

METHODS AND COMPOSITIONS FOR MODULATING A GENOME RELATED APPLICATIONS

This application claims priority to U.S. Serial No.: 62/985,285 filed Mar 4, 2020, U.S. Serial No.: 63/035,627 filed Jun 5, 2020, and U.S. Serial No.: 63/067,828 filed Aug 19, 2020, the
5 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
10 existing approaches, like CRISPR/Cas9, are more suited for small edits that rely on host repair pathways, 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 compositions (e.g., proteins and nucleic acids) and methods for inserting, altering, or deleting
15 sequences of interest in 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
20 invention features compositions, systems and methods for inserting, altering, or deleting sequences of interest in a host genome.

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

ENUMERATED EMBODIMENTS

25

1. 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 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.

5

2. 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

10 (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:

15 (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.

20 3. 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

25 (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 the RT domain comprises a sequence of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%,
30 96%, 97%, 98%, or 99% identity thereto.

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
- 5 (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 the RT domain comprises a sequence of Table 1 or 3, or a sequence of a reverse
- 10 transcriptase domain of Table 2,
- 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.
5. A system for modifying DNA comprising:
- 15 (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
- 20 genome), (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target 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.
- 25 6. 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
- 30 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 the system is capable of producing an insertion into the target site of at least 1, 2, 3, 4, 5, 10, 20, 30, 40, or 44 nucleotides.

5

7. 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

10 (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 the heterologous object sequence is at least 74, 75, 76, 77, 78, 79, 80, 81, 82, 83,
15 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.

8. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide
20 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
25 sequence, and (iv) a 3' target homology domain,

wherein the heterologous object sequence is at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, or 73 nucleotides in length.

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

10. 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
- 5 (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 the system is capable of producing a deletion into the target site of at least 81,
- 10 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides.
11. 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)
- 15 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,
- 20 wherein the system is capable of producing a deletion into the target site of at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, or 80 nucleotides.
12. A system for modifying DNA comprising:
- (a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide
- 25 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
- 30 sequence, and (iv) a 3' target homology domain,

wherein the system is capable of producing nucleotide substitutions, e.g., transitions and/or transversions, into the target site of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides.

13. A system for modifying DNA comprising:
- 5 (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 second strand of a site in a target genome),
- 10 (ii) optionally a sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target homology domain,
- wherein (a)(ii) and/or (a)(iii) comprises a TAL domain; a zinc finger domain; or a CRISPR/Cas domain chosen from Table 4 or a functional variant (e.g., mutant) thereof.
- 15 14. 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
- 20 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 the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target homology domain,
- wherein the endonuclease domain, e.g., nickase domain, cuts both the first strand and the second strand of the target site DNA, and wherein the cuts are separated from one another by at
- 25 least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or 30 nucleotides.
15. 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)
- 30 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) a sequence that specifically binds the RT domain, (iii) a heterologous object sequence, and (iv) a 3' target homology domain.

5

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

17. A system for modifying DNA comprising:

10 (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

15

(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' target homology domain.

20

18. 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;

25

(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 second strand of a site in a target genome), (e.g., wherein the first RNA comprises a gRNA);

30

(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' target homology domain.

19. The system of any of the preceding embodiments, wherein the second template RNA comprises (i).
20. The system of any of the preceding embodiments, wherein the first template RNA
5 comprises a first conjugating domain and the second template RNA comprises a second conjugating domain.
21. 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,
10 e.g., wherein the stringent conditions for hybridization includes hybridization in 4x sodium chloride/sodium citrate (SSC), at about 65°C, followed by a wash in 1xSSC, at about 65°C.
22. The system of any of the preceding embodiments, wherein the first and second conjugating domains may be joined covalently, e.g., by splint ligation, e.g., by the method
15 described by Moore, M. J., & Query, C. C. Methods in Enzymology, 317, 109–123, 2000.
23. 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.
20
24. The system of any of the preceding embodiments, wherein the reverse transcriptase (RT) domain is from a retrotransposon, or a sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.
25. A system for modifying DNA comprising:
25 (a) a polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase (RT) domain from a retrotransposon, or a sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, (ii) a DNA-binding domain (DBD); and (iii) an endonuclease domain, e.g., a nickase domain; and
30 (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 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.

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

5

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

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

10

29. The system of any of the preceding embodiments, wherein the reverse transcriptase domain comprises an amino acid sequence according to a reverse transcriptase domain of any of Table 30, Table 31, Table 41, Table 44, or Table 50, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a functional fragment thereof.

15

30. 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.

20

31. 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 an endonuclease and/or a DNA-binding domain of a polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target homology domain.

25

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

30

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

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

wherein (i) comprises a nucleic acid sequence with complementarity to a sequence of a gene of any of Tables 9-12 or with no more than 1, 2, 3, 4, or 5 differences from
10 said sequence having said complementarity.

35. 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
15 sequence, and (iv) a 3' target homology domain.

36. 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).

37. The template RNA of any of the preceding embodiments, wherein the RT domain comprises a sequence selected of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2 or a sequence that has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

38. The template RNA of any of the preceding embodiments, wherein the RT domain comprises a sequence selected of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, 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.

39. The template RNA of any of the preceding embodiments, wherein the sequence of (ii) specifically binds the RT domain.
40. 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 1 or from a domain of Table 2, or a sequence having at least 70, 75, 80, 85, 90, 95, or 99% identity thereto.
41. 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.
42. 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.
43. 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.
44. 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).
45. 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).
46. 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).

47. 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).
48. The system or template RNA of any of the preceding embodiments, wherein the sequence
5 that specifically binds the RT domain is disposed between (i) and (iii).
49. 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
10 domain (DBD) of a polypeptide, and (ii) a sequence that binds a target site (e.g., a second 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 heterologous object sequence, and (iii) 3' target homology domain.
15
50. 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.
- 20 51. 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.
52. The system of any of the preceding embodiments, wherein the system is capable of
25 producing an insertion into the target site of at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides.
53. The system of any of the preceding embodiments, 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,
30 94, 95, 96, 97, 98, 99, 100, 120, 140, 160, 180, 200, 500, or 1,000 nts in length.

54. The system of any of the preceding embodiments, 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.

5 55. The system of any of the preceding embodiments, wherein one or both of the template RNA and the RNA encoding the polypeptide of (a) comprises chemically modified mRNA, e.g., mRNA comprising a chemically modified base, e.g., mRNA comprising 5-methoxyuridine.

10 56. The system of any of the preceding embodiments, wherein one or both of the template RNA and the RNA encoding the polypeptide of (a) comprises chemically modified RNA, e.g., RNA comprising a chemically modified base, e.g., RNA comprising 2'-o-methyl phosphorothioate.

15 57. The system of any of the preceding embodiments, wherein one or both of the template RNA and the RNA encoding the polypeptide of (a) comprises chemically modified RNA, e.g., RNA comprising a chemically modified base, e.g., 2'-o-methyl phosphorothioate, at one or both of the 3, 4, or 5 bases at the 5' or 3' end of the RNA.

20 58. 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).

25 59. 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 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

30

60. 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 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, 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.
61. The polypeptide of any of the preceding embodiments, wherein the polypeptide is encoded by an mRNA, e.g., a chemically modified mRNA, e.g., an mRNA comprising a chemically modified base, e.g., an mRNA comprising 5-methoxyuridine.
62. The polypeptide of any of the preceding embodiments, wherein the polypeptide is encoded by an mRNA, e.g., a chemically modified mRNA, e.g., an mRNA comprising a chemically modified base, e.g., an mRNA comprising N1-Methyl-Pseudouridine.
63. 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 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, 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.
64. 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 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, 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; and optionally a DNA-binding domain (DBD) (e.g., a first DBD); and

5 (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.

65. 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.
10

66. 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.

15 67. A reaction mixture comprising:
a cell and any system, polypeptide, template RNA, or DNA encoding the same of any preceding embodiment.

68. A reaction mixture comprising:
20 a DNA comprising a target site and any system, polypeptide, template RNA, or DNA encoding the same of any preceding embodiment.

69. A kit comprising:
the system, polypeptide, template RNA, or DNA encoding the same of any preceding
25 embodiment;
instructions for using the system, polypeptide, template RNA, or DNA encoding the same; and
one or both of a cell or a DNA comprising a target site.

30 70. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the DBD comprises a TAL domain.

71. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the DBD comprises a zinc finger domain.
- 5 72. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the DBD comprises a CRISPR/Cas domain.
73. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the endonuclease domain is a nickase domain.
- 10 74. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the endonuclease domain comprises a CRISPR/Cas domain.
75. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments,
15 wherein the CRISPR/Cas domain comprises a domain or polypeptide from Table 4, or a functional variant (e.g., mutant) thereof.
76. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the CRISPR/Cas domain comprises a domain or polypeptide from genus/species from
20 Table 4.
77. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the endonuclease domain comprises a type IIs nuclease (e.g., FokI), a Holliday Junction resolvase, or a double-stranded DNA nuclease comprising an alteration that abrogates its ability
25 to cut one strand (e.g., transforming the double-stranded DNA nuclease into a nickase).
78. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the RT domain comprises a reverse transcriptase or functional fragment or variant thereof chosen from Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2.
- 30 79. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the RT domain comprises one or more mutations (e.g., an insertion, deletion, or

substitution) relative to a naturally occurring RT domain or an RT domain or functional fragment chosen from Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, or sequence listing SEQ ID NO: 1-67 from WO2018089860A1, incorporated herein by reference.

- 5 80. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the one or more mutations are chosen from D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K or D653N in the RT domain of murine leukemia virus reverse transcriptase or a corresponding mutation at a corresponding position of another RT domain.
- 10 81. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the one or more mutations are chosen from WO2018089860A1, incorporated herein by reference (e.g., a C952S, and/or C956S, and/or C952S, C956S (double mutant), and/or C969S, and/or H970Y, and/or R979Q, and/or R976Q, and/or R1071S, and/or R328A, and/or R329A, 15 and/or Q336A, and/or R328A, R329A, Q336A (triple mutant), and/or G426A, and/or D428A, and/or G426A,D428A (double mutant) mutation, and/or any combination thereof; positions relative to WO2018089860A1 SEQ ID NO: 52), in the RT domain of R2Bm retrotransposase or a corresponding mutation at a corresponding position of another RT domain.
- 20 82. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the DBD and/or the endonuclease domain (e.g., a CRISPR/Cas domain) comprises a domain or polypeptide from Table 4, or a functional variant (e.g., mutant) thereof.
83. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, 25 wherein the DBD and/or the endonuclease domain (e.g., CRISPR/Cas domain) comprises a domain or polypeptide from Table 4.
84. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the RT domain and the DBD and/or the endonuclease domain (e.g., CRISPR/Cas 30 domain) are fused via a peptide linker, e.g., a linker of Table 42.

85. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the linker is about 6-18, 8-16, 10-14, or 12 amino acids in length.
86. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the linker comprises glycine and serine, e.g., wherein the linker comprises solely glycine and serine residues, e.g., wherein the linker comprises a sequence of GSSGSS.
87. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the linker comprises a sequence according Table 42, e.g, linked 10 as disclosed in Table 42 to or a sequence having no more than 1, 2, or 3 substitutions relative thereto.
88. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the CRISPR/Cas domain comprises Cas9, e.g., wild-type Cas9 or nickase Cas9.
89. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the RT domain is positioned C-terminal of the DBD in the polypeptide.
90. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the RT domain is positioned C-terminal of the nickase domain in the polypeptide.
91. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the RT domain is positioned N-terminal of the DBD in the polypeptide.
92. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the RT domain is positioned N-terminal of the nickase domain in the polypeptide.
93. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the polypeptide comprises a linker, e.g., positioned between the RT domain and the DBD or the RT domain and the nickase domain.

94. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the linker is between 2-50, e.g., 2-30, amino acids in length.
95. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments,
5 wherein the linker is a flexible linker, e.g., comprising Gly and/or Ser residues.
96. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the 3' target homology domain is complementary to a sequence adjacent
to a site to be modified by the system, or comprises no more than 1, 2, 3, 4, or 5 mismatches to a
10 sequence complementary to the sequence adjacent to a site to be modified by the system.
97. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the 3' target homology domain is more than 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, 35, 40, 45, 50, 55, 60, 65, 70, 75,
15 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides long, (e.g., 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, or 30
nucleotides long).
98. The system, kit, template RNA, or reaction mixture of any of the preceding
20 embodiments, wherein the 3' target homology domain is no more than 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, 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 long.
99. The system, kit, template RNA, or reaction mixture of any of the preceding
25 embodiments, wherein the heterologous object sequence is complementary to a site to be
modified by the system except at the position or positions to be modified.
100. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the heterologous object sequence is complementary to a site to be
30 modified by the system except at positions encoding a sequence to be inserted to the site.

101. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous object sequence is complementary to a site to be modified by the system except the heterologous object sequence does not comprise nucleotides encoding a sequence to be deleted at the site.

5

102. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous object sequence is more than 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, or 30 nucleotides long, (e.g., 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, or 30 nucleotides long).

10

103. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous object sequence is no more than 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, or 30 nucleotides long.

15

104. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous object sequence substitutes at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides for non-target site nucleotides.

20

105. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous object sequence inserts 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, or 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 into the target site.

25

106. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous object sequence deletes 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, 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides.

30

107. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous object sequence is separated from the sequence that binds the polypeptide (e.g., that binds the endonuclease domain and/or DBD domain) by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 nucleotides.

5

108. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds the polypeptide (e.g., that binds the endonuclease domain and/or DBD domain) is at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, or 130 nucleotides long (and optionally no more than 150, 140, 130, 120, 110, 100, 90, 85, or 80 nucleotides long).

10

109. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds the polypeptide binds the endonuclease domain and/or DBD domain.

15

110. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds the polypeptide comprises a sequence according to one or both of a predicted 5' UTR and a predicted 3' UTR of Table 3 or Table 41, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or functional fragment thereof.

20

111. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds the polypeptide (e.g., that binds the endonuclease domain and/or DBD domain) comprises a gRNA.

25

112. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds a target site (e.g., a second strand of a site in a target genome) is at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, or 130 nucleotides long (and optionally no more than 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, or 20 nucleotides long), e.g., is 17, 18, 19, 20, 21, 22, 23, or 24 nucleotides long.

30

113. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds a target site is complementary to the second strand of the target site, or comprises no more than 1, 2, 3, 4, or 5 mismatches to a sequence complementary to the second strand of the target site.
114. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds a target site (e.g., a second strand of a site in a target genome) is separated from the sequence that binds the polypeptide (e.g., that binds the endonuclease domain and/or DBD domain) by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 nucleotides.
115. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, further comprising a second strand-targeting gRNA that directs the endonuclease domain (e.g., nickase) domain to nick the second strand (e.g., in the target genome).
116. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the template RNA further comprises the second strand-targeting gRNA.
117. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the second strand-targeting gRNA is disposed on a separate nucleic acid from the template RNA.
118. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the gRNA directs the endonuclease domain (e.g., nickase) domain to nick the second strand (e.g., in the target genome) at a site that is 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 nick on the first strand).
119. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the gRNA specifically binds the edited strand.

120. The system, method, kit, template RNA, or reaction mixture 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).
- 5 121. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous targeting domain binds to a different nucleic acid sequence than the unmodified polypeptide.
122. The system, method, kit, template RNA, or reaction mixture of any of the preceding
10 embodiments, wherein the polypeptide does not comprise a functional endogenous targeting domain (e.g., wherein the polypeptide does not comprise an endogenous targeting domain).
123. The system, method, kit, template RNA, or reaction mixture of any of the preceding
15 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).
124. The system, method, kit, template RNA, or reaction mixture 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
20 sequence comprised in the target DNA molecule).
125. The system, method, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the Cas domain is associated with a guide RNA (gRNA).
- 25 126. The system, method, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the heterologous targeting domain comprises an endonuclease domain
(e.g., a heterologous endonuclease domain).
127. The system, method, kit, template RNA, or reaction mixture of any of the preceding
30 embodiments, wherein the endonuclease domain comprises a Cas domain (e.g., a Cas9 or a
mutant or variant thereof).

128. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the Cas domain is associated with a guide RNA (gRNA).
- 5 129. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the endonuclease domain comprises a FokI domain.
130. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the template nucleic acid molecule comprises at least one (e.g., one or
10 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).
131. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein one of the at least one heterologous homology sequences is positioned at
15 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.
132. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein one of the at least one heterologous homology sequences is positioned at
20 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.
133. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous homology sequence binds within 1, 2, 3, 4, 5, 6, 7, 8, 9,
25 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.
134. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous homology sequence has less than 50%, 40%, 30%, 20%,
30 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.

135. The system, method, kit, template RNA, or reaction mixture 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
5 different the sequence bound by an endogenous homology sequence (e.g., replaced by the heterologous homology sequence).
136. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the heterologous homology sequence comprises a sequence (e.g., at its 3'
10 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).
137. The system, method, kit, template RNA, or reaction mixture of any of the preceding
15 embodiments, wherein the heterologous homology sequence comprises a sequence (e.g., at its 5' end) suitable for priming target-primed reverse transcription (TPRT) initiation.
138. The system, method, kit, template RNA, or reaction mixture of any of any of the preceding embodiments, wherein the heterologous homology sequence has at least 85%, 90%,
20 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.
- 25 139. The system, method, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the template nucleic acid molecule comprises a guide RNA (gRNA), e.g., as described herein.
140. The system, method, kit, template RNA, or reaction mixture of any of the preceding
30 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).

141. The system, method, kit, template RNA, or reaction mixture of any of the preceding
5 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.

142. The system, method, kit, template RNA, or reaction mixture of any of the preceding
10 embodiments wherein 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.

143. The system, method, kit, template RNA, or reaction mixture of any of the preceding
15 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
20 embodiments, wherein the miRNA is miR-182 or miR-183, and/or wherein the non-target cell is a dorsal root ganglion neuron.

145. 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
25 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.

146. The system, method, kit, template RNA, or reaction mixture of any of the preceding
30 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.

147. 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
5 different miRNAs.

148. 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
10 sites are recognized by the same or different miRNAs.

149. A system comprising:
an mRNA encoding the polypeptide or system of any of the preceding embodiments, and
a template RNA of any preceding embodiment.

15

150. The system of any of the preceding embodiments, wherein the mRNA encoding the polypeptide or system of any preceding embodiment and the template RNA of any preceding embodiment are disposed on different nucleic acid molecules.

20 151. A system comprising an RNA molecule comprising:
a template RNA (or RNA encoding the template RNA) of any preceding embodiment,
and
a sequence encoding the system or polypeptide of any preceding embodiment.

25 152. The system of any of the preceding embodiments, wherein the RNA molecule comprises an internal ribosome entry site, e.g., operably linked to the sequence encoding the system or polypeptide.

30 153. The system of any of the preceding embodiments, wherein the RNA molecule comprises a cleavage site, e.g., situated between the template RNA (or RNA encoding the template RNA) and the sequence encoding the system or polypeptide.

154. The system or polypeptide of any of the preceding embodiments, wherein the polypeptide comprises a split intein, e.g., two or more (e.g., all) of the RT domain, DBD, endonuclease (e.g., nickase) domain, or combinations thereof are translated as separate proteins which combine into
5 a single polypeptide by protein splicing.
155. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the system comprises one or more circular RNA molecules (circRNAs).
- 10 156. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the circRNA encodes the Gene Writer polypeptide.
157. The system of any of the preceding embodiments, wherein the circRNA comprises a template RNA.
15
158. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein circRNA is delivered to a host cell.
159. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments,
20 wherein the circRNA is capable of being linearized, e.g., in a host cell, e.g., in the nucleus of the host cell.
160. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the circRNA comprises a cleavage site.
25
161. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the circRNA further comprises a second cleavage site.
162. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments,
30 wherein the cleavage site can be cleaved by a ribozyme, e.g., a ribozyme comprised in the circRNA (e.g., by autocleavage).

163. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the circRNA comprises a ribozyme sequence.
- 5 164. The system, kit, polypeptide, or reaction mixture 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.
- 10 165. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the ribozyme is an inducible ribozyme.
- 15 166. The system, kit, polypeptide, or reaction mixture 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.
167. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the ribozyme is a nucleic acid-responsive ribozyme.
- 20 168. The system, kit, polypeptide, or reaction mixture 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).
- 25 169. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the ribozyme is responsive to a target protein (e.g., an MS2 coat protein).
- 30 170. The system, kit, polypeptide, or reaction mixture 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).

171. The system, kit, polypeptide, or reaction mixture 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.

5

172. The system, kit, polypeptide, or reaction mixture 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.

10

173. The system, kit, polypeptide, or reaction mixture 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.

15 174. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the ribozyme is activated by a moiety expressed in a target cell or target tissue.

175. The system, kit, polypeptide, or reaction mixture 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).

20

176. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the ribozyme is comprised in a circular RNA or a linear RNA.

25 177. 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.

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

30

179. 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.

180. A lipid nanoparticle (LNP) comprising the system, template RNA, polypeptide (or RNA encoding the same), or DNA encoding the system, template RNA, or polypeptide, of any preceding embodiment.

181. 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
 10 a second lipid nanoparticle comprising a nucleic acid molecule of a Gene Writing System (e.g., as described herein).

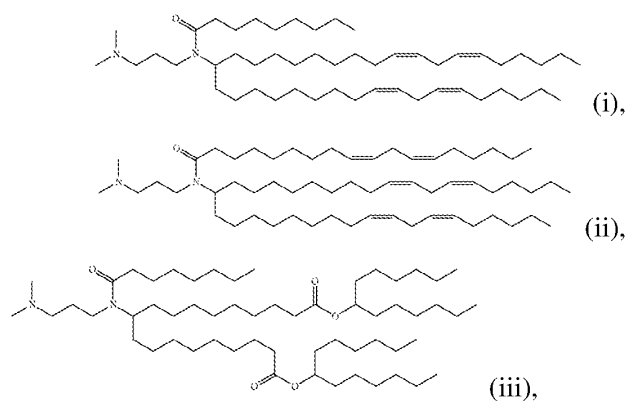
182. The system, kit, polypeptide, or reaction mixture of any preceding embodiments, wherein the system, nucleic acid molecule, polypeptide, and/or DNA encoding the same, is formulated as a lipid nanoparticle (LNP).

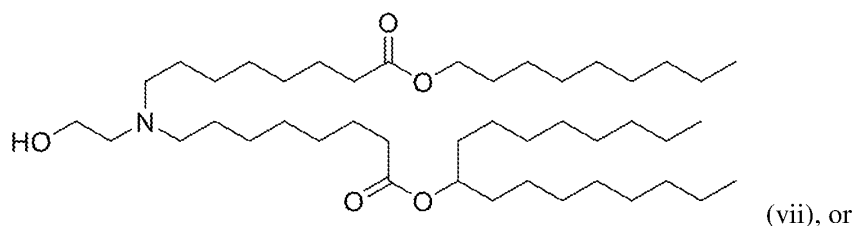
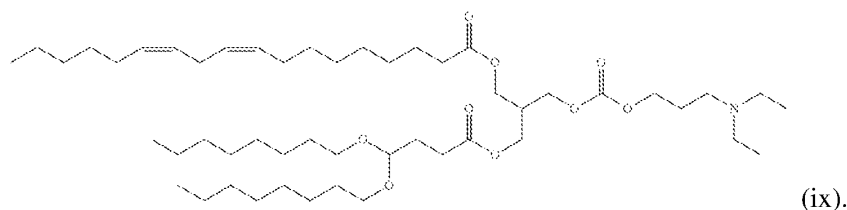
15

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

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

20





185. The LNP of any of the preceding embodiments, further comprising one or more neutral
 5 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.

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

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

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

189. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the lipid nanoparticle is comprised in a formulation comprising a plurality of the lipid nanoparticles.
- 5 190. The system, kit, polypeptide, or reaction mixture 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.
- 10 191. The system, kit, polypeptide, or reaction mixture 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.
- 15 192. The system, kit, polypeptide, or reaction mixture 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.
- 20 193. The system, kit, polypeptide, or reaction mixture 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.
- 25 194. The system, kit, polypeptide, or reaction mixture 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.
- 30 195. The system, kit, polypeptide, or reaction mixture 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.

196. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the lipid nanoparticle formulation comprises less than 3% total reactive impurity (e.g., aldehyde) content.
- 5 197. The system, kit, polypeptide, or reaction mixture 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.
- 10 198. The system, kit, polypeptide, or reaction mixture 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.
- 15 199. The system, kit, polypeptide, or reaction mixture 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.
- 20 200. The system, kit, polypeptide, or reaction mixture 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.
- 25 201. The system, kit, polypeptide, or reaction mixture 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.
- 30 202. The system, kit, polypeptide, or reaction mixture 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.

203. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments,
5 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.

204. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments,
10 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.

205. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments,
15 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 26.

206. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments,
20 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.

207. The system, kit, polypeptide, or reaction mixture of any of the preceding embodiments,
25 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),
30 e.g., in the lipid reagents, e.g., as described in Example 41.

208. The system, kit, polypeptide, or reaction mixture 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 41.

5

209. 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 preceding embodiment.

10 210. 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).

15 211. The system, kit, polypeptide, or reaction mixture of any preceding embodiment, wherein the system, nucleic acid molecule, polypeptide, and/or DNA encoding the same, is formulated as a lipid nanoparticle (LNP).

212. A system comprising:

20 a first lipid nanoparticle comprising the polypeptide (or DNA or RNA encoding the same) of a system or polypeptide of any preceding embodiment; and
a second lipid nanoparticle comprising the template RNA (or DNA encoding the same) of a system or template RNA of any preceding embodiment.

25 213. A virus, viral-like particle, fusosome, or virosome comprising the system, template RNA, polypeptide (or RNA encoding the same), or DNA encoding the system, template RNA, or polypeptide, of any preceding embodiment.

214. A system comprising:

30 a first virus, viral-like particle, fusosome, or virosome comprising the polypeptide (or DNA or RNA encoding the same) of a system or polypeptide of any preceding embodiment; and

a second virus, viral-like particle, or virosome comprising the template RNA (or DNA encoding the same) of a system or template RNA of any preceding embodiment.

215. The system, kit, template RNA, or reaction mixture of any of the preceding
5 embodiments, wherein 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, or at least 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 long (and optionally less than 15, 10, 5, or 20 kilobases long, or less than 500, 400, 300, or 200 nucleotides long).

216. The system, kit, template RNA, or reaction mixture of any of the preceding
10 embodiments, wherein 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, 40, 50, 60, 70, 80, 90, or 100 nucleotides in length).

217. The system, kit, template RNA, or reaction mixture of any of the preceding
15 embodiments, wherein at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains:

a 5' cap, e.g.: a 7-methylguanosine cap (e.g., a O-Me-m7G cap); a hypermethylated cap
20 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)), and/or

a 3' feature selected from one or more of: a polyA tail; a 16-nucleotide long stem-loop
25 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
30 reactive aldehyde followed by conjugation of the aldehyde-reactive modified nucleotide); or chemical ligation to another nucleic acid molecule.

218. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the template RNA 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).
219. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains one or more modified nucleotides.
220. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein 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, or at least 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 long) after a stability test.
221. The system, kit, or reaction mixture of any of the preceding embodiments, wherein at least 1 % of target sites are modified after the system is assayed for potency.
222. The system, kit, template RNA, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the system, polypeptide, template RNA, and/or DNA encoding the same, is formulated as a lipid nanoparticle (LNP).

223. The system, kit, template RNA, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the DNA encoding the system, polypeptide, and/or template RNA are packaged into a virus, viral-like particle, virosome, liposome, vesicle, exosome, or LNP.

224. The system, kit, template RNA, polypeptide, or reaction mixture of any of the preceding
5 embodiments, wherein the DNA encoding the system, template RNA, or polypeptide is packaged into an adeno-associated virus (AAV).

225. The system, kit, template RNA, polypeptide, or reaction mixture of any of the preceding embodiments, wherein the system, template RNA, polypeptide, lipid nanoparticle (LNP), virus,
10 viral-like particle, or virosome is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, and/or host cell protein contamination.

226. A virus, viral-like particle, or virosome comprising:
the system, template RNA, or polypeptide of any of the preceding embodiments, or DNA
15 encoding any of the same, and
an adeno-associated virus (AAV) capsid protein.

227. The system, kit, template RNA, polypeptide, virus, viral-like particle, or virosome of any of the preceding embodiments, wherein the system, template RNA, and/or polypeptide is active
20 in a target tissue and less active (e.g., not active) in a non-target tissue.

228. The system, kit, template RNA, polypeptide, virus, viral-like particle, or virosome of any of the preceding embodiments, further comprising one or more first tissue-specific expression-control sequences specific to the target tissue, wherein the one or more first tissue-specific
25 expression-control sequences specific to the target tissue are in operative association with the template RNA, the polypeptide or nucleic acid encoding the same, or both.

229. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the endonuclease domain, e.g., nickase domain, nicks the first strand of
30 the target site DNA and nicks the second strand at a site a distance from the first nick.

230. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the nicks are made in an outward orientation.
231. The system, kit, template RNA, or reaction mixture of any of the preceding
5 embodiments, wherein the nicks are made in an outward orientation.
232. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments,
10 wherein the sequence that binds a target site specifies the location of the nick to the first strand,
wherein the system further comprises an additional nucleic acid comprising a sequence that binds a site a distance from the target site, and wherein the sequence that binds a site a distance from the target site specifies the location of the nick to the second strand.
- 15 233. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the additional nucleic acid further comprises a sequence that binds the polypeptide (e.g., that binds the endonuclease domain and/or DBD), e.g., wherein the additional nucleic acid comprises a gRNA.
- 20 234. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds a site a distance from the target site (e.g., binds to the first strand of a site in a target genome) is at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, or 130 nucleotides long (and optionally no more than 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 29, 28, 27, 26,
25 25, 24, 23, 22, 21, or 20 nucleotides long), e.g., is 17, 18, 19, 20, 21, 22, 23, or 24 nucleotides long.
235. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds a site a distance from the target site is
30 complementary to the first strand of the target site, or comprises no more than 1, 2, 3, 4, or 5 mismatches to the first strand of the target site.

236. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the DBD and/or endonuclease domain comprise a CRISPR/Cas domain.
237. The system, kit, template RNA, or reaction mixture of any of the preceding
5 embodiments, wherein the CRISPR/Cas domain and the template RNA bind to the target site, and wherein the first strand of the target site comprises a first PAM site.
238. The system, kit, template RNA, or reaction mixture of any of the preceding
10 embodiments, wherein the CRISPR/Cas domain and the additional nucleic acid bind to the site a distance from the target site, and wherein the second strand of the site a distance from the target site comprises a second PAM site.
239. The system, kit, template RNA, or reaction mixture of any of the preceding
15 embodiments, wherein the first PAM site and second PAM site are positioned between the location of the nick to the first strand and the location of the nick to the second strand.
240. The system, kit, template RNA, or reaction mixture of any of the preceding
20 embodiments, wherein the location of the nick to the first strand and the location of the nick to the second strand are positioned between the first PAM site and second PAM site.
241. The system, kit, template RNA, or reaction mixture of any of the preceding
25 embodiments, further comprising an additional polypeptide comprising an additional DNA-binding domain (DBD) and an additional endonuclease domain, e.g., an additional nickase domain.
242. The system, kit, template RNA, or reaction mixture of any of the preceding
30 embodiments, wherein the additional endonuclease domain, e.g., the additional nickase domain, comprises an endonuclease or nickase domain described herein, e.g., a CRISPR/Cas domain, a type II nucleases (e.g., FokI), a Holliday Junction resolvase, a meganuclease, or a double-stranded DNA nuclease comprising an alteration that abrogates its ability to nick one strand (e.g., transforming the double-stranded DNA nuclease into a nickase).

243. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the additional DBD binds a site a distance from the target site.
- 5 244. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the endonuclease domain of (a) or (b) nicks the first strand and the additional endonuclease domain (e.g., additional nickase domain) nicks the second strand.
245. The system, kit, template RNA, or reaction mixture of any of the preceding
10 embodiments, wherein the nicks are made in an outward orientation.
246. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the nicks are made in an inward orientation.
- 15 247. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the DBD and optionally the template RNA (e.g., the sequence that binds the polypeptide) specifies the location of the nick to the first strand, and the additional DBD specifies the location of the nick to the second strand.
- 20 248. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the polypeptide (e.g., the DBD) comprises a TAL effector molecule.
249. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the polypeptide (e.g., the DBD) comprises a zinc finger molecule.
- 25 250. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the polypeptide (e.g., the DBD) comprises a CRISPR/Cas domain.
251. The system, kit, template RNA, or reaction mixture of any of the preceding
30 embodiments, wherein the additional polypeptide (e.g., the additional DBD) comprises a TAL effector molecule.

252. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the additional polypeptide (e.g., the additional DBD) comprises a zinc finger molecule.
- 5 253. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the additional polypeptide (e.g., the additional DBD) comprises a CRISPR/Cas domain.
- 10 254. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the polypeptide and the additional polypeptide bind to sites on the target DNA between the location of the nick to the first strand and the location of the nick to the second.
- 15 255. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the location of the nick to the first strand and the location of the nick to the second strand are between the sites where the polypeptide and the additional polypeptide bind to the target DNA.
- 20 256. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein, on the target DNA, the location of the nick to the second strand is positioned on the opposite side of the binding sites of the polypeptide and additional polypeptide relative to the location of the nick to the first strand.
- 25 257. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein, on the target DNA, the location of the nick to the second strand is positioned on the same side of the binding sites of the polypeptide and additional polypeptide relative to the location of the nick to the first strand.
- 30 258. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the CRISPR/Cas domain of the polypeptide and the template RNA bind to the target site, and wherein the first strand of the target site comprises a PAM site.

259. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the PAM site and the site at a distance from the target site are positioned between the location of the nick to the first strand and the location of the nick to the second strand.

260. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the location of the nick to the first strand and the location of the nick to the second strand are positioned between the PAM site and the site at a distance from the target site.

261. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, further comprising an additional nucleic acid (e.g., a gRNA) comprising a sequence that binds a site a distance from the target site, and wherein the sequence that binds a site a distance from the target site specifies the location of the nick to the second strand.

262. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the additional nucleic acid further comprises a sequence that binds the additional polypeptide (e.g., the CRISPR/Cas domain), e.g., wherein the additional nucleic acid comprises a gRNA.

263. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds a site a distance from the target site (e.g., to the first strand of a site in a target genome) is at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, or 130 nucleotides long (and optionally no more than 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, or 20 nucleotides long), e.g., is 17, 18, 19, 20, 21, 22, 23, or 24 nucleotides long.

264. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the sequence that binds a site a distance from the target site is

complementary to the first strand of the target site, or comprises no more than 1, 2, 3, 4, or 5 mismatches to the first strand of the target site.

265. The system, kit, template RNA, or reaction mixture of any of the preceding
5 embodiments, wherein the site a distance from the target site comprises a PAM site.

266. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the PAM site and the target site are positioned between the location of the
nick to the first strand and the location of the nick to the second strand.
10

267. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the location of the nick to the second strand (e.g., relative to the nick to
the first strand) is such that DNA polymerization by the RT domain proceeds toward the location
of the nick to the second strand.
15

268. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the location of the nick to the second strand (e.g., relative to the nick to
the first strand) is such that DNA polymerization by the RT domain proceeds away from the
location of the nick to the second strand.
20

269. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the first nick and the second nick are at least 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 apart.
25

270. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the first nick and the second nick are no more than 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, 200, or 250
nucleotides apart.
30

271. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the first nick and the second nick are 20-200, 30-200, 40-200, 50-200, 60-200, 70-200, 80-200, 90-200, 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 20-190, 30-190, 40-190, 50-190, 60-190, 70-190, 80-190, 90-190, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 20-180, 30-180, 40-180, 50-180, 60-180, 70-180, 80-180, 90-180, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 170-180, 20-170, 30-170, 40-170, 50-170, 60-170, 70-170, 80-170, 90-170, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 20-160, 30-160, 40-160, 50-160, 60-160, 70-160, 80-160, 90-160, 100-160, 110-160, 120-160, 130-160, 140-160, 150-160, 20-150, 30-150, 40-150, 50-150, 60-150, 70-150, 80-150, 90-150, 100-150, 110-150, 120-150, 130-150, 140-150, 20-140, 30-140, 40-140, 50-140, 60-140, 70-140, 80-140, 90-140, 100-140, 110-140, 120-140, 130-140, 20-130, 30-130, 40-130, 50-130, 60-130, 70-130, 80-130, 90-130, 100-130, 110-130, 120-130, 20-120, 30-120, 40-120, 50-120, 60-120, 70-120, 80-120, 90-120, 100-120, 110-120, 20-110, 30-110, 40-110, 50-110, 60-110, 70-110, 80-110, 90-110, 100-110, 20-100, 30-100, 40-100, 50-100, 60-100, 70-100, 80-100, 90-100, 20-90, 30-90, 40-90, 50-90, 60-90, 70-90, 80-90, 20-80, 30-80, 40-80, 50-80, 60-80, 70-80, 20-70, 30-70, 40-70, 50-70, 60-70, 20-60, 30-60, 40-60, 50-60, 20-50, 30-50, 40-50, 20-40, 30-40, or 20-30 nucleotides apart.

272. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces fewer double-stranded breaks (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein one or more of a PAM site, target site, or site a distance from the target site is not situated between the location of the first strand nick and the location of the second strand nick.

273. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces fewer double-stranded breaks (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein the polypeptide and the additional polypeptide bind to sites on

the target DNA not between the location of the nick to the first strand and the location of the nick to the second.

274. The system, kit, template RNA, or reaction mixture of any of the preceding
5 embodiments, wherein the system produces fewer double-stranded breaks (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein, on the target DNA, the location of the nick to the second strand and the location of the nick to the first strand are located between the binding sites of the polypeptide and additional polypeptide.

10 275. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces fewer double-stranded breaks (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein the location of the nick to the second strand (e.g., relative to
15 the nick to the first strand) is such that the RT domain initiates reverse transcription away from the location of the nick to the second strand.

276. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the system produces fewer deletions not encoded by the heterologous
20 object sequence (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein one or more of a PAM site, target site, or site a distance from the target site is not situated between the location of the first strand nick and the location of the second strand nick, e.g., as measured by PacBio long read sequencing, e.g., as described in Example 29.

25 277. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces fewer deletions (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein the polypeptide and the additional polypeptide bind to sites on the target DNA
30 not between the location of the nick to the first strand and the location of the nick to the second, e.g., as measured by PacBio long read sequencing, e.g., as described in Example 29.

278. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces fewer deletions not encoded by the heterologous object sequence (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein, on the target DNA, the location of the nick to the second strand and the location of the nick to the first strand are located between the binding sites of the polypeptide and additional polypeptide, e.g., as measured by PacBio long read sequencing, e.g., as described in Example 29.

279. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces fewer deletions not encoded by the heterologous object sequence (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein the location of the nick to the second strand (e.g., relative to the nick to the first strand) is such that the RT domain initiates reverse transcription away from the location of the nick to the second strand., e.g., as measured by PacBio long read sequencing, e.g., as described in Example 29.

280. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces fewer insertions not encoded by the heterologous object sequence (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein one or more of a PAM site, target site, or site a distance from the target site is not situated between the location of the first strand nick and the location of the second strand nick, e.g., as measured by PacBio long read sequencing, e.g., as described in Example 29.

281. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces fewer insertions not encoded by the heterologous object sequence (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% fewer) when modifying DNA than an otherwise similar system wherein the polypeptide and the additional polypeptide bind to sites on the target DNA not between the location of the nick to the

first strand and the location of the nick to the second, e.g., as measured by PacBio long read sequencing, e.g., as described in Example 29.

282. The system, kit, template RNA, or reaction mixture of any of the preceding
5 embodiments, wherein the system produces fewer insertions not encoded by the heterologous
object sequence (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1%
fewer) when modifying DNA than an otherwise similar system wherein, on the target DNA, the
location of the nick to the second strand and the location of the nick to the first strand are located
between the binding sites of the polypeptide and additional polypeptide, e.g., as measured by
10 PacBio long read sequencing, e.g., as described in Example 29.

283. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the system produces fewer insertions not encoded by the heterologous
object sequence (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1%
15 fewer) when modifying DNA than an otherwise similar system wherein the location of the nick
to the second strand (e.g., relative to the nick to the first strand) is such that the RT domain
initiates reverse transcription away from the location of the nick to the second strand., e.g., as
measured by PacBio long read sequencing, e.g., as described in Example 29.

20 284. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the system produces more desired Gene Writing modifications (e.g., at
least 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% more)
when modifying DNA than an otherwise similar system wherein one or more of a PAM site,
target site, or site a distance from the target site is not situated between the location of the first
25 strand nick and the location of the second strand nick, e.g., as measured by PacBio long read
sequencing, e.g., as described in Example 29.

285. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the system produces more desired Gene Writing modifications (e.g., at
30 least 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% more)
when modifying DNA than an otherwise similar system wherein the polypeptide and the

additional polypeptide bind to sites on the target DNA not between the location of the nick to the first strand and the location of the nick to the second, e.g., as measured by PacBio long read sequencing, e.g., as described in Example 29.

5 286. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces more desired Gene Writing modifications (e.g., at least 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% more) when modifying DNA than an otherwise similar system wherein, on the target DNA, the location of the nick to the second strand and the location of the nick to the first strand are located between
10 the binding sites of the polypeptide and additional polypeptide, e.g., as measured by PacBio long read sequencing, e.g., as described in Example 29.

287. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the system produces more desired Gene Writing modifications (e.g., at least 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% more)
15 when modifying DNA than an otherwise similar system wherein the location of the nick to the second strand (e.g., relative to the nick to the first strand) is such that the RT domain initiates reverse transcription away from the location of the nick to the second strand., e.g., as measured by PacBio long read sequencing, e.g., as described in Example 29.

20 288. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the first nick and the second nick are at least 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280, 300, 350, 400, 450, or 500 nucleotides apart, e.g., at least 100 nucleotides apart, (and optionally no more than 500, 400, 300, 200, 190, 180, 170, 160, 150, 140, 130, or 120 nucleotides apart).

25 289. The system, kit, template RNA, or reaction mixture of any of the preceding embodiments, wherein the first nick and the second nick are 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 100-180, 110-180, 120-180, 130-180, 140-
30 180, 150-180, 160-180, 170-180, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 100-160, 110-160, 120-160, 130-160, 140-160, 150-160, 100-150, 110-150, 120-150, 130-

150, 140-150, 100-140, 110-140, 120-140, 130-140, 100-130, 110-130, 120-130, 100-120, 110-120, or 100-110 nucleotides apart.

290. The system, kit, template RNA, or reaction mixture of any of the preceding
5 embodiments, wherein the system produces fewer insertions not encoded by the heterologous
object sequence (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1%
fewer) when modifying DNA than an otherwise similar system wherein the location of the nick
to the second strand is less than 100 nucleotides away from the location of the nick to the first
strand (and optionally at least 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides away), e.g., as
10 measured by PacBio long read sequencing, e.g., as described in Example 29.

291. The system, kit, template RNA, or reaction mixture of any of the preceding
embodiments, wherein the system produces fewer deletions not encoded by the heterologous
object sequence (e.g., at least 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1%
15 fewer) when modifying DNA than an otherwise similar system wherein the location of the nick
to the second strand is less than 100 nucleotides away from the location of the nick to the first
strand (and optionally at least 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides away), e.g., as
measured by PacBio long read sequencing, e.g., as described in Example 29.

20 292. 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.

293. A method of making a system for modifying DNA (e.g., as described herein), the method
comprising:
25 (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
30 specifically to a sequence comprised in the target DNA molecule.

294. 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.
295. The method of any of the preceding embodiments, wherein the introducing of (a) comprises inserting the homology sequence into the template nucleic acid.
296. 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.
297. 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.
298. 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.
299. 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.
300. 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.

301. 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.

5

302. 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

10 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,

15 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%,
20 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.

303. A method for manufacturing an template RNA, comprising:

25 (a) providing an template RNA of any preceding embodiment, and

(b) assaying one or more of:

(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

30 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, or 30 nucleotides in length);

5 (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 dihydrouridine, inosine, 7-methylguanosine, 5-methylcytidine (5mC), 5' Phosphate ribothymidine, 2'-O-methyl ribothymidine, 2'-O-ethyl ribothymidine, 2'-fluoro ribothymidine, C-
10 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-
15 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 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
20 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; or

25 (vii) 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 template RNA is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, or host cell protein contamination.

304. A method for manufacturing a system for modifying DNA, comprising:

- 30 (a) providing a system for modifying DNA of any preceding embodiment, and
(b) assaying one or more of:

(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;

5 (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, or 30 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.,
10 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
15 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
20 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%
25 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
30 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, glypyatyon, or lipoylation;

(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.

305. 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,

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

306. A method for modifying a target site in genomic DNA in a cell, the method comprising: contacting the cell with a system, polypeptide, template RNA, or DNA encoding the same of any preceding embodiment, thereby modifying the target site in genomic DNA in a cell.

307. The method of any of the preceding embodiments, wherein a system, polypeptide, template RNA, or DNA are delivered to the target site by electroporation, e.g., nucleofection.

308. The method of any of the preceding embodiments, which does not comprise contacting the cell with DNA, e.g., or which comprises contacting the cell with a composition that not comprise more than 10%, 5%, 4%, 3%, 2%, or 1% DNA by mass or by molar amount.

309. The method of any of the preceding embodiments, which does not comprise contacting the cell with protein, e.g., or which comprises contacting the cell with a composition that not comprise more than 10%, 5%, 4%, 3%, 2%, or 1% protein by mass or by molar amount.

310. The method of any of the preceding embodiments, which comprises contacting a target cell or population of target cells with at least two template RNAs and/or at least two GeneWriter polypeptides, such that at least two target sites (a first target site and a second target site) are modified in a target cell.

311. The method of any of the preceding embodiments, wherein the first target site and the second site are each independently edited at a frequency of at least 5%, 10%, or 15% of copies of the site in a cell population.
- 5 312. The method of any of the preceding embodiments, wherein the first target site and the second site are each independently edited at a frequency of at least 50%, 60%, 70%, or 80% of the level of editing obtained in an otherwise similar cell population contacted with an otherwise similar system targeting only one of the target sites.
- 10 313. The method of any of the preceding embodiments, wherein the resulting cell population comprises no more than 5%, 10%, or 20% unwanted indels compared to the unwanted indels obtained in an otherwise similar cell population contacted with an otherwise similar system targeting only one of the target sites.
- 15 314. The method of any of the preceding embodiments, wherein the cell is a primary cell.
315. The method of any of the preceding embodiments, wherein the cell is a T cell.
- 20 316. A method for modifying a target site in genomic DNA in a cell, the method comprising:
contacting the cell, e.g., by nucleofection or lipid particle delivery, with:
- 25 (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
- 30 (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,
- 30 thereby modifying the target site in genomic DNA in a cell,

wherein the cell is euploid, is not immortalized, is part of a tissue, is part of an organism, is a primary cell, is non-dividing, is haploid (e.g., a germline cell), is a non-cancerous polyploid cell, or is from a subject having a genetic disease.

5 317. The method of any of the preceding embodiments, wherein the template RNA comprises (i).

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

10

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

320. A method for treating a subject having a disease or condition associated with a genetic defect, the method comprising:

15 administering to the subject:

(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;
20 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
25 homology domain,

thereby treating the subject having a disease or condition associated with a genetic defect.

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

30

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

323. The method of any of the preceding embodiments, wherein the template RNA comprises
5 (i) and (ii).

324. A method for treating a subject having a disease or condition associated with a genetic defect, the method comprising:

administering to the subject a system, polypeptide, template RNA, or DNA encoding the
10 same of any preceding embodiment,
thereby treating the subject having a disease or condition associated with a genetic defect.

325. The method of any of the preceding embodiments, wherein the disease or condition associated with a genetic defect is an indication listed in any of Tables 9-12, and/or wherein the
15 genetic defect is a defect in a gene listed in any of Tables 9-12.

326. The method of any of the preceding embodiments, wherein the subject is a human patient.

20 *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
25 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
30 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.

First/Second Strand: As used herein, first strand and second strand, as used to describe the individual DNA strands of target DNA, distinguish the two DNA strands based upon which strand the reverse transcriptase domain initiates polymerization, e.g., based upon where target primed synthesis initiates. The first strand refers to the strand of the target DNA upon which the reverse transcriptase domain initiates polymerization, e.g., where target primed synthesis initiates. The second strand refers to the other strand of the target DNA. First and second strand designations do not describe the target site DNA strands in other respects; for example, in some embodiments the first and second strands are nicked by a polypeptide described herein, but the designations ‘first’ and ‘second’ strand have no bearing on the order in which such nicks occur.

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 ultraconserved 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).

Inverted Terminal Repeats: The term "inverted terminal repeats" or "ITRs" as used herein refers to AAV viral cis-elements named so because of their symmetry. These elements promote efficient multiplication of an AAV genome. It is hypothesized that the minimal elements for ITR function are a Rep-binding site (RBS; 5'-GCGCGCTCGCTCGCTC-3' (SEQ ID NO: 1538) for AAV2) and a terminal resolution site (TRS; 5'-AGTTGG-3' for AAV2) plus a variable palindromic sequence allowing for hairpin formation. According to the present invention, an ITR comprises at least these three elements (RBS, TRS and sequences allowing the formation of an hairpin). In addition, in the present invention, the term "ITR" refers to ITRs of known natural AAV serotypes (e.g. ITR of a serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 AAV), to chimeric ITRs formed by the fusion of ITR elements derived from different serotypes, and to functional variant thereof. By functional variant of an ITR, it is referred to a sequence presenting a sequence identity of at least 80%, 85%, 90%, preferably of at least 95% with a known ITR, allowing multiplication of the sequence that includes said ITR in the presence of Rep proteins.

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.

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. In various embodiments, the nucleic acids are in operative association with additional genetic elements, such as tissue-specific expression-control sequence(s) (e.g., tissue-specific promoters and tissue-specific microRNA recognition sequences), as well as additional elements, such as inverted repeats (e.g., inverted terminal repeats, such as elements from or
5 derived from viruses, e.g., AAV ITRs) and tandem repeats, inverted repeats/ direct repeats (e.g., transposon inverted repeats, e.g., transposon inverted repeats also containing direct repeats, e.g., inverted repeats also containing direct repeats), homology regions (segments with various degrees of homology to a target DNA), UTRs (5', 3', or both 5' and 3' UTRs), and various combinations of the foregoing. The nucleic acid elements of the systems provided by the
10 invention can be provided in a variety of topologies, including single-stranded, double-stranded, circular, linear, linear with open ends, linear with closed ends, and particular versions of these, such as doggybone DNA (dbDNA), close-ended DNA (ceDNA).

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
15 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
20 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
25 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
30 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.

Operative association: As used herein, “operative association” describes a functional relationship between two nucleic acid sequences, such as a 1) promoter and 2) a heterologous object sequence, and means, in such example, the promoter and heterologous object sequence (e.g., a gene of interest) are oriented such that, under suitable conditions, the promoter drives expression of the heterologous object sequence. For instance, the template nucleic acid may be single-stranded, e.g., either the (+) or (-) orientation but an operative association between promoter and heterologous object sequence means whether or not the template nucleic acid will transcribe in a particular state, when it is in the suitable state (e.g., is in the (+) orientation, in the presence of required catalytic factors, and NTPs, etc.), it does accurately transcribe. Operative association applies analogously to other pairs of nucleic acids, including other tissue-specific expression control sequences (such as enhancers, repressors and microRNA recognition sequences), IR/DR, ITRs, UTRs, or homology regions and heterologous object sequences or sequences encoding a transposase.

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.

5 **Tissue-specific expression-control sequence(s):** As used herein, a “tissue-specific expression-control sequence” means nucleic acid elements that increase or decrease the level of a transcript comprising the heterologous object sequence in the target tissue in a tissue-specific manner, e.g., preferentially in an on-target tissue(s), relative to an off-target tissue(s). In some embodiments, a tissue-specific expression-control sequence preferentially drives or represses
10 transcription, activity, or the half-life of a transcript comprising the heterologous object sequence in the target tissue in a tissue-specific manner, e.g., preferentially in an on-target tissue(s), relative to an off-target tissue(s). Exemplary tissue-specific expression-control sequences include tissue-specific promoters, repressors, enhancers, or combinations thereof, as well as tissue-specific microRNA recognition sequences. Tissue specificity refers to on-target (tissue(s)
15 where expression or activity of the template nucleic acid is desired or tolerable) and off-target (tissue(s) where expression or activity of the template nucleic acid is not desired or is not tolerable). For example, a tissue-specific promoter (such as a promoter in a template nucleic acid or controlling expression of a transposase) drives expression preferentially in on-target tissues, relative to off-target tissues. In contrast, a micro-RNA that binds the tissue-specific
20 microRNA recognition sequences (either on a nucleic acid encoding the transposase or on the template nucleic acid, or both) is preferentially expressed in off-target tissues, relative to on-target tissues, thereby reducing expression of a template nucleic acid (or transposase) in off-target tissues. Accordingly, a promoter and a microRNA recognition sequence that are specific for the same tissue, such as the target tissue, have contrasting functions (promote and repress,
25 respectively, with concordant expression levels, i.e., high levels of the microRNA in off-target tissues and low levels in on-target tissues, while promoters drive high expression in on-target tissues and low expression in off-target tissues) with regard to the transcription, activity, or half-life of an associated sequence in that tissue.

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.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic of the Gene Writing™ genome editing system.

Figure 2 is a schematic of the structure of the Gene Writer™ genome editor polypeptide.

Figure 3 is a schematic of the structure of exemplary Gene Writer™ template RNAs.

Figures 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.,
 10 comprising a DNA binding domain, e.g., of a Table herein; 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 of a Table herein), reverse transcriptase domains (e.g., comprising a reverse transcriptase domain of an element of a sequence of a Table herein; other retrotransposases (e.g.,
 15 as listed in a Table herein); a peptide containing a reverse transcriptase domain (e.g., as listed in a Table herein)), and/or an endonuclease domain (e.g., comprising an endonuclease domain of an element of a Table herein; Cas9; nickase Cas9; a restriction enzyme (e.g., a type II restriction enzyme, e.g., FokI); a meganuclease; a Holliday junction resolvase; an RLE retrotransposase; an APE retrotransposase; or a GIY-YIG retrotransposase). Exemplary GeneWriter polypeptides
 20 comprising exemplary combinations of such domains are shown in the bottom panel.

Figure 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.

30

Figure 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.

Figures 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 binding domain and introduces a DNA nick with its EN domain (*). A Cas9 nickase is then used to generate a second nick (X) at the top or bottom strand, upstream or downstream of the EN introduced nick.

Figure 8. The linker region at the C-terminus of the DNA-binding domain of R2Tg can be truncated and modified. Deletions in the Natural Linker from the myb domain at A or B to positions 1 or 2 along with replacement by 3GS (SEQ ID NO: 1024) or XTEN synthetic linkers were constructed (A). Integration efficiency was measured in HEK293T cells by ddPCR (B).

Figure 9. Landing pads designed for testing target site mutations of R2Tg Gene Writer.

Figure 10a. ddPCR assay measuring percentage of integrations from all lentiviral integrated landing pads per cell.

Figure 10b. Amplicon-sequencing and NGS analysis of indels present at landing pads sites.

Figure 11. AAVS1 ZFP replacement of DNA binding domain of a Retrotransposase Gene Writer. This Figure discloses "3GS Linker" as SEQ ID NO: 1024.

Figure 12. Cas9 or Cas9 nickase replacement of DNA binding domain of Retrotransposase GeneWriters with or without active EN domain (*= mutant)

Figure 13. AAVS1 ZFP fusion to a Retrotransposase Gene Writer with or without functional DNA binding domain.

Figure 14. Schematic of nickaseCas9-GeneWriter fusions. (A) Schematic of nickaseCas9 fused to Gene Writer protein. (B) Schematic of 3' extended gRNA.

Figure 15. Schematic of nickaseCas9-GeneWriter fusions. (A) Schematic of nickaseCas9 fused to Gene Writer protein. (B) Schematic of donor transgene flanked by UTRs and homology to the cut site.

Figure 16. Schematic of constructs. (A) Schematic of Gene Writer protein. (B) Schematic of donor transgene flanked by UTRs and homology to the cut site. (C) Schematic of Cas9 constructs used.

Figure 17. The schematics for mRNA encoding Gene Writer (A). The native untranslated regions (UTRs) were replaced by 5' and 3' UTRs optimized for the protein expression (shown as 5' UTR_{exp} and 3' UTR_{exp}). The Gene Writer protein expression was assayed by HiBit assay by probing HiBit tag expression (B). This Figure discloses "3GS" as SEQ ID NO: 1024.

Figure 18. Genome integration induced by Gene Writer protein with its native UTRs and UTRs optimized for the protein expression. The Gene Writing activity with non-native UTRs is stimulated by the presence of the RNA template bearing the retrotransposon native UTRs.

Figure 19. Delivery of Gene Writer system using mRNA encoding the polypeptide and plasmid DNA encoding the RNA template for retrotransposition.

Figure 20. Diagrams of example 5'UTR engineering strategies. HA = homology arm; K = Kozak sequence; pA = poly A signal; AMa = *A. maritima*; Rx = other species of retrotransposon.

Figure 21. Possible location of an intron (or introns) within the RNA template. Introns are shown by curved lines. 5'HA: 5' homology arm; 3' HA: 3' homology arm; 5' UTR: Retrotransposon-specific 5'UTR; 3' UTR: Retrotransposon-specific 3' UTR; GOI: gene of interest. Orange blocks correspond to the sequence designed to be expressed from the genomic location harboring its own cell specific promoter, poly(A) signal and UTRs for the protein expression (5' and 3' UTR_{exp}). The sequence can be oriented in the sense (shown above) or the antisense orientation related to retrotransposon UTRs and homology arms. The intron can be located within GOI, or within UTR_{exp}.

Figure 22. Genome integration in HEK293T cells as reported by 3' ddPCR assay. The Gene Writer mRNA at 0.5 µg/well was co-transfected with the RNA templates with or without enzymatically added cap 1 and the poly(A) tail. The Gene Writer mRNA to RNA transgene ratio was 1:1.

Figure 23. Genome integration detected by 3' ddPCR induced by expression of Gene Writer mRNA produced with either unmodified (G0) or modified nucleotides (pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U) or 5-methylcytidine (5mC)). 1 µg of Gene Writer mRNA per well was used. The non-modified RNA template was used. The Gene Writer RNA to the RNA template were co-transfected in 1:8 molar ratio.

Figure 24. Construct diagram of driver and transgene plasmids. Homology arms (HA) and stuffer sequences are variable in this set of experiments.

Figure 25. (A) Timeline of experiment. (B) Schematic of R2Tg and transgene construct configurations. (C) Western Blot against Rad51 shows loss of Rad51 protein expression at day 3.

5 Figure 26. U2OS cells were treated with a non targeting control siRNA (ctrl) or siRNA against Rad51, along with R2Tg Wt or control RT and EN mutants. ddPCR at the 3' (A) or 5' (B) junction was used to assess integration efficiency on day 3.

Figure 27. (A) Sequence map of Ribozyme of R2 element from *Taeniopygia guttata* (R2Tg) in context of modules of Gene Writer transgene molecule RNA. The Ribozyme features are denoted as: P, based paired region; P', based pair region complement strand; L, loop at end of P region; J, nucleotides joining base paired regions. Figure discloses SEQ ID NO: 1734. (B) Prediction of ribozyme secondary structure of R2Tg. Shaded box indicates a predicted catalytic position that could be used to inactivate the ribozyme. Figure discloses SEQ ID NO: 1734.

