpomene	Jockey	1,3,4,2
Jockey-8_HMM	.	Heliconius melpomene melpomene	Jockey	16,1,3,11,4,2,15

Jockey-8_LMi	.	Locusta migratoria	Jockey	16,1,3,11,4,2,15	
Jockey-8_RPr	.	Rhodnius prolixus	Jockey	1,3,4,2	
Jockey-9_BTa	.	Bemisia tabaci	Jockey	16,1,3,11,4,2,15	
Jockey-9_DAn	.	Drosophila ananassae	Jockey	16,1,87,3,88,11,4,2,15	
Jockey-9_DBp	.	Drosophila bipectinata	Jockey	1,3,2	
Jockey-9_DEu	.	Drosophila eugracilis	Jockey		
Jockey-9_DF	.	Drosophila ficusphila	Jockey		
Jockey-9_DPer	.	Drosophila persimilis	Jockey	16,1,87,3,88,4,11,2,15	
Jockey-9_DRh	.	Drosophila rhopaloa	Jockey	16,1,3,11,4,2,15	
Jockey-9_DY	.	Drosophila yakuba	Jockey	16,1,3,11,4,2,15	
Jockey-9_LMi	.	Locusta migratoria	Jockey	16,1,3,24,11,4,2	
Jockey-N1_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N1_DVi	.	Drosophila virilis	Jockey	87,88	
Jockey-N2_CQ	.	Culex quinquefasciatus	Jockey		
14					
Jockey-N3_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N4_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N5_CQ	.	Culex quinquefasciatus	Jockey		
14					
Jockey-N6_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N6B_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N6C_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N7_CQ	.	Culex quinquefasciatus	Jockey		
Jockey-N8_CQ	.	Culex quinquefasciatus	Jockey		
14					
JOCKEY2	.	Drosophila melanogaster	Jockey	1,3,4,2	
JVY_As	.	Adineta	Penelope		
82					
JY_As	.	Adineta	Penelope		
K_As	.	Adineta	Penelope	1,3	
K_Av	.	Adineta vaga	Penelope	1,3	
KARIN_TA	AF325197	Triticum aestivum	L1		
KARIN_TM	AY485644	Triticum monococcum	Non-LTRRetrotransposo	1,43,3,44,11,4,2	
Keno-1_GA	.	Gasterosteus aculeatus	Tx1	43,2,8,44,11,12,1,26,14,4,3	
Keno-1_SSa	.	Salmo salar	Tx1	1,43,3,44,11,4,2	
Keno-5_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,11,12,42,1,26,14,4,3	
Keno-6_XT	.	Xenopus tropicalis	Tx1	1,43,3,44,4,2	
Keno-7_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,11,12,1,26,14,4,15,16,3	
Keno-8_XT	.	Xenopus tropicalis	Tx1		
KenoDr1	AB097144	Danio rerio	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3,46	
KenoDr2	.	Danio rerio	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3,46	
KenoDr3	.	Danio rerio	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3,46	
KenoFr1	AB111947	Takifugu rubripes	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46	
Kibi-1_GA	.	Gasterosteus aculeatus	Tx1	1,26,3,11,4,2	
KibiDr1	.	Danio rerio	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3	
KibiDr2	.	Danio rerio	Tx1	1,26,3,11,4,2	
KibiFr1	AB097136	Takifugu rubripes	Non-LTRRetrotransposo	1,26,3,11,4,2	
KibiTn1	AB097136	Tetraodon nigroviridis	L1	1,26,43,3,44,11,4,2	
Kiri-1_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15	
Kiri-1_CQ	.	Culex quinquefasciatus	Kiri	1,3,11,4,2	
Kiri-1_DBi	.	Drosophila biarmipes	Kiri	1,3,4,2	

Kiri-1_DEl	.	.	Drosophila elegans	Kiri	1,3,11,4,2
Kiri-1_DF	.	.	Drosophila ficusphila	Kiri	1,3,4,2
Kiri-1_DK	.	.	Drosophila kikkawai	Kiri	16,1,3,24,11,4,2
Kiri-1_DPer	.	.	Drosophila persimilis	Kiri	1,3,11,4,2
Kiri-1_DRh	.	.	Drosophila rhopaloa	Kiri	1,3,11,4,2
Kiri-1_DTa	.	.	Drosophila takahashii	Kiri	1,3,4,2
Kiri-1_HMM	.	.	Heliconius melpomene melpomene	Kiri	21,1,3,4,2
Kiri-1_LMi	.	.	Locusta migratoria	Kiri	1,3,11,4,2
Kiri-1_PPo	.	.	Papilio polytes	Kiri	
Kiri-1_PXu	.	.	Papilio xuthus	Kiri	
Kiri-10_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-10_CQ	.	.	Culex quinquefasciatus	Kiri	1,3,11,4,2
Kiri-10_HMM	.	.	Heliconius melpomene melpomene	Kiri	1,3,11,4,2
Kiri-11_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-11_CQ	.	AAWU01038353	Culex quinquefasciatus	Kiri	53,2,54,11,56,57,68,1,58,4,127,3,59
Kiri-12_AAe	.	.	Aedes aegypti	Kiri	1,3,4,11,2
Kiri-12_HMM	.	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-13_AAe	.	.	Aedes aegypti	Kiri	16,1,3,24,4,11,2
Kiri-13_HMM	.	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-14_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-15_AAe	.	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-16_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-17_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-17_HMM	.	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-18_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-19_AAe	.	.	Aedes aegypti	Kiri	16,1,3,4,11,2,15
Kiri-2_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-2_CQ	.	.	Culex quinquefasciatus	Kiri	1,3,11,4,2
Kiri-2_DEl	.	.	Drosophila elegans	Kiri	16,1,3,24,11,4,2
Kiri-2_DK	.	.	Drosophila kikkawai	Kiri	16,1,3,24,11,4,2
Kiri-2_DPer	.	.	Drosophila persimilis	Kiri	1,3,11,4,2
Kiri-2_DTa	.	.	Drosophila takahashii	Kiri	16,1,3,24,11,4,2
Kiri-2_HMM	.	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-2_LMi	.	.	Locusta migratoria	Kiri	1,3,4,2
Kiri-2_PPo	.	.	Papilio polytes	Kiri	
Kiri-20_AAe	.	.	Aedes aegypti	Kiri	1,3,4,11,2
Kiri-20_HMM	.	.	Heliconius melpomene melpomene	Kiri	1,3,11,4,2
Kiri-21_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-22_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-23_AAe	.	.	Aedes aegypti	Kiri	16,1,3,24,4,11,2
Kiri-24_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-25_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-26_AAe	.	.	Aedes aegypti	Kiri	16,1,3,4,11,2,15
Kiri-27_AAe	.	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-28_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-29_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-3_AAe	.	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-3_CQ	.	.	Culex quinquefasciatus	Kiri	16,1,3,24,11,4,2
Kiri-3_DK	.	.	Drosophila kikkawai	Kiri	1,3,11,4,2
Kiri-3_DTa	.	.	Drosophila takahashii	Kiri	16,1,3,24,11,4,2
Kiri-3_HMM	.	.	Heliconius melpomene melpomene	Kiri	21,1,3,11,4,2

Kiri-3_LMi	.	Locusta migratoria	Kiri	1,3,4,2
Kiri-3_PPo	.	Papilio polytes	Kiri	
Kiri-30_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-31_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-32_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-33_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-34_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-35_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-36_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-37_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-38_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-4_AAe	.	Aedes aegypti	Kiri	1,3,4,11,2
Kiri-4_CQ	.	Culex quinquefasciatus	Kiri	53,2,54,11,56,57,68,1,58,4,127,69,3
Kiri-4_DK	.	Drosophila kikkawai	Kiri	1,3,11,4,2
Kiri-4_DTa	.	Drosophila takahashii	Kiri	1,3,11,4,2
Kiri-4_LMi	.	Locusta migratoria	Kiri	
Kiri-5_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-5_CQ	.	Culex quinquefasciatus	Kiri	1,3,11,4,2
Kiri-5_DK	.	Drosophila kikkawai	Kiri	1,3,4,2
Kiri-5_DTa	.	Drosophila takahashii	Kiri	1,3,11,4,2
Kiri-5_HMM	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-5_LMi	.	Locusta migratoria	Kiri	1,3,4,2
Kiri-6_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-6_CQ	.	Culex quinquefasciatus	Kiri	1,3,4,2
Kiri-6_DK	.	Drosophila kikkawai	Kiri	16,1,3,4,11,2,15
Kiri-6_DTa	.	Drosophila takahashii	Kiri	1,3,4,11,2
Kiri-7_AAe	.	Aedes aegypti	Kiri	1,3,11,4,2
Kiri-7_CQ	.	Culex quinquefasciatus	Kiri	1,3,4,2
Kiri-7_DTa	.	Drosophila takahashii	Kiri	1,3,11,4,2
Kiri-7_HMM	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-8_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-8_CQ	.	Culex quinquefasciatus	Kiri	1,3,4,11,2
Kiri-8B_HMM	.	Heliconius melpomene melpomene	Kiri	1,3,4,2
Kiri-9_AAe	.	Aedes aegypti	Kiri	16,1,3,11,4,2,15
Kiri-9_CQ	.	Culex quinquefasciatus	Kiri	16,1,3,11,4,2,15
KoshiTn1	AB097135	Tetraodon nigroviridis	Tx1	2,8,11,12,42,1,26,14,4,3
L1	.	Homo sapiens	L1	
L1_AC_1	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1_AC_10	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1_AC_11	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1_AC_12	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1_AC_13	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1_AC_14	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1_AC_15	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1_AC_16	.	Anolis carolinensis	L1	21,1,26,3,28,11,4,2
L1_AC_17	.	Anolis carolinensis	L1	21,16,1,26,3,24,11,4,2
L1_AC_18	.	Anolis carolinensis	L1	2,21,11,1,26,28,4,3
L1_AC_19	.	Anolis carolinensis	L1	21,1,26,3,28,11,4,2
L1_AC_2	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1_AC_20	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1_AC_3	.	Anolis carolinensis	L1	1,26,3,4,11,2

L1_AC_3B	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1_AC_4	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1_AC_5	.	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1_AC_6	.	Anolis carolinensis	L1	16,1,26,23,3,4,11,2,15
L1_AC_7	.	Anolis carolinensis	L1	1,3,11,4,2
L1_AC_8	.	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1_AC_9	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1_As	.	Adineta	Penelope	1,3,130
L1_Av	.	Adineta vaga	Penelope	1,3
L1_BT	.	Bovidae	L1	
L1_CA	.	Cicer arietinum	L1	
L1_Canid_5end	.	Canidae	L1	
L1_Canid2_5end	.	Canidae	L1	
L1_Carn1_5end	.	Carnivora	L1	
L1_Carn2_5end	.	Carnivora	L1	
L1_Carn2_orf2	.	Carnivora	L1	
L1_Carn3_5end	.	Carnivora	L1	
L1_Carn5_3end	.	Carnivora	L1	
L1_Carn5_5end	.	Carnivora	L1	
L1_Carni_orf2	.	Carnivora	L1	
L1_CC	Y00726	Cavia cutleri	L1	
L1_CF	AB012223	Canis lupus familiaris	L1	
L1_CP	M60812	Conilurus penicillatus	L1	
L1_Ele12	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,2
L1_Ele13	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1_Ele16	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1_Ele17	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,2
L1_Ele19	.	Aedes aegypti	L1	93,1,26,3,11,4,2
L1_Ele21	.	Aedes aegypti	L1	16,1,26,43,3,44,24,11,4,2
L1_Ele26	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1_Ele28	.	Aedes aegypti	L1	16,1,26,3,24,4,11,2
L1_Ele3	.	Aedes aegypti	L1	16,1,26,3,24,11,4,92,2
L1_Ele3B_AAe	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,92,2
L1_Ele3C_AAe	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1_Ele4	.	Aedes aegypti	L1	1,26,3,11,4,2
L1_Ele5	.	Aedes aegypti	L1	43,2,44,24,11,93,1,26,4,16,3
L1_Ele9	.	Aedes aegypti	L1	16,1,26,3,24,4,11,2
L1_Fc	.	Felis	L1	
L1_Fc_5end	.	Felis	L1	
L1_Fc_orf2	.	Felis	L1	
L1_Felid	.	Felidae	L1	
L1_Lme	.	Latimeria menadoensis	L1	1,3,4,2
L1_Mars1_orf2	.	Metatheria	L1	
L1_Mars1a_3end	.	Metatheria	L1	
L1_Mars1b_3end	.	Metatheria	L1	
L1_Mdo3b_3end	.	Monodelphis	L1	
L1_MM	M29324	Mus musculus	L1	
L1_Mur1_5end	.	Muridae	L1	
L1_Mur1_orf2	.	Muridae	L1	
L1_Mur2_3end	.	Muridae	L1	
L1_Mur2_5end	.	Muridae	L1	

L1_Mur2_orf2	.	Muridae	L1	
L1_Mur3_5end	.	Muridae	L1	
L1_Mur3_orf2	.	Muridae	L1	
L1_Mus1_3end	.	Mus	L1	
L1_Mus2_3end	.	Mus	L1	
L1_Opos_5end	.	Didelphidae	L1	
L1_Opos3_orf2	.	Didelphidae	L1	
L1_Rat3_3end	.	Rattus	L1	
L1_Rat4_3end	.	Rattus	L1	
L1_RN	.	Rattus norvegicus	L1	
L1_Rn2_3end	.	Rattus	L1	
L1_Rod_5end	.	Muridae	L1	
L1_Rod_orf2	.	Muridae	L1	
L1_RS1_5end	.	Cercopithecinae	L1	
L1_SS	.	Suina	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1_TC	.	Trypanosoma cruzi	Ingi	6,2,18,10,11,1,4,15,5,16,17,3,7
L1_TD	.	Triticum turgidum subsp. durum	L1	
L1-1_AB	.	Agaricus bisporus	L1	2,10,11,13,1,26,4,17,3,7
L1-1_Acar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-1_AFC	.	Cichlidae	L1	16,1,26,24,11,2
L1-1_AIp	.	Arachis ipaensis	L1	43,79,2,44,10,11,80,1,77,4,75,17,3,76,7
L1-1_ALai	.	Albugo laibachii	L1	1,3,11,4,2
L1-1_ALy	.	Arabidopsis lyrata	L1	
43,78,79,2,44,10,24,11,80,1,81,77,4,16,75,3,76				
L1-1_AMe	.	Ailuropoda melanoleuca	L1	25,21,20,23,28,30
L1-1_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-1_BDi	.	Brachypodium distachyon	L1	43,2,44,11,12,1,14,4,3
L1-1_BRa	.	Brassica rapa	L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-1_BV	.	Beta vulgaris subsp. vulgaris	L1	
43,78,79,2,8,10,44,11,12,80,1,81,77,4,17,75,3,76,7				
L1-1_CCr	.	Condylura cristata	L1	31,25,21,20,23,27,28,30
L1-1_CGr	.	Cricetulus griseus	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1-1_Cho	.	Choloepus hoffmanni	L1	16,1,26,3,24,4,11,2
L1-1_Cja	.	Callithrix jacchus	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1-1_CPa	.	Carica papaya	L1	
43,78,79,6,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76,7				
L1-1_CPB	.	Chrysemys picta bellii	L1	
L1-1_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_CQ	.	Culex quinquefasciatus	L1	93,16,1,26,3,24,11,4,2
L1-1_CR	.	Chlamydomonas reinhardtii	L1	16,1,3,24,11,4,2
L1-1_Crp	.	Crocodylus porosus	L1	
L1-1_DN	.	Dasypus novemcinctus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_Dor	.	Dipodomys ordii	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1-1_DR	AL645691	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-1_DV	.	Didelphis virginiana	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1_ECa	.	Eucalyptus camaldulensis	L1	43,2,44,11,1,77,4,75,3,76

L1-1_EE	.	Erinaceus europaeus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-1_EL	.	Esox lucius	L1	
L1-1_EPa	.	Schrenkiella parvula	L1	
L1-1_ET	.	Echinops telfairi	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1_FV	.	Fragaria vesca	L1	2,43,80,17,1,3,44,10,79,4,7
L1-1_GAr	.	Gossypium arboreum	L1	43,79,2,44,10,24,11,80,1,4,16,17,3,7
L1-1_GM	.	Glycine max	L1	43,78,2,44,11,1,81,77,4,75,3,76
L1-1_GR	.	Gossypium raimondii	L1	43,78,2,44,10,11,80,1,81,77,4,17,75,3,76
L1-1_HAE	.	Hyaloperonospora arabidopsidis	L1	6,2,10,11,80,1,26,230,4,17,3,7
L1-1_HArab	.	Hyaloperonospora arabidopsidis	L1	1,3,4,2
L1-1_HM	.	Hydra vulgaris	L1	16,1,26,3,24,11,4,2
L1-1_JC	.	Jatropha curcas	L1	1,3,11,4,2
L1-1_LA	.	Loxodonta africana	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_LA-5end	.	Loxodonta africana	L1	
L1-1_LCh	.	Latimeria chalumnae	L1	
L1-1_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,11,4,2,15
L1-1_LU	.	Linum usitatissimum	L1	1,43,3,44,4,11,2
L1-1_MAc	.	Musa acuminata	L1	1,43,75,3,76,44,77,11,4,2
L1-1_Mad	.	Malus domestica	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-1_MD	.	Monodelphis domestica	L1	20,2,21,23,11,25,1,26,28,4,30,15,16,3
L1-1_ME	.	Notamacropus eugenii	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_MN	.	Morus notabilis	L1	7,1,17,43,3,44,10,79,11,4,2,80
L1-1_Mpr	.	Mammuthus primigenius	L1	
L1-1_MPu	.	Mustela putorius furo	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-1_NS	.	Nicotiana sylvestris	L1	1,43,3,44,11,4,2
L1-1_NV	.	Nematostella vectensis	L1	
L1-1_OAf	.	Orycteropus afer	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1_OES	.	Olea europaea subsp. europaea	L1	43,79,6,2,7,10,44,24,11,80,1,77,4,16,75,17,3,76
L1-1_OL	.	Oryzias latipes	L1	2,21,23,24,11,1,26,4,16,3
L1-1_OP	.	Ochotona princeps	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_OSp	.	Pecoramyces ruminatium	L1	1,3,4,11,2
L1-1_PAb	.	Picea abies	L1	43,6,2,7,8,10,44,11,12,80,1,14,4,75,17,3,76
L1-1_PAlni	.	Phytophthora alni	L1	1,11,2
L1-1_PApha	.	Pythium aphanidermatum	L1	
L1-1_PArrh	.	Pythium arrhenomanes	L1	1,3,4,2
L1-1_Pca	.	Procavia capensis	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_PCaps	.	Phytophthora capsici	L1	6,2,10,11,80,1,230,4,17,3,7
L1-1_PCinn	.	Phytophthora cinnamomi	L1	7,1,17,3,10,11,4,6,2,80
L1-1_PD	.	Phoenix dactylifera	L1	
L1-1_PI	.	Phytophthora infestans	L1	1,8,3,11,4,12,2
L1-1_PInsi	.	Pythium insidiosum	L1	6,2,10,11,80,1,230,4,17,3,7
L1-1_PIway	.	Pythium iwayamai	L1	2,1
L1-1_PLate	.	Phytophthora lateralis	L1	
L1-1_PPini	.	Phytophthora pinifolia	L1	6,2,10,24,11,80,1,26,4,16,17,3,7
L1-1_PR	.	Phytophthora ramorum	L1	175,1,176,3,120,11,4,177,2
L1-1_PRamo	.	Phytophthora ramorum	L1	6,2,10,11,80,1,26,4,17,3,7
L1-1_PSoja	.	Phytophthora sojae	L1	
L1-1_PTr	.	Populus trichocarpa	L1	43,78,2,44,11,1,81,77,4,75,3,76

L1-1_PTr	.	Populus trichocarpa	L1	1,43,3,44,2
L1-1_PUlti	.	Pythium ultimum	L1	6,2,8,10,11,12,80,1,230,4,17,3,7
L1-1_PVa	.	Pteropus vampyrus	L1	20,2,21,23,25,1,27,28,4,30,31,3
L1-1_RC	.	Ricinus communis	L1	
L1-1_Sar	.	Sorex araneus	L1	85,20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3
L1-1_SH	.	Sarcophilus harrisii	L1	
L1-1_SMo	.	Selaginella moellendorffii	L1	16,1,43,3,44,24,11,4,2
L1-1_SP	.	Strongylocentrotus purpuratus	L1	1,26,43,44,11,4,2
L1-1_SSa	.	Salmo salar	L1	2,21,24,11,1,26,28,4,16,3
L1-1_SSc	.	Sus scrofa	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1-1_Str	.	Ictidomys tridecemlineatus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-1_STu	.	Solanum tuberosum	L1	
43,79,6,2,8,44,10,11,12,80,1,77,4,17,75,3,76,7				
L1-1_Tbel	.	Tupaia belangeri	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-1_TC	.	Theobroma cacao	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-1_TS	.	Carlito syrichta	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1-1_Ttr	.	Tursiops truncatus	L1	16,1,26,3,24,4,11,2
L1-1_VC	.	Volvox carteri	L1	16,1,26,3,24,11,4,2
L1-1_Vpa	.	Vicugna pacos	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1_VVi	.	Vitis vinifera	L1	16,1,43,3,44,24,11,4,2
L1-1_XL	.	Xenopus laevis	L1	2,21,24,11,1,26,28,4,16,3
L1-1_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,15,16,3
L1-1_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,1,14,4,16,3
L1-10_AIp	.	Arachis ipaensis	L1	43,78,2,44,11,1,81,77,4,75,3,76
L1-10_ALy	.	Arabidopsis lyrata	L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-10_ATr	.	Amborella trichopoda	L1	1,3,4,11,2
L1-10_BDi	.	Brachypodium distachyon	L1	16,1,43,3,44,24,11,4,2
L1-10_CPa	.	Carica papaya	L1	1,43,3,44,11,4,2
L1-10_CQ	.	Culex quinquefasciatus	L1	16,1,26,3,24,11,4,2
L1-10_DR	AL772154	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-10_ECa	.	Eucalyptus camaldulensis	L1	
43,78,79,2,8,44,10,11,12,80,1,81,77,4,75,17,3,76,7				
L1-10_FV	.	Fragaria vesca	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-10_GAr	.	Gossypium arboreum	L1	7,1,43,17,3,44,10,79,4,2,80
L1-10_GM	.	Glycine max	L1	16,1,43,75,3,76,44,24,11,4,2
L1-10_JC	.	Jatropha curcas	L1	1,43,17,44,10,79,4,2,6,7,80
L1-10_LCh	.	Latimeria chalumnae	L1	1,3,11,4,2
L1-10_LSal	.	Lepeophtheirus salmonis	L1	
L1-10_Mad	.	Malus domestica	L1	
L1-10_MD	.	Monodelphis domestica	L1	
L1-10_NN	.	Nelumbo nucifera	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-10_NS	.	Nicotiana sylvestris	L1	43,79,6,2,10,44,11,80,1,77,4,17,75,3,76,7
L1-10_PI	.	Phytophthora infestans	L1	
L1-10_SSa	.	Salmo salar	L1	2,21,23,11,1,26,4,3
L1-10_STu	.	Solanum tuberosum	L1	16,1,75,43,76,3,44,24,11,4,2
L1-10_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-100_CSav	.	Ciona savignyi	L1	1,14,3,11,4,12,2

L1-100_DR	.	Danio rerio	L1	1,2
L1-101_DR	.	Danio rerio	L1	
L1-102_DR	.	Danio rerio	L1	1,3,4,2
L1-103_DR	.	Danio rerio	L1	
L1-104_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-105_DR	.	Danio rerio	L1	
L1-106_DR	.	Danio rerio	L1	
L1-107_DR	.	Danio rerio	L1	20,2,21,23,11,25,1,28,4
L1-108_DR	.	Danio rerio	L1	16,21,1,26,3,24,11,4,2
L1-109_DR	.	Danio rerio	L1	21,1,26,23,3,11,4,2
L1-10a_MD	.	Monodelphis domestica	L1	
L1-11_AIp	.	Arachis ipaensis	L1	43,91,92,2,44,11,93,1,4,75,3,76
L1-11_ALy	.	Arabidopsis lyrata	L1	1,43,3,44,4,11,2
L1-11_ATr	.	Amborella trichopoda	L1	7,1,17,43,3,44,11,4,2
L1-11_BDi	.	Brachypodium distachyon	L1	14,12
L1-11_CPa	.	Carica papaya	L1	1,3,4,2
L1-11_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-11_ECa	.	Eucalyptus camaldulensis	L1	43,78,79,6,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76,7
L1-11_FV	.	Fragaria vesca	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-11_GM	.	Glycine max	L1	43,91,92,2,44,11,93,1,96,97,4,75,3,76,98
L1-11_LCh	.	Latimeria chalumnae	L1	16,1,26,3,24,11,4,2
L1-11_LSal	.	Lepeophtheirus salmonis	L1	
L1-11_Mad	.	Malus domestica	L1	43,78,79,6,2,8,10,44,11,12,80,1,81,77,4,17,75,3,76,7
L1-11_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-11_NS	.	Nicotiana sylvestris	L1	43,79,6,2,7,10,44,11,80,1,4,75,17,3,76
L1-11_PI	.	Phytophthora infestans	L1	7,1,17,3,10,4,6,2,80
L1-11_SSa	.	Salmo salar	L1	1,3,11,4,2
L1-11_STu	.	Solanum tuberosum	L1	
L1-11_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,15,16,3
L1-110_DR	.	Danio rerio	L1	
L1-111_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-112_DR	.	Danio rerio	L1	1,28,4,2
L1-113_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-114_DR	.	Danio rerio	L1	1,4,2
L1-115_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-116_DR	.	Danio rerio	L1	
L1-117_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-118_DR	.	Danio rerio	L1	1,3,11,4,2
L1-119_DR	.	Danio rerio	L1	21,1,23,11,4,2
L1-12_AIp	.	Arachis ipaensis	L1	1,43,3,44,11,4,2
L1-12_ALy	.	Arabidopsis lyrata	L1	1,3,11,4,2
L1-12_ATr	.	Amborella trichopoda	L1	1,3,11,4,2
L1-12_BDi	.	Brachypodium distachyon	L1	2,8,24,11,12,42,1,14,4,16,3
L1-12_CPa	.	Carica papaya	L1	44,43
L1-12_DR	.	Danio rerio	L1	21,1,26,23,3,11,4,2
L1-12_ECa	.	Eucalyptus camaldulensis	L1	43,78,79,6,2,7,8,10,44,24,11,12,80,1,14,81,77,4,16,75,17,3,76
L1-12_FV	.	Fragaria vesca	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-12_GM	.	Glycine max	L1	

L1-12_LCh	.	Latimeria chalumnae	L1	20,2,21,23,24,11,25,1,26,28,4,16,3
L1-12_LSal	.	Lepeophtheirus salmonis	L1	
L1-12_Mad	.	Malus domestica	L1	
L1-12_NN	.	Nelumbo nucifera	L1	43,79,6,2,44,10,11,80,1,4,100,17,3,7
L1-12_NS	.	Nicotiana sylvestris	L1	
L1-12_PI	.	Phytophthora infestans	L1	7,1,17,3,10,11,4,6,2,80
L1-12_SSa	.	Salmo salar	L1	2,21,23,11,1,26,28,30,3
L1-12_STu	.	Solanum tuberosum	L1	16,1,43,75,3,76,44,24,11,4,2
L1-12_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-120_DR	.	Danio rerio	L1	1,3,11,4,2
L1-121_DR	.	Danio rerio	L1	
L1-122_DR	.	Danio rerio	L1	1,26,3,11,4,2
L1-123_DR	.	Danio rerio	L1	44,43
L1-124_DR	.	Danio rerio	L1	1,4,2
L1-125_DR	.	Danio rerio	L1	1,3,4,2
L1-126_DR	.	Danio rerio	L1	16,1,26,3,24,4,11,2
L1-127_DR	.	Danio rerio	L1	
L1-128_DR	.	Danio rerio	L1	
L1-129_DR	.	Danio rerio	L1	
L1-12A_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-12B_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-13_AIp	.	Arachis ipaensis	L1	1,43,75,3,44,76,11,4,2
L1-13_ALy	.	Arabidopsis lyrata	L1	
L1-13_ATr	.	Amborella trichopoda	L1	1,3,4,2
L1-13_BDi	.	Brachypodium distachyon	L1	
L1-13_CPa	.	Carica papaya	L1	1,3,4,2
L1-13_DR	.	Danio rerio	L1	21,1,26,23,3,11,4,2
L1-13_FV	.	Fragaria vesca	L1	43,80,17,44,10,79,6,7
L1-13_GM	.	Glycine max	L1	
L1-13_LCh	.	Latimeria chalumnae	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-13_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,2
L1-13_Mad	.	Malus domestica	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-13_NN	.	Nelumbo nucifera	L1	43,79,6,2,44,10,24,11,80,1,4,100,16,17,3,7
L1-13_NS	.	Nicotiana sylvestris	L1	
L1-13_SSa	.	Salmo salar	L1	16,1,26,3,24,4,11,2
L1-13_STu	.	Solanum tuberosum	L1	16,1,43,75,76,44,24,11,4,2
L1-13_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,3
L1-130_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-131_DR	.	Danio rerio	L1	
L1-132_DR	.	Danio rerio	L1	1,3,11,4,2
L1-133_DR	.	Danio rerio	L1	
L1-134_DR	.	Danio rerio	L1	
L1-135_DR	.	Danio rerio	L1	
L1-136_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,3
L1-137_DR	.	Danio rerio	L1	21,16,1,26,23,24,11,2
L1-138_DR	.	Danio rerio	L1	
L1-139_DR	.	Danio rerio	L1	
L1-13B_SSa	.	Salmo salar	L1	
L1-14_AIp	.	Arachis ipaensis	L1	1,75,43,3,76,44,11,4,2
L1-14_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-14_ATr	.	Amborella trichopoda	L1	1,43,3,44,11,4,2

L1-14_BDi	.	Brachypodium distachyon	L1	1,3,11,4,2
L1-14_CPa	.	Carica papaya	L1	16,1,43,3,44,11,4,2,15
L1-14_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,15,16,3
L1-14_FV	.	Fragaria vesca	L1	43,80,17,10,44,79,6,7
L1-14_GM	.	Glycine max	L1	43,2,44,11,1,77,4,75,3,76
L1-14_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-14_Mad	.	Malus domestica	L1	
L1-14_NN	.	Nelumbo nucifera	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-14_NS	.	Nicotiana sylvestris	L1	
L1-14_SSa	.	Salmo salar	L1	21,1,3,28,11,4,2
L1-14_STu	.	Solanum tuberosum	L1	16,1,43,75,3,76,44,24,11,4,2
L1-14_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,15,16,3
L1-140_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-15_ALy	.	Arabidopsis lyrata	L1	
L1-15_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-15_BDi	.	Brachypodium distachyon	L1	1,43,3,44,11,4,2
L1-15_CPa	.	Carica papaya	L1	1,43,3,44,4,2
L1-15_DR	.	Danio rerio	L1	1,26,8,14,3,11,4,12,2
L1-15_FV	.	Fragaria vesca	L1	43,80,17,10,44,79,6,7
L1-15_GM	.	Glycine max	L1	43,2,44,11,1,77,4,75,3,76
L1-15_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-15_Mad	.	Malus domestica	L1	
L1-15_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,15,16,17,3,7
L1-15_NS	.	Nicotiana sylvestris	L1	
L1-15_SSa	.	Salmo salar	L1	1,26,3,11,4,2
L1-15_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,16,75,3,76
L1-15_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-15a_Lch	.	Latimeria chalumnae	L1	
L1-15B_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-15b_Lch	.	Latimeria chalumnae	L1	
L1-15c_Lch	.	Latimeria chalumnae	L1	
L1-16_ALy	.	Arabidopsis lyrata	L1	43,80,17,44,10,79,7
L1-16_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,7
L1-16_BDi	.	Brachypodium distachyon	L1	
L1-16_CPa	.	Carica papaya	L1	1,43,3,44,4,2
L1-16_DR	.	Danio rerio	L1	16,1,26,3,24,11,4,2
L1-16_FV	.	Fragaria vesca	L1	1,43,3,44,4,2
L1-16_GMa	.	Glycine max	L1	16,1,3,24,11,4,2
L1-16_Lch	.	Latimeria chalumnae	L1	
L1-16_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-16_Mad	.	Malus domestica	L1	1,43,3,44,10,11,4,2,80
L1-16_NN	.	Nelumbo nucifera	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-16_NS	.	Nicotiana sylvestris	L1	
L1-16_SSa	.	Salmo salar	L1	1,3,4,11,2
L1-16_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,16,75,3,76
L1-16_XT	.	Xenopus tropicalis	L1	1,3,11,4,2
L1-17_ALy	.	Arabidopsis lyrata	L1	
43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76				
L1-17_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-17_BDi	.	Brachypodium distachyon	L1	16,1,3,11,4,2,15

L1-17_CPa	.	Carica papaya	L1	44,43
L1-17_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-17_FV	.	Fragaria vesca	L1	17,10,79,6,7,80
L1-17_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,2
L1-17_Mad	.	Malus domestica	L1	7,6,1,80,17,43,3,44,10,79,4,2
L1-17_NN	.	Nelumbo nucifera	L1	1,43,3,44,11,4,2
L1-17_NS	.	Nicotiana sylvestris	L1	43,79,6,2,7,8,10,44,11,12,80,1,4,75,17,3,76
L1-17_SSa	.	Salmo salar	L1	16,26,24,11
L1-17_STu	.	Solanum tuberosum	L1	
L1-17_XT	.	Xenopus tropicalis	L1	16,1,26,3,24,4,11,2
L1-17a_Lch	.	Latimeria chalumnae	L1	
L1-17b_Lch	.	Latimeria chalumnae	L1	
L1-18_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,24,11,1,81,77,4,16,75,3,76
L1-18_ATr	.	Amborella trichopoda	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-18_BDi	.	Brachypodium distachyon	L1	
L1-18_CPa	.	Carica papaya	L1	16,1,3,24,11,4,2
L1-18_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-18_FV	.	Fragaria vesca	L1	17,10,79,6,7,80
L1-18_Lch	.	Latimeria chalumnae	L1	
L1-18_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,11,4,2,15
L1-18_Mad	.	Malus domestica	L1	
L1-18_NN	.	Nelumbo nucifera	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-18_NS	.	Nicotiana sylvestris	L1	
L1-18_SSa	.	Salmo salar	L1	2,21,24,11,1,26,28,4,30,16,3
L1-18_STu	.	Solanum tuberosum	L1	
43,78,79,6,2,7,10,44,11,80,1,81,77,4,75,17,3,76				
L1-18_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-19_ALy	.	Arabidopsis lyrata	L1	
43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7				
L1-19_ATr	.	Amborella trichopoda	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-19_CPa	.	Carica papaya	L1	1,43,3,44,4,2
L1-19_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-19_FV	.	Fragaria vesca	L1	43,80,17,44,10,79,6,7
L1-19_Lch	.	Latimeria chalumnae	L1	
L1-19_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-19_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-19_SSa	.	Salmo salar	L1	21,1,26,23,3,11,4,2
L1-19_STu	.	Solanum tuberosum	L1	16,1,75,43,3,76,44,24,4,11,2
L1-19_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-1a_Cis	.	Ciona savignyi	L1	1,14,3,11,4,12,2
L1-1A_Cpo	.	Cavia porcellus	L1	
L1-1a_EF	.	Eptesicus fuscus	L1	
L1-1a_MD	.	Monodelphis domestica	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1A_VC	.	Volvox carteri	L1	16,1,26,3,24,4,11,2
L1-1A2_Sar	.	Sorex araneus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1B_Cho	.	Choloepus hoffmanni	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-1b_Cis	.	Ciona savignyi	L1	
L1-1B_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1B_LA-5end	.	Loxodonta africana	L1	

L1-1B_MAc	.	Musa acuminata	L1	1,75,43,3,76,44,77,11,4,2
L1-1b_MD	.	Monodelphis domestica	L1	20,2,21,23,11,25,1,26,28,4,30,15,16,3
L1-1B_Opr	.	Ochotona princeps	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1b_TS	.	Carlito syrichta	L1	
L1-1C_Cho	.	Choloepus hoffmanni	L1	
L1-1C_Cpo	.	Cavia porcellus	L1	
L1-1D_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3,39
L1-1E_Cpo	.	Cavia porcellus	L1	
L1-1F_Cpo	.	Cavia porcellus	L1	
L1-1G_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-1H_Cpo	.	Cavia porcellus	L1	
L1-1N_Tbel	.	Tupaia belangeri	L1	25,21,20,23,28,30
L1-2_AIp	.	Arachis ipaensis	L1	43,79,2,7,10,44,11,80,1,77,4,75,17,3,76
L1-2_ALai	.	Albugo laibachii	L1	1,3,11,4,2
L1-2_ALy	.	Arabidopsis lyrata	L1	
L1-2_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-2_BDi	.	Brachypodium distachyon	L1	1,11,4,2
L1-2_BRa	.	Brassica rapa	L1	16,1,43,3,44,24,4,11,2
L1-2_BT	.	Bos taurus	L1	22,29
L1-2_BV	.	Beta vulgaris subsp. vulgaris	L1	7,1,17,43,3,44,10,79,11,4,2,80
L1-2_CGr	.	Cricetulus griseus	L1	2,22,24,11,1,26,4,29,16,3
L1-2_Cho	.	Choloepus hoffmanni	L1	25,21,20,23,28,30
L1-2_Cis	.	Ciona savignyi	L1	
L1-2_Cja	.	Callithrix jacchus	L1	
L1-2_CPa	.	Carica papaya	L1	16,1,43,3,44,24,11,4,2
L1-2_CPB	.	Chrysemys picta bellii	L1	
L1-2_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_CQ	.	Culex quinquefasciatus	L1	16,1,26,3,24,11,4,2
L1-2_CR	.	Chlamydomonas reinhardtii	L1	16,1,26,24,4,11,2
L1-2_Crp	.	Crocodylus porosus	L1	
L1-2_Dor	.	Dipodomys ordii	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_DR	.	Danio rerio	Tx1	2,8,24,11,12,42,1,26,14,4,16,3
L1-2_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-2_ECa	.	Eucalyptus camaldulensis	L1	1,75,43,3,76,44,77,11,4,2
L1-2_EE	.	Erinaceus europaeus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-2_EPa	AFAN01000035	Schrenkiella parvula	L1	1,43,78,3,44,81,77,11,4,2
L1-2_Fc	.	Felis catus	L1	25,21,20,23,28,30
L1-2_FR	.	Takifugu rubripes	L1	16,1,26,3,24,11,4,2
L1-2_FV	.	Fragaria vesca	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-2_GAr	.	Gossypium arboreum	L1	43,79,2,10,44,24,11,80,1,4,16,17,3
L1-2_GM	.	Glycine max	L1	16,1,43,75,76,44,24,11,2
L1-2_HAE	.	Hyaloperonospora arabidopsidis	L1	6,2,10,11,80,1,4,17,3,7
L1-2_HM	.	Hydra vulgaris	L1	1,26,3,11,4,2
L1-2_JC	.	Jatropha curcas	L1	7,1,17,43,3,44,10,79,11,4,2,80
L1-2_LA	.	Loxodonta africana	L1	20,2,21,23,11,25,1,26,28,4,30,15,16,3
L1-2_LCh	.	Latimeria chalumnae	L1	16,1,26,3,24,11,4,2
L1-2_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-2_MAc	.	Musa acuminata	L1	
L1-2_Mad	.	Malus domestica	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-2_MD	.	Monodelphis domestica	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_ME	.	Notamacropus eugenii	L1	2,21,23,24,11,1,26,4,30,16,3

L1-2_MPu	.	Mustela putorius furo	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-2_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-2_NS	.	Nicotiana sylvestris	L1	43,79,6,2,10,44,11,80,1,77,4,17,75,3,76,7
L1-2_NV	.	Nematostella vectensis	L1	
L1-2_Opr	.	Ochotona princeps	L1	84,20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_PAb	.	Picea abies	L1	16,1,3,24,4,11,2
L1-2_Pca	.	Procavia capensis	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_PD	.	Phoenix dactylifera	L1	7,16,1,17,43,3,44,24,11,4,2
L1-2_PI	.	Phytophthora infestans	L1	175,16,1,176,3,24,11,4,2
L1-2_PR	.	Phytophthora ramorum	L1	1,8,3,11,4,12,2
L1-2_PTr	.	Populus trichocarpa	L1	43,91,2,44,24,11,93,1,4,16,75,3,76
L1-2_PTr	.	Populus trichocarpa	L1	43,91,2,44,24,11,93,1,4,16,75,3,76
L1-2_PVa	.	Pteropus vampyrus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-2_Sar	.	Sorex araneus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2_SBi	.	Sorghum bicolor	L1	
L1-2_SMo	.	Selaginella moellendorffii	L1	16,1,43,3,44,24,11,4,2
L1-2_SP	.	Strongylocentrotus purpuratus	L1	1,43,3,44,4,2
L1-2_SSa	.	Salmo salar	L1	21,1,26,23,3,11,4,2
L1-2_SSc	.	Sus scrofa	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-2_STr	.	Ictidomys tridecemlineatus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-2_STu	.	Solanum tuberosum	L1	
L1-2_Tbel	.	Tupaia belangeri	L1	20,2,21,23,11,25,1,28,4,30,3
L1-2_TC	.	Theobroma cacao	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-2_TML	.	Trichechus manatus latirostris	L1	20,83,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-2_TS	.	Carlito syrichta	L1	
L1-2_Ttr	.	Tursiops truncatus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-2_VC	.	Volvox carteri	L1	16,1,26,3,24,4,11,2
L1-2_Vpa	.	Vicugna pacos	L1	20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3
L1-2_VVi	.	Vitis vinifera	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-2_XL	.	Xenopus laevis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-2_XT	.	Xenopus tropicalis	L1	1,3,11,4,2
L1-2_ZM	.	Zea mays	L1	43,2,8,44,11,12,1,14,4,75,3,76
L1-20_ALy	.	Arabidopsis lyrata	L1	43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76
L1-20_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,75,3,76,7
L1-20_CPa	.	Carica papaya	L1	43,44,11
L1-20_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,30,16,3
L1-20_FV	.	Fragaria vesca	L1	43,17,44,10,79,7,80
L1-20_LSal	.	Lepeophtheirus salmonis	L1	1,8,3,11,4,12,2
L1-20_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7
L1-20_SSa	.	Salmo salar	L1	
L1-20_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,16,3
L1-20a_Lch	.	Latimeria chalumnae	L1	
L1-20b_Lch	.	Latimeria chalumnae	L1	
L1-20c_Lch	.	Latimeria chalumnae	L1	
L1-21_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,30,3

L1-21_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-21_ATr	.	Amborella trichopoda	L1	
L1-21_CPa	.	Carica papaya	L1	44,43
L1-21_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-21_FV	.	Fragaria vesca	L1	43,17,10,44,79,6,7,80
L1-21_Lch	.	Latimeria chalumnae	L1	
L1-21_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,24,11,4,2
L1-21_NN	.	Nelumbo nucifera	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-21_SSa	.	Salmo salar	L1	2,21,23,24,11,1,26,4,16,3
L1-21_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-21B_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,30,2
L1-22_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-22_ALy	.	Arabidopsis lyrata	L1	16,1,17,43,3,44,10,79,24,11,4,2,80
L1-22_ATr	.	Amborella trichopoda	L1	1,3,4,2
L1-22_CPa	.	Carica papaya	L1	1,43,3,44,4,2
L1-22_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-22_FV	.	Fragaria vesca	L1	
L1-22_Lch	.	Latimeria chalumnae	L1	
L1-22_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-22_NN	.	Nelumbo nucifera	L1	43,79,6,2,7,10,44,24,11,80,1,4,94,16,17,3,95
L1-22_SSa	.	Salmo salar	L1	1,3,11,4,2
L1-22_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-23_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-23_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-23_ATr	.	Amborella trichopoda	L1	7,1,17,43,3,44,10,79,11,4,2,80
L1-23_CPa	.	Carica papaya	L1	44,43
L1-23_DR	.	Danio rerio	L1	2,21,218,23,24,11,1,26,219,4,30,16,3
L1-23_FV	.	Fragaria vesca	L1	75,8,76,78,81,77,12
L1-23_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-23_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,75,3,76,95,7
L1-23_SSa	.	Salmo salar	L1	2,21,23,11,1,4,30,3
L1-23_XT	.	Xenopus tropicalis	L1	
L1-23B_ACar	.	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1-24_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-24_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,17,75,3,76,7
L1-24_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-24_CPa	.	Carica papaya	L1	44,43
L1-24_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-24_FV	.	Fragaria vesca	L1	43,17,10,44,79,7,80
L1-24_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-24_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-24_SSa	.	Salmo salar	L1	21,1,26,3,28,11,4,2
L1-24_XT	.	Xenopus tropicalis	L1	16,1,26,3,24,11,4,2
L1-24B_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-24C_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-25_ACar	.	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1-25_ALy	.	Arabidopsis lyrata	L1	43,78,79,6,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7
L1-25_ATr	.	Amborella trichopoda	L1	16,1,24,11,2

L1-25_CPa	.	Carica papaya	L1	44,43
L1-25_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-25_FV	.	Fragaria vesca	L1	17,10,79,6,7,80
L1-25_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-25_NN	.	Nelumbo nucifera	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-25_SSa	.	Salmo salar	L1	21,1,26,3,28,11,4,2
L1-25_XT	.	Xenopus tropicalis	L1	21,1,23,3,11,4,30,2
L1-26_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-26_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,24,11,80,1,81,77,4,16,75,3,76
L1-26_ATr	.	Amborella trichopoda	L1	7,1,80,17,43,3,44,10,79,4,6,2
L1-26_CPa	.	Carica papaya	L1	44,43
L1-26_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-26_FV	.	Fragaria vesca	L1	
11				
L1-26_LSal	.	Lepeophtheirus salmonis	L1	
L1-26_NN	.	Nelumbo nucifera	L1	
43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,3,7				
L1-26_SSa	.	Salmo salar	L1	1,3,11,4,2
L1-26_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-26B_ACar	.	Anolis carolinensis	L1	21,16,26,1,3,24,11,4,2
L1-27_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-27_ALy	.	Arabidopsis lyrata	L1	
43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7				
L1-27_ATr	.	Amborella trichopoda	L1	43,79,2,44,10,11,80,1,77,4,17,75,3,76,7
L1-27_CPa	.	Carica papaya	L1	
L1-27_DR	.	Danio rerio	L1	
L1-27_FV	.	Fragaria vesca	L1	43,17,44,10,79,6,7,80
L1-27_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,2
L1-27_NN	.	Nelumbo nucifera	L1	7,1,80,17,3,10,79,4,6,2
L1-27_SSa	.	Salmo salar	L1	1,26,3,11,4,2
L1-27_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-27B_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-28_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-28_ALy	.	Arabidopsis lyrata	L1	43,79,2,44,10,11,80,1,77,4,75,17,3,76
L1-28_ATr	.	Amborella trichopoda	L1	
L1-28_CPa	.	Carica papaya	L1	44,43
L1-28_DR	.	Danio rerio	L1	16,26,24,11
L1-28_FV	.	Fragaria vesca	L1	
L1-28_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-28_NN	.	Nelumbo nucifera	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-28_SSa	.	Salmo salar	L1	1,2
L1-28_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-29_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-29_ALy	.	Arabidopsis lyrata	L1	1,17,43,3,44,10,79,4,11,2,80
L1-29_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,75,17,3,76,7
L1-29_CPa	.	Carica papaya	L1	44,43
L1-29_DR	.	Danio rerio	L1	16,26,24,11
L1-29_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-29_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,10,44,11,80,1,4,17,3,7

L1-29_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,3
L1-2A_Cpo	.	Cavia porcellus	L1	
L1-2a_EF	.	Eptesicus fuscus	L1	
L1-2a_MD	.	Monodelphis domestica	L1	21,23,30
L1-2B_ALy	.	Arabidopsis lyrata	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-2B_CPB	.	Chrysemys picta bellii	L1	
L1-2B_Cpo	.	Cavia porcellus	L1	
L1-2B_EC	.	Equus caballus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2B_LA-5end	.	Loxodonta africana	L1	
L1-2B_MPu	.	Mustela putorius furo	L1	
L1-2B1_Cho	.	Choloepus hoffmanni	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-2C1_Cho	.	Choloepus hoffmanni	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-3_AIp	.	Arachis ipaensis	L1	43,79,2,44,10,11,80,1,4,17,3,7
L1-3_ALai	.	Albugo laibachii	L1	1,3,11,4,2
L1-3_ALy	.	Arabidopsis lyrata	L1	
L1-3_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-3_BDi	.	Brachypodium distachyon	L1	43,2,8,44,11,12,1,14,4,75,3,76
L1-3_BRa	.	Brassica rapa	L1	7,1,43,17,3,44,10,79,4,2,80
L1-3_BT	.	Bos taurus	L1	20,2,21,23,25,1,27,28,4,30,31,3
L1-3_BV	.	Beta vulgaris subsp. vulgaris	L1	
L1-3_Cho	.	Choloepus hoffmanni	L1	16,1,26,3,24,4,11,2
L1-3_Cis	.	Ciona savignyi	L1	16,1,14,3,24,11,4,12,2
L1-3_CLan	.	Chinchilla lanigera	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-3_CPa	.	Carica papaya	L1	1,43,3,44,11,4,2
L1-3_CPB	.	Chrysemys picta bellii	L1	21,1,26,23,3,28,11,4,2
L1-3_Cpo	.	Cavia porcellus	L1	20,2,21,23,11,25,1,28,4,30,3
L1-3_CQ	.	Culex quinquefasciatus	L1	16,1,26,43,3,44,24,11,4,92,2
L1-3_CR	.	Chlamydomonas reinhardtii	L1	
L1-3_DR	.	Danio rerio	Tx1	2,8,11,12,42,1,26,14,4,3
L1-3_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-3_ECa	.	Eucalyptus camaldulensis	L1	1,43,75,3,76,44,77,11,4,2
L1-3_EE	.	Erinaceus europaeus	L1	20,2,21,23,25,1,28,4,30,3
L1-3_EPa	.	Schrenkiella parvula	L1	
L1-3_FV	.	Fragaria vesca	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-3_GAr	.	Gossypium arboreum	L1	16,1,43,3,44,24,11,4,2
L1-3_GM	.	Glycine max	L1	16,1,75,43,3,76,44,24,11,4,2
L1-3_HM	.	Hydra vulgaris	L1	1,26,3,11,4,2
L1-3_JC	.	Jatropha curcas	L1	
43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76				
L1-3_LA	.	Loxodonta africana	L1	1,22,3,4,29,2
L1-3_LCh	.	Latimeria chalumnae	L1	2,21,23,11,1,26,4,30,15,16,3
L1-3_LSal	.	Lepeophtheirus salmonis	L1	1,8,14,3,11,4,12,2
L1-3_Mad	.	Malus domestica	L1	43,79,6,2,7,10,44,11,80,1,77,4,75,17,3,76
L1-3_MD	.	Monodelphis domestica	L1	
11				
L1-3_ME	.	Notamacropus eugenii	L1	25,21,20,23,28,30
L1-3_NN	.	Nelumbo nucifera	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-3_NS	.	Nicotiana sylvestris	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-3_NV	.	Nematostella vectensis	L1	
L1-3_Opr	.	Ochotona princeps	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3

L1-3_PAb	.	Picea abies	L1	
43,6,2,10,44,24,11,80,1,14,77,230,4,16,17,75,3,76,7				
L1-3_PI	.	Phytophthora infestans	L1	1,3,11,4,2
L1-3_PR	.	Phytophthora ramorum	L1	1,3,4,2
L1-3_PTr	.	Populus trichocarpa	L1	16,1,3,24,11,4,2
L1-3_PTr	.	Populus trichocarpa	L1	16,1,3,24,11,4,2
L1-3_PVa	.	Pteropus vampyrus	L1	
L1-3_SBi	.	Sorghum bicolor	L1	1,43,44,11,4,2
L1-3_SMo	.	Selaginella moellendorffii	L1	16,1,43,3,44,24,11,4,2
L1-3_SSa	.	Salmo salar	L1	16,21,1,26,3,24,11,4,2
L1-3_SSc	.	Sus scrofa	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1-3_STu	.	Solanum tuberosum	L1	43,2,44,11,1,26,4,94,75,3,76,95
L1-3_Tbel	.	Tupaia belangeri	L1	
L1-3_TC	.	Theobroma cacao	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-3_TML	.	Trichechus manatus latirostris	L1	16,1,26,3,11,4,2,15
L1-3_TS	.	Carlito syrichta	L1	
20,2,21,22,23,11,25,1,26,27,28,4,29,30,15,31,16,3				
L1-3_VC	.	Volvox carteri	L1	16,1,26,3,24,11,4,2
L1-3_Vpa	.	Vicugna pacos	L1	16,1,26,3,24,11,4,2
L1-3_VVi	.	Vitis vinifera	L1	1,43,3,44,11,4,2
L1-3_XL	.	Xenopus laevis	L1	2,21,24,11,1,26,28,4,16,3
L1-3_XT	.	Xenopus tropicalis	L1	2,21,11,1,28,4,3
L1-3_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,16,3
L1-30_AAe	.	Aedes aegypti	L1	16,1,26,43,3,44,24,11,4,2
L1-30_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-30_ALy	.	Arabidopsis lyrata	L1	
43,78,79,2,44,10,24,11,80,1,81,77,4,16,75,3,76				
L1-30_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,11,80,1,4,75,17,3,76,7
L1-30_CPa	.	Carica papaya	L1	44,43
L1-30_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-30_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-30_NN	.	Nelumbo nucifera subsp. nucife	L1	1,3,4,11,2
L1-30_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-31_AAe	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1-31_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-31_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-31_ATr	.	Amborella trichopoda	L1	
L1-31_CPa	.	Carica papaya	L1	44,43
L1-31_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-31_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-31_NN	.	Nelumbo nucifera subsp. nucife	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-31_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-31B_LSal	.	Lepeophtheirus salmonis	L1	
L1-32_AAe	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1-32_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-32_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,11,80,1,81,77,4,75,3,76
L1-32_ATr	.	Amborella trichopoda	L1	1,75,43,3,76,44,11,4,2
L1-32_CPa	.	Carica papaya	L1	44,43
L1-32_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-32_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,11,4,2,15

L1-32_NN	.	Nelumbo nucifera subsp. nucife	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-32_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-33_AAe	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,2
L1-33_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-33_ALy	.	Arabidopsis lyrata	L1	43,78,79,2,44,10,11,80,1,81,77,4,75,3,76
L1-33_ATr	.	Amborella trichopoda	L1	1,11,4,2
L1-33_CPa	.	Carica papaya	L1	44,43
L1-33_DR	.	Danio rerio	L1	2,21,11,1,26,28,4,3
L1-33_LSal	.	Lepeophtheirus salmonis	L1	1,43,3,44,11,4,2
L1-33_NN	.	Nelumbo nucifera subsp. nucife	L1	94,16,1,3,95,24,11,4,2
L1-33_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-34_AAe	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,2
L1-34_ACar	.	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1-34_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,24,11,1,81,77,4,16,75,3,76
L1-34_ATr	.	Amborella trichopoda	L1	
L1-34_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-34_LSal	.	Lepeophtheirus salmonis	L1	1,26,3,11,4,2
L1-34_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-34_XT	.	Xenopus tropicalis	L1	1,26,3,11,4,2
L1-35_AAe	.	Aedes aegypti	L1	16,1,26,43,3,44,24,11,4,2
L1-35_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-35_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,11,1,81,77,4,75,3,76
L1-35_ATr	.	Amborella trichopoda	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-35_DR	.	Danio rerio	L1	21,1,3,28,11,4,2
L1-35_LSal	.	Lepeophtheirus salmonis	L1	1,26,3,4,11,2
L1-35_NN	.	Nelumbo nucifera subsp. nucife	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-35_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-35B_ACar	.	Anolis carolinensis	L1	
L1-36_AAe	.	Aedes aegypti	L1	1,26,3,11,4,2
L1-36_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-36_ALy	.	Arabidopsis lyrata	L1	
43,79,2,10,44,24,11,80,1,77,4,16,17,75,3,76,7				
L1-36_ATr	.	Amborella trichopoda	L1	1,43,3,44,4,2
L1-36_DR	.	Danio rerio	L1	2,21,23,11,1,28,4,3
L1-36_LSal	.	Lepeophtheirus salmonis	L1	
L1-36_NN	.	Nelumbo nucifera subsp. nucife	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-36_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,28,4,15,16,3
L1-37_AAe	.	Aedes aegypti	L1	93,16,1,3,96,24,11,4,98,2
L1-37_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-37_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,10,24,11,80,1,81,77,4,16,75,3,76
L1-37_ATr	.	Amborella trichopoda	L1	
L1-37_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,16,3
L1-37_LSal	.	Lepeophtheirus salmonis	L1	
L1-37_NN	.	Nelumbo nucifera subsp. nucife	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-37_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-38_AAe	.	Aedes aegypti	L1	16,1,26,3,24,11,4,2
L1-38_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2

L1-38_ALy	.	Arabidopsis lyrata	L1	
43,78,2,44,10,24,11,80,1,81,77,4,16,75,17,3,76				
L1-38_ATr	.	Amborella trichopoda	L1	43,79,6,2,7,10,44,11,80,1,77,4,75,17,3,76
L1-38_DR	.	Danio rerio	L1	21,1,26,23,3,11,4,2
L1-38_LSal	.	Lepeophtheirus salmonis	L1	
L1-38_NN	.	Nelumbo nucifera subsp. nucife	L1	1,43,3,44,4,2
L1-38_XT	.	Xenopus tropicalis	L1	21,1,23,3,11,4,2
L1-39_AAe	.	Aedes aegypti	L1	16,1,26,3,24,11,4,92,2
L1-39_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-39_ALy	.	Arabidopsis lyrata	L1	
43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76				
L1-39_ATr	.	Amborella trichopoda	L1	
L1-39_DR	.	Danio rerio	L1	1,43,3,44,4,2
L1-39_LSal	.	Lepeophtheirus salmonis	L1	
L1-39_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-39_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-3A_Cho	.	Choloepus hoffmanni	L1	
L1-3A_Cpo	.	Cavia porcellus	L1	
L1-3a_EF	.	Eptesicus fuscus	L1	
L1-3A_ME	.	Notamacropus eugenii	L1	25,21,20,23,28,30
L1-3B_ALy	.	Arabidopsis lyrata	L1	75,8,76,77,12
L1-3B_SS_3end	.	Sus scrofa	L1	
L1-3C_ALy	.	Arabidopsis lyrata	L1	
43,78,79,6,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7				
L1-4_AIp	.	Arachis ipaensis	L1	
43,79,2,7,101,10,44,11,102,80,1,77,4,75,17,3,76				
L1-4_ALy	.	Arabidopsis lyrata	L1	
L1-4_ATr	.	Amborella trichopoda	L1	94,16,1,43,3,44,95,24,11,4,2
L1-4_BDi	.	Brachypodium distachyon	L1	16,1,3,24,11,4,2
L1-4_BV	.	Beta vulgaris subsp. vulgaris	L1	
43,78,79,2,7,10,44,11,80,1,81,77,4,75,17,3,76				
L1-4_Cho	.	Choloepus hoffmanni	L1	
L1-4_Cis	.	Ciona savignyi	L1	
L1-4_CPa	.	Carica papaya	L1	16,1,43,3,44,11,4,2,15
L1-4_CPB	.	Chrysemys picta bellii	L1	2,21,23,11,1,26,28,4,3
L1-4_Cpo	.	Cavia porcellus	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1-4_CQ	.	Culex quinquefasciatus	L1	93,16,1,26,3,24,11,4,2
L1-4_CR	.	Chlamydomonas reinhardtii	L1	
L1-4_DR	AL807749	Danio rerio	Tx1	2,8,24,11,12,1,26,4,16,3
L1-4_EC	.	Equus caballus	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1-4_ECa	.	Eucalyptus camaldulensis	L1	43,2,44,11,1,77,4,75,3,76
L1-4_EPa	.	Schrenkiella parvula	L1	7,1,43,17,3,44,10,79,4,11,2,80
L1-4_FV	.	Fragaria vesca	L1	
43,78,79,6,2,8,10,44,11,12,80,1,81,77,4,17,75,3,76,7				
L1-4_GAr	.	Gossypium arboreum	L1	43,79,2,10,44,11,80,1,4,17,3,7
L1-4_GM	.	Glycine max	L1	
L1-4_HM	.	Hydra vulgaris	L1	16,1,26,43,3,44,24,11,4,2
L1-4_JC	.	Jatropha curcas	L1	43,78,79,2,10,44,24,11,80,1,81,4,16,17,3,7
L1-4_LA	.	Loxodonta africana	L1	
L1-4_LCh	.	Latimeria chalumnae	L1	16,1,26,3,24,11,4,2

L1-4_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-4_Mad	.	Malus domestica	L1	
43,79,6,2,7,8,10,44,11,12,80,1,77,4,75,17,3,76				
L1-4_MD	.	Monodelphis domestica	L1	
L1-4_ME	.	Notamacropus eugenii	L1	
L1-4_NN	.	Nelumbo nucifera	L1	
43,79,6,2,7,10,44,24,11,80,1,4,94,16,75,17,3,76,95				
L1-4_NS	.	Nicotiana sylvestris	L1	43,79,6,2,7,10,44,11,80,1,77,4,75,17,3,76
L1-4_NV	.	Nematostella vectensis	L1	
L1-4_PAb	.	Picea abies	L1	
43,79,6,2,10,44,24,11,80,1,77,4,16,17,75,3,76,7				
L1-4_PI	.	Phytophthora infestans	L1	1,8,3,11,4,12,2
L1-4_PR	.	Phytophthora ramorum	L1	
L1-4_PTr	.	Populus trichocarpa	L1	16,1,43,3,44,24,11,4,2
L1-4_PTr	.	Populus trichocarpa	L1	16,1,43,3,44,24,11,4,2
L1-4_PVa	.	Pteropus vampyrus	L1	
L1-4_SBi	.	Sorghum bicolor	L1	
L1-4_SMo	.	Selaginella moellendorffii	L1	16,1,43,3,44,24,11,4,2
L1-4_SSa	.	Salmo salar	L1	2,21,23,11,1,26,4,3
L1-4_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-4_TML	.	Trichechus manatus latirostris	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-4_TS	.	Carlito syrichta	L1	
L1-4_VC	.	Volvox carteri	L1	16,1,26,3,24,11,4,2
L1-4_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-4_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,16,3
L1-40_AAe	.	Aedes aegypti	L1	16,1,26,43,3,44,24,11,4,2
L1-40_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-40_ALy	.	Arabidopsis lyrata	L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-40_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-40_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-40_NN	.	Nelumbo nucifera subsp. nucife	L1	1,43,3,44,11,4,2
L1-40_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-40B_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-41_AAe	.	Aedes aegypti	L1	16,1,3,24,11,4,2
L1-41_ACar	.	Anolis carolinensis	L1	21,1,26,3,11,4,2
L1-41_ALy	.	Arabidopsis lyrata	L1	
43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76				
L1-41_LSal	.	Lepeophtheirus salmonis	L1	1,3,2
L1-41_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,44,10,80,1,4,17,3,7
L1-41_XT	.	Xenopus tropicalis	L1	1,26,3,11,4,2
L1-42_AAe	.	Aedes aegypti	L1	1,3,11,4,2
L1-42_ACar	.	Anolis carolinensis	L1	2,21,24,11,1,26,28,4,16,3
L1-42_ALy	.	Arabidopsis lyrata	L1	43,78,2,44,11,12,1,14,81,77,4,75,3,76
L1-42_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-42_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,11,4,2,15
L1-42_NN	.	Nelumbo nucifera subsp. nucife	L1	1,43,3,44,11,4,2
L1-42_XT	.	Xenopus tropicalis	L1	21,1,26,23,3,11,4,2
L1-42A_XT	.	Xenopus tropicalis	L1	21,1,23,3,11,4,2
L1-43_AAe	.	Aedes aegypti	L1	93,16,1,3,24,11,4,92,2
L1-43_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-43_DR	.	Danio rerio	L1	21,1,3,28,11,4,2

L1-43_LSal	.	Lepeophtheirus salmonis	L1	1,14,3,11,4,2
L1-43_NN	.	Nelumbo nucifera subsp. nucife	L1	
43,79,6,2,44,10,24,11,80,1,4,94,16,17,3,95,7				
L1-43_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,16,3
L1-44_AAe	.	Aedes aegypti	L1	16,1,3,24,11,4,2
L1-44_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-44_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-44_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-44_NN	.	Nelumbo nucifera subsp. nucife	L1	1,43,75,3,76,44,11,4,2
L1-44_XT	.	Xenopus tropicalis	L1	21,1,26,23,3,11,4,2
L1-45_AAe	.	Aedes aegypti	L1	93,16,1,26,3,24,11,4,2
L1-45_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-45_DR	.	Danio rerio	L1	26,11
L1-45_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,11,2
L1-45_NN	.	Nelumbo nucifera subsp. nucife	L1	
43,78,79,2,7,8,10,44,11,12,80,1,81,77,4,17,3				
L1-45_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-46_AAe	.	Aedes aegypti	L1	16,1,26,43,3,44,24,11,4,2
L1-46_ACar	.	Anolis carolinensis	L1	21,1,26,3,11,4,2
L1-46_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-46_LSal	.	Lepeophtheirus salmonis	L1	1,11,2
L1-46_NN	.	Nelumbo nucifera subsp. nucife	L1	1,43,3,44,11,4,2
L1-46_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,16,3
L1-47_AAe	.	Aedes aegypti	L1	16,1,26,3,24,4,11,2
L1-47_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-47_DR	.	Danio rerio	L1	21,26,23,28,11,30
L1-47_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-47_NN	.	Nelumbo nucifera subsp. nucife	L1	16,1,43,3,44,24,11,4,2
L1-47_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-47B_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-48_AAe	.	Aedes aegypti	L1	1,43,3,44,11,4,2
L1-48_ACar	.	Anolis carolinensis	L1	21,1,26,3,11,4,2
L1-48_DR	.	Danio rerio	L1	2,21,23,228,11,1,26,4,3,229
L1-48_LSal	.	Lepeophtheirus salmonis	L1	
11				
L1-48_XT	.	Xenopus tropicalis	L1	20,2,21,23,11,25,1,26,4,3
L1-49_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1-49_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-49_LSal	.	Lepeophtheirus salmonis	L1	
L1-49_NN	.	Nelumbo nucifera subsp. nucife	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-49_XT	.	Xenopus tropicalis	L1	
L1-4A_Cpo	.	Cavia porcellus	L1	
L1-4a_EF	.	Eptesicus fuscus	L1	
L1-4A_MD	.	Monodelphis domestica	L1	
L1-4A_ME	.	Notamacropus eugenii	L1	
L1-4B_DR	.	Danio rerio	Tx1	43,2,8,44,11,12,1,26,4,15,16,3
L1-4B_EC	.	Equus caballus	L1	20,2,21,23,25,1,27,28,4,30,31,3
L1-4B_LA-5end	.	Loxodonta africana	L1	
L1-4B_NS-3end	.	Nicotiana sylvestris	L1	
L1-4C_NS-5end	.	Nicotiana sylvestris	L1	

L1-5_A5end_EC	.	Equus caballus	L1	
L1-5_AIp	.	Arachis ipaensis	L1	43,79,6,2,10,44,11,80,1,77,4,17,75,3,76,7
L1-5_ALy	.	Arabidopsis lyrata	L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-5_ATr	.	Amborella trichopoda	L1	43,79,6,2,44,10,11,80,1,77,4,17,75,3,76,7
L1-5_B5end_EC	.	Equus caballus	L1	
L1-5_BDi	.	Brachypodium distachyon	L1	1,43,3,44,11,4,2
L1-5_BV	.	Beta vulgaris	L1	
L1-5_Cis	.	Ciona savignyi	L1	
L1-5_CPa	.	Carica papaya	L1	1,3,4,2
L1-5_CPB	.	Chrysemys picta bellii	L1	
L1-5_CQ	.	Culex quinquefasciatus	L1	16,1,26,3,24,11,4,2
L1-5_CR	.	Chlamydomonas reinhardtii	L1	
L1-5_DR	.	Danio rerio	Tx1	2,8,11,12,42,1,26,14,4,3
L1-5_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-5_ECa	.	Eucalyptus camaldulensis	L1	75,76,78,81,77
L1-5_EPa	.	Schrenkiella parvula	L1	44,43
L1-5_FV	.	Fragaria vesca	L1	1,43,17,44,10,79,2,6,7,80
L1-5_GAr	.	Gossypium arboreum	L1	16,1,43,3,44,10,24,11,4,2,80
L1-5_GM	.	Glycine max	L1	43,2,44,11,1,77,4,75,3,76
L1-5_HM	.	Hydra vulgaris	L1	16,1,26,43,3,44,24,11,4,2
L1-5_JC	.	Jatropha curcas	L1	
L1-5_LA	.	Loxodonta africana	L1	1,3,4,2
L1-5_LCh	.	Latimeria chalumnae	L1	2,21,23,24,11,1,26,4,30,16,3
L1-5_LSal	.	Lepeophtheirus salmonis	L1	1,8,14,3,11,4,12,2
L1-5_Mad	.	Malus domestica	L1	43,78,79,2,44,10,11,80,1,81,77,4,17,75,76,7
L1-5_MD	.	Monodelphis domestica	L1	
L1-5_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-5_PI	.	Phytophthora infestans	L1	6,2,10,11,80,1,26,4,17,3,7
L1-5_PR	.	Phytophthora ramorum	L1	1,3,4,2
L1-5_PTr	.	Populus trichocarpa	L1	43,79,6,2,8,10,44,11,12,80,1,77,4,17,75,3,76,7
L1-5_SBi	.	Sorghum bicolor	L1	1,43,44,11,4,2
L1-5_SSa	.	Salmo salar	L1	16,1,26,3,24,11,4,2
L1-5_STu	.	Solanum tuberosum	L1	
L1-5_TS	.	Carlito syrichta	L1	
L1-5_VC	.	Volvox carteri	L1	16,1,26,3,24,11,4,2
L1-5_XT	.	Xenopus tropicalis	L1	16,1,26,3,24,4,11,2
L1-5_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,94,16,3,95
L1-50_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-50_DR	.	Danio rerio	L1	21,1,23,3,11,4,2
L1-50_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-50_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,7,10,44,24,11,80,1,4,94,16,17,3,95
L1-50_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,16,3
L1-50A_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,16,3
L1-51_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-51_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-51_LSal	.	Lepeophtheirus salmonis	L1	11
L1-51_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7

L1-51_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,16,3
L1-52_ACar	.	Anolis carolinensis	L1	21,1,23,3,11,4,2
L1-52_DR	.	Danio rerio	L1	21,1,3,28,11,4,2
L1-52_LSal	.	Lepeophtheirus salmonis	L1	16,1,3,11,15,2
L1-52_NN	.	Nelumbo nucifera subsp. nucife	L1	43,79,6,2,44,10,11,80,1,4,17,3,7
L1-52_XT	.	Xenopus tropicalis	L1	20,2,21,23,24,11,25,1,26,4,16,3
L1-53_ACar	.	Anolis carolinensis	L1	21,1,23,3,11,4,2
L1-53_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-53_LSal	.	Lepeophtheirus salmonis	L1	
11				
L1-53_NN	.	Nelumbo nucifera subsp. nucife	L1	
43,79,6,2,7,44,10,24,11,80,1,77,4,16,75,17,3,76				
L1-53_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3
L1-54_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-54_DR	.	Danio rerio	L1	1,11,2
L1-54_LSal	.	Lepeophtheirus salmonis	L1	
L1-54_NN	.	Nelumbo nucifera subsp. nucife	L1	1,43,3,44,4,2
L1-54_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3
L1-55_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-55_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,30,16,3
L1-55_LSal	.	Lepeophtheirus salmonis	L1	
L1-55_NN	.	Nelumbo nucifera subsp. nucife	L1	
43,79,6,2,44,10,24,11,80,1,4,94,16,17,3,95,7				
L1-55_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
L1-56_ACar	.	Anolis carolinensis	L1	2,21,24,11,1,26,28,4,16,3
L1-56_DR	.	Danio rerio	L1	21,1,23,3,11,4,2
L1-56_LSal	.	Lepeophtheirus salmonis	L1	
L1-56_NN	.	Nelumbo nucifera subsp. nucife	L1	7,1,43,17,3,44,10,79,4,6,2,80
L1-56_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,11,12,1,26,14,4,3
L1-57_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-57_DR	.	Danio rerio	L1	26,11
L1-57_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-57_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,30,16,3
L1-58_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,15,16,3
L1-58_DR	.	Danio rerio	L1	
L1-58_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-58_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,16,3
L1-59_ACar	.	Anolis carolinensis	L1	1,26,23,3,4,11,2
L1-59_DR	.	Danio rerio	L1	21,16,1,26,3,4,15,11,2
L1-59_LSal	.	Lepeophtheirus salmonis	L1	
L1-59_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,28,4,16,3
L1-5A_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-5A2_EC	.	Equus caballus	L1	21,16,26,23,24,11,30
L1-5B_EC	.	Equus caballus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-6_AIp	.	Arachis ipaensis	L1	43,79,2,10,44,11,80,1,77,4,17,75,3,76,7
L1-6_ALy	.	Arabidopsis lyrata	L1	43,79,2,10,44,24,11,80,1,4,16,17,3,7
L1-6_ATr	.	Amborella trichopoda	L1	
43,79,6,2,44,10,24,11,80,1,4,94,16,17,3,95,7				
L1-6_BDi	.	Brachypodium distachyon	L1	1,43,3,44,11,4,2
L1-6_BV	.	Beta vulgaris	L1	
43,78,79,2,7,8,10,44,11,12,80,1,81,4,75,17,3,76				

L1-6_Cis	.	Ciona savignyi	L1	
L1-6_CPa	.	Carica papaya	L1	1,3,11,4,2
L1-6_CPB	.	Chrysemys picta bellii	L1	2,21,23,11,1,26,4,30,3
L1-6_CQ	.	Culex quinquefasciatus	L1	16,1,26,3,24,11,4,2
L1-6_CR	.	Chlamydomonas reinhardtii	L1	16,1,26,24,11,4,2
L1-6_DR	AL627249	Danio rerio	L1	2,21,23,24,11,1,26,28,4,16,3
L1-6_ECa	.	Eucalyptus camaldulensis	L1	16,1,43,3,44,24,11,4,2
L1-6_FV	.	Fragaria vesca	L1	7,1,80,17,43,3,44,10,79,6,2
L1-6_GAr	.	Gossypium arboreum	L1	
43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7				
L1-6_GM	.	Glycine max	L1	1,75,43,3,76,44,11,4,2
L1-6_JC	.	Jatropha curcas	L1	43,79,2,10,44,11,80,1,4,17,3,7
L1-6_LA	.	Loxodonta africana	L1	
L1-6_Laf	.	Loxodonta africana	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1-6_LCh	.	Latimeria chalumnae	L1	2,21,23,24,11,1,26,4,16,3
L1-6_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,11,2
L1-6_Mad	.	Malus domestica	L1	7,1,17,43,3,44,10,79,4,11,2,80
L1-6_MD	.	Monodelphis domestica	L1	
L1-6_NN	.	Nelumbo nucifera	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-6_PI	.	Phytophthora infestans	L1	6,2,10,11,80,1,230,4,17,3,7
L1-6_PR	.	Phytophthora ramorum	L1	1,3,4,2
L1-6_SBi	.	Sorghum bicolor	L1	1,43,75,3,76,44,11,4,2
L1-6_SSa	.	Salmo salar	L1	1,26,3,11,4,2
L1-6_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-6_TS	.	Carlito syrichta	L1	
L1-6_VC	.	Volvox carteri	L1	
L1-6_XT	.	Xenopus tropicalis	L1	2,21,11,1,26,28,4,30,3
L1-6_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,4,16,3
L1-60_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,11,4,2,15
L1-60_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-60_LSal	.	Lepeophtheirus salmonis	L1	1,8,3,4,11,12,2
L1-61_ACar	.	Anolis carolinensis	L1	21,1,26,3,28,11,4,2
L1-61_DR	.	Danio rerio	L1	21,1,26,11,2
L1-61_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-62_DR	.	Danio rerio	L1	
L1-62_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-62_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-63_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-63_DR	.	Danio rerio	L1	1,26,28,11,4,2
L1-63_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-63_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-64_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-64_DR	.	Danio rerio	L1	1,26,11,2
L1-64_LSal	.	Lepeophtheirus salmonis	L1	
L1-64_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-65_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-65_DR	.	Danio rerio	L1	
L1-65_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-65_XT	.	Xenopus tropicalis	L1	
L1-66_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2

L1-66_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-66_LSal	.	Lepeophtheirus salmonis	L1	
L1-66_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-67_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-67_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,30,16,3
L1-67_LSal	.	Lepeophtheirus salmonis	L1	1,3,4,11,2
L1-67_XT	.	Xenopus tropicalis	L1	16,1,26,3,24,11,4,2
L1-68_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-68_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-68_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-68_XT	.	Xenopus tropicalis	L1	2,21,23,11,1,26,4,15,118,16,3
L1-69_ACar	.	Anolis carolinensis	L1	1,3,4,2
L1-69_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-69_LSal	.	Lepeophtheirus salmonis	L1	1,2
L1-6A_XT	.	Xenopus tropicalis	L1	16,1,26,3,24,11,4,2
L1-6B_JC	.	Jatropha curcas	L1	
L1-7_AIp	.	Arachis ipaensis	L1	1,75,43,3,76,44,11,4,2
L1-7_ALy	.	Arabidopsis lyrata	L1	7,1,43,17,3,44,10,79,4,2,80
L1-7_ATr	.	Amborella trichopoda	L1	1,3,11,4,2
L1-7_BDi	.	Brachypodium distachyon	L1	
L1-7_BV	.	Beta vulgaris	L1	
L1-7_Cis	.	Ciona savignyi	L1	1,14,3,11,4,12,2
L1-7_CPa	.	Carica papaya	L1	1,43,3,44,4,2
L1-7_CPB	.	Chrysemys picta bellii	L1	
L1-7_CQ	.	Culex quinquefasciatus	L1	16,1,3,24,11,4,2
L1-7_CR	.	Chlamydomonas reinhardtii	L1	
L1-7_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-7_ECa	.	Eucalyptus camaldulensis	L1	1,43,3,44,4,2
L1-7_FV	.	Fragaria vesca	L1	43,17,10,44,79,6,7,80
L1-7_GAr	.	Gossypium arboreum	L1	
43,78,79,2,7,10,44,24,11,80,1,81,77,4,16,75,17,3,76				
L1-7_GM	.	Glycine max	L1	
L1-7_HM	.	Hydra vulgaris	L1	1,26,3,11,4,2
L1-7_JC	.	Jatropha curcas	L1	7,1,17,43,3,44,10,79,4,11,2,80
L1-7_LCh	.	Latimeria chalumnae	L1	16,1,3,24,4,11,2
L1-7_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-7_Mad	.	Malus domestica	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-7_MD	.	Monodelphis domestica	L1	
L1-7_NN	.	Nelumbo nucifera	L1	
43,79,6,2,10,44,24,11,80,1,4,94,16,17,3,95,7				
L1-7_PI	.	Phytophthora infestans	L1	1,3,11,4,2
L1-7_PR	.	Phytophthora ramorum	L1	8,11,12
L1-7_SBi	.	Sorghum bicolor	L1	1,3,11,4,2
L1-7_SSa	.	Salmo salar	L1	2,21,23,24,11,1,26,4,30,16,3
L1-7_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
L1-7_XT	.	Xenopus tropicalis	L1	2,21,23,24,11,1,26,4,16,3
L1-7_ZM	.	Zea mays	L1	
43,78,79,2,8,10,44,11,12,80,1,81,77,4,17,75,3,76,7				
L1-70_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-70_DR	.	Danio rerio	L1	44,43
L1-70_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2

L1-71_ACar	.	Anolis carolinensis	L1	2,21,23,11,1,26,4,3
L1-71_DR	.	Danio rerio	L1	
L1-72_ACar	.	Anolis carolinensis	L1	1,26,3,4,11,2
L1-72_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,16,3
L1-73_ACar	.	Anolis carolinensis	L1	21,1,23,3,11,4,2
L1-73_DR	.	Danio rerio	L1	
L1-74_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1-74_DR	.	Danio rerio	L1	43,2,21,44,24,11,1,26,4,16,3
L1-75_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-75_DR	.	Danio rerio	L1	1,26,3,11,4,2
L1-76_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-77_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,30,16,3
L1-78_ACar	.	Anolis carolinensis	L1	
L1-78_DR	.	Danio rerio	L1	2,21,23,11,1,26,28,4,3
L1-79_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,11,4,2,15
L1-79_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,16,3
L1-79B_ACar	.	Anolis carolinensis	L1	
L1-7A_XT	.	Xenopus tropicalis	L1	
L1-8_AIp	.	Arachis ipaensis	L1	16,1,43,3,44,24,11,4,2
L1-8_ALy	.	Arabidopsis lyrata	L1	1,43,3,44,11,4,2
L1-8_BDi	.	Brachypodium distachyon	L1	
L1-8_Cis	.	Ciona savignyi	L1	1,3,11,4,2
L1-8_CPa	.	Carica papaya	L1	1,43,3,44,11,4,2
L1-8_CQ	.	Culex quinquefasciatus	L1	16,1,43,3,44,24,11,4,2
L1-8_DR	.	Danio rerio	L1	2,21,23,24,11,1,26,4,16,3
L1-8_ECa	.	Eucalyptus camaldulensis	L1	43,2,44,11,1,77,4,75,3,76
L1-8_FV	.	Fragaria vesca	L1	43,17,44,10,79,6,7,80
L1-8_GAr	.	Gossypium arboreum	L1	43,79,2,10,44,11,80,1,4,17,3,7
L1-8_GM	.	Glycine max	L1	1,43,75,3,44,76,11,4,2
L1-8_HMa	.	Hydra vulgaris	L1	1,43,3,44,11,4,2
L1-8_JC	.	Jatropha curcas	L1	43,17,10,44,79,7,80
L1-8_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-8_Mad	.	Malus domestica	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
L1-8_MD	.	Monodelphis domestica	L1	
L1-8_NN	.	Nelumbo nucifera	L1	
43,79,6,2,7,10,44,24,11,80,1,4,94,16,17,3,95				
L1-8_PI	.	Phytophthora infestans	L1	16,1,3,24,11,4,2
L1-8_SBi	.	Sorghum bicolor	L1	43,2,8,44,11,12,1,14,4,15,16,75,3,76
L1-8_SSa	.	Salmo salar	L1	1,26,3,11,4,2
L1-8_STu	.	Solanum tuberosum	L1	43,2,44,24,11,1,4,16,75,3,76
L1-8_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,30,16,3
L1-80_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-81_ACar	.	Anolis carolinensis	L1	1,26,23,3,11,4,2
L1-81_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-82_ACar	.	Anolis carolinensis	L1	1,23,3,11,4,2
L1-82_DR	.	Danio rerio	L1	2,21,24,11,1,26,28,4,16,3
L1-83_ACar	.	Anolis carolinensis	L1	116,2,21,23,11,1,26,4,3
L1-83_DR	.	Danio rerio	L1	20,2,21,11,25,1,26,28,4,3
L1-84_ACar	.	Anolis carolinensis	L1	21,1,26,23,3,11,4,2
L1-84_DR	.	Danio rerio	L1	21,1,26,3,28,11,4,2
L1-85_ACar	.	Anolis carolinensis	L1	1,26,3,28,11,4,2

L1-85_DR	.	Danio rerio	L1	1,26,43,44,11,4,2
L1-86_DR	.	Danio rerio	L1	21,1,23,3,11,4,30,2
L1-87_ACar	.	Anolis carolinensis	L1	
L1-87_DR	.	Danio rerio	L1	
L1-88_ACar	.	Anolis carolinensis	L1	16,1,26,23,3,24,11,4,2
L1-88_DR	.	Danio rerio	L1	21,1,3,28,4,2
L1-89_ACar	.	Anolis carolinensis	L1	1,3,4,2
L1-89_DR	.	Danio rerio	L1	2,1
L1-9_AIp	.	Arachis ipaensis	L1	43,78,2,44,24,11,1,81,77,4,16,75,3,76
L1-9_ALy	.	Arabidopsis lyrata	L1	43,2,44,24,11,1,77,4,16,75,3,76
L1-9_ATr	.	Amborella trichopoda	L1	
L1-9_CPa	.	Carica papaya	L1	1,43,3,44,4,2
L1-9_CQ	.	Culex quinquefasciatus	L1	
92				
L1-9_DR	.	Danio rerio	L1	
L1-9_ECa	.	Eucalyptus camaldulensis	L1	1,43,3,44,11,4,2
L1-9_FV	.	Fragaria vesca	L1	1,43,17,44,10,79,2,7,80
L1-9_GAr	.	Gossypium arboreum	L1	
43,78,79,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76,7				
L1-9_GM	.	Glycine max	L1	
L1-9_JC	.	Jatropha curcas	L1	75,8,78,76,81,77,12
L1-9_LCh	.	Latimeria chalumnae	L1	16,1,26,3,24,4,11,2
L1-9_LSal	.	Lepeophtheirus salmonis	L1	1,3,11,4,2
L1-9_Mad	.	Malus domestica	L1	
43,78,79,2,8,44,10,11,12,80,1,81,77,4,17,75,76,7				
L1-9_NN	.	Nelumbo nucifera	L1	43,79,6,2,10,44,24,11,80,1,4,16,17,3,7
L1-9_NS	.	Nicotiana sylvestris	L1	1,75,43,3,76,44,10,79,11,4,2,80
L1-9_PI	.	Phytophthora infestans	L1	1,3,11,4,2
L1-9_STu	.	Solanum tuberosum	L1	94,16,1,43,3,44,95,24,11,4,2
L1-9_XT	.	Xenopus tropicalis	L1	2,21,24,11,1,26,28,4,16,3
L1-90_ACar	.	Anolis carolinensis	L1	16,1,26,3,24,11,4,2
L1-90_DR	.	Danio rerio	L1	
L1-91_ACar	.	Anolis carolinensis	L1	1,26,3,11,4,2
L1-91_DR	.	Danio rerio	L1	
L1-92_ACar	.	Anolis carolinensis	L1	1,3,4,2
L1-92_Cis	.	Ciona savignyi	L1	1,26,8,3,11,4,12,2
L1-92_DR	.	Danio rerio	L1	21,1,26,23,11,4,30,2
L1-93_ACar	.	Anolis carolinensis	L1	21,1,26,3,28,11,4,2
L1-93_Cis	.	Ciona savignyi	L1	16,1,26,3,24,4,11,2
L1-94_ACar	.	Anolis carolinensis	L1	1,3,4,2
L1-94_Cis	.	Ciona savignyi	L1	1,26,43,3,44,11,4,2
L1-94_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,3
L1-95_ACar	.	Anolis carolinensis	L1	2,21,23,24,11,1,26,4,16,3
L1-95_Cis	.	Ciona savignyi	L1	16,1,8,26,3,11,4,12,2,15
L1-95_DR	.	Danio rerio	L1	20,2,21,23,11,25,1,26,28,4,30,3
L1-96_ACar	.	Anolis carolinensis	L1	
23				
L1-96_Cis	.	Ciona savignyi	L1	187,16,1,26,3,24,11,4,188,2
L1-96_DR	.	Danio rerio	L1	
L1-97_Cis	.	Ciona savignyi	L1	43,2,8,44,11,185,12,1,26,4,15,16,3
L1-97_DR	.	Danio rerio	L1	2,21,23,11,1,26,4,30,15,16,3

L1-98_Cis	.	Ciona savignyi	L1	43,1,26,3,44,11,4,2
L1-98_DR	.	Danio rerio	L1	
L1-99_Cis	.	Ciona savignyi	L1	
L1-99_DR	.	Danio rerio	L1	
L1-99B_DR	.	Danio rerio	L1	
L1-99ext_Cis	.	Ciona savignyi	L1	
L1-BT	DQ000238	Bos taurus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1-N1_CQ	.	Culex quinquefasciatus	L1	
L1-N1_FV	.	Fragaria vesca	L1	75,76,77
L1-N2_CQ	.	Culex quinquefasciatus	L1	
L1-N2_FV	.	Fragaria vesca	L1	75,76,78,81,77
L1-N3_CQ	.	Culex quinquefasciatus	L1	
L1-N3_FV	.	Fragaria vesca	L1	75,76,77
L1-N4_CQ	.	Culex quinquefasciatus	L1	
L1-N4_FV	.	Fragaria vesca	L1	75,76,77
L1-N5_CQ	.	Culex quinquefasciatus	L1	
L1-N6_CQ	.	Culex quinquefasciatus	L1	
L1-Y_CF	.	Canis lupus familiaris	L1	
L1A_Mim	.	Microcebus murinus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1A_OC	.	Oryctolagus cuniculus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1A-1_MD	.	Monodelphis domestica	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1A-2_MD	.	Monodelphis domestica	L1	20,2,21,23,24,11,25,1,26,28,4,30,16,3
L1A2_OC	.	Oryctolagus cuniculus	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1B_BM	.	Bombyx mori	Jockey	
L1B_Mim	.	Microcebus murinus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1B_MM	.	Mus musculus	L1	
L1B_OC	M94361	Oryctolagus cuniculus	L1	
L1B_RN	.	Rattus norvegicus	L1	
20,2,21,23,22,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1B_SS	.	Sus scrofa	L1	20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3
L1B-1_Cho	.	Choloepus hoffmanni	L1	20,2,21,23,11,25,1,26,28,4,30,15,16,3
L1BM	AB002279	Bombyx mori	Jockey	16,1,11,2,15
L1C_Mim	.	Microcebus murinus	L1	20,2,21,23,24,11,25,1,26,27,28,4,30,31,16,3
L1C_OC	M18818, X07786, X077	Oryctolagus cuniculus	L1	
L1C_SS_5end	.	Sus scrofa	L1	
L1HS	.	Homo sapiens	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1Lx_I	.	Mus musculus	L1	
L1Lx_II	.	Mus musculus	L1	
L1Lx_III	.	Mus musculus	L1	
L1Lx_IV	.	Mus musculus	L1	
L1M1_5	.	Eutheria	L1	
L1M1B_5	.	Homo sapiens	L1	
L1M2_5	.	Homo sapiens	L1	
L1M2_5	.	Homo sapiens	L1	
L1M2A_5	.	Homo sapiens	L1	
L1M2A_5	.	Homo sapiens	L1	
L1M2A1_5	.	Homo sapiens	L1	
L1M2B_5	.	Homo sapiens	L1	
L1M2B_5	.	Homo sapiens	L1	
L1M2C_5	.	Homo sapiens	L1	

L1M2C_5	Homo sapiens	L1
L1M3A_5	Eutheria	L1
L1M3A_5	Eutheria	L1
L1M3B_5	Homo sapiens	L1
L1M3C_5	Homo sapiens	L1
L1M3C_5	Homo sapiens	L1
L1M3D_5	Eutheria	L1
L1M3D_5	Eutheria	L1
L1M3DE_5	Eutheria	L1
L1M3DE_5	Homo sapiens	L1
L1M4B	Mammalia	L1
L1M6_5end	Mammalia	L1
L1M6_5end	Mammalia	L1
L1M6B_5end	Homo sapiens	L1
L1M7_5end	Mammalia	L1
L1M7_5end	Mammalia	L1
L1MA1	Homo sapiens	L1
L1MA10	Eutheria	L1
L1MA10	Eutheria	L1
L1MA10	Eutheria	L1
L1MA2	Homo sapiens	L1
L1MA2	Homo sapiens	L1
L1MA2	Homo sapiens	L1
L1MA3	Homo sapiens	L1
L1MA4	Homo sapiens	L1
L1MA4A	Homo sapiens	L1
L1MA5	Homo sapiens	L1
L1MA5	Homo sapiens	L1
L1MA5	Homo sapiens	L1
L1MA5A	Homo sapiens	L1
L1MA6	Homo sapiens	L1
L1MA7	Homo sapiens	L1
L1MA8	Homo sapiens	L1
L1MA9	Homo sapiens	L1
L1MA9	Homo sapiens	L1
L1MA9	Homo sapiens	L1
L1MA9_5	Eutheria	L1
L1MAB_5end_EC	Equus caballus	L1
L1MAB_EC	Equus caballus	L1
L1MAB_ML	Myotis lucifugus	L1
L1MAB1_5_EC	Equus caballus	L1
L1MAB2_ML	Myotis lucifugus	L1
L1MB1	Homo sapiens	L1
L1MB2	Homo sapiens	L1
L1MB3	Homo sapiens	L1
L1MB3	Homo sapiens	L1
L1MB3	Homo sapiens	L1
L1MB3_5	Homo sapiens	L1
L1MB3_EC	Equus caballus	L1

20,2,21,22,23,24,11,25,1,26,28,4,29,30,16,3

1,22,3,4,11,29,2

L1MB4	.	Homo sapiens	L1
L1MB4_5	.	Eutheria	L1
L1MB4_Laf	.	Loxodonta africana	L1
L1MB5	.	Homo sapiens	L1
L1MB6_5	.	Eutheria	L1
L1MB6_5	.	Eutheria	L1
L1MB6_5	.	Eutheria	L1
L1MB7	.	Eutheria	L1
L1MB7	.	Eutheria	L1
L1MB7	.	Eutheria	L1
L1MB8	.	Homo sapiens	L1
L1MC1	.	Homo sapiens	L1
L1MC1_EC	.	Equus caballus	L1
L1MC2	.	Homo sapiens	L1
L1MC2	.	Homo sapiens	L1
L1MC2	.	Homo sapiens	L1
L1MC3	.	Eutheria	L1
L1MC3	.	Eutheria	L1
L1MC3	.	Eutheria	L1
L1MC4	.	Homo sapiens	L1
L1MC4	.	Homo sapiens	L1
L1MC4	.	Homo sapiens	L1
L1MC4_5end	.	Eutheria	L1
L1MC4_5end	.	Eutheria	L1
L1MC4_5end	.	Eutheria	L1
L1MC4B	.	Eutheria	L1
L1MC4B	.	Eutheria	L1
L1MC4B	.	Eutheria	L1
L1MC5	.	Eutheria	L1
L1MC5	.	Eutheria	L1
L1MCA_5	.	Homo sapiens	L1
L1MCA_5	.	Homo sapiens	L1
L1MCB_5	.	Homo sapiens	L1
L1MCB_5	.	Homo sapiens	L1
L1MCC_5	.	Homo sapiens	L1
L1MCC_5	.	Homo sapiens	L1
L1Md_F_5end	.	Mus	L1
L1Md_Gf_5end	.	Mus	L1
L1MD1	.	Homo sapiens	L1
L1MD1	.	Homo sapiens	L1
L1MD1_5	.	Eutheria	L1
L1MD2	.	Homo sapiens	L1
L1MD2	.	Homo sapiens	L1
L1MD2	.	Homo sapiens	L1
L1MD3	.	Homo sapiens	L1
L1MD3	.	Homo sapiens	L1
L1MDA_5	.	Eutheria	L1
L1MDA_5	.	Eutheria	L1
L1MdA_I	.	Mus musculus	L1
L1MdA_II	.	Mus musculus	L1
L1MdA_III	.	Mus musculus	L1

21,1,23,3,11,4,30,2

L1MdA_IV	Mus musculus	L1
L1MdA_V	Mus musculus	L1
L1MdA_VI	Mus musculus	L1
L1MdA_VII	Mus musculus	L1
L1MDB_5	Homo sapiens	L1
L1MDB_5	Homo sapiens	L1
L1MdF_I	Mus musculus	L1
L1MdF_II	Mus musculus	L1
L1MdF_III	Mus musculus	L1
L1MdF_IV	Mus musculus	L1
L1MdF_V	Mus musculus	L1
L1MdFanc_I	Mus musculus	L1
L1MdFanc_II	Mus musculus	L1
L1MdGf_I	Mus musculus	L1
L1MdGf_II	Mus musculus	L1
L1MdMus_I	Mus musculus	L1
L1MdMus_II	Mus musculus	L1
L1MdN_I	Mus musculus	L1
L1MdTf_I	Mus musculus	L1
L1MdTf_II	Mus musculus	L1
L1MdTf_III	Mus musculus	L1
L1MdV_I	Mus musculus	L1
L1MdV_II	Mus musculus	L1
L1MdV_III	Mus musculus	L1
L1ME_ORF2	Homo sapiens	L1
L1ME_ORF2	Homo sapiens	L1
L1ME1	Homo sapiens	L1
L1ME2	Homo sapiens	L1
L1ME2	Homo sapiens	L1
L1ME3	Homo sapiens	L1
L1ME3A	Homo sapiens	L1
L1ME3A	Homo sapiens	L1
L1ME3C_3end	Eutheria	L1
L1ME3C_3end	Eutheria	L1
L1ME3C_3end	Eutheria	L1
L1ME3D_3end	Eutheria	L1
L1ME3D_3end	Eutheria	L1
L1ME3D_3end	Eutheria	L1
L1ME3E_3end	Eutheria	L1
L1ME3E_3end	Eutheria	L1
L1ME3E_3end	Eutheria	L1
L1ME3F_3end	Eutheria	L1
L1ME3F_3end	Eutheria	L1
L1ME3F_3end	Eutheria	L1
L1ME4	Eutheria	L1
L1ME4A	Homo sapiens	L1
L1ME4A	Homo sapiens	L1
L1ME5	Eutheria	L1
L1ME5_3end	Mammalia	L1
L1ME5_3end	Mammalia	L1

L1ME5_3end	.		Mammalia	L1
L1MEA_5	.		Homo sapiens	L1
L1MEA_5	.		Homo sapiens	L1
L1MEB_5	.		Homo sapiens	L1
L1MEB_5	.		Homo sapiens	L1
L1MEC_5	.		Eutheria	L1
L1MEC_5	.		Eutheria	L1
L1MEC_5	.		Eutheria	L1
L1MED_5	.		Homo sapiens	L1
L1MED_5	.		Homo sapiens	L1
L1MEE_5	.		Homo sapiens	L1
L1MEe_5end	.		Eutheria	L1
L1MEe_5end	.		Eutheria	L1
L1MEe_5end	.		Eutheria	L1
L1MEf_5end	.		Mammalia	L1
L1MEf_5end	.		Mammalia	L1
L1MEf_5end	.		Mammalia	L1
L1MEg_5end	.		Mammalia	L1
L1MEg_5end	.		Mammalia	L1
L1MEg_5end	.		Mammalia	L1
L1MM_F	X57795		Mus musculus	L1
L1N1_MD	.		Monodelphis domestica	L1
L1NH_RN	X05455, Y00164		Rattus norvegicus	L1
L1P_MA2	.		Eutheria	L1
L1P_MA2	.		Eutheria	L1
L1P_MA2	.		Eutheria	L1
L1P4a_5end	.		Primates	L1
L1P4a_5end	.		Primates	L1
L1P4b_5end	.		Primates	L1
L1P4b_5end	.		Primates	L1
L1P4c_5end	.		Primates	L1
L1P4c_5end	.		Primates	L1
L1P4d_5end	.		Primates	L1
L1P4d_5end	.		Primates	L1
L1P4e_5end	.		Primates	L1
L1P4e_5end	.		Primates	L1
L1PA10	.		Homo sapiens	L1
L1PA11	.		Homo sapiens	L1
L1PA12	.		Homo sapiens	L1
L1PA12_5	.		Homo sapiens	L1
L1PA13	.		Homo sapiens	L1
L1PA13_5	.		Homo sapiens	L1
L1PA14	.		Homo sapiens	L1
L1PA14_5	.		Homo sapiens	L1
L1PA15	.		Homo sapiens	L1
L1PA16	.		Homo sapiens	L1
L1PA16_5	.		Homo sapiens	L1
L1PA17_5	.		Homo sapiens	L1
L1PA2	.		Homo sapiens	L1
L1PA3	.		Homo sapiens	L1
L1PA4	.		Homo sapiens	L1

L1PA5	.	Homo sapiens	L1	
L1PA6	.	Homo sapiens	L1	
L1PA7	.	Homo sapiens	L1	
L1PA7_5	.	Homo sapiens	L1	
L1PA8	.	Homo sapiens	L1	
L1PB1	.	Homo sapiens	L1	
L1PB1	.	Homo sapiens	L1	
L1PB2	.	Homo sapiens	L1	
L1PB2c	.	Homo sapiens	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1PB3	.	Homo sapiens	L1	
L1PB4	.	Homo sapiens	L1	
L1PBA_5	.	Homo sapiens	L1	
L1PBA1_5	.	Homo sapiens	L1	
L1PBB_5	.	Homo sapiens	L1	
L1PREC1	.	Homo sapiens	L1	
L1PREC2	.	Homo sapiens	L1	
20,2,21,22,23,24,11,25,1,26,27,28,4,29,30,31,16,3				
L1Pt	.	Pan troglodytes	L1	
L1Pt_5end	.	Pan troglodytes	L1	
L1R2_RN	M36995	Rattus norvegicus	L1	
L1S1_SS	X15410	Sus scrofa	L1	
L1S2_SS	X15411	Sus scrofa	L1	
L1SU_SP	007C10	Strongylocentrotus purpuratus	Tx1	1,26,43,3,44,11,4,2
L1VL1_5	X53728	Mus musculus	L1	
L1VL1_5end	.	Mus	L1	
L1VL2_5	.	Mus musculus	L1	
L1VL2_5end	.	Mus	L1	
L1VL4_5end	.	Muridae	L1	
L1VL4a_5end	.	Rattus	L1	
L2	.	Eutheria	CR1	
L2	.	Eutheria	CR1	
L2	.	Eutheria	CR1	
L2_AC_1	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_10	.	Anolis carolinensis	L2	16,1,3,24,4,11,2
L2_AC_11	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_12	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_13	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_14	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_15	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_16	.	Anolis carolinensis	L2	16,1,3,11,4,2,15
L2_AC_17	.	Anolis carolinensis	L2	16,1,3,24,4,11,2
L2_AC_2	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_3	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_4	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_5	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_6	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_7	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_AC_8	.	Anolis carolinensis	L2	16,1,3,11,4,2,15
L2_AC_9	.	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2_Av	.	Adineta vaga	Penelope	1,3

L2_Ele3	.	Aedes aegypti	L2
L2_Platla	.	Ornithorhynchus	CR1
L2_Platlb	.	Ornithorhynchus	CR1
L2_Platlc	.	Ornithorhynchus	CR1
L2_Platld	.	Ornithorhynchus	CR1
L2_Platle	.	Ornithorhynchus	CR1
L2_Platlf	.	Ornithorhynchus	CR1
L2_Platlg	.	Ornithorhynchus	CR1
L2_Platlh	.	Ornithorhynchus	CR1
L2_Platli	.	Ornithorhynchus	CR1
L2_Platlm	.	Ornithorhynchus	L2
L2_Platln	.	Ornithorhynchus	CR1
L2_Platlo	.	Ornithorhynchus	CR1
L2_Platlq	.	Ornithorhynchus	CR1
L2_Platlr	.	Ornithorhynchus	CR1
L2_Platls	.	Ornithorhynchus	CR1
L2_Platlt	.	Ornithorhynchus	CR1
L2_Platlu	.	Ornithorhynchus	CR1
L2-1_AC	.	Aplysia californica	L2
L2-1_AFC	.	Cichlidae	L2
L2-1_AMi	.	Alligator mississippiensis	L2
L2-1_BM	.	Bombyx mori	Kiri
L2-1_CCa	.	Cyprinus carpio	L2
L2-1_CPB	.	Chrysemys picta bellii	L2
L2-1_Croc	.	Crocodylidae	L2
L2-1_Crp	.	Crocodylus porosus	L2
L2-1_CTe	.	Capitella teleta	L2
L2-1_DL	.	Dicentrarchus labrax	L2
L2-1_DPu	.	Daphnia pulex	L2
L2-1_DR	.	Danio rerio	L2
L2-1_EL	.	Esox lucius	L2
L2-1_FR	.	Takifugu rubripes	L2
L2-1_FR1	.	Takifugu rubripes	L2
L2-1_GA	.	Gasterosteus aculeatus	L2
L2-1_HMa	.	Hydra vulgaris	L2
L2-1_Hmel	.	Heliconius melpomene	L2
L2-1_HRo	.	Helobdella robusta	L2
L2-1_HSal	.	Harpegnathos saltator	L2
L2-1_LCh	.	Latimeria chalumnae	L2
L2-1_NVi	.	Nasonia vitripennis	CR1
L2-1_OL	.	Oryzias latipes	L2
L2-1_PM	.	Petromyzon marinus	L2
L2-1_PSi	.	Pelodiscus sinensis	L2
L2-1_RPr	.	Rhodnius prolixus	Kiri
L2-1_SIn	.	Solenopsis invicta	L2
L2-1_SK	.	Saccoglossus kowalevskii	L2
L2-1_SP	.	Strongylocentrotus purpuratus	L2
L2-1_SSa	.	Salmo salar	L2
L2-1_XL	.	Xenopus laevis	L2
L2-10_CTe	.	Capitella teleta	L2

L2_Ele3	16,1,3,11,4,2,15
L2_Platla	
L2_Platlb	
L2_Platlc	
L2_Platld	
L2_Platle	
L2_Platlf	
L2_Platlg	
L2_Platlh	
L2_Platli	
L2_Platlm	
L2_Platln	
L2_Platlo	
L2_Platlq	
L2_Platlr	
L2_Platls	
L2_Platlt	
L2_Platlu	
L2-1_AC	1,3,4,2
L2-1_AFC	
L2-1_AMi	
L2-1_BM	1,3,11,4,2
L2-1_CCa	16,1,3,11,4,48,2,15
L2-1_CPB	
L2-1_Croc	1,3,11,4,2
L2-1_Crp	
L2-1_CTe	1,3,4,2
L2-1_DL	2,1
L2-1_DPu	16,1,3,4,11,2,15
L2-1_DR	63,1,64,3,11,4,48,2
L2-1_EL	1,3,11,4,2
L2-1_FR	1,3,4,2
L2-1_FR1	16,1,3,11,4,51,2,15
L2-1_GA	16,1,3,4,11,48,2,15
L2-1_HMa	1,4,2
L2-1_Hmel	1,3,11,4,2
L2-1_HRo	16,1,3,11,4,2,15
L2-1_HSal	
L2-1_LCh	
L2-1_NVi	16,1,3,11,4,2,15
L2-1_OL	16,1,3,11,4,2,15
L2-1_PM	1,3,11,4,2
L2-1_PSi	1,3,11,4,2
L2-1_RPr	
L2-1_SIn	1,8,14,3,11,4,12,2
L2-1_SK	1,3,4,11,2
L2-1_SP	1,3,4,2
L2-1_SSa	1,3,11,4,2
L2-1_XL	1,3,11,4,2
L2-10_CTe	1,3,11,4,2

66,67,53,144,2,54,55,11,56,57,68,1,58,4,3,59,60,61

L2-10_EL	.	Esox lucius	L2	
L2-10_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-10_HRo	.	Helobdella robusta	L2	53,2,54,24,11,56,57,68,1,58,4,16,3,59,60
L2-10_LCh	.	Latimeria chalumnae	L2	16,1,3,24,11,4,2
L2-11_CTe	.	Capitella teleta	L2	1,3,4,2
L2-11_DRe	.	Danio rerio	L2	1,3,2
L2-11_EL	.	Esox lucius	L2	1,3,11,4,2
L2-11_Hmel	.	Heliconius melpomene	L2	16,1,3,11,4,2,15
L2-11_HRo	.	Helobdella robusta	L2	53,2,54,11,56,57,68,1,58,4,3,59
L2-12_CTe	.	Capitella teleta	L2	1,3,4,2
L2-12_EL	.	Esox lucius	L2	
L2-12_Hmel	.	Heliconius melpomene	L2	16,1,3,11,4,2,15
L2-12_HRo	.	Helobdella robusta	L2	53,2,54,11,57,68,1,4,15,16,3,59,60
L2-12a_Lch	.	Latimeria chalumnae	L2	
L2-12b_Lch	.	Latimeria chalumnae	L2	
L2-13_CTe	.	Capitella teleta	L2	1,3,4,2
L2-13_DRe	.	Danio rerio	L2	1,3,4,2
L2-13_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-13_HRo	.	Helobdella robusta	L2	1,3,11,4,2
L2-13a_Lch	.	Latimeria chalumnae	L2	
L2-13B_DRe	.	Danio rerio	L2	2,1
L2-13b_Lch	.	Latimeria chalumnae	L2	1,3,11,4,2
L2-13c_Lch	.	Latimeria chalumnae	L2	
L2-14_CTe	.	Capitella teleta	L2	1,3,4,2
L2-14_DRe	.	Danio rerio	L2	
L2-14_EL	.	Esox lucius	L2	16,1,3,11,4,2,15
L2-14_Hmel	.	Heliconius melpomene	L2	1,3,11,4,2
L2-14_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,15,16,3,59
L2-15_CTe	.	Capitella teleta	L2	1,3,4,2
L2-15_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-15_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,15,16,3,60
L2-16_CTe	.	Capitella teleta	L2	1,3,4,2
L2-16_DRe	.	Danio rerio	L2	63,16,1,64,3,11,4,48,2,15
L2-16_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,15,16,3,59
L2-16_NVe	.	Nematostella vectensis	L2	
L2-17_CTe	.	Capitella teleta	L2	1,3,4,2
L2-17_DRe	.	Danio rerio	L2	16,1,3,4,11,2,15
L2-17_HRo	.	Helobdella robusta	L2	53,2,54,11,56,57,68,1,58,4,15,16,3,59,60
L2-17_NV	.	Nematostella vectensis	L2	
L2-18_ACar	.	Anolis carolinensis	L2	16,1,3,24,4,11,2
L2-18_CTe	.	Capitella teleta	L2	1,3,4,2
L2-18_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,15,16,3
L2-19_ACar	.	Anolis carolinensis	L2	1,3,11,4,2
L2-19_CTe	.	Capitella teleta	L2	1,3,11,4,2
L2-19_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,3,59,60
L2-1a_Cis	.	Ciona savignyi	CR1	21,16,1,3,24,11,4,2
L2-1b_Cis	.	Ciona savignyi	CR1	53,2,21,54,24,11,56,57,1,58,4,16,3,59,60
L2-1B_DR	.	Danio rerio	L2	
L2-1c_Cis	.	Ciona savignyi	CR1	2,21,24,11,56,57,1,58,4,16,3,60
L2-1d_3UTR_Cis	.	Ciona savignyi	CR1	
L2-2_CCa	.	Cyprinus carpio	L2	16,1,3,11,4,2,15

L2-2_Cis	Ciona savignyi	CR1	2,21,55,11,56,57,1,58,4,3,59,60,61
L2-2_CM	Callorhinchus milii	L2	16,1,3,24,11,4,2
L2-2_CPB	Chrysemys picta bellii	L2	
L2-2_Croc	Crocodylidae	L2	1,3,4,2
L2-2_CTe	Capitella teleta	L2	1,3,2
L2-2_DL	Dicentrarchus labrax	L2	1,3,4,2
L2-2_DRe	Danio rerio	L2	16,1,3,24,11,4,2
L2-2_EL	Esox lucius	L2	1,3,4,11,2
L2-2_GA	Gasterosteus aculeatus	L2	1,3,11,4,2
L2-2_Hmel	Heliconius melpomene	L2	1,3,4,2
L2-2_HRo	Helobdella robusta	L2	53,2,54,11,56,57,68,1,4,15,16,127,3,59
L2-2_LMe	Latimeria menadoensis	L2	1,3,11,4,2
L2-2_MD	Monodelphis domestica	CR1	
L2-2_ME	Notamacropus eugenii	CR1	
L2-2_NVi	Nasonia vitripennis	CR1	1,3,4,2
L2-2_OL	Oryzias latipes	L2	16,1,3,4,11,2,15
L2-2_PM	Petromyzon marinus	L2	16,1,3,11,4,2,15
L2-2_SIn	Solenopsis invicta	L2	16,1,8,14,3,11,4,12,2,15
L2-2_SP	Strongylocentrotus purpuratus	L2	16,1,3,11,4,2,15
L2-2_SSa	Salmo salar	L2	16,1,3,11,4,2,15
L2-2_XL	Xenopus laevis	L2	
L2-2_XT	Xenopus tropicalis	L2	16,1,3,24,4,11,2
L2-20_ACar	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2-20_CTe	Capitella teleta	L2	16,1,3,11,4,2,15
L2-20_HRo	Helobdella robusta	L2	2,11,57,68,1,4,3,59,60
L2-21_ACar	Anolis carolinensis	L2	
L2-21_CTe	Capitella teleta	L2	1,3,4,2
L2-21_HRo	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-22_ACar	Anolis carolinensis	L2	1,3,4,2
L2-22_CTe	Capitella teleta	L2	1,3,4,2
L2-22_HRo	Helobdella robusta	L2	2,11,56,57,68,1,159,58,4,15,16,160,3,59
L2-22_NV	Nematostella vectensis	CR1	1,3,4,11,2
L2-23_ACar	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2-23_CTe	Capitella teleta	L2	1,3,4,2
L2-23_HRo	Helobdella robusta	L2	16,1,23,3,4,11,2,15
L2-24_ACar	Anolis carolinensis	L2	1,3,4,2
L2-24_CTe	Capitella teleta	L2	1,3,2
L2-24_HRo	Helobdella robusta	L2	54,1,3,59,53,11,4,57,2
L2-24_NV	Nematostella vectensis	L2A	1,3,11,4,2
L2-25_ACar	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2-25_CTe	Capitella teleta	L2	1,3,4,2
L2-25_HRo	Helobdella robusta	L2	2,24,11,56,57,1,58,4,16,3,59
L2-26_ACar	Anolis carolinensis	L2	
L2-26_CTe	Capitella teleta	L2	1,3,11,4,2
L2-26_HRo	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-26B_HRo	Helobdella robusta	L2	16,11,15
L2-27_ACar	Anolis carolinensis	L2	1,3,4,2
L2-27_HRo	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-29_ACar	Anolis carolinensis	L2	16,1,3,24,11,4,2
L2-29_HRo	Helobdella robusta	L2	1,3,4,11,2
L2-29B_ACar	Anolis carolinensis	L2	1,3,11,4,2

L2-2B_SSa	.	Salmo salar	L2	16,1,3,11,4,2,15
L2-2C_SSa	.	Salmo salar	L2	16,1,3,11,4,2,15
L2-3_AFC	.	Cichlidae	L2	
L2-3_AMi	.	Alligator mississippiensis	L2	
L2-3_Cis	.	Ciona savignyi	CR1	53,2,21,54,11,56,57,68,1,58,4,3,59
L2-3_CM	.	Callorhinchus milii	L2	16,1,3,24,11,4,2
L2-3_CPB	.	Chrysemys picta bellii	L2	1,3,4,2
L2-3_Crp	.	Crocodylus porosus	L2	
L2-3_CTe	.	Capitella teleta	L2	1,3,4,2
L2-3_DL	.	Dicentrarchus labrax	L2	
L2-3_DRe	.	Danio rerio	L2	1,3,11,4,2
L2-3_EL	.	Esox lucius	L2	1,3,11,4,2
L2-3_GA	.	Gasterosteus aculeatus	L2	16,1,3,11,4,2,15
L2-3_Hmel	.	Heliconius melpomene	L2	16,1,3,11,4,2,15
L2-3_HRo	.	Helobdella robusta	L2	60,57
L2-3_LCh	.	Latimeria chalumnae	L2	1,3,4,2
L2-3_LMe	.	Latimeria menadoensis	L2	1,3,11,4,2
L2-3_NVi	.	Nasonia vitripennis	CR1	16,1,3,11,2,15
L2-3_OL	.	Oryzias latipes	L2	1,3,4,2
L2-3_PM	.	Petromyzon marinus	L2	
L2-3_SSa	.	Salmo salar	L2	16,1,3,11,4,48,2,15
L2-3_XL	.	Xenopus laevis	L2	1,3,11,4,2
L2-3_XT	.	Xenopus tropicalis	CR1	1,3,11,4,2
L2-30_ACar	.	Anolis carolinensis	L2	
L2-30_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-31_ACar	.	Anolis carolinensis	L2	
L2-31_HRo	.	Helobdella robusta	L2	2,11,57,1,159,4,160,3,60
L2-32_DRe	.	Danio rerio	L2	1,3,4,2
L2-32_HRo	.	Helobdella robusta	L2	16,1,3,24,11,4,2
L2-33_DRe	.	Danio rerio	L2	2,1
L2-33_HRo	.	Helobdella robusta	L2	1,3,11,4,2
L2-34_DR	.	Danio rerio	L2	16,1,3,11,4,2,15
L2-34_HRo	.	Helobdella robusta	L2	16,1,3,4,11,2,15
L2-35_DRe	.	Danio rerio	L2	1,3,4,2
L2-35_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-36_DRe	.	Danio rerio	L2	1,3,4,2
L2-36_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-36B_DRe	.	Danio rerio	L2	
L2-37_HRo	.	Helobdella robusta	L2	16,1,3,24,11,4,2
L2-38_HRo	.	Helobdella robusta	L2	16,1,3,24,11,4,2
L2-39_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-3b_Cis	.	Ciona savignyi	CR1	
L2-3B_CM-3end	.	Callorhinchus milii	L2	
L2-3C_CM-3end	.	Callorhinchus milii	L2	
L2-3D_CM	.	Callorhinchus milii	L2	
L2-3E_CM	.	Callorhinchus milii	L2	
L2-4_AFC	.	Cichlidae	L2	1,3,2
L2-4_Cis	.	Ciona savignyi	Crack	53,2,21,54,11,56,57,68,1,159,58,4,15,16,160,3,59,60
L2-4_CM	.	Callorhinchus milii	L2	16,1,3,24,11,4,2
L2-4_CPB	.	Chrysemys picta bellii	L2	

L2-4_CTe	.	Capitella teleta	L2	1,3,4,2
L2-4_DL	.	Dicentrarchus labrax	L2	
L2-4_DRe	.	Danio rerio	L2	16,1,3,24,11,4,2
L2-4_EL	.	Esox lucius	L2	2,111,11,1,4,112,15,16,3
L2-4_GA	.	Gasterosteus aculeatus	L2	1,3,11,4,2
L2-4_Hmel	.	Heliconius melpomene	L2	1,3,4,11,2
L2-4_HRo	.	Helobdella robusta	L2	1,3,11,4,2
L2-4_LCh	.	Latimeria chalumnae	L2	
L2-4_NVi	.	Nasonia vitripennis	CR1	54,1,3,53,4,11,2
L2-4_OL	.	Oryzias latipes	L2	1,11,2
L2-4_PM	.	Petromyzon marinus	L2	
L2-4_SSa	.	Salmo salar	L2	1,3,11,4,2
L2-4_XL	.	Xenopus laevis	L2	
L2-4_XT	.	Xenopus tropicalis	CR1	1,3,11,4,2
L2-40_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-41_DRe	.	Danio rerio	L2	16,1,3,24,11,4,2
L2-41_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-42_DRe	.	Danio rerio	L2	1,3,11,4,2
L2-43_DR	.	Danio rerio	L2	1,3,4,2
L2-43_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-44_DR	.	Danio rerio	L2	1,3,2
L2-45_DR	.	Danio rerio	L2	
L2-46_DR	.	Danio rerio	L2	1,3,2
L2-47_DR	.	Danio rerio	L2	
L2-48_DR	.	Danio rerio	L2	1,3,2
L2-49_DR	.	Danio rerio	L2	1,3,4,2
L2-4B_CM	.	Callorhinchus milii	L2	
L2-4B_EL	.	Esox lucius	L2	
L2-4C_CM-3end	.	Callorhinchus milii	L2	
L2-4D_CM	.	Callorhinchus milii	L2	
L2-5_AFC	.	Cichlidae	L2	1,3,11,4,2
L2-5_CPB	.	Chrysemys picta bellii	L2	
L2-5_CTe	.	Capitella teleta	L2	1,3,4,2
L2-5_DRe	.	Danio rerio	L2	1,3,4,2
L2-5_EL	.	Esox lucius	L2	1,3,11,4,48,2
L2-5_GA	.	Gasterosteus aculeatus	L2	16,1,3,24,11,4,2
L2-5_Hmel	.	Heliconius melpomene	L2	16,1,3,11,4,2,15
L2-5_HRo	.	Helobdella robusta	L2	53,2,54,11,57,1,4,15,16,3
L2-5_LCh	.	Latimeria chalumnae	L2	16,1,3,24,11,4,2
L2-5_NVi	.	Nasonia vitripennis	CR1	16,1,3,4,11,2,15
L2-5_OL	.	Oryzias latipes	L2	63,16,1,64,3,11,4,2,15
L2-5_PM	.	Petromyzon marinus	L2	1,3,11,4,2
L2-5_SSa	.	Salmo salar	L2	16,1,3,11,4,2,15
L2-5_XT	.	Xenopus tropicalis	CR1	16,1,3,11,4,2,15
L2-50_DR	.	Danio rerio	L2	
L2-51_DR	.	Danio rerio	L2	
L2-52_DR	.	Danio rerio	L2	
L2-53_DR	.	Danio rerio	L2	
L2-54_DR	.	Danio rerio	L2	
L2-55_DR	.	Danio rerio	L2	
L2-56_DR	.	Danio rerio	L2	1,3,2

L2-57_DR	.	Danio rerio	L2	
L2-58_DR	.	Danio rerio	L2	
L2-59_DR	.	Danio rerio	L2	
L2-5B_EL	.	Esox lucius	L2	
L2-5B_HRo	.	Helobdella robusta	L2	16,1,3,11,4,57,2,15
L2-5N1_SSa	.	Salmo salar	L2	
L2-5N2_SSa	.	Salmo salar	L2	
L2-6_AFC	.	Cichlidae	L2	
L2-6_CPB	.	Chrysemys picta bellii	L2	
L2-6_CTe	.	Capitella teleta	L2	1,3,4,2
L2-6_DRe	.	Danio rerio	L2	1,3,4,2
L2-6_EL	.	Esox lucius	L2	16,1,3,11,4,2,15
L2-6_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-6_HRo	.	Helobdella robusta	L2	1,3,4,2
L2-6_LCh	.	Latimeria chalumnae	L2	1,3,11,4,2
L2-6_NVi	.	Nasonia vitripennis	CR1	1,3,11,4,2
L2-6_OL	.	Oryzias latipes	L2	1,3,11,4,2
L2-6_PM	.	Petromyzon marinus	L2	
L2-6_SSa	.	Salmo salar	L2	1,2
L2-6_XT	.	Xenopus tropicalis	CR1	1,3,4,2
L2-68_HMa	.	Hydra vulgaris	L2	16,1,3,11,4,2,15
L2-7_CPB	.	Chrysemys picta bellii	L2	
L2-7_CTe	.	Capitella teleta	L2	1,3,2
L2-7_DRe	.	Danio rerio	L2	
L2-7_EL	.	Esox lucius	L2	1,3,11,2
L2-7_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-7_HRo	.	Helobdella robusta	L2	53,2,54,24,11,56,57,68,1,58,4,16,3,59
L2-7_LCh	.	Latimeria chalumnae	L2	16,1,3,24,11,4,2
L2-7_NVi	.	Nasonia vitripennis	CR1	1,3,11,4,2
L2-7_OL	.	Oryzias latipes	L2	1,3,4,2
L2-7_PM	.	Petromyzon marinus	L2	16,1,3,11,4,2,15
L2-7_SSa	.	Salmo salar	L2	
L2-8_AAe	.	Aedes aegypti	L2	53,2,54,11,56,57,68,1,58,4,3,59
L2-8_CTe	.	Capitella teleta	L2	1,3,11,4,2
L2-8_DR	.	Danio rerio	L2	1,3,11,4,2
L2-8_EL	.	Esox lucius	L2	1,3,11,4,2
L2-8_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-8_HRo	.	Helobdella robusta	L2	16,1,3,11,4,2,15
L2-8_LCh	.	Latimeria chalumnae	L2	1,3,11,4,2
L2-8_NVi	.	Nasonia vitripennis	CR1	1,3,60,4,11,57,2
L2-8_OL	.	Oryzias latipes	L2	2,1
L2-9_AAe	.	Aedes aegypti	L2	53,2,54,11,56,57,68,1,58,4,15,16,127,3,59
L2-9_CTe	.	Capitella teleta	L2	16,1,3,11,4,2,15
L2-9_DRe	.	Danio rerio	L2	16,1,3,11,4,2,15
L2-9_EL	.	Esox lucius	L2	1,3,2
L2-9_Hmel	.	Heliconius melpomene	L2	1,3,4,2
L2-9_HRo	.	Helobdella robusta	L2	53,2,54,11,57,68,1,4,15,16,3,59
L2-9_LCh	.	Latimeria chalumnae	L2	1,3,4,11,2
L2-9_NVi	.	Nasonia vitripennis	CR1	1,3,11,4,2
L2-9_OL	.	Oryzias latipes	L2	1,3,11,4,2
L2-N1_HRo	.	Helobdella robusta	L2	

L2A-1_NVe	Nematostella vectensis	L2A	66,67,2,11,56,57,68,1,58,15,16,59,60
L2A-1_PPo	Papilio polytes	L2	
L2A-1_SK	Saccoglossus kowalevskii	L2A	1,3,11,4,2
L2A-2_SK	Saccoglossus kowalevskii	L2A	1,3,4,2
L2B	Eutheria	CR1	
L2B	Eutheria	CR1	
L2B	Eutheria	CR1	
L2B_ME	Notamacropus eugenii	CR1	
L2B-1_AAe	Aedes aegypti	L2B	16,1,3,4,11,2,15
L2B-1_AG	Anopheles gambiae	L2B	54,1,8,3,59,53,11,4,12,2
L2B-1_BTa	Bemisia tabaci	L2B	1,3,11,4,2
L2B-1_CP	Culex pipiens	L2B	2,8,11,57,12,1,4,15,16,3,59
L2B-1_CQ	Culex quinquefasciatus	L2B	1,3,59,11,4,2
L2B-1_HM	Hydra vulgaris	L2B	1,3,59,4,11,57,2,68
L2B-1_LMi	Locusta migratoria	L2B	1,3,11,2
L2B-1_MLe	Mnemiopsis leidyi	L2B	
L2B-1_PBa	Pogonomyrmex barbatus	L2B	1,3,11,4,2
L2B-1B_CQ	Culex quinquefasciatus	L2B	1,3,4,2
L2B-2_AAe	Aedes aegypti	L2B	54,1,3,59,53,11,4,57,2
L2B-2_BTa	Bemisia tabaci	L2B	1,54,3,59,53,4,11,57,2
L2B-2_CQ	Culex quinquefasciatus	L2B	54,1,3,59,53,11,4,57,2
L2B-2_PBa	Pogonomyrmex barbatus	L2B	16,1,3,11,4,2,15
L2B-3_AAe	Aedes aegypti	L2B	53,2,54,11,57,1,4,15,16,3,59
L2B-3_CQ	Culex quinquefasciatus	L2B	16,1,8,14,3,11,4,12,2,15
L2B-3_PBa	Pogonomyrmex barbatus	L2B	1,3,11,4,2
L2B-3B_AAe	Aedes aegypti	L2B	1,3,11,4,2
L2B-3C_AAe	Aedes aegypti	L2B	1,3,11,4,2
L2B-4_AAe	Aedes aegypti	L2B	53,2,54,8,24,11,57,12,68,1,14,4,16,3,59
L2B-4_CQ	Culex quinquefasciatus	L2B	1,3,59,11,4,2
L2B-4B_AAe	Aedes aegypti	L2B	42,1,8,3,59,4,11,12,2
L2B-5_AAe	Aedes aegypti	L2B	53,2,54,11,57,1,4,15,16,3,59
L2B-5_CQ	Culex quinquefasciatus	L2B	1,3,11,4,2
L2B-6_AAe	Aedes aegypti	L2B	1,3,11,4,2
L2B-6_CQ	Culex quinquefasciatus	L2B	1,3,11,4,2
L2B-7_AAe	Aedes aegypti	L2B	1,3,4,11,2
L2C	Eutheria	L2	
L2C	Eutheria	L2	
L2C	Eutheria	L2	
L2D	Eutheria	L2	
L2D	Eutheria	L2	
L2D	Eutheria	L2	
L3	Homo sapiens	CR1	16,1,3,4,11,2,15
L3_Av	Adineta vaga	Penelope	165,1,166,167,3,168,130,169
L3_ME	Notamacropus eugenii	CR1	
L3b_3end	Mammalia	CR1	
L3b_3end	Mammalia	CR1	
L3b_3end	Mammalia	CR1	
L4	Mammalia	RTEX	1,3,4,2
L4	Mammalia	RTEX	1,3,4,2
L4	Mammalia	RTEX	1,3,4,2
L5	Mammalia	RTEX	

L5	.	Mammalia	RTEX		
L5	.	Mammalia	RTEX		
Laml	.	Lethenteron reissneri	SINE2/tRNA		
LDT1	AF081103	Lymantria dispar	Jockey	16,1,87,3,88,11,4,2,15	
LFR1_LP	.	Lepidosiren paradoxa	Non-LTRRetrotransposo		
LFSINE_AMi	.	Alligator mississippiensis	SINE2/tRNA		
LFSINE_Vert	.	Latimeria menadoensis	SINE2/tRNA		
Lgi-Nin-DC-SINE1	.	Lottia gigantea	SINE2/tRNA		
Lgi-Nin-DC-SINE3	.	Lottia gigantea	SINE2/tRNA		
Lian-Aa1	.	Aedes aegypti	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7	
LIN10_SM	.	Schmidtea mediterranea	L1	1,3,120,4,2	
LIN10B_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2	
LIN11_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2	
LIN13_SM	.	Schmidtea mediterranea	NeSL	1,215,3,4,2	
LIN14_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,149,120,4,2	
LIN15_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2	
LIN2_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,149,120,4,2	
LIN21_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,149,120,4,2	
LIN23_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,120,149,4,2	
LIN24_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2	
LIN24B_SM	.	Schmidtea mediterranea	NeSL		
LIN25_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,120,149,4,2	
LIN26_SM	.	Schmidtea mediterranea	NeSL	1,3,120,82,4,2	
LIN3_SM	.	Schmidtea mediterranea	NeSL	147,1,148,149,120,3,4,2	
LIN4_SM	.	Schmidtea mediterranea	NeSL	2,1,3,120,4	
LIN4b_SM	.	Schmidtea mediterranea	NeSL	1,148,3,149,120,4,2	
LIN5_SM	.	Schmidtea mediterranea	NeSL	1,3,4,2	
LIN6_SM	.	Schmidtea mediterranea	NeSL	1,181,3,4,2	
LIN7_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2	
LIN7B_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2	
LIN9_SM	.	Schmidtea mediterranea	NeSL	147,18,5,1,148,3,149,120,4,2	
LINE_CA	AJ411792, AJ411792.1	Cicer arietinum	L1		
LINE_CH	X72221	Capra hircus	L1		
LINE_LS	Z17425	Lilium speciosum	L1		
LINE-05_OS	.	Oryza sativa	L1		
LINE-1	.	Oryza sativa	L1		
LINE-1_AA	M95171	Aedes aegypti	Jockey		
LINE-4_OS	.	Oryza sativa	L1	43,2,8,44,11,12,42,1,14,4,3	
LINE-6_OS	.	Oryza sativa	L1		
LINE-7_OS	.	Oryza sativa	L1	43,2,44,11,12,1,14,4,3	
LINE-8_CR	.	Chlamydomonas reinhardtii	L1		
LINE1_BM	D26009	Bombyx mori	Jockey		
LINE1_CA	AJ411789	Cicer arietinum	L1		
LINE1-1_SBi	.	Sorghum bicolor	L1		
LINE1-1_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,248,2	
LINE1-10_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2	
LINE1-10_SBi	.	Sorghum bicolor	L1		
LINE1-10_ZM	.	Zea mays	L1	1,43,3,44,11,4,2	
LINE1-11_OS	.	Oryza sativa	L1	16,1,43,3,44,24,11,4,2	
LINE1-11_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2	
LINE1-11_ZM	.	Zea mays	L1	1,3,11,4,2	

LINE1-11B_OS	.	Oryza sativa	L1	
LINE1-11B_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-12_OS	.	Oryza sativa	L1	
LINE1-12_SBi	.	Sorghum bicolor	L1	16,1,43,75,3,76,44,24,11,4,2
LINE1-12_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-13_OS	.	Oryza sativa	L1	
LINE1-13_SBi	.	Sorghum bicolor	L1	
LINE1-13_ZM	.	Zea mays	L1	43,2,44,24,11,1,4,16,75,3,76
LINE1-14_OS	.	Oryza sativa	L1	
LINE1-14_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-14_ZM	.	Zea mays	L1	1,3,11,4,2
LINE1-15_OS	.	Oryza sativa	L1	
LINE1-15_SBi	.	Sorghum bicolor	L1	8,14,12
LINE1-15_ZM	.	Zea mays	L1	250,8,14,251,12
LINE1-16_OS	.	Oryza sativa	L1	
LINE1-16_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,12,1,14,4,16,75,3,76
LINE1-16_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-17_OS	.	Oryza sativa	L1	
LINE1-17_SBi	.	Sorghum bicolor	L1	1,14,3,11,4,2
LINE1-17_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-18_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-18_SBi	.	Sorghum bicolor	L1	1,14,3,4,12,2
LINE1-18_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-19_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-19_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-19_ZM	.	Zea mays	L1	1,4,2
LINE1-19B_OS	.	Oryza sativa	L1	
LINE1-2_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,1,4,16,75,3,76
LINE1-2_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-20_OS	.	Oryza sativa	L1	1,43,14,3,44,11,4,2
LINE1-20_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-20_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-21_OS	.	Oryza sativa	L1	
LINE1-21_SBi	.	Sorghum bicolor	L1	
LINE1-21_ZM	.	Zea mays	L1	1,3,11,4,2
LINE1-22_OS	.	Oryza sativa	L1	
LINE1-22_SBi	.	Sorghum bicolor	L1	
LINE1-22_ZM	.	Zea mays	L1	1,3,11,4,2
LINE1-23_OS	.	Oryza sativa	L1	
LINE1-23_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-23_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-24_OS	.	Oryza sativa	L1	
LINE1-24_SBi	.	Sorghum bicolor	L1	
LINE1-24_ZM	.	Zea mays	L1	94,16,95,24,11
LINE1-25_OS	.	Oryza sativa	L1	44,43
LINE1-25_SBi	.	Sorghum bicolor	L1	1,75,43,3,76,44,11,4,2
LINE1-25_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-26_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-26_SBi	.	Sorghum bicolor	L1	
43,78,79,2,10,44,11,80,1,81,77,4,17,75,3,76,7				
LINE1-26_ZM	.	Zea mays	L1	1,3,4,2

LINE1-27_OS	.	Oryza sativa	L1	
LINE1-27_SBi	.	Sorghum bicolor	L1	94,1,43,3,44,95,11,4,2
LINE1-27_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-28_OS	.	Oryza sativa	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-28_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-28_ZM	.	Zea mays	L1	94,16,1,3,95,24,11,4,2
LINE1-29_OS	.	Oryza sativa	L1	
LINE1-29_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-29_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-3_SBi	.	Sorghum bicolor	L1	1,43,3,44,4,2
LINE1-3_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-30_OS	.	Oryza sativa	L1	
LINE1-30_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-30_ZM	.	Zea mays	L1	1,43,3,44,4,2
LINE1-31_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-31_SBi	.	Sorghum bicolor	L1	1,3,11,4,2
LINE1-31_ZM	.	Zea mays	L1	1,3,11,4,2
LINE1-32_OS	.	Oryza sativa	L1	
LINE1-32_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-32_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-33_OS	.	Oryza sativa	L1	
LINE1-33_SBi	.	Sorghum bicolor	L1	1,75,43,3,76,44,11,4,2
LINE1-33_ZM	.	Zea mays	L1	94,16,1,3,95,24,11,4,2
LINE1-34_OS	.	Oryza sativa	L1	
LINE1-34_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,1,4,16,75,3,76
LINE1-34_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-35_OS	.	Oryza sativa	L1	
LINE1-35_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-35_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-36_OS	.	Oryza sativa	L1	
LINE1-36_SBi	.	Sorghum bicolor	L1	43,2,8,44,11,12,42,1,14,4,3
LINE1-36_ZM	.	Zea mays	L1	1,3,4,2
LINE1-37_OS	.	Oryza sativa	L1	16,1,43,3,44,24,11,4,2
LINE1-37_SBi	.	Sorghum bicolor	L1	1,43,3,44,4,2
LINE1-37_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-38_OS	.	Oryza sativa	L1	
LINE1-38_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,1,4,16,75,3,76
LINE1-38_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-39_OS	.	Oryza sativa	L1	
43,78,79,2,44,10,11,80,1,81,77,4,17,75,3,76,7				
LINE1-39_SBi	ABXC01004903.1	Sorghum bicolor	L1	1,43,3,44,11,2
LINE1-39_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-4_SBi	.	Sorghum bicolor	L1	
LINE1-4_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-40_OS	.	Oryza sativa	L1	
LINE1-40_SBi	ABXC01001255.1	Sorghum bicolor	L1	43,2,44,24,11,12,1,4,16,3
LINE1-40_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,42,1,14,4,16,75,3,76
LINE1-41_OS	.	Oryza sativa	L1	
LINE1-41_SBi	ABXC01007042.1	Sorghum bicolor	L1	1,43,3,44,4,2
LINE1-41_ZM	.	Zea mays	L1	1,43,3,44,4,2
LINE1-42_OS	.	Oryza sativa	L1	

LINE1-42_SBi	.	Sorghum bicolor	L1	43,2,8,44,11,12,42,1,14,4,3
LINE1-42_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,94,16,3,95
LINE1-43_OS	.	Oryza sativa	L1	
LINE1-43_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-43_ZM	.	Zea mays	L1	1,43,3,44,4,2
LINE1-44_OS	.	Oryza sativa	L1	16,1,43,3,44,24,4,11,2
LINE1-44_SBi	.	Sorghum bicolor	L1	43,2,44,24,11,12,1,14,4,16,3
LINE1-44_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-45_OS	.	Oryza sativa	L1	43,2,44,24,11,1,14,4,16,75,3,76
LINE1-45_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-45_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-46_OS	.	Oryza sativa	L1	
LINE1-46_SBi	.	Sorghum bicolor	L1	1,43,75,3,44,76,11,4,2
LINE1-46_ZM	.	Zea mays	L1	43,2,44,24,11,1,4,94,16,3,95
LINE1-47_OS	.	Oryza sativa	L1	
LINE1-47_SBi	.	Sorghum bicolor	L1	44,43
LINE1-47_ZM	.	Zea mays	L1	94,16,1,43,3,44,95,24,11,4,2
LINE1-48_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-48_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-48_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,14,4,16,75,3,76
LINE1-49_OS	.	Oryza sativa	L1	
LINE1-49_SBi	.	Sorghum bicolor	L1	1,43,3,44,2
LINE1-49_ZM	.	Zea mays	L1	43,2,44,24,11,12,1,4,16,3
LINE1-5_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-5_ZM	.	Zea mays	L1	43,2,44,24,11,1,14,4,16,75,3,76
LINE1-50_OS	.	Oryza sativa	L1	16,1,43,3,44,24,11,4,2
LINE1-50_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-50_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-51_OS	.	Oryza sativa	L1	
LINE1-51_SBi	.	Sorghum bicolor	L1	1,43,14,3,44,11,4,2
LINE1-51_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,1,14,4,94,16,3,95
LINE1-52_OS	.	Oryza sativa	L1	
LINE1-52_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,4,11,2
LINE1-52_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-53_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-53_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-53_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-54_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-54_SBi	.	Sorghum bicolor	L1	7,1,43,17,3,44,10,79,4,2,80
LINE1-54_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-55_OS	.	Oryza sativa	L1	
LINE1-55_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,4,11,2
LINE1-55_ZM	.	Zea mays	L1	16,1,43,3,44,24,11,4,2
LINE1-56_OS	.	Oryza sativa	L1	
LINE1-56_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-56_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-57_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-57_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-57_ZM	.	Zea mays	L1	1,43,75,3,76,44,11,4,2
LINE1-58_OS	.	Oryza sativa	L1	

43,78,79,2,7,8,10,44,24,11,12,80,1,81,77,4,16,75,17,3,76

LINE1-58_SBi	.	Sorghum bicolor	L1	1,75,43,3,76,44,11,4,2
LINE1-59_OS	.	Oryza sativa	L1	43,2,44,11,12,1,14,4,3
LINE1-59_SBi	.	Sorghum bicolor	L1	1,43,75,76,44,11,4,2
LINE1-59_ZM	.	Zea mays	L1	1,43,3,44,11,4,2
LINE1-6_OS	AC113336.2	Oryza sativa	L1	43,78,79,2,8,44,10,11,12,80,1,81,77,4,17,75,3,76,7
LINE1-6_SBi	.	Sorghum bicolor	L1	
LINE1-6_ZM	.	Zea mays	L1	43,2,8,44,11,12,1,14,4,75,3,76
LINE1-60_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-60_SBi	.	Sorghum bicolor	L1	43,78,79,2,7,10,44,11,80,1,81,77,4,249,75,17,3,76
LINE1-61_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-61_SBi	.	Sorghum bicolor	L1	
LINE1-62_OS	.	Oryza sativa	L1	43,2,8,44,11,12,1,14,4,3
LINE1-62_SBi	.	Sorghum bicolor	L1	
LINE1-62B_SBi	.	Sorghum bicolor	L1	43,78,2,44,24,11,1,81,77,4,16,75,3,76
LINE1-63_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-63_SBi	.	Sorghum bicolor	L1	1,43,3,44
LINE1-64_OS	.	Oryza sativa	L1	43,79,2,10,44,11,80,1,4,17,3,7
LINE1-64_SBi	.	Sorghum bicolor	L1	1,43,3,44,4,2
LINE1-65_OS	.	Oryza sativa	L1	16,1,43,3,44,24,11,4,2
LINE1-65_SBi	.	Sorghum bicolor	L1	
LINE1-66_OS	.	Oryza sativa	L1	43,2,44,11,12,1,14,4,3
LINE1-66_SBi	.	Sorghum bicolor	L1	43,78,79,2,8,44,10,11,12,80,1,14,81,77,4,17,75,3,76,7
LINE1-67_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-67_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-68_OS	.	Oryza sativa	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-68_SBi	.	Sorghum bicolor	L1	43,78,2,10,44,11,80,1,81,77,4,17,75,3,76,7
LINE1-69_OS	.	Oryza sativa	L1	43,78,79,2,7,8,10,44,11,12,80,1,81,77,4,75,17,3,76
LINE1-69_SBi	.	Sorghum bicolor	L1	43,2,44,11,12,1,14,4,3
LINE1-7_OS	AC146581.1	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-7_SBi	.	Sorghum bicolor	L1	43,82,2,8,44,24,11,12,1,14,4,16,3
LINE1-7_ZM	.	Zea mays	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-70_OS	.	Oryza sativa	L1	43,2,44,24,11,1,4,16,75,3,76
LINE1-70_SBi	.	Sorghum bicolor	L1	43,78,79,2,44,10,11,80,1,81,77,4,17,75,3,76,7
LINE1-71_OS	.	Oryza sativa	L1	1,43,3,44,4,11,2
LINE1-71_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-72_OS	.	Oryza sativa	L1	
LINE1-72_SBi	.	Sorghum bicolor	L1	7,1,17,43,3,44,10,11,4,2,80
LINE1-73_OS	.	Oryza sativa	L1	16,1,43,3,44,24,11,4,2
LINE1-73_SBi	.	Sorghum bicolor	L1	1,43,3,44,4,2
LINE1-74_OS	.	Oryza sativa	L1	43,2,8,44,11,12,1,14,4,75,3,76
LINE1-74_SBi	.	Sorghum bicolor	L1	1,43,3,44,11,4,2
LINE1-75_OS	.	Oryza sativa	L1	43,78,79,2,10,44,11,80,1,81,77,4,17,75,76,7
LINE1-76_OS	.	Oryza sativa	L1	43,78,79,2,10,44,11,80,1,81,77,4,17,75,3,76,7
LINE1-77_OS	.	Oryza sativa	L1	43,78,79,6,2,7,8,10,44,11,12,80,1,81,77,4,75,17,3,76

LINE1-78_OS	.	Oryza sativa	L1	
43,78,79,2,10,44,11,80,1,81,77,4,17,75,3,76,7				
LINE1-79_OS	.	Oryza sativa	L1	16,1,43,3,44,24,11,4,2
LINE1-8_OS	AP004587.3	Oryza sativa	L1	43,2,44,24,11,12,1,14,4,16,3
LINE1-8_SBi	.	Sorghum bicolor	L1	43,2,8,44,24,11,12,1,14,4,16,3
LINE1-8_ZM	.	Zea mays	L1	1,43,3,44,11
LINE1-80_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-81_OS	.	Oryza sativa	L1	
LINE1-82_OS	.	Oryza sativa	L1	1,43,3,44,4,2
LINE1-83_OS	.	Oryza sativa	L1	
LINE1-84_OS	.	Oryza sativa	L1	
LINE1-9_OS	.	Oryza sativa	L1	
LINE1-9_SBi	.	Sorghum bicolor	L1	16,1,43,3,44,24,11,4,2
LINE1-9_ZM	.	Zea mays	L1	1,43,3,44,4,2
LINE1-9B_OS	.	Oryza sativa	L1	1,43,3,44,11,4,2
LINE1-N1_OS	.	Oryza sativa	L1	
LINE1-N1B_OS	.	Oryza sativa	L1	
LINE1-N1C_OS	.	Oryza sativa	L1	
LINE1-N1D_OS	.	Oryza sativa	L1	
LINE1-N1E_OS	.	Oryza sativa	L1	
LINE1-N1F_OS	.	Oryza sativa	L1	
LINE1-N1FB_OS	.	Oryza sativa	L1	
LINE1-N1G_OS	.	Oryza sativa	L1	
LINE1-N1H_OS	.	Oryza sativa	L1	
LINE1-N1I_OS	.	Oryza sativa	L1	
LINE1-N1J_OS	.	Oryza sativa	L1	
LINE1-N1K_OS	.	Oryza sativa	L1	
LINE1-N1L_OS	.	Oryza sativa	L1	
LINE1-N1N_OS	.	Oryza sativa	L1	
LINE1-N1N2_OS	.	Oryza sativa	L1	
LINE1-N1P_OS	.	Oryza sativa	L1	
LINE1-N1Q_OS	.	Oryza sativa	L1	
LINE1-N1QB_OS	.	Oryza sativa	L1	
LINE1-N1R_OS	.	Oryza sativa	L1	
LINE1-N1S_OS	.	Oryza sativa	L1	
LINE1-N1T_OS	.	Oryza sativa	L1	
LINE1-N1U_OS	.	Oryza sativa	L1	
LINE1-N1UB_OS	.	Oryza sativa	L1	
LINE1-N1V_OS	.	Oryza sativa	L1	
LINE1-N1W_OS	.	Oryza sativa	L1	
LINE1-N1WB_OS	.	Oryza sativa	L1	
LINE1-N1X_OS	.	Oryza sativa	L1	
LINE1-N1X2_OS	.	Oryza sativa	L1	
LINE1-N1Y_OS	.	Oryza sativa	L1	
LINE1-N1Z_OS	.	Oryza sativa	L1	
LINE1-N2_OS	.	Oryza sativa	L1	
LINE1-N3_OS	.	Oryza sativa	L1	
LINE1A_MT	AC148776	Medicago truncatula	L1	1,43,3,44,4,2
LINE1A_MT	AC148776	Medicago truncatula	L1	1,43,3,44,4,2
LINE1A1_MT	AC122166	Medicago truncatula	L1	16,1,24,11,4,2
LINE1A1_MT	AC122166	Medicago truncatula	L1	16,1,24,11,4,2

LINE1A2_MT	.	Medicago truncatula	L1	1,43,3,44,4,2
LINE1A2_MT	.	Medicago truncatula	L1	1,43,3,44,4,2
LINE1B_MT	AC135415	Medicago truncatula	L1	7,1,43,17,3,44,10,79,4,6,2,80
LINE1B_MT	AC135415	Medicago truncatula	L1	7,1,43,17,3,44,10,79,4,6,2,80
LINE1C_MT	AC137081	Medicago truncatula	L1	43,79,6,2,7,8,10,44,11,12,80,1,77,4,75,17,3,76
LINE1C_MT	AC137081	Medicago truncatula	L1	43,79,6,2,7,8,10,44,11,12,80,1,77,4,75,17,3,76
LINE1D_MT	AC153004	Medicago truncatula	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
LINE1D_MT	AC153004	Medicago truncatula	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
LINE1E_MT	AC148239	Medicago truncatula	L1	43,78,79,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76,7
LINE1E_MT	AC148239	Medicago truncatula	L1	43,78,79,2,8,10,44,24,11,12,80,1,81,77,4,16,17,75,3,76,7
LINE1F_MT	AC140545	Medicago truncatula	L1	43,79,2,8,44,10,11,12,80,1,77,4,75,17,3,76,7
LINE1F_MT	AC140545	Medicago truncatula	L1	43,79,2,8,44,10,11,12,80,1,77,4,75,17,3,76,7
LINE1G_MT	.	Medicago truncatula	L1	43,75,76,44
LINE1G_MT	.	Medicago truncatula	L1	43,75,76,44
LINE1H_MT	.	Medicago truncatula	L1	43,79,2,44,10,11,80,1,4,17,3,7
LINE1H_MT	.	Medicago truncatula	L1	43,79,2,44,10,11,80,1,4,17,3,7
LINE1I_MT	AC165446	Medicago truncatula	L1	43,1,3,44,11,2
LINE1I_MT	AC165446	Medicago truncatula	L1	43,1,3,44,11,2
LINE1J_MT	AC144515	Medicago truncatula	L1	16,1,3,24,11,4,2
LINE1J_MT	AC144515	Medicago truncatula	L1	16,1,3,24,11,4,2
LINE2_CH1	.	Crotalus horridus	CR1	
LINE2_EC1	.	Echis coloratus	CR1	
LINE2_NT1	.	Natrix tessellata	CR1	
LINE2_PM1	.	Python molurus	CR1	
LINE2_VA1	.	Vipera ammodytes	CR1	
LINE2_VP1	.	Daboia palaestinae	CR1	
LINE2_WA1	.	Walterinnesia aegyptia	CR1	
LINE2A_CB	.	Caenorhabditis briggsae	CR1	
LINE2A_CE	.	Caenorhabditis elegans	CR1	
LINE2B_CB	.	Caenorhabditis briggsae	CR1	1,3,4,2
LINE2B_CE	.	Caenorhabditis elegans	CR1	
LINE2C_CB	.	Caenorhabditis briggsae	CR1	
LINE2C_CE	.	Caenorhabditis elegans	CR1	
LINE2C1_CE	U13643	Caenorhabditis elegans	CR1	16,1,3,11,4,2,15
LINE2D_CE	.	Caenorhabditis elegans	CR1	
LINE2E_CE	AF077538	Caenorhabditis elegans	CR1	
LINE2F_CE	.	Caenorhabditis elegans	CR1	
LINE2G_CE	.	Caenorhabditis elegans	CR1	1,3,4,2
LINE2H_CE	.	Caenorhabditis elegans	CR1	1,3,4,2
LINER1	.	Nasonia vitripennis	R1	2,8,11,86,12,42,1,14,15,16,3
LINER1-2_NVi	.	Nasonia vitripennis	R1	2,8,11,86,12,42,1,14,4,15,16,3
LINER1-3_NVi	.	Nasonia vitripennis	R1	16,1,3,11,4,86,2,15
LINEX-1_DR	.	Danio rerio	Rex1	
LINEX-2_DR	.	Danio rerio	Rex1	
LinSINE1	.	Lingula anatina	SINE2/tRNA	

LitSINE1	.	Littorina saxatilis	SINE2/tRNA
Lm1	.	Locusta migratoria	SINE2/tRNA
Lm1B	.	Locusta migratoria	SINE2/tRNA
Lm2	.	Locusta migratoria	SINE2/tRNA
Lm2B	.	Locusta migratoria	SINE2/tRNA
Lm2B2	.	Locusta migratoria	SINE2/tRNA
Lm2C	.	Locusta migratoria	SINE2/tRNA
Lm2C1	.	Locusta migratoria	SINE2/tRNA
Lm2C2	.	Locusta migratoria	SINE2/tRNA
Lm2D	.	Locusta migratoria	SINE2/tRNA
Lm2E	.	Locusta migratoria	SINE2/tRNA
LmeSINE1a	.	Latimeria menadoensis	SINE3/5S
LmeSINE1b	.	Metazoa Eukaryota.	SINE3/5S
LmeSINE1c	.	Metazoa Eukaryota.	SINE3/5S
LmeSINE1d	.	Latimeria chalumnae	SINE3/5S
LmeSINE1e	.	Latimeria chalumnae	SINE3/5S
LMR1_LM	.	Leptosphaeria maculans	Non-LTRRetrotransposo
LmSIDER2	.	Leishmania	Ingi
LOA	X60177	Drosophila silvestris	Loa
199,200,201,2,8,202,11,12,203,1,204,14,89,4,90,15,16,17,3			
LOA_Ele2	.	Aedes aegypti	Loa
LOA_Ele2B_AAe	.	Aedes aegypti	Loa
LOA_Ele3	.	Aedes aegypti	Loa
LOA_Ele3B_AAe	.	Aedes aegypti	Loa
LOA_Ele3C_AAe	.	Aedes aegypti	Loa
LOA_Ele4	.	Aedes aegypti	Loa
LOA_Ele5	.	Aedes aegypti	Loa
LOA_Ele6	.	Aedes aegypti	Loa
LOA_Ele6B_AAe	.	Aedes aegypti	Loa
LOA_Ele6C_AAe	.	Aedes aegypti	Loa
LOA_Ele7	.	Aedes aegypti	Loa
Loa-1_BTa	.	Bemisia tabaci	Loa
LOA-1_CQ	.	Culex quinquefasciatus	Loa
Loa-1_DAn	.	Drosophila ananassae	Loa
LOA-1_DBi	.	Drosophila biarmipes	Loa
Loa-1_DBp	.	Drosophila bipectinata	Loa
Loa-1_DEl	.	Drosophila elegans	Loa
LOA-1_DGri	.	Drosophila grimshawi	Loa
Loa-1_DK	.	Drosophila kikkawai	Loa
LOA-1_DPer	.	Drosophila persimilis	Loa
Loa-1_DPs	.	Drosophila pseudoobscura	Loa
Loa-1_DRh	.	Drosophila rhopaloa	Loa
LOA-1_DSer	.	Drosophila serrata	Loa
Loa-1_DT	.	Drosophila takahashii	Loa
LOA-1_DYa	.	Drosophila yakuba	Loa
LOA-1_FoCa	.	Folsomia candida	Loa
Loa-1_LHu	.	Linepithema humile	Loa
Loa-1_LMi	.	Locusta migratoria	Loa
Loa-1_LSal	.	Lepeophtheirus salmonis	Loa
LOA-1_MLe	.	Mnemiopsis leidyi	Loa
Loa-1_MMa	.	Mesobuthus martensii	Loa
			6,2,9,10,11,13,1,4,15,16,17,3,7
			6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
			6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
			6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
			6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
			6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
			6,2,9,10,11,86,13,1,4,15,16,17,3,7
			6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
			6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
			6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
			7,1,17,9,3,10,4,6,2,13
			6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
			6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
			6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
			6,2,7,9,10,11,13,1,89,4,90,15,16,17,3
			6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
			6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
			6,2,7,9,10,11,86,13,1,4,15,16,17,3
			6,2,9,10,11,13,1,4,15,16,17,3,7
			6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
			16,1,89,11,86,2,90,15
			6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
			6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
			6,2,7,9,10,11,86,13,1,4,15,16,17,3
			6,2,9,10,11,13,1,4,15,16,17,3,7
			6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
			6,2,7,9,10,11,13,1,89,4,90,15,16,17,3
			7,1,17,3,10,4,6,2,13
			6,2,9,10,11,13,1,89,4,90,15,16,17,3,7

LOA-1_PPo	.	Papilio polytes	Loa	16,1,3,11,4,2,15
Loa-1_RPr	.	Rhodnius prolixus	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
Loa-1_SIn	.	Solenopsis invicta	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
LOA-10_AAe	.	Aedes aegypti	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA-1B_DPer	.	Drosophila persimilis	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-1B_DT	.	Drosophila takahashii	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA-1B_FoCa	.	Folsomia candida	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-1C_DT	.	Drosophila takahashii	Loa	
Loa-2_BTa	.	Bemisia tabaci	Loa	1,3,4,2
LOA-2_CQ	.	Culex quinquefasciatus	Loa	6,2,9,10,11,86,13,1,4,15,16,17,3,7
Loa-2_DEl	.	Drosophila elegans	Loa	
Loa-2_DF	.	Drosophila ficusphila	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA-2_DGri	.	Drosophila grimshawi	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-2_DK	.	Drosophila kikkawai	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA-2_DPer	.	Drosophila persimilis	Loa	6,2,7,9,10,11,13,1,89,4,90,15,16,17,3
LOA-2_DSer	.	Drosophila serrata	Loa	6,2,9,10,11,86,13,1,15,16,17,3,7
Loa-2_DTa	.	Drosophila takahashii	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
LOA-2_DYa	.	Drosophila yakuba	Loa	6,2,9,10,11,13,1,15,16,17,7
LOA-2_FoCa	.	Folsomia candida	Loa	
Loa-2_LSal	.	Lepeophtheirus salmonis	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
LOA-2_MLe	.	Mnemiopsis leidyi	Loa	16,1,3,11,4,2,15
Loa-2_MMa	.	Mesobuthus martensii	Loa	16,1,3,11,4,2,15
Loa-2_RPr	.	Rhodnius prolixus	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-2_SIn	.	Solenopsis invicta	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
LOA-3_CQ	.	Culex quinquefasciatus	Loa	6,2,9,10,11,86,13,1,4,15,16,17,3,7
Loa-3_DEl	.	Drosophila elegans	Loa	
LOA-3_DPer	.	Drosophila persimilis	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
Loa-3_DTa	.	Drosophila takahashii	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-3_LSal	.	Lepeophtheirus salmonis	Loa	
Loa-3_RPr	.	Rhodnius prolixus	Loa	16,1,3,89,11,15,4,86,2,90
LOA-4_CQ	.	Culex quinquefasciatus	Loa	6,2,9,10,11,13,1,4,15,16,17,3,7
Loa-4_DTa	.	Drosophila takahashii	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
Loa-4_LSal	.	Lepeophtheirus salmonis	Loa	7,1,17,9,3,10,11,4,2,6,13
LOA-5_CQ	.	Culex quinquefasciatus	Loa	6,2,9,10,11,86,13,1,89,4,90,15,16,17,3,7
Loa-5_LSal	.	Lepeophtheirus salmonis	Loa	7,16,1,17,3,11,4,2,15
LOA-6_CQ	.	Culex quinquefasciatus	Loa	7,1,17,9,3,10,4,6,2,13
Loa-6_LSal	.	Lepeophtheirus salmonis	Loa	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
LOA-7_CQ	.	Culex quinquefasciatus	Loa	7,1,17,9,3,10,4,6,2,13
LOA-8_AAe	.	Aedes aegypti	Loa	6,2,7,9,10,11,86,13,1,89,4,90,15,16,17,3
LOA-9_AAe	.	Aedes aegypti	Loa	6,2,9,10,11,13,1,89,4,90,15,16,17,3,7
Loner_Ele1	.	Aedes aegypti	I	6,2,7,9,10,11,12,13,1,4,15,16,17,3
Loner_Ele2	.	Aedes aegypti	I	6,2,7,9,10,11,13,1,4,15,16,17,3
Loner_Ele3	.	Aedes aegypti	I	6,2,7,9,10,11,12,13,1,4,15,16,17,3
Loner_Ele4	.	Aedes aegypti	I	6,2,9,10,11,13,1,4,15,16,17,3,7
Lun1	.	Lepidosiren paradoxa	SINE2/tRNA	
LX	.	Muridae	L1	
Lx_3end	.	Muridae	L1	
LX10	.	Mus musculus	L1	
LX2	.	Sciuromorpha	L1	
Lx2_3end	.	Muridae	L1	
Lx2A_3end	.	Muridae	L1	

Lx2A1_3end	.	Muridae	L1		
LX2B	.	Mus musculus	L1		
Lx2B_3end	.	Muridae	L1		
Lx2B2_3end	.	Muridae	L1		
LX3	.	Sciuromorpha	L1		
Lx3B_3end	.	Muridae	L1		
Lx3C_3end	.	Muridae	L1		
LX4	.	Mus musculus	L1		
LX5	.	Mus musculus	L1		
LX5b	.	Mus musculus	L1		
LX5c	.	Mus musculus	L1		
LX6	.	Mus musculus	L1		
LX7	.	Mus musculus	L1		
Lx7_3end	.	Rodentia	L1		
LX8	.	Mus musculus	L1		
LX8B	.	Mus musculus	L1		
LX9	.	Mus musculus	L1		
M_Av	.	Adineta vaga	Penelope	1,3,2	
M1_As	.	Adineta	Penelope	1,3	
M2_As	.	Adineta	Penelope	1,3	
M3_As	.	Adineta	Penelope	1,3	
MamSINE1	.	Mammalia	SINE2/tRNA		
MamSINE1	.	Mammalia	SINE2/tRNA		
MamSINE1	.	Mammalia	SINE2/tRNA		
ManaSINE1	.	Manacus vitellinus	SINE2/tRNA		
ManaSINE1_ZCh	.	Zimmerius chrysops	SINE2/tRNA		
ManaSINE2	.	Manacus vitellinus	SINE		
MAR1	.	Metatheria	SINE2/tRNA		
MAR1_MD	.	Monodelphis domestica	SINE2/tRNA		
MAR1c_Mdo	.	Monodelphis domestica	SINE2/tRNA		
MAR4_MD	.	Monodelphis domestica	RTE		
MARE3	.	Mammalia	SINE		
MARE3	.	Mammalia	SINE		
MARE3	.	Mammalia	SINE		
MARE6	.	Mammalia	Tx1		
MARE6	.	Mammalia	Tx1		
MARE6	.	Mammalia	Tx1		
MAUI	.	Takifugu rubripes	CR1	16,1,3,24,4,11,48,2	
MdoRep1	.	Metatheria	SINE		
Meg_PVa	.	Pteropus vampyrus	SINE3/5S		
MEG-RL	EU853278.1	Rousettus aegyptiacus	SINE3/5S		
MEG-RS	EU853265.1	Rousettus aegyptiacus	SINE3/5S		
MEG-T2	.	Megachiroptera	SINE2/tRNA		
MEG-TR	.	Megachiroptera	SINE2/tRNA		
MeloSINE	.	Melopsittacus undulatus	SINE		
MEN	.	Menetes berdmorei	SINE2/tRNA		
MER129	.	Euteleostomi	SINE		
MER131	.	Euteleostomi	SINE		
MER25	.	Eutheria	L1		
Metulj-0_Hmel	.	Heliconius melpomene	SINE		
Metulj-1_Hmel	.	Heliconius melpomene	SINE		

Metulj-2_Hmel	.	Heliconius melpomene	SINE		
Metulj-3_Hmel	.	Heliconius melpomene	SINE		
Metulj-4_Hmel	.	Heliconius melpomene	SINE		
Metulj-5_Hmel	.	Heliconius melpomene	SINE		
Metulj-6_Hmel	.	Heliconius melpomene	SINE		
Metulj-7_Hmel	.	Heliconius melpomene	SINE		
MGR583	AF018033	Magnaporthe grisea	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
MinoAg1	AB090816	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3	
MIR	.	Mammalia	SINE2/tRNA		
MIR	.	Mammalia	SINE2/tRNA		
MIR	.	Mammalia	SINE2/tRNA		
MIR_Aves1	.	Aves	SINE2/tRNA		
MIR_Aves2	.	Aves	SINE2/tRNA		
MIR_Xt	.	Xenopus tropicalis	SINE2/tRNA		
MIR1_AMi	.	Alligator mississippiensis	SINE2/tRNA		
MIR1_Crp	.	Crocodylus porosus	SINE2/tRNA		
MIR3	.	Mammalia	SINE2/tRNA		
MIR3	.	Mammalia	SINE2/tRNA		
MIR3	.	Mammalia	SINE2/tRNA		
MIR3_MarsA	.	Metatheria	SINE2/tRNA		
MIR3_MarsB	.	Metatheria	SINE2/tRNA		
MIRb	.	Mammalia	SINE2/tRNA		
MIRb	.	Mammalia	SINE2/tRNA		
MIRb	.	Mammalia	SINE2/tRNA		
MIRc	.	Mammalia	SINE2/tRNA		
MIRc	.	Mammalia	SINE2/tRNA		
MIRc	.	Mammalia	SINE2/tRNA		
MIUSE1_TM	AF459639	Triticum monococcum	L1		
MON1	.	Monotremata	SINE2/tRNA		
MonoRep87A	.	Ornithorhynchus	SINE2/tRNA		
MonoRep87B	.	Ornithorhynchus	SINE2/tRNA		
MosquI_Aa2	AF134900	Aedes aegypti	I		
MoTeR1	JQ747487	Magnaporthe oryzae	CRE	1,3,2	
MoTeR2	JQ747488	Magnaporthe oryzae	CRE		
MRE1_OL	.	Oryzias latipes	SINE		
MRT	.	Monascus pilosus	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,239,7	
MSTA1	.	Primates	ERV3		
MusHAL1	.	Mus	L1		
MusHAL1_5end	.	Mus	L1		
MUSID2	.	Rodentia	SINE2/tRNA		
MUSID3	.	Rodentia	SINE2/tRNA		
MUSID4	.	Rodentia	SINE2/tRNA		
MUSID5	.	Rodentia	SINE2/tRNA		
Mutsu-1_SK	.	Saccoglossus kowalevskii	Tx1	43,2,8,44,11,12,42,1,26,14,4,3	
Mutsu-2_SK	.	Saccoglossus kowalevskii	Tx1	2,8,11,12,42,1,26,14,4,3	
MVB2	X52381	Neovison vison	SINE2/tRNA		
MyrSINE	.	Myrmecophagidae	SINE2/tRNA		
N_As	.	Adineta	Penelope	1,3	
NO_As	.	Adineta	Penelope		
NAVIRTE1	.	Nasonia vitripennis	RTE	1,3,2	
Nematis_C4	.	Caenorhabditis brenneri	Penelope	1,3,103,2	

Nematis_Cr	.	Caenorhabditis remanei	Penelope	1,3,103,72,2
Nematis_Pp	.	Pristionchus pacificus	Penelope	1,3,103
Neptune_Hyd	.	Hydra vulgaris	Penelope	1,3,2
Neptune_Sp	.	Sphenodon punctatus	Penelope	
Neptune1_Ac	.	Anolis carolinensis	Penelope	73,1,3,103,72,2
Neptune1_Ap	.	Acropora palmata	Penelope	73,1,3,71,72,74
Neptune1_Nv	.	Nematostella vectensis	Penelope	1,3,70,71,72
Neptune1_Ren	.	Amphimedon queenslandica	Penelope	73,1,3,70,71,72,62,74
Neptune2_Ap	.	Acropora palmata	Penelope	
Neptune2_Ren	.	Amphimedon queenslandica	Penelope	205,73,74,1,3,70,206,71,72,2
Neptune3_Ren	.	Amphimedon queenslandica	Penelope	74,73,1,3,70,71,72,2
NeSL-1	Z82058	Caenorhabditis elegans	NeSL	147,1,148,3,149,120,4,2
NeSL-1_C11	.	Caenorhabditis tropicalis	NeSL	147,1,3,149,120,4,2
NeSL-1_CA	.	Caenorhabditis angaria	NeSL	147,1,148,3,149,120,4,2
NeSL-1_CBre	.	Caenorhabditis brenneri	NeSL	147,1,3,149,120,4,2
NeSL-1_CBri	.	Caenorhabditis briggsae	NeSL	1,2
NeSL-1_CJap	.	Caenorhabditis japonica	NeSL	147,1,148,3,149,120,4,2
NeSL-1_CRem	.	Caenorhabditis remanei	NeSL	147,1,3,149,120,4,2
NeSL-1_SM	.	Schmidtea mediterranea	NeSL	1,3,4,2
NeSL-1_TV	.	Trichomonas vaginalis	NeSL	18,5,1,19,3,4,2
NeSL-2_CBre	.	Caenorhabditis brenneri	NeSL	147,1,148,3,149,120,4,2
NeSL-2_CRem	.	Caenorhabditis remanei	NeSL	147,5,19,1,148,3,120,149,4,2
NeSL-2_SM	.	Schmidtea mediterranea	NeSL	1,3,4,2
NeSL-3_CBre	.	Caenorhabditis brenneri	NeSL	147,1,148,3,149,120,4,2
NeSL-3_CRem	.	Caenorhabditis remanei	NeSL	147,1,148,3,120,149,4,2
NeSL-4_CRem	chrUn	Caenorhabditis remanei	NeSL	147,1,148,3,149,120,4,2
NeSL-4_SM	.	Schmidtea mediterranea	NeSL	1,3,120,4,2
Ngaro1_DR	.	Danio rerio	Non-LTRRetrotransposo	220,221,222,2,223,224,1,225,226,17,3
Nimb-1_AAe	.	Aedes aegypti	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-1_BTa	.	Bemisia tabaci	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-1_CPB	.	Chrysemys picta bellii	Nimb	
Nimb-1_CQ	.	Culex quinquefasciatus	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-1_CTe	.	Capitella teleta	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-1_DPu	.	Daphnia pulex	Nimb	
6,2,9,8,10,11,12,13,42,1,14,4,15,16,17,3,60,7				
Nimb-1_DR	.	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-1_EL	.	Esox lucius	Nimb	7,1,17,9,3,10,4,11,6,2,13
Nimb-1_GA	.	Gasterosteus aculeatus	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-1_HRo	.	Helobdella robusta	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-1_LG	.	Lottia gigantea	Nimb	6,2,9,10,11,1,4,15,16,17,3,7
Nimb-1_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-1_LVa	.	Litopenaeus vannamei	Nimb	6,2,7,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-1_MMa	.	Mesobuthus martensii	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-1_PH	.	Parhyale hawaiensis	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-1_RPr	.	Rhodnius prolixus	Nimb	12,8
Nimb-1_SSa	.	Salmo salar	Nimb	6,2,7,9,10,11,13,1,4,15,16,3
Nimb-10_DR	.	Danio rerio	Nimb	
Nimb-10_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-11_DR	.	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-11_LMi	.	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
Nimb-12_DR	.	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7

Nimb-12_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,86,13,1,14,4,15,16,17,3,7
Nimb-13_DR	.	Danio rerio	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-13_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-14_DR	.	Danio rerio	Nimb	
Nimb-14_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,86,12,13,1,14,4,15,16,17,3
Nimb-15_DR	.	Danio rerio	Nimb	2,8,11,12,1,14,4,15,16,17,3,7
Nimb-15_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-16_DR	.	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7,227
Nimb-16_LMi	.	Locusta migratoria	Nimb	6,2,9,8,10,12,13,1,14,4,17,3,7
Nimb-17_DRe	.	Danio rerio	Nimb	6,2,9,10,11,13,1,14,4,15,16,17,3,7
Nimb-17_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-18_LMi	.	Locusta migratoria	Nimb	6,2,7,9,8,10,11,86,12,13,1,14,4,15,16,17
Nimb-19_LMi	.	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,13,1,14,4,15,16,17,3
Nimb-19B_LMi	.	Locusta migratoria	Nimb	
Nimb-1N1_LVa	.	Litopenaeus vannamei	Nimb	
Nimb-2_AAe	.	Aedes aegypti	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_BTa	.	Bemisia tabaci	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_CPB	.	Chrysemys picta bellii	Nimb	
Nimb-2_CQ	.	Culex quinquefasciatus	Nimb	6,2,8,9,10,11,12,1,4,15,16,17,3,7
Nimb-2_CTe	.	Capitella teleta	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_DR	.	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_LG	.	Lottia gigantea	Nimb	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7
Nimb-2_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_LVa	.	Litopenaeus vannamei	Nimb	6,2,9,10,11,13,214,1,4,15,16,17,3,7
Nimb-2_PH	.	Parhyale hawaiensis	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2_RPr	.	Rhodnius prolixus	Nimb	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Nimb-2_SSa	.	Salmo salar	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-20_LMi	.	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,13,1,14,4,15,16,17,3
Nimb-21_LMi	.	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Nimb-22_LMi	.	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,1,14,4,15,16,17,3
Nimb-23_LMi	.	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,13,1,14,4,15,16,17,3
Nimb-24_LMi	.	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-25_LMi	.	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-26_LMi	.	Locusta migratoria	Nimb	16,1,14,3,11,4,2,15
Nimb-27_LMi	.	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,14,4,15,16,3,7
Nimb-28_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-29_LMi	.	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Nimb-2B_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-2B_SSa	.	Salmo salar	Nimb	8,14,12
Nimb-2C_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-3_BTa	.	Bemisia tabaci	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-3_CTe	.	Capitella teleta	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-3_DR	.	Danio rerio	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-3_LG	.	Lottia gigantea	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-3_LVa	.	Litopenaeus vannamei	Nimb	6,2,10,11,12,1,14,4,17,3,7
Nimb-3_PH	.	Parhyale hawaiensis	Nimb	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Nimb-3_RPr	.	Rhodnius prolixus	Nimb	
Nimb-3_SSa	.	Salmo salar	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-30_LMi	.	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-31_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17
Nimb-32_LMi	.	Locusta migratoria	Nimb	16,1,17,9,10,11,4,2,6,7,13,15