10 Figure 28. Sequence map of Ribozyme of R2 element from *Taeniopygia guttata* (R2Tg) in context of modules of Gene Writer transgene molecule RNA. The Ribozyme features are denoted as: P, based paired region; P', based pair region complement strand; L, loop at end of P region; J, nucleotides joining base paired regions. Figure discloses SEQ ID NO: 1734.

Figure 29. Prediction of ribozyme secondary structure of R2 element from *Taeniopygia guttata*. Figure discloses SEQ ID NO: 1734.

20 Figure 30. Gene Writing system for treating an exemplary repeat expansion disorder. Figure discloses SEQ ID NOS 1645, 1599, 1645, 1635-1636, 1645 and 1686-1688, respectively, in order of appearance.

Figure 31. An illustration of two orientations of second strand nicking in an exemplary Gene Writing system.

25 Figure 32. An illustration of the orientation and position of second strand nicking in an exemplary Gene Writing system and their effect on editing.

Figure 33. Shows generation and expression of Cas9-RT fusion proteins. To assess expression of novel Gene Writer polypeptides in human cells, U2OS cells were transfected with Cas-RT expression plasmids harboring various RT domains from Tables 1 and 30 fused to a wild-type (WT) or Cas9(N863A) nickase. Cell lysates were collected on day 2 post-transfection

and analyzed by Western blot using a primary antibody against Cas9. A primary antibody against GADPH was included as a loading control.

Figure 34. Shows improving expression of Cas-RT fusions through choice of linker sequence. To assess how linkers can alter the expression of novel Gene Writer polypeptides in human cells, U2OS cells were transfected with Cas-RT expression plasmids harboring various linkers from Table 42 fusing the Cas9(N863A) nickase to the RT domain of an RNA-binding domain mutated R2Bm retrotransposase. Cell lysates were collected and analyzed by Western blot using a primary antibody against Cas9. A primary antibody against vinculin (left) or GADPH (right) was included as a loading control. Cas9 controls on the left represent titration of a Cas9 expression plasmid. Empty arrows indicate the original linker tested, while the filled arrow represents a linker (Linker **10**) found to substantially improve expression of the fusion polypeptide. Sample numbers correspond to linker sequence identifiers in Table 42.

Figure 35. Shows Cas/gRNA DNA targeting activity is preserved in Cas-RT fusions. Various RT domains were fused to Cas9(WT) and electroporated into U2OS cells. Genomic DNA was harvested and analyzed for mutational signatures by next generation sequencing. Mutations in the RNA or DNA-binding domains (RBD or DBD) of R2 retrotransposase domains is indicated, where relevant. Indel frequency is used here as a proxy for Cas activity preservation in the context of the RT fusion.

Figures 36 A and B Disclose application of mutations improving reverse transcriptase domains. Conserved reverse transcriptase domains from the retrovirus genera Betaretrovirus, Deltaretrovirus, Gammaretrovirus, Epsilonretrovirus, and Spumavirus were aligned and compared to mutations previously shown to improve RT activity (Anzalone et al Nat Biotechnol 38(7):824-844 (2020); Baranauskas et al Protein Eng Des Sel 25(10):657-668 (2012); Arezi and Hogrefe Nucleic Acids Res 37(2):473-481 (2009)). Figure 36 A shows a set of 3 core mutations was identified and applied to RTs from these genera as indicated in. Figure 36B discloses additional mutations were applied with first priority from the set of T306K/W313F, or alternately from L139P/E607K where neither of the first set were deemed transferrable. Selected mutations are shown in Table 45.

Figure 37. U2OS cells were nucleofected with various Cas-RT fusion vectors in which the RT domain was selected from a database of monomeric retroviral reverse transcriptase domains. Editing of a HEK3 locus using a Template described in Table 43 was assessed by

amplicon sequencing and analysis of precise editing vs indel signatures. Data are represented here as Activity Ratios, which are calculated as the ratio of the frequency of reads with the precisely intended edit (CTT insertion at the target nick site) to the frequency of reads with any other mutations (indels). Three Template RNA configurations assayed resulted in similar outcomes, so the results for a single template (Template P2 from Table 43) are shown.

Figure 38. shows targeting multiple loci simultaneously results in efficient Gene Writing activity. HEK293 cells were nucleofected with Gene Writing systems comprising different compositions of Template plasmids to enable targeting of: 1) HEK3 alone, 2) HBB alone, or 3) both HBB and the HEK3 locus. Percent of editing is indicated for each locus upon delivery of one or both locus-specific Template RNA expression plasmids. Filled bars represent Perfect Writing events, while unfilled bars represent the frequency of indels. Target-locus-specific editing was seen when delivering either Template independently, and highly efficient and specific edits were seen at both loci when co-delivering the Templates.

Figure 39. Shows effect of length on Gene Writing activity. HEK293T cells were nucleofected with all-RNA Gene Writing systems comprising various Template RNAs (Table 48) to test editing efficiency of the DNA-free approach at the HEK3 locus. Template 4, which encoded the same edit as Template 1, but with an addition of 20 nt at the 3' end of the RT template, showed an approximately 3.1-fold drop in precise Writing activity and an approximately 2.4-fold drop in the ratio of precise corrections to indels.

Figure 40. Shows effect of all-RNA delivery of Gene Writer using different mRNA compositions. Nucleofection of various Cas9-RT(MMLV) mRNAs (Table 49) into HEK293T using Template 1 (Table 48A). No strong effects were observed here in varying capping and UTR compositions.

Figure 41. HEK293T cells were nucleofected with a Gene Writing system using a set Template (Template 1, Table 48) for editing the HEK3 locus and two different Cas-RT constructs. Sequence analysis indicated that both Cas-RT fusions made edits in a very precise and efficient manner. In both systems, there was an increase in efficiency under conditions including the optional secondary nick. These data show successful cloning and Precise Writing by the PERV RT domain in the context of these Cas-RT fusions.

Figure 42. Shows the effect of all-RNA delivery of Gene Writer employing modified nucleotides. mRNA molecules encoding the Cas-RT(MMLV) polypeptide were varied in

composition to determine effects (Table 49). Here, Template 1 is used to edit the HEK3 locus after incorporating modified nucleotides in the mRNA component. Gene Writing activity with a 5moU-modified mRNA component was found to both high and precise.

Figures 43 A, B, and C show the effect of all-RNA delivery of Gene Writer using different mRNA compositions delivered into the cell via lipid particles. Figure 43 A shows all-RNA lipofection of various Cas9-RT(MMLV) mRNAs into HEK293T was performed using Template 1 (Table 48) and delivering via Lipofectamine 3000. Figure 43 B shows all-RNA lipofection of various Cas9-RT(MMLV) mRNAs into HEK293T was performed using Template 1 (Table 48) and delivering via MessengerMax reagent. These data indicated higher precise editing efficiencies with the MessengerMax reagent. Figure 45C shows assay of two Templates differing in total length using MessengerMax reagent. No major changes in efficiency of editing were found to be associated with the template change in this experiment. Where included head-to-head, the addition of the second-nick gRNA resulted in an increase in efficiency of the system.

Figure 44. shows all-RNA delivery of Cas-RT using lipid-based systems. The Cas9-RT(MMLV) and Cas9-RT(PERV) were delivered into HEK293T cells with Template 1 (Table 48) using MessengerMax lipid reagent. Here, activity for both enzymes was around 5% Precise Writing.

Figures 45A and B show expression of all-RNA Gene Writer system in primary human CD4+ T cells. Figure 45A shows Gene Writer protein expression from mRNAs with varying doses delivered into primary human CD4+ T cells at day 1 post-nucleofection. Gene Writer was detected by an antibody targeting a Cas9 part of the polypeptide. GAPDH, a housekeeping gene, was detected by an antibody against GAPDH. Increasing expression levels were observed with increasing doses of nucleofected mRNA encoding the polypeptide were delivered, e.g., 0, 2.5, 5, and 10 μ g Gene Writer mRNAs. Data for the detection of protein expression shown comprised 2 replicate. Figure 45B shows Cell viability after nucleofection of 6 Template RNAs. Viability of primary CD4+ T cells after RNA delivery of the Gene Rewriter system at day 3 post nucleofection. Cell viability was assessed by flow cytometry after live/dead staining of harvested T cells (mean $\pm$ s.d., n = 2 replicates).

[Gate: Live cells in a singlet population of cell population selected by FSC/SSC size plot]

Figures 46 A and B show Gene Writing in primary human CD4+ T cells. Figure 46 A shows precise editing of the HEK3 genomic locus by a Gene Writer system in primary human CD4+ T cells, without addition of second-nick gRNA. Figure 46 B shows precise editing of the HEK3 genomic locus by a Gene Writer system in primary human CD4+ T cells. Genomic DNA was extracted from cells at day 3 post-nucleofection. Genome editing of HEK3 was examined by PCR-based amplicon-sequencing assay. DNA amplicons containing the expected genomic alteration were identified as Precise Write events, whereas amplicons with unintended editing (e.g. insertion, deletion) were counted as Indels. The percentage of each was calculated based on total reads per condition (mean $\pm$ s.d., n = 2 replicates).

Figures 47 A and B show use of a second-nick gRNA for Gene Writing in primary human CD4+ T cells. The data generated in Figure 46 are shown here for a direct comparison of potential effects of second-nick gRNA on efficiency. Figure 47A shows in this experiment, the addition of a second-nick gRNA did not result in an enhanced precise writing signal. Figure 47 B shows rather, the use of a second-nick gRNA may have increased the frequency of indels. Thus, in some embodiments, a second nick gRNA sequence may be absent from a system described herein. Precise editing of HEK3 genomic site by the Gene Writer system in primary human CD4+ T cells, without (Figure 47 A) or with addition of second-nick gRNA (Figure 47 B). Genomic DNA was extracted from cells at day 3 post-nucleofection. Genome editing of HEK3 was examined by PCR-based amplicon-sequencing assay. DNA amplicons containing the expected genomic alteration by Gene Writer system were identified as Precise Write events, whereas amplicons with unintended editing (e.g. insertion, deletion) were counted as Indels. The percentage of each was calculated based on total reads per condition (mean $\pm$ s.d., n = 2 replicates).

Figure 48. shows 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.

Figure 49. Screening of candidate retrotransposons identifies 25 candidates working to integrate a trans payload in human cells. A total of 163 retrotransposon systems were assayed for activity in human cells as described in Example 39. 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.

FIG.50 A & B show luciferase activity assay for primary cells. LNPs formulated as according to Example 44 were analyzed for delivery of cargo to primary human (A) and mouse (B) hepatocytes, as according to Example 45. 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.51 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.

25

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 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 heterologous object DNA sequence may include,

30

e.g., a substitution, a deletion, an insertion, e.g., a coding sequence, a regulatory sequence, or a gene expression unit.

More specifically, the disclosure provides reverse transcriptase-based 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. This disclosure is based, in part, on a bioinformatic analysis to identify reverse transcriptase sequences, for example in retrotransposons from a variety of organisms (see Table 1 or 3).

The disclosure provides, in part, Gene Writer™ genome editors comprising a polypeptide component and a template nucleic acid (e.g., template RNA) component. In some embodiments, a Gene Writer™ genome editor can be used to introduce an alteration into a target site in a genome. In some embodiments, the polypeptide component comprises a writing domain (e.g., a reverse transcriptase domain), a DNA-binding domain, and an endonuclease domain (e.g., nickase domain). In some embodiments, the template nucleic acid (e.g., template RNA) comprises a sequence that binds a target site in the genome (e.g., that binds to a second strand of the target site), a sequence that binds the polypeptide component, a heterologous object sequence, and a 3' target homology domain. Without wishing to be bound by theory, it is thought that the template nucleic acid (e.g., template RNA) binds to the second strand of a target site in the genome, and binds to the polypeptide component (e.g., localizing the polypeptide component to the target site in the genome). It is thought that the endonuclease (e.g., nickase) of the polypeptide component cuts the target site (e.g., the first strand of the target site), e.g., allowing the 3' homology domain to bind to a sequence adjacent to the site to be altered on the first strand of the target site. It is thought that the writing domain (e.g., reverse transcriptase domain) of the polypeptide component uses the 3' target homology domain as a primer and the heterologous object sequence as a template to, e.g., polymerize a sequence complementary to the heterologous object sequence. Without wishing to be bound by theory, it is thought that selection of an appropriate heterologous object sequence can result in substitution, deletion, or insertion of one or more nucleotides at the target site.

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™ genome editors

Gene Writer™ genome editors are systems that are capable of modifying a host cell's genome and can be applied for the mutation, deletion, or other modification of a genomic target sequence, including the insertion of heterologous payloads. In some embodiments, these systems take inspiration from a group of naturally evolved mobile genetic elements known as retrotransposons. Gene Writer™ polypeptides can also comprise RT domains derived from sources other than retrotransposons, e.g., from viruses.

Non-long terminal repeat (LTR) retrotransposons are a type of mobile genetic elements that are widespread in eukaryotic genomes. They include two classes: the apurinic/apyrimidinic endonuclease (APE)-type and the restriction enzyme-like endonuclease (RLE)-type. The APE class retrotransposons are comprised of two functional domains: an endonuclease/DNA binding domain, and a reverse transcriptase domain. The RLE class are comprised of three functional domains: a DNA binding domain, a reverse transcription domain, and an endonuclease domain. 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 transposase, and a variable 5' region generally having Open Reading Frame(s) ("ORF") encoding transposase proteins. The RNA sequence template may also comprise a 5' untranslated region which specifically binds the retrotransposase.

In some embodiments, as described herein, the elements of such non-LTR 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

5 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 some embodiments, the DNA-binding function may involve an RNA component that directs the protein to a DNA sequence, e.g., a gRNA. In other embodiments, the Gene Writer™ genome editor protein may comprise a reverse transcriptase domain and an endonuclease domain. In

10 certain embodiments, the elements of the Gene Writer™ gene editor polypeptide can be derived from sequences of non-LTR retrotransposons, e.g., APE-type or RLE-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

15 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™

20 genome editor protein is capable of target primed reverse transcription. In some embodiments, the Gene Writer genome editor protein is capable of second strand synthesis. Table 50 shows 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.

25 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

30 predicted retrotransposase amino acid sequence. 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 50.

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_000002235.5	Strongylocentrotus purpuratus	AAGTGGTTCAGTT CTCCGTTTAGAAA CAGGGGCTTAAAA TTGCCCATTTGAA TATTCATAGCTTAG TTGCTAAGATTGAC GAAGTATGATTACT TCTGAGAAATGCAT CATTTGATGTTATT GGTTTAAATGAAAC ATTTTGCAGCTCC	TGAATGGCTCAGCCGTAC ATACACTACGTGCATGTTCC TGTAATATTTACATGCTCT ATTCACATAAATGCGACTT AATTAATATTTCTCTATTG ACATTACTGCTTACTAGGT TAATGATATATGTTAATTT TGCAACGAGTATTATTTT TGTAAGTCTTCTGTATAAA TGTTATAACGATGTAAT ATGTTTCAATCAGGGCCCCA TGGAAGACCACACTACTGCT GGTGAATGGGCTACCCCTG TCAAAAAGATCAGAAATAA ATAAATAAA	MIDSELYDDYILLRRDRTRHGGVAVYISKR FNYKRDDLLISSTLECLWFELHPNKSPLVGE FYRPNSHVDFFFEELDGNLDSALSSEDMSCIL GDFNCDCMPHQLDLNAEKLIFYTNMYGLS QLIDTATRVNTSSTLIDLFVSKELYNEWGIF ETSDHYLVTVTRDFSAKIHNKSEYAEFRSFK NLDEKFYSELSTPWQV/KIDIDNVDAWD MWLTFNNIUNKHIPVCKKRIRKKACPWISDEI VKLMTTERDKVLKAAKANKSSDDLWDYKYLK NTVILMRKRREYNADQISQNHNNKQM WRCLNLPKSLPISPNSINVDGANVSDNANI AEAFNDFISNATKITKDLRRNKSQLRVDEDN NIGPTHSQFNLLITDFVLKQIDNLSIDKATG EDHISCKLLKMTKNVIAESLDCIINKSLSTGVVP REWKKARVVPFIKSGDISLNNYRPISILPIVSKI IERAVHQQLESEFLDVNKLHPNQSGFRPKHST ATALAKLVNQWSLNDNKEISGVAFIDLRKAF DTVDHELLHLKLAGICSNRTIYWFKSYLGDR QQVTQFKGAKSHPTIKLGVPPQSGILGPLFSI YVNSLPECTPEGIIDMYADDTTLTVSANDATV VEEKLTVALNRVMAWIKENRLVNLTEKTCVM LIGSPANLKKIESFTISLNGNLIRVQFTKCLGLL IDEELKWSKQIENVSKLTQKNISMIKRAKSLP QKSLKLLYNSLV/LPRLDYCSVVWSNRFHSHTT KLQKIQKRAARIILNKTYDTPSADLFTSLKWM TLDKRFELSRVMIMYKCVHNLAPSILQVGLIN PNDVHEHHTRQIDSGCLVRPRFHTDCYKCSPI VSSIFEWNKLDNSLRTASSVNSFKRLFQKSCN L	cd01650	452	710

CR1-2_IS	GCF_002 892825.2	Ixodes scapularis	TT	TGATCAGTTGAAGCGCTG ATGTAGATTGCTTTTGG CTCACACTGGTATGT CTCTGTTTGTGCTTTTGT TTCCCTGTAACTCTGAGT GTTTTTCATACTGTTAATA TTTTTATCTCAGACTGGA GAACAGGTTCTCTCTG CACTCGTGTATGCGGT TTGTATGTGTGTTTGT CTATTTTTTTTAGTGTC CAGACTAAGACCGATTG GTTGTCTCGGACACTTTA ATAACATTGATTGTTGA TTGATTGA	MRQHNNILNDQGNALDLCFSNTDGIHVARS ELSMVPEDSYHPALDVLIFDPPVLSPNGP SPHTSYKYSAGDYTGMYHHLGVDWASILES SDINGQVLLTTVVQEAMDMYIPLKTSKSRF PIWFSRELRTLLKRKDFLHRRFKKSGLPKWEV EFKICRKRCKKILNRDRQAHSDSVESDLVKNP KKFWRYAKSRGLGKNIIEIVLNQHSDSAKGM SELFADHFAVYNNNNCCSSILPHSDVFGDSL GSVVDEECVRLCIKKMKPTLSAGHDGIPAAI VKAYHELLPTICTIFNNSLTSGIFPDDWKLAV WVPLKSGNSGDVLYNRPISLSSFSKLFEMVI HQFLSFRSIVPNQHGMGRSTSTNLVSF MSAASEVVCNRGQLDVFYDLSKAFDVVNH NILLYKSSYGVCGSVNLLKSYLSDRHAYVRV NNTSSLYKASSGVPQGSVLPLLFNIFVNDV SRVILNSSFLQYADDIKLYKKISTLEDICLQND AYSFGRWCNDLRLNHSKTKIMTYSRKTHD IFFPYSYHGELLSRVSELRLGWFDSLRFDT HVIKVVQSALRTLGAVSRTKEFKSPSAFFLYC SLVRSQLEYASVWVWNGISKNTSSSERVQNK FISIFKYRYLENGDPCLVGNSQLVNLISLQQ RRVKADLLFLFKTLHGSINSPSLLSEVSLRVRV PTRLQSSFYARHFSNLPNPIQSAECYNRYSDT IDIFNSCGSKFGNAVQIFLLQKQG	cd01 650	316	567
Jockey- 1_DSi	GCA_004 382185.1	Drosophila simulans	TCGAGCTTTATTGC AAGACCAACCAAG CAGATATTCTTCTC ATATCCGAAACCCA TTTTTCAGACAGGT CCTACTTCACCTTA AGGGGGTATGA	TAAAATTAATATATACA TACATTATATATACATCAT ACA	MPRVSHPPPEWFSILIRSSIKFDTQPIQEDWV QCARISVSTTHGDIITLAAAYCPRFRATDDCF SSLSDSLGPKFVIGGDFNAKHTWWGSRVCN PKGRALCRQVLSRLNCHGTGTPTVWPTDP AKTPDILDAISGGLDPAIRIQTRVEDLLSDHS AIAVSVAAMLSAGKKLIFKTSDISKYHHW LNNSTVTPPLDTPELIDAAVNLTSQIEAA TYAAPRFSNQTRDRSRDVFHFWNPRVNLKS EKRLRRVWFSLRNPRDKTALNRATKELKETL VRLKEALDSFVGNLAPGDPHHNLWNVTKRI KRPLKRIPIRRQDNWCRTETERANAFAEHL EEVFTPANRCSNPEAAETARLLSEPLFDSSEIP QATEEEISNLIAAMGNHKAAGDDAINAIALKL LPPISKSLITTIINKCLQLGHFPPSSWKRAEVVMI	cd01 650	438	693

					LKPGKHEANPASVYRPIISLLPVLSKILERVFLTRV LPVIEEAGLIPDHOQGFRRRHGTPSEQHRVA QYILDAFENKKYCLAVMLDIKQAFDRVWHPG LLLKLNGLPQPYFKFLKSFLEGRFAVRCEA RSDVKEARAGVPQGSVGLPLYNAYTADLPVI PSQYIMAAATYADDTAFLTADTPAEASQTM QEQLDLNTWHTRWNI SVNDSKSTATTATR RGTCPPVSFNFGFPIEVANPNVLGFTLDRRLT WRPHLLKKRKQAENRVREFVWL MGRQSKLP TSTKLLIYKAIRPIWTYGIQLWG TASKSNIIQ TFENKT LRTVTNAHPFDNATHIEVLSFPWV KDEIKSSRQYMQRQLQDHNNLAQDLLDTS QRT RRLHRSPLDLSGIRHAN			
CR1- 4_MUn	GCF_012 275295.1	Melopsitta cus undulatus	TGATGAAGGGAGTGGAAC ACCTCCCTTATGAGGAGA GGCTGAGGAGCTGGGTC TCTTTAGCTTGAGAGA GGAGACTGAGGGGTACC TCATCAGTGTTCACAAATA TGTAAGGGTGGGTGTCA GGATGATGAGCTAGGCT TTTTTCAGTGATATCCAGT GATAGGACAAGGGGCAAT GGGTGTAACTGGAGCAT AGGAGGTTCCACGTTAAC ATCAGGAAGAACTTCTTTA CTGTAAGAGTGACAGAGC ACTGGAACAGGTTGCCCA GGGGGGTTGTGGAGTCTC CTACGTTGGAGATATTCAA GGCCCCCTGGACAAGTT CCTGTGTGATGTACTCTAG GTTACCTGCTCTTG CAGG GGGTTGGACTAGATGAT CTTTTGAGGTCCTTCCAA CCCTTGGGATCTGTGATT CTGTGA	SSF5 6672	2	126		

	Tad1- 65B_BG	GCA_900 519115.1	Blumeria graminis	AT	TAAGTCCTGCGAAATAG AGGTTTTTCAGGACAAAG AGTTGTGAGGAAGGAGG GAATGCGCAGGGCGCTTA ATTTCTTGAACTTATTTTC TTTATTTATGTTTTACGTAT AGATTGACAAAGCAAACT ATGTACTTAGGTTGTAGTG TGTCCTCGAGGGAGCATAG ACCCGTATCACGCCTATTG GGTGAGGGCGTAAAGCCT GATACGAAATGAAATCAA ATCAAATCAAA			SSSF5 6672	6	148
	RTE-1_EL	GCF_011 004845.1	Esox lucius	CTCCTTGAACTGCC ACCTTAACGTGGTG GAGGGGTTTGAGT ACCCGAGTGACCCT AGGAGCTATGTTGT CTGGGGCTATATGC CCCTGGTAGGGTCT CCCAAGGCAAAACA GGTCCTAGGCGAT GGGTCAGACTAAG AGCGGTTCAAAACC CCTTAATGAGAAAT TGTCCTGTGACGTC GCCCGGCATGGCG CAGCCGGGGCCCC ACCCTGGAGCCAG GCCCGAGGAGCGGA	TAAGCGGAAGATGATGCA ATGCAATGC	MKIAECGVPQGSPLPLFLLYIAILLQDENKN TKFGYADDIAILSIGNTASEAVAEAQKEVDKL VQLASENRIDFSTKSELLVGGGPKKKLDTS LSVQIQGHSVPTSPYVRWLGVWLDSQLNFK QHVEWCGKAQRITQFMQRQNSVQRSAAAP GPMIKVVQACVLSTALYGAEAWWPGLTRITT QRGKVKCTGVGWHTDLLDSTILKAVRAALPV WRTTPNMLHRESGITPSAVILQQRNILAAA RIQRLLDWHPLVFRSREDPKDTSIRLGLCICQ KNRSFKFEWYLRLEISAKNVPMAGPGKLY FPEASIRKNTGSDFQDEADYAKQWIASIAE TVCAYSDSSCGPARSAWGYTLKGGQVIGS DSGPLLGAEVYDAEIFGAQKALERALGKAGN CPIKVLDDNTTAVRALRTGQTKSSQEADEVFT RLARQARSVEIRWIPAHAGVMGNEDTDRLA KEALRNLPDEDTSPSEMRDRPKYTFALTRF VRQRCEVWESWWRDHRPSRYEELDLMK RKRPELALPRWAHRLITARTGHGDFIKYHE RFGHEFENCEVCVCGREKRPWHFAECRIALQQ WRNLKHEPPPRGREMLAEGKWERFLQLFTV TKCYERQTLHRED	cd01 650	529	784	

	ORTE-3_AE	GCF_002204515.2	Aedes aegypti		CATGGGGCGGCC TCCGTGGGCTCAC CACCACAGGAGG GACCATAAGGGC CGGTGCGAGAGGA TCGGCGGCAGTC GAAGCGGGGGCC TGA	TGATCGGATTGAGTGA CTACGTTGAGAGC CTACGTTGAGAGC GGACGTGCTGTG CTTCGTCGTGTA CAGTGTTCGAGT GAAAAGATCGTTTT TGCGATTGTCGTAG GAGGAATTGCT GGGTTTTTGGGT TCCGGAAGTGCCG TGTTGGAGCAGTT GGGAGGTTACCGT GCTGGTTTCCGAC GGATTGGACTGA TCCGTGATTGGCC AGCAACTCGGCTG GGTGAATAGGCT TTTCGGAGAGTAGT GAGAAAACGAAAA AGCAGCTGGTAGT GGAGTGGGAGGA GGTAGAACCTAG TGGAGGAGTGCGG	<p> TGATCGGATTGAGTGA CTGAGTTGAGAGGAGA GGAGAGAGATCACCATA TTGTGATTGCCAGATAA AGGTATTGTAAGTACAAAT CAATGTTTTAAATATGTA AATGGAAGTCAACCAAGT ATCGTTACTGGGAG GTTTAATTAATGGTAAGT GAAATTATTACAAAACATA ATCAATAGCACCAATCGTA ATTTTCATATATTATATC AAACAGACACCAATTG AAGCAATATGGATGAGGA GAGTGGAAATGCCAATC CTTATCGCTAACATACTGT TGATAGAGGGATGGTCT TAAACCCACTATTGAGA CATCTTTGTATGATGAA ATGACCGGTAGATCGGTG ACCAGTATCTAGGTTATAA TTTTGGAAATGCCTTAGGT AAACCTTAAATAATAGCAA </p>	<p> WQTGVVPLYKKGDRRVCNRYGITLLSPG KVARVLERRPIVEPIQEEQCGRGRGRTL DQLYTLRVMEGSWEFAQPVHMCVFDLEKA FDCVPRGILWRVIREYGVGLPRLRAVRSLYDR SRSLVRIAGSKSDLFPVHVGLRQGCPLSPVLFV IFMDRISRSQGPGEVRFGDHTISSLLFADDV WVLAPSSQDLQHALGRFAACEAVGMKISTS KSEAMVLSRKRVPCLQVGGECPLQVEEFKY LGVLTSEGRMERIDRRIGAAASAVMRSMYR SVVKKELSRKAKLSIYRSIVVPTLYGHLELV MTERTRSRIQAAEMSFLLRVAGRSRLDRVRS SVTREELRVEPLLLHIERGQLRWLGHILFRMPP ERLPGKVFRSPTGKPRGRPRTRWRWDYVSR LAWERLGVPPPEELEEVGEREVWASLLRLLPP RPPGG </p>	cd01650	1078	1338
--	-----------	-----------------	---------------	--	---	--	--	--	---------	------	------

[illegible]

[illegible]

					CGCTATTCTCTTTTTAA CTATTTAAATTTCAITGTTG TATGCCACAAGTTTTATAG TCATCATTTTAAACATCTTC TCTTATTATTATTATTT GTTAGCAACCAATCGATGT TGTTTCACAAAACCTCTGT CTCGGCACACGCTCTGTA	ISEMLHIKQMIHGLNKQSDTELLSDAYLPIE RIPPS			
PENELOPE	GCA_007 989325.2	virilis group	GGTCGCCAGAGCC ATCAATAAATATCA ACGGAAGGCACGC CGTATGCACAGCAA CCAAC		TAAATAATGATGTAAAA TGTTCGAAATAATGTTTAA TTTATTGATTATAATGTT AATTGTTTTTGTATCTTG GTGTTAGTGCCCTGAAGA CGGTTTGCCGATGTGCAA CCGAAATATATCGAAGA GAATTGAATAAAATGTTT TTCAITGTTTGTTTTAAACAA ACACGGACCTCGAGCCAG CCAACAAATAAATATTGAA ATATGGAAA	MSYAKIKTKYKDSKRTINKFQLTLVKLTCLKSSL KFLCKRKSNIIPNFIKNLTQHILTLITTDNKTHP DITRTLTRHTFYHTKIILNLIKHKHNLLEQET KHMEKAKTNIIEQLMTTDDAKAFFESERNIEN KITTLKKRQETKHDKLRDQRNLALADNNTQ REWFVNKTIEFPNVVALLAKGPKFALPISK RDFPLLKYADGEELVQTIKEKETQESARTKFSL LVKEHKTNNQNSRRAILDVTEQTRKLLKEN INIKILSSDKGNKTVMDEDEYKNMTNILD LCAYRTLRLDPTSLRQTKNNTFVAQLFKMGLI SKDERNKMTTAVPPRIYGLPKIHKEGTPLR PICSSIGSPSYGLCKYIIQILKNTLMDSRYNIKN AVDFKRVNNSQIREETLVSFVVSLFSPSIPI ELALDTRQKWTKEEHTNIPKQLFMDIVRFCI QENRYFKYEDKIYQLKGMMPMGSPASPIADI LMEELDKITDKLIKPRLLTKYVDDLFAITNKI DVENILKELNSFHQKFTMELEKDGKLPFLDS IVSRMDNTLIKWYRKPIASGRILNFNNSHPK SMIINTALGCMNRMIMKISDTIYHKEIEHEIKE LLTKNDFPNIKIITLLKRRQIERKKPTEPAKIYKS LIYVPRLSERLTNSDCYNKQDIKVAHKPTNTL QKFFNKIKSIPMIEKSNVVYQIPCGGDNNNK CNSVYIGTTSKSLKTRISQHKSDFKLRHQNNI QKTALMTHCIRSNHTPNFDETTILQQEQHYN KRHTLEMLHIINTPTYKRLNYKTDTENCAHLY RHLNSQTSVTISTSKSADV	cd00 304	470	558
I-3_SPur	GCF_000 002235.5	Strongyloc entrobus purpuratus	GGTAAAAAGTCGC CTCATATAACTAAT AAGTCCATATCAGA CCTATTGATTGTAC	TAGCTTTTAAAGCTATCAG GGATCCCTTAACTATGT CGATCTGTAAATAATAT AAAATGAAGTATTTCGGG	MCIQETWFNDFNVLSIPNVVVCVKRNPDRG RGGLVMYLNKNDITFQAMDTPRNAEYQKVDI FCKNGVISIFNYNPKKIELIDLNNMIRNVPG KYLILGDFNSHNPLWGSNKLDNTNGRTVESFIE	cd01 650	462	717	

				AATGGAACGCACA AGGCTTGACGGCG CATGGCGCAACTT TTTATCTCCCTTAT TGATAACAATTGT TAGGTATTCACCCA GATATA	TTTTCCATGAACAATTGC AACTAACTCCATCTCGCC TTTTATGTGCATTTTATGC AATTTTAAATTGCAGCTT TTTTTTTTTTTTTTTTTAT AAAACCATATCACTATTAA TTTCATTTACCAAGGAAA GTCGCTCGTGGCTTGATT TCCGAAGCCACGCCAAA GCTATACAGAAATTGACT GGCGCAAATAATCTTTGA GATTGAAGAGGCCCATCT AAACGAAGAACCAACCAA CC	NNDLVLLNDGSPTRLDPYGYTYSNLDLSLVSK ELARTEWVKVLSNFGSDHNLVCLNFKHVY NRPGENMGQRWSFRKVNWEEYREEIEEAL RKEGYENLGAELNIQGRYNVIVKVIKETSNCI PRVQJAKRRKVNPNWNNACKKAVQERNRL KNRLRRYYDLNDVKKYLKKAQVAQKVIRRAQ YDYWLEYCRMRNRTFDIGKVWEQVKRMIRD STMNKHIFHNLFVDGNILKESEKANAMATY FKSVDESLSVTKNENENILYNNSTNDDDCIN DDFNMSEMGNVLKEMKNSSSGKDDIYLIMI KNLPNKALVLLNLYNDIWKISLFPHQWYVEI QIPILKSGKDPTQVTSYRPSISLISTFSKIMEKMI KNRLTWFEKNRLLNPIQSGFRKYRNVKDHL VRLETDIHISLKSXYTVAVFLDLEKAYDSLWK TGLLRKLSILGIGGKMLLFKAFLLSRISNVRRIN YMSRHYTLENGIPQGSLSPLFTLMLNDLKI NVSDVKLSYADDIAIWCSNKGDELEARRKLQE ALHELEIWCWTKLVNVSPTSASVLFVTNKQK KDILLQIKGHNLPICEQYKFLGVRFDNLTWS VHIDSIIDCKKRINLLRCVSGTPWGSKEYELM MIYKGLIILDYGSEMYDSACISLTKLDSIQY RALKIITGTHNVLSLEALQVLTGELPLRYRRLL NDNFRVYLLSSTLNHPTKEVLGDDYDLNNLE WKIGEGPKLRSFRDEVVVEGQTNLLPKIPE WRLARPQVSFEHHGKVQKKEMPPEIKSIVL DSIGNSWNGFLHIYTDGSKLEDGKTGAFFV SHYVVKFFRLSHV/CIMRAELVAILMALEWIN TFIESSVVIFSDALSALQMIDSYFIRCAIVNEIL SIINILQFKDVCVAFEWIPSHCDIRGNEIVDRA AKKGAMKRCIDIVIPNTRSELAASRLIHYNKL WQAEWDNAEHGRFLYALHSDVSKPIHRQGL SRREECLIQLRVGNCSLNYVLFQINQHDHTGL CEVCOEETIKHFIFYCTKYDKDRLLKERIKLN NPNLKQLLMSNPITRELLGYVDNTRFR	cd01 650	702	964
	L2-3_XT	GCF_000 004195.4	Xenopus tropicalis	TATCCCTCTGCTGT ATACTGCCATATAA CCCACTATATTGCA GCCTCTTTATTAAT	TAGCAATACCTGTGCCCC ACCTCTCAAACTCTGGTC ATGCCCATTTCCACACCTT GTGTCTCAACCCCTTTCTCCT	MFYPALLFALSCLFKHLQHPPPSQKGITITA SLLSSPSLSAHEVFKLLCSITPVNSFSSHKTKTY KSRSHLAFSLFLLAAGDISPNPGPIPVLLGT RSPNPPCTPLKYRTLTPIPKASIQISCALWNSR			

	L1-1_VVi	GCA_011 037095.1	Vitis vinifera	TTTTCTGTTAGGC CTACCTTGATTTAG TACTTGTGATAGCT GTGTCAAACCTCACA TCTGTGTCCACTCC TCAGTGAGCAAGA GATATTTTTACAGC CCAAAACATTGGTT CTGGTGCTTGTA AA AGTCTCCCCCTCCC CACACAGCCTGCTT ACTTTGATCCCTGA TTGAACTGATT AAT TAGTCAATTAGTTT CTGGCTTTCCCCAA TCAGCTGCATATAA ATTCAGATGCAGCC TTGGGATGGCAC TAGTGT CAGGGGC AGGCACTGGTGCT ACCATTTCCTAAGT GCCAACATTT CACA AGTGCTAGGAAGA AGGGAATGGCAGT TAACAAGGCGAGTT GGACAAGGCATTG ACAAGGCTACAAA CTATCTCCTATCAA CACTGAACAACGTG GAATCGCAAGAAA ACA ACTTTGCTTAT TCATGGTAGTTTGT TATTTCTATCCCA AA	TAAGGGTGTGGATAGGGT GGCTCTTCGTTCTCTGGCA GGCTGCTAGGGAGCGTGG GGTGTTTTCTTTCTTTTC	MKLRVLSWNVRGANDSEKRKVIKALIKSQKV DVVCLQETKMQMESRMIVRSLGVGRCLDW KVLNSRGSSGGVLVFWDKRVLHLEAEVGNF SVSCREKNCEDEFCWI ESGVVGPISI MKEREN	cd01 650	496	763
--	----------	---------------------	-------------------	--	---	--	-------------	-----	-----

TGCTAACTGAAGTT TGGTTTGTGGTC CAGTGAGGGGAGA GGGGCTTATTAAG ATGGATGTT	TTTTCTCTTCTGTACTTT TAGCCACTGCTTCGGTGGT TTTTAATACAATCTCTTTAT TTATCAAAAAAAAAA	FWAELGAVRGLWSDPWCVAGDFNVWRFPV ESSRGGRLSASMRRFSEIMEDLELDLPQGG SFTWKGLNNQSHRLDRFLISNEWEDHFSG SVQYVLPKPTSDHPILLDGGGVRSGMPFRF ENMWLKEEGFKEKMQGWWVWGLNFGSAS YVLSKALKALLRDWNKLEFGKVEVNKALA LSQDFWDKMEISRTLVSQEVDAARRGAKED FKKWMEEISWRQKSREVWLREGDRNTKF FHKMANAHRGNMLSRVKINGVWLTKE VKEGVNEFKAMLSAGGWRPSLRGLSFERL EAVDAASLEEPFSEQEVMKALGFCGDKAPG PDGFSMAFWQSSWEFVKEEVLGFFREFHNI GRFVKSINATFIVLIPKKGAEGLDRFRPISLV GGLYKWLAIVLNLKRVGKVSKAQNAF VQGRQILDVILVANEVLDVILKNKEEVVMCK LDIEKAYDHVEWSFLFSVMRKMFGGEKWIR WMKWICISTVSFVSVNGSSGFFQSSRGLRQ GDPLSPYLFVLVMEAFSSLLRKAVAGGFVSAC KARSGGEGVNVSHLLFADDTLVFCGASKVQ LLHLNWLMMWFAMSGRLNLDKSELIPVGC VNNVEELAAIGCKVGSILPTSYLGLPLGAQYR SRVWDGMEERMIRKKLARWKSQYISKGGRI TLIRSTLANMPIYFMSMLSPRKVRLRLRIQ REFLWGGGALERKIHVLKWLVCLEKDNGL GVKSLILNKALLCKWSWRFAMERAFWNQ VIRGKYGEEQGGWSSKEARGETHGVGLWKT LRKEWVVKSRLVFVVGNGKRIKFWKDIWC GDEPLCVSPSLFALAVSKDAWVKDVWRCN EGGSWSPLFSRPENDWELEEVCSFFVALNR KQIQQGVDDRVWRETNCCKFSVKSLYKSLV SGHPISFPSSAIWKVTQPRVSFFGWEATWG KALTDQLQRRGWALANRCYLCQRHEESIDH ILLHCDKVRTLWVLLYSMFGVQWVLPATVKE TLGWNGSFVGKKRGVWVKASPLCLFWTV WKRTRNKVAFEEEEELSIQRLKASFVYFLWSETK RSIKDGPSTLVDVFGWVASR
---	--	--

R2-2_SMed	GCA_002 600895.1	Schmidtea mediterranea	GTCTAGGGTGAA CTGCAATCTGACA CGATGACCGAGCT GTGTCAGTTTGACG CTAGTCGCTAAAGA CTCGATCAGTCCGC CAAGTGAGGTGGC CGGTATCTGCAG CACTAGAGCCACTG GTATCAAGAGCAG AGATACGCGAGTG GAAGTTGAGTACG ACTACCTTCACGGG GTCCTCTGATAAC CACAGTGGAAGTG GGGAACATAAGTG GTGTCAGCGTTCC CTACTTTCTCGTAG GGTAAAGGGTAAAG ATAACCCAGAGAAT ATCCCATGGGAGA TATCCATGGAAAAA GCACCACGTTAGAC AATCCGATGGTCTA ACTCGGCTCCGAG GGGCTAACTATCCC AAAGGGCTTAA	TAGAAGGGAAACAAAAGGA AAAACGAATGACTGGTA ACTATGAAGGATATAGCT GAAAGCCGAAAGGAAGG CTAAGTCTGAAACCGCATC TACATCTTCGATCCCAAGA GGAACGTGGGTTAAGCT TGAGCCGACGAAAAAGC GAATGCATGTTAGACGAC GAGGTACAGTCACCTCCTC GTGGTATTTGGCGGGCAA TGCTCACTAAATTAAGTGT GAGTAGCTGAGAACTGTA TGTGTATCATGAAAAAAA AA	MRKELVTIKNLIEESGATAPAPVPLEVAVEVH QSSSVPEITDQSTTTTQEGSYSEPPHRCENCG REFRTRAGVQQHRRKAHTNEFMEKEKAAP TKLRLWTDEEKEILIESEIKIIEGSLKEQHEINK ILASRMPGRSQDGIKIRQKQEHKAEIQRRLH GTVTDTDEGRSRTSEITEPIRSIPINTKNWSED ELKRLAEVKLRNKNKEDINKKLAEIFPNRT MGSIKSRTKDKEYQDLVKLTMQTISENPDN ETDFNTSENTNSTDAAKEVKYNLNMLLSN NEEWLTSTLKEAATLALQGKKEASEKLENEY ASKTLFPGKITNQTRKREKKISKRETRRQEYA EIQLYKKNISSAAEKAINGKWSIKPEEYHNN KDLIKAWKPILEAPPFSDCRPIENIKEMDYAL MEISTAEILAIRAMGKTAPGPDGIKYSKLKN IQSMALFNTCLTSLFLPLKIARTILIPKQENP GILDYRPLTIASVVTRVFHSILAKKLDNNAQLS QRQKGRKCDGVAENIVILETILTNSRSEKRPL CMAFVDIRKAFDSVGHESIIRGAKRVGVPPM LLEYISSYQNASTNLFGEILNRRGVRQGDPL SPILNFVIDEALENLRNIGYLLKEEKVSCILAF ADDIVIAETKGGLENHIEKLEKLEKNGAGLEIN ASKCATLMVMKNGKEKSTYISTVKIKENDIP TMKATETYKVLQMGFKAREQNADEVITE GLENITRAPLPQQRIHILRDFLIPRLIHLKVLG RVAKSLKRIDQNIKKVRNWLHLPKDTTAAF IHADAGDGGGLVPALEHTIPLLRERITNLRKS NDPVTKECLRMEYTKQV/LGKWSRPTKIGETL ATNKSQLEAFRKQMLITLDGKGLKDHHTP TIHKWIRGENMTGKQFITAVKIRGNLVATKS RNSRGRPEQEKLECAQCGRPDSLGHILQGC WRTHGMRVERHNINCRRIKAIMKGKSEVW EEPRLQTNEGLRKPDDLICHKGKIIICDAQVVA DSSNCSLESENQRKIDYKKDSVVAEARKLIGR DDEDIIMAVTFNWRGAISKTSIRDLDMLLDIK SKEVIKMSRKIIRDNSIMVMEHRNRTKRR	cd01 650	474	730
Utopia-1_NVit	GCF_009 193385.2	Nasonia vitripennis	CCATTCTCTCGTAC GGGTTTTCTGTGCCG GCATAGCCGGGTG	TAGTGGGGCCATAACACC TAGGCCCCACAGTGTGGC GATGTCCATGTGTGTCGTC	MSLGTGREVKCITVNVLMQQPHTKAIREGD FIVILLPQSDDETLCPLCVGRGRYSKTRVEC LNRHVKEVHPDLTTTFCRWCGCGFAAPGDKK	cd01 650	477	740

GGAGACTCGCGG GGGAGGTCATAT CTCACCACCATCTT CGAGCTTTTGGTG CACCTGA	CTTATTTATTTATTTG TCCTGTCGAGCTGTTAAT TTATTCACATGTATGTGTT GTGTGTGTCGTCCTCGTA GCAGTTTTTATTCGTCCAA CCAGAGGTCGACCAAGAT TAAATAAGCATGGCTTG AAGCAGGCCAAGCGCCAT GTTCTAACCCGTTTTAGG GGAAGTTACTTAACCTAAA AATACAACTTTTCC	YPRKIVTQHCATCPEVSSAPSGRVDGERRV NTRRLGIAAATEASPVRRTRRNGLASPPVEQ DISQSAAPPEPARVPQHPEIVALGESADDEVF RSPVNSPPRDWRAAAPQQAASSPPAVPGIT AATPSNTRTNGSANSILAEHPITPPPTNT TEANGRADIPRSGRTPPPGQQAARRAPTTE QRRIVGLLEAATGREQLEEEATTQAMLFLARLT GRRPEPRNAIRPGPRQRHPAQEDVQAQAPD RIQEAKLQRLYRTSKKRAVQKILAGPNMNC QIDKNTITAHVELAARRDGGEDWPDVFDRE EPTAASGEALCTPTITREEVFRRLGRNNTSPG PDGITYRDLAKADPGAHVLAALYNSIWRIEAT PALWGVSNNTLYKKGDAMDINWRPISLGD TVPKLFALAILADRIKRWAVANGRYASQKGF EFEGCYEHNFVLQEAIREAKGGRKELVVAWL DLASFTSVPHSSILQALEGHGLPSKARNISSL YTGMITTRFHTAEGPTDPIIQSGVRQGCPLSP DVFNLTLEVLRQRTGEGYTIEGRRTSHLAY ADDAVAILADSPAGMRRLFAAERGARGAVGLT FNPAKCATLHIAGRGEEAVRPTFSVQGTQVPR ALASEEYEHGPIPTGYQVRQTPIINTLRDLA DIGSIDRSLAPWQKLDVAVGTFLPRLDFIMQ GAHIDKGFLTEADKIIKKAASWLSLPQRASA ELVFLPPSQGGGGLLTVASHYKMLYSSDVTVS TIAGSTLRRRTVSRLLKKNASIDIAFLSGDL PRSTSPSTFWTKVRSALRIKTLGLRWSWC QEGEVLLMACGDPRAPGTRVSPQTKHLVTTT LRRCLNRHYAESLLAKKQKQKVEFVTRRSQ SNHFLRSGSFTFCDWRFIHRARLDVLPNAA KRWQRGMIDKRCRRCGSDLETLPVLSHCGP HSAARQKRHNIIQDRLVKAASRCPGTISVNIQ TVGVGRGPDAALRPDIVVDDVNNRRVTIVDV AVPFENRLEAFDGVREAKIYPTPLARQLTDS GYTVTVFAFVVGALGAWDPNRNERSLSSIR YYAILMRRMLMVSDTIRWSRDIYVEHVSGIRQY RE
--	--	---

LIN9_SM	GCA_002 600895.1	Schmidtea mediterranea	AAACGACATCATGA ACGCTTGGCCGCA ACAATCCAGTTATC CCTGCGGTAAACATT GTGGAACCTATAA GACAAGTACTAAA AGAAGAATTAGAA AAATTAGAGAAA AAATTGAAAAATAAT TTATTTATAAAATTT AAAAATTTAAATTT AAATTTAAATGAAG ATAAAAAATTTATTT AATCCAATAAATAA TCAAGAAAAATCAA GAAA	TAAAGGGCAAAAAGATA TTTCAAGATGAATGTGGA CTCATCTAAAAAATGACCA CCTCGAGCCCAAAATAGACC TAGCTATCATGGTTGCTGA TGGAACACAGTAAGGCACC TGATAGCTAACTCTTCACC GTGGGTGCTTCAGACACT CACGGTGACAAAGAAAGGA CACCACCTAGTAAAAACAC TAGTTTCTGACACCTCTT GCTACAAACTCTGTA AAAA TCAAAAGGATCGATAGGC CGCGCTTTCACGGTCTGTA TTCGTAAGTAAAAATCAAGA TCAAGGAAGCTTTTCCCAT TTTAGTCAACACACAGGTTT CTGCTAGTTGAGCTTCC CTTGGGACATCTCGGTTAC CATTTGACAGATGTACCGC CCCAGTCAAACTCCCCACC TGCTCTCAAAACAGTTCAA TTGCATCCAAAGATCGCAA TTTTTTCATAAAAAAAAT TAACAAAAAGTTAATTATAC TGCTTCATTGAGTAAGTAG AAAAACAATC	MMDSRQLNTPKIRKYQNPKMTNDIMKSYNY AVLSDVTPQETTQTTQTHLNVVDIKETTPKQ PLTKSGPKSPVAVSYKFKDQATFIWDTTPQT NPPKDC TKLIDKTRPKRTIFKSAFQTYLKKELS NETFEVKTFLMATHKYRFKDENSRLLAYRIIN RYVMETANEFKETEFDMARFAKFFTIPENWL KHLKPYSTATETSPADRIKVKQLVDLTCRYPK TQEEQTSVANFLHFTQRSIGISRDYKFQKCI PFMARKTNGQETTSTINTTPPTTEENRLPMVII TPLEEKIEHRRPEKRGASNDTIVLSDEEFLLK RRTLPTRKSNPTGAGNVPTECTECTDEVKFI NNEYIECKEKGKVVWENVRNGMMHLRQKH DFPNRTDVMVSCVCEVPKIGAEVCNHIKNH KKDDKEESQAGSLVANTQDIPNESSLSQAAIE VYLRNLMKMKENQERNIQYLEPSTANFLNRN LRAFYNV KIEKLIGWEQVWLIHWNKCHWI VLANCDSKTSVILSDSNQMTLQQRCKIAK FDKFELEGTTEKAVLGTLEARNVPPQPNFDC GVVVIQYISDFLKPQRM DYHTPDSKRIRKEI GELILEEMKNPASKIKPNKEIQSLQLKFRLLQI NVNDVFWFAEEYQKSLPKIRTRDGLKLNKL SCSYQIQLFGLAPKRAVKEIFYQETSTADLET RVLNEHFKKDESMMECKIKNGNHYQDWIT KAQIDNKEILEALKNSTDSAPGEDNIPLRQWII WNNDGVLDFMFNYIKRTHDIPEMWKNYTT TLLIKPGKQESNIPANWRPISILPTSYRIFMKV LNKRVL DWANRGELISKWQKAVDKANGCDE HSYVIAALIEKANRSYYKNEQCNLAFLDLADA FGSIPFQVWHTLKNMGMDEKTNILLKEIKD CSTKYKCGKNESEKIKITKGVRRQGCPLSMTLFS LCIQYLIQIAEKKKGATAGQEVICILAYADDL VIVANTAKDMQMLLTITTIENLAKQADLIFKPAK CGYYRDPDRKKSMKIYGEISIVDEKNVYTY LGVRIKDTKKDLNVRFEVKKKTTAIFKSKLR SDQKLEAYNIFCQSKFVYLQGEDIAKTIETY DEEIKMKIKEDILKLDKSPFTDFVIYSPREKG GLGITKIIDEQIQTINRTAKLLNSSHRAIRAIYE ELIQVANLRGEKEINSIEEALKWLEGTNKYKK	cd01 650	793	1057
---------	---------------------	---------------------------	--	--	--	-------------	-----	------

	R2-1_TG	GCA_012 069535.1	Taeniopogon a guttata	<p>GTCTAGTTACAACT GGGCATCGCTGCA GAGATCGACCTCC TCGTGGTCCCGCTG GTAGCCCTTCGAAG GGTACTAAGTCG ATCTGCCCCAGG TACGGAGCCGTG GGACTCACCACTCC AACGTAACCTCTGC CTAAACTCGGTGAA ACAAATCTCTCGGT AAAAAGCCCC</p>	<p>TAATTACAGTTATTAGAT GCTTAGTTATGTACCTTT CTTGATTGCTAGGATTT TGATAGTGTTATTATTTT ATATTTTGTACGATTGCA TAATGTTCTTTTTTATACAG CTCTGTTTTAATAAATAG ACGATACCTAGAGACGTT AGGCAGCCCAAGCCAG TTAGGTAGCGGATAGTAG GTAGGAACAGACTTTTACT ATTTCTAACGCGTCAATT ACCACCTGATTTGGACCAA TTCACGGGATTTGTCCAAG GTGGACGGGCCACCTTTA CTTAACCAAGGAAAAGGAA CATATATAATTTATGTGTG TTCGATAAA</p>				<p>NSNAKTTWITRVREAFQTEKKHKIKIVRFVPK ENGIGYKICDTQEKVVELDNDNSKELSKLHW MIKEAYFEWKTLCQGYLSLKTSEMEWK MPRGLPDPDWRELTKVRANMLDINMIKQAN QGGRIGNTCKRCKEDKESASHVINHCASGN WSRVEKHNVQNELAKELTKRNISEFKDCIPK ETKGLRPDLVIRLKDIMIVDIKCPFDEESAIE SARNKNIDKYRELAKEIQAKTLPJTISTFVVC SLGTWDKRNNELLRQMGIRYEEKSEMIRINMI HKAIHGSRKTYDQHRNFNNNG</p>	cd01 650	690	956
--	---------	---------------------	--------------------------	---	---	--	--	--	--	-------------	-----	-----

[illegible]

ATAAGCGACTTG GTGGTTGAAGA AATGAGAAGGGTG TAGAGAAGAGCGG AAAAGAAAAGGTGA ATTGAACACAGG GGTGTCAACTGAA TTGGGCTACGTGAC AGGATTGGAGCGA TAAGAGAAAAATA AGTGAAAGGTATT AATTATTTAAACT AAGGTAAGACGAC TGATGACTGGAAT GCATTCTAATACAC TGATGAATTAAGGT AAGACTAGATATAA GCAAATATATTA TGAGTTTCTTTTC TTTATCAATTATAG CATAACATTATATA CTACACAAAGAGTA CAAA	CCGAATTGAGTGAGATTTT CACGGAGCATTGGATTTC GGCACGTGTTTTGAATG GTTTTCCGATTATGGTT TGGATGTAGTGGCGTTT GACTGAAGTGAATTTGA AATTAATACCCCTACACA AAACTGTCTTCAGCAAC TTGTTCATAGCGTCAAAAT AGACAACTTTGCTGAAAAAT ATCATCAAGCTATTCTGTT AACGTGTTGAGTTATGTCA ATAAAAAAAATCGTAA TAAGGTCGCTTTGGTCA AACCATTAATGCTTTTGA CTCGTGGTTGGGGGGATC AGTCGAGGCATGTGTTG AGGTTGCGGTTGTGCCTG GTGTTCTGTGGGGTTTCA TCGGATCGGTCGCTGGA TAGCTGGGTCGCTGCTTG CCGGGTTGTCGTTTCGT GTGGGACCGGGGAGGT TGCGAGTGTGTTGTCAAAT ACTGTAATACATACAAGA CAAAATAACACAAGCACA TGTATATATATGAGAAC ACATGCAGAAATAGACAT TGACTTATAACACAAATTA AAGCTGTTTTAAAAAGATA ATTACTCAATTTTAAATAAT AAACATAGCATTAGTAC AAGAATGACATGAGTCA TCAGTCGTTTGGTTGGTT AAAAGAAAAATAAATTGT TCGATTTAAAAAATGCAAG GAATTTGAATATCATATAA TTAAGTAAAAATAAATACT	DVAALQEVNVKAELITTRNYEWYRIGGGQGN KSRGLAILRKDGIDVMEWKEIGKYGLISLLS KGRKIILNVHGPKNKNGYIFYSSLGKLISQDHLR SNLIMVGDWNAQJGKDSVMPEDAECIGNRL GFSNNNDNGEEFKMFLTHIKIKNMSTVIGKN TDITWKGKKESQIDHVLKPREGKIEIRFIKGY WTGINTDHKMIVTGIRTRKENKEKGGKPIPI DASVLKYDTIKEKYQEALKYEIRQEKGNTIEQ DFKAVAGMKKKAANEVLSARAPLTPIKAA LNKLTAINLSNKHDPYRWKLDKRRMEF QAAIRSHNERIKKFKYKELNDFDVAVRKISY QFLKGLKRKKHKNVYIPMSQWNEALKESEG PEIQLMDEHDTCPVTPPKKEEMMQIVQQQ CNGKSAGADGIRMELLKYADKQTQQLADI WKRNVWENHLPEDMEKTIQVPIPKNAKAG VQDFRISLCNVAYKPYAKWKQRLREFTGEP DINQAFTGRSTDHMFVTRRILEEYWNA GNPLYVAALDISKAFDNVSLQLSLAKLNV SHLIDRVLCIRKEKIRVRWQNYTTECNRGK GVKQGCPLSPFLNVMQDVVRKVKPEL NLMNTGSLKPLILIFADDLITKDEEGLKRIIA LEESLREVLEINNDKSQLIRAPNATKRPNTI NLNGRNVYDKTSMRYLGAWLDTLNRPAT KQRCVNAAKVKLVVEFCFKFPTWEIGKLIY NTVIAPSLYGTATLTKRSRVQLGKYEKQILK DIWNQCRKTDERKFNVRKELKGTINRRVRV NRISYGHQIRREETHPIKIMAMKLFNKKKH GRPSFTWKKSLQDFNRYNGITEEWSQIAQ DRDKIKKAAEEIYRNNESEISDGEDSTEQDM				
---	--	--	--	--	--	--

					<p>ACACTTGTTTACTTTTCGAG GAATGATTAGAAATATCTCG CGTTCGGTATCGTGGCTGT GGTGATCGGTTTCACTTCG TCGATTTTCCCTTTTAACTAA TTACTTGGCCGGGCGACT GTTTTGACTGCTATTTAG ATTGAGATTGTCCATTGTC ATGCTAAATCGTAAAGTGT TACCAAAATCTATTGAACT TCGTGTATTAAACCTTACT CGACGAAGAGAGATCGAT CACTCGGGTACGCCTGT ATGATGCTGAACGCAAGT TATGAATATAGAAAACTCT TCTCTCACTTTTGCATGG GATGTTTTAGATTAGCAGT GTTTGTCTGTTTTACTGGT GATTTTAGTGATGTCATTG TGTTGTGTGGGTTTTTCTT TTATAGCCAGTAGATTGTT GCAGAGCGCCCAAGGGG TTCAGATTATGATCATAGT GGAATTTATTATTTTCTTGT TATAAATAATATAAGAATA TAAAGAAATAATTACAATTA TCITCATTTGTACAATGCAA ATCATAAACTGAACCTCCT AATAATAAAATGTCTAGTTA TTTTTTATATGTTAACCAAT TAGAAGTTTTGAAGTAGTA CAATAAGTAGATTAGTA AATGTGTATATGCCGAAA AAAAGTCTTAATGCCAACC AAAGTTATAACAAGGTTCA GTCGTAGCATTAAAAATTCA GAATAACTTGTATGCCGA CTGAATGTTGTTTATAAA</p>
--	--	--	--	--	---