Nimb-33_LMi	.	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-34_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-35_LMi	.	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Nimb-36_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-37_LMi	.	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-38_LMi	.	Locusta migratoria	Nimb	7,16,1,17,3,11,4,2,15
Nimb-39_LMi	.	Locusta migratoria	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-4_BTa	.	Bemisia tabaci	Nimb	17,9,10,6,7,13
Nimb-4_DR	.	Danio rerio	Nimb	6,2,21,9,23,10,11,1,26,4,30,17,3,7
Nimb-4_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-4_LVa	.	Litopenaeus vannamei	Nimb	6,2,9,10,11,13,1,14,4,15,16,17,3,7
Nimb-4_RPr	.	Rhodnius prolixus	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-4_SSa	.	Salmo salar	Nimb	6,2,8,9,10,11,12,13,1,4,15,16,17,3,7
Nimb-40_LMi	.	Locusta migratoria	Nimb	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Nimb-40N1_LMi	.	Locusta migratoria	Nimb	
Nimb-41_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-42_LMi	.	Locusta migratoria	Nimb	6,2,7,9,8,10,11,12,13,1,14,4,17,3
Nimb-4B_RPr	.	Rhodnius prolixus	Nimb	16,1,8,3,11,15,4,12,2
Nimb-5_DR	.	Danio rerio	Nimb	6,2,7,9,10,11,13,1,4,15,16,17,3
Nimb-5_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,13,1,14,4,15,16,17,3,7
Nimb-6_DR	.	Danio rerio	Nimb	6,2,9,10,11,13,1,14,4,15,16,17,3,7
Nimb-6_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,86,13,1,14,4,15,16,17,3,7
Nimb-7_DR	.	Danio rerio	Nimb	17,9,10,6,7,13
Nimb-7_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3
Nimb-8_DR	.	Danio rerio	Nimb	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Nimb-8_LMi	.	Locusta migratoria	Nimb	6,2,9,10,11,13,1,14,4,15,16,17,3,7
Nimb-9_DR	.	Danio rerio	Nimb	6,2,9,10,11,13,1,4,15,16,17,3,7
Nimb-9_LMi	.	Locusta migratoria	Nimb	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Nimb-N1_SSa	.	Salmo salar	Nimb	
Nimb-N2_LVa	.	Litopenaeus vannamei	Nimb	
Nimb-N2_SSa	.	Salmo salar	Nimb	
Nimb-N2B_LVa	.	Litopenaeus vannamei	Nimb	
nimbus	EF413180	Biomphalaria glabrata	Nimb	6,2,7,9,10,11,86,12,13,1,14,4,15,16,17,3
NLA	M57436	Capra hircus	SINE2/tRNA	
Non-LTR-1_ACar	.	Anolis carolinensis	RTEX	
NonLTR-1_BG	.	Blumeria graminis	Non-LTRRetrotransposo	
NonLTR-1_CM	.	Callorhinchus milii	Non-LTRRetrotransposo	
NonLTR-1_CPB	.	Chrysemys picta bellii	Non-LTRRetrotransposo	
NonLTR-1_CR	.	Chlamydomonas reinhardtii	Non-LTRRetrotransposo	
NonLTR-1_DR	.	Danio rerio	Non-LTRRetrotransposo	
NonLTR-1_LSal	.	Lepeophtheirus salmonis	Non-LTRRetrotransposo	
NonLTR-1_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo	6,2,9,10,11,13,1,4,15,16,17,3,7
NonLTR-1_SBi	.	Sorghum bicolor	Non-LTRRetrotransposo	
NonLTR-1_STu	.	Solanum tuberosum	Non-LTRRetrotransposo	
NonLTR-10_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo	
NonLTR-1B_DR	.	Danio rerio	Non-LTRRetrotransposo	
NonLTR-2_BG	.	Blumeria graminis	Non-LTRRetrotransposo	
NonLTR-2_CM	.	Callorhinchus milii	Non-LTRRetrotransposo	
NonLTR-2_CR	.	Chlamydomonas reinhardtii	Non-LTRRetrotransposo	
NonLTR-2_LSal	.	Lepeophtheirus salmonis	Non-LTRRetrotransposo	
NonLTR-2_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo	2,16,1,14,3,4,11,12,15

NonLTR-2_PPo	.	Papilio polytes	Non-LTRRetrotransposo		
NonLTR-2_STu	.	Solanum tuberosum	Non-LTRRetrotransposo		
NonLTR-21_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo		
NonLTR-22_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo		
NonLTR-24_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo		
NonLTR-2B_LSal	.	Lepeophtheirus salmonis	Non-LTRRetrotransposo		
NonLTR-3_BG	.	Blumeria graminis	Non-LTRRetrotransposo		
NonLTR-3_CM	.	Callorhinchus milii	Non-LTRRetrotransposo	1,3,11,4,2	
NonLTR-3_CR	.	Chlamydomonas reinhardtii	Non-LTRRetrotransposo		
NonLTR-3_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo	1,3,11,4,2	
NonLTR-3_PPo	.	Papilio polytes	Non-LTRRetrotransposo		
NonLTR-4_BG	.	Blumeria graminis	Tad1		
NonLTR-4_CR	.	Chlamydomonas reinhardtii	Non-LTRRetrotransposo		
NonLTR-4_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo	6,2,7,9,10,11,12,13,1,14,4,15,16,17,3	
NonLTR-4_PPo	.	Papilio polytes	Non-LTRRetrotransposo		
NonLTR-4B_BG	.	Blumeria graminis	Tad1		
NonLTR-5_BG	.	Blumeria graminis	Non-LTRRetrotransposo		
NonLTR-5_CR	.	Chlamydomonas reinhardtii	Non-LTRRetrotransposo		
NonLTR-5_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo	43,6,2,7,9,10,44,11,13,1,14,4,15,16,17,3	
NonLTR-6_BG	.	Blumeria graminis	Non-LTRRetrotransposo		
NonLTR-6_CR	.	Chlamydomonas reinhardtii	Non-LTRRetrotransposo		
NonLTR-6_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo	16,1,3,24,11,4,2	
NonLTR-6A_CR	.	Chlamydomonas reinhardtii	Non-LTRRetrotransposo		
NonLTR-7_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo	7,1,17,9,10,11,6,2,13	
NonLTR-9_LVa	.	Litopenaeus vannamei	Non-LTRRetrotransposo	1,3,11,4,2	
NonLTR-N1_PH	.	Parhyale hawaiensis	Non-LTRRetrotransposo		
NonLTR-N1B_PH	.	Parhyale hawaiensis	Non-LTRRetrotransposo		
NonLTR-N2_PH	.	Parhyale hawaiensis	Non-LTRRetrotransposo		
NonLTR-N3_PH	.	Parhyale hawaiensis	Non-LTRRetrotransposo		
NonLTR-N4_PH	.	Parhyale hawaiensis	Non-LTRRetrotransposo		
NotoAg1	AB090823	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3	
Nova-1_PH	.	Parhyale hawaiensis	RTE	16,1,3,11,4,48,62,2,15	
Nve-Nin-DC-SINE1	.	Nematostella vectensis	SINE2/tRNA		
Nve-Nin-DC-SINE2	.	Nematostella vectensis	SINE2/tRNA		
Nve-Nin-DC-SINE3	.	Nematostella vectensis	SINE2/tRNA		
Nve-Nin-DC-SINE4	.	Nematostella vectensis	SINE2/tRNA		
Nve-Nin-DC-SINE5	.	Nematostella vectensis	SINE2/tRNA		
Nve-Nin-DC-SINE6	.	Nematostella vectensis	SINE2/tRNA		
O_As	.	Adineta vaga	Penelope	1,3	
O_Av	.	Adineta vaga	Penelope	1,3	
Odin-1	.	Oikopleura dioica	Odin		
11					
Odin-10_OD	CABV01003716	Oikopleura dioica	Odin	1,3,4,11,2	
Odin-11_OD	CABV01003634	Oikopleura dioica	Odin	1,3,11,2	
Odin-12_OD	CABW01000719	Oikopleura dioica	Odin	1,3,11,4,2	
Odin-13_OD	CABV01005259	Oikopleura dioica	Odin	1,3,11,4,2	
Odin-14_OD	CABV01001030	Oikopleura dioica	Odin	1,3,11,2	
Odin-15_OD	CABV01000726	Oikopleura dioica	Odin	1,3,11,4,2	
Odin-16_OD	CABW01000316	Oikopleura dioica	Odin	1,3,11,2	
Odin-17_OD	CABV01002100	Oikopleura dioica	Odin	1,3,11,4,2	
Odin-18_OD	CABV01002427	Oikopleura dioica	Odin	1,3,11,2	

Odin-19_OD	CABV01004164	Oikopleura dioica	Odin	1,3,11,2
Odin-2	.	Oikopleura dioica	Odin	1,3,11,4,2
Odin-20_OD	CABV01003756	Oikopleura dioica	Odin	1,3,11,2
Odin-21_OD	CABW01000073	Oikopleura dioica	Odin	1,3,11,2
Odin-22_OD	CABV01000239	Oikopleura dioica	Odin	1,3,11,4,2
Odin-23_OD	CABV01000544	Oikopleura dioica	Odin	1,3,11,2
Odin-24_OD	CABV01000555	Oikopleura dioica	Odin	1,3,11,2
Odin-25_OD	CABV01000253	Oikopleura dioica	Odin	1,3,11,4,2
Odin-26_OD	CABV01000948	Oikopleura dioica	Odin	1,3,11,4,2
Odin-27_OD	CABV01002548	Oikopleura dioica	Odin	1,3,4,2
Odin-28_OD	CABV01002941	Oikopleura dioica	Odin	1,3,4,2
Odin-29_OD	CABV01000222	Oikopleura dioica	Odin	1,3,4,11,2
Odin-3	.	Oikopleura dioica	Odin	1,3,11,2
Odin-30_OD	CABV01001101	Oikopleura dioica	Odin	1,3,2
Odin-31_OD	CABV01004482	Oikopleura dioica	Odin	1,3,11,2
Odin-32_OD	CABV01004103	Oikopleura dioica	Odin	1,11,2
Odin-33_OD	CABV01004174	Oikopleura dioica	Odin	1,3,11,2
Odin-34_OD	CABW01000201	Oikopleura dioica	Odin	1,3,11,2
Odin-35_OD	CABV01005185	Oikopleura dioica	Odin	1,3,11,2
Odin-36_OD	CABV01000234	Oikopleura dioica	Odin	1,3,11,2
Odin-37_OD	CABV01003479	Oikopleura dioica	Odin	1,3,11,2
Odin-38_OD	CABW01000121	Oikopleura dioica	Odin	1,3,11,4,2
Odin-6_OD	CABV01000386	Oikopleura dioica	Odin	1,3,11,2
Odin-7_OD	CABW01000816	Oikopleura dioica	Odin	1,3,11,2
Odin-8_OD	CABV01004396	Oikopleura dioica	Odin	1,3,11,4,2
Odin-9_OD	CABV01000574	Oikopleura dioica	Odin	1,3,11,4,2
OegopSINE	.	Idioteuthis cordiformis	SINE2/tRNA	
OK	D32083	Octopus vulgaris	SINE2/tRNA	
OM2	X59354	Oncorhynchus masou	SINE2/tRNA	
OR1	D32089	Octopus vulgaris	SINE2/tRNA	
OR2	D32092	Octopus vulgaris	SINE2/tRNA	
ORTE-1_AAe	.	Aedes aegypti	RTE	120,2,175,176,11,1,4,3,177
ORTE-2_AAe	.	Aedes aegypti	RTE	1,175,176,3,120,4,11,177,2
ORTE-3_AAe	.	Aedes aegypti	RTE	120,2,175,176,11,1,4,3,177
ORTE-4_AAe	.	Aedes aegypti	RTE	175,1,176,3,120,11,4,177,2
ORTE-5_AAe	.	Aedes aegypti	RTE	120,2,175,176,11,1,26,4,15,16,3,177
ORTE-6_AAe	.	Aedes aegypti	RTE	120,2,175,176,24,11,1,26,4,16,3,177
OrySINE1	.	Oryzias latipes	SINE2/tRNA	
OSLINE1_1	AE017112	Oryza sativa	L1	43,82,79,2,10,44,11,80,1,4,17,3,7
OSLINE1_2	AP008218	Oryza sativa	L1	43,78,79,2,7,10,44,11,80,1,81,77,4,75,17,3,76
OSLINE1-2	AC137696	Oryza sativa	L1	1,43,3,44,11,4,2
OSLINE1-3	AC136524	Oryza sativa	L1	43,2,8,44,11,12,1,14,4,75,3,76
OSLINE1-4	AC104712	Oryza sativa	L1	16,1,3,24,11,4,2
OSLINE1-5	.	Oryza sativa	L1	1,43,3,44,11,4,2
OSSINE1	.	Salmo salar	SINE3/5S	
OsSN3	.	Oryza sativa	SINE2/tRNA	
Outcast_Ele14	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast_Ele3	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-1_BF	.	Branchiostoma floridae	Outcast	53,6,2,54,9,8,10,11,57,12,216,13,68,1,14,4,15,16,17,69,3,59,7

Outcast-1_DAn	.	Drosophila ananassae	Outcast	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Outcast-1_DBi	.	Drosophila biarmipes	Outcast	7,1,17,9,3,10,4,6,2,13
Outcast-1_DBp	.	Drosophila bipectinata	Outcast	16,1,3,11,2,15
Outcast-1_DEl	.	Drosophila elegans	Outcast	7,16,1,17,3,10,11,4,6,2,15
Outcast-1_DEu	.	Drosophila eugracilis	Outcast	7,1,17,9,3,10,4,6,2,13
Outcast-1_DT	.	Drosophila takahashii	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-1_DWi	.	Drosophila willistoni	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-16_AAe	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-17_AAe	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-18_AAe	.	Aedes aegypti	Outcast	6,2,9,10,11,13,1,4,15,16,17,3,7
Outcast-19_AAe	.	Aedes aegypti	Outcast	6,2,8,9,10,11,12,1,4,15,16,17,3,7
Outcast-2_BF	.	Branchiostoma floridae	Outcast	
53,6,2,7,54,9,8,10,11,56,57,12,13,68,1,14,58,4,15,16,17,69,3,59				
Outcast-2_DAn	.	Drosophila ananassae	Outcast	17,9,10,6,7,13
Outcast-2_DBp	.	Drosophila bipectinata	Outcast	
Outcast-2_DTa	.	Drosophila takahashii	Outcast	6,2,8,10,11,12,1,4,15,16,17,3,7
Outcast-20_AAe	.	Aedes aegypti	Outcast	6,2,9,10,11,80,1,4,15,16,17,3,7
P_As	.	Adineta	Penelope	1,3
P_Av	.	Adineta vaga	Penelope	1,3
p-SINE1_OS	.	Oryza sativa	SINE2/tRNA	
Pami	.	Patiria miniata	SINE2/tRNA	
PB1	.	Rodentia	SINE1/7SL	
PB1D10	.	Rodentia	SINE1/7SL	
PB1D11	.	Euarchontoglires	SINE1/7SL	
PB1D7	.	Rodentia	SINE1/7SL	
PB1D9	.	Rodentia	SINE1/7SL	
Peci	.	Petrolisthes cinctipes	SINE2/tRNA	
Ped-1	.	Pedetes capensis	SINE3/5S	
Ped-2	.	Pedetes capensis	SINE2/tRNA	
PENEL1_NVi	.	Nasonia vitripennis	Penelope	1,3,72
PENELA_Smed	.	Schmidtea mediterranea	Penelope	73,1,3,103,2
PENELOPE	.	virilis group	Penelope	74,1,3,71,103,72,2
Penelope_Ele2	.	Aedes aegypti	Penelope	73,74,1,3,71,103,72,2
PENELOPE_SM	BK000685	Schistosoma mansoni	Penelope	1,3,103,72,2
Penelope-1_AAe	.	Aedes aegypti	Penelope	1,3,103
Penelope-1_ACar	.	Anolis carolinensis	Penelope	74,73,1,3,71,103,72,2
Penelope-1_ADi	.	Acropora digitifera	Penelope	74,73,1,3,70,71,72,2
Penelope-1_AEc	.	Acromyrmex echinatior	Penelope	1,3,103,72,2
Penelope-1_AFC	.	Cichlidae	Penelope	1,103,72,2
Penelope-1_AMi	.	Alligator mississippiensis	Penelope	1,3,2
Penelope-1_BF	.	Branchiostoma floridae	Penelope	1,3,103,72
Penelope-1_BTa	.	Bemisia tabaci	Penelope	1,3
Penelope-1_CGi	.	Crassostrea gigas	Penelope	73,74,1,3,70,71,72,2
Penelope-1_CM	.	Callorhinchus milii	Penelope	73,71,72,74
Penelope-1_CoF1	.	Copidosoma floridanum	Penelope	1,3,72,2
Penelope-1_CPB	.	Chrysemys picta bellii	Penelope	1,3,72,2
Penelope-1_CQ	.	Culex quinquefasciatus	Penelope	
Penelope-1_Crp	.	Crocodylidae	Penelope	1,3,72,2
Penelope-1_CTe	.	Capitella teleta	Penelope	1,3,2
Penelope-1_DE1	.	Drosophila elegans	Penelope	74,1,3,70,71,103,72,2
Penelope-1_DF	.	Drosophila ficusphila	Penelope	1,3,71,103,72,74

Penelope-1_DK	Drosophila kikkawai	Penelope
Penelope-1_DR	Danio rerio	Penelope
Penelope-1_DRh	Drosophila rhopaloa	Penelope
Penelope-1_FoCa	Folsomia candida	Penelope
Penelope-1_GA	Gasterosteus aculeatus	Penelope
Penelope-1_Gav	Gavialis gangeticus	Penelope
Penelope-1_HM	Hydra vulgaris	Penelope
Penelope-1_HMM	Heliconius melpomene melpomene	Penelope
Penelope-1_HRo	Helobdella robusta	Penelope
Penelope-1_LCh	Latimeria chalumnae	Penelope
Penelope-1_LG	Lottia gigantea	Penelope
Penelope-1_LHu	Linepithema humile	Penelope
Penelope-1_LMi	Locusta migratoria	Penelope
Penelope-1_LSal	Lepeophtheirus salmonis	Penelope
Penelope-1_LVa	Litopenaeus vannamei	Penelope
Penelope-1_MLe	Mnemiopsis leidyi	Penelope
Penelope-1_NV	Nematostella vectensis	Penelope
Penelope-1_NVi	Nasonia vitripennis	Penelope
Penelope-1_OD	Oikopleura dioica	Penelope
Penelope-1_OL	Oryzias latipes	Penelope
Penelope-1_PBa	Pogonomyrmex barbatus	Penelope
Penelope-1_PH	Parhyale hawaiensis	Penelope
Penelope-1_PSi	Pelodiscus sinensis	Penelope
Penelope-1_SIn	Solenopsis invicta	Penelope
Penelope-1_SK	Saccoglossus kowalevskii	Penelope
Penelope-1_SM	Schmidtea mediterranea	Penelope
Penelope-1_SP	Strongylocentrotus purpuratus	Penelope
Penelope-1_SSa	Salmo salar	Penelope
Penelope-1_TCas	Tribolium castaneum	Penelope
Penelope-1_XL	Xenopus laevis	Penelope
Penelope-10_ACar	Anolis carolinensis	Penelope
Penelope-10_CoFl	Copidosoma floridanum	Penelope
Penelope-10_FoCa	Folsomia candida	Penelope
Penelope-10_HM	Hydra vulgaris	Penelope
Penelope-10_LMi	Locusta migratoria	Penelope
Penelope-10_LVa	Litopenaeus vannamei	Penelope
Penelope-10_NV	Nematostella vectensis	Penelope
Penelope-10_PH	Parhyale hawaiensis	Penelope
Penelope-10_XL	Xenopus laevis	Penelope
Penelope-10_XT	Xenopus tropicalis	Penelope
Penelope-100_LMi	Locusta migratoria	Penelope
Penelope-101_LMi	Locusta migratoria	Penelope
Penelope-102_LMi	Locusta migratoria	Penelope
Penelope-103_LMi	Locusta migratoria	Penelope
Penelope-104_LMi	Locusta migratoria	Penelope
Penelope-105_LMi	Locusta migratoria	Penelope
Penelope-106_LMi	Locusta migratoria	Penelope
Penelope-107_LMi	Locusta migratoria	Penelope
Penelope-108_LMi	Locusta migratoria	Penelope

1,3,72
74,1,3,71,103,72,2
73,71,72,74
73,1,3,70,71,72,62,74
73,1,3,70,71,72,74
74,73,1,3,103,71,72,2
1,3,103,72,2
73,72
1,3,103,2
74,73,1,3,70,71,72,2
1,3,71,2,74
18,74,5,1,71,103,72,2
73,1,3,71,103,72,74
1,3,2
74,1,3,71,72,2
74,73,1,3,71,103,72,2
73,74,1,3,70,71,2
73,74,1,3,70,71,103,72,2
74,1,3,71,72,2
71,103,72,74
1,3,70,71,103,72,74
3,1
74,73,1,3,103,71,72,2
1,3,71,74
1,3
74,70,71,72
73,71,74
1,3,103,72,2
1,3,103,72,2
1,3,103,72
1,3,103,72,2
1,3,103,72,2
1,3,103,72,2
1,3,103,72,2
1,3,103,72,2

Penelope-109_1Mi	Locusta migratoria	Penelope	
Penelope-11_ACar	Anolis carolinensis	Penelope	73,1,3,71,72,74
Penelope-11_CoFl	Copidosoma floridanum	Penelope	1,3,2
Penelope-11_FoCa	Folsomia candida	Penelope	73,1,3,103,72
Penelope-11_HM	Hydra vulgaris	Penelope	1,3,2
Penelope-11_LMi	Locusta migratoria	Penelope	1,3,103,72
Penelope-11_LVa	Litopenaeus vannamei	Penelope	74,1,3,71,103,72,2
Penelope-11_NV	Nematostella vectensis	Penelope	
Penelope-11_PH	Parhyale hawaiensis	Penelope	1,3,2
Penelope-11_XL	Xenopus laevis	Penelope	
Penelope-11_XT	Xenopus tropicalis	Penelope	1,3,72,2
Penelope-110_1Mi	Locusta migratoria	Penelope	
Penelope-111_1Mi	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-112_1Mi	Locusta migratoria	Penelope	74,73,1,3,71,2
Penelope-113_1Mi	Locusta migratoria	Penelope	
Penelope-114_1Mi	Locusta migratoria	Penelope	
Penelope-115_1Mi	Locusta migratoria	Penelope	
Penelope-116_1Mi	Locusta migratoria	Penelope	
Penelope-117_1Mi	Locusta migratoria	Penelope	
Penelope-118_1Mi	Locusta migratoria	Penelope	
Penelope-119_1Mi	Locusta migratoria	Penelope	3,2,1
Penelope-12_HM	Hydra vulgaris	Penelope	3,1
Penelope-12_LMi	Locusta migratoria	Penelope	1,72,2
Penelope-12_LVa	Litopenaeus vannamei	Penelope	74,1,3,71,103,72,2
Penelope-12_PH	Parhyale hawaiensis	Penelope	1,3,2
Penelope-12_XL	Xenopus laevis	Penelope	
Penelope-12_XT	Xenopus tropicalis	Penelope	3,1
Penelope-120_1Mi	Locusta migratoria	Penelope	1,3,2
Penelope-120N1_1Mi	Locusta migratoria	Penelope	
Penelope-121_1Mi	Locusta migratoria	Penelope	1,3,70,71,74
Penelope-12B_1Mi	Locusta migratoria	Penelope	
Penelope-13_HM	Hydra vulgaris	Penelope	
Penelope-13_LMi	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-13_LVa	Litopenaeus vannamei	Penelope	73,1,3,71,74
Penelope-13_XL	Xenopus laevis	Penelope	1,3,72,2
Penelope-13_XT	Xenopus tropicalis	Penelope	1,3,72
Penelope-14_HM	Hydra vulgaris	Penelope	
Penelope-14_LMi	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2
Penelope-14_XL	Xenopus laevis	Penelope	
Penelope-15_HM	Hydra vulgaris	Penelope	
Penelope-15_LMi	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-15_LVa	Litopenaeus vannamei	Penelope	1,3,71,74
Penelope-15_XL	Xenopus laevis	Penelope	73,70,71,72,74
Penelope-15N1_1Mi	Locusta migratoria	Penelope	
Penelope-15N2_1Mi	Locusta migratoria	Penelope	
Penelope-16_HM	Hydra vulgaris	Penelope	
Penelope-16_LMi	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-16_XL	Xenopus laevis	Penelope	74,1,3,71,2
Penelope-17_HM	Hydra vulgaris	Penelope	
Penelope-17_LMi	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-17_XL	Xenopus laevis	Penelope	1,3,72

Penelope-18_HMa	.	Hydra vulgaris	Penelope
Penelope-18_LMi	.	Locusta migratoria	Penelope
Penelope-19_LMi	.	Locusta migratoria	Penelope
Penelope-1A_NV	.	Nematostella vectensis	Penelope
Penelope-1B_CM	.	Callorhinchus milii	Penelope
Penelope-1B_CPB	.	Chrysemys picta bellii	Penelope
Penelope-1B_XL	.	Xenopus laevis	Penelope
Penelope-1C_CPB	.	Chrysemys picta bellii	Penelope
Penelope-1N1_OD	.	Oikopleura dioica	Penelope
Penelope-1N1_PSi	.	Pelodiscus sinensis	Penelope
Penelope-1N2_CPB	.	Chrysemys picta bellii	Penelope
Penelope-2_AAe	.	Aedes aegypti	Penelope
Penelope-2_ACar	.	Anolis carolinensis	Penelope
Penelope-2_ADi	.	Acropora digitifera	Penelope
Penelope-2_AEc	.	Acromyrmex echinatior	Penelope
Penelope-2_BF	.	Branchiostoma floridae	Penelope
Penelope-2_BTa	.	Bemisia tabaci	Penelope
Penelope-2_CGi	.	Crassostrea gigas	Penelope
Penelope-2_CoF1	.	Copidosoma floridanum	Penelope
Penelope-2_CPB	.	Chrysemys picta bellii	Penelope
Penelope-2_CQ	.	Culex quinquefasciatus	Penelope
Penelope-2_Crp	.	Crocodylus porosus	Penelope
Penelope-2_DR	.	Danio rerio	Penelope
Penelope-2_FoCa	.	Folsomia candida	Penelope
Penelope-2_GA	.	Gasterosteus aculeatus	Penelope
Penelope-2_HM	.	Hydra vulgaris	Penelope
Penelope-2_HMM	.	Heliconius melpomene melpomene	Penelope
Penelope-2_HRo	.	Helobdella robusta	Penelope
Penelope-2_Lch	.	Latimeria chalumnae	Penelope
Penelope-2_LG	.	Lottia gigantea	Penelope
Penelope-2_LHu	.	Linepithema humile	Penelope
Penelope-2_LMi	.	Locusta migratoria	Penelope
Penelope-2_LVa	.	Litopenaeus vannamei	Penelope
Penelope-2_MLe	.	Mnemiopsis leidyi	Penelope
Penelope-2_NV	.	Nematostella vectensis	Penelope
Penelope-2_OD	.	Oikopleura dioica	Penelope
Penelope-2_PBa	.	Pogonomyrmex barbatus	Penelope
Penelope-2_PH	.	Parhyale hawaiensis	Penelope
Penelope-2_PSi	.	Pelodiscus sinensis	Penelope
Penelope-2_SIn	.	Solenopsis invicta	Penelope
Penelope-2_SK	.	Saccoglossus kowalevskii	Penelope
Penelope-2_SSa	.	Salmo salar	Penelope
Penelope-2_XL	.	Xenopus laevis	Penelope
Penelope-20_LMi	.	Locusta migratoria	Penelope
Penelope-21_LMi	.	Locusta migratoria	Penelope
Penelope-22_LMi	.	Locusta migratoria	Penelope
Penelope-23_LMi	.	Locusta migratoria	Penelope
Penelope-24_LMi	.	Locusta migratoria	Penelope
Penelope-26_LMi	.	Locusta migratoria	Penelope
Penelope-27_LMi	.	Locusta migratoria	Penelope
Penelope-28_LMi	.	Locusta migratoria	Penelope

74,73,1,3,103,71,72,2
1,3,103,72,2
1,3,103,72,2
1,3,71,72,2,74
1,3,103,72,2
1,3,71,72,74
73,1,3,70,71,72,74
73,74,1,3,71,72,2
1,3
73,1,3,72
3,1
1,3,72
1,3,70,71,2
1,3,71,103,72,74
73,1,3,70,71,48,62,74
103,3,1
1,3,103
74,73,1,3,70,71,72,2
1,3,103,72,2
1,3,103,72,2
1,3,2
73,70,71
1,3,71,103,74
1,3,2
73,74,1,3,71,103,72,2
1,3,70,71,72,74
1,3,72
1,3,103,72,2
1,3,103,72,2
1,3,103,72,2
1,3,103,72,2
1,3,103,72,2
74,73,1,3,71,72,2
1,3,103,72,2
73,1,3,103,2

Penelope-29_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,72,2
Penelope-2B_Crp	.	Crocodylus porosus	Penelope	1,3,71,72,2,74
Penelope-3_AAe	.	Aedes aegypti	Penelope	73,1,70,3,71,103,72,2
Penelope-3_ACar	.	Anolis carolinensis	Penelope	73,74,1,3,70,103,71,72,2
Penelope-3_ADi	.	Acropora digitifera	Penelope	
Penelope-3_BF	.	Branchiostoma floridae	Penelope	1,3,71,103,72,74
Penelope-3_BTa	.	Bemisia tabaci	Penelope	73,1,3,71,72,74
Penelope-3_CGi	.	Crassostrea gigas	Penelope	73,1,3,70,71,72,74
Penelope-3_CoFl	.	Copidosoma floridanum	Penelope	73,1,3,70,71,72,74
Penelope-3_CQ	.	Culex quinquefasciatus	Penelope	1,3,72
Penelope-3_Crp	.	Crocodylus porosus	Penelope	
Penelope-3_DR	.	Danio rerio	Penelope	
Penelope-3_FoCa	.	Folsomia candida	Penelope	
Penelope-3_HM	.	Hydra vulgaris	Penelope	
Penelope-3_HRo	.	Helobdella robusta	Penelope	1,3,103
Penelope-3_LCh	.	Latimeria chalumnae	Penelope	
Penelope-3_LG	.	Lottia gigantea	Penelope	73,70,71,72,74
Penelope-3_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-3_LVa	.	Litopenaeus vannamei	Penelope	74,1,3,71,103,2
Penelope-3_MLe	.	Mnemiopsis leidyi	Penelope	3,1
Penelope-3_NV	.	Nematostella vectensis	Penelope	
Penelope-3_OD	.	Oikopleura dioica	Penelope	
73				
Penelope-3_PH	.	Parhyale hawaiensis	Penelope	
Penelope-3_PSi	.	Pelodiscus sinensis	Penelope	74,1,3,70,71,103,72,2
Penelope-3_SIn	.	Solenopsis invicta	Penelope	1,3,103,72,2
Penelope-3_SSa	.	Salmo salar	Penelope	1,3,72,2
Penelope-3_XL	.	Xenopus laevis	Penelope	71,74
Penelope-30_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-31_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-32_LMi	.	Locusta migratoria	Penelope	1,3,72,2
Penelope-33_LMi	.	Locusta migratoria	Penelope	
Penelope-34_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-35_LMi	.	Locusta migratoria	Penelope	
103				
Penelope-36_LMi	.	Locusta migratoria	Penelope	1,3,72,2
Penelope-37_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-37B_LMi	.	Locusta migratoria	Penelope	
Penelope-38_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-39_LMi	.	Locusta migratoria	Penelope	1,3,103,72
Penelope-3B_DR	.	Danio rerio	Penelope	
Penelope-3B_PH	.	Parhyale hawaiensis	Penelope	74,1,3,71,2
Penelope-3B_PSi	.	Pelodiscus sinensis	Penelope	1,3,70,71,103,72,2
Penelope-3N1_LMi	.	Locusta migratoria	Penelope	
Penelope-4_AAe	.	Aedes aegypti	Penelope	74,73,1,3,70,103,71,72,2
Penelope-4_ACar	.	Anolis carolinensis	Penelope	1,3,71,74
Penelope-4_ADi	.	Acropora digitifera	Penelope	1,3,71,103,72,74
Penelope-4_BTa	.	Bemisia tabaci	Penelope	73,1,3,71,72,74
Penelope-4_CoFl	.	Copidosoma floridanum	Penelope	73,72
Penelope-4_CQ	.	Culex quinquefasciatus	Penelope	1,3,103
Penelope-4_Crp	.	Crocodylus porosus	Penelope	

Penelope-4_DR	.	Danio rerio	Penelope	
Penelope-4_FoCa	.	Folsomia candida	Penelope	73,70,71,72,74
Penelope-4_HM	.	Hydra vulgaris	Penelope	
Penelope-4_HRo	.	Helobdella robusta	Penelope	103,3,1
Penelope-4_LG	.	Lottia gigantea	Penelope	73,70,71,72
Penelope-4_LMi	.	Locusta migratoria	Penelope	1,3,103,72
Penelope-4_LVa	.	Litopenaeus vannamei	Penelope	1,3,103,72,2
Penelope-4_NV	.	Nematostella vectensis	Penelope	73,1,3,70,71,74
Penelope-4_OD	.	Oikopleura dioica	Penelope	1,3,103,2
Penelope-4_PH	.	Parhyale hawaiensis	Penelope	1,3,71,2,74
Penelope-4_SIn	.	Solenopsis invicta	Penelope	74,73,1,3,71,72,2
Penelope-4_SSa	.	Salmo salar	Penelope	1,3,72
Penelope-4_XL	.	Xenopus laevis	Penelope	
72				
Penelope-40_LMi	.	Locusta migratoria	Penelope	73,74,1,3,70,103,71,2
Penelope-41_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-42_LMi	.	Locusta migratoria	Penelope	73,1,3,2
Penelope-43_LMi	.	Locusta migratoria	Penelope	1,3
Penelope-44_LMi	.	Locusta migratoria	Penelope	1,3,71,74
Penelope-44B_LMi	.	Locusta migratoria	Penelope	
Penelope-45_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-46_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-47_LMi	.	Locusta migratoria	Penelope	1,3
Penelope-48_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-49_LMi	.	Locusta migratoria	Penelope	1,3,72,2
Penelope-4N1_LVa	.	Litopenaeus vannamei	Penelope	
Penelope-5_AAe	.	Aedes aegypti	Penelope	1,3,103,72,2
Penelope-5_ACar	.	Anolis carolinensis	Penelope	73,74,1,3,71,72,2
Penelope-5_BTa	.	Bemisia tabaci	Penelope	103,72
Penelope-5_CoFl	.	Copidosoma floridanum	Penelope	73,71,72,74
Penelope-5_DR	.	Danio rerio	Penelope	1,3,71,72,74
Penelope-5_FoCa	.	Folsomia candida	Penelope	1,3,103,2
Penelope-5_HM	.	Hydra vulgaris	Penelope	
Penelope-5_HRo	.	Helobdella robusta	Penelope	1,3,2
Penelope-5_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-5_LVa	.	Litopenaeus vannamei	Penelope	1,3,70,71,103,72,2
Penelope-5_NV	.	Nematostella vectensis	Penelope	1,3,70,71,72
Penelope-5_OD	.	Oikopleura dioica	Penelope	
Penelope-5_PH	.	Parhyale hawaiensis	Penelope	1,3,2
Penelope-5_SIn	.	Solenopsis invicta	Penelope	74,73,1,3,71,72,2
Penelope-5_SSa	.	Salmo salar	Penelope	73,1,3,70,71,72,74
Penelope-5_XL	.	Xenopus laevis	Penelope	74,1,3,71,103,72
Penelope-5_XT	.	Xenopus tropicalis	Penelope	74,1,71,103,72,2
Penelope-50_LMi	.	Locusta migratoria	Penelope	1,3
Penelope-51_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,2
Penelope-52_LMi	.	Locusta migratoria	Penelope	1,3,2
Penelope-52B_LMi	.	Locusta migratoria	Penelope	
Penelope-53_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-54_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-55_LMi	.	Locusta migratoria	Penelope	73,74,1,3,71,2
Penelope-56_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2

Penelope-57_LMi	.	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2
Penelope-58_LMi	.	Locusta migratoria	Penelope	74,1,3,71,2
Penelope-59_LMi	.	Locusta migratoria	Penelope	73,1,3,103,2
Penelope-5N1_BTa	.	Bemisia tabaci	Penelope	
Penelope-6_ACar	.	Anolis carolinensis	Penelope	1,3,71,72,74
Penelope-6_BTa	.	Bemisia tabaci	Penelope	1,3,2
Penelope-6_CoF1	.	Copidosoma floridanum	Penelope	73,70,71,72,74
Penelope-6_FoCa	.	Folsomia candida	Penelope	1,3,70,71,103,72,2
Penelope-6_HM	.	Hydra vulgaris	Penelope	
Penelope-6_HRo	.	Helobdella robusta	Penelope	1,3,2
Penelope-6_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-6_LVa	.	Litopenaeus vannamei	Penelope	74,1,3,71,2
Penelope-6_NV	.	Nematostella vectensis	Penelope	
Penelope-6_OD 103	.	Oikopleura dioica	Penelope	
Penelope-6_PH	.	Parhyale hawaiensis	Penelope	1,3,2
Penelope-6_SIn	.	Solenopsis invicta	Penelope	73,74,1,3,70,71,103,72,2
Penelope-6_SSa	.	Salmo salar	Penelope	73,62
Penelope-6_XL	.	Xenopus laevis	Penelope	73,72
Penelope-6_XT	.	Xenopus tropicalis	Penelope	1,3,103,72,2
Penelope-60_LMi	.	Locusta migratoria	Penelope	73,1,3,103,2
Penelope-62_LMi	.	Locusta migratoria	Penelope	1,3,2
Penelope-63_LMi	.	Locusta migratoria	Penelope	1,3,2
Penelope-64_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-65_LMi	.	Locusta migratoria	Penelope	1,3,2
Penelope-66_LMi 72	.	Locusta migratoria	Penelope	
Penelope-67_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,72,2
Penelope-68_LMi	.	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2
Penelope-69_LMi 73	.	Locusta migratoria	Penelope	
Penelope-69B_LMi	.	Locusta migratoria	Penelope	
Penelope-6N1_NV	.	Nematostella vectensis	Penelope	
Penelope-7_ACar	.	Anolis carolinensis	Penelope	73,1,3,103,72,2
Penelope-7_BTa	.	Bemisia tabaci	Penelope	1,3,71,74
Penelope-7_CoF1	.	Copidosoma floridanum	Penelope	73,1,3,71,72,74
Penelope-7_FoCa	.	Folsomia candida	Penelope	1,3,103,72,2
Penelope-7_HM	.	Hydra vulgaris	Penelope	
Penelope-7_HRo	.	Helobdella robusta	Penelope	
Penelope-7_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,72,2
Penelope-7_LVa	.	Litopenaeus vannamei	Penelope	74,1,3,71,2
Penelope-7_NV	.	Nematostella vectensis	Penelope	
Penelope-7_PH	.	Parhyale hawaiensis	Penelope	74,1,3,71,103,2
Penelope-7_SSa	.	Salmo salar	Penelope	1,3,71,72,74
Penelope-7_XL	.	Xenopus laevis	Penelope	
Penelope-7_XT	.	Xenopus tropicalis	Penelope	18,74,5,1,71,103,72,2
Penelope-70_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-71_LMi	.	Locusta migratoria	Penelope	1,3,70,71
Penelope-72_LMi	.	Locusta migratoria	Penelope	73,1,3,70,71,74
Penelope-72N1_LMi	.	Locusta migratoria	Penelope	
Penelope-73_LMi	.	Locusta migratoria	Penelope	73,74,1,3,71,2

Penelope-73N1_LMi	.	Locusta migratoria	Penelope	
Penelope-74_LMi	.	Locusta migratoria	Penelope	
Penelope-75_LMi	.	Locusta migratoria	Penelope	3,1
Penelope-76_LMi	.	Locusta migratoria	Penelope	74,1,3,71,2
Penelope-77_LMi	.	Locusta migratoria	Penelope	74,1,3,71,103,72,2
Penelope-78_LMi	.	Locusta migratoria	Penelope	
Penelope-79_LMi	.	Locusta migratoria	Penelope	1,3,103,72,2
Penelope-8_ACar	.	Anolis carolinensis	Penelope	73,70,71,72,74
Penelope-8_BTa	.	Bemisia tabaci	Penelope	1,103,72,2
Penelope-8_CoF1	.	Copidosoma floridanum	Penelope	1,3,72
Penelope-8_FoCa	.	Folsomia candida	Penelope	1,3,103,72
Penelope-8_HM	.	Hydra vulgaris	Penelope	
Penelope-8_HRo	.	Helobdella robusta	Penelope	1,3,103,2
Penelope-8_LMi	.	Locusta migratoria	Penelope	1,3,103,2
Penelope-8_LVa	.	Litopenaeus vannamei	Penelope	1,3,103,2
Penelope-8_NV	.	Nematostella vectensis	Penelope	
Penelope-8_PH	.	Parhyale hawaiensis	Penelope	1,3,103,2
Penelope-8_SSa	.	Salmo salar	Penelope	
Penelope-8_XL	.	Xenopus laevis	Penelope	
Penelope-8_XT	.	Xenopus tropicalis	Penelope	
Penelope-80_LMi	.	Locusta migratoria	Penelope	74,73,1,3,71,103,72,2
Penelope-81_LMi	.	Locusta migratoria	Penelope	
Penelope-82_LMi	.	Locusta migratoria	Penelope	3,1
Penelope-83_LMi	.	Locusta migratoria	Penelope	74,1,3,71,103,2
Penelope-84_LMi	.	Locusta migratoria	Penelope	
Penelope-85_LMi	.	Locusta migratoria	Penelope	
Penelope-86_LMi	.	Locusta migratoria	Penelope	73,1,3,103,72
Penelope-87_LMi	.	Locusta migratoria	Penelope	
Penelope-88_LMi	.	Locusta migratoria	Penelope	
Penelope-89_LMi	.	Locusta migratoria	Penelope	
Penelope-9_ACar	.	Anolis carolinensis	Penelope	73,1,3,72
Penelope-9_CoF1	.	Copidosoma floridanum	Penelope	1,3,72
Penelope-9_FoCa	.	Folsomia candida	Penelope	1,3,103,72
Penelope-9_HM	.	Hydra vulgaris	Penelope	
Penelope-9_HRo	.	Helobdella robusta	Penelope	1,3,103,2
Penelope-9_LMi	.	Locusta migratoria	Penelope	74,73,1,3,103,71,72,2
Penelope-9_LVa	.	Litopenaeus vannamei	Penelope	18,73,5,1,3,71,74
Penelope-9_NV	.	Nematostella vectensis	Penelope	
Penelope-9_PH	.	Parhyale hawaiensis	Penelope	
Penelope-9_XL	.	Xenopus laevis	Penelope	73,1,3,71,74
Penelope-9_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,103,72,2
Penelope-90_LMi	.	Locusta migratoria	Penelope	
Penelope-91_LMi	.	Locusta migratoria	Penelope	
Penelope-92_LMi	.	Locusta migratoria	Penelope	
Penelope-93_LMi	.	Locusta migratoria	Penelope	
Penelope-94_LMi	.	Locusta migratoria	Penelope	
Penelope-94N1_LMi	.	Locusta migratoria	Penelope	
Penelope-95_LMi	.	Locusta migratoria	Penelope	
Penelope-96_LMi	.	Locusta migratoria	Penelope	1,3,70,71,103,74
Penelope-97_LMi	.	Locusta migratoria	Penelope	71,73,74
Penelope-98_LMi	.	Locusta migratoria	Penelope	

Penelope-99_LMi 73	.	Locusta migratoria	Penelope		
Penelope-N1_PH	.	Parhyale hawaiensis	Penelope		
Penelope-N1_XL	.	Xenopus laevis	Penelope		
Penelope-N1_XT	.	Xenopus tropicalis	Penelope		
Penelope-N10_XL	.	Xenopus laevis	Penelope		
Penelope-N12_XL	.	Xenopus laevis	Penelope		
Penelope-N13_XL	.	Xenopus laevis	Penelope		
Penelope-N14_XL	.	Xenopus laevis	Penelope		
Penelope-N15_XL	.	Xenopus laevis	Penelope		
Penelope-N15B_XL	.	Xenopus laevis	Penelope		
Penelope-N16_XL	.	Xenopus laevis	Penelope		
Penelope-N17_XL	.	Xenopus laevis	Penelope		
Penelope-N18_XL	.	Xenopus laevis	Penelope		
Penelope-N19_XL	.	Xenopus laevis	Penelope		
Penelope-N2_XL	.	Xenopus laevis	Penelope		
Penelope-N20_XL	.	Xenopus laevis	Penelope		
Penelope-N21_XL	.	Xenopus laevis	Penelope		
Penelope-N22_XL	.	Xenopus laevis	Penelope		
Penelope-N23_XL	.	Xenopus laevis	Penelope		
Penelope-N3_XL	.	Xenopus laevis	Penelope		
Penelope-N4_XL	.	Xenopus laevis	Penelope		
Penelope-N447_AMi	.	Alligator mississippiensis	Penelope		
Penelope-N5_XL	.	Xenopus laevis	Penelope		
Penelope-N6_XL	.	Xenopus laevis	Penelope		
Penelope-N7_XL	.	Xenopus laevis	Penelope		
Penelope-N8_XL	.	Xenopus laevis	Penelope		
Penelope-N9_XL	.	Xenopus laevis	Penelope		
Penelope1_CI	.	Ciona intestinalis	Penelope		
Penelope1_Dw	.	Drosophila willistoni	Penelope	1,3,71,103,72,74	
Penelope1_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,72,2	
Penelope1A_XT	.	Xenopus tropicalis	Penelope	1,3,71,72,2,74	
Penelope1B_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,72,2	
Penelope1C_XT	.	Xenopus tropicalis	Penelope	1,3,71,72,2,74	
Penelope1D_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,72,2	
Penelope1N_CPB	.	Chrysemys picta bellii	Penelope		
Penelope2_Dw	.	Drosophila willistoni	Penelope	74,1,3,71,103,72,2	
Penelope2_XT	.	Xenopus tropicalis	Penelope	1,3,72	
Penelope3_XT	.	Xenopus tropicalis	Penelope	74,1,3,71,72,2	
Penelope4_XT	.	Xenopus tropicalis	Penelope	1,3,71,72,74	
Perere_INT	BK004067	Schistosoma mansoni	CR1	16,1,3,11,4,2,15	
Perere_Smed	.	Schmidtea mediterranea	Penelope	1,3,103,2	
PERERE-10	BN000801	Schistosoma mansoni	Penelope	1,3,103,2	
PERERE-2	BN000793	Schistosoma mansoni	CR1	1,3,59,4,57,2	
Perere-3	BN000794	Schistosoma mansoni	RTE	16,1,26,3,11,4,2,15	
PERERE-4	BN000795	Schistosoma mansoni	CR1	16,1,3,4,11,2,15	
PERERE-5	BN000796	Schistosoma mansoni	CR1	16,1,3,59,60,11,4,57,2,15	
PERERE-6	BN000797	Schistosoma mansoni	CR1	1,3,59,4,57,2	
PERERE-7	BN000798	Schistosoma mansoni	CR1	16,1,3,11,4,2,15	
PERERE-8	BN000799	Schistosoma mansoni	Rex1	1,3,4,2	
PERERE-9	BN000800	Schistosoma mansoni	R2	18,5,19,1,3,4,2	

Perere2_Smed	.	Schmidtea mediterranea	Penelope	73,1,3,103,2	
PILN1_PT	AB016145	Pinus thunbergii	L1		
PILN2_PT	AB016146	Pinus thunbergii	L1		
PILN3_PT	AB016147	Pinus thunbergii	L1		
Plat_L3	.	Ornithorhynchus	CR1	16,1,3,11,4,2,15	
Plat_L3b	.	Ornithorhynchus	CR1		
Plat_R4	.	Ornithorhynchus	R4		
Plat_RTE1	.	Ornithorhynchus	RTE		
Plat_RTE1_SINE	.	Ornithorhynchus	SINE		
PlatCR1	.	Ornithorhynchus	CR1		
PlatCR1_old1	.	Ornithorhynchus	CR1		
PlatCR1_old2	.	Ornithorhynchus	CR1		
PlatSINE1	.	Ornithorhynchus	SINE		
Platy-1-1	.	Callithrix jacchus	SINE1/7SL		
PNL_SM	.	Schmidtea mediterranea	Penelope	1,3,2	
PNL2_SM	.	Schmidtea mediterranea	Penelope	73,1,3,70,71,103,2	
POLN2_ST	AB016141, AB016141.1	Solanum tuberosum	L1		
POMSINE	.	Podarcis muralis	SINE2/tRNA		
Poseidon_Ac	.	Anolis carolinensis	Penelope	1,3,103,72,2	
Poseidon_Ap	.	Acropora palmata	Penelope	1,3,71,103,72,74	
Poseidon_Hyd	.	Hydra vulgaris	Penelope	71,72,74	
Poseidon-1_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2	
Poseidon-10_HM	.	Hydra vulgaris	Penelope	71,74	
Poseidon-1B_PM	.	Petromyzon marinus	Penelope	1,3,71,103,74	
Poseidon-1C_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2	
Poseidon-1D_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2	
Poseidon-1E_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2	
Poseidon-1F_PM	.	Petromyzon marinus	Penelope	1,3,71,103,72,74	
Poseidon-1G_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2	
Poseidon-1H_PM	.	Petromyzon marinus	Penelope		
Poseidon-1I_PM	.	Petromyzon marinus	Penelope		
Poseidon-1J_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2	
Poseidon-1K_PM	.	Petromyzon marinus	Penelope		
Poseidon-1N1_PM	.	Petromyzon marinus	Penelope		
Poseidon-1N2_PM	.	Petromyzon marinus	Penelope		
Poseidon-2_HM	.	Hydra vulgaris	Penelope	71,72,74	
Poseidon-2_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,2	
Poseidon-2B_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,2	
Poseidon-2C_PM	.	Petromyzon marinus	Penelope	74,71	
Poseidon-2D_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,2	
Poseidon-3_HM	.	Hydra vulgaris	Penelope		
Poseidon-3_PM	.	Petromyzon marinus	Penelope	74,73,1,3,71,103,72,2	
Poseidon-3B_PM	.	Petromyzon marinus	Penelope	74,73,1,3,71,103,72,2	
Poseidon-4_HM	.	Hydra vulgaris	Penelope		
Poseidon-4_PM	.	Petromyzon marinus	Penelope	74,1,3,71,103,72,2	
Poseidon-5_HM	.	Hydra vulgaris	Penelope	71,74	
Poseidon-6_HM	.	Hydra vulgaris	Penelope	1,3	
Poseidon-N1_PM	.	Petromyzon marinus	Penelope		
Poseidon-N2_PM	.	Petromyzon marinus	Penelope		
Poseidon-N3_PM	.	Petromyzon marinus	Penelope		
Poseidon-N3B_PM	.	Petromyzon marinus	Penelope		

Poseidon-N4_PM	.	Petromyzon marinus	Penelope		
Poseidon-N5_PM	.	Petromyzon marinus	Penelope		
Poseidon-N6_PM	.	Petromyzon marinus	Penelope		
Pp_NLTR-B	.	Polysphondylium pallidum	Non-LTRRetrotransposo	6,2,10,11,13,1,4,17,3,7	
Pp_NLTR-C	.	Polysphondylium pallidum	Non-LTRRetrotransposo	1,3,2	
Pp_TRE3-A	.	Polysphondylium pallidum	Non-LTRRetrotransposo	1,3,11,4,2	
Pp_TRE3-B	.	Polysphondylium pallidum	Non-LTRRetrotransposo	1,3,11,4,2	
Pre0_SS	.	Sus scrofa	SINE2/tRNA		
PRE1_SS	Y00104	Sus scrofa	SINE2/tRNA		
PRE1f2	.	Suina	SINE2/tRNA		
PRE1h	.	Suina	SINE2/tRNA		
PRE1h2	.	Suina	SINE2/tRNA		
PRE1i	.	Suina	SINE2/tRNA		
PRE1j	.	Suina	SINE2/tRNA		
PRE1k	.	Suina	SINE2/tRNA		
PRLN1_LP	AB016125	Lolium perenne	L1		
Proto1-1_NG	.	Naegleria gruberi	Proto1	7,1,17,3,10,11,2,6,13	
Proto1-2_NG	.	Naegleria gruberi	Proto1	7,1,17,3,10,11,6,2,13	
Proto1-3_NG	.	Naegleria gruberi	Proto1	7,1,17,3,10,11,2,6,13	
Proto1-4_NG	.	Naegleria gruberi	Proto1	6,2,9,10,11,13,1,4,17,3,7	
Proto1-5_NG	.	Naegleria gruberi	Proto1	6,2,7,9,10,11,13,1,4,17,3	
Proto1-6_NG	.	Naegleria gruberi	Proto1	7,1,17,9,3,10,4,6,2,13	
Proto2-1_BF	.	Branchiostoma floridae	Proto2	16,1,3,24,11,4,2	
Proto2-1_BM	.	Bombyx mori	Proto2	1,3,2	
Proto2-1_CS1	.	Capitella sp. 1	Proto2	16,1,3,24,11,4,2	
Proto2-1_CTe	.	Capitella teleta	Proto2	1,3,4,2	
Proto2-1_DR	.	Danio rerio	Proto2	2,1	
Proto2-1_Hmel	.	Heliconius melpomene	Proto2	1,3,4,2	
Proto2-1_LVa	.	Litopenaeus vannamei	Proto2	1,3,11,4,2	
Proto2-1_OL	.	Oryzias latipes	Proto2	1,3,2	
Proto2-1_PH	.	Parhyale hawaiensis	Proto2	16,1,3,24,11,4,2	
Proto2-1_PPo	.	Papilio polytes	Proto2	16,1,3,11,4,2,15	
Proto2-1_PXu	.	Papilio xuthus	Proto2		
Proto2-1_SK	.	Saccoglossus kowalevskii	Proto2	16,1,3,11,4,2,15	
Proto2-1_SPur	.	Strongylocentrotus purpuratus	Proto2	2,11,56,57,68,1,58,4,69,3,59	
Proto2-10_CTe	.	Capitella teleta	Proto2	1,3,4,2	
Proto2-11_CTe	.	Capitella teleta	Proto2	1,3,11,4,2	
Proto2-12_CTe	.	Capitella teleta	Proto2	1,3,11,4,2	
Proto2-13_CTe	.	Capitella teleta	Proto2	16,1,3,24,4,11,2	
Proto2-14_CTe	.	Capitella teleta	Proto2	1,3,11,4,2	
Proto2-15_CTe	.	Capitella teleta	Proto2	16,1,3,24,11,4,2	
Proto2-16_CTe	.	Capitella teleta	Proto2	1,3,4,2	
Proto2-2_CS1	.	Capitella sp. 1	Proto2	16,1,3,11,4,2,15	
Proto2-2_CTe	.	Capitella teleta	Proto2		
Proto2-2_Hmel	.	Heliconius melpomene	Proto2		
Proto2-2_HMM	.	Heliconius melpomene melpomene	Proto2	1,3,11,4,2	
Proto2-2_LVa	.	Litopenaeus vannamei	Proto2		
Proto2-2_PH	.	Parhyale hawaiensis	Proto2	1,3,11,4,2	
Proto2-2_PPo	.	Papilio polytes	Proto2		
Proto2-2_SK	.	Saccoglossus kowalevskii	Proto2	1,3,4,2	
Proto2-3_CS1	.	Capitella sp. 1	Proto2	16,1,3,11,4,2,15	

Proto2-3_CTe	.	Capitella teleta	Proto2	1,3,4,2	
Proto2-3_Hmel	.	Heliconius melpomene	Proto2	1,3,4,2	
Proto2-3_PPo	.	Papilio polytes	Proto2		
Proto2-3_SK	.	Saccoglossus kowalevskii	Proto2	1,3,82,11,4,2	
Proto2-3B_PPo	.	Papilio polytes	Proto2		
Proto2-4_CS1	.	Capitella sp. 1	Proto2	1,3,11,4,2	
Proto2-4_CTe	.	Capitella teleta	Proto2	1,3,4,2	
Proto2-4_PPo	.	Papilio polytes	Proto2		
Proto2-5_CS1	.	Capitella sp. 1	Proto2	16,1,3,11,4,2,15	
Proto2-5_CTe	.	Capitella teleta	Proto2	1,3,4,2	
Proto2-5_PPo	.	Papilio polytes	Proto2		
Proto2-6_CS1	.	Capitella sp. 1	Proto2	16,1,3,24,11,4,2	
Proto2-6_CTe	.	Capitella teleta	Proto2	1,3,4,2	
Proto2-7_CS1	.	Capitella sp. 1	Proto2	1,3,11,4,2	
Proto2-7_CTe	.	Capitella teleta	Proto2	1,3,4,2	
Proto2-8_CS1	.	Capitella sp. 1	Proto2	1,3,11,4,2	
Proto2-8_CTe	.	Capitella teleta	Proto2	3,2,1	
Proto2-9_CTe	.	Capitella teleta	Proto2	3,2,1	
PSINE1	.	Procavia capensis	SINE2/tRNA		
PSLINE	AB005891	Acanthochelys spixii	CR1	2,63,11,48,1,64,4,15,16,3	
Pt1	.	Phaeodactylum tricornutum	Penelope		
82					
PTr-3	.	Populus trichocarpa	SINE2/tRNA		
Q	U03849	Anopheles gambiae	CR1	53,2,54,11,56,57,68,1,58,4,15,16,3,60	
Q_As	.	Adineta	Penelope	1,3	
Q_Av	.	Adineta vaga	Penelope	1,3	
R_Av	.	Adineta vaga	Penelope	1,129,3,130	
R0_As	.	Adineta	Penelope	1,3	
R1_DEr	.	Drosophila erecta	R1	2,8,11,86,12,1,14,4,15,16,3	
R1_DM	X51968	Drosophila melanogaster	R1	2,8,11,86,12,1,14,4,15,16,3	
R1_DMo	.	Drosophila mojavensis	R1	2,8,11,12,157,1,14,4,15,16,3,158	
R1_DPe	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3	
R1_DPs	.	Drosophila pseudoobscura	R1	128,2,8,11,86,12,1,14,4,15,16,3	
R1_DSe	.	Drosophila sechellia	R1	2,8,11,86,12,1,14,4,15,16,3	
R1_DSi	.	Drosophila simulans	R1	2,8,11,86,12,1,14,4,15,16,3	
R1_DVi	.	Drosophila virilis	R1	2,8,11,12,42,1,14,4,15,16,3	
R1_DWi	.	Drosophila willistoni	R1	2,8,11,12,1,14,4,15,16,3	
R1_DYa	.	Drosophila yakuba	R1	2,8,11,86,12,1,14,4,15,16,3	
R1_Ele1	.	Aedes aegypti	R1	2,8,11,86,12,1,14,4,15,16,3	
R1_Ele4	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3	
R1_Ele7	.	Aedes aegypti	R1	1,3,4,2	
R1_Ele8	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3	
R1_Ele9	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3	
R1_SC	L00945	Bradysia coprophila	R1		
R1-1_AEc	.	Acromyrmex echinatior	R1	16,1,3,11,4,86,2,15	
R1-1_AP	Contig61185	Acyrthosiphon pisum	R1	2,8,11,86,12,1,14,4,15,16,3	
R1-1_BM	.	Bombyx mori	R1	16,1,3,11,4,86,2,15	
R1-1_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15	
R1-1_CF1	.	Camponotus floridanus	R1		
R1-1_CoF1	.	Copidosoma floridanum	R1	16,1,3,11,4,86,2,15	
R1-1_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,42,1,14,4,15,16,3	

R1-1_DBi	.	Drosophila biarmipes	R1	2,11,86,12,1,14,4,15,16,3
R1-1_DBp	.	Drosophila bipectinata	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-1_DEl	.	Drosophila elegans	R1	16,1,3,11,4,86,2,15
R1-1_DEr	.	Drosophila erecta	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DEu	.	Drosophila eugracilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DF	.	Drosophila ficusphila	R1	2,11,86,12,1,14,4,15,16,3
R1-1_DGr	.	Drosophila grimshawi	R1	2,8,11,12,42,1,14,4,15,16,3
R1-1_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DRh	.	Drosophila rhopaloa	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DSer	.	Drosophila serrata	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DSi	.	Drosophila simulans	R1	16,1,3,11,4,86,2,15
R1-1_DT	.	Drosophila takahashii	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_DVi	.	Drosophila virilis	R1	2,8,11,86,12,42,1,14,4,15,16,3,46
R1-1_DYa	.	Drosophila yakuba	R1	16,1,3,11,4,86,2,15
R1-1_Hmel	.	Heliconius melpomene	R1	1,3,11,4,2
R1-1_HMM	.	Heliconius melpomene melpomene	R1	6,2,9,10,11,86,13,1,4,15,16,17,3,7
R1-1_HSal	.	Harpegnathos saltator	R1	16,1,3,11,4,86,2,15
R1-1_LHu	.	Linepithema humile	R1	16,1,3,11,4,86,2,15
R1-1_LMi	.	Locusta migratoria	R1	6,2,9,8,10,11,12,13,42,1,14,4,15,16,17,3,7
R1-1_NVi	.	Nasonia vitripennis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1_PBa	.	Pogonomyrmex barbatus	R1	1,3,4,2
R1-1_PPo	.	Papilio polytes	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-1_PXu	.	Papilio xuthus	R1	
R1-1_SIn	.	Solenopsis invicta	R1	16,1,3,11,4,86,2,15
R1-1_TCas	.	Tribolium castaneum	R1	2,8,11,12,42,1,14,4,15,16,3
R1-10_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,12,2,15
R1-10_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-11_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-11_PPo	.	Papilio polytes	R1	
R1-12_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-12_PPo	.	Papilio polytes	R1	
R1-13_DWi	.	Drosophila willistoni	R1	2,11,86,12,1,14,4,15,16,3
R1-13_PPo	.	Papilio polytes	R1	
R1-14_DWi	.	Drosophila willistoni	R1	2,11,86,12,1,14,4,15,16,3
R1-14_PPo	.	Papilio polytes	R1	
R1-15_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-15_PPo	.	Papilio polytes	R1	
R1-15B_PPo	.	Papilio polytes	R1	
R1-16_DWi	.	Drosophila willistoni	R1	16,1,3,11,4,2,15
R1-16_PPo	.	Papilio polytes	R1	
R1-16B_PPo	.	Papilio polytes	R1	
R1-17_DWi	.	Drosophila willistoni	R1	16,1,3,11,4,86,2,15
R1-17_PPo	.	Papilio polytes	R1	
R1-18_DWi	.	Drosophila willistoni	R1	
R1-1B_CoFl	.	Copidosoma floridanum	R1	42,8,14,12
R1-1B_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,1,14,4,15,16,3
R1-1B_DVi	.	Drosophila virilis	R1	
R1-1C_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,86,2,15
R1-2_AP	Contig3122	Acyrthosiphon pisum	R1	16,1,8,3,11,4,86,12,2,15
R1-2_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15

R1-2_CoFl	.	Copidosoma floridanum	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-2_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DBi	.	Drosophila biarmipes	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DBp	.	Drosophila bipectinata	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DEl	.	Drosophila elegans	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DEu	.	Drosophila eugracilis	R1	16,1,8,3,11,4,86,12,2,15
R1-2_DF	.	Drosophila ficusphila	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DGr	.	Drosophila grimshawi	R1	2,8,11,12,42,1,14,4,15,16,3
R1-2_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,86,2,15
R1-2_DM	.	Drosophila melanogaster	R1	2,8,11,12,1,14,4,15,16,3
R1-2_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DSer	.	Drosophila serrata	R1	2,8,11,86,12,1,14,4,15,16,3
R1-2_DTa	.	Drosophila takahashii	R1	2,11,86,12,42,1,14,4,15,16,3
R1-2_DVi	.	Drosophila virilis	R1	2,11,86,12,1,14,4,15,16,3
R1-2_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-2_DYa	.	Drosophila yakuba	R1	16,1,3,11,4,86,2,15
R1-2_Hmel	.	Heliconius melpomene	R1	16,1,3,11,4,86,2,15
R1-2_LMi	.	Locusta migratoria	R1	1,8,14,3,4,12,2
R1-2_PBa	.	Pogonomyrmex barbatus	R1	16,1,3,4,11,86,2,15
R1-2_PPo	.	Papilio polytes	R1	7,1,17,9,3,10,4,6,2,13
R1-2_SIn	.	Solenopsis invicta	R1	1,3,4,2
R1-2_TCas	.	Tribolium castaneum	R1	16,1,3,4,11,86,2,15
R1-2B_AP	Contig17211	Acyrthosiphon pisum	R1	2,8,11,12,1,14,4,15,16,3
R1-2B_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,86,2,15
R1-2B_PPo	.	Papilio polytes	R1	
R1-2C_PPo	.	Papilio polytes	R1	
R1-3_AP	Contig971	Acyrthosiphon pisum	R1	16,1,14,3,11,4,86,2,15
R1-3_BM	.	Bombyx mori	R1	
R1-3_BTa	.	Bemisia tabaci	R1	2,16,1,14,3,11,4,15
R1-3_CoFl	.	Copidosoma floridanum	R1	
R1-3_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DBp	.	Drosophila bipectinata	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DEl	.	Drosophila elegans	R1	16,1,8,11,12,2,15
R1-3_DEu	.	Drosophila eugracilis	R1	
R1-3_DF	.	Drosophila ficusphila	R1	16,1,3,11,4,86,2,15
R1-3_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-3_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DTa	.	Drosophila takahashii	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DVi	.	Drosophila virilis	R1	1,3,4,2
R1-3_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_DYa	.	Drosophila yakuba	R1	2,8,11,86,12,1,14,4,15,16,3
R1-3_LMi	.	Locusta migratoria	R1	16,1,3,11,4,2,15
R1-3_PBa	.	Pogonomyrmex barbatus	R1	190,2,8,11,86,12,42,1,14,4,15,16,3
R1-3_PPo	.	Papilio polytes	R1	
R1-3_SIn	.	Solenopsis invicta	R1	1,3,11,4,2
R1-3B_DEl	.	Drosophila elegans	R1	1,3,4,2
R1-4_AP	.	Acyrthosiphon pisum	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_BM	.	Bombyx mori	R1	1,3,4,2
R1-4_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15
R1-4_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DBp	.	Drosophila bipectinata	R1	16,1,3,11,4,86,2,15

R1-4_DF	.	Drosophila ficusphila	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DVi	.	Drosophila virilis	R1	
R1-4_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-4_DYa	.	Drosophila yakuba	R1	16,1,3,11,4,86,2,15
R1-4_PPo	.	Papilio polytes	R1	
R1-4B_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,86,2,15
R1-5_BM	.	Bombyx mori	R1	2,8,24,11,86,12,42,1,14,16
R1-5_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15
R1-5_DF	.	Drosophila ficusphila	R1	2,8,11,86,12,1,14,4,15,16,3
R1-5_DK	.	Drosophila kikkawai	R1	1,3,11,4,2
R1-5_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,1,14,4,15,16,3
R1-5_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-5_DYa	.	Drosophila yakuba	R1	16,1,3,11,4,86,2,15
R1-5_PPo	.	Papilio polytes	R1	
R1-5B_DK	.	Drosophila kikkawai	R1	2,8,11,86,12,1,14,4,15,16,3
R1-6_BM	.	Bombyx mori	R1	
R1-6_BTa	.	Bemisia tabaci	R1	2,8,11,86,12,93,42,1,14,4,15,16,3
R1-6_DK	.	Drosophila kikkawai	R1	43,2,8,44,11,86,12,42,1,14,4,15,16,3
R1-6_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-6_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-6_PPo	.	Papilio polytes	R1	
R1-7_BTa	.	Bemisia tabaci	R1	16,1,3,11,4,2,15
R1-7_DK	.	Drosophila kikkawai	R1	16,1,3,11,4,2,15
R1-7_DPer	.	Drosophila persimilis	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1-7_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-7_PPo	.	Papilio polytes	R1	
R1-8_BTa	.	Bemisia tabaci	R1	2,8,11,12,1,14,4,15,16,3
R1-8_DWi	.	Drosophila willistoni	R1	2,11,86,12,1,14,4,15,16,3
R1-8_PPo	.	Papilio polytes	R1	
R1-9_BTa	.	Bemisia tabaci	R1	16,1,8,3,11,4,12,2,15
R1-9_DWi	.	Drosophila willistoni	R1	2,8,11,86,12,1,14,4,15,16,3
R1-9_PPo	.	Papilio polytes	R1	
R1-N1_CFl	.	Camponotus floridanus	R1	8,14,12
R1-N1_SIn	.	Solenopsis invicta	R1	8,14,12
R1A_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,157,1,14,4,15,16,3,158
R1A_DS	AF015489	Aphonopelma sp. WDB-1998	R1	7,1,17,9,3,10,4,6,2,13
R1A_DSer	.	Drosophila serrata	R1	2,8,11,86,12,1,14,4,15,16,3
R1A_NVi	.	Nasonia vitripennis	R1	2,8,11,86,12,1,14,4,15,16,3
R1B_DAn	.	Drosophila ananassae	R1	2,8,11,86,12,1,14,4,15,16,3
R1B_DSer	.	Drosophila serrata	R1	2,8,11,12,1,14,4,15,16,3
R1B_DVi	.	Drosophila virilis	R1	2,8,11,12,42,1,14,4,15,16,3
R1B_NVi	.	Nasonia vitripennis	R1	2,8,11,12,1,14,4,15,16,3
R1B2_DAn	.	Drosophila ananassae	R1	
R1C_NGi	.	Nasonia giraulti	R1	16,1,8,3,11,4,12,2,15
R1C_NLo	.	Nasonia longicornis	R1	16,1,8,3,11,4,86,12,2,15
R1C_NVi	.	Nasonia vitripennis	R1	2,8,11,86,12,42,1,14,4,15,16,3
R1D_NGi	.	Nasonia giraulti	R1	16,1,3,4,11,2,15
R1D_NLo	.	Nasonia longicornis	R1	16,1,3,11,4,2,15
R1D_NVi	.	Nasonia vitripennis	R1	16,1,3,4,11,2,15

R1E_NGi	.	Nasonia giraulti	R1	2,8,11,86,12,1,14,4,15,16,3
R1N_NLo	.	Nasonia longicornis	R1	
R1NS-1_CQ	.	Culex quinquefasciatus	R1	2,8,11,86,12,1,14,4,15,16,3
R2_AM	AF015815	Anurida maritima	R2	
R2_BM	M16558	Bombyx mori	R2	18,5,19,1,3,4,2
R2_CI	AB097121	Ciona intestinalis	R2	
R2_CPB	.	Chrysemys picta bellii	R2	
R2_DA	U13640	Drosophila ambigua	R2	1,3,2
R2_DAn	.	Drosophila ananassae	R2	18,5,1,3,4,2
R2_DM	X51967	Drosophila melanogaster	R2	18,5,19,1,3,4,2
R2_DPe	.	Drosophila persimilis	R2	18,5,1,19,3,4,2
R2_DPs	.	Drosophila pseudoobscura	R2	18,5,1,19,3,4,2
R2_DSe	.	Drosophila sechellia	R2	18,5,19,1,3,4,2
R2_DSer	.	Drosophila serrata	R2	1,3,4,2
R2_DSi	.	Drosophila simulans	R2	18,5,19,1,3,4,2
R2_DYa	.	Drosophila yakuba	R2	18,5,19,1,3,4,2
R2_FA	AF015819	Forficula auricularia	R2	
R2_HC	AF015816	Hippodamia convergens	R2	
R2_KF	GU949558	Kalotermes flavicollis	R2	18,5,19,1,3,4,2
R2_LP	AF015814	Limulus polyphemus	R2	
R2_PS	AF015818	Porcellio scaber	R2	
R2_RL	GU949555	Reticulitermes lucifugus	R2	106,18,104,5,19,1,3,4,2
R2_RU	GU949554	Reticulitermes urbis	R2	106,18,104,5,19,1,3,4,2
R2-1_AAm	.	Amblyomma americanum	R2	1,3,2
R2-1_ACC	.	Aquila chrysaetos canadensis	R2	
R2-1_ACh	.	Acanthisitta chloris	R2	
R2-1_AFo	.	Aptenodytes forsteri	R2	
R2-1_AMi	.	Alligator mississippiensis	R2	106,18,104,19,1,5,3,4,2
R2-1_ApA	.	Apteryx australis mantelli	R2	
R2-1_APi	.	Acyrthosiphon pisum	R2	18,5,1,3,4,2
R2-1_BRG	.	Balearica regulorum gibbericep	R2	
R2-1_BTe	.	Bombus terrestris	R2	18,5,1,19,3,4,2
R2-1_CAnn	.	Calypte anna	R2	
R2-1_CAu	.	Cathartes aura	R2	
R2-1_CBr	.	Corvus brachyrhynchos	R2	
R2-1_CCa	.	Antrostomus carolinensis	R2	
R2-1_CCan	.	Cuculus canorus	R2	
R2-1_CPu	.	Calidris pugnax	R2	
R2-1_Crp	.	Crocodylus porosus	R2	5,1,19,3,4,2
R2-1_CSt	.	Colius striatus	R2	
R2-1_CU	.	Chlamydotis macqueenii	R2	
R2-1_CVo	.	Charadrius vociferus	R2	
R2-1_DWi	.	Drosophila willistoni	R2	18,5,19,1,3,4,2
R2-1_EGa	.	Egretta garzetta	R2	
R2-1_FAl	.	Ficedula albicollis	R2	106,18,104,19,1,5,3,4,2
R2-1_FCh	.	Falco cherrug	R2	
R2-1_FPe	.	Falco peregrinus	R2	
R2-1_GA	.	Gasterosteus aculeatus	R2	18,5,1,19,3,4,2
R2-1_Gav	.	Gavialis gangeticus	R2	18,5,1,19,3,4,2
R2-1_GFo	.	Geospiza fortis	R2	106,18,104,1,19,5,108,3,4,2
R2-1_GSt	.	Gavia stellata	R2	

R2-1_HAl	.	Haliaeetus albicilla	R2	106,18,5,104,19,1,3,4,2
R2-1_IS	.	Ixodes scapularis	R2	18,5,19,1,3,4,143,2
R2-1_LCh	.	Latimeria chalumnae	R2	18,5,19,1,3,4,2
R2-1_LDi	.	Leptosomus discolor	R2	
R2-1_LSal	.	Lepeophtheirus salmonis	R2	
R2-1_LV	.	Lytechinus variegatus	R2	106,18,104,19,1,5,3,4,2
R2-1_MDe	.	Mayetiola destructor	R2	18,5,1,3,4,2
R2-1_MLe	.	Mnemiopsis leidyi	R2	18,5,1,3,4,2
R2-1_MR	.	Megachile rotundata	R2	18,5,1,19,3,4,2
R2-1_MUn	.	Melopsittacus undulatus	R2	
R2-1_MUni	.	Mesitornis unicolor	R2	18,5,104,19,107
R2-1_MVi	.	Manacus vitellinus	R2	
R2-1_NNi	.	Nipponia nippon	R2	18,5,1,19,3,2
R2-1_NV	.	Nematostella vectensis	R2	
R2-1_OHo	.	Opisthocomus hoazin	R2	106,18,5,104,19,1,3,4,2
R2-1_PAd	.	Pygoscelis adeliae	R2	
R2-1_PBa	.	Pogonomyrmex barbatus	R2	1,3,4,2
R2-1_PCar	.	Phalacrocorax carbo	R2	
R2-1_PCau	.	Priapulus caudatus	R2	1,3,4,2
R2-1_PCr	.	Podiceps cristatus	R2	
R2-1_PCri	.	Pelecanus crispus	R2	
R2-1_PGu	.	Pterocles gutturalis	R2	
R2-1_PLe	.	Phaethon lepturus	R2	
R2-1_PM	.	Petromyzon marinus	R2	18,5,1,19,3,4,2
R2-1_PPap	.	Phlebotomus papatasi	R2	18,5,19,1,3,4,2
R2-1_PPu	.	Picoides pubescens	R2	
R2-1_PRR	.	Phoenicopterus ruber ruber	R2	
R2-1_PSi	.	Pelodiscus sinensis	R2	1,3,4,2
R2-1_RMi	.	Rhipicephalus microplus	R2	18,5,1,19,3,4,2
R2-1_RPr	.	Rhodnius prolixus	R2	1,3,4,2
R2-1_RPu	.	Rhipicephalus pulchellus	R2	1,3,4,2
R2-1_SCa	.	Serinus canaria	R2	104,105,2,18,106,1,4,5,19,107,108,3
R2-1_SK	.	Saccoglossus kowalevskii	R2	1,3,4,2
R2-1_SM	.	Schmidtea mediterranea	R2	18,5,1,3,4,2
R2-1_SP	.	Strongylocentrotus purpuratus	R2	106,18,5,104,19,1,3,4,2
R2-1_SSa	AGKD01072455	Salmo salar	R2	18,5,19,1,3,4,2
R2-1_StC	.	Struthio camelus australis	R2	
R2-1_TAl	.	Tyto alba	R2	
R2-1_TCas	.	Tribolium castaneum	R2	18,5,1,3,4,2
R2-1_TG	.	Taeniopygia guttata	R2	18,5,1,19,3,4,2
R2-1_TGut	.	Tinamus guttatus	R2	106,18,104,1,19,5,3,4,2
R2-1_TSP	.	Trichinella spiralis	R2	18,5,19,1,3,4,2
R2-1_TUr	scaffold_6	Tetranychus urticae	R2	1,3,4,2
R2-1_XM	.	Xiphophorus maculatus	R2	18,5,1,19,3,4,2
R2-1_ZA	.	Zonotrichia albicollis	R2	18,19,1,5,108,3,4,2
R2-1_ZLM	.	Zosterops lateralis melanops	R2	18,5,19
R2-2_APi	.	Acyrthosiphon pisum	R2	1,3,4,2
R2-2_CCan	.	Cuculus canorus	R2	
R2-2_CMa	.	Chlamydotis macqueenii	R2	
R2-2_DWi	.	Drosophila willistoni	R2	18,5,1,19,3,4,2
R2-2_HAl	.	Haliaeetus albicilla	R2	

R2-2_IS	.	Ixodes scapularis	R2	1,4,2
R2-2_MR	.	Megachile rotundata	R2	18,5,1,19,3,4,2
R2-2_MUn	.	Melopsittacus undulatus	R2	
R2-2_MUni	.	Mesitornis unicolor	R2	
R2-2_NNi	.	Nipponia nippon	R2	
R2-2_NV	.	Nematostella vectensis	R2	1,3,4,2
R2-2_PBa	.	Pogonomyrmex barbatus	R2	18,5,1,19,3,4,2
R2-2_PM	.	Petromyzon marinus	R2	18,5,19,1,3,4,2
R2-2_RPr	.	Rhodnius prolixus	R2	1,3,4,2
R2-2_SMed	.	Schmidtea mediterranea	R2	18,5,1,19,3,4,2
R2-2_TCas	.	Tribolium castaneum	R2	18,5,1,19,3,4,2
R2-2_TUr	scaffold_37	Tetranychus urticae	R2	18,5,1,3,4,2
R2-3_IS	ABJB010555169	Ixodes scapularis	R2	
R2-3_MR	.	Megachile rotundata	R2	1,3,4,2
R2-4_MR	.	Megachile rotundata	R2	18,5,19,1,3,4,2
R2-5_MR	.	Megachile rotundata	R2	1,3,4,2
R2-6_MR	.	Megachile rotundata	R2	
R2-7_MR	.	Megachile rotundata	R2	18,5,1,19,3,4,2
R2-8_MR	.	Megachile rotundata	R2	18,5,1,19,3,4,2
R2-N1_Gav	.	Gavialis gangeticus	R2	
R2-N2_Gav	.	Gavialis gangeticus	R2	
R2-N2B_Gav	.	Gavialis gangeticus	R2	
R2A_NVi	.	Nasonia vitripennis	R2	18,5,1,19,3,4,2
R2A_TM	AF015817	Tenebrio molitor	R2	
R2Amel	.	Apis mellifera	R2	18,5,19,1,3,4,2
R2B_DM	AF015685	Drosophila mercatorum	R2	
R2B_NVi	.	Nasonia vitripennis	R2	106,18,104,19,1,5,3,4,2
R2B_TM	AF015822	Tenebrio molitor	R2	
R2C_NGi	.	Nasonia giraulti	R2	18,5,1,3,4,2
R2Ci-B	AB097122	Ciona intestinalis	R2	18,5,1,19,3,4,2
R2Ci-D	.	Ciona intestinalis	R2	18,5,19,1,3,4,2
R2CIA_CI	AB097121	Ciona intestinalis	R2	
R2Cs-D	AB097125	Ciona savignyi	R2	
R2D_NGi	.	Nasonia giraulti	R2	
R2Dr	NM_001030097	Danio rerio	R2	18,5,19,1,3,4,2
R2E_NLo	.	Nasonia longicornis	R2	
R2Eb	AB201408	Eptatretus burgeri	R2	1,3,2
R2Ha	AB201415	Hasarius adansoni	R2	1,3,2
R2La	JN937617	Lepidurus arcticus	R2	1,3,4,2
R2LcA	.	Lepidurus couesii	R2	18,5,19,1,3,4,2
R2LcB	JN937619	Lepidurus couesii	R2	18,5,1,19,3,4,2
R2LcC	.	Lepidurus couesii	R2	1,3,2
R2Ll	JN937616	Lepidurus apus lubbocki	R2	18,5,1,19,3,4,2
R2Mr	AB201414	Metacrinus rotundus	R2	
R2NS-1_CGi	.	Crassostrea gigas	R2	18,5,1,19,3,4,2
R2NS-1_CSi	.	Clonorchis sinensis	R2	18,5,19,1,3,4,143,2
R2NS-1_PMi	.	Patiria miniata	R2	18,5,19,1,3,4,2
R2NS-1_SMed	.	Schmidtea mediterranea	R2	18,5,19,1,3,4,2
R2Nvec-A	.	Nematostella vectensis	R2	18,5,1,19,3,4,2
R2Ol-A	.	Oryzias latipes	R2	106,18,104,5,19,1,107,3,117,4,2
R2Pc	AB201416	Procambarus clarkii	R2	1,3,2

R2Sm-A	.	Schistosoma mansoni	R2	18,5,1,19,3,4,2	
R2Ta	AB201409	Tanichthys albonubes	R2	3,1	
R2Tc	EU854578	Triops cancriformis	R2	18,5,1,3,4,2	
R2Tc_it	JN937621	Triops cancriformis	R2	1,3,2	
R2Tl	AB201417	Triops longicaudatus	R2		
R4_AL	U29445	Ascaris lumbricoides	R4	106,104,1,3,4,2	
R4_HC	U29590	Haemonchus contortus	R4		
R4_Hmel	.	Heliconius melpomene	R4	1,3,4,2	
R4-1_AC	.	Anolis carolinensis	R4	1,3,4,2	
R4-1_ADi	.	Acropora digitifera	R4	1,3,4,2	
R4-1_BM	.	Bombyx mori	R4	1,3,4,2	
R4-1_BX	CADV01008175	Bursaphelenchus xylophilus	R4	1,3,4,2	
R4-1_CJap	ABLE03011482	Caenorhabditis japonica	R4	1,3,4,2	
R4-1_CM	.	Callorhinchus milii	R4	2,1	
R4-1_CPB	.	Chrysemys picta bellii	R4		
R4-1_ED	.	Entamoeba dispar	R4	1,3,4,2	
R4-1_HG	.	Heterodera glycines	R4	1,3,4,2	
R4-1_HMe	.	Heliconius melpomene melpomene	R4	1,3,4,2	
R4-1_LVa	.	Litopenaeus vannamei	R4	1,3,4,2	
R4-1_MI	CABB01003843	Meloidogyne incognita	R4	1,3,4,2	
R4-1_PH	.	Parhyale hawaiensis	R4	1,3,4,2	
R4-1_SRa	CACX01002001	Strongyloides ratti	R4	1,3,2	
R4-1_TCa	.	Tribolium castaneum	R4		
196,197					
R4-1B_AC	.	Anolis carolinensis	R4	1,3,4,2	
R4-2_AS	.	Ascaris suum	R4	1,3,4,2	
R4-2_BX	CADV01009048	Bursaphelenchus xylophilus	R4	1,3,4,2	
R4-2_HG	ABLA01000389	Heterodera glycines	R4	1,3,4,2	
R4-2_LVa	.	Litopenaeus vannamei	R4	1,3,4,2	
R4-2_SRa	CACX01002006	Strongyloides ratti	R4	1,3,2	
R4-3_BX	CADV01008832	Bursaphelenchus xylophilus	R4	1,3,4,2	
R4-3_SRa	.	Strongyloides ratti	R4	1,3,2	
R4-4_BX	.	Bursaphelenchus xylophilus	R4	1,3,4,2	
R4-4_SRa	.	Strongyloides ratti	R4	1,3,2	
R4-5_BX	.	Bursaphelenchus xylophilus	R4	1,3,4,2	
R4-N1_ED	.	Entamoeba dispar	SINE		
R5	AY216701	Girardia tigrina	NeSL	1,3,4,2	
R5-1_SM	.	Schmidtea mediterranea	NeSL	1,148,3,149,120,4,2	
R5-2_SM	.	Schmidtea mediterranea	NeSL	147,1,148,3,149,120,4,2	
R6Ag1	AB090817	Anopheles gambiae	R1	2,8,11,86,12,1,14,4,15,16,3	
R6Ag3	AB090819	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3	
R7Ag1	AB090820	Anopheles gambiae	R1	2,8,11,86,12,1,14,4,242,15,16,3	
R7Ag2	AB090821	Anopheles gambiae	R1	2,8,11,86,12,1,14,4,15,16,3	
R8Hm-A	.	Hydra vulgaris	R2	5,1,3,4,2	
R8Hm-B	.	Hydra vulgaris	R2	1,43,3,44,4,2	
R9Av	.	Adineta vaga	R2	1,3,4,2	
Racla	.	Rana clamitans	SINE2/tRNA		
RandI-1	.	Chlamydomonas reinhardtii	RandI	7,1,17,3,10,11,4,6,2	
RandI-1_ACas	AEYA01000779	Acanthamoeba castellanii	RandI	7,1,3,10,11,4,6,2,13	
RandI-1_VC	.	Volvox carteri	RandI	6,2,7,33,10,11,13,1,32,4,237,17,3	
RandI-2	.	Chlamydomonas reinhardtii	RandI	6,2,10,24,11,13,1,4,16,17,3,7	