					ATCAAGAAGAAATAATCA TAATAATAAAATAACAT GGTGGAGGAAATGATAG CAACGAGAAGCATCATAA TAAITCTATGAGAACTTA TAGCAGGTATAAAATAAT ATGATGATTATTTCTACTG TTAAGTTATTCAAACTCT TAAAAATAGTTATAGAAT GAAACATTGCTATTGCCTG ATATGAAATTAAAAATAAT GAAAAACATTAAATAAAA TCAATAATTAGTTTTGAAT TCAAAAAAATAAATAAT AAATACAAATATTAAAAAT AATAAGATACATAATGAT TTGGATGATGAGGGGAG GATGGAGGTCGTTGCTA TCATTTGCTGTATGTATA AACATAAAATAAATGTTAA AATAACAAGTATGTATAA TTTTAGTCGCTATTAACTG ATGGATATACTGGTAAGA ATACATAATCGTCTGTTT GCGAGCAGTGAACAGTAA ATTACATCTATTGCAGT TTAACCTGATGAAACTACA GAGTATAATGACCGTTACA TAACCGAAATTATTAACA CTAGTATGTTAAATTCAAA TTTCTAATGAATTAAATTT CATTCTATTAAATTTATA GAGAACAAAAATAGGAATA ATCGATGGATCGATGTTG GAGCTGCAAAACAGACTAA AGCTGCATATTGCACACG GTCCTCTATGGAGTTATAT TTCATAATTTCTTCTCTA
--	--	--	--	--	--

[illegible]

					TCTGTTGGATCATATGGCT GGAAGGAGATTCTCTATTT CCCGGGGTTTCGGTAA GTGTTTCTACTACGGGTC CGCCACCCCTTTTGGCC CTTCATACAGCGATATGTT GAATGCAGATTGGTAATA AAATTTGACCTTACTGTG AGTGAGCCGCATGCAGA GCAAGACTCAAGCAAGGA TTGGTGCTACCTACATAC ATACATACA				
CRE- 1B_FoCa	GCF_002 217175.1	Folsomia candida	TGCATCATTTGCCAA A		TAAAGTGATTTCATACTTT ATCAATGTTATGTTAATA AATAATT	MPLFNRPQCRQLCPNPLSANNATIQLVDS PVSTHNETSRSTQLDAHRSNFVGGSQLAKVE SSDLDRSSPFGVFIYRIETGKIIAQPSSSLPPS SSPLLPSSPLPLSSSSSTPLPLSSSSPPPPSS SSSPLPLSSSSPPPPSSSSPLPLSSSSSTPL PLSSSSPPPPSSSSPPRPHPGPRKRKQKRT KPRKLPPPTSSQPLPHTTEGLNPVEDLFCPV GCPGSSIGTDEWILCNACLIWYHTSCLFLQAI PEGDWFCANCPQIDTQIGPQINSNTSSPANQ SYLGTIQLTQSIPTNPRASQPISTLSTPLSNFI SSKETCTSLTLEWLNHNHIVPGGVIEKAILHDAY RRDSLGIKFPINDAEMGKLIKNSFNIRTKRIRT ELHGTSKYHYVGIQFKEHSAHSVDFQTISS LRQNCPTLKQCPKACRIPVAQVLKSAVDKCS ENTPEAWHLLTFPYITLQCESKANIKNGQSL NSKIRNNLSIFESLDKNYQDYILTLSSKNSAH DKPGSTIKDIGPKVISKVSEGNIRGAIRLLTSD IIAPNSSETLETVEKHPPQPLDNFTVLDEPNIP PFTTTVEQVISAIFSPSGSAGGLDALRPQILK DLVSQHTGDVGTLLESMTKFCNFMLAGKV PLDICPFYIGASLTALKKKGGIRPIAVGCTYR IVSKICVMNHNHVFQDYFSPVQVGFNVKNG AEAGAHAAARRFFSFPKHSPTAFLKIDFKNAFN MVDRGKLLSSVHELVPITYPINQCYRFSSSLV WNSSLIESCRGVQQGDPLGPPLFCLILHPWVS NLGSEFNMMWYLDGTLGGDPTTVLDLDFKHII	SSF5 6672	674	871

					SCANEVGLIEINFDKCEVSVILGLQTNQCILES FQSVAPGISCMITQENAEALLGAPLTYESIEHVF RKKINAFSNMAGRLNSIPNHIAYYLLKHS LAIP RLVYLLRAAPIFKSPHLLFEEDSLLHDTLEQITN CQINPLSLEEIALPVKLGGLGIKQSKTLATPCFL ASSFATTDLVLSLIPSYVTALSDSLVTEAQDLW KSI SCDLPRSTISSIQSSWEIPIFNQVCSRILDS STTPINTARILASRSKESGAWLQALPSRNLGTL LDNDSFRIAIGLRGLKLCFPHKVCVCGTQIDEL GLHGLSCKKSAGRSRHTQCNDIIRALISADI PSIREPLGCCRTDGRPDGLTLPIYKQGRSLIW DFTCTDTFAPSYLSHTLRHAGTAAKLAETKXH KLYSELEKQYIFVPVAVETSGVWGM EGLKFIQ DVGRRMKDVTGPMSTTFLIQRISLAVQRGN AASIMGAVPSQRGLEELFYILNDRSFY LQKSC KGPNGDGLGRLSGSSKGFPPNPTSGFSSTPOA N				
	L2-22_NV	GCF_000 209225.1	Nematostel la vectensis	AAACGGGAGATCT AATCTCATCCTTGG CCTCAGTGAGACGT TTTTGGACGAATCA TGGACGGATACCG AGTTGGCTATAGAT GGTTACTCGCTGGT TAGAGCTGATCGC ACATACAACTCTGG TGACGGGCTGCTA GTTTTCATACCCAG CCATATTTCTTCA AAAGCGCGTGGA TCTTGAGTTTGAAG GTATTGAAGCCATT TGGCTTGAGCTATT ATTCCACCGATCGA GCCCGTGCTTTGCT TGTTTTATTACCGT CGTCTCTTCTTACC	TAAACCGGATATCCATTC TCATGTTAAATTGTTATTC ATGTTAAATTAATGTTTGT TGTTTGCACTTGTTTCTAT GTTGACGTCTGCCTGTTTT TGAGTTAGTTTTGTCGTT AGTTTTAGTTGCTTTATC AAGATGTAGTATTAGTGAT TGCTTTATAGTTTAAATTT CCAAGTTCCCTTTTGATT TCCGTAAGTGTCTAACATA GTTTTTTTATTATTATCAT AGGTACTTAGATATAGGG AGGGCCATATATTATAAAA CTGTAAAAGGTTAAATATGA AACCCTCGTTAAATAAAGT TTATTGTTATTGTTATTG	MFSSHASSAISQQVMVLGDLNFDLLERLPIAP TKQYLSIFDSLGYOQLITKSTRPISNLSLHIFVS SSDFVIEADTISNTASDHLPIFVAVAWKLKAGYTR PTGHKVLFRSRKNFCIESFMHDLISLVPWSTL DIYDDVNDALDQWYKLFNDILTVHAPIKEHRI KRPLKLDW/FNAEISAAIKARNCLHRIATSLKTD TSWSEYRRARNHVHVLIRKAKKTFYNDAIKN NLEPNKNLWKIIRNISPAKHVNLPNRLTVNDT LYEDTFDIANFFNEHFSNISQQVCNDAYPEFP DRSVLTNFINFKHQPGSYFIPPISEDFVLSNPS RLDFGKAAGLDELKVFFLKSAAPEIFRSLTKIIN LSISSCVFPDPWKSAPVPLPKDGSFLDRTNF RPISVLCILSKVLERHIHTSFYFLTQDLDLDFQ FGFRWFRSCELATINLSDSILKNMDSGLPSSG MLLIDLKAFALVDHNTLLLLKGLYGCDLTLK WFKSYLSGRTQKTCFKGVLNALPINVGVPQ GSILGPLFELLFINDIPLIFLSSDLVTDTTMFADD TTFLLISGASVPDIELKLNKATDVHKWATLNR MVLNLTKTKSMILGSHQKVSNLAESSLSVTLN SCPIEQVNIARVLGVVFDNHFTCEPHIELFKY INRSISLRSIAPELTPAGAIHVYNACIHISOEYV	cd01 650	374	636

				TCTGCATACATAGC TCTTTTGGACGAA				CASVWNGCSQTLQLLLRSQKRAARTILSADF SVPSVLSFRNLNWLYICDLIKQRKLWTLFSILIN PDCPLCLRSKFSFVSHERRTRSTPNLVLPOQ RTDAGKRTFSFSAISLFNLLPSLQQLAFSRFLN LNQLPNVTSIAKKKVRDIFHSHLNSVSHLEQL FCHACRFYIDRCYFS	cd01 650	515	769
Crack- 5_Ae	GCF_002 204515.2	Aedes aegypti		GGTATTCCAACGG GAATGTCTTCTTAC GTAAGGATGAAAG TTCTAAGGCAATCC ATGTGAGTGATGTT AGTGTCTATCATC GATTGTGTAATATT TCCTGATTAAAGTTA A	TGAGATGTCCATATTGAAT TACTGAAATGTAATGAGTT AATGTTAGTATTAAAGTAGA AAATAAGAAACATTGTTA GATTATATGCTTGCTCCCT CGAAATATTTTACTTTTCG AGTGGCTGGACACAGTTA ATTATGTAACCTTTTAAAGC TGCCTTTGTTAAATAAAGCA AAAAAAATAAAA			MITLKNYKAELINELINFCLEEIKLDKENLNI LYLNTRSCRNKEEELYQLINEFDGSHVVVFA ETWLKNEIFNLLGYDAYHGCRRDGRGGGVS FVLSDRSQLMLVINSEDNFLMVELIDQNV KIMGVYNPGRSFNLFEEQLEQILHNYKRIVCF DFNINLLDSSGELIRQYTYKIQSLGYIILNKDA RYATRISNTIATVIAHIITDIFEYDLQFGIKDTER LISDHRTLTLISMRIENHKKNPDKHKICVNYE NIIRSTQNTFEICQQFSDIVVVLNEIMTKNRT EILSRKKAMVIRKPYINNALLIAEQNKNELYRQ YRTAPGGVLKDELYLYARKRNMLKNKIFAK EQYGRKMHKNINNYKKTWSTLREIIFQNGS ENITPQALTKEGKIIDSAAQIAKCLNEYFINI GWNIIQQNGTGNDFNLILNNMPETRFESV VETEKAKNVINSKSGVAVGLNKISTKIQKCV TLIAPKLEIVNNVILSSQFEDCLKIAKIVPLYKS GDKFSKNYRPVSIIPALSKIPECIMNEQIKHHL LQNELLNKQFGFLPKKSTETASLELNFVKG LDDGEVVGCIIDLRKAFDCISHELLIAKLSFYG FTPSAVSLKSYLANRKQQCCQVNGVLEPLLK SGVPQSGIIGPILFINIFINDLFNIPKGLQCYA DDAVLYKRAKDETLQSMQHDQLQLPSWF SANKMQINNKTNFIFSLKTSNVSLDLFINNE RLSQVTEVDYLGLAIDKNLNWKSHTLVKNKI TPFIFALRRVRGCLSTKSCWQLYNFALPHLSY MINTVWGSAAASHNLNVLQNRVITKIRKLP FLFPSASLYSPEVLINLHKYNCLILVYKIKYKL VESNIRLTFEFDLHSHSTRARQNFVYSRPRTE LSLKHFIYSGITQFNALPNHLKLEIHLSSYKKKL KKYLFENN	cd01 650	515	769
CR1- 54_AE	GCF_002 204515.2	Aedes aegypti	A		TAAGTAAATCAGTCTGTAG GAAATTTATCCAAGACGTT			MWLRPDVNLINSELCSNYTIFRCDRSEQTSSFQ RGGGVLIAVNASFRCTSVSLNDCDSLEQVWVS	cd01 650	470	721

					GCAATAAATAAATAAT AAA	VNLPMISTIYLCGIVIRPTSPPEAYSMHATAIQS ICNFSSETDAIIVGDYNLPGLSWISDEELNSY LPCNASTEATTVEAVMATGLYHVNSIRNV NNRILDIAFVSDVCDIEVLEAPSYLLRMDAHH KSEVLRVQIWETRNNSSSDSNDTDFDSLCD FESLNASLSMINW/TNELANKGTDEATAHFYD KLYDVLHSNVPRKRLRPQKGKLPWW/SPELR RLRNGVRKARKRYRSHVVRQNLHVIESRY NKCKVFTFRNYINSIESNLKNNPNSFWSEVKS RKSKNRIPEQIRFQNRTSANVDESANLIFADFF SSVNSTISPTLSTATRHSIPTHDITLSLNFTEQ DVKTTLTSLNVKKGAGADRLPPSYWKECAES MKLPISLIFNRSADRKEPAVWKTASVVP/HKS GSINSVDNYRGISILCCIAKVFESEMMHMLYN ATRHLISDQHGFKKLTSTVSNLMCYTNFLSS EIERRQQQDAVYDFSKAFDKVPHDLAIKKLE QMGLPDWITELWLSYLSGRKAFVRIGDTQSR GFEDITSGVPQGSVGLSJLVLFINDSLHRLKSK LFYADDLKIYRAITSVIDCYALPSPDINELLWCS ENGMQLNNSKCKCITFTRSLSSYFNYEINGSII DRVTTINDLGVITDSKLKFSKHVRVTTAKAFA VLGFIRNSQAFQDIYTLKTYCSLVRTILEYAA CVWSPYCTTQVLHVEKVQRSLRYALRSLPW NDPVNLPDYESRCMLLNMETLSFRRTKLKQ MFLFDLITHNIDCSDLAEVGLAPSRYLRSRQ LLAPRTHRTAYGQNSPLTSCIRSFISVINDVDF NLSRVEEKRRISSIR	cd01 650	460	710
					TTCGAAGACAGAC GAGATGCGATCAA GACCCTCACTACAT TCAACTCAATCAAA AGCGCGTGACCG GTTGTCACAAACGCC ACCATTCGTGAGGA CATGTCCAAACCCGG AGCTGTCAAAATTT CGGGACTCGTGGA AGCAGGCCATCAA	Caenorhab ditis brenneri	GCA_000 143925.2	CR1- 1_CBre	

	LINE2H_C E	GCA_004 526295.1	Caenorhab ditis elegans	<p>GAACAACACGAA GCTGGACAAGGA TCTGGACTGTGAG AAACCTCGAGCTG GTCAGAAATCCCTA CAAGGACAATGAA ACTCTCACCCCTTG GACAGTGCGAAAC AGAACCAAGTAA GTTTTCTCCCTTGC CCAT</p>	<p>AGTGGTTAATTGC ATGTAGGCTAATGC TCGTTCCATTGCAT CGCTTGAAAAACTT GAAITTTGCCATAA GTTGAGTGCCTCTA ATAGTTAGACATC TTGTTGTTACCGA AACATTTCTAAGTA ATGCTTTTACTGAT TTCTCATGCTCTTCA CCTCACTATACCTG CATAAGATCTGATC GGCTTGAGTGCCAT CCAAAAAATGCGG CGAGGCTACTGCT GTGTTTTTTCATAA AAAATGGGCTCTTA CGCCTGTCTCAATT GCACCTTCCTTTTAT CTCTCGCATTTTGT</p>	<p>TGATCTCATTCTCTCACTCA TTTGACATGATGCATAT TTTTTATTTATTTTTTAT ATTTCTCATCAACTTTCT TGAAATCATTTGTATCAT CTGACCTGGACATGTTCC AATCGTGCACCTTATCAT TCACATTTCTTACCTCATCT CATTTTATCTCTCATTTTG CTCTAATTTTCAAGGCAA GAAATAGTAATAATAA ATTAA</p>	<p>NIPVNSNILKAELSDAFSKVFTVDNSSTLFP RTNNVLESMTFEPSEIEKVLLKLVKCNTPD EIPALFKKVGTSIAIPLSIFSEFRTSQVPLW KTAIVKPLHKKGPRTDPNNYRPSISLTSSVCKV MEKLVKRNLTDLNSQQLSFRAQGFGRSNM STESQLTSYFSTLISNLSLRKPTYSVVDFKKA DTVSFEKLTCLRMYSYGIKQQLDWIVCFLTNR TQRVNVNGVLSQEKVLSGVPQGSVLGPLL LLFINDIGKFSQSHLLYADDLKLFSNPTSIQT DLGLLSEWCCKWQMSVAPNKCENIIFNLTKS PRTLNDFTTIKGSKIPRTKKIRDGLILSSDL EHIDTLRSHQRINIFFNLRASFEFIFKCYTY VRPILEYGSIVFNPVLKEQIKRIESVQSFYRV FKKFIPIYLYNFDALSVCNLTSLRLRLHADLV FLYKVIVSKETLLYPPEFTLPSLPNIRKHKFYFRS NSKSNKFFSQFISNLSLNCWNLSPSVFPPT PHSINFSISYCNFQKYLTLPNPHNF</p>	cd01 650	309	557
--	---------------	---------------------	-------------------------------	---	--	---	---	-------------	-----	-----

				CAAAATTCTGGCATT AGACATAAGCAAT CCAAAAGTAGATT ATACTGATATACG CCCCATCCACTAC CGCAAGTAGAAC GCTGAGCTCTTTAA AAATCTAAACGAG TTCTGACTGCCCC TCCTATACTTATG CCTCCTCGGTGATT TCAATTTTGCAAC ACCAAGTGAATTC AAATTTTCTTCCA CGGGCTCTGAGTAT TCTATAATTGG		SKSPRFPYQMTAIGTNNNSFLHTLHTWNL VTKHFPVPVSGYTFSKRISCLPLESYV		
RTE- 1_CRem	GCA_010 183535.1	Caenorhab ditis remanei	CTTCTCTGTACTG AGCCTCTTCGCGG GAAGAAGATCAAC AGATGTCACACTTC TACAAACCGGG AGTTCGTACGCGG GGAGTGAACATC TGCCCAAGAAGTA CAACGTCTCACGCA GGTCAGAGGTAGC GTCGAATACTCTCT GTGACCGGGCGGC TTTAGCATTTGGTTT AGCAGACCATAGT AAAGTAATCAGAG ACACCCAAACCGCT CAAAACAGACCA AAACCTGCCTCAGT CCTCGCACTCGG CTCTGCAACTGTTA CGGTAGATTCACTA GCACGTCGAGAGA	TGAGCGAAGAACGGATCG ACTAAGTAACTAAGTAAGT AAGTAA	MDVYHSKRTKLTTPKTRKTDLRICTFNCRSVA TDNRLAEELLEETNRKIYDIIGLSETKRSAEVHM TNRDGTGVILGRNDSSVSGGVGFIINKSLMP KIKEIKVNHRIGYLTQVKNKNQYTIQGYAPTA DYEEEDHSNFYENLEDVYKSKRYKMVIGDF NARIGERKGNVEFIGPHSMENRNDSEGLAT FCEVNRVFMNSQFIKPTHRRWTYISPDQKH RHELDHILANGKYITDVSVPSFTNGSDHRM LRANVINAKQAKFEQVKRRKPPRRVLDPTA ALLATENLSDCEDQDIDKEYDTLVHVLKKAQD AAVTPTNHSNRNLKDSRLLLSKRFTDRSDP NFKTLSKECROAVKRDHENFAKDRLLNAANQ RKSLRKVARDINEYQSYIPCLLSGETGEKLSRE KMEIEVRRFYSNLFATKQSKATTIIPQQTALP PFLPEIQHLSNSFQNGKAAAGEDKLSDFKS CHSSVHKLANKFSRYLQEGKVPTKWSSKTT LIFKKGDKENLENYRPICLLPVLYKVFTKILNRI RKSLKLEAQPVQAGFRFSSTIDHIHSVQRLL VGREYQIPITMVIFIDFKAFDTIEPTALWESLK TQIGDSGYIKLLKCYNDCSTTTTTFYNPNIP TRGVRRQGDPIPSNLFSACLESAFRRMSWPHL KEDKDDYDNPSPGIRINGRNLTHLRFADDDILIS	cd01 650	511	779

				GGGATCTAATCCC GGAGTT			KTPQIAEKMLQELVNSCETVGLIEINASKTKVL RNKFASSHQHIRKNNTMSAIEDVEEYVYLGR LLNTKNDLEPEIHRRRRSAAALNNIKNTTD ALTCPKIRAQLFDSIVLPALTYGSETWTFKAL SERVRITHAALERKLVGITLLEQREKNLHREDI RKMSQVKDPLVFTITKRKLSWAGHVMRRNDN RWTTLLQEWIPRNEKRPVGRPPMRWADSLK KEISVRQGTQLIEPWSTIAKDRKKWIAVIRAH TN		
NeSL- 2_CRem	GCA_010 183535.1	Caenorhab ditis remanei	CCAACTCTCATCGT ATTAACCTACGGTA TTCTCTAGTGAG TGTAATAAAGGTTA ATTACGTTTCTCTT ACCAAAGAAAAAG AAATTTTGAATCCT TTTTGTGTAACCTA CAAACTGACAGAG ACCTATCGAATTTT TTTTTTCGTATAT AGGAATAGTCACTC TGGACCACGAAGT GGACAGTTGTTGG CGGACTTCCAAAGT TGAGAGAAAAGGT GTGAAGAGAGGAG GTCTAGAAACACTT CGGCTGTCTAGGA CCAGTTCTGAGTG GAAAGAGGAAGGT CTAGAAACACTTCG GCTGTCTAGGACCA GTTCTGAGATCTC TTGTGAAGAGTTG AAAACAGTTAGCTG AGGCTACTGTATT CTTGATAGCCCCG CCCCAATCCCCCTC	TGAATACCGTCAGATAAG CCCCAACATAAAAAATAA AGTCGGCGTTAGCTAATCA CTAAACCGGCTCCTCATTG GGGAGAGATATCATCCG GTGCTCTCCGTTTGGCG GTAGGGAGGAGTTGGGTA GCGACCCGGAAGTATGGA TGCCCAACACCGCAATCT GATCTGGCAATGTGTTTCG GATGGTCTCTGTCTCTAGA TCTGAAATAGAGCTCTGGC CTGAAGAACACACGCGCG GACCGTTGGATGCCGAC TTGATCTGGAGGGTGCAA ACCTGAAAGGGAAAGTTG AAGGCCGTGAGGCTCCTG GCGGGAACCTCCGTCTATA GTCAGTGGTGTGCCAACA CCCGACGACTATGACATA GTTGGAGGAATCCTGATC TGATAATAATCATTGTTCA TATAAGGGGAGGGGGATG GTAATACCCAGGGTCCG AAACCATCAAAAGCAGTA CTGACCAGCATAGTAGTG ATGAACACATAGACCCTG GGGTTCCCTGAACTCGACC CATCTGCACAAACCCACTT	MTNVYLKPVNDNQNTKTDGNSRNTMSNSQ CEMTWKPVARTYAQAASNPADDDKTIVLIG CKYNLLKLGNTPTQTTKRSPKPSRGGARISVV YTLTDELEITHREEGKITFAIDLPNKNNILCPLC RECTQTRGRSSFTKHMKLHVKEKHQLKATF YKCSMCNEVEPEKCGTKWIQTHLQKVHNY KYDESAIVVPVPPNTRQQIANELNNAAPFVDI RKPKANVVEEKTENGALLKFLTKSNKDEQE HSQSDSPNVESPEKANQALTDPKGNNSPS KSSMRSSQSASSICQEQIEITLSEDEEPPKAAR PKGINVWSLINEKGKDAVIDDMMMAFLKM RVENYDSVNIIDPLNYQFPARVDLVLQIRNL EDGKRWVPFICADEHWTLTISNGIAAFYDP TGRMSYIEELVNLGLIIPKEQDEQPRQRDS YNGCVFMKMAEAFIQDTEWEMEVEEEDV KNFRNLLEELKPNYEFAEKIKYYNSPGKSA QSRPTSSSQCAVCPTCSRATPMMDVGN MEVDPVPOQQETPKSREPEQDEGWKVVG TRKRGVWTERSPNISPEAKRQFTGPEIKVWSP GKFHPLVGETEEMEVTCDSPTTEPTTEPKVT PSLPAMRIVSPEVTKKQTSKKKGKYGKKKQQ TKKAQPPKGEPTKKAQPKGEPKALIEQVRTW FDQMKSQEQGGSNIQTLTWIADSLTAAIFK ANNGNKYLVDKITARCPPLNNEGEMATQTS RRTEAVKPKDRFVKESNEPLRIQYAKNRKATF NVIIGHKSARCEIDINVENHFRQTLKAQPV EKALINTVCSGIKKAEDVPDSIEGPISSEGEVKAILA KIKDTSPTDGVKYSDLKWFDPEGERLALLFD ECRQHGIKPSHWKEAETVLLPKDCTEERKKP	cd01 650	895	1156	

				CCCCCCCCCTCGAC AGAAATTTCTGTTT GACCTCTGGAATT TGCAGGAGTGCG CGAGAATTTTCGAA TTCTTCGCGGTTT TCTCGAAATTTTCC AGAAGATTGAGC GGAGAATCTTCGA GAAAGTGAGCCGA ATTTGCGCGGAATT TTTCGGGATTTTCA AATTATCAATTTT GTCGGAAATTTAT TTTCTGGCAAAATT TGATTTAGTTCAGG CGGAGAGAAGGA TTTGGTGAAAG GGTATTGATTTTG TGGCGGAGGAAC TCCCACTGAATCAA TAACTCTCAAAGGA GAACTCATCGAACA ACCTCGGTGACCT GAATCTTGGCGGA AATTTTCGCAATGA CACAAGATAACACA AATTACTGTTGAAA ACAAATCAGAACA AACTGTCAAAAAG AAAGACAAAAAGT ATTGATTAAACAACA TC	TGTACAAATGAACAAACT GATGAAGAGTTTAATGATT TCTTACATCACAGCTAGCG GACCTACCGTGAGGTAGA CTCCGCGCGCTGTAGCAG GCTCGCCATTG	ENWRPISLMATVYKLYSSVWNRRISSVKGVIS DCQRFQADGCNESIGILRMCIDTATVLRN LSCWLDLTNAGSVPHILRRSLAAGYPES VINIISDMYNGSSMRVTAEQKTQNMIEAG VKQGDPISTLFNICLEGIIRRHQTRKTGYNCV GNKVRCLAFADDLAILTNNKDEMVDVNLQL DKDCRSVALIFPKKPCASLTIKKSVDQYARIKI HGTPIRTMSDGDYKYLGVQVQTNNGGSESE SLTQIAAELQMVHDTDLAPNQKLDVLKAFILP RLQHMYRNATPKLTCLKFEFENTVMKSVKMY HNPIKGSPLYVQIPVKNGGGLGVMSPRFTCLI TFLASTLFLKWSDDDEYISSIHKKALSRTAKVM GLKTQTATLQEQCDYLNTHKAITKGGYNLFSR MNEASRTLNVNLGAPLKSMQFIPENGEIALEV QASENSQIKVFSKADSMIKLVTKLDLVKSAM LKNFLENKKVKSQVQVQLQHHPQSNKFVND GKNISSQKFEVHPARLLVLCNGNSYKGDHP KDCRRCGYECESQAHILQHCYTNFTSTGITQRH DRVNLRIHEVIKGRKNNDYDIMVDTEPGP TRERPDIIMIQKDGPEVLLADVTVPYENGVVA IEAAWVWKIEKYSHFIEYFARLGRVILPLVV GSLGTYPDTSNLSRLMLGLSDGQIRNLIPDIS MIALESSKQIYWRHIFGDSYRIVTELYCTKNKQ EARFGDEPMENNVQVSDRFQPKTREREKKSE EEKRRRSKKGKTWRGSKKQTDNRQSGKSY QNQGFQRSVGQGGSR			
CR1- 2_CRem	GCA_010 183535.1	Caenorhab ditis remanei	TGATTTTTTTTTTCTTCTT TATTCTCCCTCCCCCTTGCC CAACACCTATTTTATAATTC CTTATTCGCGGTTTCTCTTC TCAGGTTTAGTCTCGTGAG	MSAQSLALIDHLDYELPLKSCILAEITLTHLLES SIFNHPFSLSSLSKSLPERFPFTRILPHLQEPPLL NSDHFSVDFSLITTVSNHSSASLPTSPTRNEH RLNYKKCDFKSLNADIANVDWDLVLSISSPT DKFALFSKMLSDFIISHTPLSTPPKPHRTSFSIR	cd01 650	350	599		