RandI-2_ACas	AEYA01000457	Acanthamoeba castellanii	RandI	7,1,3,10,11,4,6,2
RandI-3	.	Chlamydomonas reinhardtii	RandI	7,1,17,3,10,11,4,6,2,13
RandI-3_ACas	AEYA01001211	Acanthamoeba castellanii	RandI	7,1,17,3,10,4,6,2,13
RandI-4	.	Chlamydomonas reinhardtii	RandI	7,1,17,3,10,4,11,6,2
RandI-4_ACas	AEYA01002376	Acanthamoeba castellanii	RandI	7,1,3,10,4,6,2,13
RandI-5	.	Chlamydomonas reinhardtii	RandI	
RandI-5_ACas	AEYA01000341	Acanthamoeba castellanii	RandI	7,1,17,3,10,4,2,13
RandI-6	.	Chlamydomonas reinhardtii	RandI	7,1,32,17,3,10,33,4,11,6,2,13
RaR2	FJ461304	Rhynchosciara americana	R2	18,5,1,19,3,4,2
Ras1	.	Rasbora paucipertorata	SINE2/tRNA	
RAVLIN_MT	.	Medicago truncatula	L1	1,3,4,2
RAVLIN_MT	.	Medicago truncatula	L1	1,3,4,2
RAVLIN2_MT	.	Medicago truncatula	L1	43,78,79,44,10,80,81,77,75,17,76,7
RAVLIN2_MT	.	Medicago truncatula	L1	43,78,79,44,10,80,81,77,75,17,76,7
RAVLIN5_MT	AC152406	Medicago truncatula	L1	1,75,8,3,76,77,11,4,12,2
RAVLIN5_MT	AC152406	Medicago truncatula	L1	1,75,8,3,76,77,11,4,12,2
RDRE1_RN	K03501	Rattus norvegicus	SINE1/7SL	
REP-1_ED	.	Entamoeba dispar	SINE	
REP2-2_TT	AY371729	Tetrahymena thermophila	Non-LTRRetrotransposo	
REP6_TT	AY371731	Tetrahymena thermophila	Non-LTRRetrotransposo	
REP6_XT	.	Xenopus tropicalis	Penelope	
REX1_CY	.	Poeciliidae	Rex1	
REX1_FS	.	Fundulus sp. Laguna de Labrado	Rex1	
REX1_FURC	.	Takifugu rubripes	CR1	
Rex1-1_ACar	.	Anolis carolinensis	Rex1	1,40,3,4,41,2
REX1-1_AFC	.	Cichlidae	Rex1	1,40,3,11,41,4,2
REX1-1_BF	.	Branchiostoma floridae	Rex1	1,3,11,4,2
Rex1-1_CCa	.	Cyprinus carpio	Rex1	1,40,3,4,41,11,2
Rex1-1_CTe	.	Capitella teleta	Rex1	1,40,3,4,41,2
REX1-1_DR	.	Dario rerio	Rex1	1,40,3,11,41,4,2
Rex1-1_EL	.	Esox lucius	Rex1	1,40,3,4,41,11,2
Rex1-1_FoCa	.	Folsomia candida	Rex1	1,3,11,2
Rex1-1_FR	.	Takifugu rubripes	Rex1	1,40,3,4,41,2
REX1-1_GA	.	Gasterosteus aculeatus	Rex1	1,40,3,11,41,4,2
Rex1-1_HRo	.	Helobdella robusta	Rex1	1,3,11,4,2
Rex1-1_OL	.	Oryzias latipes	Rex1	1,40,3,11,41,4,2
Rex1-1_PM	.	Petromyzon marinus	Rex1	1,40,3,11,41,4,2
Rex1-1_SP	.	Strongylocentrotus purpuratus	Rex1	16,1,3,24,11,4,2
Rex1-1_SSa	.	Salmo salar	Rex1	1,40,3,11,41,4,2
Rex1-1_XL	.	Xenopus laevis	Rex1	1,40,3,11,41,4,2
REX1-1_XT	.	Xenopus tropicalis	Rex1	1,40,3,4,41,2
Rex1-10_SSa	.	Salmo salar	Rex1	1,40,3,11,41,4,2
Rex1-11_SSa	.	Salmo salar	Rex1	1,40,3,4,41,2
Rex1-15_DRe	.	Danio rerio	Rex1	
Rex1-19_DRe	.	Danio rerio	Rex1	1,3,4,2
Rex1-1B_EL	.	Esox lucius	Rex1	
Rex1-1B_FR	.	Takifugu rubripes	Rex1	1,40,3,4,41,2
REX1-2_AFC	.	Cichlidae	Rex1	1,40,3,11,41,4,2
Rex1-2_CTe	.	Capitella teleta	Rex1	1,40,3,4,41,2
REX1-2_DR	.	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-2_FoCa	.	Folsomia candida	Rex1	1,3,11,2

Rex1-2_FR	Takifugu rubripes	Rex1	1,40,3,4,41,2
Rex1-2_GA	Gasterosteus aculeatus	Rex1	1,40,3,11,41,4,2
Rex1-2_OL	Oryzias latipes	Rex1	1,40,3,4,41,2
Rex1-2_PM	Petromyzon marinus	Rex1	1,40,3,4,41,11,2
Rex1-2_SSa	Salmo salar	Rex1	1,40,3,4,41,11,2
REX1-2_XT	Xenopus tropicalis	Rex1	1,40,3,11,41,4,2
Rex1-20_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-21_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-21_NV	Nematostella vectensis	CR1	41,40
Rex1-23_DRe	Danio rerio	Rex1	1,40,3,4,41,2
Rex1-23_NV	Nematostella vectensis	CR1	1,40,3,11,41,4,2
Rex1-24_NV	Nematostella vectensis	Rex1	1,40,3,4,41,2
Rex1-25_NVe	Nematostella vectensis	Rex1	1,3,11,4,2
Rex1-26_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-28_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-29_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
REX1-3_AFC	Cichlidae	Rex1	1,40,3,11,41,2
Rex1-3_CTe	Capitella teleta	Rex1	1,40,3,11,41,4,2
REX1-3_DR	Danio rerio	Rex1	1,3,4,2
Rex1-3_EL	Esox lucius	Rex1	1,40,3,4,41,2
Rex1-3_FR	Takifugu rubripes	Rex1	1,40,3,11,41,4,2
Rex1-3_GA	Gasterosteus aculeatus	Rex1	1,40,3,4,41,2
Rex1-3_OL	Oryzias latipes	Rex1	1,40,3,4,41,2
Rex1-3_SSa	Salmo salar	Rex1	1,40,3,11,41,4,2
REX1-3_XT	Xenopus tropicalis	Rex1	1,40,3,4,41,11,2
Rex1-30_DRe	Danio rerio	Rex1	1,40,3,4,41,2
Rex1-31_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-38_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-3B_SSa	Salmo salar	Rex1	1,40,3,4,41,11,2
Rex1-3C_SSa	Salmo salar	Rex1	1,40,3,11,41,4,2
Rex1-3D_SSa	Salmo salar	Rex1	1,40,3,11,41,4,2
Rex1-3E_SSa	Salmo salar	Rex1	
REX1-4_AFC	Cichlidae	Rex1	40,41
REX1-4_DR	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-4_EL	Esox lucius	Rex1	1,40,3,11,41,4,2
Rex1-4_OL	Oryzias latipes	Rex1	1,40,3,11,41,4,2
Rex1-4_SSa	Salmo salar	Rex1	1,40,3,11,41,4,2
REX1-4_XT	Xenopus tropicalis	Rex1	1,40,3,11,41,4,2
Rex1-40_DRe	Danio rerio	Rex1	1,40,3,11,41,4,2
Rex1-40B3_DR	Danio rerio	Rex1	
Rex1-43_DRe	Danio rerio	Rex1	1,40,3,41,2
Rex1-5_AFC	Cichlidae	Rex1	40,41
REX1-5_DR	Danio rerio	Rex1	1,40,3,4,41,2
Rex1-5_EL	Esox lucius	Rex1	1,40,3,11,41,4,2
Rex1-5_OL	Oryzias latipes	Rex1	1,40,3,4,41,2
Rex1-5_SSa	Salmo salar	Rex1	1,3,11,4,2
REX1-5_XT	Xenopus tropicalis	Rex1	1,40,3,11,41,4,2
Rex1-50_DR	Danio rerio	Rex1	1,40,3,4,41,2
Rex1-52_DR	Danio rerio	Rex1	
Rex1-53_DR	Danio rerio	Rex1	
Rex1-54_DR	Danio rerio	Rex1	

Rex1-55_DR	.	Danio rerio	Rex1		
Rex1-56_DR	.	Danio rerio	Rex1	1,40,3,41,2	
Rex1-57_DR	.	Danio rerio	Rex1	1,40,3,11,41,4,2	
Rex1-58_DR	.	Danio rerio	Rex1		
Rex1-59_DR	.	Danio rerio	Rex1		
Rex1-6_DR	.	Danio rerio	Rex1	1,40,3,11,41,4,2	
Rex1-6_OL	.	Oryzias latipes	Rex1	1,40,3,11,41,4,2	
Rex1-6_SSa	.	Salmo salar	Rex1	1,40,3,11,41,4,2	
REX1-6_XT	.	Xenopus tropicalis	Rex1		
Rex1-60_DR	.	Danio rerio	Rex1	1,40,3,41,2	
Rex1-61_DR	.	Danio rerio	Rex1		
REX1-7_DR	.	Danio rerio	Rex1	1,40,3,11,41,4,2	
Rex1-7_OL	.	Oryzias latipes	Rex1	1,40,3,11,41,4,2	
Rex1-7_SSa	.	Salmo salar	Rex1	1,40,3,11,41,4,2	
REX1-7_XT	.	Xenopus tropicalis	Rex1		
REX1-8_DR	.	Danio rerio	Rex1	1,40,3,11,41,4,2	
Rex1-8_SSa	.	Salmo salar	Rex1	1,40,3,4,41,2	
REX1-8_XT	.	Xenopus tropicalis	Rex1		
REX1-9_DR	.	Danio rerio	Rex1		
Rex1-9_SSa	.	Salmo salar	Rex1	1,40,113,3,4,41,2	
REX1-9_XT	.	Xenopus tropicalis	Rex1	1,40,3,11,41,4,2	
REX3	AF125983	Xiphophorus maculatus	RTE	1,3,4,2	
REX5	AY228504	Xiphophorus maculatus	L2	1,3,11,4,2	
Rex6	.	Takifugu rubripes	R4	1,3,4,2	
Rex6-1_OL	.	Oryzias latipes	R4	1,3,4,2	
Rhin-1	AC150242.3, AC152492	Rhinolophus ferrumequinum	SINE2/tRNA		
RILN1_OA	AB016096	Oryza australiensis	L1		
RILN3_OA	AB016098	Oryza australiensis	L1		
RILN4_OS	AB016099	Oryza sativa	L1		
RILN6_OS	AB016101	Oryza sativa	L1		
RILN7_OS	AB016102	Oryza sativa	L1	43,2,44,11,12,1,14,4,3	
RILN8_OS	.	Oryza sativa	L1		
RMER1B	.	Murinae	Non-LTRRetrotransposo		
RNALUIII	M13501	Rattus norvegicus	SINE1/7SL		
RSG1	M37214	Oncorhynchus mykiss	SINE		
RSINE1	.	Rodentia	SINE2/tRNA		
RSINE2	.	Rodentia	SINE2/tRNA		
RSINE2A	.	Rodentia	SINE2/tRNA		
RT_BC	AJ430549	Brassica carinata	Non-LTRRetrotransposo		
RT1	M93690	Anopheles gambiae	R1	2,245,8,11,86,12,246,1,14,4,15,16,3	
RT2	M93691	Anopheles gambiae	R1	2,8,11,12,1,14,4,243,3	
RTAg3	AB090812	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3	
RTAg4	AB090813	Anopheles gambiae	R1	43,2,8,44,11,86,12,42,1,14,4,15,16,3	
RTE_BOV_B_AC_1	.	Anolis carolinensis	RTE	16,1,26,3,4,11,2,15	
RTE_Ele2	.	Aedes aegypti	RTE	1,26,3,4,11,2	
RTE_Ele2B_AAe	.	Aedes aegypti	RTE	1,26,3,11,4,2	
RTE_Ele2C_AAe	.	Aedes aegypti	RTE	1,26,3,11,4,2	
RTE_Ele3	.	Aedes aegypti	RTE	16,1,26,3,11,4,2,15	
RTE_Ele4	.	Aedes aegypti	RTE	1,26,3,4,11,2	
RTE_Ele5	.	Aedes aegypti	RTE	1,26,3,11,4,2	
RTE_SJ	AY027869	Schistosoma japonicum	RTE		

RTE-1	AF054983	Caenorhabditis elegans	RTE	16,1,26,3,24,11,4,2
RTE-1_AC	.	Aplysia californica	RTE	16,1,26,3,24,11,4,2
RTE-1_AC_1	.	Anolis carolinensis	RTE	18,5,1,19,3,11,4,2
RTE-1_AFC	.	Cichlidae	RTE	16,1,26,3,11,4,2,15
RTE-1_AG	.	Anopheles gambiae	RTE	16,1,26,3,24,11,4,2
RTE-1_AIp	.	Arachis ipaensis	RTE	65,1,26,3,11,4,2
RTE-1_ATr	.	Amborella trichopoda	RTE	1,3,11,4,2
RTE-1_BDi	.	Brachypodium distachyon	RTE	1,3,4,2
RTE-1_BF	.	Branchiostoma floridae	RTE	1,127,3,59,60,4,56,57,2,68
RTE-1_BM	.	Bombyx mori	RTE	1,3,4,2
RTE-1_BTa	.	Bemisia tabaci	RTE	1,26,3,4,11,2
RTE-1_CBre	.	Caenorhabditis brenneri	RTE	1,26,3,11,4,2
RTE-1_CJap	.	Caenorhabditis japonica	RTE	16,1,26,3,11,4,2,15
RTE-1_CPB	.	Chrysemys picta bellii	RTE	18,5,16,1,26,3,11,4,2,15
RTE-1_CQ	.	Culex quinquefasciatus	RTE	1,26,3,4,11,2
RTE-1_CRem	.	Caenorhabditis remanei	RTE	1,26,3,11,4,2
RTE-1_Crp	.	Crocodylus porosus	RTE	18,5,16,19,1,26,3,24,11,4,2
RTE-1_CTe	.	Capitella teleta	RTE	1,3,11,4,2
RTE-1_DaCa	.	Daucus carota	RTE	1,3,11,4,2
RTE-1_DAn	.	Drosophila ananassae	RTE	1,26,3,4,11,2
RTE-1_DBi	.	Drosophila biarmipes	RTE	1,3,4,2
RTE-1_DBp	.	Drosophila bipectinata	RTE	1,26,3,11,4,2
RTE-1_DCa	.	Dendrobium catenatum	RTE	65,1,26,3,11,4,2
RTE-1_DF	.	Drosophila ficusphila	RTE	1,3,11,4,2
RTE-1_DPl	.	Danaus plexippus	RTE	1,26,3,11,4,2
RTE-1_DR	.	Danio rerio	RTE	16,1,26,3,11,4,2,15
RTE-1_DSa	.	Drosophila santomea	RTE	1,26,3,11,4,2
RTE-1_DSer	.	Drosophila serrata	RTE	1,26,3,4,11,2
RTE-1_DT	.	Drosophila takahashii	RTE	1,26,3,11,4,2
RTE-1_DYa	.	Drosophila yakuba	RTE	1,3,4,2
RTE-1_EC	.	Equus caballus	RTE	1,3,11,4,2
RTE-1_EL	.	Esox lucius	RTE	16,1,26,49,3,11,4,50,2,15
RTE-1_GA	.	Gasterosteus aculeatus	RTE	16,1,114,26,3,4,115,11,2,15
RTE-1_Gav	.	Gavialis gangeticus	RTE	18,5,1,19,3,11,4,2
RTE-1_GM	.	Glycine max	RTE	65,1,26,3,11,4,2
RTE-1_HMe	.	Heliconius melpomene melpomene	RTE	16,1,26,3,4,11,2,15
RTE-1_HRo	.	Helobdella robusta	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-1_JC	.	Jatropha curcas	RTE	
RTE-1_LVa	.	Litopenaeus vannamei	RTE	1,26,3,11,4,2
RTE-1_MAc	.	Musa acuminata	RTE	65,1,26,3,11,4,2
RTE-1_Mad	.	Malus domestica	RTE	65,1,26,3,11,4,2
RTE-1_MD	.	Monodelphis domestica	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-1_MLe	.	Mnemiopsis leidyi	RTE	1,3,11,4,2
RTE-1_MMa	.	Mesobuthus martensii	RTE	1,3,4,2
RTE-1_MN	.	Morus notabilis	RTE	65,1,26,3,11,4,2
RTE-1_NN	.	Nelumbo nucifera	RTE	1,3,4,2
RTE-1_NS	.	Nicotiana sylvestris	RTE	1,26,3,11,4,2
RTE-1_NV	.	Nematostella vectensis	RTE	
RTE-1_NVi	.	Nasonia vitripennis	RTE	1,3,4,2
RTE-1_OAf	.	Orycteropus afer	RTE	16,1,26,3,4,11,2,15
RTE-1_OL	.	Oryzias latipes	RTE	1,26,16,49,3,11,4,50,2,15

RTE-1_PBa	.	Pogonomyrmex barbatus	RTE	
RTE-1_PD	.	Phoenix dactylifera	RTE	1,3,4,2
RTE-1_PH	.	Parhyale hawaiensis	RTE	16,1,26,3,11,4,2,15
RTE-1_PM	.	Petromyzon marinus	RTE	16,1,26,3,11,4,2,15
RTE-1_PPac	.	Pristionchus pacificus	RTE	16,1,26,3,24,4,11,2
RTE-1_PPo	.	Papilio polytes	RTE	1,3,2
RTE-1_PSi	.	Pelodiscus sinensis	RTE	18,5,16,19,1,26,3,4,11,2,15
RTE-1_PXu	.	Papilio xuthus	RTE	
RTE-1_RC	.	Ricinus communis	RTE	
RTE-1_SBi	.	Sorghum bicolor	RTE	1,3,4,2
RTE-1_SIn	.	Solenopsis invicta	RTE	16,1,3,4,11,2,15
RTE-1_SIt	.	Setaria italica	RTE	65,16,1,26,3,24,11,4,2
RTE-1_SK	.	Saccoglossus kowalevskii	RTE	1,3,59,60,4,2
RTE-1_SSa	.	Salmo salar	RTE	2,18,24,11,1,26,4,65,5,16,19,3
RTE-1_STu	.	Solanum tuberosum	RTE	1,26,3,11,4,2
RTE-1_TD	TREP1820	Triticum turgidum subsp. durum	RTE	65,16,1,26,3,24,11,4,2
RTE-1_TML	.	Trichechus manatus latirostris	RTE	16,1,26,3,24,11,4,2
RTE-1_TP	.	Thalassiosira pseudonana	RTE	5,1,3,11,2
RTE-10_AMi	.	Alligator mississippiensis	RTE	1,26,3,11,4,2
RTE-10_BF	.	Branchiostoma floridae	RTE	16,1,26,3,11,4,2,15
RTE-10_CPB	.	Chrysemys picta bellii	RTE	
RTE-10_Crp	.	Crocodylus porosus	RTE	
RTE-10_Hmel	.	Heliconius melpomene	RTE	1,3,4,2
RTE-10_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-10_PPac	.	Pristionchus pacificus	RTE	16,1,3,11,4,2,15
RTE-10_PPo	.	Papilio polytes	RTE	1,3,4,2
RTE-10_PSi	.	Pelodiscus sinensis	RTE	16,1,26,3,11,4,2,15
RTE-10_PXu	.	Papilio xuthus	RTE	
RTE-10_SP	.	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-10B_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,4,11,2,15
RTE-11_AMi	.	Alligator mississippiensis	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-11_BF	.	Branchiostoma floridae	RTE	1,26,16,3,11,4,2,15
RTE-11_CPB	.	Chrysemys picta bellii	RTE	1,26,3,11,4,2
RTE-11_Crp	.	Crocodylus porosus	RTE	
RTE-11_Hmel	.	Heliconius melpomene	RTE	1,3,4,2
RTE-11_HMM	.	Heliconius melpomene melpomene	RTE	16,1,26,3,11,4,2,15
RTE-11_LCh	.	Latimeria chalumnae	RTE	16,1,26,3,11,4,2,15
RTE-11_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-11_PPac	.	Pristionchus pacificus	RTE	1,3,4,2
RTE-11_PPo	.	Papilio polytes	RTE	
RTE-11_PSi	.	Pelodiscus sinensis	RTE	18,5,16,19,1,26,3,24,4,11,2
RTE-11_PXu	.	Papilio xuthus	RTE	
RTE-11_SP	.	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-12_AMi	.	Alligator mississippiensis	RTE	
RTE-12_BF	.	Branchiostoma floridae	RTE	16,1,26,3,24,11,4,2
RTE-12_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15
RTE-12_Crp	.	Crocodylus porosus	RTE	16,1,26,3,24,11,4,2
RTE-12_Hmel	.	Heliconius melpomene	RTE	1,3,2
RTE-12_HMM	.	Heliconius melpomene melpomene	RTE	1,3,4,2
RTE-12_LCh	.	Latimeria chalumnae	RTE	18,5,16,1,19,3,11,4,2,15
RTE-12_LMi	.	Locusta migratoria	RTE	1,26,3,11,4,2

RTE-12_PPac	.	Pristionchus pacificus	RTE	16,1,26,3,24,11,4,2
RTE-12_PPo	.	Papilio polytes	RTE	
RTE-12_PSi	.	Pelodiscus sinensis	RTE	65,16,1,26,3,24,11,4,2
RTE-12_PXu	.	Papilio xuthus	RTE	
RTE-12_SP	.	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-13_AMi	.	Alligator mississippiensis	RTE	
RTE-13_BF	.	Branchiostoma floridae	RTE	16,1,26,3,4,11,2,15
RTE-13_CPB	.	Chrysemys picta bellii	RTE	65,16,1,26,3,11,4,2,15
RTE-13_Hmel	.	Heliconius melpomene	RTE	1,3,4,2
RTE-13_Lch	.	Latimeria chalumnae	RTE	65,1,16,26,3,4,11,2,15
RTE-13_LMi	.	Locusta migratoria	RTE	1,3,172,173,4,2
RTE-13_PPac	.	Pristionchus pacificus	RTE	170,1,3,4,171,2
RTE-13_PPo	.	Papilio polytes	RTE	
RTE-13_PSi	.	Pelodiscus sinensis	RTE	65,16,1,26,3,11,4,2,15
RTE-13_PXu	.	Papilio xuthus	RTE	
RTE-13_SP	.	Strongylocentrotus purpuratus	RTE	65,16,26,11,15
RTE-14_BF	.	Branchiostoma floridae	RTE	1,26,3,4,11,2
RTE-14_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,4,11,2,15
RTE-14_Hmel	.	Heliconius melpomene	RTE	1,3,2
RTE-14_Lch	.	Latimeria chalumnae	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-14_LMi	.	Locusta migratoria	RTE	16,1,26,3,4,11,2,15
RTE-14_PPac	.	Pristionchus pacificus	RTE	1,3,11,4,2
RTE-14_PPo	.	Papilio polytes	RTE	1,3,4,2
RTE-14_PSi	.	Pelodiscus sinensis	RTE	5,16,1,19,26,3,4,11,2,15
RTE-14_SP	.	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-15_BF	.	Branchiostoma floridae	RTE	16,1,26,3,11,4,2,15
RTE-15_CPB	.	Chrysemys picta bellii	RTE	
RTE-15_Hmel	.	Heliconius melpomene	RTE	1,3,4,2
RTE-15_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-15_PPo	.	Papilio polytes	RTE	1,3,4,2
RTE-15_PSi	.	Pelodiscus sinensis	RTE	1,3,11,4,2
RTE-15_PXu	.	Papilio xuthus	RTE	
RTE-15_SP	.	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-16_BF	.	Branchiostoma floridae	RTE	1,26,3,11,4,2
RTE-16_CPB	.	Chrysemys picta bellii	RTE	
RTE-16_Hmel	.	Heliconius melpomene	RTE	16,1,26,3,11,4,2,15
RTE-16_LMi	.	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-16_PPo	.	Papilio polytes	RTE	
RTE-16_PSi	.	Pelodiscus sinensis	RTE	16,1,26,3,11,4,2,15
RTE-16_SP	.	Strongylocentrotus purpuratus	RTE	1,3,2
RTE-17_BF	.	Branchiostoma floridae	RTE	5,1,19,4,2
RTE-17_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15
RTE-17_Hmel	.	Heliconius melpomene	RTE	18,5,19
RTE-17_LMi	.	Locusta migratoria	RTE	16,1,26,3,24,11,4,2
RTE-17_PPo	.	Papilio polytes	RTE	
RTE-17_PSi	.	Pelodiscus sinensis	RTE	1,26,16,3,11,4,2,15
RTE-17B_PPo	.	Papilio polytes	RTE	
RTE-18_BF	.	Branchiostoma floridae	RTE	1,3,11,4,2
RTE-18_CPB	.	Chrysemys picta bellii	RTE	
RTE-18_Hmel	.	Heliconius melpomene	RTE	65,16,1,26,3,4,11,2,15
RTE-18_HMM	.	Heliconius melpomene melpomene	RTE	1,3,2

RTE-18_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-18_PPo	.	Papilio polytes	RTE	
RTE-18_SP	.	Strongylocentrotus purpuratus	RTE	1,26,3,11,4,2
RTE-18B_PPo	.	Papilio polytes	RTE	
RTE-19_BF	.	Branchiostoma floridae	RTE	1,3,4,2
RTE-19_CPB	.	Chrysemys picta bellii	RTE	
RTE-19_Hmel	.	Heliconius melpomene	RTE	1,3,4,2
RTE-19_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-19_PPo	.	Papilio polytes	RTE	16,1,26,3,11,4,2,15
RTE-1B_DAn	.	Drosophila ananassae	RTE	1,26,3,11,4,2
RTE-1B_DCa	.	Dendrobium catenatum	RTE	1,3,4,2
RTE-1B_DSa	.	Drosophila santomea	RTE	26,11
RTE-1B_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-1B_Mad	.	Malus domestica	RTE	65,1,26,3,11,4,2
RTE-1B_PPo	.	Papilio polytes	RTE	
RTE-1BN1_DAn	.	Drosophila ananassae	RTE	
RTE-1C_DAn	.	Drosophila ananassae	RTE	1,26,16,3,11,4,2,15
RTE-1C_LMi	.	Locusta migratoria	RTE	
RTE-2_AC	.	Aplysia californica	RTE	1,26,16,3,24,11,4,2
RTE-2_AFC	.	Cichlidae	RTE	1,26,3,11,4,2
RTE-2_AG	.	Anopheles gambiae	RTE	1,26,3,4,11,2
RTE-2_BF	.	Branchiostoma floridae	RTE	16,1,26,3,4,11,2,15
RTE-2_BM	.	Bombyx mori	RTE	
RTE-2_BTa	.	Bemisia tabaci	RTE	1,26,3,11,4,2
RTE-2_CBre	.	Caenorhabditis brenneri	RTE	16,1,26,3,4,11,2,15
RTE-2_CJap	.	Caenorhabditis japonica	RTE	1,26,3,11,4,2
RTE-2_CPB	.	Chrysemys picta bellii	RTE	18,5,16,19,1,26,3,24,4,11,2
RTE-2_CQ	.	Culex quinquefasciatus	RTE	16,1,26,3,24,11,4,2
RTE-2_Croc	.	Crocodylidae	RTE	18,5,1,19,3,11,4,2
RTE-2_CTe	.	Capitella teleta	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-2_DAn	.	Drosophila ananassae	RTE	1,26,3,11,4,2
RTE-2_DBp	.	Drosophila bipectinata	RTE	1,26,3,11,4,2
RTE-2_DF	.	Drosophila ficusphila	RTE	1,26,3,11,4,2
RTE-2_DPl	.	Danaus plexippus	RTE	1,26,3,11,4,2
RTE-2_DR	.	Danio rerio	RTE	1,2
RTE-2_EL	.	Esox lucius	RTE	1,49,3,4,50,2
RTE-2_Gav	.	Gavialis gangeticus	RTE	1,3,4,2
RTE-2_Hmel	.	Heliconius melpomene	RTE	16,1,26,3,11,4,2,15
RTE-2_HRo	.	Helobdella robusta	RTE	2,11,57,68,1,26,4,3,59,60
RTE-2_LCh	.	Latimeria chalumnae	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-2_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-2_LVa	.	Litopenaeus vannamei	RTE	16,1,26,3,11,4,2,15
RTE-2_MD	.	Monodelphis domestica	RTE	1,3,4,2
RTE-2_ME	.	Notamacropus eugenii	RTE	1,26,3,11,4,2
RTE-2_MLe	.	Mnemiopsis leidyi	RTE	1,3,4,2
RTE-2_MMa	.	Mesobuthus martensii	RTE	16,1,26,3,11,4,2,15
RTE-2_NS	.	Nicotiana sylvestris	RTE	1,3,11,4,2
RTE-2_NV	.	Nematostella vectensis	RTE	
RTE-2_NVi	.	Nasonia vitripennis	RTE	16,1,3,11,15,4,2
RTE-2_OL	.	Oryzias latipes	RTE	1,26,3,11,4,2
RTE-2_PD	.	Phoenix dactylifera	RTE	

RTE-2_PH	.	Parhyale hawaiensis	RTE	65,16,1,26,3,24,11,4,2
RTE-2_PM	.	Petromyzon marinus	RTE	16,1,26,3,11,4,2,15
RTE-2_PPac	.	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-2_PPo	.	Papilio polytes	RTE	
RTE-2_PSi	.	Pelodiscus sinensis	RTE	16,1,26,3,11,4,2,15
RTE-2_PXu	.	Papilio xuthus	RTE	
RTE-2_SIn	.	Solenopsis invicta	RTE	1,3,4,11,2
RTE-2_SK	.	Saccoglossus kowalevskii	RTE	65,16,1,26,3,11,4,2,15
RTE-2_SP	.	Strongylocentrotus purpuratus	RTE	65,16,1,26,3,11,4,2,15
RTE-2_STu	.	Solanum tuberosum	RTE	
RTE-20_CPB	.	Chrysemys picta bellii	RTE	
RTE-20_Hmel	.	Heliconius melpomene	RTE	16,1,26,3,11,4,2,15
RTE-20_HMM	.	Heliconius melpomene melpomene	RTE	1,3,2
RTE-20_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-20_PPo	.	Papilio polytes	RTE	
RTE-21_Croc	.	Crocodylidae	RTE	5,16,1,19,26,3,11,4,2,15
RTE-21_LMi	.	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-21_PPo	.	Papilio polytes	RTE	1,3,11,4,2
RTE-22_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-22_PPo	.	Papilio polytes	RTE	
RTE-22B_PPo	.	Papilio polytes	RTE	
RTE-23_CPB	.	Chrysemys picta bellii	RTE	18,66,16,5,19,1,26,67,3,11,4,2,15
RTE-23_LMi	.	Locusta migratoria	RTE	16,1,3,11,4,2,15
RTE-24_CPB	.	Chrysemys picta bellii	RTE	
RTE-24_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-24_PPo	.	Papilio polytes	RTE	
RTE-25_CPB	.	Chrysemys picta bellii	RTE	
RTE-25_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-25_PPo	.	Papilio polytes	RTE	
RTE-25B_PPo	.	Papilio polytes	RTE	
RTE-26_CPB	.	Chrysemys picta bellii	RTE	
RTE-26_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-26_PPo	.	Papilio polytes	RTE	
RTE-27_CPB	.	Chrysemys picta bellii	RTE	1,3,4,2
RTE-27_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-27_PPo	.	Papilio polytes	RTE	
RTE-27B_CPB	.	Chrysemys picta bellii	RTE	
RTE-28_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-28_PPo	.	Papilio polytes	RTE	
RTE-29_CPB	.	Chrysemys picta bellii	RTE	1,3,4,2
RTE-29_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-29_PPo	.	Papilio polytes	RTE	
RTE-2B_PPo	.	Papilio polytes	RTE	
RTE-3_AC	.	Aplysia californica	RTE	16,1,26,3,4,11,2,15
RTE-3_AG	.	Anopheles gambiae	RTE	1,3,4,2
RTE-3_AMi	.	Alligator mississippiensis	RTE	18,5,16,1,19,3,11,4,2,15
RTE-3_BF	.	Branchiostoma floridae	RTE	16,1,26,3,4,11,2,15
RTE-3_BM	.	Bombyx mori	RTE	1,3,4,2
RTE-3_BTa	.	Bemisia tabaci	RTE	16,1,26,3,24,11,4,2
RTE-3_CBre	.	Caenorhabditis brenneri	RTE	
RTE-3_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15

RTE-3_CQ	.	Culex quinquefasciatus	RTE	1,26,3,11,4,2
RTE-3_Crp	.	Crocodylus porosus	RTE	18,5,16,19,1,26,3,24,4,11,2
RTE-3_CTe	.	Capitella teleta	RTE	16,1,26,3,24,11,4,2
RTE-3_DF	.	Drosophila ficusphila	RTE	1,26,3,11,4,2
RTE-3_DPl	.	Danaus plexippus	RTE	16,1,26,3,11,4,2,15
RTE-3_DR	.	Danio rerio	RTE	
RTE-3_Hmel	.	Heliconius melpomene	RTE	1,26,3,11,4,2
RTE-3_HRo	.	Helobdella robusta	RTE	2,11,57,68,1,26,4,3,59,60
RTE-3_LCh	.	Latimeria chalumnae	RTE	18,5,16,1,26,3,24,11,4,2
RTE-3_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-3_LVa	.	Litopenaeus vannamei	RTE	1,26,3,11,4,2
RTE-3_MD	.	Monodelphis domestica	RTE	
RTE-3_ME	.	Notamacropus eugenii	RTE	
RTE-3_MMa	.	Mesobuthus martensii	RTE	16,1,26,3,11,4,2,15
RTE-3_NS	.	Nicotiana sylvestris	RTE	
RTE-3_NV	.	Nematostella vectensis	RTE	
RTE-3_OL	.	Oryzias latipes	RTE	1,26,3,11,4,2
RTE-3_PM	.	Petromyzon marinus	RTE	16,1,26,3,11,4,2,15
RTE-3_PPac	.	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-3_PSi	.	Pelodiscus sinensis	RTE	
RTE-3_PXu	.	Papilio xuthus	RTE	
RTE-3_SIn	.	Solenopsis invicta	RTE	16,1,3,11,4,2,15
RTE-3_SP	048B05	Strongylocentrotus purpuratus	RTE	1,3,11,4,2
RTE-30_CPB	.	Chrysemys picta bellii	RTE	
RTE-30_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-31_CPB	.	Chrysemys picta bellii	RTE	
RTE-31_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-31_PPo	.	Papilio polytes	RTE	
RTE-32_CPB	.	Chrysemys picta bellii	RTE	
RTE-32_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-32_PPo	.	Papilio polytes	RTE	
RTE-33_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15
RTE-33_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-33_PPo	.	Papilio polytes	RTE	
RTE-34_CPB	.	Chrysemys picta bellii	RTE	1,3,11,4,2
RTE-34_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-34_PPo	.	Papilio polytes	RTE	
RTE-35_CPB	.	Chrysemys picta bellii	RTE	
RTE-35_PPo	.	Papilio polytes	RTE	
RTE-36_CPB	.	Chrysemys picta bellii	RTE	
RTE-36_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-36_PPo	.	Papilio polytes	RTE	
RTE-37_CPB	.	Chrysemys picta bellii	RTE	
RTE-37_LMi	.	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-37_PPo	.	Papilio polytes	RTE	
RTE-38_LMi	.	Locusta migratoria	RTE	1,3,2
RTE-38_PPo	.	Papilio polytes	RTE	
RTE-39_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-3B_SP	048B05	Strongylocentrotus purpuratus	RTE	1,26,3,11,4,2
RTE-4_AC	.	Aplysia californica	RTE	1,26,3,11,4,2
RTE-4_AMi	.	Crocodylidae	RTE	16,1,26,3,11,4,2,15

RTE-4_BF	.	Branchiostoma floridae	RTE	16,1,26,3,11,4,2,15
RTE-4_BM	.	Bombyx mori	RTE	1,3,11,4,2
RTE-4_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,11,4,2,15
RTE-4_Croc	.	Crocodylidae	RTE	
RTE-4_CTe	.	Capitella teleta	RTE	1,3,4,2
RTE-4_DF	.	Drosophila ficusphila	RTE	1,26,3,11,4,2
RTE-4_DPl	.	Danaus plexippus	RTE	65,16,1,26,3,11,4,2,15
RTE-4_DR	.	Danio rerio	RTE	
RTE-4_Hmel	.	Heliconius melpomene	RTE	16,1,26,3,11,4,2,15
RTE-4_HRo	.	Helobdella robusta	RTE	53,2,54,11,57,68,1,4,3,59,60
RTE-4_LCh	.	Latimeria chalumnae	RTE	16,1,26,3,11,4,2,15
RTE-4_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-4_LVa	.	Litopenaeus vannamei	RTE	1,3,4,2
RTE-4_MMa	.	Mesobuthus martensii	RTE	16,1,26,3,11,4,2,15
RTE-4_NV	.	Nematostella vectensis	RTE	
RTE-4_PM	.	Petromyzon marinus	RTE	1,16,26,3,11,4,2,15
RTE-4_PPac	.	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-4_PPo	.	Papilio polytes	RTE	16,1,26,3,11,4,2,15
RTE-4_PSi	.	Pelodiscus sinensis	RTE	65,1,26,3,11,4,2
RTE-4_PXu	.	Papilio xuthus	RTE	
RTE-4_SP	048B05	Strongylocentrotus purpuratus	RTE	1,3,11,4,2
RTE-40_LMi	.	Locusta migratoria	RTE	1,3,4,2
RTE-41_LMi	.	Locusta migratoria	RTE	16,1,3,11,4,2,15
RTE-42_LMi	.	Locusta migratoria	RTE	65,16,1,26,3,4,11,2,15
RTE-43_LMi	.	Locusta migratoria	RTE	16,1,3,11,4,2,15
RTE-44_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-45_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-4562_AMi	.	Crocodylidae	RTE	
RTE-46_LMi	.	Locusta migratoria	RTE	16,1,26,3,4,11,2,15
RTE-47_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-48_LMi	.	Locusta migratoria	RTE	
RTE-49_LMi	.	Locusta migratoria	RTE	
RTE-4B_PPo	.	Papilio polytes	RTE	
RTE-5_AC	.	Aplysia californica	RTE	65,16,1,26,3,24,11,4,2
RTE-5_AMi	.	Alligator mississippiensis	RTE	16,1,3,11,4,2,15
RTE-5_BF	.	Branchiostoma floridae	RTE	1,3,4,2
RTE-5_BM	.	Bombyx mori	RTE	65,1,26,11,2
RTE-5_CPB	.	Chrysemys picta bellii	RTE	1,26,3,11,4,2
RTE-5_CTe	.	Capitella teleta	RTE	1,3,4,2
RTE-5_DPl	.	Danaus plexippus	RTE	16,1,3,11,4,2,15
RTE-5_Hmel	.	Heliconius melpomene	RTE	18,5,16,19,1,26,3,24,11,4,2
RTE-5_HRo	.	Helobdella robusta	RTE	16,1,26,3,24,11,4,2
RTE-5_LCh	.	Latimeria chalumnae	RTE	18,5,1,3,11,4,2
RTE-5_LMi	.	Locusta migratoria	RTE	65,16,1,26,3,11,4,2,15
RTE-5_LVa	.	Litopenaeus vannamei	RTE	
RTE-5_MMa	.	Mesobuthus martensii	RTE	65,16,1,26,3,4,11,2,15
RTE-5_NV	.	Nematostella vectensis	RTE	
RTE-5_PM	.	Petromyzon marinus	RTE	5,1,19,26,11,2
RTE-5_PPac	.	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-5_PPo	.	Papilio polytes	RTE	
RTE-5_PSi	.	Pelodiscus sinensis	RTE	18,5,16,19,1,26,3,11,4,2,15

RTE-5_PXu	Papilio xuthus	RTE	
RTE-5_SP	Strongylocentrotus purpuratus	RTEX	21,1,3,11,4,2
RTE-50_LMi	Locusta migratoria	RTE	1,3,11,4,2
RTE-51_LMi	Locusta migratoria	RTE	65,1,3,4,11,2
RTE-52_LMi	Locusta migratoria	RTE	1,3,11,4,2
RTE-53_LMi	Locusta migratoria	RTE	65,16,1,26,3,11,4,2,15
RTE-54_AMi	Crocodylidae	RTE	
RTE-54_LMi	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-55_LMi	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-56_LMi	Locusta migratoria	RTE	1,26,3,4,11,2
RTE-57_LMi	Locusta migratoria	RTE	
RTE-58_LMi	Locusta migratoria	RTE	
RTE-59_LMi	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-6_AC	Aplysia californica	RTE	16,1,26,3,11,4,2,15
RTE-6_AMi	Alligator mississippiensis	RTE	1,26,3,4,11,2
RTE-6_BF	Branchiostoma floridae	RTE	16,1,26,3,11,4,2,15
RTE-6_BM	Bombyx mori	RTE	1,3,4,2
RTE-6_CPB	Chrysemys picta bellii	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-6_Crp	Crocodylus porosus	RTE	
RTE-6_DPl	Danaus plexippus	RTE	1,11,2
RTE-6_Hmel	Heliconius melpomene	RTE	1,3,4,2
RTE-6_HMM	Heliconius melpomene melpomene	RTE	16,1,26,3,11,4,2,15
RTE-6_HRo	Helobdella robusta	RTE	65,16,1,26,3,4,11,2,15
RTE-6_LCh	Latimeria chalumnae	RTE	1,26,3,11,4,2
RTE-6_LMi	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-6_LVa	Litopenaeus vannamei	RTE	
RTE-6_NV	Nematostella vectensis	RTE	1,3,2
RTE-6_PM	Petromyzon marinus	RTE	16,1,26,3,11,4,2,15
RTE-6_PPac	Pristionchus pacificus	RTE	1,3,2
RTE-6_PPo	Papilio polytes	RTE	18,5,16,19,1,26,3,24,11,4,2
RTE-6_PSi	Pelodiscus sinensis	RTE	
RTE-6_PXu	Papilio xuthus	RTE	16,1,26,3,11,15,4,2
RTE-6_SP	Strongylocentrotus purpuratus	RTE	100,16,1,3,24,4,11,2
RTE-60_LMi	Locusta migratoria	RTE	
RTE-61_LMi	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-62_LMi	Locusta migratoria	RTE	
RTE-63_LMi	Locusta migratoria	RTE	
RTE-64_LMi	Locusta migratoria	RTE	1,3,11,4,2
RTE-65_LMi	Locusta migratoria	RTE	1,3,11,4,2
RTE-66_LMi	Locusta migratoria	RTE	
RTE-67_LMi	Locusta migratoria	RTE	
RTE-68_LMi	Locusta migratoria	RTE	16,1,3,11,4,2,15
RTE-69_LMi	Locusta migratoria	RTE	
RTE-6B_HRo	Helobdella robusta	RTE	65,16,1,26,3,11,4,2,15
RTE-7_AC	Aplysia californica	RTE	1,3,11,4,2
RTE-7_AMi	Alligator mississippiensis	RTE	16,1,26,3,11,4,2,15
RTE-7_BF	Branchiostoma floridae	RTE	
RTE-7_CPB	Chrysemys picta bellii	RTE	18,5,1,19,3,4,2
RTE-7_Crp	Crocodylus porosus	RTE	
RTE-7_DPl	Danaus plexippus	RTE	1,3,4,2
RTE-7_Hmel	Heliconius melpomene	RTE	16,1,3,11,4,2,15

RTE-7_HRo	.	Helobdella robusta	RTE	1,3,4,2
RTE-7_LCh	.	Latimeria chalumnae	RTE	18,5,16,19,1,26,3,11,4,2,15
RTE-7_LMi	.	Locusta migratoria	RTE	16,1,26,3,11,4,2,15
RTE-7_LVa	.	Litopenaeus vannamei	RTE	1,3,4,11,2
RTE-7_NV	.	Nematostella vectensis	RTE	65,16,1,26,3,11,4,2,15
RTE-7_PM	.	Petromyzon marinus	RTE	16,1,3,11,4,2,15
RTE-7_PPac	.	Pristionchus pacificus	RTE	1,3,4,2
RTE-7_PPo	.	Papilio polytes	RTE	
RTE-7_PSi	.	Pelodiscus sinensis	RTE	1,3,4,2
RTE-7_PXu	.	Papilio xuthus	RTE	
RTE-7_SP	.	Strongylocentrotus purpuratus	RTE	65,1,16,26,3,4,11,2,15
RTE-70_LMi	.	Locusta migratoria	RTE	1,26,3,11,4,2
RTE-71_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-8_AMi	.	Alligator mississippiensis	RTE	16,1,26,3,11,4,2,15
RTE-8_BF	.	Branchiostoma floridae	RTE	53,2,54,11,57,68,1,26,4,65,3,59,60
RTE-8_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,24,11,4,2
RTE-8_Crp	.	Crocodylus porosus	RTE	
RTE-8_Hmel	.	Heliconius melpomene	RTE	1,3,4,2
RTE-8_HRo	.	Helobdella robusta	RTE	18,5,1,19,3,11,4,2
RTE-8_LCh	.	Latimeria chalumnae	RTE	16,1,26,3,11,4,2,15
RTE-8_LMi	.	Locusta migratoria	RTE	1,3,11,4,2
RTE-8_LVa	.	Litopenaeus vannamei	RTE	16,1,26,3,11,4,2,15
RTE-8_NV	.	Nematostella vectensis	RTE	
RTE-8_PM	.	Petromyzon marinus	RTE	18,5,16,1,19,3,11,4,2,15
RTE-8_PPac	.	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-8_PPo	.	Papilio polytes	RTE	
RTE-8_PSi	.	Pelodiscus sinensis	RTE	1,3,2
RTE-8_PXu	.	Papilio xuthus	RTE	
RTE-8_SP	.	Strongylocentrotus purpuratus	RTE	16,1,26,3,11,4,2,15
RTE-9_BF	.	Branchiostoma floridae	RTE	1,26,3,11,4,2
RTE-9_CPB	.	Chrysemys picta bellii	RTE	16,1,26,3,4,11,2,15
RTE-9_Crp	.	Crocodylus porosus	RTE	1,3,11,4,2
RTE-9_Hmel	.	Heliconius melpomene	RTE	1,3,4,2
RTE-9_LCh	.	Latimeria chalumnae	RTE	2,18,11,1,26,4,15,5,16,19,3
RTE-9_LMi	.	Locusta migratoria	RTE	16,1,26,3,24,11,4,2
RTE-9_LVa	.	Litopenaeus vannamei	RTE	1,3,11,4,2
RTE-9_NVe	.	Nematostella vectensis	RTE	1,3,4,2
RTE-9_PM	.	Petromyzon marinus	RTE	1,3,11,4,2
RTE-9_PPac	.	Pristionchus pacificus	RTE	1,26,3,11,4,2
RTE-9_PPo	.	Papilio polytes	RTE	16,1,26,3,4,11,2,15
RTE-9_PSi	.	Pelodiscus sinensis	RTE	16,1,26,3,11,4,2,15
RTE-9_PXu	.	Papilio xuthus	RTE	
RTE-9_SP	.	Strongylocentrotus purpuratus	RTE	1,3,4,2
RTE-N1_ATr	.	Amborella trichopoda	RTE	
RTE-N1_Crp	.	Crocodylus porosus	RTE	
RTE-N1_Lch	.	Latimeria chalumnae	RTE	
RTE-N1B_ATr	.	Amborella trichopoda	RTE	
RTE-N2_Lch	.	Latimeria chalumnae	RTE	
RTE0_Mars	.	Metatheria	RTE	
RTE1_ET	.	Echinops telfairi	RTE	1,3,11,4,2
RTE1_LA	.	Loxodonta africana	RTE	65,16,1,26,3,24,11,2

RTE1_MT	.	Medicago truncatula	RTE	1,3,4,2
RTE1_MT	.	Medicago truncatula	RTE	1,3,4,2
RTE1_Pca	.	Procavia capensis	RTE	16,1,26,3,24,11,4,2
RTE1_Sar	.	Sorex araneus	RTE	1,3,11,4,2
RTE1_ZM	.	Zea mays	RTE	1,3,11,4,2
RTE1-N1_LA	.	Loxodonta africana	RTE	
RTE1-N1b_LA	.	Loxodonta africana	RTE	
RTE1N_Sar	.	Sorex araneus	RTE	
RTE1X_SP	048B05	Strongylocentrotus purpuratus	RTE	16,1,26,3,24,11,4,2
RTE2_ZM	.	Zea mays	RTE	65,1,26,3,11,4,2
RTESINE1	.	Monodelphis	SINE	
RTESINE2	.	Metatheria	SINE	
RTETP-1_PT	chr_7	Phaeodactylum tricornutum	RTETP	
RTEX-1_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-1_AMi	.	Alligator mississippiensis	RTEX	1,3,11,4,2
RTEX-1_BF	.	Branchiostoma floridae	RTEX	62,2,63,24,11,1,64,4,16,3
RTEX-1_BTa	.	Bemisia tabaci	RTEX	21,16,1,3,4,11,2,15
RTEX-1_CGi	.	Crassostrea gigas	RTEX	21,1,23,3,11,4,2
RTEX-1_CoFl	.	Copidosoma floridanum	RTEX	1,3,11,4,2
RTEX-1_CPB	.	Chrysemys picta bellii	RTEX	
RTEX-1_CR	.	Chlamydomonas reinhardtii	RTEX	1,3,11,4,2
RTEX-1_Crp	.	Crocodylus porosus	RTEX	1,3,4,2
RTEX-1_DR	.	Danio rerio	RTEX	16,1,3,24,11,4,62,2
RTEX-1_DRh	.	Drosophila rhopaloa	RTEX	
RTEX-1_DT	.	Drosophila takahashii	RTEX	1,3,11,4,2
RTEX-1_EL	.	Esox lucius	RTEX	1,3,11,4,2
RTEX-1_FoCa	.	Folsomia candida	RTEX	16,1,3,11,4,2,15
RTEX-1_LG	.	Lottia gigantea	RTEX	1,3,11,4,2
RTEX-1_NV	.	Nematostella vectensis	RTEX	62,2,63,24,11,1,64,4,16,3
RTEX-1_PBa	.	Pogonomyrmex barbatus	RTEX	16,1,3,11,4,2,15
RTEX-1_RPr	.	Rhodnius prolixus	RTEX	16,1,3,11,4,2,15
RTEX-1_SK	.	Saccoglossus kowalevskii	RTEX	2,63,11,1,64,4,15,16,3
RTEX-1_SP	.	Strongylocentrotus purpuratus	RTEX	1,3,11,4,2
RTEX-1_SSa	.	Salmo salar	RTEX	1,3,11,4,2
RTEX-1_VC	.	Volvox carteri	RTEX	1,3,82,11,4,2
RTEX-10_ACar	.	Anolis carolinensis	RTEX	1,3,11,4,2
RTEX-10_BF	.	Branchiostoma floridae	RTEX	
53,62,2,63,54,11,56,57,48,68,1,64,58,4,127,3,59,60				
RTEX-10_CGi	.	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-10_RPr	.	Rhodnius prolixus	RTEX	16,1,3,11,4,2,15
RTEX-10_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-11_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-11_BF	.	Branchiostoma floridae	RTEX	1,109,3,110,11,4,62,2
RTEX-11_CGi	.	Crassostrea gigas	RTEX	
RTEX-11_RPr	.	Rhodnius prolixus	RTEX	21,11
RTEX-11_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-11A_BF	.	Branchiostoma floridae	RTEX	
RTEX-11B_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-12_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,4,11,2
RTEX-12_BF	.	Branchiostoma floridae	RTEX	16,1,3,24,11,4,2
RTEX-12_CGi	.	Crassostrea gigas	RTEX	1,3,4,2

RTEX-12_RPr	.	Rhodnius prolixus	RTEX	16,1,3,11,4,2,15
RTEX-12_SK	.	Saccoglossus kowalevskii	RTEX	16,1,3,24,11,4,2
RTEX-13_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-13_BF	.	Branchiostoma floridae	RTEX	16,1,3,11,4,2,15
RTEX-13_CGi	.	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-13_RPr	.	Rhodnius prolixus	RTEX	1,3,11,4,2
RTEX-13_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-14_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-14_BF	.	Branchiostoma floridae	RTEX	2,1,109,3,110,11,4,62
RTEX-14_CGi	.	Crassostrea gigas	RTEX	21,1,3,11,4,2
RTEX-14_SK	.	Saccoglossus kowalevskii	RTEX	1,3,11,4,62,2
RTEX-15_CGi	.	Crassostrea gigas	RTEX	1,3,4,2
RTEX-15_SK	.	Saccoglossus kowalevskii	RTEX	1,3,11,4,2
RTEX-16_BF	.	Branchiostoma floridae	RTEX	16,1,3,24,11,4,2
RTEX-16_CGi	.	Crassostrea gigas	RTEX	21,16,1,23,3,24,11,4,2
RTEX-16_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-17_BF	.	Branchiostoma floridae	RTEX	16,1,3,11,4,2,15
RTEX-17_CGi	.	Crassostrea gigas	RTEX	
RTEX-17_SK	.	Saccoglossus kowalevskii	RTEX	16,1,3,24,11,4,2
RTEX-18_CGi	.	Crassostrea gigas	RTEX	21,1,23,3,11,4,2
RTEX-18_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-19_CGi	.	Crassostrea gigas	RTEX	21,1,23,3,11,4,2
RTEX-19_SK	.	Saccoglossus kowalevskii	RTEX	16,1,3,24,11,4,2
RTEX-2_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-2_BF	.	Branchiostoma floridae	RTEX	189,190,62,2,191,192,193,11,194,1,4,15,16,3
RTEX-2_BTa	.	Bemisia tabaci	RTEX	16,1,3,11,4,2,15
RTEX-2_CGi	.	Crassostrea gigas	RTEX	1,3,4,2
RTEX-2_CoFl	.	Copidosoma floridanum	RTEX	16,1,3,4,11,2,15
RTEX-2_DR	.	Danio rerio	RTEX	1,3,4,2
RTEX-2_DT	.	Drosophila takahashii	RTEX	1,3,11,4,2
RTEX-2_FoCa	.	Folsomia candida	RTEX	16,1,3,11,4,2,15
RTEX-2_NV	.	Nematostella vectensis	RTEX	62,16,1,3,24,11,4,2
RTEX-2_PBa	.	Pogonomyrmex barbatus	RTEX	
RTEX-2_RPr	.	Rhodnius prolixus	RTEX	1,3,4,2
RTEX-2_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-2_SSa	.	Salmo salar	RTEX	2,16,1,3,24,11,4,62
RTEX-2_VC	.	Volvox carteri	RTEX	1,3,11,4,2
RTEX-20_CGi	.	Crassostrea gigas	RTEX	21,1,23,3,11,4,2
RTEX-20_SK	.	Saccoglossus kowalevskii	RTEX	1,3,11,4,2
RTEX-21_CGi	.	Crassostrea gigas	RTEX	1,23,3,4,11,2
RTEX-21_SK	.	Saccoglossus kowalevskii	RTEX	16,1,3,24,11,4,2
RTEX-22_CGi	.	Crassostrea gigas	RTEX	1,3,4,2
RTEX-22_SK	.	Saccoglossus kowalevskii	RTEX	16,1,3,11,4,2,15
RTEX-23_CGi	.	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-23_SK	.	Saccoglossus kowalevskii	RTEX	16,21,1,3,24,11,4,2
RTEX-24_CGi	.	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-24_SK	.	Saccoglossus kowalevskii	RTEX	16,1,3,4,11,2,15
RTEX-25_CGi	.	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-25_SK	.	Saccoglossus kowalevskii	RTEX	1,3,11,4,2
RTEX-26_CGi	.	Crassostrea gigas	RTEX	16,1,3,11,4,2,15
RTEX-26_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2

RTEX-27_CGi	.	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-27_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-28_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-3_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-3_BF	.	Branchiostoma floridae	RTEX	2,1,3,11,4,62
RTEX-3_CGi	.	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-3_DR	.	Danio rerio	RTEX	
RTEX-3_NV	.	Nematostella vectensis	RTEX	2,16,1,3,11,4,62,15
RTEX-3_RPr	.	Rhodnius prolixus	RTEX	1,3,4,2
RTEX-3_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,2
RTEX-3_SSa	.	Salmo salar	RTEX	62,1,3,11,4,2
RTEX-3_VC	.	Volvox carteri	RTEX	16,1,3,24,11,4,2
RTEX-4_ACar	.	Anolis carolinensis	RTEX	1,3,11,4,2
RTEX-4_BF	.	Branchiostoma floridae	RTEX	
53,62,2,63,54,11,56,57,68,1,64,58,4,15,16,3,59,60				
RTEX-4_CGi	.	Crassostrea gigas	RTEX	21,1,3,11,4,2
RTEX-4_DR	.	Danio rerio	RTEX	2,1,3,11,4,48,62
RTEX-4_NV	.	Nematostella vectensis	RTEX	16,1,3,4,11,48,62,2,15
RTEX-4_RPr	.	Rhodnius prolixus	RTEX	1,3,2
RTEX-4_SK	.	Saccoglossus kowalevskii	RTEX	1,3,11,4,62,2
RTEX-4_VC	.	Volvox carteri	RTEX	1,3,11,4,2
RTEX-5_ACar	.	Anolis carolinensis	RTEX	16,1,3,11,4,2,15
RTEX-5_BF	.	Branchiostoma floridae	RTEX	16,1,3,24,11,4,62,2
RTEX-5_CGi	.	Crassostrea gigas	RTEX	1,3,2
RTEX-5_DR	.	Danio rerio	RTEX	1,3,11,4,2
RTEX-5_NV	.	Nematostella vectensis	RTEX	
RTEX-5_RPr	.	Rhodnius prolixus	RTEX	1,3,11,4,2
RTEX-5_SK	.	Saccoglossus kowalevskii	RTEX	1,3,11,4,2
RTEX-5_VC	.	Volvox carteri	RTEX	1,3,11,4,2
RTEX-6_ACar	.	Anolis carolinensis	RTEX	1,3,11,4,2
RTEX-6_BF	.	Branchiostoma floridae	RTEX	16,1,3,24,11,4,2
RTEX-6_CGi	.	Crassostrea gigas	RTEX	1,3,11,4,2
RTEX-6_NV	.	Nematostella vectensis	RTEX	
RTEX-6_RPr	.	Rhodnius prolixus	RTEX	2,21,23,11,1,4,15,16,3
RTEX-6_SK	.	Saccoglossus kowalevskii	RTEX	1,3,4,11,2
RTEX-7_ACar	.	Anolis carolinensis	RTEX	16,1,3,4,11,2,15
RTEX-7_BF	.	Branchiostoma floridae	RTEX	1,3,4,2
RTEX-7_CGi	.	Crassostrea gigas	RTEX	16,1,3,24,11,4,2
RTEX-7_RPr	.	Rhodnius prolixus	RTEX	21,1,3,11,4,2
RTEX-7_SK	.	Saccoglossus kowalevskii	RTEX	63,2,16,1,64,3,24,11,4,62
RTEX-8_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-8_BF	.	Branchiostoma floridae	RTEX	1,3,4,2
RTEX-8_CGi	.	Crassostrea gigas	RTEX	
RTEX-8_NV	.	Nematostella vectensis	RTEX	
RTEX-8_RPr	.	Rhodnius prolixus	RTEX	1,3,11,4,2
RTEX-8_SK	.	Saccoglossus kowalevskii	RTEX	1,3,11,4,2
RTEX-9_ACar	.	Anolis carolinensis	RTEX	16,1,3,24,11,4,2
RTEX-9_BF	.	Branchiostoma floridae	RTEX	1,3,11,4,2
RTEX-9_CGi	.	Crassostrea gigas	RTEX	21,1,23,3,11,4,2
RTEX-9_RPr	.	Rhodnius prolixus	RTEX	1,3,11,4,2
RTEX-9_SK	.	Saccoglossus kowalevskii	RTEX	1,3,11,4,2

RTEY-1_CTe	.	Capitella teleta	RTEY	16,1,3,11,4,2,15
RTEY-2_CTe	.	Capitella teleta	RTEY	1,3,4,2
RTEY-3_CTe	.	Capitella teleta	RTEY	1,3,4,2
Ruka	.	Ixodida	SINE2/tRNA	
S_Av	.	Adineta vaga	Penelope	1,3
S0_Av	.	Adineta vaga	Penelope	1,3
S1_B:N	L24035	Brassica napus	SINE2/tRNA	
SacSINE1	.	Squalus acanthias	SINE2/tRNA	
Sadhu-1_ALy	.	Arabidopsis lyrata	SINE	
Sadhu-2_ALy	.	Arabidopsis lyrata	SINE	
Sadhu-3_ALy	.	Arabidopsis lyrata	Non-LTRRetrotransposo	
Sadhu-4_ALy	.	Arabidopsis lyrata	Non-LTRRetrotransposo	
Sadhu1-1	NC_003071.3	Arabidopsis thaliana	SINE	
Sadhu1-2	NC_003070.5	Arabidopsis thaliana	SINE	
Sadhu1-3	.	Arabidopsis thaliana	SINE	
Sadhu10-1	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu2-1	NC_003070.5	Arabidopsis thaliana	SINE	
Sadhu3-1	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu3-2	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu4-1	NC_003076.4	Arabidopsis thaliana	SINE	
Sadhu4-2	NC_003070.5	Arabidopsis thaliana	SINE	
Sadhu5-1	NC_003075.3	Arabidopsis thaliana	SINE	
Sadhu5-2	NC_003076.4	Arabidopsis thaliana	SINE	
Sadhu6-1	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu7-1	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu7-2	NC_003074.4	Arabidopsis thaliana	SINE	
Sadhu8-1	NC_003070.5	Arabidopsis thaliana	SINE	
Sadhu9-1	NC_003070.5	Arabidopsis thaliana	SINE	
Sake_BM	.	Bombyx mori	Daphne	16,1,3,11,4,2,15
Sake2_BM	.	Bombyx mori	Daphne	1,3,4,2
SalSINE1	.	Salmo salar	SINE2/tRNA	
SART-1_APi	.	Acyrthosiphon pisum	R1	2,8,11,86,12,1,14,4,15,16,3
SART-1_CF	.	Camponotus floridanus	R1	2,8,11,86,12,1,14,4,15,16,3
SART-1_HMM	.	Heliconius melpomene melpomene	R1	2,8,11,86,12,1,14,4,15,16,3
SART-1_LHu	.	Linepithema humile	R1	2,11,86,12,1,14,217,4,15,16,3
SART-1_MOc	AFFJ01003038	Galendromus occidentalis	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-1_PHu	.	Pediculus humanus	R1	
SART-1_PXu	.	Papilio xuthus	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-1B_HMM	.	Heliconius melpomene melpomene	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-2_APi	.	Acyrthosiphon pisum	R1	8,14,12
SART-2_MOc	AFFJ01001302	Galendromus occidentalis	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-2_PXu	.	Papilio xuthus	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-3_APi	.	Acyrthosiphon pisum	R1	8,14,12
SART-3_MOc	AFFJ01002771	Galendromus occidentalis	R1	1,3,4,2
SART-4_APi	.	Acyrthosiphon pisum	R1	2,16,1,14,3,11,4,86,15
SART-4_MOc	.	Galendromus occidentalis	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-5_APi	.	Acyrthosiphon pisum	R1	16,1,3,4,11,86,2,15
SART-5_MOc	.	Galendromus occidentalis	R1	2,8,11,86,12,42,1,14,4,15,16,3
SART-6_APi	.	Acyrthosiphon pisum	R1	2,8,11,86,12,1,14,4,15,16,3
SART-7_APi	.	Acyrthosiphon pisum	R1	1,3,4,2
SART-8_APi	.	Acyrthosiphon pisum	R1	

SART-9_APi	.	Acyrthosiphon pisum	R1	16,1,3,4,11,86,2,15
SART1	D85594	Bombyx mori	R1	2,8,24,11,86,12,1,14,4,16,3
SART1_R1_Hmel	.	Heliconius melpomene	R1	8,14,12
SARTBm3	.	Bombyx mori	R1	2,8,11,86,12,1,14,4,15,16,3
SARTPx1	.	Papilio xuthus	R1	2,8,11,86,12,1,14,4,15,16,3
SARTTc1	AB593321	Tribolium castaneum	R1	2,8,11,86,12,42,1,14,4,15,16,3
SARTTc2	AB593322	Tribolium castaneum	R1	2,8,11,86,12,1,14,4,15,16,3
SARTTc3	AB593323	Tribolium castaneum	R1	2,8,11,86,12,1,14,4,15,16,3
SARTTc4	AB593324	Tribolium castaneum	R1	2,11,86,12,1,14,4,15,16,3
SARTTc5	AB593325	Tribolium castaneum	R1	2,11,86,12,1,14,4,15,16,3
SARTTc6	AB593326	Tribolium castaneum	R1	2,8,11,86,12,42,1,14,4,15,16,3
SARTTc7	AB593327	Tribolium castaneum	R1	2,8,11,86,12,1,14,4,15,16,3
SARTTc8	.	Tribolium castaneum	R1	2,8,11,86,12,1,14,4,15,16,3
SepiaSINE	.	Sepia officinalis	SINE2/tRNA	
SepiaSINE2	.	Sepia officinalis	SINE2/tRNA	
Sepioth-SINE1	.	Sepioteuthis lessoniana	SINE2/tRNA	
Sepioth-SINE2A	.	Sepioteuthis lessoniana	SINE2/tRNA	
Sepioth-SINE2B	.	Sepioteuthis lessoniana	SINE2/tRNA	
SHALINE10_MT	.	Medicago truncatula	L1	78,2,8,11,12,1,14,81,77,4,75,3,76
SHALINE10_MT	.	Medicago truncatula	L1	78,2,8,11,12,1,14,81,77,4,75,3,76
SHALINE11_MT	.	Medicago truncatula	L1	100,1,3,11,4,2
SHALINE11_MT	.	Medicago truncatula	L1	100,1,3,11,4,2
SHALINE12_MT	.	Medicago truncatula	L1	1,75,8,3,76,77,11,4,12,2
SHALINE12_MT	.	Medicago truncatula	L1	1,75,8,3,76,77,11,4,12,2
SHALINE14_MT	.	Medicago truncatula	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
SHALINE14_MT	.	Medicago truncatula	L1	43,79,6,2,10,44,11,80,1,4,17,3,7
SHALINE15_MT	AC147498	Medicago truncatula	L1	2,93,1,75,3,76,92
SHALINE15_MT	AC147498	Medicago truncatula	L1	2,93,1,75,3,76,92
SHALINE16_MT	.	Medicago truncatula	L1	
43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7				
SHALINE16_MT	.	Medicago truncatula	L1	
43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7				
SHALINE17_MT	.	Medicago truncatula	L1	1,3,11,4,2
SHALINE17_MT	.	Medicago truncatula	L1	1,3,11,4,2
SHALINE18_MT	.	Medicago truncatula	L1	1,3,4,2
SHALINE18_MT	.	Medicago truncatula	L1	1,3,4,2
SHALINE2_MT	.	Medicago truncatula	L1	44,43
SHALINE2_MT	.	Medicago truncatula	L1	44,43
SHALINE20_MT	.	Medicago truncatula	L1	43,79,6,2,10,44,80,1,4,17,3,7
SHALINE20_MT	.	Medicago truncatula	L1	43,79,6,2,10,44,80,1,4,17,3,7
SHALINE21_MT	.	Medicago truncatula	L1	75,76,78,81,77
SHALINE21_MT	.	Medicago truncatula	L1	75,76,78,81,77
SHALINE4_MT	.	Medicago truncatula	L1	75,76,77
SHALINE4_MT	.	Medicago truncatula	L1	75,76,77
SHALINE5_MT	.	Medicago truncatula	L1	
43,79,2,8,10,44,24,11,12,80,1,77,4,16,17,75,3,76,7				
SHALINE5_MT	.	Medicago truncatula	L1	
43,79,2,8,10,44,24,11,12,80,1,77,4,16,17,75,3,76,7				
SHALINE6_MT	.	Medicago truncatula	L1	
43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7				

SHALINE6_MT	.	Medicago truncatula	L1	
43,78,79,2,10,44,24,11,80,1,81,77,4,16,17,75,3,76,7				
SHALINE7_MT	.	Medicago truncatula	L1	43,79,6,2,44,10,11,80,1,4,17,3,7
SHALINE7_MT	.	Medicago truncatula	L1	43,79,6,2,44,10,11,80,1,4,17,3,7
SHALINE8_MT	.	Medicago truncatula	L1	78,2,24,11,1,81,77,4,16,75,76
SHALINE8_MT	.	Medicago truncatula	L1	78,2,24,11,1,81,77,4,16,75,76
SHALINE9_MT	.	Medicago truncatula	L1	2,6,1,17,3,10,79,7,80
SHALINE9_MT	.	Medicago truncatula	L1	2,6,1,17,3,10,79,7,80
SHANSINE_MT	.	Medicago truncatula	SINE	
SINE_AFC	.	Lepidiolamprologus elongatus	SINE2/tRNA	
SINE_CP	X79504	Chironomus pallidivittatus	SINE	
SINE_CP1	X79507	Chironomus pallidivittatus	SINE	
SINE_DR1	.	Danio rerio	SINE	
SINE_FR1A	.	Takifugu rubripes	SINE2/tRNA	
SINE_FR1B	.	Takifugu rubripes	SINE2/tRNA	
SINE_FR1C	.	Takifugu rubripes	SINE2/tRNA	
SINE_FR1D	.	Takifugu rubripes	SINE2/tRNA	
SINE_FR2	.	Takifugu rubripes	SINE2/tRNA	
SINE_RR	.	Pelophylax ridibundus	SINE	
SINE_SM	.	Prosopium spilonotus	SINE2/tRNA	
SINE_TE	.	Teleostei	SINE2/tRNA	
SINE_UT	AB108791, AB108792,	Urotrichus talpoides	SINE	
SINE_VV	.	Vicugna vicugna	SINE2/tRNA	
SINE-1_AIp	.	Arachis ipaensis	SINE2/tRNA	
SINE-1_ALy	.	Arabidopsis lyrata	SINE	
SINE-1_ATr	.	Amborella trichopoda	SINE	
SINE-1_BM	.	Bombyx mori	SINE	
SINE-1_CQ	.	Culex quinquefasciatus	SINE	
SINE-1_ECa	.	Eucalyptus camaldulensis	SINE	
SINE-1_EPa	.	Schrenkiella parvula	SINE2/tRNA	
SINE-1_HM	.	Hydra vulgaris	SINE	
SINE-1_LMi	.	Locusta migratoria	SINE	
SINE-1_Mad	.	Malus domestica	SINE2/tRNA	
SINE-1_MD	.	Monodelphis domestica	SINE2/tRNA	
SINE-1_Pmo	.	Python molurus	SINE3/5S	
SINE-12_OS	.	Oryza sativa	SINE2/tRNA	
SINE-2_CQ	.	Culex quinquefasciatus	SINE	
SINE-2_EL	.	Esox lucius	SINE3/5S	
SINE-2_LMi	.	Locusta migratoria	SINE	
SINE-2_Mad	.	Malus domestica	SINE2/tRNA	
SINE-2_MD	.	Monodelphis domestica	SINE2/tRNA	
SINE-2019_Crp	.	Crocodylus porosus	SINE	
SINE-3_CQ	.	Culex quinquefasciatus	SINE2/tRNA	
SINE-3_LMi	.	Locusta migratoria	SINE	
SINE-3_Mad	.	Malus domestica	SINE2/tRNA	
SINE-3B_CQ	.	Culex quinquefasciatus	SINE2/tRNA	
SINE-4_CQ	.	Culex quinquefasciatus	SINE2/tRNA	
SINE-4_LMi	.	Locusta migratoria	SINE	
SINE-4_Mad	.	Malus domestica	SINE2/tRNA	
SINE-5_CQ	.	Culex quinquefasciatus	SINE2/tRNA	
SINE-5_FV	.	Fragaria vesca	SINE	

SINE-5_LMi	.		Locusta migratoria	SINE
SINE-5_Mad	.		Malus domestica	SINE2/tRNA
SINE-6_CQ	.		Culex quinquefasciatus	SINE2/tRNA
SINE-6_CR	.		Chlamydomonas reinhardtii	SINE2/tRNA
SINE1_BB	AF327249		Bubalus bubalis	SINE
SINE1_CE	.		Caenorhabditis elegans	SINE
SINE1_MT	.		Medicago truncatula	SINE
SINE1_MT	.		Medicago truncatula	SINE2/tRNA
SINE1_SO	.		Solanum demissum	SINE
SINE1_SS	AC129961,	AF018149,	Sus scrofa	SINE2/tRNA
SINE16_OS	D85060		Oryza sativa	SINE2/tRNA
SINE1A_SS	.		Sus scrofa	SINE2/tRNA
SINE1a1_Cis	.		Ciona savignyi	SINE
SINE1a2_Cis	.		Ciona savignyi	SINE
SINE1B_SS	.		Sus scrofa	SINE2/tRNA
SINE1b1_Cis	.		Ciona savignyi	SINE
SINE1b2_Cis	.		Ciona savignyi	SINE
SINE1C_SS	.		Sus scrofa	SINE2/tRNA
SINE1D_SS	.		Sus scrofa	SINE2/tRNA
SINE1OS	D10673		Oryza sativa	SINE2/tRNA
SINE1R5_OS	D10674		Oryza sativa	SINE2/tRNA
SINE2_Gav	.		Crocodylidae	SINE2/tRNA
SINE2-1_AC	.		Aplysia californica	SINE2/tRNA
SINE2-1_ACar	.		Anolis carolinensis	SINE2/tRNA
SINE2-1_ADi	.		Acropora digitifera	SINE2/tRNA
SINE2-1_AFC	.		Cichlidae	SINE2/tRNA
SINE2-1_ALy	.		Arabidopsis lyrata	SINE2/tRNA
SINE2-1_AMi	.		Alligator mississippiensis	SINE2/tRNA
SINE2-1_AP	.		Acyrthosiphon pisum	SINE2/tRNA
SINE2-1_BDi	.		Brachypodium distachyon	SINE2/tRNA
SINE2-1_BG	.		Blumeria graminis	SINE2/tRNA
SINE2-1_BT	.		Bos taurus	SINE2/tRNA
SINE2-1_BTa	.		Bemisia tabaci	SINE2/tRNA
SINE2-1_BV	.		Beta vulgaris	SINE2/tRNA
SINE2-1_CCa	.		Cyprinus carpio	SINE2/tRNA
SINE2-1_CCr	.		Condylura cristata	SINE2/tRNA
SINE2-1_CM	.		Callorhinchus milii	SINE2/tRNA
SINE2-1_CoFl	.		Copidosoma floridanum	SINE2/tRNA
SINE2-1_CQ	.		Culex quinquefasciatus	SINE2/tRNA
SINE2-1_Croc	.		Crocodylidae	SINE2/tRNA
SINE2-1_DaCa	.		Daucus carota	SINE2/tRNA
SINE2-1_DL	.		Dicentrarchus labrax	SINE2/tRNA
SINE2-1_DR	.		Danio rerio	SINE2/tRNA
SINE2-1_EC	.		Equus caballus	SINE2/tRNA
SINE2-1_EL	.		Esox lucius	SINE2/tRNA
SINE2-1_FV	.		Fragaria vesca	SINE2/tRNA
SINE2-1_GA	.		Gasterosteus aculeatus	SINE2/tRNA
SINE2-1_HM	.		Hydra vulgaris	SINE2/tRNA
SINE2-1_JC	.		Jatropha curcas	SINE2/tRNA
SINE2-1_Laf	.		Loxodonta africana	SINE2/tRNA
SINE2-1_LMi	.		Locusta migratoria	SINE2/tRNA

SINE2-1_LVa	Litopenaeus vannamei	SINE2/tRNA
SINE2-1_MAc	Musa acuminata	SINE2/tRNA
SINE2-1_NS	Nicotiana sylvestris	SINE2/tRNA
SINE2-1_NV	Nematostella vectensis	SINE2/tRNA
SINE2-1_OL	Oryzias latipes	SINE2/tRNA
SINE2-1_OS	Oryza sativa	SINE2/tRNA
SINE2-1_Pca	Procavia capensis	SINE2/tRNA
SINE2-1_PD	Phoenix dactylifera	SINE2/tRNA
SINE2-1_PH	Parhyale hawaiensis	SINE2/tRNA
SINE2-1_PL	Paracentrotus lividus	SINE2/tRNA
SINE2-1_PM	Petromyzon marinus	SINE2/tRNA
SINE2-1_PMaj	Parus major	SINE2/tRNA
SINE2-1_PPo	Papilio polytes	SINE2/tRNA
SINE2-1_PSi	Pelodiscus sinensis	SINE2/tRNA
SINE2-1_PTr	Populus trichocarpa	SINE2/tRNA
SINE2-1_PTr	Populus trichocarpa	SINE2/tRNA
SINE2-1_PXu	Papilio xuthus	SINE2/tRNA
SINE2-1_RPr	Rhodnius prolixus	SINE2/tRNA
SINE2-1_SBi	Sorghum bicolor	SINE2/tRNA
SINE2-1_SLy	Solanum lycopersicum	SINE2/tRNA
SINE2-1_SP	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-1_SSc	Sus scrofa	SINE2/tRNA
SINE2-1_STu	Solanum tuberosum	SINE2/tRNA
SINE2-1_TAe	Triticum aestivum	SINE2/tRNA
SINE2-1_VC	Volvox carteri	SINE2/tRNA
SINE2-1_VMa	Vaccinium macrocarpon	SINE2/tRNA
SINE2-1_XT	Xenopus tropicalis	SINE2/tRNA
SINE2-1_ZM	Zea mays	SINE2/tRNA
SINE2-1l_OS	Oryza sativa	SINE2/tRNA
SINE2-1l_SBi	Sorghum bicolor	SINE2/tRNA
SINE2-1z_OS	Oryza sativa	SINE2/tRNA
SINE2-1z_SBi	Sorghum bicolor	SINE2/tRNA
SINE2-1a_SBi	Sorghum bicolor	SINE2/tRNA
SINE2-1B_ACar	Anolis carolinensis	SINE2/tRNA
SINE2-1B_CCa	Cyprinus carpio	SINE2/tRNA
SINE2-1B_CCr	Condylura cristata	SINE2/tRNA
SINE2-1B_DR	Danio rerio	SINE2/tRNA
SINE2-1B_VMa	Vaccinium macrocarpon	SINE2/tRNA
SINE2-1C_CCa	Cyprinus carpio	SINE2/tRNA
SINE2-1c_SBi	Sorghum bicolor	SINE2/tRNA
SINE2-1D_CCa	Cyprinus carpio	SINE2/tRNA
SINE2-2_AC	Aplysia californica	SINE2/tRNA
SINE2-2_ACar	Anolis carolinensis	SINE2/tRNA
SINE2-2_ADi	Acropora digitifera	SINE2/tRNA
SINE2-2_AP	Acyrthosiphon pisum	SINE2/tRNA
SINE2-2_BT	Bos taurus	SINE2/tRNA
SINE2-2_BV	Beta vulgaris	SINE2/tRNA
SINE2-2_CCa	Cyprinus carpio	SINE2/tRNA
SINE2-2_CM	Callorhinchus milii	SINE2/tRNA
SINE2-2_CoFl	Copidosoma floridanum	SINE2/tRNA
SINE2-2_CQ	Culex quinquefasciatus	SINE2/tRNA

SINE2-2_DaCa	Daucus carota	SINE2/tRNA
SINE2-2_DR	Danio rerio	SINE2/tRNA
SINE2-2_FV	Fragaria vesca	SINE2/tRNA
SINE2-2_HM	Hydra vulgaris	SINE2/tRNA
SINE2-2_LMi	Locusta migratoria	SINE2/tRNA
SINE2-2_LVa	Litopenaeus vannamei	SINE2/tRNA
SINE2-2_MAc	Musa acuminata	SINE2/tRNA
SINE2-2_NV	Nematostella vectensis	SINE2/tRNA
SINE2-2_OL	Oryzias latipes	SINE2/tRNA
SINE2-2_OS	Oryza sativa	SINE2/tRNA
SINE2-2_PH	Parhyale hawaiensis	SINE2/tRNA
SINE2-2_PM	Petromyzon marinus	SINE2/tRNA
SINE2-2_PPo	Papilio polytes	SINE2/tRNA
SINE2-2_PTr	Populus trichocarpa	SINE2/tRNA
SINE2-2_PTr	Populus trichocarpa	SINE2/tRNA
SINE2-2_PXu	Papilio xuthus	SINE2/tRNA
SINE2-2_RPr	Rhodnius prolixus	SINE2/tRNA
SINE2-2_SP	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-2_STu	Solanum tuberosum	SINE2/tRNA
SINE2-2_VMa	Vaccinium macrocarpon	SINE2/tRNA
SINE2-2B_HM	Hydra vulgaris	SINE2/tRNA
SINE2-2B_PH	Parhyale hawaiensis	SINE2/tRNA
SINE2-2B_SSa	Salmo salar	SINE2/tRNA
SINE2-2C_SSa	Salmo salar	SINE2/tRNA
SINE2-2D_SSa	Salmo salar	SINE2/tRNA
SINE2-3_ALy	Arabidopsis lyrata	SINE2/tRNA
SINE2-3_BT	Bos taurus	SINE2/tRNA
SINE2-3_CCa	Cyprinus carpio	SINE2/tRNA
SINE2-3_CM	Callorhinchus milii	SINE2/tRNA
SINE2-3_CoFl	Copidosoma floridanum	SINE2/tRNA
SINE2-3_DaCa	Daucus carota	SINE2/tRNA
SINE2-3_DR	Danio rerio	SINE2/tRNA
SINE2-3_EL	Esox lucius	SINE2/tRNA
SINE2-3_FV	Fragaria vesca	SINE2/tRNA
SINE2-3_LMi	Locusta migratoria	SINE2/tRNA
SINE2-3_LVa	Litopenaeus vannamei	SINE2/tRNA
SINE2-3_MAc	Musa acuminata	SINE2/tRNA
SINE2-3_NV	Nematostella vectensis	SINE2/tRNA
SINE2-3_OL	Oryzias latipes	SINE2/tRNA
SINE2-3_OS	Oryza sativa	SINE2/tRNA
SINE2-3_PH	Parhyale hawaiensis	SINE2/tRNA
SINE2-3_PM	Petromyzon marinus	SINE2/tRNA
SINE2-3_SP	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-3_VMa	Vaccinium macrocarpon	SINE2/tRNA
SINE2-3B_DR	Danio rerio	SINE2/tRNA
SINE2-3B_OL	Oryzias latipes	SINE2/tRNA
SINE2-4_ALy	Arabidopsis lyrata	SINE2/tRNA
SINE2-4_CCa	Cyprinus carpio	SINE2/tRNA
SINE2-4_DR	Danio rerio	SINE2/tRNA
SINE2-4_FV	Fragaria vesca	SINE2/tRNA
SINE2-4_NV	Nematostella vectensis	SINE2/tRNA

SINE2-4_OS	.	Oryza sativa	SINE2/tRNA
SINE2-4_PH	.	Parhyale hawaiensis	SINE2/tRNA
SINE2-4_PM	.	Petromyzon marinus	SINE2/tRNA
SINE2-4_SP	.	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-4_VMa	.	Vaccinium macrocarpon	SINE2/tRNA
SINE2-4a_SP	.	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-4a1_SP	.	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-4B_OS	.	Oryza sativa	SINE2/tRNA
SINE2-4b_SP	.	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-4c_SP	.	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-5_ALy	.	Arabidopsis lyrata	SINE2/tRNA
SINE2-5_DR	.	Danio rerio	SINE2/tRNA
SINE2-5_NV	.	Nematostella vectensis	SINE2/tRNA
SINE2-5_OL	.	Oryzias latipes	SINE2/tRNA
SINE2-5_SP	.	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-5_SSa	.	Salmo salar	SINE2/tRNA
SINE2-5_VMa	.	Vaccinium macrocarpon	SINE2/tRNA
SINE2-6_ALy	.	Arabidopsis lyrata	SINE2/tRNA
SINE2-6_SP	.	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-6_SSa	.	Salmo salar	SINE2/tRNA
SINE2-6_VMa	.	Vaccinium macrocarpon	SINE2/tRNA
SINE2-7_SP	.	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2-8_SP	.	Strongylocentrotus purpuratus	SINE2/tRNA
SINE2B-1_CM	.	Callorhinchus milii	SINE2/tRNA
SINE3_AFC	.	Cichlidae	SINE2/tRNA
SINE3_IP	.	Ictalurus punctatus	SINE3/5S
SINE3-1	.	Danio rerio	SINE3/5S
SINE3-1_AO	.	Aspergillus oryzae	SINE3/5S
SINE3-1_CCa	.	Cyprinus carpio	SINE3/5S
SINE3-1_TC	.	Tribolium castaneum	SINE3/5S
SINE3-1a	.	Danio rerio	SINE3/5S
SINE3-2_AO	.	Aspergillus oryzae	SINE3/5S
SINE3-2_DR	.	Danio rerio	SINE3/5S
SINE6_OS	D85050	Oryza sativa	SINE2/tRNA
SINE8_OS	D85052	Oryza sativa	SINE2/tRNA
SINE9_OS	.	Oryza sativa	SINE2/tRNA
SINEB1_MU	.	Meriones unguiculatus	SINE1/7SL
SINEB2_MU	.	Meriones unguiculatus	SINE2/tRNA
SINEC_c1	.	Canidae	SINE2/tRNA
SINEC_Fc	.	Felis	SINE2/tRNA
SINEC_Fc2	.	Felis	SINE2/tRNA
SINEC_Fc3	.	Felis catus	SINE2/tRNA
SINEC_PV	Z33497	Phoca vitulina	SINE2/tRNA
SINEC-1_MPu	.	Mustela putorius furo	SINE2/tRNA
SINEC-1B_MPu	.	Mustela putorius furo	SINE2/tRNA
SINEC-1C_MPu	.	Mustela putorius furo	SINE2/tRNA
SINEC-1D_MPu	.	Mustela putorius furo	SINE2/tRNA
SINEC1_AMe	.	Ailuropoda melanoleuca	SINE2/tRNA
SINEC1_AMe2	.	Ailuropoda melanoleuca	SINE2/tRNA
SINEC1_CF	.	Canis lupus familiaris	SINE2/tRNA
SINEC1A_CF	.	Canis lupus familiaris	SINE2/tRNA