					GGACTGTACTGACCTCCA GTGACATTTTCTGTATATT TAAATAGTTAAATAAAATA TCAA	QLKRARQRYSKLLKLPNPADQIFRLRTLIAST SKRIKSNMVSEERKILSAPNSRSARLLIKRTRT SLKIPPLYVNNELISSNAAKSSIFSSVFFSNNYC SPSSSLIPVNNSSNNISPSSELFIPWIENTLRK VPPRCGFTTHANFYIRNCATTLAIPLSIYSD SFLKSHVPASWKHAIIPKPKKSLTSPENYRPI SLTDPFSRVCERILCNKINILFRNISRQHGFL SKRSCSSLVHSISEYRLLSSSHKSLDVWYDFR KAFDQVDHQFLISKLDSEFGIPSNISWLTDFLS DRTEFSVKIDDFVGTGPSASIPSGVPQGSVGPILL FLVFINDLLKLADISSLCIAAFADDIKLYSHDPL ALQMIGIDLVESWASDNSLPLAHSKTALLRLG PRNSSHPYTGGLSDSDSVDRDLGLLIEPNLKF TRHINRAVALALLRSKQLLKSKFSNSPQFYILF KTYVPLVEYCVVYSPPPSSKLAHKLETPLRFF SRKILQRCNTPYCSYSDRLSQLDLFSMRHRL KAQLLLYNLIGASYEPSLETHVRLSSSSRRPM SLICINPNCNFFSTVPIWNAITINVPFLCPS EFNSLLVNNIARF	SSF5 6672	203	380
	Penelope- 6_CoFI	GCF_000 648655.2	Copidosom a floridanum	ACTCCTGTAGCCAA CAACAACAATGATT CAAAGTGGCTGTTT AATCTTTCTAATGT TACTATACCAGAAG AAGTCTATAATATT TTGAAA	TAGAGCAACAGATTACTG TAATTTAAATTTGCGAGTT CGTTGACTTGTACTGGGGT GGGGTGGCTGTAGCGTTG TGTGACTTAAATATACTCA TGTACAGTGTACACTTAC ACTGACCGATAGTATATAA GCACGATCGTAACTTAAG AAGTGGCTATTTTCGATC TACCATCTATCGTGATGCC TATAGTCGCTATTGTCGTG GCTATGTAGTGTTCAATTA TTGTGCGATTTTCTTGTA AGTATTGATGACTTGAAT ATAATTGTTTCATATTTTA TATTTTATAATTGTTATTT AGAAAGTTGTACCTGATG AGGGCCTTGTAAACGTGT TTACTGTACGGGCCGGAA	MIRETSYFTNMNTHSKSEFFKLLSLSHFVPT VTDKIRTNLLKMFDFKFLKSTIAQKFDNQKIIND INKTALFKHNKELLVTVDKGNPTVVLDRSE HINSVEEQNLNDKYYTELLKNLLNELIKSVTTF VKRWYDLGIFKDCFEFIRSEISKNDLSKCYGL VKIHMAGNPIRIVSAVNSPTYNLSKFTKFLN DNIPKPNSCIKNSFHIKEKLSNFIIPQHHTIISLD VWSLFTNTSEIILKSIKNNWVYLRNKNLSSQE FCEGVSLNSTVFQFNKFYKQIFGSAMRSP VSPVLCDLALRCLEEDCLKLSFFVPFVVRVD DILICVRNNSIETITVFNHNDRLHLQLTVEIPV DNEISFLDLRUFESDGSIIITWYRKPTHSGRYL NYSNHPPLSQKIAIVYTLVDYSIKLSDSKFHKE NLDISKILFNNQYPLNFINKYIKKRLTILSINTP LNRVVSNNRGLMILPYFQGFDSIKRYLYKNNI NVIFKIDKFSKFIKLGKDRTKHTDRCNAVYRI VCNDCDSYIGQSSRRLLKTRINEHNMNDCKHR REKSALVDHIAKTNHNSFNDFEVKILDTESNRFK			

						CGTTGTTACAGCAATAAA CGATTTTGATGGCATCACG AAGAGAGTCATTATTTCTT GATAAAATATCTGCAAT CACGAGAGCCACTGACGA ACTA	RLFSEMLCDSYRNNINRMHDTQNLKLCYKN VITEHT			
	RTEX- 15_Cgi	GCA_902 806645.1	Crassostrea gigas	T		TGAGTTTCACTTTGTATAA CTTTTCCTAACTCTATAACC TTTATATTAATATATGTTTA CCTCTGATCAATTTATATG TATATTTTAATGTTATTTTT GTCCTCATGTACCATATGT TTGAAATGGTTTTGAGCG AATAAATGAACCTGAAC	MAVTCRNCSTVDVVLASANVFEILEDLDIHDF CNLFSDVHNVPVSFSLSTCYLSRNISTGKNHSG VKLWCPDKADQFRSNIDVHAVKSIENYLNEIS NSPGDLCDNINIKVVKDICTVFTDAARDTFGP SHFSPCIEKHNPWFNRECKSARKQYHLAKS MHNKSTDDNKNKLKEKSIYKKAMDNNAV/K TYNTGTRKLMQLRSTNTKDFWNILNSSKRDS KSPVNIKDMFDFKNINEGSLSPDSDLHEPQ VDGSSNELLNCEINESEITDAVKKLKNGKAPG YDNIVNEHIASTLNTFLPIYKLFNAVFDGLV PEEWLVGIIPIYKNKGDPSTPENVRPITLLSCL GKLFTSILSKRLEYSEEIELHNNQAGFRKGFS TLDHILTLQFLSNTLMCKKKLFCAFIDFKQAF DTVWRNGLWCKLLNYGINGKCFKYITNMYK GIKSLIKMNGTVSDFFSCNVGVRQGENLSPFL FSLYINDLENYLVEKNIAGLQSFSTSIENELCVY IIMCIYKLLVLYADDTVILAESGDDLQNALNE FECYCKEWNINVIDKTVLIFSGRMKCRVF YFNDRVLEIVKEFKYLGIIFSRAGSFLKAKKHLC DOAQKAMYGVIRKIQFNLPTCQLELFDKM WVPVLLYGCEVWGFELKDSIERVHLKFLKHLC LKSSTPNVMVYGETGRFPLSVIVFTRMVSYW CKLSSTNTSITNLYRYLLNQNTENTTSTNPWID CIRHVFNSCGLSNVWHSQFLNIVNQKWST AVKQRLQDQFIQKWHADINSSKGVIYKIFT NFGYEKYLIDLP/LKFRKILGKFTSNHHLPEL GRWEGIPLNERFPLCNTKRIADEFHYLECD AISQVRNKFIDKKFSARPVNLKFSYLMTTCNM KLLRKLCEITKIVEVCCPP	cd01 650	324	600
	CR1- 24_CQ	GCF_000 209185.1	Culex quinquefas ciatus			TAAATCATGTTTTGTTTAA ATTATGTTTTGATAACTT GACTATATGATGTTTGTA	MEDPRLFTVVPVPSHSRNYAFCTSHPSRPGPV FGPRRGVFNRSVGKYPYAENDLVSLPDARL TSSNVSTNESIRQDSAYRSQDAASQQSNDNA	cd01 650	918	1167

AAGTCTTATTAGTTTAA GCAACCAACTTCATTGAG ACCACGGGTGAGATGAT TTCAATAAACAAATC	DPRPPTCYMEDRLSLVTVPVPHSRNYAFCT SHPSRPGVPFGRPGVFNRSVGYKPYAEND LVSLPDARLTSSNVSTNESIRQDSAYRSQDAA SQSQNDNADPGRPPTCYMEDRLSLVTVPV HSRKYAFCSHPSRPGVPFGRPGVFNHVLV GKPYAENDLVSLPDARLTSSNVSTNESLRPD SAYRSQDAAQQSTDNADPGRPPIRIMEDPD PDTVVPCHCCSSRSGHQSRPGVPFGSREG VFQTPNAGKYQAAAFQGYALDPDPRLSSDV QTTVPVHSDPTTPVHHDSRRSSAGQVIFEP RFYQNVRLNTKVDFFLSCDCDHDVILLT ETWLDGRVTSQQLFGNNFVYHYTYRNPANS KRKVGGLVIAVRNSLDSVECVQAVAEILEQ MFVTIRTPSKRIFVGCWYLPPEKRFETELMEK HLSCIENIRRIAGPHDIVVAGDYNQGSVWVK HLSGNRSYVDPQLTRTRTEKERLTHGVMLDG MANCNMHQCNLVKNDNRNRLDUFVDDDC TVAAADEPLVPLDAPHPALVFDIDASRTVFV QVDNNDRLNRYKRTNFDELQRTLADIDWTP VTSAMDVDTAVREFSNLLNFHLAQHTPQFRT PPKEWSNARLRLKKEAAALKFTLLRRCR MTKLAFNAASSRYRRYRNVYVIRKTQSQ LRRNPKKFWTFVKSYYKERGLPATMALGDST ATTNDEKCELFARHFSSVFREPEISSTHLDA VPRDLVDINTFVISRDMIEKATKKLSSFSFGP DGIPSAVLKRCANQLIEPLLVIFNLSLRQAIFPT AWKKSMFPIFKKGSRSIENYRGITSLCSGSK LLEIIVSDVFLFNCRSYVAPEQYGFMPKRSVNT NLEFTSVICIDKIAEGRQVYAYITDLKAADFRI DHIDIALQKFSKLGFSPLCRWLESYLGRIQ VKIGASVSELTNGSGVAQGSNLGSIVFSGFF NDANFLFSGNCKLSYADDFKLPIDTLADCYV LQSRDLKFSEWCGHNRLSSAKCYVISFNK QSRELFTFDYHLGSHVLERLEWKDLGVLLDL KLDFHLQLSSVIDKANRKLGFMFKAQAEFDD PLCLRTLYCALVRSHLETSAVWVWAPYHQNW DRIERVQRKFVWFALRLPWNPNPAQLPSYET RCNLLGIETLQNRRTISRAVFAAKVITSEIDSPN		
---	---	--	--

	Crack- 10_CQ	GCF_000 209185.1	Culex quinquefas ciatus	C		<p>TAATATTTTGTAGTTAAT TCTCGATTATCCCTCGT TTGTATTTGTATCCATCTTG AACACATTTATCTTGGTAG TTTTCTTTTATTTTCA GCCGCTAATCGCCAGAAA CCGAACCTCTTAAAGGTC TACGCCGATGGAGTTCTG GAGCCAGCCGCATCAACTT TGTCGTTACATTTTATTTA ATTTCAATGACATGTAAT CTACTTAATTTGTATTGCA CAATTGAAAAGAAACAGTTT TTTCAGTCAGCACTTTGCT GGCTTTTCTGTCAAAATAA AAACAATCAATCAATCAAT CAATCAA</p>	<p>LLFQLNARVQPRNLRPTPGFLSRPLVRTVVAE NSPIRALATAFNEFYHLDFHEPVSRFLERLRT ELDRSRSSRTAENPQQASRTRRSSR</p>	IPR00 0477	1	182
	Crack- 10_CQ	GCF_000 209185.1	Culex quinquefas ciatus	C		<p>TAATATTTTGTAGTTAAT TCTCGATTATCCCTCGT TTGTATTTGTATCCATCTTG AACACATTTATCTTGGTAG TTTTCTTTTATTTTCA GCCGCTAATCGCCAGAAA CCGAACCTCTTAAAGGTC TACGCCGATGGAGTTCTG GAGCCAGCCGCATCAACTT TGTCGTTACATTTTATTTA ATTTCAATGACATGTAAT CTACTTAATTTGTATTGCA CAATTGAAAAGAAACAGTTT TTTCAGTCAGCACTTTGCT GGCTTTTCTGTCAAAATAA AAACAATCAATCAATCAAT CAATCAA</p>	<p>MLDNRDDVLGLFLDLSKAFDVTDRKILDKLR YAGVRGIALDLFTSYLSNRAQFVCIDGHSLLT LVNVGPQGSVLGPLFFIYLNDFSLPLKGD RLFADDDSLFYNNKSTNMNDLNRDGLTIVIE YFRLNKLTLNINKTNIINIKSSRSIPNSLSLTS KFPDLEVVS DTKYLGIILDNRLNWSAHINALIL KLHKITGIIFKIKHKLPPQKVLFLYHSLFHSILSV TAVWGNACGLINKLQVAQNRIKLILNPIRS HTVDLYVKNILPVRGIYVFQVCSFYCLNNK THSNTKFVRSNHQYTRYHDLNLRPNVTTVP GERSINFNGAQLYNYFNRFNGKSSSLNIFKNQ LKQCLSRPDIEKLLKSVNIFA</p>	IPR00 0477	7	181

L2B-6_CQ	GCF_000 209185.1	Culex quinquefas ciatus	AATTGG	TGAAGTGGCATCCCATAG GTGTGGCTGTGAAGGAGA GCATGTCATGACGGTCGG CTTCTCTACA	MLAIEIHQQNGIFGGVHSPSSDRDFLDS FENWLQVMADDKTNVISGDFNICWNNEG YSRDLKNIADAVGLEQKVLEHTRISVSRRTIID LVFANVEDCTVKVLADWRITDHTLGINVEG VCKINMPEKKTEYTCWRKYSKERLQNTLLCDE TLWRNNLPVDEKAQQFSRALATAVGQLTET RMRRSADTKRWFTPHLQAVKLKRD EAYKIFK TTNCAEHWERYKQFRNEVRELQKAKNSSV QQEIQSCSGDSKKLWRCLDLIKPGGTPEKEII FDDKTEPCSD EETANRLNHYFIKSVKEIHDSIP TVNQPPQDPLPTVRTLDISEFSQFPISMEKL RKIVAA MNKAGVENISKRVLEDAMDVVGD KLLDIINSSNLGVFPQEWKKS VVPIPKVPNS TRAEDRRPINMLPIYKVL ETVVRQLVEVD RAGLLIEEQSGFRKQHSCESALNLLMKWKQ SIENKFLT VFDL KRAFETIDRRKLLLEVLRN RIKGDVWRWFKSYLEYRTQITRYNNSVSEAE AVELGVPQGSVLGPLLFIYINDLT KALRRACV NLFADDTVIWAGDKLDECYEIMNNELAVFA DWLKWKKLKNVSKTKYVMVVTTRSESNCIT SVDGELVERVNAIKYLGVM LDEKLSFAEHVDY TIRKAARKLVLCRINRYLSFDNKIMVYKTLIA PHFDYCASILELATNQQRKRMQIVQSKAMR MILRCDLTPRILMLDSLQWMSIRQRYEYSTL VFIFKVVNGLAPQVLT KTVCQGRDVRHFTR QAGDIRLLNFKKSC TQNSLFYRGYNLFNMILPE STRATTNLR EFKKLPFVRQRPLE	cd01 650	397	652
RTE-3_CQ	GCF_000 209185.1	Culex quinquefas ciatus	TGTGTGTGCGCCGT GCACATCGGTATTG TTTTTCGTACCGTTC CCGCGAAAAGTGCA AAATTTACACGAAA ATACGCGTTAGTG GGGTGTGAAATTC ATTGCATCCGTGTT CTGAACCCGCTCC TGGCCCGGAAAG TGATAGTCCGTGTT	TAGCGCAATAAAAGTAA AGTAAGTAA	MVLRDRGLVQATRRRTKTPVQEAHDASRT NRHGTGHHMRSHDWKLG TWNCRSLKFDGS IRILSDILVRKFSIVALQEVGWIGAEVQAYP RIGCTIYQSRGENKRLGTAFIVL GEMRDRVIG WTP LDRMCLVRIGRFFNIIN VHSPhSGS EDDDKDAFYEQINW TYNSCPKH DVKVI GDF NAQVGQEEEEFRPVIGKFS AHVRTNENGLRLI DFATSKNM AVRSTCFQHNL RDKYTWRS PQG TESQIDHVVIDGRHFSDIIDV RTYRGANVDS D HYLVMMKMRQLSLAKSVRYRRPPRLDLERL KLPEVASRYAHSLEAALPGE GELLEAPLED CW	cd01 650	569	829

TTACGACCGGAGT GAGACAAATCAATC ACGAGATTGACCG AATATCGCGTTCCG CCACGGGCAATTG AAAAAAGTACAA AGAACTGCTAAGG CATCGTCGTCGCG TCGCGATCGATGGT CGTTGATTGTGTC GCGGCGCGTGTGG AAGTGACCGCGA AGACCATACTGACG CGTGTCGTCGTCGT CCTCGTCACTGTGT CTGTTGCTACGAGA GTCGTGTATAGAA GAGAAGCCGTTGG TGCACGGTGGGGT GAAATTGCTAGGAA AGCTACAGAAAAA AAAAAGTGTGCAA AGTTTTTTGGGTGA AATTAAAAAGAG GAAAAATCAAGAA GCGCAACCCCTGCTG CATACACAGCCCCC CCCCCCCCCTCT AGTTCCGTTTCTGG GCGTCTGCACCCCA TGTTAGGGGCGGC CCAAACGGGACTAG GTGGTCGATTCCCTT CTATCCGTCGAGCG GCTGTTCCCCACGT TAGGGGCGGCTCG AAAGAAGTGGCCG ACACCCCCCCCCC					RSVKAAITNAAESTIGFVERGRNRDWFDEEC RAILEEKNAARAMLQYNLRDYEEAYGQRRR QQHQLFRAKVRHQEELEFEDMEQLHRSNET RKFYKLNKSRNGFTPRVEMCRDKNGAILTN EREVIDRWKQHFDEHLNSAEAEAGVQGGRR EDFIGTAGEGEEPPVTMREVKDAIKLKNK AAGKDGIAGELIKMGPEKLASCLHRLIVRVWE SEQLPEEWKEGVICPIYKKGDKLDCENYRAITI LNAAVKVFSQILFSRLSPIAEGFVGSYQAGFV MGRSTTDQIFTVRQILQKREYQVPTHHLFID FKAAYDSVDREELWKIMDENGFPGLIRLIK MTMDGARCCVKISGAESDSFTSLGGLRQGD GISCLFNVLGVMMRRAGFNMRGTIFSASN QFICFADDDMDIVGRTFKAVADAYTDLKREAE KVGLRVNVAKTKYLLAGGTESLRARIGPSVTID GDEFEVVEEFVYLGSLVTSNCSREIRRIIA GSRAYFGLHKSLSRKFSLHTKSYKSLIRPVV LYGHETWTMLEEDLRALSVFERRVLRITFGGV YENDGWRRRMNHELALQLYNEPSIRKVKAG RLQWAGHVARMPERADHLSQRNQKINPAK LVFVSEPVGTRRRGVQRRARWVDQVESDLES VGAPRNRNNAAMDRAACWRIRVQQAKLMV	
---	--	--	--	--	--	--

[illegible]

Rex1- 20_DRe	GCA_008 692375.1	Danio rerio	ACGTGGC	<p>ATTGTGCTGAACCTTGAATT AGATATTGGATATGTACAA GTGTTTACTCGTGTGTTTT GGAACCAAGGTGTGCAGT ATGAAATGTAATCTTATTA TTGTTTTATATGTATGTA ACTTATTATCTATTTTTAG TATGGAGAGCTCTGCTGT GACATGTGAATTTCCCTCC GGGGATTAATAAAGTCAA ACTAAACTAAACTAAACTA A</p>	<p>MMAPYLWQPVAALDVFNFVLLYFLNIFLQ WLRTLYLHTFQFESIRKHWIIFILIGKSDA VNSPAIADAVIVSREQLALRHSTFSVPTME VPTELKRRRRGCRAGVKRRLKKRYKPSLPVII MGNVRS LGNKTIELTALVRTQREYRECSFMC FTETWLN DKTPDSVVALDRFHLVRADRCAKQ TGKNKGGLALFVNERWCRPAHITTKEVVCN KDIELLAVGMRPYLPREFSVIAITVYIPPSAN ASNACEHINTVASIQLKYPHSLIVISGDFNHV SLQSSLHTFTQVDCGTRDCKTLDMLYANVK DAYASTPLPLGRSDHNLIVRSVYVPAVRKQ PPTIRTVQVWSDSVVEALMDCFDSTDWEVL CKSHDEDISLTDCTIDYITFVLRTPQPKRLN ATQTI SHGLYNFCVENTVPTKKVKCYSNNKP WVTPELKALLNEKKRAF LAGDREEIKRVQKEL KIKIEGKVIINLKNINYNKIMLEMYGVIVCT KLEVTIEQVRNGLRRIKKKSGPDGISAGVLK ACADQICWVWHYFNLRLRVPILWKTS CV VPVKKPNPTFESHYRPIALTAQLMKVFERLV LNYIKLLIASEDR LQFAYKTGVGVDDAIYLLH KSLSHLETSGNTVRITFFDSSAFNTIQPALLKV KLENAGLQKGMVSWLTDYITNRPQYVRMH NCVSQVIMSSTGV PQGTVLAPYLTFTYTSDFQ YSSKNCHLQKFSDDSAIVGYIRNHDESEYRLL KSFADWCDMNCNLNLTSTKEIIVDFSRHPS QAVPVIIQSEIEIVSSRYLGVYLNKLDWSE NTIQVYKKGQSRIHLLRRLRSFGVCQTLLKTFY DSVVSSVILYAVTCWGGGLEKEKNLKNLIK RAGSVVGGALPTIEEIAQDRMLDKFVSIMNH DSHPLYDTVQACASSFSTRLIQPRCYRERYKS FLPTVIRLYNQSHTR</p>	cd01 650	578	828
L2-6_DRe	GCA_008 692375.1	Danio rerio		<p>GAAAAGAGATCATATCTCC CCTGTCTTGCCCTCATTTGC ACTGGCTACCTATAAAATA CAGGGTTGATTTTAAAT CTTCTTTCACTTATAAAGC TTTGACAAATTCTGCCCG GCCTATGTCAGCGAGCTTA</p>	<p>MLTQPVSFSRVFHENSALFNSYTNFLSNAN ASLPSCSDTEQLAETFNSACD FALNEVAPLKS KRSRVSSIPQPVWLNSSTRTLRQECRAERRW KKDGLQVSYQILKDCLNSYQDSVSEAKSKYFS NLIAKSASKPKVLFKTNVSNLPTMCSVPVWNL ETC IKFSDFFAQKIQDIKSHLFPDTDIDHSDSVPP SNSFTNFQPVSSSQLNTISQMKLTFPCPSDVL</p>	cd01 650	260	521

[illegible]

	REX1-7_DR	GCA_008692375.1	Danio rerio	<p>TGCGAGGCGCA GCTTGCTGAGCAT GTTTGACTGCTCT GCTTGCTCTAGTG TTTTAGAGTTAAT TTGTTGCTTTAATT TTAATTTTTGTTTT AGTTATTAGTGT AGTAAGCGTGCCCT CCTTAAACATGAGG GTAGAGAGCCA TCTGTGTTTTTTT GGTTACTGCTTTT GTTTGAGCTTGCCG CACGCTGCTAATGC TAACACAGCTAGCG GTTGTCGGTTTT TGTTATACCCCGG ACCAGCTGATCTCG TTGAGGCAGGATG CTTCGAGCACAGTG CGGTCTCTCGGTA TCCCCGCGGTTAC GGAGGGGGAAGAC TAAGAGAGGCTGC CGTGAGGACAAG GTGCGAGGGA AAA GAACGAGGAATGG GAAATTGCTGATG AGCAGCTGAGAAG ACGGTATAAACCTA ACCTGCCGCTCTGTT GTTATGGGCAACG</p>	<p>TAAATTTATGCAATATCTTG CTGCTGCGCCATATATTA ATAATATTACGGTTTT AATCGGTGAAGCTGTG CATTATGTTCTATGATG CAATAGTCTATTAGTGCA ATAATTTGACGTTTTTTTAC TGGGTCTTAAATATTTTA GTGCAATATCTTTGTTTTT AAATGAGTGCTGAATGAC TTTACTACTTGAATTATTGT ATGTGTAATTGTTTTATGT GTAATATTATTGTTTTGG TTAATATTATTTTCACTACT TGATATATGACTTATTAGT ATTAGTTACCACTGATTCT TTTAAGTGTAATATCTTAT ATAATAGTTTGATACTTA TGATTGTGGCTGTTTTCT TTTATAGATTGTTGGATT TATGACTTATTTCACCT ACTTAATTATTATTATTGT CTTGTTGTTTTGTTTTCT GTGTTCTGTGTGTTATGCG AGTCATTACATAAGCAATT TCCTGCGGGATAAATAAA GTTTTCTAAGTCTAAGTCT AAGCTTAA</p>	<p>VIDNMLTFQPNTOAVYKKVQQRLLFFKLRS FQVSTSMMTLIFYKQFIESVLSFCIVAWYGNLS VSSKNRLSSLVKAGKIQFNQTLGLMVIYQNH VIRRAQGILCTPDHPLYSVQQLPSGRRFRVPI CRTNRTNSFIMVAIRLSNESRGEQDVSIY</p>	cd01650	491	742
--	-----------	-----------------	-------------	--	---	--	---------	-----	-----

				TGAGGTCACTGGA AAATAAGATGGAC GAGCTTACGGCAC GGGCAGATGCAC AGAGAATTTCTGTG AGTGTAGTTTGATG TGTTTTACTGA						
Tx1- 49_DR	GCA_008 692375.1	Danio rerio	C	TAAAGGGCTTTTGTAAAGAC CTATAGCACTCAACTAATT AACAAAAATTTTAAATTTT ATTGAGACATATTCGCATA TTAATATAATGAGTAAAT AAATAAAATGAATAGTAT AAATGAATAAAATGGGTA TTAAACAAGATATAAT			MYMQEKEKGGVLIGLDLEKAFDRVEHDFL AVLKEFGFGENFRKWIQIFLYRYFNVLVNVF FFLTECLQPTRSIRQGCPLTALLTLEAEPLGLA IKSDRSIKQIENNDNREAVVQYADDTTSLGI LERYCKASGAK/VNGKSLCMRVGRVKEITQH IPFKEEKEHTKMLGIWVGMDSNKADILNWE RVNLNGIEKRLRFWKARNLTLLKGKYSKFDVKR KVFIINALMILSKVWVYVLSVTALPEVYVEKIND VVLFKLGNGRAKIAYTYTLIGEVKEGGLDLD SFLKMKTLTIKIIQRFLEKMGVWKSVDLYL KKGCHFNMGCVLLMKMKKRMIAGLPKFY VEVLESGLRIMRGIFRKCVDYTKSKGADSRIT TFTELNNFKTILSEGNVLYFKEWWDGGLRQV KDVLYEVIKGFVPLKVIIVEDGYKRVVKAKLQR QCHELKAIPKDCIVIEKKKSWAMKMNQYQR PILRRKMERSMSLCTVKSFYAIFR	SSF5 6672	8	159
L2- 42_DRe	GCA_008 692375.1	Danio rerio	G	TAATCTTTATTGTGTTGT TGCTGTAACTTTTATTCC TATTATTACTTTATTGTTA TCTTTACTCTTTAGAGTG TGTGCTATGTTTTGTTCA GCACITTTGGTCAGTGCCAA ACTGAAGTAAATGTGCTTT ATAAATAAACTTACTTAC TTACTTACTT			MAPVCLSACRLPRQFLNVCSATILSVLFLLLFM SCVDALISYDRQALLNIRGLLENQPRDCSITHD CYSSFPPTAAFVRRLLPCLLYQKRKRRRGN RGGVRVRRMRKEISISLALHTACSFVPTYGHRG LHPARRSWDVRYTCHLPFPDKRLQAEFGAP PRVIRRRGVNLQHIKSLDVWHQSDLIQVISHI NMALLNARSVNKTFLINDFFTSNLDLFLFT ETWIGAGESSVFGECLNCSFISTPRCIGKGG GVAMVFRNWFRCTISVDGFTTFESQCALIE SDTTPFLCVLVYRPPKPDGRGFIKEFSEFVSHIT SYDRLLILGDFNIHVCCPGKPLVTFEMHVLD FGFTQHIVGATHALGHTLDLIMSYGFSIENIVI EDASFSDHKPIVFNNTLSYFSGTTKSEGFHSRC INSLTANQFSTMFNSNKVSTAILAAAEHSSSP DDLISLFNSSCSLDLDCVAPLKYYKPKMKISQP	cd01 650	673	932

						WLDLRSRLQICRKAERRWRKDHILVVSLEIF RDAQVNFQAAKKARATYFSDIINKHSHTPKI LFTINAIISPTSCKIEQSSEICERKFRFTEKV ADIRSSFAPLCPSSCVDSTVSSATFCNFEPVSL SELRLGVFQTKTSTSPCDPISSNLLKETFNTIGS SIQAIINSLVSGTVPAFLKHAVVQPLKKPNA DINCLNNFRPISKLPFMSKVLERAVLSQLQPFL KSSHKLEPLOSGETMHHSHTETALLKVWNDLL LSVDSGSAVLVLLDSSAFDTV/DHTILLSRLE HWVGIGSALQWFQSYLKSRSFSVSVGQFM STSAPLLCGVPQGSILGPMFLNLYMLPLGNIIR KHNSHFYADDSQLVYPLKSONSLQSLINCLE DIKCMMAKNFLQLNNDKTEVILFGPSKTRMS YIGGLSLAPYVKSQVKNLGVIFDSELCLKQI SSWKNSYQRLRIISRLKSILSFQDLEKVIHAFIT SRLDYCNLSYVGLPQSSSLARQLVQNAARLL TGTKKHAHISPVLASLHWLPVNFRIQFKILLIV FKALNGQAPSYISELINLQSTPRSLRSSNKILLQ VPRSRLKLGDRFAFAVAAPRLWNQLPPDITS ASSLSIFKSRLKTHFYSIAFPGF	cd01 650	526	803		
						TAGGATACGTCACTCATGG TCAACACTGACCGACGGA GGCCTACTA	GGTTCCAAAGTTGT CATAGCCGGGGC GGTAATGGGGATA AGCTCCCACTATCT ATAAAGTACCCTA TACGGCGTGGTCT CAAATAGCCTCTGA CAACCAAGTCACG TCCTGGCCTTCAAG TGTGGTTTAGCTAC CAACCCGGCGGA ACTGTTTTCACTGA CAGGAGAAAGGGC GCAGGCGGGTCAC TGGCGCCTTACAAC CAGTTGCTTCGGG GAGATGATATTCGT TAGCCTGGGAAGG	Danio rerio	GCA_008 692375.1	RTE-1_DR	

[illegible]

[illegible]

AGCAACGCCCGCCAATA GCCGTGCAGCAACAGCCC ACTAATACGAGCACAGA CGCCGATCCAGCCGCCAC TGCTCCGATTACAAAACA AGCGACTGCGTCCAGGCA GAAACCCGCTACGCAGC AAGGCACATCAACACGA CGCTCAGCCACCCGGACC AGCCGTACCACTCTGGCC ACCAGACCAACCCCGA CGACGCCAGCAGAACT CGCGCGCAGCATCGCT CATCAGTATGACGCAAC CACCTGCATCAGCCGTAC TACTCTGGCTCCAGACCA AGCTACAGCAACACAGGCA GAAACCCGCCAGCAGC AACGGTTGCCAACACGC ACACATGATCTGGAACA GCCGTGCCGACTTGGCT CCAAACCAACACAGGCG TCCAGGCAGCTACATCGCA GCGCTGCAGCAACGGCCA GCCAACACGTGCACAAG AACCTAGAAAAGCAGCCA CCATTCTGACCACCAAGAC CAACAATGGCATCCAGGC AACACCCCGCACAAAGC ACAGGCCACCAACCACT GAAGAAAAACCTGAAAG CGCCGCCAATGGATTGAC CACCAACTCCCGACGAC GCCAGCAGTAACCCCGC CGCGCAGCAACGGCTCAC CAATACGACACACTGATAC CTGACCAACCGCCACTGC CCTGGCCACAGACCAAA	TGDGTSAKDWYSWTKSSEKTEAFTRHLEER FSPKLTNSESFRAHVISKKEICTQKLLMTTPC WMHRENYGHNSDKTNLPFTPNEVLAEVK LLQNKAPGPDQIDNKLKLLPHKATLHLTHY NCMLRYSYFPKWKRAAITMIKPGKTPHEV SSYRPISSLPGFSKIFETLLMHRMFKIQDFERAI PSHQGFQRKHSTEQQLRVYTHILSAYEQK MYCSAVFIDVSEAFDRVWHDGLLLKLKLLP HSLYKILQSYLTDRSFTMNSTNGIPSRGKISA GVPQGSVLGPILYTVFTSDMPLPEVISSQHTY NPSYFENMMLSTYADDTIILCTKREPQLAIQT TESFLKYITDWATSWCITLNPNTKTHVLFTRRS VANLTGNKSPKMKGILIKNEPKHRYLGLIDR KLLHHHINQLQNRIRNVAAAGLWIIIGTSKL SPQSKILLYKQIAPWQYALPLWGALTSNTQF QRIEAAQSKLIRKILKAGRYTRNQVLRDAYNIK SINEIYSEASNRLATSLASHPNTTEARKILMDPV TPNRLEMPRYLDQLNRFPLRQQLQQNQ LNTHQPALPTLSIEQQLLSQQAASIKARQQDE EREQNKRLKIEQDPHYRPSIGWVNVAMSYKLR HGITTRWEMADKIKYEPEYIQRUMSPLLPSEI PPPTPDQELDARLSTLQIINDLDLNLNYETN LAEGDSRTRFLPNSSSPILQAPQIPICINLVEDH PLGNPAIADIWKSTPEI					
--	--	--	--	--	--	--

126

TART- 1_DWi	GCF_000 005925.1	Drosophila willistoni	AATATTGGCTTATT TACTTAATATTT GGTCTTTATGTT TTTCCCACTAAA GCTAATTGCTCTAA AACTTCTAGCCTAT AAAGCTAACTCAAC TAAAAGCTGACTGC TCTAAAACCTTTAG CCTTTAAAGCCAAG CCTTGCTGCAATT GATGCGCCACAG CCAAGCTAAAAAC TAGTGTACTCAGA CCCAAAATATGGT CTCCTACTTTTTTC TCCAGTGTGGAGG AAAACAGTTTTTT CCCGACATTAAATC GCCACGTTTTTGTA GCCGCTTAGCACCC AAATTCTTATTCGC CAAGTTTATTATTA AAAAATTGCTACCT ACTTGACTACTTAT GCTAAAAGTAACAT AATCTTAAATCTAT CATACTACAACTA TCTTACCTAAATCG CTAAAAACTAAATG TTAATCTAATGTCT TATTGTATATTTTT ATTAGTATTAATC TAGTTTTAAGATAA TTTTAATCCCCCG CCACTGCTGCTGTA ACGGCTTTTAAAT TTAAATTTTGTCTAT	TGAGGAAAAAATCAGGGA CAACCACTTGCTATTCAAA GGCTATCTGCCCCCTTG GGCCAAACGAAACAAATTC AATACACGGAATAATGTAC TACTCGCTCTAAACATGGC TAGTCTGGAATCTGAGTTG CTCAAAATTAATTGATAACA GCACGAATAGCACTGATG CTCTAAAAAAGGAAAA ACTGATACTGCACCCATGA TGGCCTAACTAACAAATCA CTTAACACCAATACTAAT ACTTCTGCAACATTCACCT TACCTAACCAATAAGAAAT ATGATTTTAGCGGCCCT GCAAAGCCAGAAATCCTGC CCACTATTGAGTGGCGTC CTGGCTGGCCAGAGGGC CAGTCTCGACAGCTGTGC CTCCAGTAGACGCGCCTCC TGCACCTCTGCTCGTTTC TCAGCGTCTGCACCATCTC GTCCATCGTCTGTCCAGA TTTCCGCTCGCCTGATGCA CCAATACCTGAAAAATAAAA CAAACTAAAAATAAAACC AACTAAAAAGTACAAATA CTCACCTGAAAAACACAA CAGCAACAGTAGCTCCTT GTCCAGCGTCTGATGCATC GTCTCATCTGCTGTCCCA GAATCTGCAAGCTAGTC CGTCTGAGCTCTGTATC TGAAAAATATATAAAATTA TTTTCTAAATCAGCAATAA AACAAAAAATCACCCTTA AAATCTACGTACAGAGAGC	MNILNKEISLKNKHLNKLKIDSSDRYRPHK LWQLTNLKRQPLPNWPLRIPSGVDKKHAW TRSPGKMEFTDQLEKRFSPNLNSSLDRAK INEETNALAQKMLMEYACKTHHSIASSANF DSSNSPFVKPITLSELKKEIDSLINKKAPGYDRI NNALIKVLPKAIIFLLKIYNSILRFGHFPLVWK KAIVVMVKPKKTTSDASSYRPIISLLPAFSKILE SLIAKRLKFNRLSQSPNHHQFGFRKFHGAEH QLARVTQFLQRAFEDKLYGSPIFIDITEAFGKV WHCGLRNLIKLLPINLYKVLDSYLDKRTFSVR GNHGLMSRDGKILAGIPQGSVLGPLLYNISS DMPKPIEHEHETLIEYPTYKHGDTLLSTYADD TIFISTDKNALVAMNKNTENYLHKFSTWSNKW CFTVNVSKTVHVLFSRRSTSPAVKAATVKFNIN KPIKNTSSCRYLGLV/DSKLLFKTHIAKLCAQA KSTYTRYKWLGGGAKSHLSTRCKVTYKQILVPI WTYAIPWGSUSNTLFFKLTGTHNKLRLKLTN ANIYCPNSEIQEKHNKAPLETIVNDACLFFNS MSDHETKEIEKLYLKPFPKRLPWPRYSKQIDK LVASRTIKKAQTPPLMPPPLNPSVTVTNKLRIE SKLMEFFSKAPALIAADNCHQTYNRITEKQDVL RAKEASIAKRNKQKIQSGH	cd01 650	195	471
----------------	---------------------	--------------------------	---	--	---	-------------	-----	-----

[illegible]

					AGGCACGCGCAA ATACGAATCGTACC GAATCGGAGGAGC CATACTATAAATA AAAAATTGCACAAA TTATTAACACAAA AAATTAATAAAAT TGATATAATGCCTC TCACGAAACTTTGA ACGTAAAAAACAT TAA	GCCGCTAGCAGTTCCAAC GGAGCTCCTCTGCATCTT CTGCTGCACACAGTGCCCC GTCATCTGGTCTAAACGC CATAAAGCTGCTCTGCAC CTGAAAAAATAATAAAA TGCTAAACTATGCTACTAA ATCAACAAATTACTCACCT GCCATGCAAAATGCCGCCA GCCGC			
CRE- 1_FoCa	GCF_002 217175.1	Folsomia candida	GGTAACCTACACCG GAGATACTACTAT CTATCGTTGGAAGT CCCTCTACTTTCACT CTTGAATTCGCCCC CCGCTATATATTTT AAGCGTGGAAGGA AATTGCCAGTATTG TTGGATACAAAGTCG TCAGACCACG	TAAAGTGATTTTCATAGTTT ATCAATGTTTATGTTAATA ATAAATAATAC	MPLFKSLDIWKYVNLTLLPPLPSSSSPPLPPS SSPLPSSSPLPPSSSPLPPSSSPLPPSSSPLPPS SSPLPSSSPLPPSSSPLPPSSSPLPPSSSPLPPS SSPLPSSSPLPPSSSPLPPSSSSSRPHPGPRK KRNKQKRTGPCKPPPTSSRPLPHTEGLNP VSDDLFCPVSGCGPSSSATDEWIECSACLLW YHTSCLLLQAIPEDDWFCANCPQHITQJGPQ VNSNTPSPANQSTLGHSHSQIIQSSPCTNPRAS QPISLTSLPSNFISSKATCTSLTLEWLNNNIYP GGIEKNLLHEAYRQDSLGIKPLNDAEIGKLK NSFSNIRTKRLRTEFHGASKYHYVGIQLKSEQ VPETVFDFKSISSLRQNCPTLKQCPKACRPV AQSIKSAIDKCITENTPEAWHLLTFPVYVTLQC DSRANIKNGQSLNNKIRNNLSIFESLDKNTYQ DYLISLSKNSVFDKSGSKINDIGPKVISKVSEG NIRGAIRLLTSDDVAPNSSATLDTLVEKHPQ QSLDNFTVLDEPNIPPFTTTTTEQVISAIFSPSG SAGGLDALRPQLIKDLVSQHTGDVGTQTLLES TKFCNFLAGKVPDICPFYIGASLTALKKSG GIRPIAVGCTYRRIVSKICVKVMNHVFDQDYFS PVQVGFNVKNGAEAGAHAARRFFSFHQSP TAFLLKIDFNAMFMVDRGKLLSSVHELVTPLY PYINQCYRFSSSLVWNSSLIDSCRGVQQGDPL GPPFLCLTLHPVISNLGSEFNMMWYLDGDTLG GNPTTVLDDFQHIISCANEVGLNELNFDKCEVS VLGPQTNTOCILESFSQSVAPGISCMQTENAEL LGAPLTLESIEHFVRKKIHAFNTNMAGRHLHSIPS	SSF5 6672	619	816	

[illegible]

[illegible]

CR1-6_NVi	GCF_009193385.2	Nasonia vitripennis	T	TCAGATATATGTATCTCCG TATTATTATTATTTGTAC ATTCGAATCTAATGAAC TCTGTCACTTGTGTATATT TTTCAAAACTTGTATTATT ATAGTATCTTTCTAAAGGA CGCGTCAGCCGTTAAATTA TAAATCAATAAATATATA AATAT	<p> RKYGLDRLPKPLVFSQVAPWTVSPPIVDTSLA SRLDKRRDDPCIAIRMIALTHMDARWSESIRY TDGSHNPDDGTTGLGIVPSSSTRISKRLPNN CSVVTSEMLAILYAVKYSAGHYHSNVVLSDS LSSLTILNDSSNSQAEVLDEIROACSLVHSQG KSIELAWIPSHVGIQNEIADELARIGLGLSEID RIDYSLSEIRSMITKSIKSTWEATLIDANAFRR LNDSSSVFLHHPNRQICRTVARLRMMGEV RNHKVELKLCPCGEAGTAEHLIHCEQLREL KSLVEIMKRYRLTYTLKNILSHTGEGMSTICET VYNFLESSEIRMIT </p>	cd01650	711	962
-----------	-----------------	---------------------	---	---	---	---------	-----	-----

[illegible]

	L2-8_OL	GCA_004 348135.1	Oryzias latipes	T	TAGTGGGACTTGTTTCATTT TATTAATTTAATCTTTTAT TGTTTTTGCTTTTATGTT CAATTGATTTTATGTTTTA TTGTGATTGTTTCTTTATTG CATCATATTGTTTGTAIT TTATTGGATTTTATGTTAA GCACTTTGGTTGCCTTGGC TTTTATAAGTGCTATACAA ATAATAAATTGATTGATTG A	IPR00 0477	3	89
	Rex1- 5_OL	GCA_004 348135.1	Oryzias latipes	G	TAACCTTTGCACATTTTTT TATTATCTCTTATTAGTT TTTTCTATTTGCACATTTC TTTTGCACATCTCTTTACCT GCACTGTTCTGAAATAGC TCTTTAAATTATAATACCTT GATGCTGCATTTATCCTA CATTTTCTGTTCTTATTT TATTTATATTCTTACTC TTTTTTCCCTCTGCTGTCA GTTGTTATGGTAACGAACA AATTTCCACGCTGTGGAT CAATAAAGTATTTCTGATT CTGATTCTGAT	cd01 650	51	301
	RTE- 15_PSi	GCF_000 230535.1	Pelodiscus sinensis	CACTACACCTTTTTT TGTC AAGGAAAC ATAGTGATGAGAA ACGCCTGCATGGA GTTGGTTTGGCCAT TCGCAACTCGCTGC TCAAGTCAATAGAC	cd01 650	445	722	

[illegible]

						RNAFGSVHQDMLWVYLLRLGVERAFVERCD DIYEDSVFVGNADGATEPVRQEVGVYQG CPLSPLFLIAALVPLRLALEKLDGVGVADGV RPCTTAYADDLKVFSDSAAGIARCHAVVEQFL EWTGLQANPGKCAFLAVTRNARGNPAHDK DMKLSLHDEEVSSIKLHDSYRYLGVGDGFDH VRHRLQLEPKLQIKREAVALMQSGLATWQ VWKALKTYVPKVEYALRHLRPLQSLQSFDR VWAKGLRHLRLPRSATNEVLYAPTSSGGGLG QPLVEMHRAIQ/AHAWQMLHSHKDPAREVA RAQVCQVARKRHLREEHWRGRDDELVRCE LNSELAASPHAEALRRHGDIGSLWSDVQRWL RIYHLSLVQDDRNGLDPLGLRVPHHAKWLD HKSVLRHVKLHLKIRHQTRWKLADQDGKTVR AHGGVGAKFVSTGAGLSDDDYRFGVKARLN QIDTNAVLRKRRLRSHKTCRDPCTSSAETLAH VLNHCEPNMIDAIRQRHDDALEQIGSKIRNAL KRGESTAEIRLNQTVPEYTGAAALRPDIVLRNV AAKMMVIADLAVTTEEHAAGARHSSLQLSHD HKTLYQPIVAELQLKGWQVQQTAAIVYGSIG SVQPSNSTPTRRS	cd01 650	912	1175
					GCCCGCTTCCCCCG CGGCCACGACGAC AGCCCCGATGCG	TAGGAACCTAGCCGATTG CAGCAGCACCTGCGCCCC GCCAGCAGCACCAAGCGC CCCAGCCAAAGACACGAG GGCAGTCGGGTGACGGAG ACCGGGGGGACTCCGTGCG CGCTCCGGCCGCGCTAG ACGGCAACGAGGCCATT GCGGCCGACAGGCCACA CCCATGTAGGGAACCGCC CTCGAAACCCCTGCACCG GCTCACATGGTCCGACCCC CCGATTGGTTCAACCCGTA CACCAACTACACCAATT AGGGACTCCCAAGGTAGTC ACTACCAGGACGCAACGG	Phytophth ora sojae	GCA_009 848525.1	Utopia- 2_PS

[illegible]

[illegible]

[illegible]

							ESVQRRTKIPIGLRNLTQERLTCLKLPTLVH RRKRGDMIDVYKYVNGLYTVQEMPFKMSTE DRTRGHSRLEKKYSRDLIRKYFFSNRVVELW NGLPENVAAGSVNSFKNRLDSHWKSPV SLLYEPV				
Crack- 8_SP	GCF_000 002235.5			Strongyloc entrotus purpuratus			TAAATATAACCAAGTACT CGAAATAAGTAGCCACT ATCAAACTTCAATATT AGTAAATTACCAATTGCAG ATCACTGATTTCTCATAA ATATCCTACAAATTACAAA TAATTATGTTTAGACACAC TGTCACCTCAATAATAT TTTTCTTCTCAAAATTTG TTTATGATGATTTCTTGAT GACTTGATTTTTTGGATT TTTTTGTGTTACTGCTGAG TTTAAATTACATTTGAGA TTTAATCTGTAATCTTTCT GGTATGTTTACTACTCACC ATGTGATGATGTTTTTCATG TATATGATGTCCTTATATA TCTATTTTATGTTATGATG GGCCCCAAGAATAACAG TATTTTAACTACTGGTGGG GCTACCTGTTGAAATAAA ACGAAATAAATAAATAAA TAAATA		cd01 650	26	284
Daphne- 19_SP	GCF_000 002235.5			Strongyloc entrotus purpuratus			ACCAGAAGGTGTT AGCTGCATGGACTC GTGATGGTATGATT ATTGCCTCAATTCA AGCCTCTGGTGA AAATCGACCTCTAA AGTTATACGTTTCA ATATCGACCTCGAA AACCTTTGATAGTT ATGTCACGTTAGCG		cd01 650	550	804
							MIISKPITRLNLRLGLVVPQEWKQARVTPIH KGGDKDNVNNYRPIISILTIMKVLEKAVHSQL HDHILLEFDLLNAQQSGFRGYSTNTLLTYVSD NLHRNIDSGKLTGIVFDIRKAFDSVDHSHKLE KLYTYGVRGNTLSWFESYLGRYQCTLVQRA RSDLVQLSGVPQGSILGPLFTLFMDLPNS VEHSKVMLYADDTAIMFSALNPKEIESTMINS DLDRVKSWLCHNKLNLNTDKNFLLVGTPQK LRHIEAPIDIKIDGLQIRQVEHVYKVLGVWLDQ HLNWNEHMKHTSSKIASRLALLRRVRKYLTD SAKLLANALVLPYFDYCCTSWNSNCNQLRDT LIKLRHQMARIVLQMNVLTPSSEMFSKLKW MNMKRWQFQKCKLMFDIVNGYSPSYLQD MFTQSAATHRYSTRIADTRGLVIPKVKTEGR RAFSFEGAMTWNNLPSELRNAQTQSFCSTY FRNM				
							MPGSTVNSIINDQSFVEFSPRNGENDFYGPS HGSYTCNEFNDLVNDSNDGLFSILHINCRLS KNYDNISTFLSSLTVQFTALTEIWIKEYSYV DIHNYTLVLKGRQNKGGGIGFYVRNDVKFN RRLLDINTPYIESLFIENVVTANISAILGVYRPP DQIQSFLESFENVITVTNTENRIFYMIGDYNI NLFKVNQCQSHVKDFIEMFMAVGMYPVLVHGP TRVNNISLIDNIFTNLENCSSTGLTDSIDH FPIFCYSQSTQVTETCTMKRQFSKINNTFI TNIGNTHWILKENDANSSYDCFLKQFQHVVEI				

				ATATTATGAACGAT CGAGCTGACATTTT AATGACTTATATTT CTCCCTGGAGTGA GAAGTTTTTTTACAT TTATCTTTTTGGAT TTTTCTAATACTCG AACTTTTAAATGTA AAATACCTTTTGTGA TTTCTAATTGTGTTT TTGTATACTGCTTG T	ATTTTATGTGTGATTATTC CTCATTTAATTGTATATGTT CTTGATTTTTTAATGATGA GTGTAAATAAACGAACTT GAACTTGAACTTG	CFPMKCSKSKKRFDPKWFNDELRLQNRKK SLYYKYLNNRTAYRLDAYKKCRNKLNRKYV KYYFKQSFVDAKGDTKQTKWIKINGLLGTSK KNEIPSTLQYNNETIDNDTDICESFNEYFINIGP KLSTLDIPTGDFKQYLKKMNLFFFYFFFTPV TPREINIVDKLKSNGCSGFSVDVDMVKRVIH LICYPLCAFNSSLLETGIVPDQLKIAQIPIFKSGT KTEMSNYRPSILLPVFSKILERAVYNNRLYSFLNL HNILNNSQFGFRTGHSTSSASVDFTCQVSQV FDKSEVMMGLFDLSKAFDLDHNIILAKLFN YGIRGVLLNWFKSYLSNRKQFVKIKGNTSSM QTVVCGVPQGSILGPLLFILSINDMCMISDKLK YILFADDTSVFMSHKNIQSLQDQDFGNEINKLL HWLSLNKLIUNVKNYMTVMFTKQKIDFNEV DLQVNNHEFKRVSVNKFLEIGIEIDSKLTWNEH TGAVCNKISKNTGILNKLRLPKCVLRMLYNTL ILPHLYYCEVWAMTSQTNLNRILKMQKKAI RIISNSHYRAHSKPIFYKLLNVDFICEYHIAVF MYLCRNESLPCALLKYFNLNSTFHDYETRNAS NFHLPQIKTTLMKKSISFDGPKIWNISPINIRN CRSLQTKFRKYKMYLIESYVL	cd01 650	424	681
Daphne- 26_SPur	GCF_000 002235.5	Strongyloc entrotus purpuratus	CCAAACCGTTTTC GTCATAGGAGTCA CAGAAACGTGGCT TAATGATACATCTC CATCTACTTTATGT AATATTGACAAATTA TTCGATTGTACGTT CTGATAGAAAAACC GGAAGGGGTGGTG GAGCTGCT	TGACAAATCTCTAAATTA ACACATGATATCATATATG TTTGTCAGCATTTCTTGA TATAGCAGGAAAAATCCCC TCCCATGCACCTCTATCC CATTTTTTTTTTTTTGGTT GCAAGTCACTGTTTATTTT TTGTTATTTTCCCCCAATTA TGTTCTGTAATTTACTTTAC TTTACAATCGTTATGATTTT AAATGTTTCTGGGCTGCAA GCGTATCAAGCTGTCAACGC TTAGGCTGAGCCCTTCTG TATCATCACTCTTCGAGC GTATGTATCTATGATTTAC TGAATGTGGAATGTAATT TAACCTATGTTTATACTCA	MFISNLLTYKIRSDLSKSDGTDSDICIEINMAQR KNILICTIYRPPHTNVELFDEFDRMLHKVNLE NKDVLIMGDFNIDLLQEDNTFHNMLRSNSF NAMIDKPTRLANNSSTLIDNIFTNVTNNYIGS GLLYSEISDHLPIFLICNHVRNYPEDEKNQAMI RKITAEKIAMFNDELKLVTVNDVVACTDTNK AYALFWNLFSLLNKHIPMTKKRSARNRIKQP WITKGIFCSIKRNRLYKYSLRHPGEEESLNNYK SYRNKLTIIIRQSRQMVFYAKQFENANGNASS TWNVINQALKKQKSKCHITKDNNTMVDSD KEIANCFNAYFVGVPNQANTIASNGVDFSQ FLQTPCNKSLFLKPTNLIELIINARSLKISYANG HDGISTYLLKQVIGNIUPLVHIFNLSTTGTTFPE AFKIAKVVPYKDDSSQSVSNYRVPVSISSLSKV LERLIYNRLYCFLTENDLLNQDQYGRKYHST DLALAQLYDKVSNALASRKHVIGVFMDSLKA FDTLDHKVLSKSLHYHYGVGRGVALDWFNDYLS				

						GRKQWVSYESVDSSLLPMKCGVPQGSILGPLL FLLYVNDISNTSLQYIIFADDTNVFCSNDF RSLQRILNTELPKLSLWFRSNKLSLNLKNTNYI HFQSRKNATDKPIPEIVLDGEYLEKKQNTKFL GTIINENLKWSDHVVHVYKRVSRVTGLLYKLLK HYVPLNLYMLYNSLILPYMSYCNILWADSKV YTDKMLLLQKKAIRVCTNAGYRDHTDPLFAR LKTLYVDINFLQKALFMFRLHNDSLPSHFPP MFCLNRNIHTYPTRQSNKYHLCNPKTTMAM KSIRHSGPDIWNLSLVDIRNLKSVTFKSKVKS ILLSKY	cd01 650	746	998
	CR1- 30_SP	GCF_000 002235.5	Strongyloc entrotus purpuratus	GTGAGGTGATGGC TACAGGGAGCAGA CATGGTCTGGTTTT GGTGATATAAAATC CACTTAATCTTGCA CTTTAACTGCAGA TCGTGATTTTATTT ACCTGATTTTGCTT CTTGAACATATCTT CAAGCTTTTGATAT GATTATTTTGTTCTT TGATACCTTCTGGA GGTAGAATATTGT GATAGTTTACAGC CATTTTCTGCCAAG GAATTTGAACACAC ACCTGTAGTCCCAT ATGGAGGATCCTG AGTTTGCTTATTTT GCCAGCCTTCTTC ATACCTACACTACA GGCAACAATAACT GTAATGTATATTTT CTGACTGAATTATG CTTATATTATCTTA TGATATAGAAGCA ATTACTCACTGGGA	TGTAAGACATTGTACAT TATTGTTATAACCTATGAT GATTAGTAACCAATAAAT ATTTCAACTATTTCAACTAT TTCAACTA	TAAGCTTACTAGCTTGGTT TTTACTTGCACTGGACTTA GCATATTTTACTGTAAAT GACAAGTGCTTTGTCCACA CACCATTCAGCAATTCAT GACACTCTGTAGTCACT TGCCTCACCTACCATCCCA GTATAGCTTCTCGAGTGA GCTGCCAAAGGGATTACC CAACCAAGTACCAAGTAA			

			<div>TTTAAATCATTTAC CAGATGATTTAGCT AACCGTCTACTACT ACTAATAGCTAACCC TACATAAAATGGGT GTAGCTGAGTGAG TGAGACTGCTGAG CTTTCTGACTTCCC TTTGTGTTGTTGGA ATACGTATTGAATA TTGAATACTGTCCT GCCAAATTTAATCA CCGGTATTATCAC CGTTGCACTTGCAA ACGCCATC</div>		<div>QQAQVVVDGSKSTWRDVTSGVPQGSVIGPTL FLLYINDIQDNISQSTMRLFADDSLTLYREIRCPE DHQILKDDLQQLSRWSSNMWIMDFNVKKA ILTTTRKTPSIREYHLSNGLIPRVKEYKLVGTV AADLRWNKHCQTIRHKASRTLGLLRRLSPCT RDVKSRAVTALVRPQLEYASEAWNPHSVTVT KMILEQVQRTAARFVYRDYRTTSPSALVAAL GWDTLHTRRLDQCTLLYKIHRLVPSIPTIVI PATYFGRHDHNLKVVPVATIDSKFESYPRAI RIWNHLPGSAVNATGITNFKEAALPSIRMMQ PPV/GSHII</div>		
L2-1_SP	GCF_000 002235.5	Strongyloc entrotus purpuratus	<div>AG</div>	<div>TACCTGTTTCATAAGAGAA AATCTATAAAGTTTCACTA TTTGTGTTATAAATCCATGT ATAACCTTATCCCATCCTAT ATTTCTAGTCTGTTTTGTCC CATCCAGCAATCAAGATAT CCATTACGTTCAAGGTCAG ACCCCTCTCTTGCAAT CCCCGCATCAATTCAAAG AAGGGTCAACAAATATA ATTTTTCTGGTCCCACTAT TTGGAATAGTCTCCCTTTA AGCATTGCTTCTTGCACT CAGTCGCATTATTTAAGAA ACATTTAAAAACCCATTG TTTTCTTTGTAATTGT</div>	cd01 650	306	575

	CR1Y_SP	GCF_000 002235.5	Strongy/loc entrotus purpuratus	<p>TGTGTGTGTGTCGC GGGAGTCGGATGT ACACTGCAATGCT TCCTGGAACTGCC TACATCGACACCT GGGATGTCCTTATC TGGAATTCCTCAT GAATTTAGGAATTT CACCTTCCTCTCA TGGCAGATTGAATT TTATGTTCTTTGG ACCAATTCGTCAA CTGAAATCTAAGCT AGATTTTGCCACTC ATTTCAATTAAGAGT GCCATCAAAATCTG TAGTCCACACGCAA GACTCTGATGTCTA CAGCTTTCACATGG GCTACAGTCATACA TGTACAGACATGTA CAGGGAACCACT AGCTTCTAGTTCTA CACTGTACAACTGG CAACCATGGTCAAT ACTGTAATGTATTA TTATTATTACATTCC TTTCTCCACTATGA GTTGATTGAATTA CATGCCCAAGAAAC CTAAATTTCCATGT GGTTCATGTGGGA AAATGTGGCTTG GAAGTCAAGGGCA TTCAGTGTGAAGGT GAATGTGAACAGT GGTTCATGCAGA GTGTATTAACTTTA</p>	<p>CAAGACCCTCAGGACCT GCCCATGCACATAGCTACC AAGTCTCCGTTCAAGAAC TGAGCAGCATAAAGCAGTC CTTCTTTCCAGGACTGTG AGAGTGAACAACCTCCA TCAGAAACAGTATGGCCTC ATCGTCAATGCCTTCCGT CAATGCCTGACGGGAGGC AGCAACTAAGTCCGGGAA CTTTTACTTGACACCATGT AAATAGCATTGACCATGC ACCTTGCTCTCGTAAGCAC ACACACACTCATCAGTCTG CAAGAATCTTCAACTGTGA AGGAGGTAGACTTAACCC AAAGAAGAAGAAGAA</p>	<p>MAPAQLCGSATSVEYTSAGQIYFPVPVSAT SAAKICHLNHTDIEPCTTNPDSKDEGESDDSL CLGSPKFSSTPRTLKPKINPTKPKAKVLTLNV NCQSIKAKKQSFALIHQHPNDVWVETESWLT EDIRSSEVFPTDHYNIARRDRPNDLHGGVFLA IRKDLLASREIETDCELIWCKISMPGAKTFH VAMKRAHLVQDQHARSQDFPCCEDEMSTLE LESLSLSPTHSVVIGGFNLPGWDRNQ QVGNCRYPNLHRRLSLLDDHGLTQVVTEPT RYQNTLDLLVSNPTVNVNEVTVPVGSVDHC PLVLNVRPTRPDEEESFTLQSSVGEVWDQ PGQMRRKVLVYSKAQWEKFGESIEAVRDEIE AKVTYASVNEIWLIFKTGIEKGINDFIPQKLTP HKANLPWVTNGIRRLIKKRDRLSNECHLRKR EGLIPVYVEERLRLKRTLQFALRKAYWNVVE SIFTVEDDDEGPGKYNCKMRFWQFIKNRN DSAGIAELKSDGVLFSEPAEKADALNRQFKKV FTQEIPHPHLLPDTSEHESMPDIVITLKGEM MLKLLKPHKAPGPDGITPRILKQLSRVIAPILRI YRLSYDTGEIPDDWHKANVVPVYKKGDRTD PSNYRPSILTCICCKMMEHIFASSIMRHGDEK DILCPLOHGFQAKKSCQLTRLCHLGFVTDLL NNMELSQQTDVLITDFSKAFDKVGHQRLKK MEYGVGRKSIDWISSFLPNRKQVVDGHT SPTVEVQSGVPQGSVLGPCLFLFYINDLPNQL STKVRLFADDTVLYHTIQETHEDAEALLQQDLD QLCGWSSKWMMLNPKKQVITVSRKKSQ VCPSVFPQWRNHWSLCQLPNTLALLSTST GTQHVSACTCKKANNTLAFLRNLRVNSTQLK TTAYQALVRPILLEYGSVWDPHTIRCYNQLE MVQRRAARYVNLNRYHHTSSVSDMLSQLGW TSIQKRREHQRLLTYKMHYSLVDFDIHPPTS NRQQS</p>	cd01 650	621	854
--	---------	---------------------	---------------------------------------	--	---	--	-------------	-----	-----