SINEC1B_AMe	.	Ailuropoda melanoleuca	SINE2/tRNA	
SINEC1B1_CF	.	Canis lupus familiaris	SINE2/tRNA	
SINEC1B2_CF	.	Canis lupus familiaris	SINE2/tRNA	
SINEC1C1_CF	.	Canis lupus familiaris	SINE2/tRNA	
SINEC1C2_CF	.	Canis lupus familiaris	SINE2/tRNA	
SINEC1D_CF	.	Canis lupus familiaris	SINE2/tRNA	
SINEC2_AMe	.	Ailuropoda melanoleuca	SINE2/tRNA	
SINEC2A1_CF	.	Canis lupus familiaris	SINE2/tRNA	
SINEC2A2_CF	.	Canis lupus familiaris	SINE2/tRNA	
SINEL_SM	.	Schmidtea mediterranea	NeSL	
SINEU-1_CM	.	Callorhinchus milii	SINEU/snRNA	
SINEU-1A_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1B_CM	.	Callorhinchus milii	SINEU/snRNA	
SINEU-1B_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1C_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1D_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1E_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1F_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1G_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1G2_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1H_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1I_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-1J_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-2_Crp	.	Crocodylus porosus	SINEU/snRNA	
SINEU-2_Gav	.	Gavialis gangeticus	SINEU/snRNA	
SINEU-3A_Croc	.	Crocodylidae	SINEU/snRNA	
SINEU-3B_Croc	.	Crocodylidae	SINEU/snRNA	
SINEX-1_AG	.	Anopheles gambiae	SINE	
SINEX-1_CM	DQ524330	Callorhinchus milii	SINE2/tRNA	
SINEX-1_CR	.	Chlamydomonas reinhardtii	SINE	
SINEX-2_AG	.	Anopheles gambiae	SINE2/tRNA	
SINEX-2_CR	.	Chlamydomonas reinhardtii	SINE	
SINEX-3_CR	.	Chlamydomonas reinhardtii	L1	
SINEX-4_CR	.	Chlamydomonas reinhardtii	SINE	
SINEX-5_CR	.	Chlamydomonas reinhardtii	SINE2/tRNA	
Sismon-1_PH	.	Parhyale hawaiensis	RTE	18,5,1,4,11,2
SK	D14862	Heterololigo bleekeri	SINE2/tRNA	
SLACS	X17078	Trypanosoma brucei	CRE	1,3,4,2
SlmI	.	Salmoniformes	SINE2/tRNA	
SlmII	.	Salmoniformes	SINE2/tRNA	
Sm1	.	Selaginella moellendorffii	Penelope	
Sm2	.	Selaginella moellendorffii	Penelope	
235,236				
SMAIDIV	.	Oncorhynchus keta	SINE2/tRNA	
SOD_SINE	.	Solanum demissum	SINE	
SOD_SINE	.	Solanum demissum	SINE	
SOR1	.	Sorex	SINE2/tRNA	
SP-D-Geo	.	Dipodomys	SINE1/7SL	
SPEC1	J01207	Strongylocentrotus purpuratus	SINE2/tRNA	
SPLN3_SO	AB016137, AB016137.1	Spinacia oleracea	Non-LTRRetrotransposo	
SPOCK_DMi	AY144571	Drosophila miranda	Outcast	16,1,3,11,4,2,15

Squam2	.	Squamata	SINE2/tRNA	
SR1	U66331	Schistosoma mansoni	CR1	1,3,4,2
SR2	AF025672	Schistosoma mansoni	RTE	1,26,3,11,4,2
SR2A	AF025680	Schistosoma mansoni	RTE	
SR2B	AF025681	Schistosoma mansoni	RTE	
STR1	.	Ictidomys tridecemlineatus	SINE1/7SL	
STRID1	.	Ictidomys tridecemlineatus	SINE2/tRNA	
STRID2	.	Ictidomys tridecemlineatus	SINE2/tRNA	
STRID3	.	Ictidomys tridecemlineatus	SINE2/tRNA	
STRID4	.	Ictidomys tridecemlineatus	SINE2/tRNA	
STRIDE_Cpo	.	Cavia porcellus	SINE1/7SL	
STRIDE1	.	Ictidomys tridecemlineatus	SINE1/7SL	
STRIDE2	.	Ictidomys tridecemlineatus	SINE1/7SL	
STRIDM	.	Ictidomys tridecemlineatus	SINE2/tRNA	
SURF1	X07085	Strongylocentrotus purpuratus	SINE2/tRNA	
SUSINE2	.	Sus scrofa	SINE2/tRNA	
SVA_A	.	Homo sapiens	SINE	
SVA_A	.	Homo sapiens	SINE	
SVA_B	.	Homo sapiens	SINE	
SVA_C	.	Homo sapiens	SINE	
SVA_D	.	Homo sapiens	SINE	
SVA_E	.	Homo sapiens	SINE	
SVA_F	.	Homo sapiens	SINE	
SVA2	.	Eutheria	MSAT	
Swimmer	AF055641	Oryzias latipes	L1	
Syrinx_DS	.	Darwinula stevensoni	Jockey	16,1,3,11,4,2,15
T_Av	.	Adineta vaga	Penelope	1,3
T1	M93689	Anopheles gambiae	CR1	54,1,3,59,53,11,4,2
T1_As	.	Adineta	Penelope	1,3,184
T2	X15618	Schistosoma mansoni	SINE2/tRNA	
T2_As	.	Adineta	Penelope	1,3
T3_As	.	Adineta	Penelope	1,3
TA11	L47193	Arabidopsis thaliana	L1	43,78,2,44,24,11,1,81,77,4,16,75,3,76
TA12	.	Arabidopsis thaliana	L1	
Tad1	L25662	Neurospora crassa	Tad1	16,1,14,3,11,4,2,15
Tad1-1_ACap	.	Histoplasma capsulatum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_AF	.	Aspergillus fumigatus	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-1_CCi	AACS02000007	Coprinopsis cinerea	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_CGlo	.	Chaetomium globosum	Tad1	16,1,3,11,4,2,15
Tad1-1_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_GLuc	.	Ganoderma lucidum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-1_NC	.	Neurospora crassa	Tad1	
Tad1-1_PGr	.	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Tad1-1_PPl	.	Postia placenta	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-1_PSt	.	Puccinia striiformis	Tad1	
Tad1-1_PTrit	.	Puccinia triticina	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-1_RGl	AEVR01000393	Rhodotorula toruloides	Tad1	6,2,9,10,11,1,15,16,17,3,7
Tad1-1_TMe	.	Tuber melanosporum	Tad1	16,1,3,11,4,2,15
Tad1-1_TSt	.	Talaromyces stipitatus	Tad1	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7

Tad1-10_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-10_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-10_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-10_PGr	.	Puccinia graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-10_TSt	.	Talaromyces stipitatus	Tad1	
Tad1-100_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-101_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-102_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-11_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-11_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-11_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-11_TSt	.	Talaromyces stipitatus	Tad1	
Tad1-12_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-12_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-13_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-13_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-13B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-14_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-14_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-15_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-15_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-15B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-15C_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-16_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-16_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-17_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-17_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-17B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-18_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-18_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-19_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-19_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-1B_PGr	AAWC01000331	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-2_ACap	.	Histoplasma capsulatum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-2_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-2_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,4,11,2,15
Tad1-2_CGlo	.	Chaetomium globosum	Tad1	16,1,3,11,4,2,15
Tad1-2_EP	.	Erysiphe pisi	Tad1	7,16,1,17,9,3,10,11,4,6,2,15
Tad1-2_PGr	AAWC01004263	Puccinia graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-2_PTrit	.	Puccinia triticina	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-2_TMe	.	Tuber melanosporum	Tad1	16,1,3,11,4,2,15
Tad1-2_TSt	.	Talaromyces stipitatus	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-20_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-20_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-21_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-21_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-22_BG	.	Blumeria graminis	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-22_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-23_BG	.	Blumeria graminis	Tad1	1,3,11,4,2
Tad1-23_EP	.	Erysiphe pisi	Tad1	7,1,17,16,3,11,4,2,15
Tad1-24_BG	.	Blumeria graminis	Tad1	16,1,8,3,11,4,12,2,15

Tad1-24_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25B_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-25C_BG	.	Blumeria graminis	Tad1	17,9,10,6,7,13
Tad1-26_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-26_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-26B_BG	.	Blumeria graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-26C_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-27_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-27_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-28_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-28_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-28B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-29_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-29_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-3_ACap	.	Histoplasma capsulatum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-3_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-3_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-3_CGlo	.	Chaetomium globosum	Tad1	6,2,8,10,11,12,42,1,14,4,15,16,17,3
Tad1-3_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-3_PGr	AAWC01001353	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-3_PTrit	.	Puccinia triticina	Tad1	16,1,3,11,4,2,15
Tad1-3_TMe	.	Tuber melanosporum	Tad1	16,1,3,4,11,2,15
Tad1-3_TSt	.	Talaromyces stipitatus	Tad1	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3
Tad1-30_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-30_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-30B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-31_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-31_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-31B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-32_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-32_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-32B_EP	.	Erysiphe pisi	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-33_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-33_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-34_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-34_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-35_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-35_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-35B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-36_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-36_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-37_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-37_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-38_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-38_EP	.	Erysiphe pisi	Tad1	16,42,1,8,3,4,11,12,2,15
Tad1-38B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-39_BG	.	Blumeria graminis	Tad1	16,1,8,14,3,11,4,12,2,15
Tad1-39_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7

Tad1-4_ACap	.	Histoplasma capsulatum	Tad1	2,8,11,12,1,14,4,15,16,3
Tad1-4_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-4_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-4_CGlo	.	Chaetomium globosum	Tad1	6,2,10,11,1,14,4,15,16,17,3,7
Tad1-4_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-4_PGr	AAWC01002745	Puccinia graminis	Tad1	6,2,9,10,11,1,4,15,16,17,3,7
Tad1-4_PTrit	.	Puccinia triticina	Tad1	
Tad1-4_TMe	.	Tuber melanosporum	Tad1	6,2,18,7,9,10,11,13,1,4,5,17,3
Tad1-4_TSt	.	Talaromyces stipitatus	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-40_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-40_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-41_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-41_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-41B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-42_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-42_EP	.	Erysiphe pisi	Tad1	1,17,9,3,10,6,7,13
Tad1-43_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-43_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-43B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-44_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-44_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-45_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-45B_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-46_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-46B_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-46C_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-46D_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-47_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-48_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-48_EP	.	Erysiphe pisi	Tad1	6,2,7,9,10,11,1,4,15,16,17,3
Tad1-49_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-49_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-5_ACap	.	Histoplasma capsulatum	Tad1	6,2,9,8,10,11,12,13,1,4,15,16,17,3,7
Tad1-5_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-5_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-5_CGlo	.	Chaetomium globosum	Tad1	2,8,11,12,42,1,14,4,15,16,17,3
Tad1-5_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-5_PGr	AAWC01000958	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-5_TMe	.	Tuber melanosporum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-5_TSt	.	Talaromyces stipitatus	Tad1	
Tad1-50_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-50_EP	.	Erysiphe pisi	Tad1	
Tad1-51_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-52_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-52_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-53_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-53_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-53B_BG	.	Blumeria graminis	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-53C_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-54_EP	.	Erysiphe pisi	Tad1	16,1,3,11,4,2,15
Tad1-55_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7

Tad1-55_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-56_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-56_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-57_BG	.	Blumeria graminis	Tad1	16,1,8,3,11,4,12,2,15	
Tad1-57_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-58_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-58_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-59_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-59_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-5B_TMe	.	Tuber melanosporum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-6_ACap	.	Histoplasma capsulatum	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-6_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-6_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15	
Tad1-6_CGlo	.	Chaetomium globosum	Tad1	1,3,11,4,2	
Tad1-6_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-6_PGr	.	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3	
Tad1-6_TSt	.	Talaromyces stipitatus	Tad1	7,1,17,9,16,3,10,11,4,6,2,15	
Tad1-60_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-60_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-61_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-61_EP	.	Erysiphe pisi	Tad1	16,1,8,3,11,4,12,2,15	
Tad1-62_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-62_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-63_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-63_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-63B_BG	.	Blumeria graminis	Tad1	7,1,17,9,3,10,4,6,2,13	
Tad1-63C_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-64_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-64_EP	.	Erysiphe pisi	Tad1	7,1,17,9,3,10,4,6,2,13	
Tad1-65_BG	.	Blumeria graminis	Tad1		
Tad1-65_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-65B_BG	.	Blumeria graminis	Tad1	7,1,17,9,3,10,6,2,13	
Tad1-66_BG	.	Blumeria graminis	Tad1	6,2,241,9,10,11,13,1,4,15,16,17,3,7	
Tad1-66_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-66B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-67_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-67_EP	.	Erysiphe pisi	Tad1	16,1,3,11,4,2,15	
Tad1-68_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-68_EP	.	Erysiphe pisi	Tad1	16,1,3,11,4,2,15	
Tad1-69_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-69_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-69B_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-69C_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-7_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,239,7	
Tad1-7_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15	
Tad1-7_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7	
Tad1-7_PGr	.	Puccinia graminis	Tad1	6,2,9,10,11,13,1,14,4,15,16,17,3,7	
Tad1-7_TSt	.	Talaromyces stipitatus	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3	
Tad1-70_BG	.	Blumeria graminis	Tad1	16,1,8,3,11,4,12,2,15	
Tad1-70_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,1,4,15,16,17,3,7	
Tad1-70B_BG	.	Blumeria graminis	Tad1	16,1,3,11,4,2,15	

Tad1-71_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-71_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-72_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-72_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-73_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-73_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-74_BG	.	Blumeria graminis	Tad1	
Tad1-74_EP	.	Erysiphe pisi	Tad1	6,2,7,9,10,11,1,4,15,16,17,3
Tad1-75_BG	.	Blumeria graminis	Tad1	16,1,3,11,4,2,15
Tad1-75_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-76_BG	.	Blumeria graminis	Tad1	7,1,17,9,3,10,4,6,2,13
Tad1-76_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-77_BG	.	Blumeria graminis	Tad1	17,9,10,6,7,13
Tad1-77_EP	.	Erysiphe pisi	Tad1	16,1,3,4,11,2,15
Tad1-78_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-79_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-7B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-7B_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-8_BG	.	Blumeria graminis	Tad1	6,2,7,9,10,11,13,1,4,15,16,17,3
Tad1-8_BoBo	.	Botryobasidium botryosum	Tad1	
Tad1-8_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-8_PGr	.	Puccinia graminis	Tad1	6,2,7,9,10,11,13,1,14,4,15,16,17,3
Tad1-8_TSt	.	Talaromyces stipitatus	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-80_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-81_EP	.	Erysiphe pisi	Tad1	7,16,1,17,3,10,11,4,6,2,15
Tad1-82_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-83_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-84_EP	.	Erysiphe pisi	Tad1	16,11,15
Tad1-85_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-86_EP	.	Erysiphe pisi	Tad1	16,1,8,3,11,4,12,2,15
Tad1-87_EP	.	Erysiphe pisi	Tad1	7,1,17,9,3,10,4,6,2
Tad1-88_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-89_EP	.	Erysiphe pisi	Tad1	16,11,15
Tad1-9_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-9_BoBo	.	Botryobasidium botryosum	Tad1	16,1,3,11,4,2,15
Tad1-9_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-9_PGr	.	Puccinia graminis	Tad1	7,1,17,9,3,10,6,2,13
Tad1-9_TSt	.	Talaromyces stipitatus	Tad1	16,1,17,3,11,4,2,15
Tad1-90_EP	.	Erysiphe pisi	Tad1	2,8,11,12,42,1,14,4,15,16,3
Tad1-91_EP	.	Erysiphe pisi	Tad1	7,1,17,16,3,11,4,2,15
Tad1-92_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-93_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-94_EP	.	Erysiphe pisi	Tad1	
Tad1-95_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-96_EP	.	Erysiphe pisi	Tad1	2,8,11,12,42,1,14,4,15,16,3
Tad1-97_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-98_EP	.	Erysiphe pisi	Tad1	16,1,3,11,4,2,15
Tad1-99_EP	.	Erysiphe pisi	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-9B_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
Tad1-9C_BG	.	Blumeria graminis	Tad1	6,2,9,10,11,13,1,4,15,16,17,3,7
TAHRE	AJ5425818	Drosophila melanogaster	Jockey	16,1,87,88,3,11,4,2,15

TAL	.	Talpidae	SINE2/tRNA		
TAL-1	AB108790, AB108813,	Mogera robusta	SINE2/tRNA		
Talua	.	Reticulitermes lucifugus	SINE2/tRNA		
Talub	.	Termitoidae	SINE2/tRNA		
Taluc	.	Termitoidae	SINE2/tRNA		
Talud	.	Termitoidae	SINE2/tRNA		
TAR2_TA	AB012296	Triticum aestivum	L1	16,1,3,11,4,2,15	
TART_B1	U14101	Drosophila melanogaster	Jockey	16,1,87,3,88,11,4,2,15	
TART_DV	AY219709	Drosophila virilis	Jockey	1,3,4,2	
TART-1_DWi	.	Drosophila willistoni	Jockey	16,1,87,88,3,11,4,2,15	
TART-A	AJ566116	Drosophila melanogaster	Jockey	1,4,2	
TDD3	AF002669	Dictyostelium discoideum	L1		
TE-18_LVa	.	Litopenaeus vannamei	Penelope		
TE-3_XT	.	Xenopus tropicalis	Penelope		
TE-4_LVa	.	Litopenaeus vannamei	Jockey		
TE-5_LVa	.	Litopenaeus vannamei	Penelope		
TE-6_LVa	.	Litopenaeus vannamei	Penelope		
TguSINE1	.	Estrildidae	SINE2/tRNA		
TguSINE1_Ucy	.	Uraeginthus cyanocephalus	SINE2/tRNA		
THER1	.	Mammalia	SINE2/tRNA		
THER1	.	Mammalia	SINE2/tRNA		
THER1	.	Mammalia	SINE2/tRNA		
THER1_MD	.	Monodelphis domestica	SINE2/tRNA		
THER2	.	Mammalia	SINE2/tRNA		
THER2	.	Mammalia	SINE2/tRNA		
THER2	.	Mammalia	SINE2/tRNA		
THER2_MD	.	Monodelphis domestica	SINE2/tRNA		
THER2_ME	.	Notamacropus eugenii	SINE2/tRNA		
Togen-1_CCa	.	Cyprinus carpio	Non-LTRRetrotransposo	1,3,2	
Togen-1_DR	.	Danio rerio	Non-LTRRetrotransposo	1,3,4,2	
Togen-1_SSa	.	Salmo salar	Non-LTRRetrotransposo		
TP2	X52770	Physarum polycephalum	Non-LTRRetrotransposo		
TRAS-1_HMM	.	Heliconius melpomene melpomene	R1	7,1,17,9,10,4,6,2,13	
TRAS-1_MSe	.	Manduca sexta	R1	6,2,9,10,11,86,1,4,15,16,17,3,7	
TRAS-1_PXu	.	Papilio xuthus	R1		
TRAS-1_TUr	scaffold_45	Tetranychus urticae	R1	6,2,9,10,11,13,1,4,15,16,17,3,7	
TRAS-2_PXu	.	Papilio xuthus	R1	6,2,9,10,11,86,13,1,4,15,16,17,3,7	
TRAS-2_TUr	.	Tetranychus urticae	R1	6,2,7,8,9,10,11,12,13,1,4,15,16,17,3	
TRAS-3_TUr	.	Tetranychus urticae	R1	6,2,9,10,11,13,1,4,15,16,17,3,7	
TRAS-4_TUr	.	Tetranychus urticae	R1	6,2,9,10,11,13,1,4,15,16,17,3,7	
TRAS-5_TUr	.	Tetranychus urticae	R1	6,2,7,8,9,10,11,12,13,1,14,4,15,16,17,3	
TRAS-6_TUr	.	Tetranychus urticae	R1		
TRAS-7_TUr	.	Tetranychus urticae	R1	6,2,9,8,10,11,12,13,1,14,4,15,16,17,3,7	
TRAS-8_TUr	.	Tetranychus urticae	R1		
TRAS1	D38414	Bombyx mori	R1	6,2,7,9,10,11,86,13,1,4,15,16,17,3	
TRAS1_R1_Hmel 133,130,134,135,136,000	.	Heliconius melpomene	R1		
TRAS3_BM	AB046668	Bombyx mori	R1	6,2,9,10,11,86,13,1,4,15,16,17,3,7	
TRAS3_SC	AB046673	Samia cynthia	R1		
TRAS4_SC	AB046674	Samia cynthia	R1		
TRAS6_BM	AB046671	Bombyx mori	R1		

TRAS9_SC	AB046675	Samia cynthia	R1	1,3,11,4,2
TRE3B	AF134170	Dictyostelium discoideum	L1	16,1,3,24,4,11,2
TRE3C	AF134171	Dictyostelium discoideum	L1	1,3,11,4,2
TRE3D	NW_633189	Dictyostelium discoideum	L1	1,3,11,4,2
TREP20	AY013246	Hordeum vulgare	L1	
TREP51	AJ270053	Hordeum vulgare	L1	
TREP52	AJ270056	Hordeum vulgare	L1	
TREP53	AJ270054	Hordeum vulgare	L1	
TREP58	AJ270055	Hordeum marinum	L1	
TREP80	Y13390	Secale cereale	L1	
TrepLINE1	.	Trifolium repens	L1	43,2,8,44,24,11,12,1,77,4,16,75,3,76
TRIM	X59239	Drosophila miranda	Loa	
TS	D17453	Nicotiana tabacum	SINE	
TS2	.	Solanum demissum	SINE	
TSCL	.	Solanum demissum	SINE	
tSINE_Fc	U65470	Arabidopsis thaliana	Non-LTRRetrotransposo	
Tu-III	.	Felis catus	SINE2/tRNA	
TUB1	AB090261AB090262AB09	Tupaia glis	SINE2/tRNA	
TUB2	.	Tupaia belangeri	SINE2/tRNA	
TUB2B	.	Tupaia belangeri	SINE2/tRNA	
TUB3	.	Tupaia belangeri	SINE2/tRNA	
TUB4	.	Tupaia belangeri	SINE2/tRNA	
TUBE1	.	Tupaia belangeri	SINE2/tRNA	
TUBE2	.	Tupaia belangeri	SINE2/tRNA	
TUBE3	.	Tupaia belangeri	SINE2/tRNA	
TUBE4	.	Tupaia belangeri	SINE2/tRNA	
TUBE5	.	Tupaia belangeri	SINE2/tRNA	
TUS	.	Tupaia belangeri	SINE2/tRNA	
TWIN	.	Culex pipiens	SINE2/tRNA	
TWIN2_CP	AF282725	Culex pipiens	SINE2/tRNA	
twinID-Spe	.	Spermophilus	SINE2/tRNA	
TX1_FR1	.	Takifugu rubripes	L1	
Tx1_FR1B	.	Takifugu rubripes	Tx1	1,26,45,3,11,4,2
TX1_FR2	.	Takifugu rubripes	L1	
Tx1_XT	.	Xenopus tropicalis	Tx1	1,26,8,14,3,11,4,12,2
Tx1-1_AAe	.	Aedes aegypti	Tx1	2,8,24,11,12,42,1,14,4,16,3
Tx1-1_AC	.	Aplysia californica	Tx1	16,1,26,3,24,4,11,2
Tx1-1_ACar	.	Anolis carolinensis	Tx1	16,1,3,46,24,4,11,12,2
Tx1-1_ADi	.	Acropora digitifera	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-1_ASi	.	Alligator sinensis	Tx1	16,1,26,8,3,24,11,4,12,2
Tx1-1_BF	.	Branchiostoma floridae	Tx1	43,2,8,44,11,12,1,26,14,4,3
Tx1-1_CGi	.	Crassostrea gigas	Tx1	
150,43,2,44,11,1,26,151,152,4,153,15,16,3,154				
Tx1-1_CM	.	Callorhinchus milii	Tx1	
Tx1-1_CPB	.	Chrysemys picta bellii	Tx1	2,8,24,11,12,42,1,26,14,4,16,3
Tx1-1_CQ	.	Culex quinquefasciatus	Tx1	1,8,14,3,11,4,12,2
Tx1-1_Crp	.	Crocodylus porosus	Tx1	16,1,26,8,3,24,4,11,12,2
Tx1-1_DPu	.	Daphnia pulex	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-1_DR	.	Danio rerio	Tx1	1,8,26,3,11,4,12,2
Tx1-1_FoCa	.	Folsomia candida	Tx1	16,1,26,8,3,24,11,4,12,2

Tx1-1_Gav	.	Gavialis gangeticus	Tx1	1,8,11,12,2
Tx1-1_HM	.	Hydra vulgaris	Tx1	1,26,3,11,4,2
Tx1-1_LCh	.	Latimeria chalumnae	Tx1	2,8,24,11,12,1,26,14,4,16,3,46
Tx1-1_LMi	.	Locusta migratoria	Tx1	16,1,3,24,11,4,2
Tx1-1_NV	.	Nematostella vectensis	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-1_PM	.	Petromyzon marinus	Tx1	2,8,11,12,42,1,26,14,4,3
Tx1-1_PSi	.	Pelodiscus sinensis	Tx1	
Tx1-1_SK	.	Saccoglossus kowalevskii	Tx1	43,2,21,23,44,24,11,1,26,4,30,16,3
Tx1-1_SP	.	Strongylocentrotus purpuratus	Tx1	1,43,3,44,4,2
Tx1-1_SSa	.	Salmo salar	Tx1	2,8,24,11,12,42,1,26,14,4,16,3
Tx1-10_AMi	.	Alligator mississippiensis	Tx1	
Tx1-10_BF	.	Branchiostoma floridae	Tx1	1,26,43,3,44,11,4,2
Tx1-10_CGi	.	Crassostrea gigas	Tx1	138,2,11,1,26,139,140,4,15,16,3,141
Tx1-10_CQ	.	Culex quinquefasciatus	Tx1	92,16,1,8,3,24,11,4,12,2
Tx1-10_Crp	.	Crocodylus porosus	Tx1	
Tx1-10_DR	.	Danio rerio	Tx1	2,8,24,11,12,42,1,26,14,4,16,3
Tx1-10_LCh	.	Latimeria chalumnae	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46
Tx1-10_LMi	.	Locusta migratoria	Tx1	16,1,26,3,24,11,4,2
Tx1-10_SK	.	Saccoglossus kowalevskii	Tx1	16,1,26,3,24,11,4,2
Tx1-10_SSa	.	Salmo salar	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-10B_Crp	.	Crocodylus porosus	Tx1	
Tx1-11_AMi	.	Alligator mississippiensis	Tx1	
Tx1-11_BF	.	Branchiostoma floridae	Tx1	43,2,21,44,24,11,1,26,4,16,3
Tx1-11_CGi	.	Crassostrea gigas	Tx1	
150,43,2,44,11,1,26,151,152,4,153,15,16,3,154				
Tx1-11_CQ	.	Culex quinquefasciatus	Tx1	2,1
Tx1-11_Crp	.	Crocodylus porosus	Tx1	
Tx1-11_DR	.	Danio rerio	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-11_LCh	.	Latimeria chalumnae	Tx1	16,1,26,3,24,11,4,2
Tx1-11_LMi	.	Locusta migratoria	Tx1	16,1,26,3,24,11,4,2
Tx1-11_SK	.	Saccoglossus kowalevskii	Tx1	1,26,43,3,44,11,4,2
Tx1-11_SSa	.	Salmo salar	Tx1	1,43,44,4,2
Tx1-12_AMi	.	Alligator mississippiensis	Tx1	
Tx1-12_BF	.	Branchiostoma floridae	Tx1	43,2,44,11,12,1,26,14,4,15,16,3
Tx1-12_CGi	.	Crassostrea gigas	Tx1	16,1,43,26,3,44,4,15,11,2
Tx1-12_CQ	.	Culex quinquefasciatus	Tx1	16,1,26,3,24,11,4,2
Tx1-12_Crp	.	Crocodylus porosus	Tx1	16,1,3,24,11,4,2
Tx1-12_DR	.	Danio rerio	Tx1	21,1,26,23,43,3,44,11,4,2
Tx1-12_Lch	.	Latimeria chalumnae	Tx1	43,2,8,44,11,12,42,1,14,4,3,46
Tx1-12_LMi	.	Locusta migratoria	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-12_SK	.	Saccoglossus kowalevskii	Tx1	1,8,26,14,3,11,4,12,2
Tx1-12_SSa	.	Salmo salar	Tx1	2,8,11,12,42,1,26,14,4,15,16,3
Tx1-13_BF	.	Branchiostoma floridae	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
Tx1-13_CGi	.	Crassostrea gigas	Tx1	16,1,26,3,24,11,4,2
Tx1-13_CQ	.	Culex quinquefasciatus	Tx1	1,3,11,4,2
Tx1-13_DR	.	Danio rerio	Tx1	1,43,3,44,11,4,2
Tx1-13_Lch	.	Latimeria chalumnae	Tx1	
Tx1-13_LMi	.	Locusta migratoria	Tx1	16,1,26,43,44,24,11,4,2
Tx1-13_SK	.	Saccoglossus kowalevskii	Tx1	1,26,3,11,4,2
Tx1-13_SSa	.	Salmo salar	Tx1	
Tx1-14_BF	.	Branchiostoma floridae	Tx1	

Tx1-14_CGi	.	Crassostrea gigas	Tx1	1,3,11,4,2
Tx1-14_CQ	.	Culex quinquefasciatus	Tx1	93,16,1,26,3,24,11,4,92,2
Tx1-14_DR	.	Danio rerio	Tx1	1,26,43,3,44,11,4,2
Tx1-14_LMi	.	Locusta migratoria	Tx1	16,1,43,3,44,24,11,4,2
Tx1-14_SK	.	Saccoglossus kowalevskii	Tx1	1,43,3,44,11,4,2
Tx1-14_SSa	.	Salmo salar	Tx1	1,26,11,2
Tx1-14a_Lch	.	Latimeria chalumnae	Tx1	
Tx1-14b_Lch	.	Latimeria chalumnae	Tx1	
Tx1-15_BF	.	Branchiostoma floridae	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-15_CQ	.	Culex quinquefasciatus	Tx1	16,8,24,11,12
Tx1-15_DR	.	Danio rerio	Tx1	1,43,3,44,4,2
Tx1-15_Lch	.	Latimeria chalumnae	Tx1	
Tx1-15_LMi	.	Locusta migratoria	Tx1	1,8,3,11,4,12,2
Tx1-15_SK	.	Saccoglossus kowalevskii	Tx1	1,43,3,44,4,2
Tx1-16_BF	.	Branchiostoma floridae	Tx1	1,43,23,3,44,4,2
Tx1-16_DR	.	Danio rerio	Tx1	43,2,21,23,44,24,11,1,26,4,16,3
Tx1-16_SK	.	Saccoglossus kowalevskii	Tx1	1,43,3,44,11,4,2
Tx1-16a_Lch	.	Latimeria chalumnae	Tx1	
Tx1-16b_Lch	.	Latimeria chalumnae	Tx1	
Tx1-17_BF	.	Branchiostoma floridae	Tx1	16,1,26,43,3,44,11,4,2,15
Tx1-17_DR	.	Danio rerio	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-17_Lch	.	Latimeria chalumnae	Tx1	
Tx1-17_SK	.	Saccoglossus kowalevskii	Tx1	1,43,3,44,4,2
Tx1-18_BF	.	Branchiostoma floridae	Tx1	2,1,8,26,14,3,11,4,12
Tx1-18_DR	.	Danio rerio	Tx1	21,1,26,23,43,3,44,11,4,2
Tx1-18_Lch	.	Latimeria chalumnae	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46
Tx1-18_SK	.	Saccoglossus kowalevskii	Tx1	1,43,3,44,4,2
Tx1-19_DR	.	Danio rerio	Tx1	43,2,21,23,44,11,1,26,4,15,16,3
Tx1-19_SK	.	Saccoglossus kowalevskii	Tx1	26,11
Tx1-19a_Lch	.	Latimeria chalumnae	Tx1	43,2,8,44,11,12,1,14,4,3,46
Tx1-19b_Lch	.	Latimeria chalumnae	Tx1	43,2,8,44,11,12,1,14,4,3,46
Tx1-1B_CM	.	Callorhinchus milii	Tx1	
Tx1-1B_CPB	.	Chrysemys picta bellii	Tx1	2,8,24,11,12,1,26,14,4,16,3
Tx1-1B_PM	.	Petromyzon marinus	Tx1	2,8,11,12,42,1,26,14,4,3
Tx1-1B_PSi	.	Pelodiscus sinensis	Tx1	16,1,26,3,24,11,4,2
Tx1-1C_CPB	.	Chrysemys picta bellii	Tx1	
Tx1-1N1_ACar	.	Anolis carolinensis	Tx1	46,12
Tx1-2_AAe	.	Aedes aegypti	Tx1	2,8,24,11,12,42,1,14,4,16,3
Tx1-2_AMi	.	Alligator mississippiensis	Tx1	1,26,3,11,4,2
Tx1-2_ASi	.	Alligator sinensis	Tx1	1,26,8,3,11,4,12,2
Tx1-2_BF	.	Branchiostoma floridae	Tx1	21,1,26,23,3,11,4,2
Tx1-2_CGi	.	Crassostrea gigas	Tx1	2,16,1,8,26,3,11,4,12,15
Tx1-2_CPB	.	Chrysemys picta bellii	Tx1	
Tx1-2_CQ	.	Culex quinquefasciatus	Tx1	2,8,24,11,12,1,14,4,16,3
Tx1-2_Crp	.	Crocodylus porosus	Tx1	42,8,14,46,12
Tx1-2_DR	.	Danio rerio	Tx1	21,1,26,23,3,11,4,2
Tx1-2_FoCa	.	Folsomia candida	Tx1	42,1,26,8,14,3,11,4,12,2
Tx1-2_LCh	.	Latimeria chalumnae	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
Tx1-2_LMi	.	Locusta migratoria	Tx1	16,1,26,3,24,11,4,2
Tx1-2_NV	.	Nematostella vectensis	Tx1	
Tx1-2_PM	.	Petromyzon marinus	Tx1	1,26,3,11,4,2

Tx1-2_SK	.	Saccoglossus kowalevskii	Tx1	1,26,3,11,4,2
Tx1-2_SPur	.	Strongylocentrotus purpuratus	Tx1	1,26,43,3,44,11,4,2
Tx1-2_SSa	.	Salmo salar	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-2_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,24,11,12,1,26,14,4,16,3
Tx1-20_DR	.	Danio rerio	Tx1	21,1,26,43,3,44,11,4,2
Tx1-20_SK	.	Saccoglossus kowalevskii	Tx1	1,43,3,44,11,4,2
Tx1-21_DR	.	Danio rerio	Tx1	1,26,11,2
Tx1-21_SK	.	Saccoglossus kowalevskii	Tx1	1,8,26,14,43,3,44,11,4,12,2
Tx1-22_DR	.	Danio rerio	Tx1	1,43,3,44,11,4,2
Tx1-23_DR	.	Danio rerio	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-24_DR	.	Danio rerio	Tx1	1,3,4,2
Tx1-25_DR	.	Danio rerio	Tx1	1,26,8,14,3,11,4,12,2
Tx1-26_DR	.	Danio rerio	Tx1	1,26,8,14,3,4,11,12,2
Tx1-27_DR	.	Danio rerio	Tx1	1,43,3,44,2
Tx1-28_DR	.	Danio rerio	Tx1	44,43
Tx1-29_DR	.	Danio rerio	Tx1	21,1,43,23,3,44,11,4,2
Tx1-2B_SSa	.	Salmo salar	Tx1	1,43,44,11,4,2
Tx1-3_AAe	.	Aedes aegypti	Tx1	2,1,8,3,11,4,12
Tx1-3_AMi	.	Alligator mississippiensis	Tx1	16,1,26,8,3,24,4,11,12,2
Tx1-3_ASi	.	Alligator sinensis	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46
Tx1-3_BF	.	Branchiostoma floridae	Tx1	43,2,21,23,44,24,11,1,26,4,16,3
Tx1-3_CGi	.	Crassostrea gigas	Tx1	21,1,26,3,11,4,2
Tx1-3_CPB	.	Chrysemys picta bellii	Tx1	43,2,8,44,24,11,12,1,26,14,4,16
Tx1-3_CQ	.	Culex quinquefasciatus	Tx1	2,8,11,12,93,1,14,4,3
Tx1-3_DR	.	Danio rerio	Tx1	21,1,26,3,11,4,2
Tx1-3_FoCa	.	Folsomia candida	Tx1	16,1,26,3,24,11,4,2
Tx1-3_FR	.	Takifugu rubripes	Tx1	2,8,11,12,42,1,26,14,4,3
Tx1-3_LCh	.	Latimeria chalumnae	Tx1	16,1,26,43,3,44,46,24,11,4,2
Tx1-3_LMi	.	Locusta migratoria	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-3_PM	.	Petromyzon marinus	Tx1	2,8,24,11,12,1,26,4,16,3
Tx1-3_SK	.	Saccoglossus kowalevskii	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-3_SPur	.	Strongylocentrotus purpuratus	Tx1	1,43,3,44,11,4,2
Tx1-3_SSa	.	Salmo salar	Tx1	93,1,8,14,26,3,11,4,12,2
Tx1-3_XT	.	Xenopus tropicalis	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46
Tx1-30_DR	.	Danio rerio	Tx1	1,3,2
Tx1-31_DR	.	Danio rerio	Tx1	1,3,4,2
Tx1-32_DR	.	Danio rerio	Tx1	1,26,3,11,4,2
Tx1-33_DR	.	Danio rerio	Tx1	1,26,3,11,4,2
Tx1-34_DR	.	Danio rerio	Tx1	2,8,24,11,12,1,26,14,4,16,3
Tx1-35_DR	.	Danio rerio	Tx1	
Tx1-36_DR	.	Danio rerio	Tx1	44,43
Tx1-37_DR	.	Danio rerio	Tx1	44,43
Tx1-38_DR	.	Danio rerio	Tx1	21,1,43,23,3,44,11,4,30,2
Tx1-39_DR	.	Danio rerio	Tx1	93,1,26,3,11,4,2
Tx1-4_AAe	.	Aedes aegypti	Tx1	92,2,8,11,12,42,1,14,4,3
Tx1-4_AMi	.	Alligator mississippiensis	Tx1	43,2,8,44,11,12,42,1,26,14,4,3,46
Tx1-4_BF	.	Branchiostoma floridae	Tx1	43,2,44,24,11,12,1,26,14,4,16,3,46
Tx1-4_CGi	.	Crassostrea gigas	Tx1	138,2,24,11,1,26,14,139,140,4,16,141
Tx1-4_CPB	.	Chrysemys picta bellii	Tx1	1,8,3,46,12,2
Tx1-4_CQ	.	Culex quinquefasciatus	Tx1	2,8,24,11,12,42,1,14,4,16,3
Tx1-4_DR	.	Danio rerio	Tx1	21,1,26,43,3,44,11,4,2

Tx1-4_FoCa	.	Folsomia candida	Tx1	16,1,8,3,24,11,4,12,2
Tx1-4_FR	.	Takifugu rubripes	Tx1	1,26,3,11,4,2
Tx1-4_LCh	.	Latimeria chalumnae	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
Tx1-4_LMi	.	Locusta migratoria	Tx1	16,1,43,44,24,11,2
Tx1-4_PM	.	Petromyzon marinus	Tx1	1,26,3,11,4,2
Tx1-4_SK	.	Saccoglossus kowalevskii	Tx1	43,2,21,44,24,11,1,26,4,16,3
Tx1-4_SPur	.	Strongylocentrotus purpuratus	Tx1	16,1,26,43,3,44,82,24,11,4,2
Tx1-4_SSa	.	Salmo salar	Tx1	2,21,23,11,1,26,4,15,16,3
Tx1-40_DR	.	Danio rerio	Tx1	
Tx1-41_DR	.	Danio rerio	Tx1	1,26,8,14,3,11,4,12,2
Tx1-42_DR	.	Danio rerio	Tx1	
Tx1-43_DR	.	Danio rerio	Tx1	1,8,26,3,11,4,12,2
Tx1-44_DR	.	Danio rerio	Tx1	1,26,3,11,4,2
Tx1-45_DR	.	Danio rerio	Tx1	16,1,26,3,24,11,4,2
Tx1-46_DR	.	Danio rerio	Tx1	44,43
Tx1-47_DR	.	Danio rerio	Tx1	2,8,11,12,42,1,26,14,4,3
Tx1-48_DR	.	Danio rerio	Tx1	2,8,11,12,42,1,26,14,4,3
Tx1-49_DR	.	Danio rerio	Tx1	1,3,2
Tx1-4B_Crp	.	Crocodylus porosus	Tx1	
Tx1-5_AAe	.	Aedes aegypti	Tx1	2,8,24,11,12,42,1,26,14,4,16,3
Tx1-5_AMi	.	Alligator mississippiensis	Tx1	16,1,26,3,24,11,4,2
Tx1-5_BF	.	Branchiostoma floridae	Tx1	2,8,24,11,12,1,26,14,4,16,3,46
Tx1-5_CGi	.	Crassostrea gigas	Tx1	16,1,43,26,3,44,11,4,2,15
Tx1-5_CQ	.	Culex quinquefasciatus	Tx1	16,1,3,24,11,4,12,92,2
Tx1-5_Crp	.	Crocodylus porosus	Tx1	1,8,26,3,11,4,12,2
Tx1-5_DR	.	Danio rerio	Tx1	21,1,26,23,3,11,4,2
Tx1-5_FR	.	Takifugu rubripes	Tx1	1,26,43,3,44,11,4,2
Tx1-5_LCh	.	Latimeria chalumnae	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
Tx1-5_LMi	.	Locusta migratoria	Tx1	92,2,8,11,12,93,42,1,26,14,4,3
Tx1-5_PM	.	Petromyzon marinus	Tx1	16,1,26,3,24,11,4,2
Tx1-5_SK	.	Saccoglossus kowalevskii	Tx1	43,155,2,23,44,11,1,26,4,15,16,156,3
Tx1-5_SPur	.	Strongylocentrotus purpuratus	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-5_SSa	.	Salmo salar	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-50_DR	.	Danio rerio	Tx1	2,8,24,11,12,1,26,14,4,16,3
Tx1-51_DR	.	Danio rerio	Tx1	
Tx1-52_DR	.	Danio rerio	Tx1	21,1,26,43,3,44,11,4,2
Tx1-53_DR	.	Danio rerio	Tx1	
Tx1-54_DR	.	Danio rerio	Tx1	1,26,43,44,11,2
Tx1-55_DR	.	Danio rerio	Tx1	16,1,26,43,8,3,44,24,11,4,12,2
Tx1-56_DR	.	Danio rerio	Tx1	2,8,24,11,12,1,26,14,4,16,3
Tx1-57_DR	.	Danio rerio	Tx1	1,43,3,44,11,2
Tx1-58_DR	.	Danio rerio	Tx1	1,8,26,14,3,11,4,12,2
Tx1-6_AAe	.	Aedes aegypti	Tx1	16,1,26,3,24,11,4,2
Tx1-6_AMi	.	Crocodylidae	Tx1	42,8,14,46,12
Tx1-6_BF	.	Branchiostoma floridae	Tx1	2,8,24,11,12,42,1,26,14,4,16,3,46
Tx1-6_CGi	.	Crassostrea gigas	Tx1	21,1,198,26,43,3,44,11,4,2
Tx1-6_CQ	.	Culex quinquefasciatus	Tx1	2,8,24,11,12,42,1,14,4,16,3
Tx1-6_Crp	.	Crocodylus porosus	Tx1	
Tx1-6_DR	.	Danio rerio	Tx1	1,26,43,3,44,11,4,2
Tx1-6_LCh	.	Latimeria chalumnae	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-6_LMi	.	Locusta migratoria	Tx1	16,1,26,3,24,11,4,2

Tx1-6_SK	.	Saccoglossus kowalevskii	Tx1	43,2,21,23,44,24,11,1,26,4,16,3
Tx1-6_SPur	.	Strongylocentrotus purpuratus	Tx1	1,26,43,3,44,11,4,2
Tx1-6_SSa	.	Salmo salar	Tx1	2,8,24,11,12,1,26,14,4,16,3
Tx1-6B_AMi	.	Alligator mississippiensis	Tx1	
Tx1-7_AMi	.	Crocodylidae	Tx1	
Tx1-7_BF	.	Branchiostoma floridae	Tx1	2,8,24,11,12,42,1,26,14,4,16,182,3,46
Tx1-7_CGi	.	Crassostrea gigas	Tx1	16,1,26,3,24,11,4,2
Tx1-7_CQ	.	Culex quinquefasciatus	Tx1	2,1,8,3,11,4,12,92
Tx1-7_Crp	.	Crocodylus porosus	Tx1	
Tx1-7_DR	.	Danio rerio	Tx1	21,1,26,23,3,11,4,30,2
Tx1-7_LCh	.	Latimeria chalumnae	Tx1	2,1,43,3,44,46,11,4,12
Tx1-7_LMi	.	Locusta migratoria	Tx1	16,1,26,3,24,11,4,2
Tx1-7_SK	.	Saccoglossus kowalevskii	Tx1	1,26,3,11,4,2
Tx1-7_SSa	.	Salmo salar	Tx1	43,2,21,23,44,11,1,26,4,3
Tx1-8_AMi	.	Alligator mississippiensis	Tx1	16,1,8,26,3,24,11,4,12,2
Tx1-8_BF	.	Branchiostoma floridae	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-8_CGi	.	Crassostrea gigas	Tx1	138,2,11,1,26,139,140,4,15,16,3,141
Tx1-8_CQ	.	Culex quinquefasciatus	Tx1	42,1,8,14,3,11,4,12,2
Tx1-8_Crp	.	Crocodylus porosus	Tx1	
Tx1-8_DR	.	Danio rerio	Tx1	1,26,3,11,4,2
Tx1-8_LCh	.	Latimeria chalumnae	Tx1	43,2,8,44,11,12,42,1,14,4,3,46
Tx1-8_LMi	.	Locusta migratoria	Tx1	142,16,1,26,3,24,11,4,2
Tx1-8_SK	.	Saccoglossus kowalevskii	Tx1	1,26,43,3,44,11,4,2
Tx1-8_SSa	.	Salmo salar	Tx1	21,1,26,43,3,44,4,11,2
Tx1-9_AMi	.	Crocodylidae	Tx1	1,26,8,3,4,11,12,2
Tx1-9_BF	.	Branchiostoma floridae	Tx1	16,1,26,43,3,44,24,11,4,2
Tx1-9_CGi	.	Crassostrea gigas	Tx1	1,3,4,2
Tx1-9_CQ	.	Culex quinquefasciatus	Tx1	1,3,11,4,2
Tx1-9_DR	.	Danio rerio	Tx1	43,2,8,44,11,12,1,26,14,4,3
Tx1-9_LMi	.	Locusta migratoria	Tx1	16,1,8,26,3,24,11,4,12,2
Tx1-9_SK	.	Saccoglossus kowalevskii	Tx1	43,2,21,23,44,24,11,1,26,4,16,3
Tx1-9_SSa	.	Salmo salar	Tx1	43,2,8,44,24,11,12,42,1,26,14,4,16,3,46
Tx1-N1_CQ	.	Culex quinquefasciatus	Tx1	8,14,12,92
Tx1-Sat-1_SK	.	Saccoglossus kowalevskii	Tx1	16,1,26,8,3,4,11,12,2,15
Tx1-TGTA-1_SK	.	Saccoglossus kowalevskii	Tx1	16,1,26,3,11,4,2,15
Tx1-U5-1_SK	.	Saccoglossus kowalevskii	Tx1	1,26,43,3,44,11,4,2
Tx1-U5-2_SK	.	Saccoglossus kowalevskii	Tx1	43,2,44,11,12,1,26,14,4,3,46
UCON13	.	Euteleostomi	Penelope	
UCON3	.	Euteleostomi	SINE2/tRNA	
UCON3_CM	.	Callorhinchus milii	SINE2/tRNA	
UCON49	.	Mammalia	L2	
UCON49	.	Mammalia	L2	
UCON49	.	Mammalia	L2	
UCON55	.	Mammalia	SINE2/tRNA	
UCON55	.	Mammalia	SINE2/tRNA	
UCON55	.	Mammalia	SINE2/tRNA	
UCON82	.	Mammalia	RTE	
UCON82	.	Mammalia	RTE	
UCON82	.	Mammalia	RTE	
UCON86	.	Mammalia	L2	
UCON86	.	Mammalia	L2	

UCON86	.	Mammalia	L2	
UnaL2	.	Anguilla japonica	L2	16,1,3,24,4,11,2
UnaSINE1	.	Anguilla japonica	SINE2/tRNA	
UnaSINE2	.	Anguilla japonica	SINE2/tRNA	
Utopia-1_ACa	.	Acanthamoeba castellanii	NeSL	18,5,1,19,3,4,2
Utopia-1_ACar	scaffold_474	Anolis carolinensis	NeSL	1,3,4,2
Utopia-1_AEc	.	Acromyrmex echinatior	NeSL	1,3,4,2
Utopia-1_AMi	.	Alligator mississippiensis	NeSL	1,3,4,2
Utopia-1_APi	.	Acyrthosiphon pisum	NeSL	1,3,4,2
Utopia-1_APl	.	Agrilus planipennis	NeSL	1,3,4,2
Utopia-1_CFl	.	Camponotus floridanus	NeSL	1,3,4,2
Utopia-1_CMy	.	Chelonia mydas	NeSL	1,3,4,2
Utopia-1_CPB	.	Chrysemys picta bellii	NeSL	1,3,4,2
Utopia-1_Crp	.	Crocodylus porosus	NeSL	1,3,4,2
Utopia-1_DPo	.	Dendroctonus ponderosae	NeSL	1,3,4,2
Utopia-1_DPu	.	Daphnia pulex	NeSL	1,3,4,2
Utopia-1_DYak	.	Drosophila yakuba	NeSL	1,3,4,2
Utopia-1_EBr	.	Eimeria brunetti	NeSL	
Utopia-1_EMi	.	Eimeria mitis	NeSL	1,3,4,2
Utopia-1_ENe	.	Eimeria necatrix	NeSL	5,1,19,3,4,2
Utopia-1_Gav	.	Gavialis gangeticus	NeSL	1,3,4,2
Utopia-1_GGl	.	Ganaspis	NeSL	1,3,4,2
Utopia-1_HAra	.	Hyaloperonospora arabidopsidis	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_HG	.	Heterodera glycines	NeSL	1,3,4,2
Utopia-1_HMM	.	Heliconius melpomene melpomene	NeSL	1,3,4,2
Utopia-1_HSal	.	Harpegnathos saltator	NeSL	1,3,4,2
Utopia-1_IS	.	Ixodes scapularis	NeSL	1,3,4,2
Utopia-1_LAl	.	Lasioglossum albipes	NeSL	1,3,4,2
Utopia-1_LFu	.	Ladona fulva	NeSL	1,3,4,2
Utopia-1_LV	AGCV01358106	Lytechinus variegatus	NeSL	18,5,1,19,3,4,2
Utopia-1_MRo	.	Megachile rotundata	NeSL	
Utopia-1_NVit	.	Nasonia vitripennis	NeSL	18,5,1,3,4,2
Utopia-1_PAlni	.	Phytophthora alni	NeSL	7,1,17,10,230,6,2,80
Utopia-1_PArrh	.	Pythium arrhenomanes	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_PBa	.	Pogonomyrmex barbatus	NeSL	1,3,4,2
Utopia-1_PCa	.	Phytophthora capsici	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_PCinn	.	Phytophthora cinnamomi	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_PCu	AHJF01004292	Pseudoperonospora cubensis	NeSL	1,3,4,2
Utopia-1_PI	.	Phytophthora infestans	NeSL	6,2,10,80,1,230,4,5,19,17,3,7
Utopia-1_PInsi	.	Pythium insidiosum	NeSL	7,1,17,3,10,230,4,2,80
Utopia-1_PKern	.	Phytophthora kernoviae	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-1_PLate	.	Phytophthora lateralis	NeSL	7,1,17,3,10,4,2,80
Utopia-1_PMi	.	Patiria miniata	NeSL	18,5,1,19,3,4,2
Utopia-1_PPac	.	Pristionchus pacificus	NeSL	1,3,4,2
Utopia-1_PPini	.	Phytophthora pinifolia	NeSL	80,17,10,6,7
Utopia-1_PR	.	Phytophthora ramorum	NeSL	7,1,80,17,3,10,230,4,2,6
Utopia-1_PRe	.	Panagrellus redivivus	NeSL	1,3,4,2
Utopia-1_PS	.	Phytophthora sojae	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-1_PSi	.	Pelodiscus sinensis	NeSL	1,3,4,2
Utopia-1_PT	.	Parasteatoda tepidariorum	NeSL	18,5,19,1,3,4,2
Utopia-1_PU	ADOS01001321	Pythium ultimum	NeSL	7,1,80,17,3,10,230,4,6,2

Utopia-1_PVexa	.	Phytopythium aff. vexans	NeSL	7,1,17,3,10,4,2,80
Utopia-1_SaPa	.	Saprolegnia parasitica	NeSL	1,3,4,2
Utopia-1_SDicl	.	Saprolegnia diclina	NeSL	1,3,4,2
Utopia-1_SM	.	Strigamia maritima	NeSL	1,3,4,2
Utopia-1_SP	AAGJ02140537	Strongylocentrotus purpuratus	NeSL	18,5,1,19,3,4,2
Utopia-1_TSP	.	Trichinella spiralis	NeSL	147,1,148,3,149,120,4,2
Utopia-1B_CPB	.	Chrysemys picta bellii	NeSL	1,3,4,2
Utopia-2_APi	.	Acyrthosiphon pisum	NeSL	1,3,4,2
Utopia-2_CMy	.	Chelonia mydas	NeSL	1,3,4,2
Utopia-2_CPB	.	Chrysemys picta bellii	NeSL	1,3,4,2
Utopia-2_DPu	.	Daphnia pulex	NeSL	1,3,4,2
Utopia-2_LFu	.	Ladona fulva	NeSL	
Utopia-2_PCa	.	Phytophthora capsici	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-2_PI	.	Phytophthora infestans	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-2_PR	.	Phytophthora ramorum	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-2_PS	.	Phytophthora sojae	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-2_PU	.	Pythium ultimum	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-3_CPB	.	Chrysemys picta bellii	NeSL	1,3,4,2
Utopia-3_DPu	.	Daphnia pulex	NeSL	1,3,4,2
Utopia-3_LFu	.	Ladona fulva	NeSL	
Utopia-3_PCa	.	Phytophthora capsici	NeSL	
Utopia-3_PI	.	Phytophthora infestans	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-3_PR	.	Phytophthora ramorum	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-4_LFu	.	Ladona fulva	NeSL	
Utopia-4_PI	AATU01001281.1	Phytophthora infestans	NeSL	6,2,10,80,1,230,4,5,19,17,3,7
Utopia-4_PR	.	Phytophthora ramorum	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-5_LFu	.	Ladona fulva	NeSL	1,3,4,2
Utopia-5_PI	.	Phytophthora infestans	NeSL	7,1,17,3,10,230,4,6,2,80
Utopia-5_PR	.	Phytophthora ramorum	NeSL	7,1,80,17,3,10,230,4,6,2
Utopia-6_LFu	.	Ladona fulva	NeSL	1,3,4,2
UVIR_DV	.	Drosophila virilis	Jockey	16,1,3,11,4,2,15
Ves	Y19045Y19046Y19047Y1	Myotis daubentonii	SINE2/tRNA	
Ves2_ML	.	Myotis lucifugus	SINE2/tRNA	
Ves2B_ML	.	Myotis lucifugus	SINE2/tRNA	
Ves3_ML	.	Myotis lucifugus	SINE2/tRNA	
Ves4_ML	.	Myotis lucifugus	SINE2/tRNA	
Vingi-1_AAl	.	Atelerix albiventris	Vingi	16,1,3,11,4,2,15
Vingi-1_Acar	.	Anolis carolinensis	Vingi	16,1,3,11,4,2,15
Vingi-1_AMi	.	Crocodylidae	Vingi	
Vingi-1_APis	.	Acyrthosiphon pisum	Vingi	16,1,3,11,4,2,15
Vingi-1_BF	.	Branchiostoma floridae	Vingi	16,1,3,11,4,2,15
Vingi-1_BM	.	Bombyx mori	Vingi	16,1,3,11,4,2,15
Vingi-1_BTa	.	Bemisia tabaci	Vingi	16,1,3,11,4,2,15
Vingi-1_CE	.	Caenorhabditis elegans	Vingi	16,1,3,11,4,2,15
Vingi-1_CJ	.	Caenorhabditis japonica	Vingi	94,16,1,3,95,24,11,4,2
Vingi-1_CPB	.	Chrysemys picta bellii	Vingi	16,1,3,11,4,2,15
Vingi-1_Croc	.	Crocodylidae	Vingi	16,1,3,11,4,2,15
Vingi-1_EE	.	Erinaceus europaeus	Vingi	16,1,3,11,4,2,15
Vingi-1_Gav	.	Gavialis gangeticus	Vingi	1,3,4,2
Vingi-1_Hmel	.	Heliconius melpomene	Vingi	16,1,3,11,4,2,15
Vingi-1_HR	.	Helobdella robusta	Vingi	16,1,3,11,4,2,15

Vingi-1_Lch	.	Latimeria chalumnae	Vingi	1,3,11,4,2
Vingi-1_Lme	.	Latimeria menadoensis	Vingi	1,3,11,4,2
Vingi-1_LVa	.	Litopenaeus vannamei	Vingi	1,3,11,4,2
Vingi-1_PH	.	Parhyale hawaiensis	Vingi	16,1,3,11,4,2,15
Vingi-1_PMa	.	Petromyzon marinus	Vingi	16,1,3,11,4,2,15
Vingi-1_Pp	.	Phlebotomus papatasi	Vingi	16,1,3,11,4,2,15
Vingi-1_PSi	.	Pelodiscus sinensis	Vingi	1,3,4,2
Vingi-1_SPur	.	Strongylocentrotus purpuratus	Vingi	16,1,3,4,11,2,15
Vingi-1_Tcas	.	Tribolium castaneum	Vingi	16,1,3,11,4,2,15
Vingi-10_AMi	.	Alligator mississippiensis	Vingi	
Vingi-10_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-10_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-11_AMi	.	Alligator mississippiensis	Vingi	1,3,4,2
Vingi-11_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-11_LMi	.	Locusta migratoria	Vingi	16,1,3,4,11,2,15
Vingi-12_CPB	.	Chrysemys picta bellii	Vingi	16,1,3,11,4,2,15
Vingi-12_LMi	.	Locusta migratoria	Vingi	2,1
Vingi-13_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-13_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-14_CPB	.	Chrysemys picta bellii	Vingi	16,1,3,11,4,2,15
Vingi-14_LMi	.	Locusta migratoria	Vingi	1,3,11,4,2
Vingi-15_LMi	.	Locusta migratoria	Vingi	1,2
Vingi-16_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-17_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-18_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-19_LMi	.	Petromyzon marinus	Vingi	
Vingi-1B_PM	.	Bombyx mori	Vingi	
Vingi-1N1_BM	.	Erinaceus europaeus	Vingi	
Vingi-1N1_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N10_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N11_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N2_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N3_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N4_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N5_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N6_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N7_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N8_EE	.	Erinaceus europaeus	Vingi	
Vingi-1N9_EE	.	Erinaceus europaeus	Vingi	
Vingi-2_Acar	.	Anolis carolinensis	Vingi	16,1,3,11,4,2,15
Vingi-2_AMi	.	Alligator mississippiensis	Vingi	
Vingi-2_BF	.	Branchiostoma floridae	Vingi	16,1,3,4,11,2,15
Vingi-2_CE	.	Caenorhabditis elegans	Vingi	1,3,2
Vingi-2_CJap	.	Caenorhabditis japonica	Vingi	16,1,3,11,4,2,15
Vingi-2_CPB	.	Chrysemys picta bellii	Vingi	16,1,3,11,4,2,15
Vingi-2_Crp	.	Crocodylus porosus	Vingi	
Vingi-2_Gav	.	Gavialis gangeticus	Vingi	16,1,3,11,4,2,15
Vingi-2_HR	.	Helobdella robusta	Vingi	16,1,3,11,4,2,15
Vingi-2_LCh	.	Latimeria chalumnae	Vingi	16,1,3,11,4,2,15
Vingi-2_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-2_LVa	.	Litopenaeus vannamei	Vingi	16,1,3,4,11,2,15

Vingi-2_PH	.	Parhyale hawaiensis	Vingi	16,1,11,2,15
Vingi-2_PM	.	Petromyzon marinus	Vingi	16,1,3,11,4,2,15
Vingi-2_SPur	.	Strongylocentrotus purpuratus	Vingi	16,1,3,11,4,2,15
Vingi-20_LMi	.	Locusta migratoria	Vingi	16,1,3,4,11,2,15
Vingi-21_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-22_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-23_LMi	.	Locusta migratoria	Vingi	16,1,3,11,2,15
Vingi-24_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-25_LMi	.	Locusta migratoria	Vingi	16,1,3,11,2,15
Vingi-26_LMi	.	Locusta migratoria	Vingi	1,3,2
Vingi-27_LMi	.	Locusta migratoria	Vingi	16,1,3,4,11,2,15
Vingi-28_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-29_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-3_AMi	.	Alligator mississippiensis	Vingi	1,3,4,2
Vingi-3_BF	.	Branchiostoma floridae	Vingi	1,3,4,2
Vingi-3_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-3_Crp	.	Crocodylus porosus	Vingi	
Vingi-3_Gav	.	Gavialis gangeticus	Vingi	1,3,4,2
Vingi-3_HR	.	Helobdella robusta	Vingi	16,1,3,11,4,2,15
Vingi-3_LCh	.	Latimeria chalumnae	Vingi	16,1,3,11,4,2,15
Vingi-3_LMi	.	Locusta migratoria	Vingi	16,1,3,4,11,2,15
Vingi-3_PH	.	Parhyale hawaiensis	Vingi	
Vingi-3_PM	.	Petromyzon marinus	Vingi	1,3,4,2
Vingi-3_SPur	.	Strongylocentrotus purpuratus	Vingi	16,1,3,11,4,2,15
Vingi-30_LMi	.	Locusta migratoria	Vingi	16,1,3,24,11,4,2
Vingi-31_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-32_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-33_LMi	.	Locusta migratoria	Vingi	1,3,2
Vingi-35_LMi	.	Locusta migratoria	Vingi	
Vingi-36_LMi	.	Locusta migratoria	Vingi	
Vingi-37_LMi	.	Locusta migratoria	Vingi	16,1,3,24,4,11,2
Vingi-38_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-39_LMi	.	Locusta migratoria	Vingi	1,4,2
Vingi-4_AMi	.	Alligator mississippiensis	Vingi	16,1,11,4,2,15
Vingi-4_CPB	.	Chrysemys picta bellii	Vingi	1,3,4,2
Vingi-4_Crp	.	Crocodylus porosus	Vingi	
Vingi-4_HR	.	Helobdella robusta	Vingi	16,1,3,60,11,4,2,15
Vingi-4_LCh	.	Latimeria chalumnae	Vingi	16,1,3,24,11,4,2
Vingi-4_LMi	.	Locusta migratoria	Vingi	16,1,3,11,4,2,15
Vingi-4_PM	.	Petromyzon marinus	Vingi	
Vingi-4_SPur	.	Strongylocentrotus purpuratus	Vingi	16,1,3,24,11,4,2
Vingi-40_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-41_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-42_LMi	.	Locusta migratoria	Vingi	1,3,2
Vingi-43_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-44_LMi	.	Locusta migratoria	Vingi	
Vingi-45_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-46_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-47_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-48_LMi	.	Locusta migratoria	Vingi	
Vingi-49_LMi	.	Locusta migratoria	Vingi	1,3,4,2

Vingi-5_AMi	.	Alligator mississippiensis	Vingi	16,1,3,11,4,2,15
Vingi-5_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-5_Crp	.	Crocodylus porosus	Vingi	
Vingi-5_LMi	.	Locusta migratoria	Vingi	16,1,3,24,11,4,2
Vingi-5_PM	.	Petromyzon marinus	Vingi	1,3,4,2
Vingi-6_AMi	.	Alligator mississippiensis	Vingi	1,3,4,2
Vingi-6_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-6_LCh	.	Latimeria chalumnae	Vingi	16,1,3,11,4,2,15
Vingi-6_LMi	.	Locusta migratoria	Vingi	1,3,11,4,2
Vingi-6_PM	.	Petromyzon marinus	Vingi	
Vingi-7_AMi	.	Alligator mississippiensis	Vingi	
Vingi-7_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-7_LCh	.	Latimeria chalumnae	Vingi	16,1,3,11,4,2,15
Vingi-7_LMi	.	Locusta migratoria	Vingi	16,1,3,4,11,2,15
Vingi-7_PM	.	Petromyzon marinus	Vingi	
Vingi-8_AMi	.	Alligator mississippiensis	Vingi	1,3,4,2
Vingi-8_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-8_LCh	.	Latimeria chalumnae	Vingi	
Vingi-8_LMi	.	Locusta migratoria	Vingi	16,1,24,11,4,2
Vingi-8_PM	.	Petromyzon marinus	Vingi	
Vingi-9_AMi	.	Alligator mississippiensis	Vingi	
Vingi-9_CPB	.	Chrysemys picta bellii	Vingi	
Vingi-9_LCh	.	Latimeria chalumnae	Vingi	
Vingi-9_LMi	.	Locusta migratoria	Vingi	1,3,4,2
Vingi-9_PM	.	Petromyzon marinus	Vingi	1,3,4,2
VINSINE	.	Varanus indicus	SINE2/tRNA	
VLINE1_VV	.	Vitis vinifera	L1	1,43,3,44,4,2
VLINE1_VV	.	Vitis vinifera	L1	1,43,3,44,4,2
VLINE10_VV	.	Vitis vinifera	L1	43,2,44,24,11,1,4,94,16,75,3,76,95
VLINE11_VV	.	Vitis vinifera	L1	16,1,75,43,3,76,44,24,11,4,2
VLINE2_VV	.	Vitis vinifera	L1	16,1,3,24,11,4,2
VLINE2_VV	.	Vitis vinifera	L1	16,1,3,24,11,4,2
VLINE3_VV	.	Vitis vinifera	L1	16,1,43,3,44,24,11,2
VLINE3_VV	.	Vitis vinifera	L1	16,1,43,3,44,24,11,2
VLINE4_VV	.	Vitis vinifera	L1	1,3,4,2
VLINE4_VV	.	Vitis vinifera	L1	1,3,4,2
VLINE5_VV	.	Vitis vinifera	L1	94,1,3,95,11,4,2
VLINE5_VV	.	Vitis vinifera	L1	94,1,3,95,11,4,2
VLINE6_VV	.	Vitis vinifera	L1	1,43,75,3,44,76,11,4,2
VLINE6_VV	.	Vitis vinifera	L1	1,43,75,3,44,76,11,4,2
VLINE7_VV	.	Vitis vinifera	L1	16,1,43,3,44,24,4,11,2
VLINE8_VV	.	Vitis vinifera	L1	16,1,43,3,44,24,11,4,2
VLINE9_VV	.	Vitis vinifera	L1	16,1,75,43,3,76,44,24,11,4,2
W_As	.	Adineta	Penelope	3,1
W_Av	.	Adineta vaga	Penelope	1,3,130
W1_As	.	Adineta	Penelope	1,3
W1_Av	.	Adineta vaga	Penelope	1,3,167,130,169
W2_As	.	Adineta	Penelope	1,3
W2_Av	.	Adineta vaga	Penelope	1,3
Waldo-1_AAe	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3
Waldo-1_CQ	.	Culex quinquefasciatus	R1	2,8,11,86,12,42,1,14,4,15,16,3

Waldo-2_AAe	.	Aedes aegypti	R1	2,8,11,86,12,1,14,4,15,16,3
Waldo-3_AAe	.	Aedes aegypti	R1	2,8,11,86,12,1,14,4,15,16,3
Waldo-4_AAe	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3
Waldo-5_AAe	.	Aedes aegypti	R1	2,8,11,86,12,42,1,14,4,15,16,3
Waldo-6_AAe	.	Aedes aegypti	R1	2,8,11,86,12,1,14,4,15,16,3
Waldo1_AG	AB090814	Anopheles gambiae	R1	2,8,11,86,12,1,4,15,16,3
WaldoAg2	AB090815	Anopheles gambiae	R1	2,8,11,86,12,42,1,14,4,15,16,3
WaldoFs1	AB078939	Forficula scudderi	R1	16,1,3,4,11,2,15
WALLSI1	.	Notamacropus eugenii	SINE	
WALLSI1A	.	Notamacropus eugenii	SINE	
WALLSI2	.	Notamacropus eugenii	SINE2/tRNA	
WALLSI3	.	Notamacropus eugenii	SINE	
WALLSI4	.	Notamacropus eugenii	SINE2/tRNA	
WHLN1_TA	AB016116	Triticum aestivum	L1	
WHLN2_TA	AB016117	Triticum aestivum	L1	
WHLN3_TA	AB016118	Triticum aestivum	L1	1,3,11,4,2
WORF_DMi	AY144572	Drosophila miranda	CR1	Non-LTRRetrotransposo
WRETRO	.	Notamacropus eugenii	SINE2/tRNA	
WSINE1	.	Notamacropus eugenii	SINE2/tRNA	
WSINE1a	.	Notamacropus eugenii	SINE2/tRNA	
X1_LINE	.	Euteleostomi	CR1	
X11_LINE	.	Mammalia	RTE	
X11_LINE	.	Mammalia	RTE	
X11_LINE	.	Mammalia	RTE	
X12_LINE	.	Mammalia	Vingi	
X12_LINE	.	Mammalia	Vingi	
X12_LINE	.	Mammalia	Vingi	
X13_LINE	.	Mammalia	Crack	
X15_LINE	.	Mammalia	L2	
X15_LINE	.	Mammalia	L2	
X15_LINE	.	Mammalia	L2	
X17_LINE	.	Mammalia	CR1	
X17_LINE	.	Mammalia	CR1	
X17_LINE	.	Mammalia	CR1	
X18_LINE	.	Mammalia	CR1	
X18_LINE	.	Mammalia	CR1	
X18_LINE	.	Mammalia	CR1	
X19_LINE	.	Mammalia	CR1	
X19_LINE	.	Mammalia	CR1	
X19_LINE	.	Mammalia	CR1	
X2_LINE	.	Euteleostomi	CR1	
X20_LINE	.	Mammalia	CR1	
X20_LINE	.	Mammalia	CR1	
X20_LINE	.	Mammalia	CR1	
X21_LINE	.	Mammalia	CR1	
X21_LINE	.	Mammalia	CR1	
X21_LINE	.	Mammalia	CR1	
X24_LINE	.	Mammalia	L2	
X24_LINE	.	Mammalia	L2	
X24_LINE	.	Mammalia	L2	
X3_LINE	.	Mammalia	RTE	

X4_LINE	.	Vertebrata	R4	
X5A_LINE	.	Euteleostomi	CR1	
X5B_LINE	.	Euteleostomi	CR1	
X6A_LINE	.	Euteleostomi	CR1	
X6B_LINE	.	Euteleostomi	CR1	
X7A_LINE	.	Euteleostomi	CR1	
X7B_LINE	.	Euteleostomi	CR1	
X7C_LINE	.	Euteleostomi	CR1	
X7D_LINE	.	Euteleostomi	CR1	
X8_LINE	.	Euteleostomi	CR1	
X9_LINE	.	Euteleostomi	L1	
YJ_As	.	Adineta	Penelope	
Y11i	YLI319752	Yarrowia lipolytica	L1	3,1 16,1,26,3,24,11,4,2
Yoyo-1_PH	.	Parhyale hawaiensis	CR1	1,3,82,4,11,2
YURE_CSa	.	Ciona savignyi	NeSL	5,1,3,4,2
YURE-2_Cis	.	Ciona savignyi	R2	5,1,19,3,4,2
YURECi	.	Ciona intestinalis	NeSL	5,1,3,4,2
YW_As	.	Adineta	Penelope	1,3
ZENON_BM	.	Bombyx mori	CR1	16,1,3,11,4,2,15
Zenon-1_Hmel	.	Heliconius melpomene	CR1	16,1,3,11,4,2,15
ZENON-1_PPo	.	Papilio polytes	CR1	1,3,4,2
ZENON-1_PXu	.	Papilio xuthus	CR1	
ZENON-1B_PPo	.	Papilio polytes	CR1	
Zenon-2_Hmel	.	Heliconius melpomene	CR1	16,1,3,11,4,2,15
ZENON-2_PPo	.	Papilio polytes	CR1	
ZENON-2_PXu	.	Papilio xuthus	CR1	
ZENON-2B_PPo	.	Papilio polytes	CR1	
ZENON-2C_PPo	.	Papilio polytes	CR1	
ZENON-2D_PPo	.	Papilio polytes	CR1	
ZENON-2E_PPo	.	Papilio polytes	CR1	
Zenon-3_Hmel	.	Heliconius melpomene	CR1	16,1,3,11,4,2,15
ZENON-3_PPo	.	Papilio polytes	CR1	
ZENON-3_PXu	.	Papilio xuthus	CR1	
ZENON-4_PPo	.	Papilio polytes	CR1	
ZENON-4_PXu	.	Papilio xuthus	CR1	
ZENON-5_PPo	.	Papilio polytes	CR1	
ZENON-5_PXu	.	Papilio xuthus	CR1	
Zepp	AB008896, D83919	Chlorella vulgaris	L1	16,1,26,24,11,4,2
Zorro	AF254443	Candida albicans	L1	1,3,11,4,2

Table Y. Exemplary InterPro domains from Table X *(following pages)*

NID	Domain	Name	Description
1	IPR000477	Reverse transcriptase domain	"The use of an RNA template to produce DNA, for integration into the host genome and exploitation of a host cell, is a strategy employed in the replication of retroid elements, such as the retroviruses and bacterial retrons. The enzyme catalysing polymerisation is an RNA-directed DNA-polymerase, or reverse trancriptase (RT) (). Reverse transcriptase occurs in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses.Retroviral reverse transcriptase is synthesised as part of the POL polyprotein that contains; an aspartyl protease, a reverse transcriptase, RNase H and integrase. POL polyprotein undergoes specific enzymatic cleavage to yield the mature proteins. The discovery of retroelements in the prokaryotes raises intriguing questions concerning their roles in bacteria and the origin and evolution of reverse transcriptases and whether the bacterial reverse transcriptases are older than eukaryotic reverse transcriptases.Several crystal structures of the reverse transcriptase (RT) domain have been determined."
2	PF00078	Reverse transcriptase (RNA-dependent DNA polymerase)	"A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses."
3	PS50878	Reverse transcriptase (RT) catalytic domain profile	"Retroviral reverse transcriptases (EC 2.7.7.49) (prototypical RTs) exhibit three enzymatic activities: an RNA-dependent DNA polymerase, Ribonuclease H, and a DNA-dependent DNA polymerase. These are employed in copying the plus strand RNA genome to produce a minus strand of DNA, removal of the RNA template, and synthesis of the plus strand of DNA using the minus-strand DNA as a template, respectively[PMID:7526778]. Both polymerase activities are performed by the RT domain whereas the ribonuclease activity is achieved by the RNase H domain. Several crystal structures of the RT domain have been determined (see for example {PDB:3HVT})[PMID:13774403]. Structurally, the RT domain can be divided in four subdomains. Its anatomical resemblance to a right hand has led to naming the subdomains as fingers, palm, and thumb. The fourth subdomain lies between the rest of the RT and the RNase H domain leading it to be called the connection subdomain. The palm domain contains the catalytic triad critical for polymerase activity, and its folded structure resembles that of the corresponding catalytic domain in other DNA and RNA polymerases. The position of fingers and thumb defined a deep cleft, with polymerase active residues at its base. Proteins known to contain an RT domain are listed below: - Retroviral reverse transcriptase. - Catalytic subunit of the telomerase. - Eukaryotic retrotransposons. - Escherichia coli retroelement retron, which is responsible for the synthesis of multicopy single-stranded DNA (msDNA). - Bacterial mobile group II intron, that can reverse splice into DNA target substrate. The finger, thumb and connection subdomains are less conserved among all RT domain subfamilies[PMID:9703504]and could not be identified by the profile method. The profile we developed is thus directed against the catalytic region (palm subdomain)."
4	cd01650	RT_nLTR: Non-LTR	"RT_nLTR: Non-LTR (long terminal repeat) retrotransposon and non-LTR retrovirus reverse transcriptase (RT). This subfamily contains both non-LTR retrotransposons and non-LTR retrovirus RTs. RTs catalyze the conversion of single-stranded RNA into

		(long terminal repeat) retrotransposon and non-LTR retrovirus...	double-stranded DNA for integration into host chromosomes. RT is a multifunctional enzyme with RNA-directed DNA polymerase, DNA directed DNA polymerase and ribonuclease hybrid (RNase H) activities. [PMID: 1698615, PMID: 10605110, PMID: 10628860, PMID: 11734649, PMID: 12117499, PMID: 12777502, PMID: 14871946, PMID: 15939396, PMID: 16271150, PMID: 16356661, PMID: 2463954, PMID: 3040362, PMID: 3656436, PMID: 7512193, PMID: 7534829, PMID: 7659515, PMID: 8524653, PMID: 9190061, PMID: 9218812, PMID: 9332379, PMID: 9364772, PMID: 8828137]"
5	IPR013087	Zinc finger C2H2-type	"C2H2-type (classical) zinc fingers (Znf) were the first class to be characterised. They contain a short beta hairpin and an alpha helix (beta/beta/alpha structure), where a single zinc atom is held in place by Cys(2)His(2) (C2H2) residues in a tetrahedral array. C2H2 Znf's can be divided into three groups based on the number and pattern of fingers: triple-C2H2 (binds single ligand), multiple-adjacent-C2H2 (binds multiple ligands), and separated paired-C2H2. C2H2 Znf's are the most common DNA-binding motifs found in eukaryotic transcription factors, and have also been identified in prokaryotes. Transcription factors usually contain several Znf's (each with a conserved beta/beta/alpha structure) capable of making multiple contacts along the DNA, where the C2H2 Znf motifs recognise DNA sequences by binding to the major groove of DNA via a short alpha-helix in the Znf, the Znf spanning 3-4 bases of the DNA. C2H2 Znf's can also bind to RNA and protein targets.Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.This entry represents the classical C2H2 zinc finger domain."
6	PS50879	RNase H domain profile	"Ribonuclease H (RNase H) (EC 3.1.26.4) recognizes and cleaves the RNA strand of RNA-DNA heteroduplexes. In prokaryotes RNase H is involved in removing the RNA of RNA-DNA hybrids, particularly the primers of Okazaki fragments in lagging-strand DNA synthesis[PMID:2177653]. An additional function of the bacterial RNase H domain is to generate primers for initiation of DNA replication[PMID:2446774]. The RNase H domain is also present in retroviral reverse transcriptase proteins where it performs three functions : (a) generation of a polypurine tract which acts as the primer for plus strand DNA synthesis, (b) removal of RNA primers from nascent minus and plus DNA strands, and (c) degradation of the original RNA template. For HIV, inactivation of the RNase H renders the virus non-infectious, making this enzymatic activity a target for therapy of AIDS[PMID:2177653]. In eukaryotes, the RNase H1 protein is present in nuclei, but it has also been isolated from mitochondria of many organisms. In mouse inactivation of the RNaseH1 gene arrest development about day E8.5 because of a failure to generate mtDNA, indicating an essential role for RNase H1 during mammalian embryogenesis[PMID:12667461]. The 3D structure of the RNase H domain from diverse bacteria and retroviruses has been solved (see for example {PDB:1RNH}) [PMID:2169648,PMID:8108376,PMID:1707186]. All have four beta strands and four to five alpha helices. The Escherichia coli RNase H1 protein binds a single Mg(2+) ion cofactor in the active site of the enzyme. The divalent cation is bound by the

			carboxyl groups of four acidic residues, Asp-10, Glu-48, Asp-70, and Asp-134[PMID:8108376]. The first three acidic residues are highly conserved in all bacterial and retroviral RNase H sequences. The profile we developed covers the whole RNase H domain."
7	IPR012337	Ribonuclease H-like superfamily	"The catalytic domain of several polynucleotidyl transferases share a similar structure, consisting of a 3-layer alpha/beta/alpha fold that contains mixed beta sheets, suggesting that they share a similar mechanism of catalysis. Polynucleotidyl transferases containing this domain include ribonuclease H class I (RNase HI) and class II (RNase HII), HIV RNase (reverse transcriptase domain), retroviral integrase (catalytic domain), Mu transposase (core domain), transposase inhibitor Tn5 (containing additional all-alpha subdomains), DnaQ-like 3'-5' exonucleases (exonuclease domains), RuvC resolvase, and mitochondrial resolvase ydc2 (catalytic domain).This superfamily also includes the YqgF domain, described as RNase H-like and typified by the Escherichia coli protein YqgF. YqgF domain-containing proteins are predicted to be ribonucleases or resolvases based on homology to RuvC Holliday junction resolvases."
8	PS50158	Zinc finger CCHC-type profile	"The 18 residues CCHC zinc finger is mainly found in the nucleocapsid protein of retrovirus. It is required for viral genome packaging and for early infection process [PMID:2695083,PMID:10074407]. It is also found in eukaryotic proteins involved in RNA binding or single-stranded DNA binding. Some proteins known to contain a CCHC zinc finger are listed below. - Retroviral GAG protein. Nucleocapside protein. - Human CNBP protein. CNBP is involved in sterol-mediated repression. It has also been shown to bind the 5' untranslated region of ribosomal proteins mRNA (a region known to be important in translational regulation). - Yeast BYR3; the CNBP homolog. - Caenorhabditis elegans protein glh-1; a probable ATP-binding RNA helicase. - Leishmania major hexamer-binding protein (HEXP). It binds the 5' untranslated region of the GP63 gene. - Tobbacco transposon TNT1 PolX protein. Retrovirus-related pol polyprotein. The profile we developed covers the whole domain."
9	cd09276	non-LTR RNase HI domain of reverse transcriptases	"Ribonuclease H (RNase H) is classified into two families, type 1 (prokaryotic RNase HI, eukaryotic RNase H1 and viral RNase H) and type 2 (prokaryotic RNase HII and HIII, and eukaryotic RNase H2). Ribonuclease HI (RNase HI) is an endonuclease that cleaves the RNA strand of an RNA/DNA hybrid in a sequence non-specific manner. RNase H is widely present in various organisms, including bacteria, archaea and eukaryotes. RNase HI has also been observed as an adjunct domain to the reverse transcriptase gene in retroviruses, long-term repeat (LTR)-bearing retrotransposons and non-LTR retrotransposons. RNase HI in LTR retrotransposons perform degradation of the original RNA template, generation of a polypurine tract (the primer for plus-strand DNA synthesis), and final removal of RNA primers from newly synthesized minus and plus strands. The catalytic residues for RNase H enzymatic activity, three aspartatic acids and one glutamic acid residue (DEDD), are unvaried across all RNase H domains. The position of the RNase domain of non-LTR and LTR transposons is at the carboxyl terminal of the reverse transcriptase (RT) domain and their RNase domains group together, indicating a common evolutionary origin. Many non-LTR transposons have lost the RNase domain because their activity is at the nucleus and cellular RNase may suffice; however LTR retrotransposons always encode their own RNase domain because it requires RNase activity in RNA-protein particles in the cytoplasm. RNase H inhibitors have been explored as an anti-HIV drug target because RNase H inactivation inhibits reverse transcription. [PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265, PMID: 19228198, PMID: 11435400]"
10	IPR002156	Ribonuclease H domain	"The RNase H domain is responsible for hydrolysis of the RNA portion of RNA x DNA hybrids, and this activity requires the presence of divalent cations (Mg2+ or Mn2+) that bind its active site. This domain is a part of a large family of homologous RNase H enzymes of which the RNase HI protein from Escherichia coli is the best characterised. Secondary structure predictions for the enzymes from E. coli, yeast, human liver and diverse retroviruses (such as Rous sarcoma virus and the Foamy viruses) supported, in every case, the five beta-strands (1 to 5) and four or five alpha-helices (A, B/C, D, E) that have been identified by crystallography in the RNase H domain of Human immunodeficiency virus 1 (HIV-1) reverse transcriptase

				and in E. coli RNase H. Reverse transcriptase (RT) is a modular enzyme carrying polymerase and ribonuclease H (RNase H) activities in separable domains. Reverse transcriptase (RT) converts the single-stranded RNA genome of a retrovirus into a double-stranded DNA copy for integration into the host genome. This process requires ribonuclease H as well as RNA- and DNA-directed DNA polymerase activities.Retroviral RNase H is synthesised as part of the POL polyprotein that contains; an aspartyl protease, a reverse transcriptase, RNase H and integrase. POL polyprotein undergoes specific enzymatic cleavage to yield the mature proteins. Bacterial RNase Hcatalyses endonucleolytic cleavage to 5'-phosphomonoester acting on RNA-DNA hybrids.The 3D structure of the RNase H domain from diverse bacteria and retroviruses has been solved. All have four beta strands and four to five alpha helices. The E. coli RNase H1 protein binds a single Mg2+ ion cofactor in the active site of the enzyme. The divalent cation is bound by the carboxyl groups of four acidic residues, Asp-10, Glu-48, Asp-70, and Asp-134. The first three acidic residues are highly conserved in all bacterial and retroviral RNase H sequences."
1 1	IPR03 6691	Endonuclease /exonuclease/ phosphatase superfamily		"This domain superfamily is found in a large number of proteins including magnesium dependent endonucleases and phosphatases involved in intracellular signalling. This domain is found in many proteins including AP endonuclease proteins (DNA-(apurinic or apyrimidinic site) lyase), deoxyribonuclease-1 (DNase I;), synaptojanin an inositol-1,4,5-trisphosphate phosphatase () and sphingomyelinase ().The DNase I domain structure contains a beta-sandwich fold and a duplication of the alpha+beta motif."
1 2	IPR00 1878	"Zinc finger, CCHC-type"		"Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.This entry represents the CysCysHisCys (CCHC) type zinc finger domains, and have the sequence:C-X2-C-X4-H-X4-Cwhere X can be any amino acid, and number indicates the number of residues. These 18 residues CCHC zinc finger domains are mainly found in the nucleocapsid protein of retroviruses. It is required for viral genome packaging and for early infection process. It is also found in eukaryotic proteins involved in RNA binding or single-stranded DNA binding."
1 3	PF000 75	RNase H		"RNase H digests the RNA strand of an RNA/DNA hybrid. Important enzyme in retroviral replication cycle, and often found as a domain associated with reverse transcriptases. Structure is a mixed alpha+beta fold with three a/b/a layers."
1 4	IPR03 6875	"Zinc finger, CCHC-type superfamily"		"Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf

				motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.This entry represents the CysCysHisCys (CCHC) type zinc finger domains superfamily, and have the sequence:C-X2-C-X4-H-X4-Cwhere X can be any amino acid, and number indicates the number of residues. These 18 residues CCHC zinc finger domains are mainly found in the nucleocapsid protein of retroviruses. It is required for viral genome packaging and for early infection process. It is also found in eukaryotic proteins involved in RNA binding or single-stranded DNA binding."
1 5	PF145 29	Endonuclease -reverse transcriptase	"This domain represents the endonuclease region of retrotransposons from a range of bacteria, archaea and eukaryotes. These are enzymes largely from class EC:2.7.7.49."
1 6	IPR00 5135	Endonuclease /exonuclease/ phosphatase	"This domain is found in a large number of proteins including magnesium dependent endonucleases and phosphatases involved in intracellular signalling. Proteins this domain is found in include: AP endonuclease proteins (), DNase I proteins (), Synaptojanin an inositol-1,4,5-trisphosphate phosphatase () and Sphingomyelinase ()."
1 7	IPR03 6397	Ribonuclease H superfamily	"The catalytic domain of several polynucleotidyl transferases share a similar structure, consisting of a 3-layer alpha/beta/alpha fold that contains mixed beta sheets, suggesting that they share a similar mechanism of catalysis. Polynucleotidyl transferases containing this domain include ribonuclease H class I (RNase HI) and class II (RNase HII), HIV RNase (reverse transcriptase domain), retroviral integrase (catalytic domain), Mu transposase (core domain), transposase inhibitor Tn5 (containing additional all-alpha subdomains), DnaQ-like 3'-5' exonucleases (exonuclease domains), RuvC resolvase, and mitochondrial resolvase ydc2 (catalytic domain)."
1 8	PS000 28	Putative nuclease HARBI1	NA
1 9	PS501 57	Zinc finger C2H2 type domain profile	"'Zinc finger' domains [PMID:3125980,PMID:3292287,PMID:4040853,PMID:3124104] are nucleic acid-binding protein structures first identified in the Xenopus transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid residues. There are two cysteine or histidine residues at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides. A schematic representation of a zinc finger domain is shown below: x x x x x x x x x C H x \ / x x Zn x x / \ x C H x x x x x x x x x Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc- dependent DNA or RNA binding property of some members of this class. Some of the proteins known to include C2H2-type zinc fingers are listed below. We have indicated, between brackets, the number of zinc finger regions found in each of these proteins; a '+' symbol indicates that only partial sequence data is available and that additional finger domains may be present. - Saccharomyces cerevisiae: ACE2 (3), ADR1 (2), AZF1 (4), FZF1 (5), MIG1 (2), MSN2 (2), MSN4 (2), RGM1 (2), RIM1 (3), RME1 (3), SFP1 (2), SSL1 (1), STP1 (3), SWI5 (3), VAC1 (1) and ZMS1 (2). - Emericella nidulans: brlA (2), creA (2). - Drosophila: AEF-1 (4), Cf2 (7), ci-D (5), Disconnected (2), Escargot (5), Glass (5), Hunchback (6), Kruppel (5), Kruppel-H (4+), Odd-skipped (4), Odd-paired (4), Pep (3), Snail (5), Spalt-major (7), Serependity locus beta (6), delta (7), h-1 (8), Suppressor of hairy wing su(Hw) (12), Suppressor of variegation

				suvar(3)7 (5), Teashirt (3) and Tramtrack (2). - Xenopus: transcription factor TFIIIA (9), p43 from RNP particle (9), Xfin (37 !), Xsna (5), gastrula XlcGF5.1 to XlcGF71.1 (from 4+ to 11+), Oocyte XlcOF2 to XlcOF22 (from 7 to 12). - Mammalian: basonuclin (6), BCL-6/LAZ-3 (6), erythroid krueppel-like transcription factor (3), transcription factors Sp1 (3), Sp2 (3), Sp3 (3) and Sp(4) 3, transcriptional repressor YY1 (4), Wilms' tumor protein (4), EGR1/Krox24 (3), EGR2/Krox20 (3), EGR3/Pilot (3), EGR4/AT133 (4), Evi-1 (10), GLI1 (5), GLI2 (4+), GLI3 (3+), HIV-EP1/ZNF40 (4), HIV-EP2 (2), KR1 (9+), KR2 (9), KR3 (15+), KR4 (14+), KR5 (11+), HF.12 (6+), REX-1 (4), ZfX (13), ZfY (13), Zfp-35 (18), ZNF7 (15), ZNF8 (7), ZNF35 (10), ZNF42/MZF-1 (13), ZNF43 (22), ZNF46/Kup (2), ZNF76 (7), ZNF91 (36), ZNF133 (3). In addition to the conserved zinc ligand residues it has been shown[PMID:8129873]that a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. The best conserved position is found four residues after the second cysteine; it is generally an aromatic or aliphatic residue. A profile was also developed that spans the whole domain."
20	PF17490	L1 transposable element dsRBD-like domain		This entry represents the double stranded RNA-binding-like domain.
21	IPR004244	"Transposase, L1"		"Many human L1 elements are capable of retrotransposition. Some of these have been shown to exhibit reverse transcriptase (RT) activityalthough the function of many are, as yet, unknown."
22	PF08333	Protein of unknown function (DUF1725)		This family include many eukaryotic and one bacterial sequence. Many of its members are annotated as being putative L1 retrotransposons or LINE-1 reverse transcriptase homologs. The region in question is found repeated in some family members.
23	IPR042497	"Transposase, L1 superfamily"		"This superfamily represents a conserved component in the protein ORF1 of the long interspersed nuclear element 1 (LINE-1 or L1) retrotransposon. This domain is designated RRM, which stands for RNA recognition motif, and is involved in RNA binding. Cooperation between this RRM domain and the C-terminal domain is required to achieve this, despite the fact that this two domains do not contact each other."
24	PF03372	Endonuclease /Exonuclease/ phosphatase family		"This large family of proteins includes magnesium dependent endonucleases and a large number of phosphatases involved in intracellular signalling[PMID:10838565]. This family includes: AP endonuclease proteins EC:4.2.99.18 e.g Swiss:P27695, DNase I proteins EC:3.1.21.1 e.g. Swiss:P24855, Synaptojanin an inositol-1,4,5-trisphosphate phosphatase EC:3.1.3.56 Swiss:O43426, Sphingomyelinase EC:3.1.4.12 Swiss:P11889 and Nocturnin Swiss:O35710."
25	IPR035300	"L1 transposable element, dsRBD-like domain"		This entry represents the double stranded RNA-binding-like domain of LINE-1 (L1) transposable elements.
26	cd09076	"Endonuclease domain (L1-EN) of the non-LTR retrotransposon LINE-1		"This family contains the endonuclease domain (L1-EN) of the non-LTR retrotransposon LINE-1 (L1), and related domains, including the endonuclease of Xenopus laevis Tx1. These retrotranspons belong to the subtype 2, L1-clade. LINES can be classified into two subtypes. Subtype 2 has two ORFs: the second (ORF2) encodes a modular protein consisting of an N-terminal apurine/apyrimidine endonuclease domain (EN), a central reverse transcriptase, and a zinc-finger-like domain at the C-terminus. LINE-1/L1 elements (full length and truncated) comprise about 17% of the human genome. This endonuclease nicks the genomic DNA at the consensus target sequence 5'TTTT-AA3' producing a ribose 3'-hydroxyl end as a primer for reverse

		(L1), and related domains"	transcription of associated template RNA. This subgroup also includes the endonuclease of Xenopus laevis Tx1, another member of the L1-clade. This family belongs to the large EEP (exonuclease/endonuclease/phosphatase) superfamily that contains functionally diverse enzymes that share a common catalytic mechanism of cleaving phosphodiester bonds. [PMID: 15274918, PMID: 17626046, PMID: 9922177, PMID: 12832649, PMID: 11810275, PMID: 8945517, PMID: 11700292, PMID: 15983781, PMID: 16336743, PMID: 18767160, PMID: 12870715]"
27	PF17489	L1 transposable element trimerization domain	This entry represents the trimerization domain.
28	IPR042566	"L1 transposable element, C-terminal domain"	This superfamily represents a conserved component in the protein ORF1 of the long interspersed nuclear element 1 (LINE-1 or L1) retrotransposon. This domain is designated C-terminal and its sole function is to assist RNA binding to the central RRM domain.
29	IPR013544	Domain of unknown function DUF1725	This domain is found at the C terminus of many eukaryotic and a few bacterial sequences. Many of its members are annotated as being putative L1 retrotransposons or LINE-1 reverse transcriptase homologues. The region in question is found repeated in some family members.
30	PF02994	L1 transposable element RBD-like domain	This entry represents the RBD-like domain.
31	IPR035301	"L1 transposable element, trimerization domain"	This entry represents the trimerization domain found in protein ORF1 of the long interspersed nuclear element 1 (LINE-1 or L1).
32	PF02099	Josephin	NA
33	IPR006155	Josephin domain	"The Josephin domain is an eukaryotic protein module of about 180 residues, which occurs in stand-alone form in Josephin-like proteins, and as an amino- terminal domain associated with two or three copies of the ubiquitin- interacting motif (UIM) in ataxin 3-like proteins. Josephin domain-containing proteins function as de-ubiquitination enzymes. Although it has originally been proposed that the Josephin domain could be an all-alpha helical domain distantly related to ENTH and VHS domains involved in membrane trafficking and regulatory adaptor function, it is now believed that it is a mainly alpha helical cysteine-protease domain predicted to be active against ubiquitin chains or related substrates.The Josephin domain contains two conserved histidines and one cysteine that is required for the ubiquitin protease activityand two ubiquitin-binding sites.Some proteins known to contain a Josephin domain are: * Animal Machado-Joseph disease protein 1 (Ataxin 3). It interacts with key regulators (CBP, p300 and PCAF) of transcription and represses transcription. * Plant Machado-Joseph disease-like protein

			(MJD1a-like) (Ataxin 3 homologue). * Mammalian Josephin 1 and 2. * Drosophila melanogaster Josephin-like protein. * Arabidopsis thaliana Josephin-like protein."
33 34	IPR022035	"PCIF1, WW domain"	"This domain is found in bacteria and eukaryotes, and is approximately 180 amino acids in length. It is the WW domain of PCIF1. PCIF1 interacts with phosphorylated RNA polymerase II carboxy-terminal domain (CTD). The WW domain of PCIF1 can directly and preferentially bind to the phosphorylated CTD compared to the unphosphorylated CTD. PCIF1 binds to the hyperphosphorylated RNAP II (RNAP IIO) in vitro and in vivo. Double immunofluorescence labeling in HeLa cells demonstrated that PCIF1 and endogenous RNAP IIO are co-localized in the cell nucleus. PCIF1 is a cap-specific adenosine methyltransferase that catalyzes formation of N6,2'-O-dimethyladenosine cap (m6A(m)) by methylating the adenosine at the second transcribed position of capped mRNAs."
33 35	PF12237	Phosphorylated CTD interacting factor 1 WW domain	"This domain family is found in bacteria and eukaryotes, and is approximately 180 amino acids in length. This domain is the WW domain of PCIF1. PCIF1 interacts with phosphorylated RNA polymerase II carboxy-terminal domain (CTD). The WW domain of PCIF1 can directly and preferentially bind to the phosphorylated CTD compared to the unphosphorylated CTD. PCIF1 binds to the hyperphosphorylated RNAP II (RNAP IIO) in vitro and in vivo. Double immunofluorescence labeling in HeLa cells demonstrated that PCIF1 and endogenous RNAP IIO are co-localized in the cell nucleus. Thus, PCIF1 may play a role in mRNA synthesis by modulating RNAP IIO activity."
33 36	cd14845	"L-Ala-D-Glu peptidase, also known as L-alanyl-D-glutamate endopeptidase "	"This L-Ala-D-Glu peptidase family includes L-alanyl-D-glutamate peptidase (bacteriophage T5) (also known as L-alanoyl-D-glutamate endopeptidase), and Ply118 and Ply500 L-Ala-D-Glu peptidase. Bacteriophage endolysin degrades the peptidoglycan of the bacterial host from within, leading to cell lysis and release of progeny virions. The bacteriophage endolysin Ply118 cleaves between L-Ala and D-Glu residues of Listeria cell wall peptidoglycan. This family belongs to the MEROPS peptidase M15 subfamily C. [PMID: 15044722, PMID: 18560152, PMID: 16789876, PMID: 9702193, PMID: 9614968, PMID: 8852545, PMID: 8631706, PMID: 7810996, PMID: 7121588, PMID: 10500118, PMID: 9265630]"
33 37	IPR015898	G-protein gamma-like domain	"This entry represents the G protein gamma subunit and the GGL (G protein gamma-like) domain, which are related in sequence and are comprised of an extended alpha-helical polypeptide. The G protein gamma subunit forms a stable dimer with the beta subunit, but it does not make any contact with the alpha subunit, which contacts the opposite face of the beta subunit. The GGL domain is found in several RGS (regulators of G protein signaling) proteins. GGL domains can interact with beta subunits to form novel dimers that prevent gamma subunit binding, and may prevent heterotrimer formation by inhibiting alpha subunit binding. The interaction between G protein beta-5 neuro-specific isoforms and RGS GGL domains may represent a general mode of binding between beta-propeller proteins and their partners."
33 38	PS50058	G-protein gamma subunit domain profile	"Guanine nucleotide-binding proteins (G proteins) [1] act as intermediaries in the transduction of signals generated by transmembrane receptors. G proteins consist of three subunits (alpha, beta, and gamma). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. The gamma subunits are small proteins (from 70 to 110 residues) that are bound to the membrane via a isoprenyl group (either a farnesyl or a geranyl- geranyl) covalently linked to their C-terminus. In mammals there are at least 12 different isoforms of gamma subunits. The Caenorhabditis elegans protein egl-10, which is a regulator of G-protein signalling, contains a G-protein gamma-like domain. We have developed a profile that spans the complete length of the gamma subunit."
33 39	cd16857	"Inhibitor of growth (ING) domain of	"ING1 is an epigenetic regulator and a type II tumor suppressor that impacts cell growth, aging, apoptosis, and DNA repair, by affecting chromatin conformation and gene expression. It acts as a reader of the active chromatin mark, the trimethylation of histone H3 lysine 4 (H3K4me3). It binds and directs growth arrest and DNA damage inducible protein 45 a (Gadd45a) to target

		inhibitor of growth protein ING1, ING2, and similar proteins"	sites, thus linking the histone code with DNA demethylation. It interacts with the proliferating cell nuclear antigen (PCNA) via the PCNA-interacting protein (PIP) domain in a UV-inducible manner. It also interacts with a PCNA-interacting protein, p15 (PAF). Moreover, ING1 associates with members of the 14-3-3 family, which is necessary for cytoplasmic relocalization. Endogenous ING1 protein specifically interacts with the pro-apoptotic BCL2 family member BAX and colocalizes with BAX in a UV-inducible manner. It stabilizes the p53 tumor suppressor by inhibiting polyubiquitination of multi-monoubiquitinated forms via interaction with and colocalization of the herpesvirus-associated ubiquitin-specific protease (HAUSP)-deubiquitinase with p53. It is also involved in trichostatin A-induced apoptosis and caspase 3 signaling in p53-deficient glioblastoma cells. In addition, tyrosine kinase Src can bind and phosphorylate ING1 and further regulates its activity. ING2, also termed inhibitor of growth 1-like protein (ING1Lp), or p32, or p33ING2, is a core component of a multi-factor chromatin-modifying complex containing the transcriptional co-repressor SIN3A and histone deacetylase 1 (HDAC1). It has been implicated in the control of cell cycle, in genome stability, and in muscle differentiation. ING2 independently interacts with H3K4me3 (Histone H3 trimethylated on lysine 4) and PtdIns(5)P, and modulates crosstalk between lysine methylation and lysine acetylation on histone proteins through association with chromatin in the presence of DNA damage. It collaborates with SnoN to mediate transforming growth factor (TGF)-beta-induced Smad-dependent transcription and cellular responses. It is upregulated in colon cancer and increases invasion by enhanced MMP13 expression. It also acts as a cofactor of p300 for p53 acetylation and plays a positive regulatory role during p53-mediated replicative senescence. Both ING1 and ING2 contain an N-terminal leucine zipper-like (LZL) motif-containing ING domain, and a well-characterized C-terminal plant homeodomain (PHD)-type zinc-finger domain. [PMID: 18450387, PMID: 18836436, PMID: 18655775, PMID: 19085961, PMID: 19442115, PMID: 20066899, PMID: 20179197, PMID: 20681406, PMID: 21078114, PMID: 21227930, PMID: 21286670, PMID: 21707358, PMID: 21731648, PMID: 21896275, PMID: 22163049, PMID: 23388825, PMID: 23459830, PMID: 23585863, PMID: 23803967, PMID: 19766113, PMID: 19730436, PMID: 19442116, PMID: 19437536, PMID: 19184981, PMID: 18334480, PMID: 18160212, PMID: 17208481, PMID: 16024799, PMID: 18533182, PMID: 8944021, PMID: 10626813, PMID: 10866301, PMID: 10807544, PMID: 9440695, PMID: 24008732, PMID: 19962781, PMID: 20676127, PMID: 20621832, PMID: 20890119, PMID: 21124965, PMID: 20948283, PMID: 21767350, PMID: 22808232, PMID: 23823870, PMID: 23864195, PMID: 16728974, PMID: 16387653, PMID: 16893883, PMID: 12859901, PMID: 11784859, PMID: 11481424, PMID: 10072587, PMID: 16728977, PMID: 23412501]"
40	PF09004	Domain of unknown function (DUF1891)	This domain is found in a set of hypothetical eukaryotic proteins.
41	IPR015095	"Alkylated DNA repair protein AlkB, homologue 8, N-terminal"	This domain is found at the extreme N terminus of eukaryotic alkylated DNA repair protein homologues.
42	PF00098	Zinc knuckle	"The zinc knuckle is a zinc binding motif composed of the the following CX2CX4HX4C where X can be any amino acid. The motifs are mostly from retroviral gag proteins (nucleocapsid). Prototype structure is from HIV. Also contains members involved in eukaryotic gene regulation, such as C. elegans GLH-1. Structure is an 18-residue zinc finger."
43	IPR026960	Reverse transcriptase	This domain would appear to be a zinc-binding region of a putative reverse transcriptase.

		zinc-binding domain	
4 4	PF139 66	zinc-binding in reverse transcriptase	This domain would appear to be a zinc-binding region of a putative reverse transcriptase.
4 5	cd059 20	"2,3-dihydroxyben zoate-AMP ligase"	"2,3-dihydroxybenzoate-AMP ligase activates 2,3-dihydroxybenzoate (DHB) by ligation of AMP from ATP with the release of pyrophosphate. However, it can also catalyze the ATP-PPi exchange for 2,3-DHB analogs, such as salicyclic acid (o-hydrobenzoate), as well as 2,4-DHB and 2,5-DHB, but with less efficiency. Proteins in this family are the stand-alone adenylation components of non-ribosomal peptide synthases (NRPSs) involved in the biosynthesis of siderophores, which are low molecular weight iron-chelating compounds synthesized by many bacteria to aid in the acquisition of this vital trace elements. In Escherichia coli, the 2,3-dihydroxybenzoate-AMP ligase is called EntE, the adenylation component of the enterobactin NRPS system. [PMID: 12221282, PMID: 21166461, PMID: 16678412, PMID: 2531000, PMID: 11018148, PMID: 16632253, PMID: 9373621, PMID: 17762044, PMID: 11255012]"
4 6	IPR04 2509	Zinc finger CCHC domain-containing protein 3	"Zinc finger CCHC domain-containing protein 3 (ZCCHC3) is a zinc-finger protein that acts as a co-sensor of cGAS in innate immune response to cytosolic dsDNA and DNA virus. It has been shown to bind DNA and RNA in the cytoplasm and acts by promoting recognition of viral nucleic acids by virus sensors, such as DDX58/RIG-I, IFIH1/MDA5 and CGAS."
4 7	cd058 44	glycosyltransf erase family 4 proteins	"Glycosyltransferases catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds. The acceptor molecule can be a lipid, a protein, a heterocyclic compound, or another carbohydrate residue. This group of glycosyltransferases is most closely related to glycosyltransferase family 4 (GT4). The members of this family may transfer UDP, ADP, GDP, or CMP linked sugars. The diverse enzymatic activities among members of this family reflect a wide range of biological functions. The protein structure available for this family has the GTB topology, one of the two protein topologies observed for nucleotide-sugar-dependent glycosyltransferases. GTB proteins have distinct N- and C- terminal domains each containing a typical Rossmann fold. The two domains have high structural homology despite minimal sequence homology. The large cleft that separates the two domains includes the catalytic center and permits a high degree of flexibility. [PMID: 12624205, PMID: 7596293, PMID: 16564135, PMID: 16037492, PMID: 16263268, PMID: 15653326, PMID: 12691742]"
4 8	cd002 29	"SGNH_hydr olase, or GDSL_hydrol ase, is a diverse family of lipases and esteras..."	"SGNH_hydrolase, or GDSL_hydrolase, is a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the typical Ser-His-Asp(Glu) triad from other serine hydrolases, but may lack the carboxlic acid. [PMID: 7610479, PMID: 15522763, PMID: 12842470, PMID: 10801485]"
4 9	PS003 54	Putative nuclease HARBI1	NA
5 0	IPR00 0637	"HMG-I/HMG-Y,	"High mobility group (HMG) proteins are a family of relatively low molecular weight non-histone components in chromatin. HMG-I and HMG-Y (HMGA) are proteins of about 100 amino acid residues which are produced by the alternative splicing of a

		DNA-binding, conserved site"	single gene. HMG-I/Y proteins bind preferentially to the minor groove of AT-rich regions in double-stranded DNA in a non-sequence specific manner. It is suggested that these proteins could function in nucleosome phasing and in the 3' end processing of mRNA transcripts. They are also involved in the transcription regulation of genes containing, or in close proximity to, AT-rich regions. DNA-binding of these, and several related, proteins is effected by an 11-residue domain known as an AT-hook. Within known HMG-I/Y proteins are found three highly conserved regions, closely related to the consensus sequence TPKRPRGRPKK. A synthetic oligopeptide with this sequence specifically binds to substrate DNA in a manner reminiscent of intact HMG-I/Y proteins. Structure predictions suggest that the peptide has a secondary structure similar to the anti-tumour and anti-viral drugs netropsin and distamycin, and to the dye Hoechst 33258. These ligands, which also preferentially bind to AT-rich DNA, effectively compete with both the synthetic peptide and the HMG-I/Y proteins for DNA binding. The peptide also contains novel structural features such as a predicted Asx bend, or 'hook', at its N terminus, and laterally-projecting cationic Arg/Lys 'bristles', which may play a role in the binding of HMG-I/Y proteins. The predicted peptide structure, the AT-hook, is a previously undescribed DNA-binding motif."
5 1	IPR001952	Alkaline phosphatase	"This entry represents alkaline phosphatases () (ALP), which act as non-specific phosphomonoesterases to hydrolyse phosphate esters, optimally at high pH. The reaction mechanism involves the attack of a serine alkoxide on a phosphorus of the substrate to form a transient covalent enzyme-phosphate complex, followed by the hydrolysis of the serine phosphate. Alkaline phosphatases are found in all kingdoms of life, with the exception of some plants. Alkaline phosphatases are metalloenzymes that exist as a dimer, each monomer binding metal ions. The metal ions they carry can differ, although zinc and magnesium are the most common. For example, Escherichia coli alkaline phosphatase (encoded by phoA) requires the presence of two zinc ions bound at the M1 and M2 metal sites, and one magnesium ion bound at the M3 site. However, alkaline phosphatases from Thermotoga maritima and Bacillus subtilis require cobalt for maximal activity.In mammals, there are four alkaline phosphatase isozymes: placental, placental-like (germ cell), intestinal and tissue-nonspecific (liver/bone/kidney). All four isozymes are anchored to the outer surface of the plasma membrane by a covalently attached glycosylphosphatidylinositol (GPI) anchor. Human alkaline phosphatases have four metal binding sites: two for zinc, one for magnesium, and one for calcium ion. Placental alkaline phosphatase (ALPP or PLAP) is highly polymorphic, with at least three common alleles. Its activity is down-regulated by a number of effectors such as l-phenylalanine, 5'-AMP, and by p-nitrophenyl-phosphonate (PNPPate). The placental-like isozyme (ALPPL or PLAP-like) is elevated in germ cell tumours. The intestinal isozyme (ALPI or IAP) has the ability to detoxify lipopolysaccharide and prevent bacterial invasion across the gut mucosal barrier. The tissue-nonspecific isozyme (ALPL) is, and may play a role in skeletal mineralisation. Defects in ALPL are a cause of hypophosphatasia, including infantile-type (OMIM:241500), childhood-type (OMIM:241510) and adult-type (OMIM:146300). Hhypophosphatasia is an inherited metabolic bone disease characterised by defective skeletal mineralisation.This entry also contains the related enzyme streptomycin-6-phosphate phosphatase () (encoded by strK) from Streptomyces species. This enzyme is involved in the synthesis of the antibiotic streptomycin, specifically cleaving both streptomycin-6-phosphate and, more slowly, streptomycin-3-phosphate."
5 2	IPR040260	Replication factor A protein-like	"Rfa2 (also known as RPA32) is a component of the replication protein A (RPA) complex, which binds to and removes secondary structure from ssDNA. The RPA complex is involved in DNA replication, repair, and recombination.Stn1 is a component of the CST complex, a complex that binds to single-stranded DNA and is required to protect telomeres from DNA degradation. The CST complex binds single-stranded DNA with high affinity in a sequence-independent manner, while isolated subunits bind DNA with low affinity by themselves. In addition to telomere protection, the CST complex has probably a more general role in DNA metabolism at non-telomeric sites.This entry contains proteins with a winged helix DNA-binding domain, including replication factor A protein 2 (RFA2) and CST complex subunit STN1."

5353	PS01359	CUB domain profile	NA
5354	IPR019786	"Zinc finger, PHD-type, conserved site"	"Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.This entry represents the PHD (homeodomain) zinc finger domain, which is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation. The PHD finger motif is reminiscent of, but distinct from the C3HC4 type RING finger.The function of this domain is not yet known but in analogy with the LIM domain it could be involved in protein-protein interaction and be important for the assembly or activity of multicomponent complexes involved in transcriptional activation or repression. Alternatively, the interactions could be intra-molecular and be important in maintaining the structural integrity of the protein. In similarity to the RING finger and the LIM domain, the PHD finger is thought to bind two zinc ions.The signature of this entry starts at the first cysteine of the zinc finger region and ends at the last one. The spacing between cysteines in the PHD finger is closely related to that in the RING finger. Discrimination between these two domains with either a pattern or a profile is therefore difficult, and some rare domains are recognised by both the RING and PHD patterns and profiles."
5355	PF03258	Baculovirus FP protein	The FP protein is missing in baculovirus (Few Polyhedra) mutants[PMID:8760443].
5356	IPR019787	"Zinc finger, PHD-finger"	"Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.This entry represents the PHD (homeodomain) zinc finger domain, which is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation. The PHD finger motif is reminiscent

			of, but distinct from the C3HC4 type RING finger.The function of this domain is not yet known but in analogy with the LIM domain it could be involved in protein-protein interaction and be important for the assembly or activity of multicomponent complexes involved in transcriptional activation or repression. Alternatively, the interactions could be intra-molecular and be important in maintaining the structural integrity of the protein. In similarity to the RING finger and the LIM domain, the PHD finger is thought to bind two zinc ions."
5 7	IPR01 1011	"Zinc finger, FYVE/PHD-type"	"Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.The FYVE zinc finger domain is conserved from yeast to man, and is named after four proteins that it has been found in: Fab1, YOTB/ZK632.12, Vac1, and EEA1. It functions in the membrane recruitment of cytosolic proteins by binding to phosphatidylinositol 3-phosphate (PI3P), which is found mainly on endosomes.The plant homeodomain (PHD) zinc finger domain has a C4HC3-type motif, and is widely distributed in eukaryotes, being found in many chromatin regulatory factors.Both the FYVE and the PHD zinc finger motifs display strikingly similar dimetal(zinc)-bound alpha+beta folds."
5 8	PS500 16	Zinc finger PHD-type profile	"The Plant Homeodomain (PHD) type zinc finger (C4HC3)[PMID:7701562]is found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation. The PHD finger motif is reminiscent of, but distinct from the C3HC4-type RING finger. A general function of the PHD finger is not yet known, but in the CREB-binding protein (CBP) it is an integral part of the enzymatic core of the acetyltransferase domain and is also a key contributor to the transcriptional activity of CBP. The 3D structure of Williams-Beuren Syndrome transcription factor reveals that the PHD finger coordinates two atoms of zinc in a cross-braced manner[PMID:11124022]. This atypical conformation is also shared by the FYVE. The two zinc sites are located at opposite ends of a short beta-sheet. This feature is also shared by the RING finger and the FYVE domain, but the structure differ in the surrounding loops that vary significantly in length and sequence and could explain the wide variety of functions between these three domains. Single or multiple PHD fingers are found in the following proteins: - Drosophila trithorax protein and the homologous mammalian protein HRX (also known as ALL1, MLL and Htrx). The trithorax gene is required for proper regulation of the homeotic genes in flies[PMID:8098166]. The HRX gene is involved in translocations in acute lymphoblastic leukemia. Trithorax has 4 PHD fingers. The most C-terminal one in HRX is degenerated. - Drosophila Polycomblike, a protein encoded by a Polycomb-group gene. Polycomblike and the Polycomb-group genes are required for correct regulation of the homeotic genes in flies. These genes counteract the effects of the trithorax-group genes presumably by creating an inactive chromatin structure[PMID:7907493]. Polycomblike has 2 PHD fingers. - Human Xe169 and murine Smcx, proteins encoded by X-linked genes that escape X-chromosome inactivation. Murine Smcy is a related gene on the Y chromosome. Xe169 has 2 PHD fingers. - Human RBP2, a Xe169-related protein thought to bind to the Rb protein. - Human Br140 (also known as peregrin), a protein which also has a bromodomain. - Medicago sativa Alfin-1, a presumptive transcription

			factor. - A group of plant homeobox-containg proteins such as Arabidopsis HAT3 and PRHA, maize ZMHOX1a, parsley PRHP. - Human Mi-2 autoantigen. - Drosophila shuttle craft protein. - Human BS69, a protein that binds adenovirus E1A. - Murine CBP (CREB-binding protein), the closely related human E1A binding protein p300, and YNJ1, a related protein from C.elegans - Mammalian zinc-finger protein of the 'requiem' family. - Human TIF1, a putative transcriptional mediator for nuclear receptors. The consensus pattern we developed start at the first cysteine and end at the last one, whereas the profile we developed spans the complete domain."
59	IPR013083	"Zinc finger, RING/FYVE/PHD-type"	"Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.This superfamily represents RING-, PHD-, and FYVE-type zinc finger domains, which share a common dimetal (zinc)-bound alpha/beta structural fold, as well as the non-zinc-containing U-box domain, which is similar to the RING zinc finger only lacking the metal ion-binding residues (U-box associated with multi-ubiquitination)."
60	cd15489	PHD finger superfamily	"The PHD finger superfamily includes a canonical plant homeodomain (PHD) finger typically characterized as Cys4HisCys3, and a non-canonical extended PHD finger, characterized as Cys2HisCys5HisCys2His. Variations include the RAG2 PHD finger characterized by Cys3His2Cys2His and the PHD finger 5 found in nuclear receptor-binding SET domain-containing proteins characterized by Cys4HisCys2His. The PHD finger is also termed LAP (leukemia-associated protein) motif or TTC (trithorax consensus) domain. Single or multiple copies of PHD fingers have been found in a variety of eukaryotic proteins involved in the control of gene transcription and chromatin dynamics. PHD fingers can recognize the unmodified and modified histone H3 tail, and some have been found to interact with non-histone proteins. They also function as epigenome readers controlling gene expression through molecular recruitment of multi-protein complexes of chromatin regulators and transcription factors. The PHD finger domain SF is structurally similar to the RING and FYVE_like superfamilies. [PMID: 12480335, PMID: 16297627, PMID: 17949986, PMID: 18682256, PMID: 19011376, PMID: 19442115, PMID: 19925420, PMID: 20048137, PMID: 21813457, PMID: 24291127, PMID: 11884585, PMID: 7568208, PMID: 7701562, PMID: 7583761, PMID: 21514168, PMID: 11124022, PMID: 17142463]"
61	IPR004941	Baculovirus FP protein	The function of the FP protein is not known. The protein is missing in baculovirus FP (Few Polyhedra) mutants.
62	IPR036514	SGNH hydrolase superfamily	"SGNH hydrolase has a similar fold to flavoproteins, namely a three-layer alpha/beta/alpha structure, where the beta-sheets are composed of five parallel strands. Enzymes containing this domain act as esterases and lipases, but have little sequence homology to true lipases. Proteins containing this type of esterase domain have been found in a variety of hydrolases; those with structural information include an esterase from Streptomyces scabies; the esterase domain of viral haemagglutinin-esterase surface glycoproteins (influenza C virus, coronaviruses and toroviruses); mammalian acetylhydrolases; fungal

				rhamnogalacturonan acetylesterase; and the multifunctional enzyme thioesterase I (TAP) from Escherichia coli. SGNH hydrolase-type esterase domains contain unique hydrogen bond network that stabilises their catalytic centres; they usually contain the catalytic triad Ser/Acid/His."
63	PF13472	GDSL-like Lipase/Acylhydrolase family	This family of presumed lipases and related enzymes are similar to Pfam:PF00657.	
64	IPR013830	SGNH hydrolase-type esterase domain	"This entry represents the SGNH hydrolase-type esterase domain, which has a similar fold to flavoproteins, namely a three-layer alpha/beta/alpha structure, where the beta-sheets are composed of five parallel strands. Enzymes containing this type of esterase domain act as esterases and lipases, but have little sequence homology to true lipases. Proteins containing this type of esterase domain have been found in a variety of hydrolases; those with structural information include an esterase from Streptomyces scabies; the esterase domain of viral haemagglutinin-esterase surface glycoproteins (influenza C virus, coronaviruses and toroviruses); mammalian acetylhydrolases; fungal rhamnogalacturonan acetylesterase; and the multifunctional enzyme thioesterase I (TAP) from Escherichia coli. SGNH hydrolase-type esterase domains contain unique hydrogen bond network that stabilises their catalytic centres; they usually contain the catalytic triad Ser/Acid/His."
65	IPR027124	SWR1-complex protein 5/Craniofacial development protein 2	"This family includes SWR1-complex protein 5 (Swc5), and craniofacial development protein 2 (CFDP2).Swc5 is a component of the SWR1 complex which mediates the ATP-dependent exchange of histone H2A for the H2A variant HZT1 leading to transcriptional regulation of selected genes by chromatin remodeling. It is involved in chromosome stability.Gene duplication of the ancestral bnct (bucentaur) gene leads to the h-type bnct(cfdp1) gene and the p97bcnt (cfdp2) gene. Cfdp2 is rumiant-specific, containing a region derived from the endonuclease domain of a retrotransposable element RTE-1 not found in human and mouse. For this reason bcnt has been used as a model for molecular evolution."
66	IPR018527	"Rubredoxin, iron-binding site"	"Rubredoxin is a low molecular weight iron-containing protein involved in electron transfer, sometimes replacing ferredoxin as an electron carrier. It is found in bacteria, archaea and plants.The 3-D structures of a number of rubredoxins have been solved. The fold belongs to the alpha+beta class, with 2 alpha-helices and 2-3 beta-strands. Its active site contains an iron ion which is co-ordinated by the sulphurs of four conserved cysteine residues forming an almost regular tetrahedron. The conserved cysteines reside on two loops, which are the most conserved regions of the protein. In addition, a ring of acidic residues in the proximity of the [Fe(Cys)4] centre is also well-conserved.This entry represents a conserved region which includes two of the cysteine residues that bind the iron atom."
67	PS00202	Putative nuclease HARBI1	NA	
68	IPR001965	"Zinc finger, PHD-type"	"Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved

			specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.This entry represents the PHD (homeodomain) zinc finger domain, which is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation. The PHD finger motif is reminiscent of, but distinct from the C3HC4 type RING finger.The function of this domain is not yet known but in analogy with the LIM domain it could be involved in protein-protein interaction and be important for the assembly or activity of multicomponent complexes involved in transcriptional activation or repression. Alternatively, the interactions could be intra-molecular and be important in maintaining the structural integrity of the protein. In similarity to the RING finger and the LIM domain, the PHD finger is thought to bind two zinc ions."
6 9	cd15517	"PHD finger found in Transcription factor 19 (TCF-19), Lysine-specific demethylase KDM5A and KDM5B, and other similar proteins"	"TCF-19 was identified as a putative trans-activating factor with expression beginning at the late G1-S boundary in dividing cells. It functions as a novel islet factor necessary for proliferation and survival in the INS-1 beta cell line. It plays an important role in susceptibility to both Type 1 Diabetes Mellitus (T1DM) and Type 2 Diabetes Mellitus (T2DM); it has been suggested that it may positively impact beta cell mass under conditions of beta cell stress and increased insulin demand. KDM5A was originally identified as a retinoblastoma protein (Rb)-binding partner and its inactivation may be important for Rb to promote differentiation. It is involved in transcription through interaction with TBP, p107, nuclear receptors, Myc, Sin3/HDAC, Mad1, RBP-J, CLOCK, and BMAL1. KDM5B has a restricted expression pattern in the testis, ovary, and transiently in the mammary gland of the pregnant female and has been shown to be upregulated in breast cancer, prostate cancer, and lung cancer, suggesting a potential role in tumorigenesis. Both KDM5A and KDM5B function as trimethylated histone H3 lysine 4 (H3K4me3) demethylases. This family also includes Caenorhabditis elegans Lysine-specific demethylase 7 homolog (ceKDM7A). ceKDM7A (also termed JmjC domain-containing protein 1.2, PHD finger protein 8 homolog, or PHF8 homolog) is a plant homeodomain (PHD)- and JmjC domain-containing protein that functions as a histone demethylase specific for H3K9me2 and H3K27me2. The binding of the PHD finger to H3K4me3 guides H3K9me2- and H3K27me2-specific demethylation by its catalytic JmjC domain in a trans-histone regulation mechanism. In addition, this family includes plant protein OBERON 1 and OBERON 2, Alfin1-like (AL) proteins, histone acetyltransferases (HATs) HAC, and AT-rich interactive domain-containing protein 4 (ARID4). [PMID: 1868030, PMID: 10777095, PMID: 23860123, PMID: 10336460, PMID: 10616211, PMID: 10878660, PMID: 12237901, PMID: 12657635, PMID: 15803180, PMID: 16645588, PMID: 17320161, PMID: 17311883, PMID: 17373667, PMID: 17363312, PMID: 23907094, PMID: 23777767, PMID: 23637629, PMID: 23408432, PMID: 22020125, PMID: 22420752, PMID: 22615382, PMID: 21980403, PMID: 21937684, PMID: 21821892, PMID: 21369698, PMID: 21105039, PMID: 20863814, PMID: 20403335, PMID: 19336002, PMID: 18591252, PMID: 18048344, PMID: 17973255, PMID: 17709396, PMID: 20406991, PMID: 21348942, PMID: 21562575, PMID: 21960634, PMID: 20231316, PMID: 22937203, PMID: 23093672, PMID: 23112189, PMID: 17573780, PMID: 18483221, PMID: 18270511, PMID: 18722178, PMID: 19762557, PMID: 19850045, PMID: 21788502, PMID: 23722541, PMID: 17320160, PMID: 17320163, PMID: 15949438, PMID: 11358960, PMID: 9129143, PMID: 7935440, PMID: 1857421, PMID: 8414517, PMID: 17351631, PMID: 17351630, PMID: 20567261, PMID: 20567262, PMID: 20346720, PMID: 18403411, PMID: 19392692, PMID: 14963126, PMID: 19154204, PMID: 12466527, PMID: 11160878]"
7 0	PF01541	GIY-YIG catalytic domain	"This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded

				antiparallel beta-sheet flanked by three helices[PMID:12379841]. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site."
7 1	IPR00 0305	GIY-YIG endonuclease		"Nucleases of the GIY-YIG family are involved in many cellular processes, including DNA repair and recombination, transfer of mobile genetic elements, and restriction of incoming foreign DNA. The GIY-YIG superfamily groups together nucleases characterised by the presence of a domain of typically ~100 amino acids, with two short motifs ""GIY"" and ""YIG"" in the N-terminal part, followed by an Arg residue in the centre and a Glu residue in the C-terminal part.The GIY-YIG domain forms a compact structural domain, which serves as a scaffold for the coordination of a divalent metal ion required for catalysis of the phosphodiester bond cleavage. The GIY-YIG domain has an alpha/bera-sandwich architecture with a central three-stranded antiparallel beta-sheet flanked by three-helices. The three-stranded anti-parallel beta-sheet contains the GIY-YIG sequence elements. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site.The GIY-YIG domain has been implicated in a variety of cellular processes involving DNA cleavage, from self-propagation with or without introns, to restriction of foreign DNA, to DNA repair and maintenance of genome stability.Some proteins known to contain a GIY-YIG domain include: * Eukaryotic Slx-1 proteins, involved in the maintenance of the rDNA copy number. They have a C-terminal RING finger Zn-binding domain. * Mamalian ankyrin repeat and LEM domain- containing protein 1 (ANKLE1). * Bacterial and archaeal UvrC subunits of (A)BC excinucleases, which remove damaged nucleotides by incising the damaged strand on both sides of the lesion. Paramecium bursaria Chlorella virus 1 (pbvc1). * Phage T4 endonucleases SegA to E, probably involved in the movement of the endonuclease-encoding DNA. * Phage T4 intron-associated endonuclease 1 (I-TevI), specific to the thymidylate synthase (td) gene splice junction and involved in intron homing."
7 2	cd104 42	Catalytic GIY-YIG endonuclease domain of penelope-like elements and similar proteins		"This model corresponds to the EN domain of PLEs that contains catalytic module of the GIY-YIG endonucleases of group I bacterial/organellar introns, as well as bacterial UvrC DNA repair proteins. It can cleave DNA with low nucleotide sequence specificity. However, the PLEs EN domain is distinct from other GIY-YIG endonucleases by the presence of a well-conserved CCHH motif (CX(2-7)CX(33-39)HX(3-5)H, X can be any residue). The role of the CCHH motif has not yet been identified. Penelope-like elements (PLEs) represent a novel class of eukaryotic retroelements, which do not belong to either long terminal repeat (LTR) retrotransposons or non-LTR retrotransposons (often called LINEs), but instead form a sister clade to telomerase reverse transcriptases (TERTs), highly specialized non-mobile reverse transcriptases (RTs) which are responsible for the addition of telomeric repeats to the ends of eukaryotic chromosomes. The single open reading frame (ORF) encoded by PLE consists of two principal domains, RT domain and endonuclease (EN) domain, jointed by a linker region of variable length. Both of these two domains are functionally active. [PMID: 17345670, PMID: 16093704, PMID: 15465912, PMID: 18261821, PMID: 12743378, PMID: 16646971]"
7 3	IPR03 5901	GIY-YIG endonuclease superfamily		"Nucleases of the GIY-YIG family are involved in many cellular processes, including DNA repair and recombination, transfer of mobile genetic elements, and restriction of incoming foreign DNA. The GIY-YIG superfamily groups together nucleases characterised by the presence of a domain of typically ~100 amino acids, with two short motifs ""GIY"" and ""YIG"" in the N-terminal part, followed by an Arg residue in the centre and a Glu residue in the C-terminal part.The GIY-YIG domain forms a compact structural domain, which serves as a scaffold for the coordination of a divalent metal ion required for catalysis of the phosphodiester bond cleavage. The GIY-YIG domain has an alpha/bera-sandwich architecture with a central three-stranded antiparallel beta-sheet flanked by three-helices. The three-stranded anti-parallel beta-sheet contains the GIY-YIG sequence elements. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site.The GIY-YIG domain has been implicated in a variety of cellular processes involving DNA cleavage, from self-propagation with or without introns, to restriction of foreign DNA, to DNA repair and maintenance of genome

				stability.Some proteins known to contain a GIY-YIG domain include: * Eukaryotic Slx-1 proteins, involved in the maintenance of the rDNA copy number. They have a C-terminal RING finger Zn-binding domain. * Mamalian ankyrin repeat and LEM domain- containing protein 1 (ANKLE1). * Bacterial and archaeal UvrC subunits of (A)BC excinucleases, which remove damaged nucleotides by incising the damaged strand on both sides of the lesion. Paramecium bursaria Chlorella virus 1 (pbvc1). * Phage T4 endonucleases SegA to E, probably involved in the movement of the endonuclease-encoding DNA. * Phage T4 intron-associated endonuclease 1 (I-TevI), specific to the thymidylate synthase (td) gene splice junction and involved in intron homing.This entry represents GIY-YIG nuclease superfamily, and identifies the GIY-YIG domain in diverse and distantly related superfamily members, such as Slx1."
74	PS50164	GIY-YIG domain profile		"Nucleases of the GIY-YIG family are involved in many cellular processes, including DNA repair and recombination, transfer of mobile genetic elements, and restriction of incoming foreign DNA. The GIY-YIG superfamily groups together nucleases characterized by the presence of a domain of typically ~100 amino acids, with two short motifs ""GIY"" and ""YIG"" in the N-terminal part, followed by an Arg residue in the center and a Glu residue in the C-terminal part [PMID:10219084,PMID:12379841,PMID:15692561,PMID:16646971,PMID:19361436]. The GIY-YIG domain forms a compact structural domain, which serves as a scaffold for the coordination of a divalent metal ion required for catalysis of the phosphodiester bond cleavage. The GIY-YIG domain has an alpha/beta- sandwich architecture with a central three-stranded antiparallel beta-sheet flanked by three-helices. The three-stranded anti-parallel beta-sheet contains the GIY-YIG sequence elements. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site [PMID:12379841,PMID:15692561,PMID:16646971,PMID:19361436]. Some proteins known to contain a GIY-YIG domain are listed below: - Eukaryotic Slx-1 proteins, involved in the maintenance of the rDNA copy number. They have a C-terminal RING finger Zn-binding domain. - Mamalian ankyrin repeat and LEM domain- containing protein 1 (ANKLE1). - Bacterial and archaeal UvrC subunits of (A)BC excinucleases, which remove damaged nucleotides by incising the damaged strand on both sides of the lesion. - Paramecium bursaria Chlorella virus 1 (pbvc1). - Phage T4 endonucleases SegA to E, probably involved in the movement of the endonuclease-encoding DNA. - Phage T4 intron-associated endonuclease 1 (I-TevI), specific to the thymidylate synthase (td) gene splice junction and involved in intron homing. The profile we developed covers the entire GIY-YIG domain."
75	PF14111	Domain of unknown function (DUF4283)		"This domain family is found in plants, and is approximately 100 amino acids in length. Considering the very diverse range of other domains it is associated with it is possible that this domain is a binding/guiding region. There are two highly conserved tryptophan residues."
76	IPR025558	Domain of unknown function DUF4283		"This domain is found in plant proteins. Considering the very diverse range of other domains it is associated with, it is possible that this domain is a binding/guiding region. It contains two highly conserved tryptophan residues."
77	IPR040256	Uncharacterized protein At4g02000-like		"This entry represents a group of plant proteins, including At4g02000. Their function is not clear."
78	PF14392	Zinc knuckle		The zinc knuckle is a zinc binding motif composed of the the following CX2CX4HX4C where X can be any amino acid. This particular family is found in plant proteins.

79	cd06222	"Ribonuclease H-like superfamily, including RNase H, HI, HII, HIII, and RNase-like domain IV of spliceosomal protein Prp8"	"Ribonuclease H (RNase H) enzymes are divided into two major families, Type 1 and Type 2, based on amino acid sequence similarities and biochemical properties. RNase H is an endonuclease that cleaves the RNA strand of an RNA/DNA hybrid in a sequence non-specific manner in the presence of divalent cations. It is widely present in various organisms, including bacteria, archaea, and eukaryotes. Most prokaryotic and eukaryotic genomes contain multiple RNase H genes. Despite the lack of amino acid sequence homology, type 1 and type 2 RNase H share a main-chain fold and steric configurations of the four acidic active-site residues and have the same catalytic mechanism and functions in cells. RNase H is involved in DNA replication, repair and transcription. An important RNase H function is to remove Okazaki fragments during DNA replication. RNase H inhibitors have been explored as anti-HIV drug targets since RNase H inactivation inhibits reverse transcription. This model also includes the Prp8 domain IV, which adopts the RNase fold but shows low sequence homology; domain IV is implicated in key spliceosomal interactions. [PMID: 19228198, PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265, PMID: 14734815]"
80	PF13456	Reverse transcriptase-like	This domain is found in plants and appears to be part of a retrotransposon.
81	IPR025836	Zinc knuckle CX2CX4HX4C	This zinc knuckle is a zinc binding motif composed of the sequence CX2CX4HX4C where X can be any amino acid.
82	PS51257	Prokaryotic membrane lipoprotein lipid attachment site profile	"In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached[PMID:2202727]. Some of the proteins known to undergo such processing currently include (for recent listings see [PMID:2202727,PMID:3253732]): - Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp). - Escherichia coli lipoprotein-28 (gene nlpA). - Escherichia coli lipoprotein-34 (gene nlpB). - Escherichia coli lipoprotein nlpC. - Escherichia coli lipoprotein nlpD. - Escherichia coli osmotically inducible lipoprotein B (gene osmB). - Escherichia coli osmotically inducible lipoprotein E (gene osmE). - Escherichia coli peptidoglycan-associated lipoprotein (gene pal). - Escherichia coli rare lipoproteins A and B (genes rplA and rplB). - Escherichia coli copper homeostasis protein cutF (or nlpE). - Escherichia coli plasmids traT proteins. - Escherichia coli Col plasmids lysis proteins. - A number of Bacillus beta-lactamases. - Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA). - Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB). - Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7). - Chlamydia trachomatis outer membrane protein 3 (gene omp3). - Fibrobacter succinogenes endoglucanase cel-3. - Haemophilus influenzae proteins Pal and Pcp. - Klebsiella pullulunase (gene pulA). - Klebsiella pullulunase secretion protein pulS. - Mycoplasma hyorhinis protein p37. - Mycoplasma hyorhinis variant surface antigens A, B, and C (genes vlpABC). - Neisseria outer membrane protein H.8. - Pseudomonas aeruginosa lipopeptide (gene lppL). - Pseudomonas solanacearum endoglucanase egl. - Rhodopseudomonas viridis reaction center cytochrome subunit (gene cytC). - Rickettsia 17 Kd antigen. - Shigella flexneri invasion plasmid proteins mxiJ and mxiM. - Streptococcus pneumoniae oligopeptide transport protein A (gene amiA). - Treponema pallidium 34 Kd antigen. - Treponema pallidium membrane protein A (gene tmpA). - Vibrio harveyi chitobiase (gene chb). - Yersinia virulence plasmid protein yscJ. - Halocyanin from Natrobacterium pharaonis[PMID:8195126], a membrane associated copper- binding protein. This is the first archaebacterial protein known to be modified in such a fashion). From the precursor sequences of all these proteins, we derived a profile that starts at the beginning of the sequence and ends after the post-translationally modified cysteine."

83	cd05654	"M20 Peptidase arginine utilization protein, RocB"	"Peptidase M20 family, ArgE RocB (arginine utilization protein, RocB; arginine degradation protein, RocB) subfamily. This group of proteins is possibly related to acetylornithine deacetylase (ArgE) and may be involved in the arginine and/or ornithine degradation pathway. In Bacillus subtilis, RocB is one of the three genes found in the rocABC operon, which is sigma L dependent and induced by arginine. The function of members of this family is as yet unknown, although they are predicted as deacetylases. [PMID: 7674922, PMID: 8113162]"
84	cd17374	Organic anion transporters of the Major Facilitator Superfamily of transporters	"Organic anion transporters (OATs) generally display broad substrate specificity and they facilitate the exchange of extracellular with intracellular organic anions (OAs). Several OATs have been characterized including OAT1-10 and urate anion exchanger 1 (URAT1, also called SLC22A12). Many OATs occur in renal proximal tubules, the site of active drug secretion. OATs mediate the absorption, distribution, and excretion of a diverse array of environmental toxins, and clinically important drugs, including anti-HIV therapeutics, anti-tumor drugs, antibiotics, anti-hypertensives, and anti-inflammatories, and therefore is critical for the survival of the mammalian species. OAT falls into the SLC22 (solute carrier 22) family of the Major Facilitator Superfamily (MFS) of membrane transport proteins, which are thought to function through a single substrate binding site, alternating-access mechanism involving a rocker-switch type of movement. [PMID: 24745985, PMID: 23789957, PMID: 23506881, PMID: 23054972, PMID: 22841915, PMID: 16189627, PMID: 17245646, PMID: 17103332, PMID: 25540139, PMID: 27614971, PMID: 17502342, PMID: 17038320, PMID: 23920220, PMID: 26358290, PMID: 25904762, PMID: 25481222, PMID: 21446918, PMID: 23196129, PMID: 24646860, PMID: 19854166, PMID: 26758938, PMID: 26234418, PMID: 26098515, PMID: 23530251, PMID: 23403214, PMID: 18537473]"
85	cd00548	cytochrome c nitrite reductase and similar proteins	"This family contains cytochrome c nitrite reductase (also known as cytochrome c552, or NrfA) and similar proteins. The pentaheme enzyme NrfA catalyzes the electron reduction of nitrite to ammonia in the nitrogen cycle. This enzyme can also transform nitrogen monoxide and hydroxylamine, two potential bound reaction intermediates, into ammonia. It is a homodimer, with each monomer containing four classical CXXCH type heme-binding sites along with an alternative CXXCK heme-binding motif, which is important for catalysis. This family also includes octaheme nitrite reductase (TvNiR) from the haloalkaliphilic bacterium Thioalkalivibrio paradoxus which catalyzes the reduction of nitrite and hydroxylamine to ammonia as well as the reduction of sulfite to sulfide. [PMID: 25658043, PMID: 10440380, PMID: 11863430, PMID: 12618432, PMID: 18201106, PMID: 18597779, PMID: 22382353, PMID: 12165429, PMID: 12196156]"
86	cd09077	Endonuclease domain encoded by various R1- and I-clade non-long terminal repeat retrotransposons	"This family contains the endonuclease (EN) domain of various non-long terminal repeat (non-LTR) retrotransposons, long interspersed nuclear elements (LINEs) which belong to the subtype 2, R1- and I-clade. LINES can be classified into two subtypes. Subtype 2 has two ORFs: the second (ORF2) encodes a modular protein consisting of an N-terminal apurine/apyrimidine endonuclease domain (EN), a central reverse transcriptase, and a zinc-finger-like domain at the C-terminus. Most non-LTR retrotransposons are inserted throughout the host genome; however, many retrotransposons of the R1 clade exhibit target-specific retrotransposition. This family includes the endonucleases of SART1 and R1bm, from the silkworm Bombyx mori, which belong to the R1-clade. It also includes the endonuclease of snail (Biomphalaria glabrata) Nimbus/Bgl and mosquito Aedes aegypti (MosquI), both which belong to the I-clade. This family belongs to the large EEP (exonuclease/endonuclease/phosphatase) superfamily that contains functionally diverse enzymes that share a common catalytic mechanism of cleaving phosphodiester bonds. [PMID: 12644555, PMID: 12949131, PMID: 15247245, PMID: 17537809, PMID: 10605110, PMID: 9364772, PMID: 9092665, PMID: 2554335, PMID: 2447482, PMID: 1690812, PMID: 17521654]"
87	PF07530	Associated with zinc fingers	This function of this domain is unknown[PMID:11779830]and is often found associated with Pfam:PF00096.

88	IPR006579	Pre-C2HC domain	"This domain (PRE_C2HC) is present in proteins associated with nuclear functionsand is found exclusively in the arthropods, including a number of Drosophila species, the silk moth and the gypsy moth. These proteins are possibly involved in RNA binding or single strand DNA binding."
89	IPR031961	Domain of unknown function DUF4780	"This presumed domain is functionally uncharacterised. This domain is found in insect proteins, and is typically between 132 and 144 amino acids in length. There is a single completely conserved residue W that may be functionally important."
90	PF16012	Domain of unknown function (DUF4780)	"This presumed domain is functionally uncharacterised. This domain family is found in eukaryotes, and is typically between 132 and 144 amino acids in length. There is a single completely conserved residue W that may be functionally important."
91	IPR012677	Nucleotide-binding alpha-beta plait domain superfamily	"This superfamily represents nucleotide-binding domains with an alpha-beta plait structure, which consists of either a ferredoxin-like (beta-alpha-beta)2 fold, such as that found in RNA-binding domains of various ribonucleoproteins or in viral DNA-binding domains; or a beta-(alpha)-beta-alpha-beta(2) fold, such as that found in the ribosomal protein L23."
92	cd00590	RNA recognition motif (RRM) superfamily	"RRM, also known as RBD (RNA binding domain) or RNP (ribonucleoprotein domain), is a highly abundant domain in eukaryotes found in proteins involved in post-transcriptional gene expression processes including mRNA and rRNA processing, RNA export, and RNA stability. This domain is 90 amino acids in length and consists of a four-stranded beta-sheet packed against two alpha-helices. RRM usually interacts with ssRNA, but is also known to interact with ssDNA as well as proteins. RRM binds a variable number of nucleotides, ranging from two to eight. The active site includes three aromatic side-chains located within the conserved RNP1 and RNP2 motifs of the domain. The RRM domain is found in a variety heterogeneous nuclear ribonucleoproteins (hnRNPs), proteins implicated in regulation of alternative splicing, and protein components of small nuclear ribonucleoproteins (snRNPs). [PMID: 15853797, PMID: 22516180, PMID: 22278943, PMID: 18515081, PMID: 20458169, PMID: 7509757, PMID: 17473849, PMID: 18059478, PMID: 17589525, PMID: 16936729, PMID: 16432215, PMID: 16179478, PMID: 17036044, PMID: 16362037, PMID: 16885237, PMID: 10323862, PMID: 15769465, PMID: 8807823, PMID: 7535921, PMID: 10896472]"
93	IPR035979	RNA-binding domain superfamily	"Many eukaryotic proteins containing one or more copies of a putative RNA-binding domain of about 90 amino acids are known to bind single-stranded RNAs. The largest group of single strand RNA-binding proteins is the eukaryotic RNA recognition motif (RRM) family that contains an eight amino acid RNP-1 consensus sequence. RRM proteins have a variety of RNA binding preferences and functions, and include heterogeneous nuclear ribonucleoproteins (hnRNPs), proteins implicated in regulation of alternative splicing (SR, U2AF, Sxl), protein components of small nuclear ribonucleoproteins (U1 and U2 snRNPs), and proteins that regulate RNA stability and translation (PABP, La, Hu). The RRM in heterodimeric splicing factor U2 snRNP auxiliary factor (U2AF) appears to have two RRM-like domains with specialised features for protein recognition. The motif also appears in a few single stranded DNA binding proteins.The typical RRM consists of four anti-parallel beta-strands and two alpha-helices arranged in a beta-alpha-beta-beta-alpha-beta fold with side chains that stack with RNA bases. Specificity of RNA binding is determined by multiple contacts with surrounding amino acids. A third helix is present during RNA binding in some cases. The RRM is reviewed in a number of publications."
94	IPR020847	"AP endonuclease	"DNA damaging agents such as the anti-tumour drugs bleomycin and neocarzinostatin or those that generate oxygen radicals produce a variety of lesions in DNA. Amongst these is base-loss which forms apurinic/apyrimidinic (AP) sites or strand breaks

		1, binding site"	with atypical 3' termini. DNA repair at the AP sites is initiated by specific endonuclease cleavage of the phosphodiester backbone. Such endonucleases are also generally capable of removing blocking groups from the 3' terminus of DNA strand breaks.AP endonucleases can be classified into two families based on sequence similarity. This family contains members of AP endonuclease family 1. Except for Rrp1 and arp, these enzymes are proteins of about 300 amino-acid residues. Rrp1 and arp both contain additional and unrelated sequences in their N-terminal section (about 400 residues for Rrp1 and 270 for arp).This entry represents a motif containing a glutamate residue which has been shown in the Escherichia coli enzyme to bind a divalent metal ion such as magnesium or manganese."
9 5	PS007 26	Putative nuclease HARBI1	NA
9 6	PS501 02	Eukaryotic RNA Recognition Motif (RRM) profile	"Many eukaryotic proteins that are known or supposed to bind single-stranded RNA contain one or more copies of a putative RNA-binding domain of about 90 amino acids [PMID:2470643,PMID:3072706]. This domain is known as the RNA recognition motif (RRM). This region has been found in the following proteins: ** Heterogeneous nuclear ribonucleoproteins ** - hnRNP A1 (helix destabilizing protein) (twice). - hnRNP A2/B1 (twice). - hnRNP C (C1/C2) (once). - hnRNP E (UP2) (at least once). - hnRNP G (once). ** Small nuclear ribonucleoproteins ** - U1 snRNP 70 Kd (once). - U1 snRNP A (once). - U2 snRNP B" (once). ** Pre-RNA and mRNA associated proteins ** - Protein synthesis initiation factor 4B (eIF-4B)[PMID:2390971], a protein essential for the binding of mRNA to ribosomes (once). - Nucleolin (4 times). - Yeast single-stranded nucleic acid-binding protein (gene SSB1) (once). - Yeast protein NSR1 (twice). NSR1 is involved in pre-rRNA processing; it specifically binds nuclear localization sequences. - Poly(A) binding protein (PABP) (4 times). ** Others ** - Drosophila sex determination protein Sex-lethal (Sxl) (twice). - Drosophila sex determination protein Transformer-2 (Tra-2) (once). - Drosophila 'elav' protein (3 times), which is probably involved in the RNA metabolism of neurons. - Human paraneoplastic encephalomyelitis antigen HuD (3 times)[PMID:1655278], which is highly similar to elav and which may play a role in neuron-specific RNA processing. - Drosophila 'bicoid' protein (once)[PMID:2752425], a segment-polarity homeobox protein that may also bind to specific mRNAs. - La antigen (once), a protein which may play a role in the transcription of RNA polymerase III. - The 60 Kd Ro protein (once), a putative RNP complex protein. - A maize protein induced by abscisic acid in response to water stress, which seems to be a RNA-binding protein. - Three tobacco proteins, located in the chloroplast[PMID:16986606], which may be involved in splicing and/or processing of chloroplast RNAs (twice). - X16[PMID:2030943], a mammalian protein which may be involved in RNA processing in relation with cellular proliferation and/or maturation. - Insulin-induced growth response protein CI-4 from rat (twice). - Nucleolysins TIA-1 and TIAR (3 times)[PMID:1326761]which possesses nucleolytic activity against cytotoxic lymphocyte target cells. may be involved in apoptosis. - Yeast RNA15 protein, which plays a role in mRNA stability and/or poly-(A) tail length [9]. Inside the RRM there are two regions which are highly conserved. The first one is a hydrophobic segment of six residues (which is called the RNP-2 motif), the second one is an octapeptide motif (which is called RNP-1 or RNP-CS). The position of both motifs in the domain is shown in the following schematic representation: xxxxxx######xxxxxxxxxxxxxxxxxxxxxxxx######xxxxxxxxxxxxxxxxxxxxxxxxx RNP-2 RNP-1 We have developed a profile that spans the RRM domain."
9 7	PF000 76	"RNA recognition motif. (a.k.a. RRM, RBD,	"The RRM motif is probably diagnostic of an RNA binding protein. RRMs are found in a variety of RNA binding proteins, including various hnRNP proteins, proteins implicated in regulation of alternative splicing, and protein components of snRNPs. The motif also appears in a few single stranded DNA binding proteins. The RRM structure consists of four strands and two helices arranged in an alpha/beta sandwich, with a third helix present during RNA binding in some cases The C-terminal beta strand (4th strand) and final helix are hard to align and have been omitted in the SEED alignment The LA proteins

		or RNP domain)"	(Swiss:P05455) have an N terminal rrm which is included in the seed. There is a second region towards the C terminus that has some features characteristic of a rrm but does not appear to have the important structural core of a rrm. The LA proteins (Swiss:P05455) are one of the main autoantigens in Systemic lupus erythematosus (SLE), an autoimmune disease."
9 8	IPR000504	RNA recognition motif domain	"Many eukaryotic proteins containing one or more copies of a putative RNA-binding domain of about 90 amino acids are known to bind single-stranded RNAs. The largest group of single strand RNA-binding proteins is the eukaryotic RNA recognition motif (RRM) family that contains an eight amino acid RNP-1 consensus sequence. RRM proteins have a variety of RNA binding preferences and functions, and include heterogeneous nuclear ribonucleoproteins (hnRNPs), proteins implicated in regulation of alternative splicing (SR, U2AF, Sxl), protein components of small nuclear ribonucleoproteins (U1 and U2 snRNPs), and proteins that regulate RNA stability and translation (PABP, La, Hu). The RRM in heterodimeric splicing factor U2 snRNP auxiliary factor (U2AF) appears to have two RRM-like domains with specialised features for protein recognition. The motif also appears in a few single stranded DNA binding proteins.The typical RRM consists of four anti-parallel beta-strands and two alpha-helices arranged in a beta-alpha-beta-beta-alpha-beta fold with side chains that stack with RNA bases. Specificity of RNA binding is determined by multiple contacts with surrounding amino acids. A third helix is present during RNA binding in some cases. The RRM is reviewed in a number of publications."
9 9	IPR004263	Exostosin-like	"There are five identified human EXT family proteins (EXT1, EXT2, EXTL1, EXTL2 and EXTL3), which are members of the hereditary multiple exostoses family of tumor suppressors. They are glycosyltransferases required for the biosynthesis of heparan sulfate. Hereditary multiple exostoses (EXT) is an autosomal dominant disorder that is characterised by the appearance of multiple outgrowths of the long bones (exostoses) at their epiphyses. Mutations in two homologous genes, EXT1 and EXT2, are responsible for the EXT syndrome. The human and mouse EXT genes have at least two homologues in the invertebrate Caenorhabditis elegans, indicating that they do not function exclusively as regulators of bone growth. EXT1 and EXT2 have both been shown to encode glycosyltransferases involved in the chain elongation step of heparan sulphate biosynthesis.This entry also includes Arabidopsis Xyloglucan galactosyltransferase KATAMARI1and Drosophila melanogaster EXT homologues."
1 0 0	IPR004808	AP endonuclease 1	AP endonucleases can be classified into two families based on sequence similarity. This family contains members of AP endonuclease family 1. They are endonucleases that remove the damaged DNA at cytosines and guanines by cleaving on the 3'-side of the AP site by a beta-elimination reaction. The proteins contain glutamate which has been shownin the Escherichia coli enzyme to bind a divalent metal ion such as magnesium or manganese.DNA damaging agents such as the anti-tumour drugs bleomycin and neocarzinostatin or those that generate oxygen radicals produce a variety of lesions in DNA. Amongst these is base-loss which forms apurinic/apyrimidinic (AP) sites or strand breaks with atypical 3' termini. DNA repair at the AP sites is initiated by specific endonuclease cleavage of the phosphodiester backbone. Such endonucleases are also generally capable of removing blocking groups from the 3' terminus of DNA strand breaks.
1 0 1	IPR000595	Cyclic nucleotide-binding domain	"Proteins that bind cyclic nucleotides (cAMP or cGMP) share a structural domain of about 120 residues[cite:PUB00013967],[cite:PUB00013968],[cite:PUB00005543]. The best studied of these proteins is the prokaryotic catabolite gene activator (also known as the cAMP receptor protein) (gene crp) where such a domain is known to be composed of three alpha-helices and a distinctive eight-stranded, antiparallel beta-barrel structure. There are six invariant amino acids in this domain, three of which are glycine residues that are thought to be essential for maintenance of the structural integrity of the beta-barrel. cAMP- and cGMP-dependent protein kinases (cAPK and cGPK) contain two tandem copies of the cyclic nucleotide-binding domain. The cAPK's are composed of two different subunits, a catalytic chain and a regulatory chain, which contains both copies of the domain. The cGPK's are single chain enzymes that include the two copies of the domain in their N-

				terminal section. Vertebrate cyclic nucleotide-gated ion-channels also contain this domain. Two such cations channels have been fully characterised, one is found in rod cells where it plays a role in visual signal transduction."
1 0 2	PS500 42	cAMP/cGMP binding motif profile		"Proteins that bind cyclic nucleotides (cAMP or cGMP) share a structural domain of about 120 residues [PMID:2550070,PMID:1710853,PMID:13134l6]. The best studied of these proteins is the prokaryotic catabolite gene activator (also known as the cAMP receptor protein) (gene crp) where such a domain is known to be composed of three alpha-helices and a distinctive eight-stranded, antiparallel beta-barrel structure. Such a domain is known to exist in the following proteins: - Prokaryotic catabolite gene activator protein (CAP). - cAMP- and cGMP-dependent protein kinases (cAPK and cGPK). Both types of kinases contains two tandem copies of the cyclic nucleotide-binding domain. The cAPK's are composed of two different subunits: a catalytic chain and a regulatory chain which contains both copies of the domain. The cGPK's are single chain enzymes that include the two copies of the domain in their N- terminal section. The nucleotide specificity of cAPK and cGPK is due to an amino acid in the conserved region of beta-barrel 7: a threonine that is invariant in cGPK is an alanine in most cAPK. - Vertebrate cyclic nucleotide-gated ion-channels. Two such cations channels have been fully characterized. One is found in rod cells where it plays a role in visual signal transduction. It specifically binds to cGMP leading to an opening of the channel and thereby causing a depolarization of rod photoreceptors. In olfactory epithelium a similar, cAMP-binding, channel plays a role in odorant signal transduction. There are six invariant amino acids in this domain, three of which are glycine residues that are thought to be essential for maintenance of the structural integrity of the beta-barrel. We developed two signature patterns for this domain. The first pattern is located within beta-barrels 2 and 3 and contains the first two conserved Gly. The second pattern is located within beta- barrels 6 and 7 and contains the third conserved Gly as well as the three other invariant residues."
1 0 3	cd003 04	"RT_like: Reverse transcriptase (RT, RNA-dependent DNA polymerase)_1ike family..."		"RT_like: Reverse transcriptase (RT, RNA-dependent DNA polymerase)_like family. An RT gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. RTs occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. These elements can be divided into two major groups. One group contains retroviruses and DNA viruses whose propagation involves an RNA intermediate. They are grouped together with transposable elements containing long terminal repeats (LTRs). The other group, also called poly(A)-type retrotransposons, contain fungal mitochondrial introns and transposable elements that lack LTRs. [PMID: 1698615, PMID: 8828137, PMID: 10669612, PMID: 9878607, PMID: 7540934, PMID: 7523679, PMID: 8648598]"
1 0 4	IPR00 1005	SANT/Myb domain		"The retroviral oncogene v-myb, and its cellular counterpart c-myb, encode nuclear DNA-binding proteins. In myb, one of the most conserved regions consisting of three tandem repeats has been shown to be involved in DNA-binding.The SANT domain is present in nuclear receptor co-repressors and in the subunits of many chromatin-remodelling complexes. It has a strong structural similarity to the DNA-binding domain of Myb-related proteins. Both consist of tandem repeats of three alpha-helices that are arranged in a helix-turn-helix motif, each alpha helix containing a bulky aromatic residue. Despite the overall similarity there is differences that indicate that the SANT domain is functionally divergent from the canonical Myb DNA-binding domain."
1 0 5	PF000 96	"Zinc finger, C2H2 type"		"The C2H2 zinc finger is the classical zinc finger domain. The two conserved cysteines and histidines co-ordinate a zinc ion. The following pattern describes the zinc finger. #-X-C-X(1-5)-C-X3-#-X5-#-X2-H-X(3-6)-[H/C] Where X can be any amino acid, and numbers in brackets indicate the number of residues. The positions marked # are those that are important for the stable fold of the zinc finger. The final position can be either his or cys. The C2H2 zinc finger is composed of two short beta strands followed by an alpha helix. The amino terminal part of the helix binds the major groove in DNA binding zinc fingers. The accepted consensus binding sequence for Sp1 is usually defined by the asymmetric hexanucleotide core GGGCGG but this

				sequence does not include, among others, the GAG (=CTC) repeat that constitutes a high-affinity site for Sp1 binding to the wt1 promoter[PMID:12683994]."
1 0 6	cd001 67	"SWI3, ADA2, N-CoR and TFIIIB' DNA-binding domains. Tandem copies o..."	"SWI3, ADA2, N-CoR and TFIIIB' DNA-binding domains. Tandem copies of the domain bind telomeric DNA tandem repeatsas part of the capping complex. Binding is sequence dependent for repeats which contain the G/C rich motif [C2-3 A (CA)1-6]. The domain is also found in regulatory transcriptional repressor complexes where it also binds DNA. [PMID: 8882580, PMID: 9048479, PMID: 10652526, PMID: 11572773]"
1 0 7	PF002 49	Myb-like DNA-binding domain	"This family contains the DNA binding domains from Myb proteins, as well as the SANT domain family[PMID:8882580]."
1 0 8	IPR00 9057	Homeobox-like domain superfamily	"Homeobox domain (also known as homeodomain) proteins are transcription factors that share a related DNA binding homeodomain. The homeodomain was first identified in a number of Drosophila homeotic and segmentation proteins, but is now known to be well conserved in many other animals, including vertebrates. The domain binds DNA through a helix-turn-helix (HTH) structure. The HTH motif is characterised by two alpha-helices, which make intimate contacts with the DNA and are joined by a short turn. The second helix binds to DNA via a number of hydrogen bonds and hydrophobic interactions, which occur between specific side chains and the exposed bases and thymine methyl groups within the major groove of the DNA. The first helix helps to stabilise the structure. Many proteins contain homeodomains, including Drosophila Engrailed, yeast mating type proteins, hepatocyte nuclear factor 1a and HOX proteins.The homeodomain motif is very similar in sequence and structure to domains in a wide range of DNA-binding proteins, including recombinases, Myb proteins, GARP response regulators, human telomeric proteins (hTRF1), paired domain proteins (PAX), yeast RAP1, centromere-binding proteins CENP-B and ABP-1, transcriptional regulators (TyrR), AraC-type transcriptional activators, and tetracycline repressor-like proteins (TetR, QacR, YcdC)."
1 0 9	IPR00 1087	GDSL lipase/esterase	"GDSL esterases and lipases are hydrolytic enzymes with multifunctional properties. This new subclass of lipolytic enzymes possesses a distinct GDSL sequence motif different from the GxSxG motif found in many lipases. Members include; Aeromonas hydrophila lipase, Vibrio mimicus arylesterase, Vibrio parahaemolyticus thermolabile haemolysin, rabbit phospholipase (AdRab-B), and Brassica napus anter-specific proline-rich protein."
1 1 0	PF006 57	GDSL-like Lipase/Acylhydrolase	NA
1 1 1	IPR03 3447	Nectin-3	NA
1 1 2	PF171 82	OSK domain	This entry represents the OSK domain defined by Jeske and colleagues[PMID:26190108]. The domain is related to SGNH hydrolases but lacks the active site residues. The domain binds to RNA[PMID:26190108].

1 1 3	cd09881	"VapC-like PIN domain of Mycobacterium tuberculosis VapC4 and VapC5, and Neisseria gonorrhoeae FitB and related proteins"	"This family includes the Virulence associated protein C (VapC)-like PIN (PilT N terminus) domain of Mycobacterium tuberculosis VapC4 and VapC5 ribonuclease toxins of the VapBC toxin/antitoxin (TA) system, and Neisseria gonorrhoeae FitB toxin of the FitAB TA system. This family belongs to the VapC-like family of the PIN domain nuclease superfamily. VapC is the PIN-domain ribonuclease toxin from prokaryotic VapBC toxin-antitoxin (TA) systems. VapB is a transcription factor-like protein antitoxin acting as an inhibitor. Other members of the VapC-like nuclease family include eukaryotic ribonucleases such as Smg6, ribosome assembly factor NOB1, exosome subunit Rrp44 endoribonuclease and rRNA-processing protein Fcf1. The structural properties of the PIN (PilT N terminus) domain indicate its active center, consisting of three highly conserved catalytic residues which coordinate metal ions, in some members, additional metal coordinating residues can be found. Some members of the superfamily lack several of these key catalytic residues. The PIN active site is geometrically similar in the active center of structure-specific 5' nucleases, PIN-domain ribonucleases of eukaryotic rRNA editing proteins, and bacterial toxins of toxin-antitoxin (TA) operons. [PMID: 21036780, PMID: 22140099, PMID: 26833558, PMID: 27998932, PMID: 19445953, PMID: 18952600, PMID: 17496075, PMID: 16982615, PMID: 15993073, PMID: 14734548, PMID: 11137022, PMID: 28575517, PMID: 28508407]"
1 1 4	IPR019793	Peroxidases heam-ligand binding site	"Peroxidases are haem-containing enzymes that use hydrogen peroxide as the electron acceptor to catalyse a number of oxidative reactions. Most haem peroxidases follow the reaction scheme: Fe3+ + H2O2 --> [Fe4+=O]R' (Compound I) + H2O [Fe4+=O]R' + substrate --> [Fe4+=O]R (Compound II) + oxidised substrate [Fe4+=O]R + substrate --> Fe3+ + H2O + oxidised substrateIn this mechanism, the enzyme reacts with one equivalent of H2O2 to give [Fe4+=O]R' (compound I). This is a two-electron oxidation/reduction reaction where H2O2 is reduced to water and the enzyme is oxidised. One oxidising equivalent resides on iron, giving the oxyferrylintermediate, while in many peroxidases the porphyrin (R) is oxidised to the porphyrin pi-cation radical (R'). Compound I then oxidises an organic substrate to give a substrate radical.Haem peroxidases include two superfamilies: one found in bacteria, fungi, plants and the second found in animals. The first one can be viewed as consisting of 3 major classes. Class I, the intracellular peroxidases, includes: yeast cytochrome c peroxidase (CCP), a soluble protein found in the mitochondrial electron transport chain, where it probably protects against toxic peroxides; ascorbate peroxidase (AP), the main enzyme responsible for hydrogen peroxide removal in chloroplasts and cytosol of higher plants; and bacterial catalase-peroxidases, exhibiting both peroxidase and catalase activities. It is thought that catalase-peroxidase provides protection to cells under oxidative stress.Class II consists of secretory fungal peroxidases: ligninases, or lignin peroxidases (LiPs), and manganese-dependent peroxidases (MnPs). These are monomeric glycoproteins involved in the degradation of lignin. In MnP, Mn2+ serves as the reducing substrate. Class II proteins contain four conserved disulphide bridges and two conserved calcium-binding sites.Class III consists of the secretory plant peroxidases, which have multiple tissue-specific functions: e.g., removal of hydrogen peroxide from chloroplasts and cytosol; oxidation of toxic compounds; biosynthesis of the cell wall; defence responses towards wounding; indole-3-acetic acid (IAA) catabolism; ethylene biosynthesis; and so on. Class III proteins are also monomeric glycoproteins, containing four conserved disulphide bridges and two calcium ions, although the placement of the disulphides differs from class II enzymes.The crystal structures of a number of these proteins show that they share the same architecture - two all-alpha domains between which the haem group is embedded.This entry represents the binding site for heam in a number of peroxidases."
1 1 5	PS00435	Putative nuclease HARBI1	NA

1 1 6	cd16965	Alpha-kinase domain of channel kinases	"This group is composed of channel kinases 1 (ChaK1) and 2 (ChaK2), and similar proteins. ChaK1 and ChaK2 are also called transient receptor potential cation channel subfamily M members 7 (TRMP7) and 6 (TRMP6), respectively. They are fusion proteins containing a transmembrane ion pore or channel and a C-terminal alpha-kinase domain, both of which are functional. They are both cation-selective channels that preferentially permeate Zn2+, Mg2+, and Ca2+ ions. They are central regulators of Mg2+ and Ca2+ homeostasis. TRMP7 is ubiquitously expressed while TRMP6 is highly expressed in specific tissues such as the kidney and intestine. Alpha-kinase is an atypical protein kinase catalytic domain with no detectable similarity to conventional protein serine/threonine kinases. The alpha-kinase family was named after the unique mode of substrate recognition by its initial members, the Dictyostelium heavy chain kinases, which targeted protein sequences that adopt an alpha-helical conformation. More recently, alpha-kinases were found to also target residues in non-helical regions. [PMID: 11161216, PMID: 10021370, PMID: 11389851, PMID: 26218331, PMID: 26179995, PMID: 25437439, PMID: 25204892, PMID: 24756720, PMID: 24752033, PMID: 24634592, PMID: 24906182, PMID: 24756719, PMID: 20932259, PMID: 19780399, PMID: 18660673, PMID: 20012461]"
1 1 7	cd00603	IPT domain of Plexins and Cell Surface Receptors (PCSR) and related proteins ...	"IPT domain of Plexins and Cell Surface Receptors (PCSR) and related proteins . This subgroup contains IPT domains of plexins, receptors, like the plasminogen-related growth factor receptors, the hepatocyte growth factor-scatter factors, and the macrophage-stimulating receptors and of fibrocystin. Plexins are involved in the regulation of cell proliferation and of cellular adhesion and repulsion receptors. In general, there are three copies of the IPT_PCSR domain present preceeded by SEMA (semaphorin) and PSI (plexin, semaphorin, integrin) domains. [PMID: 11919560, PMID: 12367632, PMID: 10027283, PMID: 10520995]"
1 1 8	cd17772	Two tandem repeats of the cystathionine beta-synthase (CBS pair) domains associated with the DHH and nucleotidyltransferase (NT) domains	"This cd contains two tandem repeats of the cystathionine beta-synthase (CBS pair) domains associated with an upstream DHH domain which performs a phosphoesterase function and a downstream nucleotidyltransferase (NT) domain of family X DNA polymerases. The CBS domain, named after human CBS, is a small domain originally identified in cystathionine beta-synthase and is subsequently found in a wide range of different proteins. CBS domains usually occur in tandem repeats. They associate to form a so-called Bateman domain or a CBS pair based on crystallographic studies in bacteria. The CBS pair was used as a basis for this cd hierarchy since the human CBS proteins can adopt the typical core structure and form an intramolecular CBS pair. The interface between the two CBS domains forms a cleft that is a potential ligand binding site. The CBS pair coexists with a variety of other functional domains and this has been used to help in its classification here. It has been proposed that the CBS domain may play a regulatory role, although its exact function is unknown. Mutations of conserved residues within this domain are associated with a variety of human hereditary diseases, including congenital myotonia, idiopathic generalized epilepsy, hypercalciuric nephrolithiasis, and classic Bartter syndrome (CLC chloride channel family members), Wolff-Parkinson-White syndrome (gamma 2 subunit of AMP-activated protein kinase), retinitis pigmentosa (IMP dehydrogenase-1), and homocystinuria (cystathionine beta-synthase). [PMID: 16275737, PMID: 9020585, PMID: 9106071, PMID: 9590298, PMID: 10200156, PMID: 11230183, PMID: 11524006, PMID: 14722619]"
1 1 9	IPR006939	SNF5/SMARCB1/INI1	"SNF5 is a component of the yeast SWI/SNF complex, which is an ATP-dependent nucleosome-remodelling complex that regulates the transcription of a subset of yeast genes. SNF5 is a key component of all SWI/SNF-class complexes characterised so far. This family consists of the conserved region of SNF5, including a direct repeat motif. SNF5 is essential for the assembly promoter targeting and chromatin remodelling activity of the SWI-SNF complex. SNF5 is also known as SMARCB1, for SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily b, member 1, and also INI1 for integrase interactor 1. Loss-of function mutations in SNF5 are thought to contribute to oncogenesis in malignant rhabdoid tumours (MRTs)."

1 2 0	IPR03 8765	Papain-like cysteine peptidase superfamily	This entry represents the papain-like cysteine peptidase superfamily. Its members typically consist of one alpha-helix and 4 strands of antiparallel beta-sheet and contain the catalytic triad Cys-His-Asn.
1 2 1	PF022 57	RFX DNA-binding domain	"RFX is a regulatory factor which binds to the X box of MHC class II genes and is essential for their expression. The DNA-binding domain of RFX is the central domain of the protein and binds ssDNA as either a monomer or homodimer[PMID:22538877]. It recognize X-boxes (DNA of the sequence 5'-GTNRCC(0-3N)RGYAAC-3', where N is any nucleotide, R is a purine and Y is a pyrimidine) using a highly conserved 76-residue DNA-binding domain (DBD)[PMID:10706293]."
1 2 2	IPR03 9779	RFX-like DNA-binding protein	"This entry includes Rfx1 from budding yeasts, Sak1 from fission yeasts and RFX1-8 from humans. They are transcription factors that contain a RFX DNA binding domain and a conserved C-terminal region."
1 2 3	IPR03 6390	Winged helix DNA-binding domain superfamily	"Winged helix DNA-binding proteins share a related winged helix-turn-helix DNA-binding motif, where the ""wings"", or loops, are small beta-sheets. The winged helix motif consists of two wings (W1, W2), three alpha helices (H1, H2, H3) and three beta-sheets (S1, S2, S3) arranged in the order H1-S1-H2-H3-S2-W1-S3-W2. The DNA-recognition helix makes sequence-specific DNA contacts with the major groove of DNA, while the wings make different DNA contacts, often with the minor groove or the backbone of DNA. Several winged-helix proteins display an exposed patch of hydrophobic residues thought to mediate protein-protein interactions."
1 2 4	IPR00 3150	DNA-binding RFX-type winged-helix domain	"The RFX family of transcription factors is characterised by a unique approximately 75-residue DNA-binding domain. RFX genes have been isolated in yeasts, nematode and vertebrates. The characteristic RFX-type HTH DNA binding domain has been recruited into otherwise very divergent regulatory factors functioning in a diverse spectrum of unrelated systems, including regulation of the mitotic cell cycle in fission yeast, the control of the immune response in mammals,and infection by human hepatitis B virus.The RFX-type DNA-binding domain is an unusual member of the winged-helix subfamily of helix-turn-helix domains because it uses a beta-hairpin (or wing) to recognise DNA instead of the recognition helix typical of helix-tun-helix domains. It consists of three alpha-helices (H), three beta-strands (S) and three connecting loops (L), arranged in the order H1-S1-H2-L1-H3-L2-S2-W1-S3. The third loop, connecting beta-strands S2 and S3, forms wing W1 of the winged-helix motif. In contrast to previously described winged-helix DBDs (typically 110 residues long with two wings, W1 and W2), the shorter RFX-type winged-helix DNA-binding domain has only one wing."
1 2 5	IPR03 6388	Winged helix-like DNA-binding domain superfamily	"Winged helix DNA-binding proteins share a related winged helix-turn-helix DNA-binding motif, where the ""wings"", or loops, are small beta-sheets. The winged helix motif consists of two wings (W1, W2), three alpha helices (H1, H2, H3) and three beta-sheets (S1, S2, S3) arranged in the order H1-S1-H2-H3-S2-W1-S3-W2. The DNA-recognition helix makes sequence-specific DNA contacts with the major groove of DNA, while the wings make different DNA contacts, often with the minor groove or the backbone of DNA. Several winged-helix proteins display an exposed patch of hydrophobic residues thought to mediate protein-protein interactions.Many different proteins with diverse biological functions contain a winged helix DNA-binding domain, including transcriptional repressors such as biotin repressor, LexA repressor and the arginine repressor; transcription factors such as the hepatocyte nuclear factor-3 proteins involved in cell differentiation, heat-shock transcription factor, and the general transcription factors TFIIE and TFIIF; helicases such as RuvB that promotes branch migration at the Holliday junction, and CDC6 in the pre-replication complex; endonucleases such as FokI and TnsA; histones; and Mu transposase, where the flexible wing of the enhancer-binding domain is essential for efficient transposition."

1 2 6	IPR037800	Histone acetyltransferase GCN5	"The transcription regulatory histone acetylation complex Spt-Ada-Gcn5 acetyltransferase (SAGA) is involved in RNA polymerase II-dependent transcriptional regulation of approximately 10% of yeast genes. SAGA preferentially acetylates histones H3 and H2B and deubiquitinates histone H2B. SAGA is known as PCAF in vertebrates and PCAF acetylates nucleosomal histone H3. The SAGA complex consists of at least TRA1, CHD1, SPT7, TAF5, ADA3, SGF73, SPT20/ADA5, SPT8, TAF12, TAF6, HFI1/ADA1, UBP8, GCN5, ADA2, SPT3, SGF29, TAF10, TAF9, SGF11 and SUS1, and some of these components are present as two copies. The complex is built up from distinct modules, each of which has a separate function and crosslinks with either other proteins or other modules in the complex.SLIK (SAGA-like) is a multi-subunit histone acetyltransferase complex that preferentially acetylates histones H3 and H2B and deubiquitinates histone H2B. It is an embellishment of the SAGA complex. The yeast SLIK complex consists of at least TRA1, CHD1, SPT7, CC TAF5, ADA3, SPT20, RTG2, TAF12, TAF6, HFI1, UBP8 (a deubiquitinase), GCN5, ADA2, SPT3, SGF29, TAF10 and TAF9.The yeast SALSA complex is an altered form of the SAGA complex and consists of at least TRA1, SPT7 (C-terminal truncated form), TAF5, ADA3, SPT20, TAF12, TAF6, HFI1, GCN5, ADA2 and SPT3.The ADA complex is a transcription regulatory histone acetylation (HAT) complex. ADA preferentially acetylates nucleosomal histones H3 (at 'Lys-14' and 'Lys-18') and H2B. The complex consists of at least ADA2, ADA3, AHC1, and GCN5. AHC1 is required for the overall structural integrity of the ADA complex.This entry includes histone acetyltransferases GCN5, KAT2A and KAT2B (all of which are included in EC:2.3.1.48). GCN5 acetylates histones H2B, H3 and H4, providing a specific tag for epigenetic transcription activation. GCN5 is a component of the transcription regulatory histone acetylation (HAT) complexes SAGA, SLIK, SALSAand ADA. Mammals have two paralogues: KAT2A (also known as GCN5) and KAT2B. KAT2A acetylates core histones to provide a specific tag for epigenetic transcription activation, but not nucleosome core particles. It also acetylates proteins such as CEBPB. KAT2A is a component of the ATAC complex, which has acetyltransferase activity on histones H3 and H4. KAT2B (also known as P300/calcium-binding protein (CBP)-associated factor or PCAF) can acetylate the core histones H3 and H4 as well as nucleosome core particles and non-histone proteins such as ACLY."
1 2 7	PF00628	PHD-finger	PHD folds into an interleaved type of Zn-finger chelating 2 Zn ions in a similar manner to that of the RING and FYVE domains[PMID:11124022]. Several PHD fingers have been identified as binding modules of methylated histone H3[PMID:17142463].
1 2 8	IPR000104	"Antifreeze protein, type I"	"Marine teleosts from polar oceans can be protected from freezing in icy sea-water by serum antifreeze proteins (AFPs) or glycoproteins (AFGPs): these function by binding to, and preventing the growth of, ice crystals within the fish and depressing the non-equilibrium freezing point to below that of the melting point.Despite functional similarity, antifreeze proteins are structurally diverse and include glycosylated and at least 3 non-glycosylated forms: the AFGP of notothenioids and cods are polymers of a tripeptide repeat, Ala-Ala-Thr, with a disaccharide attached to the threonine residue; type I AFPs are found in flounder and sculpin; type II AFPs of sea-raven, smelt and herring are Cys-rich proteins; and type III AFPs, found in eel pouts, are rich in beta-structure. Non-homologous antifreeze proteins have also been identified in insects and plants.Type I AFPs are Ala-rich, amphiphilic, alpha-helical proteins. The ice-binding sites of all AFPs are relatively flat and hydrophobic and have an uninterupted section of alanines running the length of the approximately 16.5A helix repeat. Based on the energy-minimised structure, a model has been proposed to describe the binding of the protein to ice crystals, whereby the protein binds to an ice nucleation structure, in a zipper-like fashion, via hydrogen bonding of the methyl-group of threonine side chains (with an 11-residue period) to oxygen atoms in the ice lattice. The growth of ice crystals is thus stopped, or retarded, and the freezing point depressed. The high lysince content of these peptides may serve to promote the solubility of these proteins."

1 2 9	cd000 09	The AAA+ (ATPases Associated with a wide variety of cellular Activities) supe...	"The AAA+ (ATPases Associated with a wide variety of cellular Activities) superfamily represents an ancient group of ATPases belonging to the ASCE (for additional strand, catalytic E) division of the P-loop NTPase fold. The ASCE division also includes ABC, RecA-like, VirD4-like, PilT-like, and SF1/2 helicases. Members of the AAA+ ATPases function as molecular chaperons, ATPase subunits of proteases, helicases, or nucleic-acid stimulated ATPases. The AAA+ proteins contain several distinct features in addition to the conserved alpha-beta-alpha core domain structure and the Walker A and B motifs of the P-loop NTPases. [PMID: 18208389, PMID: 15037234, PMID: 18466635, PMID: 9927482, PMID: 11473577, PMID: 11916378, PMID: 15381417]"
1 3 0	IPR02 7417	P-loop containing nucleoside triphosphate hydrolase	"The P-loop NTPase fold is the most prevalent domain of the several distinct nucleotide-binding protein folds.The most common reaction catalysed by enzymes of the P-loop NTPase fold is the hydrolysis of the beta-gamma phosphate bond of a bound nucleoside triphosphate (NTP). The energy from NTP hydrolysis is typically utilised to induce conformational changes in other molecules, which constitutes the basis of the biological functions of most P-loop NTPases. P-loop NTPases show substantial substrate preference for either ATP or GTP.P-loop NTPases are characterised by two conserved sequence signatures, the Walker A motif (the P-loop proper) and Walker B motifs which bind, respectively, the beta and gamma phosphate moieties of the bound NTP, and a Mg2+ cation.P-loop ATPase domains belong to one of the two major divisions. The kinase-GTPase (KG) division includes the kinases and GTPases, and the ASCE division, characterised by an additional strand in the core sheet, which is located between the P-loop strand and the Walker B strand. Most members of the ASCE division utilise ATP and members of this group include AAA+, ABC, PilT, HerA-FtsK, superfamily 1/2 (SF1/2) helicases, and the RecA/ATP-synthase superfamilies of ATPases, etc."
1 3 1	PF152 88	Zinc knuckle	This Zinc knuckle is found in FAM90A mammalian proteins.
1 3 2	IPR04 1670	Zinc knuckle	This Zinc knuckle is found in FAM90A and transcription initiation factor TFIID proteins.
1 3 3	PF014 43	Viral (Superfamily 1) RNA helicase	"Helicase activity for this family has been demonstrated[PMID:8269709]and NTPase activity[PMID:10217401]. This helicase has multiple roles at different stages of viral RNA replication, as dissected by mutational analysis[PMID:8057461]."
1 3 4	IPR02 7351	(+) RNA virus helicase core domain	"Helicases have been classified in 6 superfamilies (SF1-SF6). All of the proteins bind ATP and, consequently, all of them carry the classical Walker A (phosphate-binding loop or P-loop) and Walker B (Mg2+-binding aspartic acid) motifs. The two largest superfamilies, commonly referred to as SF1 and SF2, share similar patterns of seven conserved sequence motifs, some of which are separated by long poorly conserved spacers. Helicase motifs appear to be organised in a core domain which provides the catalytic function, whereas optional inserts and amino- and carboxy-terminal sequences may comprise distinct domains with diverse accessory roles. The helicase core contains two structural domains, an N- terminal ATP-binding domain and a C-terminal domain. Putative SF1 helicases are extremely widespread among positive-stranded (+)RNA viruses. They have been identified in a variety of plant virus families, as well as alpha- rubi-, arteri-, hepatitis E, and coronaviruses. A number of these viral enzymes have been implicated in diverse aspects of transcription and replication but also in RNA stability and cell-to-cell movement.The (+) RNA virus helicase core contains two RecA-like alpha/beta domains. The N-terminal ATP-binding domain contains a parallel six- stranded beta-sheet surrounded by four helices on one side and two helices on the other. The C-terminal

				domain contains a parallel four-stranded beta-sheet sandwiched between two helices on each of its sides. The (+)RNA virus helicase core is likely to bind NTP in cleft between the N terminus of the ATP-binding domain and the beginning of the C-terminal domain.This entry represents the (+)RNA virus helicase core domain."
1 3 5	PS516 57	(+)RNA virus helicase core domain profile	"Helicases have been classified in 6 superfamilies (SF1-SF6)[PMID:22573863]. All of the proteins bind ATP and, consequently, all of them carry the classical Walker A (phosphate-binding loop or P-loop) and Walker B (Mg2+-binding aspartic acid) motifs [1]. The two largest superfamilies, commonly referred to as SF1 and SF2, share similar patterns of seven conserved sequence motifs, some of which are separated by long poorly conserved spacers [1]. Helicase motifs appear to be organized in a core domain which provides the catalytic function, whereas optional inserts and amino- and carboxy-terminal sequences may comprise distinct domains with diverse accessory roles. The helicase core contains two structural domains, an N- terminal ATP-binding domain and a C-terminal domain. Putative SF1 helicases are extremely widespread among positive-stranded (+)RNA viruses. They have been identified in a variety of plant virus families, as well as alpha- rubi-, arteri-, hepatitis E, and coronaviruses. A number of these viral enzymes have been implicated in diverse aspects of transcription and replication but also in RNA stability and cell-to-cell movement[PMID:11000230]. The (+)RNA virus helicase core contains two RecA-like alpha/beta domains. The N-terminal ATP-binding domain contains a parallel six- stranded beta-sheet surrounded by four helices on one side and two helices on the other. The C-terminal domain contains a parallel four-stranded beta-sheet sandwiched between two helices on each of its sides. The (+)RNA virus helicase core is likely to bind NTP in cleft between the N-terminus of the ATP-binding domain and the beginning of the C-terminal domain[PMID:22573863]. The profile we developed covers the entire (+)RNA virus helicase core."
1 3 6	cd188 09	C-terminal helicase domain of RecD family helicases	"RecD is a member of the RecBCD (EC 3.1.11.5, Exonuclease V) complex. It is the alpha chain of the complex and functions as a 3'-5' helicase. The RecBCD enzyme is both a helicase that unwinds, or separates the strands of DNA, and a nuclease that makes single-stranded nicks in DNA. RecD family helicases are DEAD-like helicases belonging to superfamily (SF)1, a diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. Similar to SF2 helicases, SF1 helicases do not form toroidal structures like SF3-6 helicases. Their helicase core consists of two similar protein domains that resemble the fold of the recombination protein RecA. This model describes the C-terminal domain, also called HelicC. [PMID: 3526335, PMID: 7603978, PMID: 1552844, PMID: 2563148, PMID: 10322435, PMID: 11545728, PMID: 16935882, PMID: 19747077, PMID: 21779027, PMID: 23542735, PMID: 23416748, PMID: 20225155, PMID: 20941364]"
1 3 7	cd179 33	DEXS-box helicase domain of RecD and similar proteins	"RecD is a member of the RecBCD (EC 3.1.11.5, Exonuclease V) complex. It is the alpha chain of the complex and functions as a 3'-5' helicase. The RecBCD enzyme is both a helicase that unwinds, or separates the strands of DNA, and a nuclease that makes single-stranded nicks in DNA. RecD is a member of the DEAD-like helicase superfamily, a diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region. [PMID: 3526335, PMID: 7603978, PMID: 1552844, PMID: 2563148, PMID: 10322435, PMID: 11545728, PMID: 16935882, PMID: 19747077, PMID: 21779027, PMID: 23542735, PMID: 23416748, PMID: 20225155, PMID: 20941364]"
1 3 8	PF087 19	Domain of unknown function (DUF1768)	This is a domain of unknown function. It is alpha helical in structure. The GO annotation for this protein suggests it is involved in nematode larval development and has a positive regulation on growth rate.
1 3 9	IPR01 2816	NADAR	"This family contains E. coli swarming motility protein YbiA. Mutations in YbiA cause defects in Escherichia coli swarming, but not necessarily in motility. This family was predicted to be involved in NAD-utilizing pathways, likely to act on ADP-ribose derivatives, and was been named NADAR (NAD and ADP-ribose). More recently, YbiA has been shown to be involved in the

			disposal of riboflavin intermediates. It catalyzes the hydrolysis of the N-glycosidic bond in the first two intermediates of riboflavin biosynthesis, which are highly reactive metabolites, yielding relatively innocuous products."
140	IPR037238	YbiA-like superfamily	"This superfamily contains E. coli swarming motility protein YbiA. Mutations in YbiA cause defects in Escherichia coli swarming, but not necessarily in motility. This family was predicted to be involved in NAD-utilizing pathways, likely to act on ADP-ribose derivatives, and was been named NADAR (NAD and ADP-ribose). More recently, YbiA has been shown to be involved in the disposal of riboflavin intermediates. It catalyzes the hydrolysis of the N-glycosidic bond in the first two intermediates of riboflavin biosynthesis, which are highly reactive metabolites, yielding relatively innocuous products."
141	cd15457	Escherichia coli swarming motility protein YbiA and related proteins	"This family of uncharacterized domains was initially classified as Domain of Unknown Function (DUF1768). It contains members such as the E. coli swarming motility protein YbiA. Mutations in YbiA cause defects in Escherichia coli swarming, but not necessarily in motility. More recently, this family has been predicted to be involved in NAD-utilizing pathways, likely to act on ADP-ribose derivatives, and has been named the NADAR (NAD and ADP-ribose) superfamily. [PMID: 22399070, PMID: 17122336]"
142	PF16588	C2H2 zinc-finger	NA
143	IPR036236	Zinc finger C2H2 superfamily	"C2H2-type (classical) zinc fingers (Znf) were the first class to be characterised. They contain a short beta hairpin and an alpha helix (beta/beta/alpha structure), where a single zinc atom is held in place by Cys(2)His(2) (C2H2) residues in a tetrahedral array. C2H2 Znf's can be divided into three groups based on the number and pattern of fingers: triple-C2H2 (binds single ligand), multiple-adjacent-C2H2 (binds multiple ligands), and separated paired-C2H2. C2H2 Znf's are the most common DNA-binding motifs found in eukaryotic transcription factors, and have also been identified in prokaryotes. Transcription factors usually contain several Znf's (each with a conserved beta/beta/alpha structure) capable of making multiple contacts along the DNA, where the C2H2 Znf motifs recognise DNA sequences by binding to the major groove of DNA via a short alpha-helix in the Znf, the Znf spanning 3-4 bases of the DNA. C2H2 Znf's can also bind to RNA and protein targets.Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target."
144	cd17335	Major facilitator superfamily domain-	"Human Major facilitator superfamily domain-containing protein 6 (MFSD6) is also called macrophage MHC class I receptor 2 homolog (MMR2). It has been postulated as a possible receptor for human leukocyte antigen (HLA)-B62. MFSD6 is conserved through evolution and appeared before bilateral animals. It belongs to the Major Facilitator Superfamily (MFS) of membrane transport proteins, which are thought to function through a single substrate binding site, alternating-access mechanism involving

		containing protein 6	a rocker-switch type of movement. [PMID: 20123006, PMID: 21044875, PMID: 20648054, PMID: 17010536, PMID: 26758938, PMID: 26234418, PMID: 26098515, PMID: 23530251, PMID: 23403214, PMID: 18537473, PMID: 16842212]"
1 4 5	IPR000571	"Zinc finger, CCCH-type"	"Zinc finger (Znf) domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule. Some of these domains bind zinc, but many do not; instead binding other metals such as iron, or no metal at all. For example, some family members form salt bridges to stabilise the finger-like folds. They were first identified as a DNA-binding motif in transcription factor TFIIIA from Xenopus laevis (African clawed frog), however they are now recognised to bind DNA, RNA, protein and/or lipid substrates. Their binding properties depend on the amino acid sequence of the finger domains and of the linker between fingers, as well as on the higher-order structures and the number of fingers. Znf domains are often found in clusters, where fingers can have different binding specificities. There are many superfamilies of Znf motifs, varying in both sequence and structure. They display considerable versatility in binding modes, even between members of the same class (e.g. some bind DNA, others protein), suggesting that Znf motifs are stable scaffolds that have evolved specialised functions. For example, Znf-containing proteins function in gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing, to name but a few. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.This entry represents C-x8-C-x5-C-x3-H (CCCH) type Zinc finger (Znf) domains. Proteins containing CCCH Znf domains include Znf proteins from eukaryotes involved in cell cycle or growth phase-related regulation, e.g. human TIS11B (butyrate response factor 1), a probable regulatory protein involved in regulating the response to growth factors, and the mouse TTP growth factor-inducible nuclear protein, which has the same function. The mouse TTP protein is induced by growth factors. Another protein containing this domain is the human splicing factor U2AF 35kDa subunit, which plays a critical role in both constitutive and enhancer-dependent splicing by mediating essential protein-protein interactions and protein-RNA interactions required for 3' splice site selection. It has been shown that different CCCH-type Znf proteins interact with the 3'-untranslated region of various mRNA. This type of Znf is very often present in two copies."
1 4 6	PS50103	Zinc finger C3H1-type profile	"The well conserved cysteine-histidine arrangement, C-x(8)-C-x(5)-C-x(3)-H, of the C3H1-type zinc finger is present in many eukaryotic proteins from yeast to mammals. In various proteins the C3H1-type zinc finger has been shown to interact with the 3'-untranslated region of mRNAs (see below). It is very often present in two copies [PMID:15963892,PMID:15718139]. The structure of the two C3H1-type zinc fingers of the protein TIS11d bound to its target mRNA has been solved[PMID:14981510]. The two zinc fingers fold independently to form small, compact domains connected by the linker sequence. The polypeptide backbone adopts little regular secondary structure. In each zinc finger, there is a short alpha-helix immediately after the first cysteine ligand and a turn of 3/10 helix between the second and third cysteine ligands. All amino acid residues that exhibit intermolecular contacts with the RNA are positioned on a single surface of each zinc finger. Some proteins known to contain a C3H1-type zinc finger are listed below: - Mammalian tristetraprolin (TTP) proteins, prototype of the C3H1 zinc finger proteins, inhibit TNF alpha production from macrophages by destabilizing its messenger RNA[PMID:15718139]. TTP is a specific RNA-binding protein; the region encompassing the two C3H1-type zinc fingers recognizes nonameric UUAUUUAUU sequences present in the 3'-untranslated region (UTR) of TNF alpha [PMID:10330172,PMID:17030620]. - Eukaryotic TIS/CTH family. Regulatory protein involved in regulating the response to growth factors similar to TTP. The tandem zinc finger domain of the protein TIS11d binds to the class II AU-rich element (ARE) in the 3' untranslated region (3' UTR) of target mRNAs and promotes their deadenylation and degradation[PMID:14981510]. - Eukaryotic U2AG, U2R1, U2R2 proteins, small nuclear ribonucleoproteins (1 or 2 copies). - Caenorhabditis Elegans PIE1 protein. It is required for specification of embryonic germ cells lineage and for the lack of mRNA transcription in these cells (2 copies). - Caenorhabditis Elegans POS-1 protein. A cytoplasmic protein similar to TIS11, essential for germ line specification in C. elegans (2 copies). - Eukaryotic cleavage and

				polyadenylation specificity factor subunit 4 (Cpsf4) protein. It plays a key role in pre-mRNA 3'-end formation, recognizing the AAUAAA signal sequence and interacting with poly(A) polymerase and other factors to bring about cleavage and poly(A) addition. Cpsf4 binds RNA polymers with a preference for poly(U) (5 copies). - Schizosaccharomyces pombe ZFS1 protein. It was isolated as a suppressor of the sterility caused by overexpression of a double-stranded RNase (2 copies). - Human Z183 protein, contains one ring finger and one C3H1-type zinc finger. - Mammalian Nucleoporin-like 2, required for the export of mRNAs containing poly(A) tails from the nucleus into the cytoplasm (1 copy). The profile we developed covers the entire C3H1-type zinc finger."
1 4 7	PS5060 0	Ubiquitin-like protease family profile		"Deubiquitinating enzymes (DUB) form a large family of cysteine protease that can deconjugate ubiquitin or ubiquitin-like proteins from ubiquitin-conjugated proteins. They can be classified in 3 families according to sequence homology [PMID:10603300,PMID:8982460]: ubiquitin carboxyl-terminal hydrolases (UCH), ubiquitin-specific processing proteases (UBP), and ubiquitin-like proteases (ULP) (EC 3.4.22.68) specific for deconjugating ubiquitin-like proteins. In contrast to the UBP pathway, which is very redundant (16 UBP enzymes in yeast), there is few ubiquitin-like protease (only one in yeast, ULP1). ULP1 catalyses two critical functions in the SUMO/Smt3 pathway via its cysteine protease activity. ULP1 processes the Smt3 C-terminal sequence (-GGATY) to its mature form (-GG), and it deconjugates Smt3 from the lysine epsilon-amino group of the target protein[PMID:10094048]. Crystal structure of yeast ULP1 bound to Smt3[PMID:10882122]revealed that the catalytic and interaction interface is situated in a shallow and narrow cleft where conserved residues recognize the Gly-Gly motif at the C-terminal extremity of Smt3 protein. Ulp1 adopts a novel architecture despite some structural similarity with other cysteine protease. The secondary structure is composed of seven alpha helices and seven beta strands. The catalytic domain includes the central alpha helix, beta-strands 4 to 6, and the catalytic triad (Cys-His-Asp). We developed a profile directed against the C-terminal part of ULP proteins that displays full proteolytic activity[PMID:10882122]."
1 4 8	PF0290 2	"Ulp1 protease family, C-terminal catalytic domain"		This domain contains the catalytic triad Cys-His-Asn.
1 4 9	IPR003 653	"Ulp1 protease family, C-terminal catalytic domain"		"This entry represents the C-terminal part of ubiquitin-like proteases that displays full proteolytic activity.Deubiquitinating enzymes (DUB) form a large family of cysteine protease that can deconjugate ubiquitin or ubiquitin-like proteins from ubiquitin-conjugated proteins. They can be classified in 3 families according to sequence homology: ubiquitin carboxyl-terminal hydrolases (UCH), ubiquitin-specific processing proteases (UBP), and ubiquitin-like proteases (ULP) () specific for deconjugating ubiquitin-like proteins. In contrast to the UBP pathway, which is very redundant (16 UBP enzymes in yeast), there is few ubiquitin-like protease (only one in yeast, ULP1).ULP1 catalyses two critical functions in the SUMO/Smt3 pathway via its cysteine protease activity. ULP1 processes the Smt3 C-terminal sequence (-GGATY) to its mature form (-GG), and it deconjugates Smt3 from the lysine epsilon-amino group of the target protein.Crystal structure of yeast ULP1 bound to Smt3revealed that the catalytic and interaction interface is situated in a shallow and narrow cleft where conserved residues recognise the Gly-Gly motif at the C-terminal extremity of Smt3 protein. Ulp1 adopts a novel architecture despite some structural similarity with other cysteine protease. The secondary structure is composed of seven alpha helices and seven beta strands. The catalytic domain includes the central alpha helix, beta-strands 4 to 6, and the catalytic triad (Cys-His-Asp).Cysteine peptidases with a chymotrypsin-like fold are included in clan PA, which also includes serine peptidases. Cysteine peptidases that are N-terminal nucleophile hydrolases are included in clan PB. Cysteine peptidases with a tertiary structure similar to that of

			the serine-type aspartyl dipeptidase are included in clan PC. Cysteine peptidases with an intein-like fold are included in clan PD, which also includes asparagine lyases.A cysteine peptidase is a proteolytic enzyme that hydrolyses a peptide bond using the thiol group of a cysteine residue as a nucleophile. Hydrolysis involves usually a catalytic triad consisting of the thiol group of the cysteine, the imidazolium ring of a histidine, and a third residue, usually asparagine or aspartic acid, to orientate and activate the imidazolium ring. In only one family of cysteine peptidases, is the role of the general base assigned to a residue other than a histidine: in peptidases from family C89 (acid ceramidase) an arginine is the general base. Cysteine peptidases can be grouped into fourteen different clans, with members of each clan possessing a tertiary fold unique to the clan. Four clans of cysteine peptidases share structural similarities with serine and threonine peptidases and asparagine lyases. From sequence similarities, cysteine peptidases can be clustered into over 80 different families. Clans CF, CM, CN, CO, CP and PD contain only one family.Cysteine peptidases are often active at acidic pH and are therefore confined to acidic environments, such as the animal lysosome or plant vacuole. Cysteine peptidases can be endopeptidases, aminopeptidases, carboxypeptidases, dipeptidyl-peptidases or omega-peptidases. They are inhibited by thiol chelators such as iodoacetate, iodoacetic acid, N-ethylmaleimide or p-chloromercuribenzoate.Clan CA includes proteins with a papain-like fold. There is a catalytic triad which occurs in the order: Cys/His/Asn (or Asp). A fourth residue, usually Gln, is important for stabilising the acyl intermediate that forms during catalysis, and this precedes the active site Cys. The fold consists of two subdomains with the active site between them. One subdomain consists of a bundle of helices, with the catalytic Cys at the end of one of them, and the other subdomain is a beta-barrel with the active site His and Asn (or Asp). There are over thirty families in the clan, and tertiary structures have been solved for members of most of these. Peptidases in clan CA are usually sensitive to the small molecule inhibitor E64, which is ineffective against peptidases from other clans of cysteine peptidases.Clan CD includes proteins with a caspase-like fold. Proteins in the clan have an alpha/beta/alpha sandwich structure. There is a catalytic dyad which occurs in the order His/Cys. The active site His occurs in a His-Gly motif and the active site Cys occurs in an Ala-Cys motif; both motifs are preceded by a block of hydrophobic residues. Specificity is predominantly directed towards residues that occupy the S1 binding pocket, so that caspases cleave aspartyl bonds, legumains cleave asparaginyl bonds, and gingipains cleave lysyl or arginyl bonds.Clan CE includes proteins with an adenain-like fold. The fold consists of two subdomains with the active site between them. One domain is a bundle of helices, and the other a beta barrell. The subdomains are in the opposite order to those found in peptidases from clan CA, and this is reflected in the order of active site residues: His/Asn/Gln/Cys. This has prompted speculation that proteins in clans CA and CE are related, and that members of one clan are derived from a circular permutation of the structure of the other.Clan CL includes proteins with a sortase B-like fold. Peptidases in the clan hydrolyse and transfer bacterial cell wall peptides. The fold shows a closed beta barrel decorated with helices with the active site at one end of the barrel. The active site consists of a His/Cys catalytic dyad."
1 5 0	IPR00 1498	"Impact, N-terminal"	"The Impact protein is a translational regulator that ensures constant high levels of translation under amino acid starvation. It acts by interacting with Gcn1/Gcn1L1, thereby preventing activation of Gcn2 protein kinases (EIF2AK1 to 4) and subsequent down-regulation of protein synthesis. It is evolutionary conserved from eukaryotes to archaea.This entry represents the N-terminal domain of the Impact proteins."
1 5 1	IPR02 3582	Impact family	"The Impact protein is a translational regulator that ensures constant high levels of translation under amino acid starvation. It acts by interacting with Gcn1/Gcn1L1, thereby preventing activation of Gcn2 protein kinases (EIF2AK1 to 4) and subsequent down-regulation of protein synthesis. It is evolutionary conserved from eukaryotes to archaea."
1 5 2	IPR02 0568	Ribosomal protein S5	"Domain 2 of the ribosomal protein S5 has a left-handed, 2-layer alpha/beta fold with a core structure consisting of beta(3)-alpha-beta-alpha. Domains with this fold are found in numerous RNA/DNA-binding proteins, as well as in kinases from the GHMP kinase family. Proteins containing this alpha/beta fold domain include: * Translational machinery components

		domain 2-type fold	(ribosomal proteins S5 and S9, and domain IV of elongation factors EF-G and eEF-2). * Ribonuclease P protein (RNase P). * Ribonuclease PH (domain 1), as well as various exosome complex exonucleases (RRP41, RRP42, RRP43, RRP45, RRP46, MTR3, ECX1, ECX2). * DNA modification proteins (DNA mismatch repair proteins MutL and PMS2, DNA gyrase B, DNA topoisomerase II, IV-B and VI-B). * GHMP kinases that transfer a phosphoryl group from ATP to an acceptor (galactokinase (), homoserine kinase (), and mevalonate kinase ()). * Caenorhabditis elegans early switch protein Xol-1 (a divergent member of the GHMP kinase family that has lost the ATP-binding site). * Hsp90 chaperone (middle domain), which is related to the DNA gyrase/MutL family; this domain contains an extra C-terminal alpha/beta subdomain. * Imidazole glycerol phosphate dehydratase, which contains a duplication consisting of two structural repeats of this fold. * The catalytic domain of ATP-dependent protease Lon (La), which contains an extra C-terminal alpha/beta subdomain. * Formaldehyde-activating enzyme FAE, which contains a modification of this fold consisting of an extra alpha/beta unit after strand 2."
1 5 3	PF01205	Uncharacterized protein family UPF0029	NA
1 5 4	IPR036956	"Impact, N-terminal domain superfamily"	"The Impact protein is a translational regulator that ensures constant high levels of translation under amino acid starvation. It acts by interacting with Gcn1/Gcn1L1, thereby preventing activation of Gcn2 protein kinases (EIF2AK1 to 4) and subsequent down-regulation of protein synthesis. It is evolutionary conserved from eukaryotes to archaea. This entry represents the N-terminal domain superfamily of the Impact proteins."
1 5 5	IPR033134	"Asp/Glu racemase, active site 2"	"Aspartate racemase () and glutamate racemase () are two evolutionary related bacterial enzymes that do not seem to require a cofactor for their activity. Glutamate racemase, which interconverts L-glutamate into D-glutamate, is required for the biosynthesis of peptidoglycan and some peptide-based antibiotics such as gramicidin S. In addition to characterised aspartate and glutamate racemases, this entry also includes a hypothetical protein from Erwinia carotovora and one from Escherichia coli (ygeA). Two conserved cysteines are present in the sequence of these enzymes. They are expected to play a role in catalytic activity by acting as bases in proton abstraction from the substrate. This entry represents a conserved region containing the second cysteine."
1 5 6	PS00924	Putative nuclease HARBI1	NA
1 5 7	cd05931	Fatty acyl-AMP ligase (FAAL)	"FAAL belongs to the class I adenylate forming enzyme family and is homologous to fatty acyl-coenzyme A (CoA) ligases (FACLs). However, FAALs produce only the acyl adenylate and are unable to perform the thioester-forming reaction, while FACLs perform a two-step catalytic reaction; AMP ligation followed by CoA ligation using ATP and CoA as cofactors. FAALs have insertion motifs between the N-terminal and C-terminal subdomains that distinguish them from the FACLs. This insertion motif precludes the binding of CoA, thus preventing CoA ligation. It has been suggested that the acyl adenylates serve as substrates for multifunctional polyketide synthases to permit synthesis of complex lipids such as phthiocerol dimycocerosate, sulfolipids, mycolic acids, and mycobactin. [PMID: 21185305, PMID: 21601529, PMID: 12406578, PMID: 15112993, PMID: 12598647, PMID: 9373621, PMID: 17762044, PMID: 11255012]"
1 5 8	IPR040097	Fatty acyl-AMP ligase	"Fatty acyl-AMP ligase (FAAL) belongs to the class I adenylate forming enzyme family and is homologous to fatty acyl-coenzyme A (CoA) ligases (FACLs). However, FAALs produce only the acyl adenylate and are unable to perform the thioester-forming reaction, while FACLs perform a two-step catalytic reaction; AMP ligation followed by CoA ligation using ATP and CoA as cofactors. FAALs have insertion motifs between the N-terminal and C-terminal subdomains that distinguish them from

				the FACLs. This insertion motif precludes the binding of CoA, thus preventing CoA ligation. It has been suggested that the acyl adenylates serve as substrates for multifunctional polyketide synthases to permit synthesis of complex lipids such as phthiocerol dimycocerosate, sulfolipids, mycolic acids, and mycobactin."
15959	IPR002219	"Protein kinase C-like, phorbol ester/diacylglycerol-binding domain"		"Diacylglycerol (DAG) is an important second messenger. Phorbol esters (PE) are analogues of DAG and potent tumour promoters that cause a variety of physiological changes when administered to both cells and tissues. DAG activates a family of serine/threonine protein kinases, collectively known as protein kinase C (PKC). Phorbol esters can directly stimulate PKC. The N-terminal region of PKC, known as C1, has been shownto bind PE and DAG in a phospholipid and zinc-dependent fashion. The C1 region contains one or two copies (depending on the isozyme of PKC) of a cysteine-rich domain, which is about 50 amino-acid residues long, and which is essential for DAG/PE-binding. The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain."
16090	PS50081	Zinc finger phorbol-ester/DAG-type profile		"Diacylglycerol (DAG) is an important second messenger. Phorbol esters (PE) are analogues of DAG and potent tumor promoters that cause a variety of physiological changes when administered to both cells and tissues. DAG activates a family of serine/threonine protein kinases, collectively known as protein kinase C (PKC)[PMID:1396661]. Phorbol esters can directly stimulate PKC. The N- terminal region of PKC, known as C1, has been shown[PMID:2500657]to bind PE and DAG in a phospholipid and zinc-dependent fashion. The C1 region contains one or two copies (depending on the isozyme of PKC) of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. Such a domain has also been found in the following proteins: - Diacylglycerol kinase (EC 2.7.1.107) (DGK)[PMID:2156169], the enzyme that converts DAG into phosphatidate. It contains two copies of the DAG/PE-binding domain in its N-terminal section. At least five different forms of DGK are known in mammals. - N-chimaerin. A brain specific protein which shows sequence similarities with the BCR protein at its C-terminal part and contains a single copy of the DAG/PE-binding domain at its N-terminal part. It has been shown [PMID:2268301,PMID:1660266] to be able to bind phorbol esters. - The raf/mil family of serine/threonine protein kinases. These protein kinases contain a single N-terminal copy of the DAG/PE-binding domain. - The unc-13 protein from Caenorhabditis elegans. Its function is not known but it contains a copy of the DAG/PE-binding domain in its central section and has been shown to bind specifically to a phorbol ester in the presence of calcium[PMID:1445255]. - The vav oncogene. Vav was generated by a genetic rearrangement during gene transfer assays. Its expression seems to be restricted to cells of hematopoeitic origin. Vav seems [PMID:1660266,PMID:1614545] to contain a DAG/PE-binding domain in the central part of the protein. - The Drosophila GTPase activating protein rotund. The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain. We have developed both a signature pattern and a profile that span completely the DAG/PE domain."
16061	PS50045	Sigma-54 interaction domain profile		"Some bacterial regulatory proteins activate the expression of genes from promoters recognized by core RNA polymerase associated with the alternative sigma-54 factor. These have a conserved domain of about 230 residues involved in the ATP-dependent[PMID:2041769]interaction with sigma-54. This domain has been found in the proteins listed below: - acoR from Alcaligenes eutrophus, an activator of the acetoin catabolism operon acoXABC. - algB from Pseudomonas aeruginosa, an activator of alginate biosynthetic gene algD. - dctD from Rhizobium, an activator of dctA, the C4-dicarboxylate transport protein. - dhaR from Citrobacter freundii, a regulator of the dha operon for glycerol utilization. - fhlA from Escherichia coli, an activator of the formate dehydrogenase H and hydrogenase III structural genes. - flbD from Caulobacter crescentus, an activator of flagellar genes. - hoxA from Alcaligenes eutrophus, an activator of the hydrogenase operon. - hrpS from Pseudomonas syringae, an activator of hprD as well as other hrp loci involved in plant pathogenicity. - hupR1 from Rhodobacter capsulatus, an activator of the [NiFe] hydrogenase genes hupSL. - hydG from Escherichia coli and Salmonella typhimurium, an activator of the hydrogenase activity. - levR from Bacillus subtilis, which regulates the expression of the levanase operon (levDEFG and

			sacC). - nifA (as well as anfA and vnfA) from various bacteria, an activator of the nif nitrogen-fixing operon. - ntrC, from various bacteria, an activator of nitrogen assimilatory genes such as that for glutamine synthetase (glnA) or of the nif operon. - pgtA from Salmonella typhimurium, the activator of the inducible phospho- glycerate transport system. - pilR from Pseudomonas aeruginosa, an activator of pilin gene transcription. - rocR from Bacillus subtilis, an activator of genes for arginine utilization - tyrR from Escherichia coli, involved in the transcriptional regulation of aromatic amino-acid biosynthesis and transport. - wtsA, from Erwinia stewartii, an activator of plant pathogenicity gene wtsB. - xylR from Pseudomonas putida, the activator of the tol plasmid xylene catabolism operon xylCAB and of xylS. - Escherichia coli hypothetical protein yfhA. - Escherichia coli hypothetical protein yhgB. About half of these proteins (algB, dcdT, flbD, hoxA, hupR1, hydG, ntrC, pgtA and pilR) belong to signal transduction two-component systems[PMID:2694934]and possess a domain that can be phosphorylated by a sensor-kinase protein in their N- terminal section. Almost all of these proteins possess a helix-turn-helix DNA-binding domain in their C-terminal section. The domain which interacts with the sigma-54 factor has an ATPase activity. This may be required to promote a conformational change necessary for the interaction[PMID:15347752]. The domain contains an atypical ATP-binding motif A (P-loop) as well as a form of motif B. The two ATP-binding motifs are located in the N- terminal section of the domain; we have developed signature patterns for both motifs. Other regions of the domain are also conserved. We have selected one of them, located in the C-terminal section, as a third signature pattern."
1 6 2	IPR002078	RNA polymerase sigma factor 54 interaction domain	"Some bacterial regulatory proteins activate the expression of genes from promoters recognised by core RNA polymerase associated with the alternative sigma-54 factor. These have a conserved domain of about 230 residues involved in the ATP-dependentinteraction with sigma-54. About half of the proteins in which this domain is found (algB, dcdT, flbD, hoxA, hupR1, hydG, ntrC, pgtA and pilR) belong to signal transduction two-component systemsand possess a domain that can be phosphorylated by a sensor-kinase protein in their N-terminal section. Almost all of these proteins possess a helix-turn-helix DNA-binding domain in their C-terminal section.The domain which interacts with the sigma-54 factor has an ATPase activity. This may be required to promote a conformational change necessary for the interaction. The domain contains an atypical ATP-binding motif A (P-loop) as well as a form of motif B. The two ATP-binding motifs are located in the N-terminal section of the domain."
1 6 3	IPR000851	Ribosomal protein S5	"Ribosomes are the particles that catalyse mRNA-directed protein synthesis in all organisms. The codons of the mRNA are exposed on the ribosome to allow tRNA binding. This leads to the incorporation of amino acids into the growing polypeptide chain in accordance with the genetic information. Incoming amino acid monomers enter the ribosomal A site in the form of aminoacyl-tRNAs complexed with elongation factor Tu (EF-Tu) and GTP. The growing polypeptide chain, situated in the P site as peptidyl-tRNA, is then transferred to aminoacyl-tRNA and the new peptidyl-tRNA, extended by one residue, is translocated to the P site with the aid the elongation factor G (EF-G) and GTP as the deacylated tRNA is released from the ribosome through one or more exit sites. About 2/3 of the mass of the ribosome consists of RNA and 1/3 of protein. The proteins are named in accordance with the subunit of the ribosome which they belong to - the small (S1 to S31) and the large (L1 to L44). Usually they decorate the rRNA cores of the subunits.Many ribosomal proteins, particularly those of the large subunit, are composed of a globular, surfaced-exposed domain with long finger-like projections that extend into the rRNA core to stabilise its structure. Most of the proteins interact with multiple RNA elements, often from different domains. In the large subunit, about 1/3 of the 23S rRNA nucleotides are at least in van der Waal's contact with protein, and L22 interacts with all six domains of the 23S rRNA. Proteins S4 and S7, which initiate assembly of the 16S rRNA, are located at junctions of five and four RNA helices, respectively. In this way proteins serve to organise and stabilise the rRNA tertiary structure. While the crucial activities of decoding and peptide transfer are RNA based, proteins play an active role in functions that may have evolved to streamline the process of protein synthesis. In addition to their function in the ribosome, many ribosomal proteins have some function 'outside'

			the ribosome.Ribosomal protein S5 is one of the proteins from the small ribosomal subunit, and is a protein of 166 to 254 amino-acid residues. In Escherichia coli, S5 is known to be important in the assembly and function of the 30S ribosomal subunit. Mutations in S5 have been shown to increase translational error frequencies. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups bacterial, cyanelle, red algal chloroplast, archaeal and fungal mitochondrial S5; mammalian, Caenorhabditis elegans, Drosophila and plant S2; and yeast S4 (SUP44)."
1 6 4	PS515 26	RFX-type winged-helix DNA-binding domain profile	"The RFX family of transcription factors is characterized by a unique ~75- residue DNA-binding domain. RFX genes have been isolated in yeasts, nematode and vertebrates. The characteristic RFX-type HTH DNA binding domain has been recruited into otherwise very divergent regulatory factors functioning in a diverse spectrum of unrelated systems, including regulation of the mitotic cell cycle in fission yeast, the control of the immune response in mammals, and infection by human hepatitis B virus [PMID:8600444,PMID:10767550]. The RFX-type DNA-binding domain is an unusual member of the winged-helix subfamily of helix-turn-helix domains because it uses a beta-hairpin (or wing) to recognize DNA instead of the recognition helix typical of helix-tun-helix domains. It consists of three alpha-helices (H), three beta-strands (S) and three connecting loops (L), arranged in the order H1-S1-H2-L1-H3-L2-S2-W1-S3. The third loop, connecting beta-strands S2 and S3, forms wing W1 of the winged-helix motif. In contrast to previously described winged-helix DBDs (typically 110 residues long with two wings, W1 and W2), the shorter RFX-type winged-helix DNA-binding domain has only one wing[PMID:10706293]. Some proteins known to contain a RFX-type winged-helix DNA-binding domain are listed below: - Human RFX1, a cellular transactivator that is used by the highly pathogenic hepatitis B virus. - Human RFX2. - Human RFX3, a transcription factor required for ciliogenesis and islet cell differentiation during endocrine pancreas development. - Human RFX5, a regulator of MHC class II gene transcription. - Human ARID2, a component of the SWI/SNF-B (PBAF) chromatin remodeling complex[PMID:15640446]. - Yeast Crt1, an effector of the DNA damage and replication checkpoint. - Schizosaccharomyces pombe sak1, an essential regulatory protein in the life cycle. It allows cells to exit the mitotic cycle and enter either the stationary phase or the pathway leading to sexual differentiation. The profile we developed covers the entire RFX-type winged-helix DNA-binding domain."
1 6 5	IPR00 1487	Bromodomai n	"Bromodomains are found in a variety of mammalian, invertebrate and yeast DNA-binding proteins. Bromodomains can interact with acetylated lysine. In some proteins, the classical bromodomain has diverged to such an extent that parts of the region are either missing or contain an insertion (e.g., mammalian protein HRX, Caenorhabditis elegans hypothetical protein ZK783.4, yeast protein YTA7). The bromodomain may occur as a single copy, or in duplicate.The precise function of the domain is unclear, but it may be involved in protein-protein interactions and may play a role in assembly or activity of multi-component complexes involved in transcriptional activation."
1 6 6	cd081 97	2-deoxy-scyllo-inosose synthase (DOIS) catalyzes carbocycle formation from D-glucose-6-phosphate to 2-deoxy-scyllo-inosose	"2-deoxy-scyllo-inosose synthase (DOIS) catalyzes carbocycle formation from D-glucose-6-phosphate to 2-deoxy-scyllo-inosose through a multistep reaction in the biosynthesis of aminoglycoside antibiotics. 2-deoxystreptamine (DOS)-containing aminoglycoside antibiotics includes neomycin, kanamycin, gentamicin, and ribostamycin. They are important antibacterial agents. DOIS is a homolog of the dehydroquinate synthase which catalyzes the cyclization of 3-deoxy-D-arabino-heptutulosonate-7-phosphate to dehydroquinate (DHQ) in the shikimate pathway. [PMID: 17879343, PMID: 16258246, PMID: 15313224, PMID: 17035031, PMID: 16915821, PMID: 17368605, PMID: 15928851, PMID: 15788164, PMID: 16861807, PMID: 14757238, PMID: 14750780, PMID: 16506694]"

1 6 7	IPR02 2690	"Geminivirus AL1 replication- associated protein, catalytic domain"	"Geminiviruses are characterised by a genome of circular single-stranded DNA encapsidated in twinned (geminate) quasi-isometric particles, from which the group derives its name. Most geminiviruses can be divided into two subgroups on the basis of host range and/or insect vector: i.e. those that infect dicotyledenous plants and are transmitted by the same whitefly species, and those that infect monocotyledenous plants and are transmitted by different leafhopper vectors. The genomes of the whitefly-transmitted African cassava mosaic virus, Tomato golden mosaic virus (TGMV) and Bean golden mosaic virus (BGMV) possess a bipartite genome. By contrast, only a single DNA component has been identified for the leafhopper-transmitted Maize streak virus (MSV) and Wheat dwarf virus (WDV). Beet curly top virus (BCTV), and Tobacco yellow dwarf virus belong to a third possible subgroup. Like MSV and WDV, BCTV is transmitted by a specific leafhopper species, yet like the whitefly-transmitted geminiviruses it has a host range confined to dicotyledenous plants.Sequence comparison of the whitefly-transmitted Squash leaf curl virus (SqLCV) and Tomato yellow leaf curl virus (TYLCV) with the genomic components of TGMV and BGMV reveals a close evolutionary relationship. Amino acid sequence alignments of Potato yellow mosaic virus (PYMV) proteins with those encoded by other geminiviruses show that PYMV is closely related to geminiviruses isolated from the New World, especially in the putative coat protein gene regions. Comparison of MSV DNA-encoded proteins with those of other geminiviruses infecting monocotyledonous plants, including Panicum streak virusand Miscanthus streak virus (MiSV), reveal high levels of similarity.The AL1 proteins encodes the replication initiator protein (Rep) of geminiviruses, which is a replicon-specific initiator enzyme and is an essential component of the replisome. For geminivirus Rep protein, this N-terminal region is crucial for origin recognition and DNA cleavage and nucleotidyl transfer. It is found in association with."
1 6 8	PS500 14	Bromodomai n profile	"The bromodomain [PMID:1346755,PMID:7580139] is a conserved region of about 70 amino acids found in the following proteins: - Higher eukaryotes transcription initiation factor TFIID 250 Kd subunit (TBP-associated factor p250) (gene CCG1). P250 associated with the TFIID TATA-box binding protein and seems essential for progression of the G1 phase of the cell cycle. - Human RING3, a protein of unknown function encoded in the MHC class II locus. - Mammalian CREB-binding protein (CBP), which mediates cAMP-gene regulation by binding specifically to phosphorylated CREB protein. - Drosophila female sterile homeotic protein (gene fsh), required maternally for proper expression of other homeotic genes involved in pattern formation, such as Ubx. - Drosophila brahma protein (gene brm), a protein required for the activation of multiple homeotic genes. - Mammalian homologs of brahma. In human, three brahma-like proteins are known: SNF2a(hBRM), SNF2b, and BRG1. - Human BS69, a protein that binds to adenovirus E1A and inhibits E1A transactivation - Human peregrin (or Br140). - Yeast BDF1[PMID:7580139], a transcription factor involved in the expression of a broad class of genes including snRNAs. - Yeast GCN5, a general transcriptional activator operating in concert with certain other DNA-binding transcriptional activators, such as GCN4, HAP2/3/4 or ADA2. - Yeast NPS1/STH1, involved in G(2) phase control in mitosis. - Yeast SNF2/SWI2, which is part of a complex with the SNF5, SNF6, SWI3 and ADR6/SWI1 proteins. This SWI-complex is involved in transcriptional activation. - Yeast SPT7, a transcriptional activator of Ty elements and possibly other genes. - Caenorhabditis elegans protein cbp-1. - Yeast hypothetical protein YGR056w. - Yeast hypothetical protein YKR008w. - Yeast hypothetical protein L9638.1. Some proteins contain a region which, while similar to some extent to a classical bromodomain, diverges from it by either lacking part of the domain or because of an insertion. These proteins are: - Mammalian protein HRX (also known as All-1 or MLL), a protein involved in translocations leading to acute leukemias and which possibly acts as a transcriptional regulatory factor. HRX contains a region similar to the C- terminal half of the bromodomain. - Caenorhabditis elegans hypothetical protein ZK783.4. The bromodomain of this protein has a 23 amino-acid insertion. - Yeast protein YTA7. This protein contains a region with significant similarity to the C-terminal half of the bromodomain. As it is a member of the AAA family it is also in a functionally different context. The above proteins generally contain a single bromodomain, but some of them contain two copies, this is the case of BDF1, CCG1, fsh, RING3, YKR008w and L9638.1. The exact function of this domain is not yet

				known but it is thought to be involved in protein-protein interactions and it may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation. The consensus pattern that we have developed spans a major part of the bromodomain; a more sensitive detection is available through the use of a profile which spans the whole domain."
1 6 9	PF007 99	Geminivirus Rep catalytic domain		"The AL1 proteins encodes the replication initiator protein (Rep) of geminiviruses, which is a replicon-specific initiator enzyme and is an essential component of the replisome[PMID:12130667]. For geminivirus Rep protein, this N-terminal region is crucial for origin recognition and DNA cleavage and nucleotidyl transfer[PMID:12130667]."
1 7 0	IPR00 8162	Inorganic pyrophosphatase		"Inorganic pyrophosphatase () (PPase)is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed principally as the product of the many biosynthetic reactions that utilise ATP. All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity. Among other residues, a lysine has been postulated to be part of or close to the active site. PPases have been sequenced from bacteria such as Escherichia coli (homohexamer), Bacillus PS3 (Thermophilic bacterium PS-3) and Thermus thermophilus, from the archaebacteria Thermoplasma acidophilum, from fungi (homodimer), from a plant, and from bovine retina. In yeast, a mitochondrial isoform of PPase has been characterised which seems to be involved in energy production and whose activity is stimulated by uncouplers of ATP synthesis.The sequences of PPases share some regions of similarities, among which is a region that contains three conserved aspartates that are involved in the binding of cations."
1 7 1	PS003 87	Putative nuclease HARBI1		NA
1 7 2	PS002 90	Putative nuclease HARBI1		NA
1 7 3	IPR00 3006	"Immunoglobulin/major histocompatibility complex, conserved site"		"The basic structure of immunoglobulin (Ig) molecules is a tetramer of two light chains and two heavy chains linked by disulphide bonds. There are two types of light chains: kappa and lambda, each composed of a constant domain (CL) and a variable domain (VL). There are five types of heavy chains: alpha, delta, epsilon, gamma and mu, all consisting of a variable domain (VH) and three (in alpha, delta and gamma) or four (in epsilon and mu) constant domains (CH1 to CH4). Ig molecules are highly modular proteins, in which the variable and constant domains have clear, conserved sequence patterns. The domains in Ig and Ig-like molecules are grouped into four types: V-set (variable;), C1-set (constant-1;), C2-set (constant-2;) and I-set (intermediate;). Structural studies have shown that these domains share a common core Greek-key beta-sandwich structure, with the types differing in the number of strands in the beta-sheets as well as in their sequence patterns.Immunoglobulin-like domains that are related in both sequence and structure can be found in several diverse protein families. Ig-like domains are involved in a variety of functions, including cell-cell recognition, cell-surface receptors, muscle structure and the immune system.Major Histocompatibility Complex (MHC) glycoproteins are heterodimeric cell surface receptors that function to present antigen peptide fragments to T cells responsible for cell-mediated immune responses. MHC molecules can be subdivided into two groups on the basis of structure and function: class I molecules present intracellular antigen peptide fragments (~10 amino acids) on the surface of the host cells to cytotoxic T cells; class II molecules present exogenously derived antigenic peptides (~15 amino acids) to helper T cells. MHC class I and II molecules are assembled and loaded with their peptide ligands via different mechanisms. However, both present peptide fragments rather than entire proteins to T cells, and are required to mount an immune response.Some of the proteins in this group are responsible for the molecular basis of the blood group antigens, surface markers on the outside of the red blood cell membrane. Most of these markers are proteins, but some are carbohydrates attached to lipids or proteins [Reid M.E., Lomas-Francis C. The Blood Group Antigen FactsBook Academic Press, London / San Diego,

			(1997)]. Lutheran blood group glycoprotein (B-CAM cell surface glycoprotein) (Auberger B antigen) (F8/G253 antigen) belongs to the Lutheran blood group system and is associated with Lu(a/b), Au(a/b), LU6 to LU20 antigens."
1 7 4	IPR03 9091	Aryl hydrocarbon receptor/Aryl hydrocarbon receptor repressor	"Aryl hydrocarbon receptor (AHR) is a ligand-activated transcription factor that mediates the biological and toxic effects of dioxin-like compounds. Upon ligand binding, the receptor translocates into the nucleus where it heterodimerises with ARNT (AHR nuclear translocator). The aryl hydrocarbon receptor repressor (AHRR) competes with AHR for the binding of ARNT, and thereby repressing AHR's transcription activity."
1 7 5	PF023 38	OTU-like cysteine protease	"This family is comprised of a group of predicted cysteine proteases, homologous to the Ovarian Tumour (OTU) gene in Drosophila. Members include proteins from eukaryotes, viruses and pathogenic bacterium. The conserved cysteine and histidine, and possibly the aspartate, represent the catalytic residues in this putative group of proteases."
1 7 6	IPR00 3323	OTU domain	"An homology region containing four conserved motifs has been identified in proteins from eukaryotes, several groups of viruses and the pathogenic bacteria Chlamydia pneumoniae. None of these proteins has a known biochemical function but low sequence similarity with the polyprotein regions of arteriviruses has lead to suggest that it could possess cysteine protease activity. In this case, the conserved cysteine and aspartate in motif I and the histidine in motif IV could be the catalytic residues. Motifs II and III have a more limited sequence conservation and could be involved in substrate recognition.It has been proposed that the eukaryotic proteins containing an OTU domain could mediate proteolytic events involved in signalling associated with the modification of chromatin structure and control of cell proliferation.In viruses proteins containing this domain are annotated as replicase or RNA-dependent RNA polymerase. The eukaryotic sequences are related to the Ovarian Tumour (OTU) gene in Drosophila, cezanne deubiquitinating peptidase and tumor necrosis factor, alpha-induced protein 3 (MEROPS peptidase family C64) and otubain 1 and otubain 2 (MEROPS peptidase family C65).Cysteine peptidases with a chymotrypsin-like fold are included in clan PA, which also includes serine peptidases. Cysteine peptidases that are N-terminal nucleophile hydrolases are included in clan PB. Cysteine peptidases with a tertiary structure similar to that of the serine-type aspartyl dipeptidase are included in clan PC. Cysteine peptidases with an intein-like fold are included in clan PD, which also includes asparagine lyases.A cysteine peptidase is a proteolytic enzyme that hydrolyses a peptide bond using the thiol group of a cysteine residue as a nucleophile. Hydrolysis involves usually a catalytic triad consisting of the thiol group of the cysteine, the imidazolium ring of a histidine, and a third residue, usually asparagine or aspartic acid, to orientate and activate the imidazolium ring. In only one family of cysteine peptidases, is the role of the general base assigned to a residue other than a histidine: in peptidases from family C89 (acid ceramidase) an arginine is the general base. Cysteine peptidases can be grouped into fourteen different clans, with members of each clan possessing a tertiary fold unique to the clan. Four clans of cysteine peptidases share structural similarities with serine and threonine peptidases and asparagine lyases. From sequence similarities, cysteine peptidases can be clustered into over 80 different families. Clans CF, CM, CN, CO, CP and PD contain only one family.Cysteine peptidases are often active at acidic pH and are therefore confined to acidic environments, such as the animal lysosome or plant vacuole. Cysteine peptidases can be endopeptidases, aminopeptidases, carboxypeptidases, dipeptidyl-peptidases or omega-peptidases. They are inhibited by thiol chelators such as iodoacetate, iodoacetic acid, N-ethylmaleimide or p-chloromercuribenzoate.Clan CA includes proteins with a papain-like fold. There is a catalytic triad which occurs in the order: Cys/His/Asn (or Asp). A fourth residue, usually Gln, is important for stabilising the acyl intermediate that forms during catalysis, and this precedes the active site Cys. The fold consists of two subdomains with the active site between them. One subdomain consists of a bundle of helices, with the catalytic Cys at the end of one of them, and the other subdomain is a beta-barrel with the active site His and Asn (or Asp). There are over thirty families in the clan, and tertiary structures have been solved for members of most of these.

			Peptidases in clan CA are usually sensitive to the small molecule inhibitor E64, which is ineffective against peptidases from other clans of cysteine peptidases.Clan CD includes proteins with a caspase-like fold. Proteins in the clan have an alpha/beta/alpha sandwich structure. There is a catalytic dyad which occurs in the order His/Cys. The active site His occurs in a His-Gly motif and the active site Cys occurs in an Ala-Cys motif; both motifs are preceded by a block of hydrophobic residues. Specificity is predominantly directed towards residues that occupy the S1 binding pocket, so that caspases cleave asparty1 bonds, legumains cleave asparaginyl bonds, and gingipains cleave lysyl or arginyl bonds.Clan CE includes proteins with an adenain-like fold. The fold consists of two subdomains with the active site between them. One domain is a bundle of helices, and the other a beta barrell. The subdomains are in the opposite order to those found in peptidases from clan CA, and this is reflected in the order of active site residues: His/Asn/Gln/Cys. This has prompted speculation that proteins in clans CA and CE are related, and that members of one clan are derived from a circular permutation of the structure of the other.Clan CL includes proteins with a sortase B-like fold. Peptidases in the clan hydrolyse and transfer bacterial cell wall peptides. The fold shows a closed beta barrel decorated with helices with the active site at one end of the barrel. The active site consists of a His/Cys catalytic dyad."
1 7 7	PS50802	OTU domain profile	"An homology region containing four conserved motifs has been identified in proteins from eukaryotes, several groups of viruses and the pathogenic bacteria Chlamydia pneumoniae[PMID:10664582]. None of these proteins has a known biochemical function but low sequence similarity with the polyprotein regions of arteriviruses has lead to suggest that it could possess cysteine protease activity[PMID:10664582]. In this case, the conserved cysteine and aspartate in motif I and the histidine in motif IV could be the catalytic residues. Motifs II and III have a more limited sequence conservation and could be involved in substrate recognition[PMID:10664582]. It has been proposed that the eukaryotic proteins containing an OTU domain could mediate proteolytic events involved in signaling associated with the modification of chromatin structure and control of cell proliferation[PMID:10664582]. Some proteins containing a OTU domain are listed below[PMID:10664582]: - Ovarian Tumor (OTU), a Drosophila protein involved in oocyte morphogenesis. As the putative catalytic cysteine is replaced by a serine it is not clear whether OTU is an active protease or in inactivated protease homologue. - Putative HIV-1 induced protein (HIN1), the human homologue of OTU. - B0546.2 protein, the Caenorhabditis elegans homologue of OTU. - M2 protein from La France disease virus. - Tumor-necrosis-factor-alpha-induced protein A20 (TNP3) from mammals. It protects cells from apoptosis and blocks activation of lymphoid cells. In addition to the OTU domain, TNP3 contains several copies of a zinc finger like motif that is also present in other proteins implicated in ubiquitin- dependent proteolysis. - Hypothetical protein YFL044C from yeast, a protein that also contains a C2H2 zinc finger. - Hypothetical protein F12K2.7 protein from Arabidopsis thaliana. - Hypothetical protein C25D7.8 protein from Caenorhabditis elegans. - Hypothetical protein CPN0483 from Chlamydia pneumoniae. The profile we developed covers the entire OTU domain."
1 7 8	IPR001849	Pleckstrin homology domain	"Pleckstrin homology (PH) domains are small modular domains that occur in a large variety of proteins. The domains can bind phosphatidylinositol within biological membranes and proteins such as the beta/gamma subunits of heterotrimeric G proteinsand protein kinase C. Through these interactions, PH domains play a role in recruiting proteins to different membranes, thus targeting them to appropriate cellular compartments or enabling them to interact with other components of the signal transduction pathways.PH domains have been found to possess inserted domains (such as in PLC gamma, syntrophins) and to be inserted within other domains. Mutations in Brutons tyrosine kinase (Btk) within its PH domain cause X-linked agammaglobulinaemia (XLA) in patients. Point mutations cluster into the positively charged end of the molecule around the predicted binding site for phosphatidylinositol lipids.The 3D structure of several PH domains has been determined. All known cases have a common structure consisting of two perpendicular anti-parallel beta sheets, followed by a C-terminal amphipathic helix. The loops connecting the beta-strands differ greatly in length, making the PH domain relatively difficult to detect. There are no totally invariant residues within the PH domain.Proteins reported to contain one more PH domains belong to the

			following families: * Pleckstrin, the protein where this domain was first detected, is the major substrate of protein kinase C in platelets. Pleckstrin is one of the rare proteins to contains two PH domains. * Ser/Thr protein kinases such as the Akt/Rac family, the beta-adrenergic receptor kinases, the mu isoform of PKC and the trypanosomal NrkA family. * Tyrosine protein kinases belonging to the Btk/Itk/Tec subfamily. * Insulin Receptor Substrate 1 (IRS-1). * Regulators of small G-proteins like guanine nucleotide releasing factor GNRP (Ras-GRF) (which contains 2 PH domains), guanine nucleotide exchange proteins like vav, dbl, SoS and Saccharomyces cerevisiae CDC24, GTPase activating proteins like rasGAP and BEM2/IPL2, and the human break point cluster protein bcr. * Cytoskeletal proteins such as dynamin (see), Caenorhabditis elegans kinesin-like protein unc-104 (see), spectrin beta-chain, syntrophin (2 PH domains) and S. cerevisiae nuclear migration protein NUM1. * Mammalian phosphatidylinositol-specific phospholipase C (PI-PLC) (see) isoforms gamma and delta. Isoform gamma contains two PH domains, the second one is split into two parts separated by about 400 residues. * Oxysterol binding proteins OSBP, S. cerevisiae OSH1 and YHR073w. * Mouse protein citron, a putative rho/rac effector that binds to the GTP-bound forms of rho and rac. * Several S. cerevisiae proteins involved in cell cycle regulation and bud formation like BEM2, BEM3, BUD4 and the BEM1-binding proteins BOI2 (BEB1) and BOI1 (BOB1). * C. elegans protein MIG-10. * C. elegans hypothetical proteins C04D8.1, K06H7.4 and ZK632.12. * S. cerevisiae hypothetical proteins YBR129c and YHR155w."
1 7 9	PS500 03	PH domain profile	"The 'pleckstrin homology' (PH) domain is a domain of about 100 residues that occurs in a wide range of proteins involved in intracellular signaling or as constituents of the cytoskeleton [PMID:8500161,PMID:8497315,PMID:8236453,PMID:7985225,PMID:7531822,PMID:7890802,PMID:7583640]. The function of this domain is not clear, several putative functions have been suggested: - binding to the beta/gamma subunit of heterotrimeric G proteins, - binding to lipids, e.g. phosphatidylinositol-4,5-bisphosphate, - binding to phosphorylated Ser/Thr residues, - attachment to membranes by an unknown mechanism. It is possible that different PH domains have totally different ligand requirements. The 3D structure of several PH domains has been determined[PMID:7634082]. All known cases have a common structure consisting of two perpendicular anti-parallel beta sheets, followed by a C-terminal amphipathic helix. The loops connecting the beta-strands differ greatly in length, making the PH domain relatively difficult to detect. There are no totally invariant residues within the PH domain. Proteins reported to contain one more PH domains belong to the following families: - Pleckstrin, the protein where this domain was first detected, is the major substrate of protein kinase C in platelets. Pleckstrin is one of the rare proteins to contains two PH domains. - Ser/Thr protein kinases such as the Act/Rac family, the beta-adrenergic receptor kinases, the mu isoform of PKC and the trypanosomal NrkA family. - Tyrosine protein kinases belonging to the Btk/Itk/Tec subfamily. - Insulin Receptor Substrate 1 (IRS-1). - Regulators of small G-proteins like guanine nucleotide releasing factor GNRP (Ras-GRF) (which contains 2 PH domains), guanine nucleotide exchange proteins like vav, dbl, SoS and yeast CDC24, GTPase activating proteins like rasGAP and BEM2/IPL2, and the human break point cluster protein bcr. - Cytoskeletal proteins such as dynamin, Caenorhabditis elegans kinesin-like protein unc-104, spectrin beta- chain, syntrophin (2 PH domains) and yeast nuclear migration protein NUM1. - Mammalian phosphatidylinositol-specific phospholipase C (PI-PLC) isoforms gamma and delta. Isoform gamma contains two PH domains, the second one is split into two parts separated by about 400 residues. - Oxysterol binding proteins (OSBPs). - Mouse protein citron, a putative rho/rac effector that binds to the GTP-bound forms of rho and rac, - Several yeast proteins involved in cell cycle regulation and bud formation like BEM2, BEM3, BUD4 and the BEM1-binding proteins BOI2 (BEB1) and BOI1 (BOB1). - Caenorhabditis elegans protein mig-10. - Caenorhabditis elegans hypothetical proteins C04D8.1, K06H7.4 and ZK632.12. - Yeast hypothetical proteins YBR129c and YHR155w. The profile for the PH domain, which has been developed by Toby Gibson at the EMBL, covers the total length of domain. Several proteins contain large insertions in the PH domain and are thus difficult to detect with this profile. In some of these cases, the profile will align only to one half of the PH domain."

1880	cd15551	PHD finger found in PYGO proteins	"The family includes Drosophila melanogaster protein pygopus (dPYGO) and its two homologs, PYGO1 and PYGO2. dPYGO is a fundamental Wnt signaling transcriptional component in Drosophila. PYGO1 is essential for the association with Legless (Lgs)/Bcl9 that acts an adaptor between Pygopus (Pygo) and Arm/beta-catenin. dPYGO and PYGO2 function as context-dependent beta-catenin coactivators, and they bind di- and trimethylated lysine 4 of histone H3 (H3K4me2/3). Moreover, PYGO2 acts as a histone methylation reader, and a chromatin remodeler in a testis-specific and Wnt-unrelated manner. It also mediates chromatin regulation and links Wnt signaling and Notch signaling to suppress the luminal/alveolar differentiation competence of mammary stem and basal cells. PYGO2 also plays a new role in rRNA transcription during cancer cell growth. It regulates mammary tumor initiation and heterogeneity in MMTV-Wnt1 mice. All family members contain a plant homeodomain (PHD) finger. [PMID: 11955446, PMID: 19555349, PMID: 17499269, PMID: 17113272, PMID: 18498752, PMID: 23637336, PMID: 11988739, PMID: 20637214, PMID: 23684539, PMID: 23517060, PMID: 23334328, PMID: 20361361, PMID: 24183574, PMID: 12015286, PMID: 18404694, PMID: 24265259, PMID: 24046329]"
1881	cd17065	ubiquitin-like (Ubl) domain found in ubiquitin carboxyl-terminal hydrolase 24 (UBP24) and similar proteins	"UBP24 (EC 3.4.19.12), also termed deubiquitinating enzyme 24, or ubiquitin thioesterase 24, or ubiquitin-specific-processing protease 24 (USP24), is a deubiquitinating protein that interacts with damage-specific DNA-binding protein 2 (DDB2) and regulates DDB2 stability. It may also play a role in the pathogenesis of Parkinson's disease (PD). UBP24 proteins contain an N-terminal ubiquitin-associated (UBA) domain, a ubiquitin-like (Ubl) domain, and a C-terminal peptidase C19 domain. [PMID: 16917932, PMID: 20302855, PMID: 23159851, PMID: 22923019, PMID: 18854859]"
1882	cd18013	DEXQ-box helicase domain of bacterial SNF2 family proteins	"Proteins belonging to the SNF2 family of DNA dependent ATPases are important members of the chromatin remodeling complexes that are implicated in epigenetic control of gene expression. The Snf2 family comprise a large group of ATP-hydrolyzing proteins that are ubiquitous in eukaryotes, but also present in eubacteria and archaea. The bacterial SNF2 present in this family are members of the DEAD-like helicase superfamily, a diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region. [PMID: 15627498, PMID: 16738128, PMID: 18508648, PMID: 19234488, PMID: 18786849, PMID: 15207721, PMID: 16935875, PMID: 16568948, PMID: 1552844, PMID: 2563148, PMID: 11545728, PMID: 10322435, PMID: 16935882, PMID: 19747077, PMID: 21779027, PMID: 23542735, PMID: 23416748, PMID: 20225155, PMID: 20941364]"
1883	IPR037931	Chromatin complexes subunit BAP18	BAP18 is a component of chromatin complexes such as the MLL1/MLLand NURF. It has been shown to facilitate the recruitment of MLL1 subcomplex and androgen receptor (AR) to androgen-response element (ARE) of AR target genes. BAP18 contains a SANT domain.
1884	cd04643	Two tandem repeats of the cystathionine beta-synthase (CBS pair) domains	"The CBS domain, named after human CBS, is a small domain originally identified in cystathionine beta-synthase and is subsequently found in a wide range of different proteins. CBS domains usually occur in tandem repeats. They associate to form a so-called Bateman domain or a CBS pair based on crystallographic studies in bacteria. The CBS pair was used as a basis for this cd hierarchy since the human CBS proteins can adopt the typical core structure and form an intramolecular CBS pair. The interface between the two CBS domains forms a cleft that is a potential ligand binding site. The CBS pair coexists with a variety of other functional domains and this has been used to help in its classification here. It has been proposed that the CBS domain may play a regulatory role, although its exact function is unknown. Mutations of conserved residues within this domain are

		present in bacteria	associated with a variety of human hereditary diseases, including congenital myotonia, idiopathic generalized epilepsy, hypercalciuric nephrolithiasis, and classic Bartter syndrome (CLC chloride channel family members), Wolff-Parkinson-White syndrome (gamma 2 subunit of AMP-activated protein kinase), retinitis pigmentosa (IMP dehydrogenase-1), and homocystinuria (cystathionine beta-synthase). [PMID: 16275737, PMID: 9020585, PMID: 9106071, PMID: 9590298, PMID: 10200156, PMID: 11230183, PMID: 11524006, PMID: 14722619]"
1885	cd18357	"first BTB (Broad-Complex, Tramtrack and Bric a brac)/POZ (poxvirus and zinc finger) domain found in Rho-related BTB domain-containing protein 3 (RhoBTB3)"	"RhoBTB3 is an atypical Rho family small guanosine triphosphatase (GTPase) and is a member of the RhoBTB subfamily, which is characterized by containing a GTPase domain (in most cases, non-functional) followed by a proline-rich region, tandem BTB domains, and a conserved C-terminal region. The carboxyl terminal extension that harbors two BTB domains is capable of assembling cullin 3-dependent ubiquitin ligase complexes. RhoBTB3 is a Golgi-associated Rho-related ATPase that regulates the S/G2 transition of the cell cycle by targeting cyclin E for ubiquitylation. It is involved in vesicle trafficking and in targeting proteins for degradation in the proteasome. It binds directly to Rab9 GTPase and functions with Rab9 in protein transport from endosomes to the trans Golgi network. It also promotes proteasomal degradation of Hypoxia-inducible factor alpha (HIFalpha) through facilitating hydroxylation and suppresses the Warburg effect. This model corresponds to the first BTB domain. The BTB/POZ domain is a common protein-protein interaction motif of about 100 amino acids. [PMID: 27314390, PMID: 18298893, PMID: 12426103, PMID: 26215701, PMID: 25670879, PMID: 24923387, PMID: 24145166, PMID: 22245496, PMID: 19531346, PMID: 19490898, PMID: 18835386]"
1866	cd06503	"F-type ATP synthase, membrane subunit b"	"Membrane subunit b is a component of the Fo complex of FoF1-ATP synthase. The F-type ATP synthases (FoF1-ATPase) consist of two structural domains: the F1 (assembly factor one) complex containing the soluble catalytic core, and the Fo (oligomycin sensitive factor) complex containing the membrane proton channel, linked together by a central stalk and a peripheral stalk. F1 is composed of alpha (or A), beta (B), gamma (C), delta (D) and epsilon (E) subunits with a stoichiometry of 3:3:1:1:1, while Fo consists of the three subunits a, b, and c (1:2:10-14). An oligomeric ring of 10-14 c subunits (c-ring) make up the Fo rotor. The flux of protons through the ATPase channel (Fo) drives the rotation of the c-ring, which in turn is coupled to the rotation of the F1 complex gamma subunit rotor due to the permanent binding between the gamma and epsilon subunits of F1 and the c-ring of Fo. The F-ATP synthases are primarily found in the inner membranes of eukaryotic mitochondria, in the thylakoid membranes of chloroplasts or in the plasma membranes of bacteria. The F-ATP synthases are the primary producers of ATP, using the proton gradient generated by oxidative phosphorylation (mitochondria) or photosynthesis (chloroplasts). Alternatively, under conditions of low driving force, ATP synthases function as ATPases, thus generating a transmembrane proton or Na(+) gradient at the expense of energy derived from ATP hydrolysis. This group also includes F-ATP synthase that has also been found in the archaea Candidatus Methanoperedens. [PMID: 28001127, PMID: 24650630, PMID: 11533110, PMID: 12370007, PMID: 10580496, PMID: 26460036, PMID: 27791192, PMID: 26439008]"
1887	IPR038738	Nedd8-like ubiquitin	"Nedd8 (also known as Rub1) has a single conserved ubiquitin-like domain that is part of a protein modification pathway similar to that of ubiquitin. Nedd8 modifies a family of molecular scaffold proteins called cullins that are responsible for assembling the ROC1/Rbx1 RING-based E3 ubiquitin ligases, of which several play a direct role in tumorigenesis. This entry also includes Nedd8-like proteins from fungi and plants."
1888	cd01806	ubiquitin-like (Ubl) domain found in	"NEDD8, also termed Neddylin, or RELATED TO UBIQUITIN (RUB/Rub1p) in plant and yeast, is a ubiquitin-like protein that conjugates to nuclear proteins in a manner analogous to ubiquitination and sentrinization. It modifies a family of molecular scaffold proteins called cullins that are responsible for assembling the ROC1/Rbx1 RING-based E3 ubiquitin ligases, of which

		neural precursor cell expressed developmentally down-regulated protein 8 (NEDD8) and similar proteins	several play a direct role in tumorigenesis. NEDD8 deamidation and its inhibition of Cullin-RING ubiquitin ligases (CRLs) activity are responsible for Cycle-inhibiting factor (Cif)/Cif homolog in Burkholderia pseudomallei (CHBP)-induced cytopathic effect. NEDD8 contains a single conserved ubiquitin-like (Ubl) domain with a beta-grasp Ubl fold, a common structure involved in protein-protein interactions. Polyubiquitination, signals for a diverse set of cellular events via different isopeptide linkages formed between the C terminus of one ubiquitin (Ub) and the epsilon-amine of K6, K11, K27, K29, K33, K48, or K63 of a second Ub. Ubl NEDD8, contains many of the same lysines (K6, K11, K27, K33, K48) as Ub, where K27 has an role (other than conjugation) in the mechanism of protein neddylation. [PMID: 24773014, PMID: 3041007, PMID: 9857029, PMID: 15775960, PMID: 17220875, PMID: 18652489, PMID: 18805092, PMID: 20129059, PMID: 23175788, PMID: 15567417, PMID: 14690597, PMID: 9857030, PMID: 15242646, PMID: 12759363, PMID: 12730221, PMID: 12215427, PMID: 11953428, PMID: 11259415, PMID: 10597293, PMID: 9790970, PMID: 9353319, PMID: 10318914, PMID: 9694792, PMID: 22708563, PMID: 25833379, PMID: 18793140]"
1 8 9	IPR001623	DnaJ domain	"The hsp70 chaperone machine performs many diverse roles in the cell, including folding of nascent proteins, translocation of polypeptides across organelle membranes, coordinating responses to stress, and targeting selected proteins for degradation. DnaJ is a member of the hsp40 family of molecular chaperones, which is also called the J-protein family, the members of which regulate the activity of hsp70s. DnaJ (hsp40) binds to dnaK (hsp70) and stimulates its ATPase activity, generating the ADP-bound state of dnaK, which interacts stably with the polypeptide substrate.Structurally, the DnaJ protein consists of an N-terminal conserved domain (called 'J' domain) of about 70 amino acids, a glycine-rich region ('G' domain') of about 30 residues, a central domain containing four repeats of a CXXCXGXG motif ('CRR' domain) and a C-terminal region of 120 to 170 residues.Such a structure is shown in the following schematic representation: +----------+-+------+-+------+----------+ | |J-domain | | Gly-R | |CXXCXGXG |C-terminal | +----------+-+------+-+------+----------+ which has a charged surface that includes basic residues that are essential for interaction with the ATPase domain of hsp70.J-domains are found in many prokaryotic and eukaryotic proteins. In yeast, three J-like proteins have been identified containing regions closely resembling a J-domain, but lacking the conserved HPD motif - these proteins do not appear to act as molecular chaperones."
1 9 0	cd14686	Basic leucine zipper (bZIP) domain of bZIP transcription factors: a DNA-binding and dimerization domain	"Basic leucine zipper (bZIP) factors comprise one of the most important classes of enhancer-type transcription factors. They act in networks of homo and heterodimers in the regulation of a diverse set of cellular processes including cell survival, learning and memory, lipid metabolism, and cancer progression, among others. They also play important roles in responses to stimuli or stress signals such as cytokines, genotoxic agents, or physiological stresses. The bZIP structural motif contains a basic region and a leucine zipper, composed of alpha helices with leucine residues 7 amino acids apart, which stabilize dimerization with a parallel leucine zipper domain. Dimerization of leucine zippers creates a pair of the adjacent basic regions that bind DNA and undergo conformational change. Dimerization occurs in a specific and predictable manner resulting in hundreds of dimers having unique effects on transcription. [PMID: 16731568, PMID: 16580748, PMID: 19519454, PMID: 19150388, PMID: 18424222, PMID: 21454302, PMID: 21438114, PMID: 21257317, PMID: 21061914, PMID: 23661758, PMID: 19880309, PMID: 8504253, PMID: 7780801, PMID: 16410410]"
1 9 1	cd06257	DnaJ domain or J-domain. DnaJ/Hsp40 (heat shock protein 40)	"DnaJ domain or J-domain. DnaJ/Hsp40 (heat shock protein 40) proteins are highly conserved and play crucial roles in protein translation, folding, unfolding, translocation, and degradation. They act primarily by stimulating the ATPase activity of Hsp70s, an important chaperonine family. Hsp40 proteins are characterized by the presence of a J domain, which mediates the interaction with Hsp70. They may contain other domains as well, and the architectures provide a means of classification.

		proteins are hig...	[PMID: 16952052, PMID: 15170475, PMID: 10830586, PMID: 10441399, PMID: 10221986, PMID: 8016869, PMID: 9585179, PMID: 17919282, PMID: 17428722, PMID: 8764403, PMID: 9271376, PMID: 9644977]"
1 9 2	PF002 26	DnaJ domain	"DnaJ domains (J-domains) are associated with hsp70 heat-shock system and it is thought that this domain mediates the interaction. DnaJ-domain is therefore part of a chaperone (protein folding) system. The T-antigens, although not in Prosite are confirmed as DnaJ containing domains from literature[PMID:9271376]."
1 9 3	PS500 76	dnaJ domain profile	"The hsp70 chaperone machine performs many diverse roles in the cell, including folding of nascent proteins, translocation of polypeptides across organelle membranes, coordinating responses to stress, and targeting selected proteins for degradation. DnaJ is a member of the hsp40 family of molecular chaperones, which is also called the J-protein family, the members of which regulate the activity of hsp70s. DnaJ (hsp40) binds to dnaK (hsp70) and stimulates its ATPase activity, generating the ADP-bound state of dnaK, which interacts stably with the polypeptide substrate [1,2]. DnaJ consists of an N-terminal conserved domain (called 'J' domain) of about 70 amino acid residues, a glycine and phenylalanine-rich domain ('G/F' domain), a central cysteine rich domain (CR-type zinc finger) containing four repeats of a CXXCXGXG motif which can coordinate two zinc atom and a C-terminal domain (CTD) [2]. Such a structure is shown in the following schematic representation: +----------+-+--------+----+-----+----------+	J-domain		Gly/Phe-R		CXXCXGXG	CTD	+----------+-+--------+----+-----+----------+ The structures of the 'J' domain and the 'CR' domain have been solved [PMID:8764403,PMID:10891270]. The J domain consists of four helices, the second of which has a charged surface that includes basic residues that are essential for interaction with the ATPase domain of hsp70[PMID:12454054]. The CR-type zinc finger has an overall V-shaped extended beta-hairpin topology and two symmetrical zinc binding sites, designated as Zn1 and Zn2: Zn1 is formed by the two cysteine motifs that are furthest apart in the primary sequence, while Zn2 is formed by the two central, adjacent cysteine motifs[PMID:10891270]. It has been shown that Zn1 is important for the autonomous, dnaK-independent chaperone activity, while Zn2 is a necessary interaction site with dnaK, which seems to be crucial for in vivo function in the dnaJ/dnaK system[PMID:12941935]. J-protein family are classified in three classes [2]: The type 1 contains proteins with a J-domain, a G/F domain, a CR zinc finger and a CTD domain (true homologues of dnaJ): - Yeast protein MAS5/YDJ1 that is involved in protein folding and mitochondrial protein import. - Yeast protein SCJ1 that is involved in protein folding and mitochondrial protein import. - Yeast protein APJ1/YNL077w, that might modulate folding reaction. - Cucumber dnaJ homolog anchored at the glyoxysomal membrane. - Yeast protein XDJ1. - Human protein HDJ2. The type 2 contains proteins with a J-domain, a G/F domain and a CTD domain: - Rhizobium fredii nolC, a protein involved in cultivar-specific nodulation of soybean. - Escherichia coli cbpA, a protein that binds curved DNA. - Yeast protein SIS1, required for nuclear migration during mitosis. - Yeast protein CAJ1. - Yeast hypothetical protein YFR041c. - Yeast hypothetical protein YIR004w. - Yeast hypothetical protein YJL162c. - Plasmodium falciparum ring-infected erythrocyte surface antigen (RESA). RESA, whose function is not known, is associated with the membrane skeleton of newly invaded erythrocytes. - Human HDJ1. - Human HSJ1, a neuronal protein. - Drosophila cysteine-string protein (csp). The type 3 subgroup contains proteins that have only the J-domain: - Yeast protein SEC63/NPL1. It is important for protein assembly into the endoplasmic reticulum and the nucleus. - Eukaryotic Tim14 protein. An essential component of the PAM complex, a complex required for the translocation of transit peptide-containing proteins from the inner membrane into the mitochondrial matrix in an ATP-dependent manner. In the complex, it is required to stimulate activity of mtHSP70 (SSC1). - Yeast Jac1 protein and HscB eukary"
1 9 4	IPR03 6869	Chaperone J-domain superfamily	"The hsp70 chaperone machine performs many diverse roles in the cell, including folding of nascent proteins, translocation of polypeptides across organelle membranes, coordinating responses to stress, and targeting selected proteins for degradation. DnaJ is a member of the hsp40 family of molecular chaperones, which is also called the J-protein family, the members of which

				regulate the activity of hsp70s. DnaJ (hsp40) binds to dnaK (hsp70) and stimulates its ATPase activity, generating the ADP-bound state of dnaK, which interacts stably with the polypeptide substrate.Structurally, the DnaJ protein consists of an N-terminal conserved domain (called 'J' domain) of about 70 amino acids, a glycine-rich region ('G' domain') of about 30 residues, a central domain containing four repeats of a CXXCXGXG motif ('CRR' domain) and a C-terminal region of 120 to 170 residues.Such a structure is shown in the following schematic representation: +----------+-+------+-+----+----------+----------+ | J-domain | | Gly-R | | CXXCXGXG | C-terminal | +----------+-+------+-+----+----------+----------+ +The structure of the J-domain has been solved. The J domain consists of four helices, the second of which has a charged surface that includes basic residues that are essential for interaction with the ATPase domain of hsp70.J-domains are found in many prokaryotic and eukaryotic proteins. In yeast, three J-like proteins have been identified containing regions closely resembling a J-domain, but lacking the conserved HPD motif - these proteins do not appear to act as molecular chaperones."
1 9 5	cd17282	"Type I restriction-modification system specificity (S) subunit Target Recognition Domain-ConseRved domain (TRD-CR), similar to Escherichia coli G4/9 S subunit (S.Eco16444ORF1681P) TRD1-CR1 and Zobellia galactanivorans DsiJT S subunit (S.ZgaJTORF2697P)TRD2-CR2"	"The recognition sequences of Escherichia coli G4/9 S subunit (S.Eco16444ORF1681P) and Zobellia galactanivorans DsiJT S subunit (S.ZgaJTORF2697P) are undetermined. The restriction-modification (RM) system S subunit consists of two variable target recognition domains (TRD1 and 2) and two conserved regions (CR1 and CR2) which separate the TRDs. The TRDs each bind to different specific sequences in the DNA. RM systems protect a bacterial cell against invasion of foreign DNA by endonucleolytic cleavage of DNA that lacks a site specific modification. The host genome is protected from cleavage by methylation of specific nucleotides in the target sites. In type I systems, both restriction and modification activities are present in one heteromeric enzyme complex composed of one DNA specificity (S) subunit (this family), two modification (M) subunits and two restriction (R) subunits. This model contains both TRD1-CR1 and TRD2-CR2. It also includes TRD-CR-like sequence-recognition domains of various putative type II restriction enzymes and methyltransferases and may also include type I DNA methyltransferases. [PMID: 8336674, PMID: 9837717, PMID: 25378308]"
1 9 6	IPR001611	Leucine-rich repeat	"Leucine-rich repeats (LRR) consist of 2-45 motifs of 20-30 amino acids in length that generally folds into an arc or horseshoe shape. LRRs occur in proteins ranging from viruses to eukaryotes, and appear to provide a structural framework for the formation of protein-protein interactions.Proteins containing LRRs include tyrosine kinase receptors, cell-adhesion molecules,	

			virulence factors, and extracellular matrix-binding glycoproteins, and are involved in a variety of biological processes, including signal transduction, cell adhesion, DNA repair, recombination, transcription, RNA processing, disease resistance, apoptosis, and the immune response.Sequence analyses of LRR proteins suggested the existence of several different subfamilies of LRRs. The significance of this classification is that repeats from different subfamilies never occur simultaneously and have most probably evolved independently. It is, however, now clear that all major classes of LRR have curved horseshoe structures with a parallel beta sheet on the concave side and mostly helical elements on the convex side. At least six families of LRR proteins, characterised by different lengths and consensus sequences of the repeats, have been identified. Eleven-residue segments of the LRRs (LxxLxLxxN/CxL), corresponding to the beta-strand and adjacent loop regions, are conserved in LRR proteins, whereas the remaining parts of the repeats (herein termed variable) may be very different. Despite the differences, each of the variable parts contains two half-turns at both ends and a ""linear"" segment (as the chain follows a linear path overall), usually formed by a helix, in the middle. The concave face and the adjacent loops are the most common protein interaction surfaces on LRR proteins. 3D structure of some LRR proteins-ligand complexes show that the concave surface of LRR domain is ideal for interaction with alpha-helix, thus supporting earlier conclusions that the elongated and curved LRR structure provides an outstanding framework for achieving diverse protein-protein interactions. Molecular modeling suggests that the conserved pattern LxxLxL, which is shorter than the previously proposed LxxLxLxxN/CxL is sufficient to impart the characteristic horseshoe curvature to proteins with 20- to 30-residue repeats."
1979	PS51450	Leucine-rich repeat profile	"The leucine-rich repeat (LRR) domain is formed from tandem arrays (2 to 52 repeats) containing a leucine-rich consensus sequence. LRR proteins have been identified in viruses, bacteria, archae and eukaryotes. LRR domains mediate macromolecular interactions in processes as diverse as bacterial invasion of host cells, the plant immune response, and inhibition of RNA binding. All LRR repeats can be divided into a highly conserved segment and a variable segment. The highly conserved segment consists either of an 11-residue stretch, LxxLxLxxNxL, or a 12-residue stretch, LxxLxLxxCxxL. Variable sequences connecting the consensus provide functional diversity for binding [PMID:16231091,PMID:11751054]. Various crystallographic studies of leucine-rich repeats revealed that each LRR contains a beta-strand and an alpha-helix connected by loops (see for example {PDB:2BNH})[PMID:8264799]. Multiple LRRs are arranged so that they form a nonglobular, horseshoe-shaped structure, wherein parallel beta-sheets line the inner circumference of the horseshoe (the concave side) and alpha-helices decorate the outer circumference (the convex side). Although the sequences and numbers of LRRs can differ, diverse LRRs share this overall horseshoe shape. Mutagenesis studies and structural analyses of LRR-ligand complexes have revealed that the concave surface of LRRs is involved mainly in ligand binding[PMID:11751054]. In various types of repeats it has been shown that they fold in a highly cooperative manner. In general folding of an individual repeat is energetically unfavorable, whereas formation of interrepeat interfaces is highly favorable. Thus, folding is initiated only when two adjacent repeats fold forming a favorable interface. Once this activation energy barrier is surpassed, folding proceeds rapidly, via a nucleation-propagation mechanism. In some LRR repeats it has been show that nucleation initiate in the N-terminal capping domain [PMID:18462667,PMID:18462675]. The profile we developed covers the 22 amino acids of the LRR repeat."
1988	cd05676	M20 cytosolic nonspecific dipeptidases including anserinase and serum carnosinase	"Peptidase M20 family, CNDP (cytosolic nonspecific dipeptidase) subfamily including anserinase (Xaa-methyl-His dipeptidase, EC 3.4.13.5), 'serum' carnosinase (beta-alanyl-L-histidine dipeptidase; EC 3.4.13.20), and some uncharacterized proteins. Two genes, CN1 and CN2, coding for proteins that degrade carnosine (beta-alanyl-L-histidine) and homocarnosine (gamma-aminobutyric acid-L-histidine), two naturally occurring dipeptides with potential neuroprotective and neurotransmitter functions, have been identified. CN1 encodes for serum carnosinase and has narrow substrate specificity for Xaa-His dipeptides, where Xaa can be beta-alanine (carnosine), N-methyl beta-alanine, alanine, glycine and gamma-aminobutyric acid (homocarnosine). CN2 corresponds to the cytosolic nonspecific dipeptidase (CNDP; EC 3.4.13.18) and is not limited to Xaa-His

			dipeptides. CNDP requires Mn(2+) for full activity and does not hydrolyze homocarnosine. Anserinase is a dipeptidase that mainly catalyzes the hydrolysis of N-alpha-acetylhistidine. [PMID: 18550540, PMID: 21262300, PMID: 7674922, PMID: 22525515, PMID: 17012794, PMID: 16181134, PMID: 16302965, PMID: 16722645, PMID: 10493853]"
1999	IPR005162	Retrotransposon gag domain	"Transposable elements (TEs) promote various chromosomal rearrangements more efficiently, and often more specifically, than other cellular processes. Retrotransposons are structurally similar to retroviruses and are bounded by long terminal repeats. This entry represents eukaryotic Gag or capsid-related retrotranspon-related proteins. There is a central motif QGXXEXXXXXFXXLXXH that is common to Retroviridae gag-proteins, but is poorly conserved."
2000	PF03732	Retrotransposon gag protein	"Gag or Capsid-like proteins from LTR retrotransposons. There is a central motif QGXXEXXXXXFXXLXXH that is common to Retroviridae gag-proteins, but is poorly conserved[PMID:11600699]."
2001	PS00141	Putative nuclease HARBI1	NA
2002	IPR021109	Aspartic peptidase domain superfamily	"Aspartic peptidases, also known as aspartyl proteases ([intenz:3.4.23.-]), are widely distributed proteolytic enzymesknown to exist in vertebrates, fungi, plants, protozoa, bacteria, archaea, retroviruses and some plant viruses. All known aspartic peptidases are endopeptidases. A water molecule, activated by two aspartic acid residues, acts as the nucleophile in catalysis. Aspartic peptidases can be grouped into five clans, each of which shows a unique structural fold. * Peptidases in clan AA are either bilobed (family A1 or the pepsin family) or are a homodimer (all other families in the clan, including retropepsin from HIV-1/AIDS). Each lobe consists of a single domain with a closed beta-barrel and each lobe contributes one Asp to form the active site. Most peptidases in the clan are inhibited by the naturally occurring small-molecule inhibitor pepstatin. * Clan AC contains the single family A8: the signal peptidase 2 family. Members of the family are found in all bacteria. Signal peptidase 2 processes the premurein precursor, removing the signal peptide. The peptidase has four transmembrane domains and the active site is on the periplasmic side of the cell membrane. Cleavage occurs on the amino side of a cysteine where the thiol group has been substituted by a diacylglyceryl group. Site-directed mutagenesis has identified two essential aspartic acid residues which occur in the motifs GNXXDRX and FNXAD (where X is a hydrophobic residue). No tertiary structures have been solved for any member of the family, but because of the intramembrane location, the structure is assumed not to be pepsin-like. * Clan AD contains two families of transmembrane endopeptidases: A22 and A24. These are also known as ""GXGD peptidases"" because of a common GXGD motif which includes one of the pair of catalytic aspartic acid residues. Structures are known for members of both families and show a unique, common fold with up to nine transmembrane regions. The active site aspartic acids are located within a large cavity in the membrane into which water can gain access. * Clan AE contains two families, A25 and A31. Tertiary structures have been solved for members of both families and show a common fold consisting of an alpha-beta-alpha sandwich, in which the beta sheet is five stranded. * Clan AF contains the single family A26. Members of the clan are membrane-proteins with a unique fold. Homologues are known only from bacteria. The structure of omptin (also known as OmpT) shows a cylindrical barrel containing ten beta strands inserted in the membrane with the active site residues on the outer surface. * There are two families of aspartic peptidases for which neither structure nor active site residues are known and these are not assigned to clans. Family A5 includes thermopsin, an endopeptidase found only in thermophilic archaea. Family A36 contains sporulation factor SpoIIGA, which is known to process and activate sigma factor E, one of the transcription factors that controls sporulation in bacteria.This domain superfamily is found in aspartic peptidases, including pepsin A and other peptidase A1 family members, the yeast DNA-damage inducible protein 1 (Ddi1) (MEROPS peptidase subfamily A28A), animal

			retroviral-like aspartic protease (SASPase; MEROPS peptidase subfamily A28B) and Caulobacter PerP peptidase (MEROPS peptidase family A32). The tertiary structure shows a retropepsin-like fold and the peptidase is active as a homodimer."
2003	IPR001969	"Aspartic peptidase, active site"	"Aspartic peptidases, also known as aspartyl proteases ([intenz:3.4.23.-]), are widely distributed proteolytic enzymesknown to exist in vertebrates, fungi, plants, protozoa, bacteria, archaea, retroviruses and some plant viruses. All known aspartic peptidases are endopeptidases. A water molecule, activated by two aspartic acid residues, acts as the nucleophile in catalysis. Aspartic peptidases can be grouped into five clans, each of which shows a unique structural fold. * Peptidases in clan AA are either bilobed (family A1 or the pepsin family) or are a homodimer (all other families in the clan, including retropepsin from HIV-1/AIDS). Each lobe consists of a single domain with a closed beta-barrel and each lobe contributes one Asp to form the active site. Most peptidases in the clan are inhibited by the naturally occurring small-molecule inhibitor pepstatin. * Clan AC contains the single family A8: the signal peptidase 2 family. Members of the family are found in all bacteria. Signal peptidase 2 processes the premurein precursor, removing the signal peptide. The peptidase has four transmembrane domains and the active site is on the periplasmic side of the cell membrane. Cleavage occurs on the amino side of a cysteine where the thiol group has been substituted by a diacylglyceryl group. Site-directed mutagenesis has identified two essential aspartic acid residues which occur in the motifs GNXXDRX and FNXAD (where X is a hydrophobic residue). No tertiary structures have been solved for any member of the family, but because of the intramembrane location, the structure is assumed not to be pepsin-like. * Clan AD contains two families of transmembrane endopeptidases: A22 and A24. These are also known as ""GXGD peptidases"" because of a common GXGD motif which includes one of the pair of catalytic aspartic acid residues. Structures are known for members of both families and show a unique, common fold with up to nine transmembrane regions. The active site aspartic acids are located within a large cavity in the membrane into which water can gain access. * Clan AE contains two families, A25 and A31. Tertiary structures have been solved for members of both families and show a common fold consisting of an alpha-beta-alpha sandwich, in which the beta sheet is five stranded. * Clan AF contains the single family A26. Members of the clan are membrane-proteins with a unique fold. Homologues are known only from bacteria. The structure of omptin (also known as OmpT) shows a cylindrical barrel containing ten beta strands inserted in the membrane with the active site residues on the outer surface. * There are two families of aspartic peptidases for which neither structure nor active site residues are known and these are not assigned to clans. Family A5 includes thermopsin, an endopeptidase found only in thermophilic archaea. Family A36 contains sporulation factor SpoIIGA, which is known to process and activate sigma factor E, one of the transcription factors that controls sporulation in bacteria.This signature contains the active site residues which are conserved in eukaryotic and viral aspartyl proteases.Aspartic peptidase, also known as aspartyl proteases ([intenz:3.4.23.-]) are a widely distributed family of proteolytic enzymesknown to exist in vertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centred on a catalytic aspartyl residue. The two domains most probably evolved from the duplication of an ancestral gene encoding a primordial domain. Currently known eukaryotic aspartyl proteases are: * Vertebrate gastric pepsins A and C (also known as gastricsin). Vertebrate chymosin (rennin), involved in digestion and used for making cheese. * Vertebrate lysosomal cathepsins D (EC 3.4.23.5) and E (EC 3.4.23.34). * Mammalian renin (EC 3.4.23.15) whose function is to generate angiotensin I from angiotensinogen in the plasma. * Fungal proteases such as aspergillopepsin A (EC 3.4.23.18), candidapepsin (EC 3.4.23.24), mucoropepsin (EC 3.4.23.23) (mucor rennin), endothiapepsin (EC 3.4.23.22), polyporopepsin (EC 3.4.23.29), and rhizopuspepsin (EC 3.4.23.21). * Yeast saccharopepsin (EC 3.4.23.25) (proteinase A) (gene PEP4). PEP4 is implicated in posttranslational regulation of vacuolar hydrolases. * Yeast barrierpepsin (EC 3.4.23.35) (gene BAR1); a protease that cleaves alpha-factor and thus acts as an antagonist of the mating pheromone. * Fission yeast sxa1 which is involved in degrading or processing the mating pheromones."

2004	cd00303	Retropepsins; pepsin-like aspartate proteases	"The family includes pepsin-like aspartate proteases from retroviruses, retrotransposons and retroelements, as well as eukaryotic dna-damage-inducible proteins (DDIs), and bacterial aspartate peptidases. While fungal and mammalian pepsins are bilobal proteins with structurally related N and C-terminals, retropepsins are half as long as their fungal and mammalian counterparts. The monomers are structurally related to one lobe of the pepsin molecule and retropepsins function as homodimers. The active site aspartate occurs within a motif (Asp-Thr/Ser-Gly), as it does in pepsin. Retroviral aspartyl protease is synthesized as part of the POL polyprotein that contains an aspartyl protease, a reverse transcriptase, RNase H, and an integrase. The POL polyprotein undergoes specific enzymatic cleavage to yield the mature proteins. In aspartate peptidases, Asp residues are ligands of an activated water molecule in all examples where catalytic residues have been identified. This group of aspartate peptidases is classified by MEROPS as the peptidase family A2 (retropepsin family, clan AA), subfamily A2A. [PMID: 1851433, PMID: 17199106, PMID: 2194475, PMID: 7674916, PMID: 1455179, PMID: 8841139, PMID: 16970402]"
2005	PS50035	Phospholipase D phosphodiesterase active site profile	"Phospholipase D (PLD) is an ubiquitous enzyme found in bacteria, fungi, plants and mammals. PLD is involved in vesicle formation, protein transport, signal transduction and mitosis. It catalyzes the breakdown of phosphatidylcholine into choline and the putative second messenger phosphatidic acid, which in turn is broken down into two second messengers (diacylglycerol and lysophosphatidic acid). PLD also catalyzes a phosphatidyl transfer reaction using primary alcohols as nucleophilic acceptors to produce phosphatidylalcohols. Sequence analysis revealed that PLD belongs to a superfamily that includes cardiolipin synthases, phosphatidylserine synthase, poxvirus envelope proteins, a Yersinia murine toxin and several endonucleases. All members of this superfamily of phosphodiesterases contain the sequence motif H-x-K-x(4)-D-x(6)-G-S-x-N denoted 'HKD'. Despite the distinct substrate specificities of the superfamily members, the consensus HKD motif appears to be essential for their enzymatic activity. Most of the enzymes in the superfamily contain two copies of this consensus sequence, but the bacterial endonuclease contain only a single copy [PMID:8669123,PMID:8732763,PMID:8755242]. Two HKD motifs are supposed to associate to produce a single active site. It has been proposed that upon substrate binding, one of the histidine residues function as the nucleophile attacking the phosphodiester bond, while the other histidine could serve as a general acid protonating the oxygen atom of the leaving group [PMID:10074947,PMID:10825182]. The profile we developed is centered around the phospholipase D phosphodiesterase active site."
2006	IPR001736	Phospholipase D/Transphosphatidylase	"Phosphatidylcholine-hydrolysing phospholipase D (PLD) isoforms are activated by ADP-ribosylation factors (ARFs). PLD produces phosphatidic acid from phosphatidylcholine, which may be essential for the formation of certain types of transport vesicles or may be constitutive vesicular transport to signal transduction pathways. PC-hydrolysing PLD is a homologue of cardiolipin synthase, phosphatidylserine synthase, bacterial PLDs, and viral proteins. Each of these appears to possess a domain duplication which is apparent by the presence of two motifs containing well-conserved histidine, lysine, and/or asparagine residues which may contribute to the active site aspartic acid. An Escherichia coli endonuclease (nuc) and similar proteins appear to be PLD homologues but possess only one of these motifs."
2007	PS50020	WW/rsp5/WWP domain profile	"The WW domain [PMID:7846762,PMID:7802651,PMID:7828727,PMID:7641887] (also known as rsp5 or WWP) has been originally discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown[PMID:7644498]to bind proteins with particular proline- motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes. Proteins containing the WW domain are listed below. - Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C- terminal globular

			domain. Dystrophin form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats. - Utrophin, a dystrophin-like protein of unknown function. - Vertebrate YAP protein is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains[PMID:7782338]. - Mouse NEDD-4 plays a role in the embryonic development and differentiation of the central nervous system. It contains 3 WW modules followed by a HECT domain. The human ortholog contains 4 WW domains, but the third WW domain is probably spliced resulting in an alternate NEDD-4 protein with only 3 WW modules[PMID:7828727]. - Yeast RSP5 is similar to NEDD-4 in its molecular organization. It contains an N-terminal C2 domain. A related protein, dodo (gene dod) exists in Drosophila and in mammals (gene PIN1). - Tobacco DB10 protein. The WW domain is located N-terminal to the region with similarity to ATP-dependent RNA helicases. - IQGAP, a human GTPase activating protein acting on ras. It contains an N- terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain. - Yeast pre-mRNA processing protein PRP40, Caenorhabditis elegans ZK1098.1 and fission yeast SpAC13C5.02 are related proteins with similarity to MYO2- type myosin, each containing two WW-domains at the N-terminus. - Caenorhabditis elegans hypothetical protein C38D4.5, which contains one WW module, a PH domain and a C-terminal phosphatidylinositol 3-kinase domain. - Yeast hypothetical protein YFL010c. For the sensitive detection of WW domains, we have developed a profile which spans the whole homology region as well as a pattern."
2008	IPR001202	WW domain	"Synonym(s): Rsp5 or WWP domainThe WW domain is a short conserved region in a number of unrelated proteins, which folds as a stable, triple stranded beta-sheet. This short domain of approximately 40 amino acids, may be repeated up to four times in some proteins[cite:PUB00005431],[cite:PUB00000239],[cite:PUB00001680],[cite:PUB00001690]. The name WW or WWP derives from the presence of two signature tryptophan residues that are spaced 20-23 amino acids apart and are present in most WW domains known to date, as well as that of a conserved Pro. The WW domain binds to proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and/or phosphoserine- phosphothreonine-containing motifs. It is frequently associated with other domains typical for proteins in signal transduction processes.A large variety of proteins containing the WW domain are known. These include; dystrophin, a multidomain cytoskeletal protein; utrophin, a dystrophin-like protein of unknown function; vertebrate YAP protein, substrate of an unknown serine kinase; Mus musculus (Mouse) NEDD-4, involved in the embryonic development and differentiation of the central nervous system; Saccharomyces cerevisiae (Baker's yeast) RSP5, similar to NEDD-4 in its molecular organisation; Rattus norvegicus (Rat) FE65, a transcription-factor activator expressed preferentially in liver; Nicotiana tabacum (Common tobacco) DB10 protein, amongst others."
2009	IPR036020	WW domain superfamily	"Synonym(s): Rsp5 or WWP domainThe WW domain is a short conserved region in a number of unrelated proteins, which folds as a stable, triple stranded beta-sheet. This short domain of approximately 40 amino acids, may be repeated up to four times in some proteins[cite:PUB00005431],[cite:PUB00000239],[cite:PUB00001680],[cite:PUB00001690]. The name WW or WWP derives from the presence of two signature tryptophan residues that are spaced 20-23 amino acids apart and are present in most WW domains known to date, as well as that of a conserved Pro. The WW domain binds to proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and/or phosphoserine- phosphothreonine-containing motifs. It is frequently associated with other domains typical for proteins in signal transduction processes.A large variety of proteins containing the WW domain are known. These include; dystrophin, a multidomain cytoskeletal protein; utrophin, a dystrophin-like protein of unknown function; vertebrate YAP protein, substrate of an unknown serine kinase; Mus musculus (Mouse) NEDD-4, involved in the embryonic development and differentiation of the central nervous system; Saccharomyces cerevisiae (Baker's yeast) RSP5, similar to

			NEDD-4 in its molecular organisation; Rattus norvegicus (Rat) FE65, a transcription-factor activator expressed preferentially in liver; Nicotiana tabacum (Common tobacco) DB10 protein, amongst others."
2 1 0	cd002 01	Two conserved tryptophans domain; also known as the WWP or rsp5 domain; aroun...	"Two conserved tryptophans domain; also known as the WWP or rsp5 domain; around 40 amino acids; functions as an interaction module in a diverse set of signalling proteins; binds specific proline-rich sequences but at low affinities compared to other peptide recognition proteins such as antibodies and receptors; WW domains have a single groove formed by a conserved Trp and Tyr which recognizes a pair of residues of the sequence X-Pro; variable loops and neighboring domains confer specificity in this domain; there are five distinct groups based on binding: 1) PPXY motifs 2) the PPLP motif; 3) PGM motifs; 4) PSP or PTP motifs; 5) PR motifs. [PMID: 9401204, PMID: 9407065, PMID: 9305852, PMID: 8757138, PMID: 7828727, PMID: 7782338, PMID: 7802651, PMID: 7846762]"
2 1 1	PF003 97	WW domain	The WW domain is a protein module with two highly conserved tryptophans that binds proline-rich peptide motifs in vitro.
2 1 2	IPR00 1368	TNFR/NGFR cysteine-rich region	"A number of proteins, some of which are known to be receptors for growth factors, have been found to contain a cysteine-rich domain of about 110 to 160 amino acids in their N-terminal part, that can be subdivided into four (or in some cases, three) modules of about 40 residues containing 6 conserved cysteines. Some of the proteins containing this domain are listed below: * Tumor Necrosis Factor type I and type II receptors (TNFR). Both receptors bind TNF-alpha and TNF-beta, but are only similar in the cysteine-rich region. TNFR contains four cysteine-rich domain modules (CRDs), termed CRD1 through CRD4. CRD2 and CRD3 are known as TNF-binding domains. * Shope fibroma virus soluble TNF receptor (protein T2) * Lymphotoxin alpha/beta receptor * Low-affinity nerve growth factor receptor (LA-NGFR) (p75) * CD40 (Bp50), the receptor for the CD40L (or TRAP) cytokine * CD27, the receptor for the CD27L cytokine * CD30, the receptor for the CD30L cytokine * T-cell protein 4-1BB, the receptor for the 4-1BBL putative cytokine * FAS antigen (or APO-1), the receptor for FASL, a protein involved in apoptosis (programmed cell death) * T-cell antigen OX40, the receptor for the OX40L cytokine * Wsl-1, a receptor (for a yet undefined ligand) that mediates apoptosis * Vaccinia virus protein A53 (SalF19R)It has been shownthat the six cysteines all involved in intrachain disulphide bonds. A schematic representation of the structure of the 40 residue module of these receptors is shown below: +----------+ +----------+ | | | xCxxxxxxxxxxxxCxCxxCxxxxxxxxCxxxxCxx | | +----------+'C': conserved cysteine involved in a disulphide bond."
2 1 3	PS006 52	Putative nuclease HARBI1	NA
2 1 4	cd008 90	"Prefoldin is a hexameric molecular chaperone complex, found in both eukaryote..."	"Prefoldin is a hexameric molecular chaperone complex, found in both eukaryotes and archaea, that binds and stabilizes newly synthesized polypeptides allowing them to fold correctly. The complex contains two alpha and four beta subunits, the two subunits being evolutionarily related. In archaea, there is usually only one gene for each subunit while in eukaryotes there two or more paralogous genes encoding each subunit adding heterogeneity to the structure of the hexamer. The structure of the complex consists of a double beta barrel assembly with six protruding coiled-coils. [PMID: 11884745, PMID: 9630229, PMID: 10550210, PMID: 11171552, PMID: 12456645]"

2155	cd17756	DNA replication licensing factor Mcm5	"Mcm5 is a helicase that play an important role in replication. It is part of the heterohexameric ring-shaped Mcm2-7 complex, which is part of the replicative helicase that unwinds parental double-stranded DNA at a replication fork to provide single-stranded DNA templates for the replicative polymerases. [PMID: 8332451, PMID: 28383499, PMID: 22918583, PMID: 26222030]"
2116	cd17059	ubiquitin-like (Ubl) domain found in ubiquitin thioesterase OTU1 and similar proteins	"OTU1 (EC 3.4.19.12), also termed YOD1, or DUBA-8, or HIV-1-induced protease 7 (HIN-7), or OTU domain-containing protein 2 (OTUD2), is a p97-associated deubiquitinylase that functions as a key player in endoplasmic reticulum-associated degradation (ERAD). Its deubiquitinylase activity is also required for negatively regulating cholera toxin A1 (CTA1) retro-translocation. OTU1 contains a conserved ubiquitin-like (Ubl) domain with a beta-grasp Ubl fold, a C2H2-type zinc finger, and an OTU domain. [PMID: 19818707, PMID: 23827681, PMID: 26463207, PMID: 25747718, PMID: 24068323, PMID: 24610782]"
2117	cd17218	Ras-associating (RA) domain found in Ras-association domain-containing protein 1 (RASSF1)	"RASSF1 is a member of a family of six related RASSF1-6 proteins (the classical RASSF proteins). RASSF1 has eight transcripts (A-H) arising from alternative splicing and differential promoter usage. With the exception of some minor splice variants (RASSF1F and RASSF1G), RASSF1 contains an RA domain and a C-terminal SARAH protein-protein interaction motif. The RA domain of the classical RASSF proteins has a beta-grasp ubiquitin-like fold with low sequence similarity to ubiquitin. The RA domain mediates interactions with Ras and other small GTPases, and the SARAH domain mediates protein-protein interactions crucial in the pathways that induce cell cycle arrest and apoptosis. RASSF1A and 1C are the most extensively studied RASSF1 with both localized to microtubules and involved in regulation of growth and migration. [PMID: 10888881, PMID: 11333291, PMID: 12024041, PMID: 11857081, PMID: 15205320, PMID: 14729613, PMID: 14743218, PMID: 15109305, PMID: 15753381, PMID: 21134643, PMID: 21199877, PMID: 18566590, PMID: 16510573, PMID: 15949439, PMID: 17692468]"
2158	IPR025913	Cep57 centrosome localisation domain	"The CLD or centrosome localisation domain of Cep57 is found at the N terminus, and lies approximately between residues 58 and 239. This region lies within the first alpha-helical coiled-coil segment of Cep57, and localises to the centrosome internally to gamma-tubulin, suggesting that it is either on both centrioles or on a centromatrix component. This N-terminal region can also multimerise with the N terminus of other Cep57 molecules."
2119	PF14073	Centrosome localisation domain of Cep57	"The CLD or centrosome localisation domain of Cep57 is found at the N-terminus, and lies approximately between residues 58 and 239. This region lies within the first alpha-helical coiled-coil segment of Cep57, and localises to the centrosome internally to gamma-tubulin, suggesting that it is either on both centrioles or on a centromatrix component. This N-terminal region can also multimerise with the N-terminus of other Cep57 molecules. The C-terminal part, Family Cep57_MT_bd, Pfam:PF06657, is the microtubule-binding region of Cep57."
2220	IPR011010	"DNA breaking-rejoining enzyme, catalytic core"	"Phage integrases are enzymes that mediate unidirectional site-specific recombination between two DNA recognition sequences, the phage attachment site, attP, and the bacterial attachment site, attB. Integrases may be grouped into two major families, the tyrosine recombinases and the serine recombinases, based on their mode of catalysis. Tyrosine family integrases, such as Bacteriophage lambda integrase, utilise a catalytic tyrosine to mediate strand cleavage, tend to recognise longer attP sequences, and require other proteins encoded by the phage or the host bacteria.The 356 amino acid lambda integrase consists of two domains: an N-terminal domain that includes residues 1-64 and is responsible for binding the arm-type sites of attP, and a C-terminal domain (CTD) that binds the lower affinity core-type sites and contains the catalytic site. The CTD can be further divided into the core-type binding domain (residues 65-169) and the catalytic core domain (170-356), the later representing this entry. The catalytic core adopts an alpha3-beta3-alpha4 fold, where one side of the beta sheet is exposed.The recombinases Cre

			from phage P1, XerD from Escherichia coli and Flp from yeast are members of the tyrosine recombinase family, and have a two-domain motif resembling that of lambda integrase, as well as sharing a conserved binding mechanism. The structural fold of their catalytic core domains resemble that of Lambda integrase.The catalytic core of the eukaryotic DNA topoisomerase I shares significant structural similarity with the bacteriophage family of DNA integrases. Topoisomerases I promote the relaxation of DNA superhelical tension by introducing a transient single-stranded break in duplex DNA and are vital for the processes of replication, transcription and recombination."
2 2 1	cd09275	DIRS1 family of RNase HI in long-term repeat retroelements	"Ribonuclease H (RNase H) enzymes are divided into two major families, Type 1 and Type 2, based on amino acid sequence similarities and biochemical properties. RNase H is an endonuclease that cleaves the RNA strand of an RNA/DNA hybrid in a sequence non-specific manner in the presence of divalent cations. RNase H is widely present in various organisms, including bacteria, archaea and eukaryotes. RNase HI has also been observed as adjunct domains to the reverse transcriptase gene in retroviruses, in long-term repeat (LTR)-bearing retrotransposons and non-LTR retrotransposons. RNase HI in LTR retrotransposons perform degradation of the original RNA template, generation of a polypurine tract (the primer for plus-strand DNA synthesis), and final removal of RNA primers from newly synthesized minus and plus strands. The catalytic residues for RNase H enzymatic activity, three aspartatic acids and one glutamic acid residue (DEDD), are unvaried across all RNase H domains. Phylogenetic patterns of RNase HI of LTR retroelements is classified into five major families, Ty3/Gypsy, Ty1/Copia, Bel/Pao, DIRS1 and the vertebrate retroviruses. The structural features of DIRS1-group elements are different from typical LTR elements. RNase H inhibitors have been explored as an anti-HIV drug target because RNase H inactivation inhibits reverse transcription. [PMID: 19228198, PMID: 11435400, PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265]"
2 2 2	IPR013762	"Integrase-like, catalytic domain superfamily"	"Phage integrases are enzymes that mediate unidirectional site-specific recombination between two DNA recognition sequences, the phage attachment site, attP, and the bacterial attachment site, attB. Integrases may be grouped into two major families, the tyrosine recombinases and the serine recombinases, based on their mode of catalysis. Tyrosine family integrases, such as lambda integrase, utilise a catalytic tyrosine to mediate strand cleavage, tend to recognise longer attP sequences, and require other proteins encoded by the phage or the host bacteria.The 356 amino acid lambda integrase consists of two domains: an N-terminal domain that includes residues 1-64 and is responsible for binding the arm-type sites of attP, and a C-terminal domain (CTD) that binds the lower affinity core-type sites and contains the catalytic site. The CTD can be further divided into the core-type binding domain (residues 65-169) and the catalytic core domain (170-356), the later representing this entry. The catalytic core adopts an alpha3-beta3-alpha4 fold, where one side of the beta sheet is exposed.The recombinases Cre from phage P1, XerD from Escherichia coli and Flp from yeast are members of the tyrosine recombinase family, and have a two-domain motif resembling that of lambda integrase, as well as sharing a conserved binding mechanism. The structural fold of their catalytic core domains resemble that of Lambda integrase."
2 2 3	PF13495	"Phage integrase, N-terminal SAM-like domain"	NA
2 2 4	IPR004107	"Integrase, SAM-like, N-terminal"	"Proteins containing this domain cleave DNA substrates by a series of staggered cuts, during which the protein becomes covalently linked to the DNA through a catalytic tyrosine residue at the carboxy end of the alignment.The phage integrase N-terminal SAM-like domain is almost always found with the signature that defines the phage integrase family (see)."

2255	cd03714	RT_DIRS1: Reverse transcriptases (RTs) occurring in the DIRS1 group of retran...	"RT_DIRS1: Reverse transcriptases (RTs) occurring in the DIRS1 group of retransposons. Members of the subfamily include the Dictyostelium DIRS-1, Volvox carteri kangaroo, and Panagrellus redivivus PAT elements. These elements differ from LTR and conventional non-LTR retrotransposons. They contain split direct repeat (SDR) termini, and have been proposed to integrate via double-stranded closed-circle DNA intermediates assisted by an encoded recombinase which is similar to gamma-site-specific integrase. [PMID: 11606703, PMID: 12524337, PMID: 2175908, PMID: 1318387, PMID: 1315955, PMID: 1698615, PMID: 8828137]"
2256	IPR010998	"Integrase/recombinase, N-terminal"	"Phage integrases are enzymes that mediate unidirectional site-specific recombination between two DNA recognition sequences, the phage attachment site, attP, and the bacterial attachment site, attB. Integrases may be grouped into two major families, the tyrosine recombinases and the serine recombinases, based on their mode of catalysis. Tyrosine family integrases, such as Bacteriophage lambda integrase, utilise a catalytic tyrosine to mediate strand cleavage, tend to recognise longer attP sequences, and require other proteins encoded by the phage or the host bacteria.The 356 amino acid lambda integrase consists of two domains: an N-terminal domain (NTD) that includes residues 1 - 64 and is responsible for binding the arm-type sites of attP, and a C-terminal domain that binds the lower affinity core-type sites and contains the catalytic site. The NTD adopts a 4-5 helical bundle fold with two orthogonally packed alpha-hairpins.The recombinases Cre from P1 phage, XerD from Escherichia coli, and Flp from yeast are members of the tyrosine recombinase family, and have a two-domain motif resembling that of lambda integrase, as well as sharing a conserved binding mechanism.The phage integrase NTD is almost always found with the signature that defines the phage integrase family (see)"
2257	IPR024064	FdhE-like superfamily	"This superfamily represents a domain found in formate dehydrogenase accessory protein FdhE and homologues. It consists of an N-terminal five-helical bundle, a tandem repeat of two rubredoxin-like domains and a C-terminal helix packed together."
2258	IPR000727	Target SNARE coiled-coil homology domain	"The process of vesicular fusion with target membranes depends on a set of SNAREs (SNAP-Receptors), which are associated with the fusing membranes. Target SNAREs (t-SNAREs) are localised on the target membrane and belong to two different families, the syntaxin-like family and the SNAP-25 like family. One member of each family, together with a v-SNARE localised on the vesicular membrane, are required for fusion.The N- and C-terminal coiled-coil domains of members of the SNAP-25 family and the most C-terminal coiled-coil domain of the syntaxin family are related to each other and form a new homology domain of approximately 60 amino acids. This domain is also found in other known proteins involved in vesicular membrane traffic, some of which belong to different protein families."
2259	PS50192	t-SNARE coiled-coil homology domain profile	"The process of vesicular membrane fusion in eukaryotic cells depends on a conserved fusion machinery called SNARE (soluble N-ethylmaleimide-sensitive factor (NSF) attachment protein (SNAP) receptors). In the process of vesicle docking, proteins present on the vesicle (v-SNARE) have to bind to their counterpart on the target membrane (t-SNARE) to form a core complex that can then recruit the soluble proteins NSF and SNAP. This so called fusion complex can then disassemble after ATP hydrolysis mediated by the ATPase NSF in a process that leads to membrane fusion and the release of the vesicle contents. t-SNAREs consist of two different families of proteins: the type II integral membrane proteins syntaxins and SNAP-25 (synaptosome-associated protein of 25 kDa), which is anchored in the plasma membrane by attached lipids and does not span the membrane[PMID:9096343]. The N- and C-terminal coiled-coil domains of members of the SNAP-25 family and the most C-terminal coiled-coil domain of the syntaxin family are related to each other and form a new homology domain of approximately 60 amino acids. This domain is also found in other known proteins involved in vesicular membrane traffic, some

			of which belong to different protein families[PMID:9096343]: - Yeast SFT1, UFE1, BET1, VAM7 and YDR468c proteins. - Mammalian TSL-4 proteins. The profile we developed covers the entire t-SNARE coiled-coil homology domain."
2 3 0	cd09279	"RNAse HI family that includes archaeal, some bacterial as well as plant RNase HI"	"Ribonuclease H (RNase H) is classified into two evolutionarily unrelated families, type 1 (prokaryotic RNase HI, eukaryotic RNase H1 and viral RNase H) and type 2 (prokaryotic RNase HII and HIII, and eukaryotic RNase H2). RNase H is an endonuclease that cleaves the RNA strand of an RNA/DNA hybrid in a sequence non-specific manner. RNase H is involved in DNA replication, repair and transcription. RNase H is widely present in various organisms, including bacteria, archaea and eukaryotes and most prokaryotic and eukaryotic genomes contain multiple RNase H genes. Despite the lack of amino acid sequence homology, type 1 and type 2 RNase H share a main-chain fold and steric configurations of the four acidic active-site (DEDD) residues and have the same catalytic mechanism and functions in cells. One of the important functions of RNase H is to remove Okazaki fragments during DNA replication. Most archaeal genomes contain only type 2 RNase H (RNase HII); however, a few contain RNase HI as well. Although archaeal RNase HI sequences conserve the DEDD active-site motif, they lack other common features important for catalytic function, such as the basic protrusion region. Archaeal RNase HI homologs are more closely related to retroviral RNase HI than bacterial and eukaryotic type I RNase H in enzymatic properties. [PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265, PMID: 19228198]"
2 3 1	IPR001509	NAD-dependent epimerase/dehydratase	This domain is found in proteins that utilise NAD as a cofactor and use nucleotide-sugar substrates for a variety of chemical reactions. One of the best studied of these proteins is UDP-galactose 4-epimerase which catalyses the conversion of UDP-galactose to UDP-glucose during galactose metabolism.
2 3 2	IPR005886	UDP-glucose 4-epimerase	"Synonym: UDP-galactose 4-epimeraseUDP-glucose 4-epimerase () interconverts UDP-glucose and UDP-galactose which are precursors of glucose- and galactose-containing exopolysaccharides (EPS). Arabidopsis thaliana has five genes encoding functional UDP-D-glucose/UDP-D-galactose 4-epimerase.A set of related proteins, some of which are tentatively identified as UDP-glucose-4-epimerase in Thermotoga maritima, Bacillus halodurans, and several archaea, but deeply branched from this set and lacking experimental evidence, are not included in this family."
2 3 3	PF01370	NAD dependent epimerase/dehydratase family	This family of proteins utilise NAD as a cofactor. The proteins in this family use nucleotide-sugar substrates for a variety of chemical reactions.
2 3 4	IPR036291	NAD(P)-binding domain superfamily	"This entry represents NAD- and NADP-binding domains superfamily with a core Rossmann-type fold, which consists of three layers (alpha/beta/alpha) with parallel beta-sheet of 6 strands. Many different enzymes contain an NAD/NADP-binding domain, including: * C-terminal domain of alcohol dehydrogenases * Tyrosine-dependent oxidoreductases (also known as short-chain dehydrogenases) * N-terminal domain of glyceraldehyde-3-phosphate dehydrogenase * NAD-binding domain of formate/glycerate dehydrogenases * N-terminal domain of sirohaem synthase * N-terminal domain of lactate dehydrogenase * N-terminal domain of 6-phosphogluconate dehydrogenase (the beta-sheet is extended to 8 strands) * C-terminal domain of amino acid dehydrogenases (an extra N-terminal helix displaces the C-terminal helix * NAD-binding domain of certain potassium channels * C-terminal domain of the transcriptional repressor Rex * Ornithine cyclodeaminase * CoA-binding N-terminal domain of the alpha chain of succinyl-CoA synthetase"
2 3 5	PF05879	Root hair defective 3 GTP-binding	This family consists of several eukaryotic root hair defective 3 like GTP-binding proteins. It has been speculated that the RHD3 protein is a member of a novel class of GTP-binding proteins that is widespread in eukaryotes and required for regulated cell

		protein (RHD3)	enlargement[PMID:9087433]. The family also contains the homologous yeast synthetic enhancement of YOP1 (SEY1) protein which is involved in membrane trafficking[PMID:12427979].
2 3 6	IPR008803	RHD3/Sey1	"This family consists of plant root hair defective 3 (RHD3) protein and its homologues from other eukaryotes. RHD3 is a conserved protein with putative GTP-binding motifs that is implicated in the control of vesicle trafficking between the endoplasmic reticulum and the Golgi compartments. It is required for appropriate root and root hair cells enlargement. It may play a role in cell wall biosynthesis and actin organisation. Its homologue from Saccharomyces cerevisiae, Sey1, is a dynamin-like GTPase that mediates homotypic ER fusion."
2 3 7	PS50957	Josephin domain profile	"The Josephin domain is an eukaryotic protein module of about 180 residues, which occurs in stand-alone form in Josephin-like proteins, and as an amino- terminal domain associated with two or three copies of the ubiquitin- interacting motif (UIM) in ataxin 3-like proteins. Although it has originally been proposed that the Josephin domain could be an all-alpha helical domain distantly related to ENTH and VHS domains involved in membrane trafficking and regulatory adaptor function[PMID:12486728], it is now believed that it is a mainly alpha helical cysteine- protease domain predicted to be active against ubiquitin chains or related substrates [PMID:12944423,PMID:14559776]. The Josephin domain contains two conserved histidines and one cysteine that is required for the ubiquitin protease activity [PMID:12944423,PMID:14559776]. Some proteins known to contain a Josephin domain are listed below: - Animal Machado-Joseph disease protein 1 (Ataxin 3). It interacts with key regulators (CBP, p300 and PCAF) of transcription and represses transcription. - Plant Machado-Joseph disease-like protein (MJD1a-like) (Ataxin 3 homolog). - Mammalian Josephin 1 and 2. - Drosophila melanogaster Josephin-like protein. - Arabidopsis thaliana Josephin-like protein. The profile we developed spans the entire Josephin domain."
2 3 8	IPR010301	"Nucleolar, Nop52"	Nop52 is believed to be involved in the generation of 28S rRNA.
2 3 9	IPR036280	Multiheme cytochrome superfamily	"In proteins belonging to the c-type cytochrome family, the haem group is covalently attached by thioether bonds to two conserved cysteine residues located in the cytochrome c centre. Cytochromes c typically function in electron transfer, but c-type cytochrome centres are also found in the active sites of many enzymes.This domain contains multiple CxxCH motifs. There are a variable number of helices, as well as a little beta structure present in multihaem cytochromes, but they do not form a true structural fold. Cytochrome (cyt) c3-like proteins are multihaem cytochromes, including cyt c3 with four haem groups, cyt c7 (cyt c551.5) with three haem groups (deletion of one cyt c3 haem-binding site), nine-haem cyt c (tandem repeat of four cyt c3-like domains with an additional haem-binding site), and 16-haem cyt c HmcA (tandem repeat of four cyt c3-like domains). The photosynthetic reaction centre composed of a cytochrome subunit is also a multihaem cytochrome. In addition, the di-haem elbow motif shows a similar structure, the main characteristic feature of this motif being the packing of its two haems; many members of this group of proteins contain one or more complete motifs flanked by incomplete motifs and/or other domains. For example, the di-haem elbow motif is present in the multihaem cytochrome domain found in the periplasmic nitrate reductase subunit NapB, in hydroxylamine oxioreductase HAO (contains 3 complete motifs), in cyt c554 (contains one complete motif), in cyt c nitrite reductase, and in the N-terminal domain of flavocytochrome c3 (respiratory fumarate reductase)."
2 4 0	cd09280	Eukaryotic RNase H is essential and is longer and more complex than their	"Ribonuclease H (RNase H) is classified into two families, type 1 (prokaryotic RNase HI, eukaryotic RNase H1 and viral RNase H) and type 2 (prokaryotic RNase HII and HIII, and eukaryotic RNase H2). RNase H is an endonuclease that cleaves the RNA strand of an RNA/DNA hybrid in a sequence non-specific manner. RNase H is involved in DNA replication, repair and transcription. One of the important functions of RNase H is to remove Okazaki fragments during DNA replication. RNase H is widely present in various organisms, including bacteria, archaea and eukaryote and most prokaryotic and eukaryotic genomes contain multiple RNase H genes. Despite the lack of amino acid sequence homology, type 1 and type 2 RNase H share a main-

		prokaryotic counterparts	chain fold and steric configurations of the four acidic active-site (DEDD) residues and have the same catalytic mechanism and functions in cells. Eukaryotic RNase H is longer and more complex than in prokaryotes. Almost all eukaryotic RNase HI have highly conserved regions at their N-termini called hybrid binding domain (HBD). It is speculated that the HBD contributes to binding the RNA/DNA hybrid. Prokaryotes and some single-cell eukaryotes do not require RNase H for viability, but RNase H is essential in higher eukaryotes. RNase H knockout mice lack mitochondrial DNA replication and die as embryos. [PMID: 16093691, PMID: 11083878, PMID: 16601679, PMID: 15989951, PMID: 17964265, PMID: 19228198]"
2 4 1	IPR03 6864	Zn(2)-C6 fungal-type DNA-binding domain superfamily	"The N-terminal region of a number of fungal transcriptional regulatory proteins contains a Cys-rich motif that is involved in zinc-dependent binding of DNA. The region forms a binuclear Zn cluster, in which two Zn atoms are bound by six Cys residues. A wide range of proteins are known to contain this domain. These include the proteins involved in arginine, proline, pyrimidine, quinate, maltose and galactose metabolism; amide and GABA catabolism; leucine biosynthesis, amongst others."
2 4 2	IPR01 6067	"S- adenosylmeth ionine decarboxylase , core"	"S-adenosylmethionine decarboxylase (AdoMetDC)catalyzes the removal of the carboxylate group of S-adenosylmethionine to form S-adenosyl-5'-3-methylpropylamine which then acts as the n-propylamine group donor in the synthesis of the polyamines spermidine and spermine from putrescine.The catalytic mechanism of AdoMetDC involves a covalently-bound pyruvoyl group. This group is post-translationally generated by a self-catalyzed intramolecular proteolytic cleavage reaction between a glutamate and a serine. This cleavage generates two chains, beta (N-terminal) and alpha (C-terminal). The N-terminal serine residue of the alpha chain is then converted by nonhydrolytic serinolysis into a pyruvyol group."
2 4 3	IPR03 7794	Transcription initiation factor TFIID subunit 12	"The transcription regulatory histone acetylation complex Spt-Ada-Gcn5 acetyltransferase (SAGA) is involved in RNA polymerase II-dependent transcriptional regulation of approximately 10% of yeast genes. SAGA preferentially acetylates histones H3 and H2B and deubiquitinates histone H2B. SAGA is known as PCAF in vertebrates and PCAF acetylates nucleosomal histone H3. The SAGA complex consists of at least TRA1, CHD1, SPT7, TAF5, ADA3, SGF73, SPT20/ADA5, SPT8, TAF12, TAF6, HFI1/ADA1, UBP8, GCN5, ADA2, SPT3, SGF29, TAF10, TAF9, SGF11 and SUS1, and some of these components are present as two copies. The complex is built up from distinct modules, each of which has a separate function and crosslinks with either other proteins or other modules in the complex.SLIK (SAGA-like) is a multi-subunit histone acetyltransferase complex that preferentially acetylates histones H3 and H2B and deubiquitinates histone H2B. It is an embellishment of the SAGA complex. The yeast SLIK complex consists of at least TRA1, CHD1, SPT7, CC TAF5, ADA3, SPT20, RTG2, TAF12, TAF6, HFI1, UBP8 (a deubiquitinase), GCN5, ADA2, SPT3, SGF29, TAF10 and TAF9.The yeast SALSA complex is an altered form of the SAGA complex and consists of at least TRA1, SPT7 (C-terminal truncated form), TAF5, ADA3, SPT20, TAF12, TAF6, HFI1, GCN5, ADA2 and SPT3.The DNA-binding general transcription factor complex TFIID is central to the initiation of DNA-dependent RNA polymerase II transcription. TFIID is the only general transcription initiation factor that bind the to the TATA box. The binding of TFIID to the TATA-box is the first step in the formation of a complex able to initiate transcription. TFIID consists of the TATA binding protein (TBP) and 14 TBP-associated factors (TAFs). One copy of each TAF1, TAF2, TAF3, TAF7, TAF8, TAF11, TAF13, two copies of each TAF4, TAF5, TAF6, TAF9, TAF10, TAF12, and three copies of TAF14.Transcription initiation factor TFIID subunit 12 (TAF12) is a component of the DNA-binding general transcription factor complex TFIIDand the transcription regulatory histone acetylation (HAT) complexes SAGA, SALSAand SLIK."
2 4 4	cd049 49	accessory Sec system glycosyltransf erase GtfA	"This family is most closely related to the GT4 family of glycosyltransferases and is named after gtfA in Streptococcus gordonii, where it plays a role in the O-linked glycosylation of GspB, a cell surface glycoprotein involved in platelet binding. In general glycosyltransferases catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds. The acceptor molecule can be a lipid, a protein, a heterocyclic compound, or another carbohydrate

		and similar proteins	residue. This group of glycosyltransferases is most closely related to the previously defined glycosyltransferase family 1 (GT1). The members of this family may transfer UDP, ADP, GDP, or CMP linked sugars. The diverse enzymatic activities among members of this family reflect a wide range of biological functions. The protein structure available for this family has the GTB topology, one of the two protein topologies observed for nucleotide-sugar-dependent glycosyltransferases. GTB proteins have distinct N- and C- terminal domains each containing a typical Rossmann fold. The two domains have high structural homology despite minimal sequence homology. The large cleft that separates the two domains includes the catalytic center and permits a high degree of flexibility. The members of this family are found in bacteria. [PMID: 15489421, PMID: 12691742, PMID: 9004408, PMID: 16564135, PMID: 16037492, PMID: 16263268, PMID: 15653326]"
2 4 5	IPR014012	Helicase/SANT-associated domain	"The helicase/SANT-associated (HSA) domain is a predicted DNA-binding domain of ~75 amino acids, which is found in the eukaryotic SRCAP/p400/DOM and SNF2/brahma families. While each family has the core sequences that define the HSA domain, they each also have additional sequences that distinguish these families from one another. For example, the sequence HWDY(L/C)EEEM(Q/V) is found in the SRCAP/p400/DOM family, whereas the sequence HQE(Y/F)LNSILQ is found in the SNF2 /brahma family. In addition to the SANT and helicase domains, the HSA domain is also found in association with the bromo domain."
2 4 6	PS51204	HSA domain profile	"The helicase/SANT-associated (HSA) domain is a predicted DNA-binding domain of ~75 amino acids[PMID:11779830], which is found in the eukaryotic SRCAP/p400/DOM and SNF2/brahma families[PMID:16024792]. While each family has the core sequences that define the HSA domain, they each also have additional sequences that distinguish these families from one another. For example, the sequence HWDY(L/C)EEEM(Q/V) is found in the SRCAP/p400/DOM family, whereas the sequence HQE(Y/F)LNSILQ is found in the SNF2 /brahma family[PMID:16024792]. In addition to the SANT and helicase domains, the HSA domain is also found in association with bromo[PMID:11779830]. Some proteins known to contain a HSA domain are listed below[PMID:16024792]: - Yeast transcription regulatory protein SWI2/SNF2. - Drosophila brahma. - Mammalian SNF2-alpha (BRM). - Mammalian SNF2-beta (BRG). - Yeast SNF2-related CBP-activating protein (SRCAP), a coactivator for CREB- and nuclear hormone receptor-mediated transcriptional activation. - Drosophila domino (DOM), the homolog of SCRAP. It is essential for organismal viability and has been implicated in several aspects of fly development, including hematopoeisis, wing development, and female fertility. - Mammalian SRCAP. - Mammalian SWI2/SNF2-like ATPase-containing p400/Domino. The profile we developed covers the entire HSA domain."
2 4 7	cd17248	"Type I restriction-modification system specificity (S) subunit Target Recognition Domain-ConseRved domain (TRD-CR), similar to Actinosynne	"The S. AmiI S subunit recognizes 5'... CAGNNNNNNTCGA ... 3'. The restriction-modification (RM) system S subunit consists of two variable target recognition domains (TRD1 and 2) and two conserved regions (CR1 and CR2) which separate the TRDs. The TRDs each bind to different specific sequences in the DNA. For example, S. AmiI -TRD1 recognizes CAG/CTG, and TRD2 recognizes TCGA. RM systems protect a bacterial cell against invasion of foreign DNA by endonucleolytic cleavage of DNA that lacks a site specific modification. The host genome is protected from cleavage by methylation of specific nucleotides in the target sites. In type I systems, both restriction and modification activities are present in one enzyme complex composed of one DNA specificity (S) subunit (this family), two modification (M) subunits and two restriction (R) subunits. This model contains both TRD1-CR1 and TRD2-CR2. It may also include TRD-CR-like sequence-recognition domains of various type II restriction enzymes and methyltransferases and type I DNA methyltransferases. [PMID: 8336674, PMID: 9837717, PMID: 25378308]"

		ma mirum DSM 43827 S subunit (S. AmiI) TRD2-CR2"	
2 4 8	cd01776	Ras-associating (RA) domain of Ras and Rab interactor (Rin) protein family	"Family of Ras-interaction/interference (Rin) proteins, also known as Ras and Rab interactors, is composed of Rin1, Rin2, and Rin3, which have multifunctional domains, including SH2 and proline-rich domains in the N-terminal region, and RH, VPS9, and RA domains in the C-terminal region. RA domain-containing proteins function by interacting with Ras proteins directly or indirectly and are involved in several different functions ranging from tumor suppression to being oncoproteins. Ras proteins are small GTPases that are involved in cellular signal transduction. The RA domain has the beta-grasp ubiquitin-like (Ubl) fold with low sequence similarity to ubiquitin; ubiquitin is a protein modifier in eukaryotes that is involved in various cellular processes including transcriptional regulation, cell cycle control, and DNA repair. The RA domains of Rin1, Rin2, and Rin3 are well conserved and they all have Ras binding characteristics. [PMID: 12574403, PMID: 23765536, PMID: 21209314, PMID: 22249975, PMID: 11784866, PMID: 11703925, PMID: 11305092, PMID: 18796434, PMID: 22976291, PMID: 16757336, PMID: 15886098, PMID: 9208849, PMID: 9144171, PMID: 7862125, PMID: 11733506, PMID: 19631308, PMID: 24449201, PMID: 22890385, PMID: 22825554, PMID: 20424861, PMID: 12972505, PMID: 21586568, PMID: 26296892, PMID: 18486601, PMID: 23185384]"
2 4 9	cd17305	Phosphatidylinositol phosphate kinase (PIPK) catalytic domain found in type II phosphatidylinositol 5-phosphate 4-kinase (PIP5KII) and similar proteins	"PIP5KIIs, also known as PIPKIIs, or PI4P5KIIs, are responsible for the synthesis of phosphatidylinositol-4,5-bisphosphate (PtdIns4,5P2), an essential lipid molecule in various cellular processes, from phosphatidylinositol-5-phosphate (PtdIns5P). Three distinct PIP5KIs have been characterized in erythrocytes, PIP5K2A, PIP5K2B, and PIP5K2C isoforms. [PMID: 8869744, PMID: 11294622, PMID: 16801950, PMID: 18314871, PMID: 18545987, PMID: 19644675, PMID: 24389605, PMID: 25548108, PMID: 7852364, PMID: 7639683, PMID: 9367159, PMID: 18364242, PMID: 20583997, PMID: 19896968, PMID: 9038203, PMID: 9753329, PMID: 27132569, PMID: 25866244, PMID: 24127122, PMID: 26774281, PMID: 27313209, PMID: 25495341, PMID: 26710750, PMID: 22496355, PMID: 23758345]"
2 5 0	PS00018	Putative nuclease HARBI1	NA
2 5 1	IPR018247	"EF-Hand 1, calcium-binding site"	"Many calcium-binding proteins belong to the same evolutionary family and share a type of calcium-binding domain known as the EF-hand. This type of domain consists of a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand). Ca2 binding induces a conformational change in the EF-hand motif, leading to the activation or inactivation of target proteins. EF-hands tend to occur in pairs or higher copy

		numbers.This signature pattern includes the complete EF-hand loop as well as the first residue which follows the loop and which seem to always be hydrophobic. Note: positions 1 (X), 3 (Y) and 12 (-Z) are the most conserved. The 6th residue in an EF-hand loop is, in most cases a Gly, but the number of exceptions to this 'rule' has gradually increased, therefore, this signature pattern includes all the different residues which have been shown to exist in this position in functional Ca-binding sites. The pattern is known, in some cases, to miss one of the EF-hand regions in some proteins with multiple EF-hand domains."

All publications, patent applications, patents, and other publications and references (e.g., sequence database reference numbers) cited herein are incorporated by reference in their entirety. For example, all GenBank, Unigene, and Entrez sequences referred to herein, e.g., in any Table herein, are incorporated by reference. Unless otherwise specified, the sequences specified herein (e.g., by gene name in RepBase or by accession number), including in any Table herein, refer to the database entries current as of March 4, 2020. When one gene or protein references a plurality of sequence accession numbers, all of the sequence variants are encompassed.

EXAMPLES

The invention is further illustrated by the following examples. The examples are provided for illustrative purposes only and are not to be construed as limiting the scope or content of the invention in any way.

Example 1: Delivery of a Gene Writer system to mammalian cells

This example describes a Gene Writer genome editing system delivered to a mammalian cell for the insertion of exogenous DNA into a mammalian cell genome. The Gene Writer component parts are derived from a retrotransposon, e.g., a retrotransposon listed in Table X.

In this example, the polypeptide component of the Gene Writer™ system is the Vingi_1_EE retrotransposase from *Erinaceus europaeus* encoded in a plasmid with expression driven by a CMV promoter. The template component comprises a plasmid with a heterologous object sequence between the 5' UTR and 3' UTR from the Vingi_1_EE retrotransposon, all of which is transcribed by a CMV promoter. The heterologous object sequence is an EF1a-driven GFP reporter cassette that is in antisense orientation relative to the CMV-driven template transcript. Additionally, the GFP gene is interrupted by a forward intron, such that splicing of the RNA only occurs when transcribed from the CMV promoter, while the intron remains present when transcribed in the opposite direction from the EF1a promoter. As a result, the GFP ORF is only intact when transcription first occurs from the CMV promoter and the spliced transcript is

then reverse transcribed and integrated into the genome, allowing subsequent transcription driven by EF1a to generate an mRNA encoding functional GFP.

U2OS cells are nucleofected with the following test agents:

1. DNA coding for the polypeptide described above
2. DNA coding for the polypeptide described above with endonuclease-inactivating mutation
3. DNA coding for the template RNA described above
4. Combination of 1 and 3
5. Combination of 2 and 3

After transfection, U2OS cells are cultured for at least 3 days and then assayed for genomic insertions. For detection of expression, cells are assayed by flow cytometry to measure the levels of GFP, which should only be produced after successful integration events. For a molecular readout, genomic DNA samples isolated from each group of U2OS cells are subjected to unidirectional sequencing to determine the number and identity of insertion sites. To perform unidirectional sequencing, genomic DNA is first extracted and then fragmented. Adapters are ligated to the fragments, and a PCR is performed using an adapter primer and a primer specific to the Gene Writer template. A primer designed to span the intronic junction, which thus binds only to integrated GFP, is used to reduce background reads from the unspliced GFP present in the template plasmid. First round amplicons are then subjected to a second round of PCR to amplify the signal and subsequently analyzed on a MiSeq to determine the genomic locations of successful integration events. Incorporation of unique barcodes into primers permits an estimate of the number of integration events. Retrotransposase-mediated Gene Writing is indicated by the presence of integration events enriched in condition (4) above.

Example 2: Directing a retrotransposon to a specific site in human cells

This example describes a Gene Writer genome editing system delivered to a mammalian cell for the insertion of exogenous DNA into a mammalian cell genome at a pre-determined site. The Gene Writer component parts are derived from a retrotransposon, e.g., a retrotransposon listed in Table X, and additionally comprise a heterologous DNA binding domain.

In this example, the polypeptide component of the Gene Writer™ system is the Vingi_1_EE retrotransposase from *Erinaceus europaeus* encoded in a plasmid with expression driven by a CMV promoter. The Vingi_1_EE retrotransposase additionally includes an N-terminal dCas9 domain. The Gene Writer polypeptide is directed by the addition of a U6-driven gRNA with 20 nt homology to the AAVS1 site, thus the dCas9 domain in combination with the gRNA provides heterologous DNA binding activity to guide the polypeptide to the AAVS1 site for programmed retrotransposition of the template. The template component comprises a plasmid with a heterologous object sequence between the 5' UTR and 3' UTR from the Vingi_1_EE retrotransposon, all of which is transcribed by a CMV promoter. The heterologous object sequence is an EF1a-driven GFP reporter cassette that is in antisense orientation relative to the CMV-driven template transcript. Additionally, the GFP gene is interrupted by a forward intron, such that splicing of the RNA only occurs when transcribed from the CMV promoter, while the intron remains present when transcribed in the opposite direction from the EF1a promoter. As a result, the GFP ORF is only intact when transcription first occurs from the CMV promoter and the spliced transcript is then reverse transcribed and integrated into the genome, allowing subsequent transcription driven by EF1a to generate an mRNA encoding functional GFP.

U2OS cells are nucleofected with the following test agents:

1. DNA coding for the polypeptide and gRNA described above
2. DNA coding for the polypeptide and gRNA described above with endonuclease-inactivating mutation
3. DNA coding for the polypeptide described above, but lacking the gRNA
4. DNA coding for the template RNA described above
5. Combination of 1 and 4
6. Combination of 2 and 4
7. Combination of 3 and 4

After transfection, U2OS cells are cultured for at least 3 days and then assayed for genomic insertions. For detection of expression, cells are assayed by flow cytometry to measure the levels of GFP, which should only be produced after successful integration events. For a molecular readout, genomic DNA samples isolated from each group of U2OS cells are subjected to ddPCR

to determine the frequency of insertions at the AAVS1 site. A primer annealing to the 3' end of the template is paired with a primer that anneals downstream of the AAVS1 insertion site in the genome and a probe located between the primer pair. Thus, amplification events occur when there is an integration at the AAVS1 site and the primers can anneal on each side of the integration junction. By this design, signal in the ddPCR assay, normalized to a standard RPP30 reference, indicates the copy number of integration events at the AAVS1 site. Retrotransposase-mediated Gene Writing is indicated by enrichment of the integration signal in condition (5) above. As compared to condition (7), where the system lacks a gRNA to drive the targeting specificity of the polypeptide, condition (5) demonstrates at least a five-fold increase in integration efficiency at the AAVS1 site as measured by ddPCR.

Example 3: Gene Writer™ enabling nucleotide substitution in genomic DNA to correct alpha-1 antitrypsin deficiency

A Gene Writer system can be used to alter a target DNA, e.g., genomic DNA, by insertion, deletion, or substitution of nucleotides. Without wishing to be limited to a single approach, this example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence at a single nucleotide.

In this example, the Gene Writer™ polypeptide and writing template are provided as DNA transfected into HEK293T cells that possess the PiZ genotype (E342L), a common allele associated with alpha-1 antitrypsin deficiency. The Gene Writer™ polypeptide uses a Cas9 nickase for both DNA-binding and endonuclease functions. The writing template is designed to have homology to the target sequence, while incorporating additional nucleotides at the desired position, such that reverse transcription of the template RNA results in the generation of a new DNA strand containing the substitution.

To create the transversion in the affected human *SERPINA1* gene that restores the GAG triplet coding for glutamate in healthy patients, the Gene Writer™ polypeptide is used with a specific template, which encodes a spacer for polypeptide homing, target homology domain to set up TPRT, and a template sequence for reverse transcription that includes the G substitution.

After transfection, cells are incubated for three days to allow for expression of the Gene Writing™ system and conversion of the genomic DNA target. After the incubation period, genomic DNA is extracted from cells. Genomic DNA is then subjected to PCR-based

amplification using site-specific primers and amplicons are sequenced on an Illumina MiSeq according to manufacturer's protocols. Sequence analysis is then performed to determine the frequency of reads containing the desired edit.

Example 4: Detection of Retrotransposaseon-mediated integration in human cells

This example describes the identification of retrotransposons demonstrating functionality in human cells. By assaying native or modified retrotransposons for integration activity, this example demonstrates a method for the selection of retrotransposases comprising protein domains that can be used to recreate retrotransposases in their native domain composition or as components of chimeric or synthetic Gene Writers for engineering the genome of human cells. For example, a retrotransposon successfully producing an integration signal is expected to comprise functional DNA binding, endonuclease, reverse transcriptase, and, optionally, second-strand synthesis activities. In some embodiments, a reverse transcriptase domain from a retrotransposon that has been shown to demonstrate activity as described in this example is used to provide the reverse transcriptase activity in a Gene Writer polypeptide, e.g., as the RT of a Cas-RT fusion polypeptide, e.g., as disclosed in Table 9. The screen described here employs the nucleofection of a two-plasmid system comprising a retrotransposon polypeptide and an inactivated reporter template into human cells to characterize the RT-dependent retrotransposition efficiency of computationally selected retrotransposons.

In this example, a two-plasmid system was employed comprising: 1) a retrotransposon-encoded protein expression driver plasmid, e.g., a plasmid encoding a retrotransposase polypeptide from Table 3B, comprising a human codon-optimized retrotransposase coding sequence fused with a HiBit tag for detection of protein expression and driven by the mammalian CMV promoter, and 2) a template plasmid, e.g., a plasmid comprising (i) a promoter for expression in mammalian cells to drive transcription of the RNA template molecule, e.g., a CMV promoter, with the template molecule further comprising (ii) a reporter cassette that is inactive in the context of plasmid-derived expression, e.g., an EGFP expression cassette with coding sequence disrupted by an intron encoded in the opposite orientation (GFPai) flanked by (iii) the untranslated regions (UTRs) of the native retrotransposon that naturally comprises the

retrotransposase of (1) (see Figure 8). Here, the GFP reporter is encoded in the absence of a promoter to drive its expression to avoid any loss of signal due to GFP toxicity (see Figure 8).

To deliver the two-plasmid system into U2OS cells, ~400,000 cells were nucleofected with 88.3 ng driver plasmid (1) and 161.7 ng template plasmid (2) using the Lonza SE Cell Line 96-well Nucleofector™ Kit as per manufacturer's instructions. Three days post-nucleofection, integration efficiency was measured using ddPCR to determine the copy number of integrations per genome. Reverse transcription-dependent retrotransposition activity was measured by using a ddPCR approach that utilized the antisense intron as described below. Expression of the driver protein was measured by HiBit-based bioluminescence assay.

When employing an antisense intron reporter containing intronic sequence within the reporter cassette of the template plasmid, e.g., the GFPai system described here, the intron is present in the plasmid but is spliced out during transcription, thus only reporter DNA derived from the transcript by reverse transcription would lack the intron sequence (Figure 8). To limit detection to only events derived from reverse transcription, a ddPCR Taqman probe was designed to span the splicing junction to hybridize to DNA lacking the intron but not to plasmid DNA still containing the intact intron. The forward and reverse primers were designed upstream and downstream of the probe and within the GFP sequence. This design avoids the possible background from template plasmid directly recombined into the genome without a first transcription step, or from intact template plasmid contaminating the gDNA extraction samples.

Gene Writing systems derived from retrotransposases in Table 3B were assayed as following this example to determine activity in human cells. Analysis of the integration efficiency of 163 candidate retrotransposon systems by ddPCR is shown in Figure 9. From the assay described in this example, 25 retrotransposase candidates demonstrated successful trans-integration of the retrotransposon UTR-flanked Template sequence at greater than 0.01 copies/genome on average (see Fig. 10).

In some embodiments, candidate systems are further optimized as described elsewhere in this application. In some embodiments, two driver plasmids are designed for one retrotransposon system based on different consensus approaches (see, e.g., Example 8). In some embodiments, driver plasmids are codon-optimized for human codon preferences. In some embodiments, the RNA template molecule is designed to further comprise sequences flanking the retrotransposon UTRs to facilitate integration, e.g., homology arms comprising sequence identity to the regions

upstream and downstream of the retrotransposon in its endogenous genomic context or to a target DNA sequence in the human genome (see, e.g., Example 9).

Example 5: Consensus genome mapping for retrotransposase amino acid sequence with improved activity in human cells

This example describes the use of genome mining consensus approaches to engineer improved retrotransposon activity. Given a retrotransposon may exist in many copies within a genome, a comparison of all copies could be used to reconstruct an ancestral sequence or one expected to have the highest activity within the host.

In this example, retrotransposons from Table 3B were computationally identified in their host genome. Genomic nucleotide sequences were identified that contained high matches to the termini of each retrotransposon and at least 80% overall nucleotide identity to the complete element. The predicted coding sequence of each pairwise alignment to the target retrotransposon was translated and used to generate a consensus amino acid sequence for the corresponding retrotransposase informed by all copies within the host genome.

To generate constructs for the consensus rederived retrotransposases, each amino acid sequence was back-translated and human optimized and cloned into a driver plasmid as in Example 7. Consensus-generated retrotransposases were then compared with the first-generation retrotransposase sequences for integration activity. For each retrotransposon system, the integration efficiencies driven by consensus-generated and original retrotransposases were compared using the two plasmids system described in Example 7. The same template plasmids were co-nucleofected with one of the two drivers into U2OS cells. ~400,000 cells were nucleofected with 88.3 ng driver plasmid and 161.7 ng template plasmid using the Lonza SE Cell Line 96-well Nucleofector Kit as according to manufacturer's protocols. Genomic DNA was extracted 3 days post nucleofection to quantify the integration efficiency by ddPCR as described in Example 7.

Several retrotransposons showed rescued or improved activity based on the ddPCR integration readout. Specifically, consensus-generated sequences for CR1-3_IS, CR-10_Ami, CR1-54_AAe, and TART-1_DWi only integrated into human cells with the consensus-generated drivers (Figure 11). Additionally, L2-18_ACar and RTE-2_LMi showed markedly improved activity using the consensus-generated sequence. Improved sequences are shown in Table11.

Example 6: Addition of homology arms derived from endogenous integration sites in the native host improves integration of a Gene Writer template in human cells

This example describes the use of genome mining motif approaches to engineer improved Gene Writing activity.

More specifically, retrotransposons from Table 3B were computationally identified in their host genome. Genomic nucleotide sequences were identified that contained high matches to the termini of each native retrotransposon and at least 80% overall nucleotide identity to the complete element. The 3' and 5' neighboring sequences from the host genome were extracted from each identified occurrence of a given element and motifs corresponding to the 3' and 5' ends were generated. Consensus motifs were next aligned to the human genome (assembly hg38) and homology arms were designed with identity to human sequences based on best matches to each motif. Human-specific homology arms used here are indicated in Table 11.

To generate the new template plasmids, the 3' and 5' homology arms, described above, were inserted immediately downstream and upstream of the 3' and 5' retrotransposon UTR sequences, as described in Example 4. Template plasmids comprising or lacking the homology arms were transfected with a matching driver plasmid as described in Example 4 and the performances of the two template versions were compared to characterize any changes in activity. Specifically, ~400,000 U2OS cells were nucleofected with 88.3 ng driver plasmid and 161.7 ng template plasmid using the Lonza SE Cell Line 96-well Nucleofector Kit, as according to manufacturer's protocols. gDNA was extracted 3 days post nucleofection to quantify the integration efficiency of each system by ddPCR, as described in Example 4.

Several Gene Writing templates with added homology arms showed improved or rescued activity based on the ddPCR integration readout. Specifically, homology arm design and addition improved integration for L2-18_ACar, Dong-1_MMa, L2-2_Dre, CR1_AC_1, and RTE-1_Aip (Figure 12). Homology arm addition rescued integration for Tad1-65B_BG compared to a design lacking the homology arm, in which signal had previously been undetectable.

Example 7: Selection of lipid reagents with reduced aldehyde content

In this example, lipids are selected for downstream use in lipid nanoparticle formulations containing Gene Writing component nucleic acid(s), and lipids are selected based at least in part on having an absence or low level of contaminating aldehydes. Reactive aldehyde groups in lipid reagents may cause chemical modifications to component nucleic acid(s), e.g., RNA, e.g., template RNA, during LNP formulation. Thus, in some embodiments, the aldehyde content of lipid reagents is minimized.

Liquid chromatography (LC) coupled with tandem mass spectrometry (MS/MS) can be used to separate, characterize, and quantify the aldehyde content of reagents, e.g., as described in Zurek et al. The Analyst 124(9):1291-1295 (1999), incorporated herein by reference. Here, each lipid reagent is subjected to LC-MS/MS analysis. The LC/MS-MS method first separates the lipid and one or more impurities with a C8 HPLC column and follows with the detection and structural determination of these molecules with the mass spectrometer. If an aldehyde is present in a lipid reagent, it is quantified using a staple-isotope labeled (SIL) standard that is structurally identical to the aldehyde, but is heavier due to C13 and N15 labeling. An appropriate amount of the SIL standard is spiked into the lipid reagent. The mixture is then subjected to LC-MS/MS analysis. The amount of contaminating aldehyde is determined by multiplying the amount of SIL standard and the peak ratio (unknown/SIL). Any identified aldehyde(s) in the lipid reagents is quantified as described. In some embodiments, lipid raw materials selected for LNP formulation are not found to contain any contaminating aldehyde content above a chosen level. In some embodiments, one or more, and optionally all, lipid reagents used for formulation comprise less than 3% total aldehyde content. In some embodiments, one or more, and optionally all, lipid reagents used for formulation comprise less than 0.3% of any single aldehyde species. In some embodiments, one or more, and optionally all, lipid reagents used in formulation comprise less than 0.3% of any single aldehyde species and less than 3% total aldehyde content.

Example 8: Quantification of RNA modification caused by aldehydes during formulation

In this example, the RNA molecules are analyzed post-formulation to determine the extent of any modifications that may have happened during the formulation process, e.g., to detect chemical modifications caused by aldehyde contamination of the lipid reagents (see, e.g., Example 7).

RNA modifications can be detected by analysis of ribonucleosides, e.g., as according to the methods of Su et al. Nature Protocols 9:828-841 (2014), incorporated herein by reference in its entirety. In this process, RNA is digested to a mix of nucleosides, and then subjected to LC-MS/MS analysis. RNA post-formulation is contained in LNPs and must first be separated from lipids by coprecipitating with GlycoBlue in 80% isopropanol. After centrifugation, the pellets containing RNA are carefully transferred to a new Eppendorf tube, to which a cocktail of enzymes (benzonase, Phosphodiesterase type 1, phosphatase) is added to digest the RNA into nucleosides. The Eppendorf tube is placed on a preheated Thermomixer at 37□C for 1 hour. The resulting nucleosides mix is directly analyzed by a LC-MS/MS method that first separates nucleosides and modified nucleosides with a C18 column and then detects them with mass spectrometry.

If aldehyde(s) in lipid reagents have caused chemical modification, data analysis will associate the modified nucleoside(s) with the aldehyde(s). A modified nucleoside can be quantified using a SIL standard which is structurally identical to the native nucleoside except heavier due to C13 and N15 labeling. An appropriate amount of the SIL standard is spiked into the nucleoside digest, which is then subjected to LC-MS/MS analysis. The amount of the modified nucleoside is obtained by multiplying the amount of SIL standard and the peak ratio (unknown/SIL). LC-MS/MS is capable of quantifying all the targeted molecules simultaneously. In some embodiments, the use of lipid reagents with higher contaminating aldehyde content results in higher levels of RNA modification as compared to the use of higher purity lipid reagents as materials during the lipid nanoparticle formulation process. Thus, in preferred embodiments, higher purity lipid reagents are used that result in RNA modification below an acceptable level.

Example 9: Gene Writer™ enabling large insertion into genomic DNA

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence by insertion of a large string of nucleotides.

In this example, the Gene Writer™ polypeptide, gRNA, and writing template are provided as DNA transfected into HEK293T cells. The Gene Writer™ polypeptide uses a Cas9 nickase for both DNA-binding and endonuclease functions. The reverse transcriptase function is

derived from the highly processive RT domain of an R2 retrotransposase. The writing template is designed to have homology to the target sequence, while incorporating the genetic payload at the desired position, such that reverse transcription of the template RNA results in the generation of a new DNA strand containing the desired insertion.

To create a large insertion in the human HEK293T cell DNA, the Gene Writer™ polypeptide is used in conjunction with a specific gRNA, which targets the Cas9-containing Gene Writer™ to the target locus, and a template RNA for reverse transcription, which contains an RT-binding motif (3' UTR from an R2 element) for associating with the reverse transcriptase, a region of homology to the target site for priming reverse transcription, and a genetic payload (GFP expression unit). This complex nicks the target site and then performs TPRT on the template, initiating the reaction by using priming regions on the template that are complementary to the sequence immediately adjacent to the site of the nick and copying the GFP payload into the genomic DNA.

After transfection, cells are incubated for three days to allow for expression of the Gene Writing™ system and conversion of the genomic DNA target. After the incubation period, genomic DNA is extracted from cells. Genomic DNA is then subjected to PCR-based amplification using site-specific primers and amplicons are sequenced on an Illumina MiSeq according to manufacturer's protocols. Sequence analysis is then performed to determine the frequency of reads containing the desired edit.

Example 10: Gene Writers can integrate genetic cargo independently of the single-stranded template repair pathway

This example describes the use of a Gene Writer system in a human cell wherein the single-stranded template repair (SSTR) pathway is inhibited.

In this example, the SSTR pathway will be inhibited using siRNAs against the core components of the pathway: FANCA, FANCD2, FANCE, USP1. Control siRNAs of a non-target control will also be included. 200k U2OS cells will be nucleofected with 30pmols (1.5μM) siRNAs, as well as R2Tg driver and transgene plasmids (trans configuration). Specifically, 250 ng of Plasmids expressing R2Tg, control R2Tg with a mutation in the RT domain, or control R2Tg with an endonuclease inactivating mutation) are used in conjunction with transgene at a

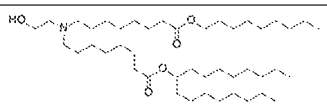
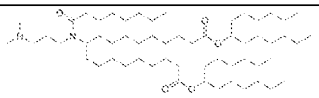
1:4 molar ratio (driver to transgene). Transfections of U2OS cells is performed in SE buffer using program DN100. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3 and ddPCR is performed to assess integration at the rDNA site. Transgene integration at rDNA is detected in the absence of core SSTR pathway components.

Example 11: Formulation of Lipid Nanoparticles encapsulating Firefly Luciferase mRNA

In this example, a reporter mRNA encoding firefly luciferase was formulated into lipid nanoparticles comprising different ionizable lipids. Lipid nanoparticle (LNP) components (ionizable lipid, helper lipid, sterol, PEG) were dissolved in 100% ethanol with the lipid component. These were then prepared at molar ratios of 50:10:38.5:1.5 using ionizable lipid LIPIDV004 or LIPIDV005 (Table A1), DSPC, cholesterol, and DMG-PEG 2000, respectively. Firefly Luciferase mRNA-LNPs containing the ionizable lipid LIPIDV003 (Table A1) were prepared at a molar ratio of 45:9:44:2 using LIPIDV003, DSPC, cholesterol, and DMG-PEG 2000, respectively. Firefly luciferase mRNA used in these formulations was produced by in vitro transcription and encoded the Firefly Luciferase protein, further comprising a 5′ cap, 5′ and 3′ UTRs, and a polyA tail. The mRNA was synthesized under standard conditions for T7 RNA polymerase in vitro transcription with co-transcriptional capping, but with the nucleotide triphosphate UTP 100% substituted with N1-methyl-pseudouridine triphosphate in the reaction. Purified mRNA was dissolved in 25 mM sodium citrate, pH 4 to a concentration of 0.1 mg/mL.

Firefly Luciferase mRNA was formulated into LNPs with a lipid amine to RNA phosphate (N:P) molar ratio of 6. The LNPs were formed by microfluidic mixing of the lipid and RNA solutions using a Precision Nanosystems NanoAssemblr™ Benchtop Instrument, using the manufacturer's recommended settings. A 3:1 ratio of aqueous to organic solvent was maintained during mixing using differential flow rates. After mixing, the LNPs were collected and dialyzed in 15 mM Tris, 5% sucrose buffer at 4°C overnight. The Firefly Luciferase mRNA-LNP formulation was concentrated by centrifugation with Amicon 10 kDa centrifugal filters (Millipore). The resulting mixture was then filtered using a 0.2 μm sterile filter. The final LNP was stored at −80°C until further use.

Table A1: Ionizable Lipids used in Example 11 (Formula (ix), (vii), and (iii))

LIPID ID	Chemical Name	Molecular Weight	Structure
LIPIDV003	(9Z,12Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9, 12-dienoate	852.29	
LIPIDV004	Heptadecan-9-yl 8-((2-hydroxyethyl)(8-(nonyloxy)-8-oxooctyl)amino)octanoate	710.18	
LIPIDV005		919.56	

Prepared LNPs were analyzed for size, uniformity, and %RNA encapsulation. The size and uniformity measurements were performed by dynamic light scattering using a Malvern Zetasizer DLS instrument (Malvern Panalytical). LNPs were diluted in PBS prior to being measured by DLS to determine the average particle size (nanometers, nm) and polydispersity index (pdi). The particle sizes of the Firefly Luciferase mRNA-LNPs are shown in **Table A2**.

Table A2: LNP particle size and uniformity

LNP ID	Ionizable Lipid	Particle Size (nm)	pdi
LNPV019-002	LIPIDV005	77	0.04
LNPV006-006	LIPIDV004	71	0.08
LNPV011-003	LIPIDV003	87	0.08

The percent encapsulation of luciferase mRNA was measured by the fluorescence-based RNA quantification assay Ribogreen (ThermoFisher Scientific). LNP samples were diluted in 1× TE buffer and mixed with the Ribogreen reagent per manufacturer's recommendations and measured on a i3 SpectraMax spectrophotomer (Molecular Devices) using 644 nm excitation and 673 nm emission wavelengths. To determine the percent encapsulation, LNPs were measured using the Ribogreen assay with intact LNPs and disrupted LNPs, where the particles were incubated with 1× TE buffer containing 0.2% (w/w) Triton-X100 to disrupt particles to allow encapsulated RNA to interact with the Ribogreen reagent. The samples were again measured on the i3 SpectraMax spectrophotometer to determine the total amount of RNA present. Total RNA was subtracted from the amount of RNA detected when the LNPs were intact to determine the

fraction encapsulated. Values were multiplied by 100 to determine the percent encapsulation. The Firefly Luciferase mRNA-LNPs that were measured by Ribogreen and the percent RNA encapsulation is reported in Table A3.

Table A3: RNA encapsulation after LNP formulation

LNP ID	Ionizable Lipid	% mRNA encapsulation
LNPV019-002	LIPIDV005	98
LNPV006-006	LIPIDV004	92
LNPV011-003	LIPIDV003	97

Example 12: In vitro activity testing of mRNA-LNPs in Primary Hepatocytes

In this example, LNPs comprising the luciferase reporter mRNA were used to deliver the RNA cargo into cells in culture. Primary mouse or primary human hepatocytes were thawed and plated in collagen-coated 96-well tissue culture plates at a density of 30,000 or 50,000 cells per well, respectively. The cells were plated in 1x William's Media E with no phenol red and incubated at 37˚C with 5% CO_2. After 4 hours, the medium was replaced with maintenance medium (1x William's Media E with no phenol containing Hepatocyte Maintenance Supplement Pack (ThermoFisher Scientific)) and cells were grown overnight at 37˚C with 5% CO_2. Firefly Luciferase mRNA-LNPs were thawed at 4˚C and gently mixed. The LNPs were diluted to the appropriate concentration in maintenance media containing 7.5% fetal bovine serum. The LNPs were incubated at 37˚C for 5 minutes prior to being added to the plated primary hepatocytes. To assess delivery of RNA cargo to cells, LNPs were incubated with primary hepatocytes for 24 hours and cells were then harvested and lysed for a Luciferase activity assay. Briefly, medium was aspirated from each well followed by a wash with 1x PBS. The PBS was aspirated from each well and 200 µL passive lysis buffer (PLB) (Promega) was added back to each well and then placed on a plate shaker for 10 minutes. The lysed cells in PLB were frozen and stored at −80˚C until luciferase activity assay was performed.

To perform the luciferase activity assay, cellular lysates in passive lysis buffer were thawed, transferred to a round bottom 96-well microtiter plate and spun down at 15,000*g* at 4˚C for 3 min to remove cellular debris. The concentration of protein was measured for each sample using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific) according to the manufacturer's instructions. Protein concentrations were used to normalize for cell numbers and

determine appropriate dilutions of lysates for the luciferase assay. The luciferase activity assay was performed in white-walled 96-well microtiter plates using the luciferase assay reagent (Promega) according to manufacturer's instructions and luminescence was measured using an i3X SpectraMax plate reader (Molecular Devices). The results of the dose-response of Firefly luciferase activity mediated by the Firefly mRNA-LNPs are shown in FIG. 13 and indicate successful LNP-mediated delivery of RNA into primary cells in culture. As shown in Fig. 13A, LNPs formulated as according to Example 11 were analyzed for delivery of cargo to primary human (A) and mouse (B) hepatocytes, as according to Example 12. The luciferase assay revealed dose-responsive luciferase activity from cell lysates, indicating successful delivery of RNA to the cells and expression of Firefly luciferase from the mRNA cargo.

Example 13: LNP-mediated delivery of RNA to the mouse liver.

To measure the effectiveness of LNP-mediated delivery of firefly luciferase containing particles to the liver, LNPs were formulated and characterized as described in Example 11 and tested in vitro prior (Example 12) to administration to mice. C57BL/6 male mice (Charles River Labs) at approximately 8 weeks of age were dosed with LNPs via intravenous (i.v.) route at 1 mg/kg. Vehicle control animals were dosed i.v. with 300 µL phosphate buffered saline. Mice were injected via intraperitoneal route with dexamethasone at 5 mg/kg 30 minutes prior to injection of LNPs. Tissues were collected at necropsy at or 6, 24, 48 hours after LNP administration with a group size of 5 mice per time point. Liver and other tissue samples were collected, snap-frozen in liquid nitrogen, and stored at −80°C until analysis.
Frozen liver samples were pulverized on dry ice and transferred to homogenization tubes containing lysing matrix D beads (MP Biomedical). Ice-cold 1x luciferase cell culture lysis reagent (CCLR) (Promega) was added to each tube and the samples were homogenized in a Fast Prep-24 5G Homogenizer (MP Biomedical) at 6 m/s for 40 seconds. The samples were transferred to a clean microcentrifuge tube and clarified by centrifugation. Prior to luciferase activity assay, the protein concentration of liver homogenates was determined for each sample using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific) according to the manufacturer's instructions. Luciferase activity was measured with 200 µg (total protein) of liver homogenate using the luciferase assay reagent (Promega) according to manufacturer's instructions using an i3X SpectraMax plate reader (Molecular Devices). Liver samples revealed

successful delivery of mRNA by all lipid formulations, with reporter activity following the ranking LIPIDV005>LIPIDV004>LIPIDV003 (FIG. 14). As shown in FIG. 14, Firefly luciferase mRNA-containing LNPs were formulated and delivered to mice by iv, and liver samples were harvested and assayed for luciferase activity at 6, 24, and 48 hours post administration. Reporter activity by the various formulations followed the ranking LIPIDV005>LIPIDV004>LIPIDV003. RNA expression was transient and enzyme levels returned near vehicle background by 48 hours. Post-administration. This assay validated the use of these ionizable lipids and their respective formulations for RNA systems for delivery to the liver.

Without wishing to be limited by example, the lipids and formulations described in this example are support the efficacy for the in vivo delivery of other RNA molecules beyond a reporter mRNA. All-RNA Gene Writing systems can be delivered by the formulations described herein. For example, all-RNA systems employing a Gene Writer polypeptide mRNA, Template RNA, and an optional second-nick gRNA are described for editing the genome in vitro by nucleofection, by using modified nucleotides, by lipofection), and editing cells, e.g., primary T cells. As described in this application, these all-RNA systems have many unique advantages in cellular immunogenicity and toxicity, which is of importance when dealing with more sensitive primary cells, especially immune cells, e.g., T cells, as opposed to immortalized cell culture cell lines. Further, it is contemplated that these all RNA systems could be targeted to alternate tissues and cell types using novel lipid delivery systems as referenced herein, e.g., for delivery to the liver, the lungs, muscle, immune cells, and others, given the function of Gene Writing systems has been validated in multiple cell types in vitro here, and the function of other RNA systems delivered with targeted LNPs is known in the art. The in vivo delivery of Gene Writing systems has potential for great impact in many therapeutic areas, e.g., correcting pathogenic mutations), instilling protective variants, and enhancing cells endogenous to the body, e.g., T cells. Given an appropriate formulation, all-RNA Gene Writing is conceived to enable the manufacture of cell-based therapies in situ in the patient.

What is claimed is:

1. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, wherein one or both of (i) or (ii) have an amino acid sequence encoded by a nucleic acid sequence of an element of Table 3B, 10, or 11 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

2. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) an endonuclease domain, wherein one or both of (i) or (ii) have an amino acid sequence encoded by a nucleic acid sequence of an element of Table 3B, 10, or 11 or a sequence having no more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotide differences therefrom; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

3. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid (e.g., DNA or mRNA) encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain and (ii) a target DNA binding domain, wherein one or both of (i) or (ii) have an amino acid sequence encoded by a nucleic acid sequence of an element of Table 3B, 10, or 11 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(b) a template RNA (or DNA encoding the template RNA) comprising (i) a sequence that binds the polypeptide and (ii) a heterologous object sequence.

4. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and

(b) a template RNA (or DNA encoding the template RNA) comprising (e.g., from 5' to 3') (i) optionally a sequence that binds a target site (e.g., a non-edited strand of a site in a target genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' homology domain,

wherein the RT domain has a sequence of Table 3B, 10, or 11, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.