[illegible]

					CTTTTCTCCTTTTCTCCTTTT CACATCCTTTTCCCTCTCTC TTTATCCCGGTCTCCTTG TTATCACATATATTAT	LHEILSDAKCKKRSLHTTFFDLADAFGSVSHQL ITSSMERYAIPTPVARYVATLYSRLNGRVGP GWITKPFKRGVFGQDPLSPTIFLMCFNPVL EYLKEQQDQFGYKWGDRKIITTPYADDFNLIT GNKLQHQLMNGINDRLKSMGLTLKPRKCS SLSTSGKPNIVDFTIGSDVSSVHGHVPVKFLG SLITFSMTSEIFEYLGKIESTLENINKNIRNE YKLCILTRYALPSRLYMLTVHDVHASHLKSJDII LDRIKMWLGIPIRGANLAVVHDPDGLNIPSI SDLYRTCHCLAYSRSRVKGGDDVWNHALDRRL ERERGWSRQTSFSLEAHEIHSKIPDSVSWNQ TKTVIKKAVQDVCRENWKRVIPELLVQGEFM KVHMENTDVTWKSIIYNLPQSVMKFSVNA AIDSLPTNKNLYRWGKRISDKCNLCNCTGTLH HVLNNCERMLDRYTWRRHNNILRIMLHKIKNT SNNSTKVYADIEGESINGGTIPPNIIRTQHKPD ICHIDNKNKDIIIIELTVPFETNIDKAHQKQDK YASLTHDITSTGWKSKLCIEIGSRGLITPQNKY RLKEILKITNSTRYPESRDEILKAVLFSSVVFQGG RREPAAWQFDQYL			
	Crack- 2_SP	GCF_000 002235.5	Strongyloc entrotus purpuratus	GGTCTGTTGTATTA CGATATTTCTGACC ATTACCGATATTT ATTATTTCTAAAGT TCACCTTGTTAAAA AAATTACTAAAAAT AATTGTGCA	TTTCAACATGCAGTAGGCA ATTGTCAAAATTCGGTCAC ATTGCAACTCAGTTTTGCA AATTTTGTTTTCTATTTCGT ATTTCTTAATTTCTTTTATT GTAAAGGCTGTATTTGTTT CTACTCGACTATACCACAC AGAGCTACGTACACTTCCA CTATTTGTTACAAACACAT ACAAACACGCACGCTCGC GTGCACGTGCACACACAC ACACACACACCCACACAG CACACACGCACACACACA ATATATATATATATATA CACTTGCACAAACACTCA CCAACTCATACACAAAGAT CGCTTGAACACTTTTTCAT TTTAAGTTTTACCTTAAGA	cd01 650	263	519	

	RTE-11_SP	GCF_000 002235.5	Strongyloc entrotus purpuratus		<p>ACAAATGGATACA AGATGCAAAAAAG TGAAACCTCACGT CAGGTACCAAATC CAAGCAGTCTACA CAACAAACAAACA ATCACTAAGAAATC TAGCATCAGAAATGC AGGCAGTGAAAAAG CAAAGATGGAGAA ATCCTTACCGAGAT GAAAGATGTTAAA GAGAGATGGAAAG AAAGCTTTGAGGA ACTGTACAATCACC AGAACCAACAAA TGGAGAAAGCACTA CTCAGATGCCAG CCTCGAACCGGTTTC CAGACATCTGCGA GATGAAGTAGAAT</p>	<p>TTTTTTTTGTTTGTCTT GTTTTTTTTAGGCGCCA TCCTTGCAAGCTTTGCTT ATGATTGCGGTCTCTGCA CTTTCCACCAGAAAGTT AGGATAATACTTTAAATTT ATGATAATATGTTATGGAA ATAATAAGTTTCGTTATTA TAATTATGTACATATATTA TGTAACAAAAACATTTTGAC ATGAAGAAATTAATGTTTIG TTTATGTTGTTATATACCTT ACATGTATATTTATGTTATT TGTTATTCAGTTATATTTTA TTTGAAAAAAATGCAGAA ATAAAATCAAAATCAAAATC AAA</p> <p>GTCAAGTAAAGTCAATTAA</p>	<p>MPAKPCRESIRIILQRRIMKKTTEVLSEAQAGF RPGRSTMDQLFTLRQIMKRLERNQSLCCFI DYEKAFDSVWQEGWLWKAMNFFGYSSKYIHLL QALYQSSSAVRVNGELTDWFKTTVGVRQG CALSPQLNILLEAVMLYAIHDTKIGVRVQGG LNNLRFADDIVLIADNDRDLQTIVDLVHQSSS NFGKINITKTEVOAISKQPQLLNINVGGEQLK QVDKFTYLGGTISQGTGSCSEEVKHRIGKAIGA FOALKKIWASQDIRCSTKVELYKVLVLSILYLG AETWTLKKVDENRLHTEFEMACLRIMGVTRF DRLRNTHIRTQLNMEETIIDRVATKRLRYFGHI NRMNSKRYPHILLNGNIHGRPRGRPAKRW TDCIKADCKNRQVDSLTKATRLTEDRKVWQA IVKQKPSYIPVMVWTA</p>	cd01 650	12	235
					<p>YISHTEPLFQRLKTLKIQDHIHQTCIFMVKYSH NQLPSLFHNIYSLNSNIHTYPTRCSTNYHLENP KILLAQKSVRRHHGPDVWNSLPNSIRHCSSLYS FKASLKKTLSKYLHN</p>					

	REX1- 2_XT	GCF_000 004195.4	Xenopus tropicalis	<p>TTGCTGTACGAAAA CTGCAGAAGAGA AAGGATTGACAG CATCACAGGAGAG GAACTCAAAGCTAC TGGCCAACACGGA ATCACCATCTTCA CAAAATATGCGG GAAGAAAGTTTTC GGCTGACTGGGA AAAGCCATAATAAC ACCAATCTACAAGA AAAAACAATCAAC TGTGAAACTACA GGGGTATTAG</p>	<p>TGACATTCTGCCCAACCA ACACACTCCCCATAACT ACTGACTCTTCGAGTGT CACTTAAGCAAGCCACT TAAATCCTCACTGCACAA TTTCAAAATATGCATTGCA TTTTATTATTGAAATACTT GTACCTTTATGCACACTT TTATTATATTAATATTGT ATTTTTTTTTTTTGT TGTTTGTGTCTATGTA AAGTTGGGAGGAACACGG GTCAAAAAATGTCTATTAC AGTAATGATGCTTCATTGT TATTGTATATGACAATAA ACTTGAACTTGAAACT</p>	<p>MAPVRMAAVAQPPQLCFFLMFVVLFFSFF ATCNALISYDRQTLTLDIRQNSQSGPSQTEFFM NRGDLPLWPQKHKRRKRGKRAIGILVRLRN RQHRAPLSILLANVQSLLENKLDLRLARITFHR DIKNCNIFVTETWLVPSPDIAIVPEGLSIHRR DRTIDSGKCRGGGVCVLTNRKRWCTDVQTVS SGCSPDLEYLMIRCRPFYLPREFTAIFCTAVYIP PHADTTRALDELDAIDRQETLHPEAFIVAG DFNKANMKKVLPKFYQHVNFPTRGENILDH VYTPYAHAYKARPRAFGKSDHSSVLLPAYR QKLKREQPMMRVQVRWTDQSDSTLRHCFNI TDWEVFRTAADNININNYAEYVMDFIHKCIN DVVPRVNVRTYPNQKPWVNGEVRAALRART AAFKSGDPSEYKSSRYALRKAITAKGQYREK VESCYSNSNIRNMWSGLKTTIDYKGKSCRVG EVTASLPDELNTFYARFDKHLSQLLEEDGRP YQISQADVCKHFRKVNTRKAPGPDNIPGRVF KACAHADVFTDIFNLSQLQSAVPTCFKRTTII PVPKMNVSCLNDYRPPVALTSIAMKCFERIVK AHITSLPASLDPLQFAYRSNRSTDDAIAIALH TALSHLDQRNTYVRMLFIDYSSAFNTIVPSRLV MKLRDLNIGSSLCSWILDFTLRNPQVVRIGNI TSSTLTSTGAPQGCVLSPLLYTLFTHDCTATH</p>	cd01 650	564	818
--	---------------	---------------------	-----------------------	--	--	---	-------------	-----	-----

[illegible]

				<p>TACAAAGCGCGG AGTCCAGGCAGG CTCCCACTCTACA CCGGCTTTCGCC AGTCCAAGTTTG ATGATCAAGGATG GACAACGGTGGA TCCAACAAGGTA GAAGACGCCGTT GAGGCTTACGA CGCCGAGTCGAG TCCCTGCTGAGCTT GTTGGAAGGTGCT TCAACTGCTTGGC TGCATCACAGG CTCGGGGTGTA CAGTGAGCCTCGG TGCTTCTCTGCTC GGAGCTTGCCAC AAGTCTTCTCCTC CTGCCCAAGACCC CGGTGTGGAAGCG GCTCGGCGCCCC AGGCCTGTGAAGT CCGGTCTGGGAG CGGCTTTCACGGC AGTTGAGGGGAGA GTCCAACGAGCGG CAGGCCACCGTTTCG GTTTGGCAGCGTAT CGCTACGCCCGAA GCTGTTCCGATCGC CCCTTCTGTCTGG TGGGAAGC</p>		<p>ILNVLTWLFGRLLKCSLSIASLIRSMARTSRRI AWLKDG DANTRLFHSWSRYRKRKNFIKIQD GEQLVTSHSDKAEVLSFYSDLIGTGQQRST VNLDTLEMQHTDLQGLEQPISEEEVWNTIKQ LPPDKAPGPDGFTGRFYKVCWSIIKKDIMAII MAIWNKDFRNFYLNATFITLLPKKEDAILAK DFRPSLIHSFAKLTKIMASRLSGYMDKLVS NQSFIKGRFIQDNFMLVQQQTARLLHAQKQ PRILLKDISKAFDSVSWSFLEVLKVGFGPI WRDVISGLMTSSTQILLNGVPGFQHLRGL RQGDLSPLLIIVMDVLNGLIVKAADEGLLQ PLSTRSIQHSIYADDVIFLRPAASDLHLVTR ILELFGDASGLKTNVQKSSVAPIHCSPVDME MVTQIFPCKVETLPLKYLGLPLSLRKLSEQLQ PLIDRLADLLPNWKADLMTRAGRAIQVQHVL TATVIYHAMALDLPKWAIIKAIKIRRGFLWK GRKDLVGGHCLVAWGKVTRPKELGGLGIDSL QKLNWALVRWLWLEKTDPSKSWADLPQS SSILKSILSTALITEIGDGRNTLFWDRDWLMGL RIEDLTPHIFAIVPTRITNRRTVAEGIVEMAWI RDLRGTTWDLVSLDFLTSEVIADYSINVGVPD KHIWKFSPTGQYSSKAYDMLFLGSTSLGAYE RIWESWAPSKCSFFLWLMHNRCWTADRL ARRGLPHPSVCPLCDQEEETIQHLLISCVFAR QFWYSLLSRSGLPNLSPDNASCSLDDWWSR AVEKIPRDLQSGFNSLVILGAWSLWRHRNDC VFNGEIAKPGECSHPRWR</p>		
LINE1- 24_ZM	GCA_902 373975.1	Zea mays	<p>TAATTC CAGGGGACTC TTCCAAGCTTTATCTCTTA ACTCGGCGTATGTGACTCT CTTACCAAAAAAGCTGAG</p>	<p>CCTGCATCTTGAG TTCCTCACTGCTAT GGCTCGGAGGAG GCGGCTCTTCGCCG</p>				

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

				TTACTGCTGTGATG GTTCCGACGAATTT GATCCAGCTCAT CCCTGGCATGGAC GGAGGCATGAGG ATGGCGCGGGAG GCATTGGAATCCA CGTCGGGCATCGC GCCCAATTTTCAG CGTCTGGGCCCAAC GTTGCTGGTGCCCA GCGCTCGGCAGC GGCCCGGCTGCA GGACGAGGCGCGC GGTCTTTGGACGA GGAGGTGGCTGCA GCCCCCTGTGCT GCGGGCTGTGCG GCGGCTGGCGGTT ATGGCGGAGTCG AGGAGGTGGCCGC ACAGGAACCGAGG GCCTGCGCTCCAG CGCACTGCTGGGG TGCTGGCGGAGGA GGTTGGGCCACTG GAATGGGCCAGGG AGGTGTGGCCCA GATGTCACGGCAG CAGATGGGTCCCG GGACAGTTTCGCC CGTCACTCTGTG CAACAAGATGATG GGCAGGGCTGTGT TGGTGGGTGCT GAAATCGCTTGGG TCGGGCTCTCCCCG ACAGAGGTTTGTC	TGCAATGTTGTCTCTTTCC ATGTAATATCTAGGTCTT CTCTCTCTGTTAAAAAGC TCCCTAGAAGATCCTTTCT GGAGCTTATTGATAAGGT GGCAGCAAAGCTTCTGG CTGGAAAGCTGCTAATCTT AACCTGGCCGGAAGGGCA ACTTTGGTTAAAGCTGTGC TCACTGCTATTCCATTAT CATCTAATTGCTTCAATGT CCAAGTGGGTAATCAAG GCAATAGACAAGATTCGC AGGGTTTTCTTTGGAAA GGAAGAACTGACATTAGA GGAGGACATTTGTCTGTT GGGTGGTCCAAGGTCTGT CGTCCCTCTCTCTTGGTG GGTTGGGAATTCACAATTT GGAAGTTATGGGATGGTC TTTAAATTTGAGATGGCTT TGCTCCAGAAAACCTGAA CCAGTAGGCCTTGGGCA GATTTAAAAATAAGGTGC ACAGATGCTTCAGCTCTCT TCGCAGCCTCTGTGTGTT GGTGGTTGGTAATGGGGC AAACACTCTCTCTGGACA GATAAATGGCTTCAGGGC CAATCTTTCTCAGATTGCC CCCGCTCTAATAAATCGGG TGCCCGCGAATATAAGAG GCGCAGATCTGTCCAAGC CGCCTTAACAGCCAATGGT TGGGTGCATGATATCTCTG GAAGCCTTACGGCGCAGG TGCTGTTTGAATTTTAT GGTTGGGATTACTCCAA				
--	--	--	--	---	--	--	--	--	--

				GGGTTGCAGAGGC TGACCAAAITTCGC CCCCTTCAATGTCG GCTTGGTCCCCGG CTGACCTCAACTCA AAATCTTTGGTTGA CTGGCGACAGAA CCTCCCTGAATGTT TATGCTAGGAGGC CAGCTGATTACAGA TTGAAAAAGCAGA CTTCAGATACCGCA CCGGTTCAATCCCT TCGGCCAGATGAG GTGCGGCACCTCT GGTACTGCTGGTA TTGGAGAGGCGAC CCCCGTTATTGGTG GCTAATATCCAAGA ACATTGACTGCTTG CTTCGCAACCTCC AATTCAAAAAGA AGAACTAAACAGCT TCCGCCAGATTTTG TCCACGTCGCAGT AGTCGCTTGTCAAA GAAAAGGAAGGC TTAAATCAACAGT CAGACAAGTCCAG GCGGAGCTCATTTG AAGTTGAACGTAC AAACCAAGGTGGC TGTAAGTATGAA ATGCTGGAGGAAT TTGGCAGCTGTTTG ACAAACCCCTTCC AACTCTACATCAA GGCCCTGGCAGCTT	GAGGTGCAACTGTACCCT AATGTGGAGGATCAACAT CGGTGGTTGCCATCCTCT CGGGGCTGTATCCGCCA GCTCTGCTTATTTGCGCTT CCATGGTGGTGCAACAAC CTTTGAGCCTGCTACAAGA ATCTGGAAGACTTGGGCC CCCCCGCGCTGCAAGTATT TCATTTGGCTTGCACTCT CAACCGATGTTGGACGGC TGATCGACTGGCACGTCA GGAATGGATCATCCAGAT AGTTGTCCCTTGTGTGATC AGCAGGATGAGTCAGTGC ATCACTTGCTGGTGGATTG TGTTTTACGAGGGAAGTT TGGTTACAACTAAGCA GGCTAGGTTTGCAACATCT TGCTCCACGCTCTGATGCT ACAGGGTTTCAAGATTGG TGAGTGAGGCAGAAAAGA CAAGTGTCTAAGCATAAG AAAAAAGGTTTCAACTCTA TTGTAATTTTAGTAGCTTG GCAGATTTGGAAGCATAG AAATGCGTGTGTTTGAC AATGCTTCCCAACATTC ACGCAATTTTGCTGTCCAT CCATGATGAGGCTTTATTA TGGGGTCTTGGAGGGCT GCTGCTCTTAGGCAGGTTT GGTTGTAGTTTTTTTATTT TGGTTTTTTTAAATTTTTGG GGTGCTTCTGGGCACCTT TGTAAGGCTCTTTTCTAG ACCTTTTCTTATCTTCT AATATTTGATGCGTGGT
--	--	--	--	---	---

				TGTTGGATGGTCG GTCCAGATAATGT GCAAGATGCTTG GTGTCCTTTGCTA GGGAGAGCATTG AGTGTGCTTGCTGC TGAGCCTCACCTTA GTGGTTGATTTCT GTTATCA	TCTCCACGTTGTTCCGAGAA AAAAAA				
LINE1- 41_ZM	GCA_902 373975.1	Zea mays	G	TAATTACGTAACCTCTATG TACCCTTCTCGCGTGTCT CTTGGGCACCTGCTTCACTC TCTTCTTTATTAAATAGAAA CGTGACAAATCTCGTGCCG TTCGTTCAAAAAA		cd01 650	23	286	
						MADVMAVMNSLHDQRFLNFDLLNRANIVLL PKHDGAENLSSFRPISLIHSAKLFMKLLSMRL APTMRDIISKQCTAFIKGRSHDNFLFVRNMA RRYHRNGTPMLLVKLDISKAFDSVRWDYLLA LLAHLGFPKWRDWVASLLATCSSQVLLNGI PGQSITHGRGLRQGDPLSPFLVLAIDPLQRLL HLATEAGILSRVAKARTRLRIAMYADDAIIFVK PEKRELDSFATLLHQFGEATGLRTNIQSSV PIKCAGLNDEILAGFPASRSQFIKYLGIPLSV ERLKKVDFQYLLDKVSSKLTWCWQGRNFTFAG RLVLKSVLSAQPVHTLTINVPKVWLEKIDKF RKRLWAGNENLTGGCKKNWPTVARPLDL GGLGVDIHRFARALRLRWLWREWEVPEAP WAGLQPCNDTDKLLFAAATRIAIGDGAKTS FWHCAWVNGRRPKDIAPDIFAASSDCRLSVR EALTDSTWVNRIDLHAIHTADHLKQFVELWS FAHSVSLVQNTEDRIDWIYNASGEYSAAAYR IQFMGHTCNLKALIWKPWGPSCKLFAWLI IRNRVWTSRDLAARGWPHSPSPCLCKRHPET AHHIIAGCRYRRIWDALASWLSIGPLAPAT WPDSTSJKQWWFMVGSVAVPKKGRSLLLL VWWHIWLERNNRTFQRQERSVADLLAAIKAE ARLWGVAGAKHLLALLPML			
LINE1- 35_ZM	GCA_902 373975.1	Zea mays	ACAAACATGGTAC TGCGCTACGACGG AGCCGACGGGTG CAGGGGCAGGGGT GGAATTTAACATGC AGGATCTGGAGAA	TAGCTCTGGTTTGTGTTTTT AGGATCTGTTTGCTCTTGT TTTTGTTTTTCTCTCTT TTTTCTTTGGTTTTTTTATC CCCTCGGGGAGGTCTTTT GTACTATGGTCTATTTTGG		cd01 650	498	760	
						MDPSKIIICWNVNRLNSADRQNSVRNLVSSK ADVCLQETKMTVISTRTILAAALGSDFCQFSFL PSCGASGGVLVAVKRWHLGVTGLSRIDRYISI QFCNAHNRCAWWLTCVYGPQSNDSKMLFL QELRDVRAACTGPWLGIAGDFNLIIYKADKNN SNINRAMMGRFRKRLVEDLSLNEVPLFRKRYT			

				<p>WSNQDQSPVLVCLDRVLCSEVEDNFPNVL LQSAASEGSDHCPDLLGIRDNKLGRRCFEA FWPKEGFDVIEVVDVSAVAVNCPFMTLN LKIKATARKLQSWSGKHVGHVRTQLAKEL HRLEIAQDCRALSAEIWLSNLLKKHSLFLSSL QRTIARSRSRWLKEGDANTKLFHMHSHR KRRNFAQLVNGDQILTNHDDKAVLVDNFI ARLGECGDREQTNLEALHPTFNLAHLEP VTELELWNTIKGLPNDKAPGPDGFTGRFYKS CWNIIKMDLMVVAAVWNRKEDNFCLLNS AFITLIPKHDGAVHVKDFRPSLVHSAKLITKL LANRISQKLNDMVSPISQAFIKGRFIQDNFM LVQQTARLFHQQRPRLLKLDITKAFDSVSW PFLIEVMKYLFGQIWRDIICGLSSSTQVLL NGIPGGHIEYRGLRQGDPLSPMLFILVMDV LGLLFSKAEHAGLLDLSPPRPLHRVSMYADD VALFLHPSAADIAITLDILHLFGEASGLRNEQ KSNVYIQCGGVELEVMQDLLPCGISTFPCKY LGLPLSLKNTKVQVQPIVDRIADQLPGWKA DLMTRAGRKVQVQHVLTMILYLAMAVNIP SWAIKAIDKIRRSFLWRGRKEARGGHCLVAV GRVCRPLQLGGLGISSKELIWAIRMRWLWL QKTEPNRPWCSLPIKVPDKARAFSMVLTSK VNGANTLFWTDKWIHGQKIADLVPRLFDI PKKITNRRTVQEALDRRWIADITGSLSVGL LDYHLWNLSIDFELLDPDIEDKHTFSIAANGTY SAKAAYDGLFAGVQFGHYERVWKTWAPSK CHFFLWLSVLNKCWTADRLAKRGLDHPERCP LCDQSESLDHILVSCVFAREFWFRLMCIERL QNLAPTPSTVSTFGWWEIMIDNGSGDMIMIK GVNSLIAGAWIWNHRNIRIVFDGISPSVAA LCQAREEQQLWEMAGAKGLSFLAATIRGT</p>	cd01 650	496	759
<p>ACCCTCTTTTCTCTACTT AATATAATGCGCAGTT CTCCTGCGCTTCGAGAAA AAAAA</p>	<p>CAGAGCCACAAA CAAGTAATGCGGA CTCTGAGAGTCATC ACAGAGAAATGCAA GCATCAACCAGCAA GCTGTTAAGGATTA TGAGAACTCTTTTC AGCATCCTTTAACT CAAAGTCATGTTGA AGCGTTAGCGGCT CTCTTTGGATGGGA GACTCCGGGGAAT CTCTCCCGTAGTG TGTCTTCTTCTGG GTTGTAAATTATAA TTGTTGTA</p>	<p>TAGCAGTTTTCTGCTGGT AGTCGTTTTGGTCGGGT GTTACCTTGTGTGTGTG TGGTGCCTTGTGTATTGC CGGAGGGGTCTCGGCCTA GGGGCCTTTGTAATTAGG</p>	<p>GGACCTGTTGATG AGCCAGGGACCGT GGAGCAGCTGGCG CCCAGTACTGAGG ATTCTATAGCTGTC ACCTCTATTCTTCC</p>	Zea mays	GCA_902 373975.1	LINE1- 2_ZM	

GC GAATTTACCTCT GCTTACTTGGGAA GAGGATAGACTTG GTGCTGCTGTGGTT CCAGACATTTGTGC TAACACGGGTATTC CTCACCTGTTGAG AGGCATAGGCTG TTGAGCCGACGGA TGCTGCCATTAATC GCAGAGGTGGCCT CCCTTCCAACCTTC AGGTATATTTAGG CGAAAGCGTCACTC TAAGAAATGTTCTG CGCGGATGCGCC TGCTGTTTCTTTCA ATGACCTGGAGAA AATCATCAAGCCAG TGGTACAGCTGCTG CCCATGCCGTTGC CAAGATTAGAAC AGTAAGACTGTTGT TGCTGTCCCTCGTC GGAGTAGGAGGGT GGCTGGTGTCAAG CCTTGTTCTCCAGG TCCGATCATCACTG AAGCCCAAAAGAA AGTCATCAGGAGTT TGGGTCTCGCAGCC ACGGAAGAAGCCA TTGATCAAAAAGCT CAAGACAACTATTT TAAGCTGTTTCGAG GGCGCCTGACGG ACTCTCATCTTGCT GCACTTGCTGCTAT	GGTTGACAAAGTCTTGCTCTC TCCTATGTTCTTTTTTCTT CTTAATAATAAGACAGCA GCTCTCCTGCGTTTGTTCA AAAAAAAAAA	SNERSPTLRLDRVFCNDWEDFFPNSLLQS ASSGVSDHCPILSLNVQTHKKRRHFESWVP KLPGLEAVSHSWNAPVSTACPVVEGIFKLQR LSKGLQKWSHHKVGNVKLQATAKEILHRLEI ARDNRVLSPEGEWLKRLKLIHCLGLASLERTI ARLSRILYLQEGDANTSFFHQARVKKKNF IAKFQVEERIVTSQEEKQQAALDFYDNLGTA EHRDYLTLHSLGIQHQHDLGDLVFSMEEV WSVWKDLPDKAPGPDGFTGRFYKVCWDIIK EDMMGALLAVQCCHVSKLKLNTAFITLLPK KVDALLVKDYRPSILVHSFAKLVTKVLAARLAP HLPMSVINSQSAFIKGRSIQDNFLLVQQLTRS LHHSKEPHILLKDISKAFDSISWPFLLDMLQH LGFGRSWCNLISILLCTSSTRILVNGEPGEYILH HRGLRQGDPLSPMLFILVMEALNAMVSYAF REHMLQPIANQHAHCHISFYADDVVLFMRPS HQDISAIIIDIFGHATGLKTNISKSSVTPIQCG ANELTTISALLPCATKNFPCTYLGIPLSIRLAN SDDLPLVDKVDKLPGWKANLLSRAGRLVMV KTVLSSVPYLMIALELPKWVFKAIKRRRGFL WKGQGDANGNCLVSWEAVQRPKYGGL GIINLEMFGWALRARWLWLQKTDASRPWA GLPIRVHRNAKALVDVAITSVVSGGESVKFW SDRWLLGKTVAEHCPQLIQVISRRALKRRTVA EGLTNRQWVSDIRGGLSVTVLVEYLQWNLV DGWVLPDIADQHIWRLSAHGTYSKSAYDA LFVGSIPFGPWRVWKTWAPLRCKFFVWLAI KNRVWTADRLAKRGLPHPVACPLCDQAEETI QHILVSCVFARQIWTSMLHNLGLLAIVPQHG CTFSNWWCQSIKKVEKSLRKLNSLIILVAW EIWKHRNACVFEGVVPCTQRVQSAVIEGNNV WCLAGASALQDILLRLQLPRGP				
---	---	---	--	--	--	--

[illegible]

LINE1-28_ZM	GCA_902373975.1	Zea mays	GGGGGTCAATTCG GCGGACGTTCCG CTCCCGCCCTCC GCCTCCCTGCATG CCCTCCGCTCC CGCCCATCGCTCC GTTAACGCGCCC GCCCACTCCCTCC GCACACTAGCCC CTGCGCGGCTCA CCCGACTCCGGG ACCACTTCCTCCTC CCCTCCCGTGTGG CGACTTTGCTCC GCGGCTTGCTCGTT CTCCCGCTGGGC GTTCCAAGGCGCA GCGGTGGCGTCT TCCTGCCTACCG CCCGTCTCATCGG CCTTGGCGTTGGT GCCGTTGGAAGCG GTGCTCGCAGAG GACTTCAAGGAC GTGCTGCTGCTCC TCCGCCCTGCCGT CTGCCGCGAGTCC CTCCTCGACCTCCG CACAGACCGAAGTT TGCGGTGGTGCAA CGTCCCTGGAACG GTGCGCAGGCCTG TTCTGACCGGGA GGGCTGGTTAGG GTGGAACCTCGTGG ACGTCAACTCCCC	TAATGTTTTTCTATCCGAT GTTCTGAGCGGATCTTG CTTGTTGTTACGAGCTCCT TCCATGTTCCATTTCCGAA TTCCCTGTACAGTACCTTG GCCTCCGCTTTCTCTTGG AAAGCTTAAGAGAAGTCA AATTCAAGCAATTGTCGAT AAAACTGCAGCACAGTTA CCTGGTGGAAGCTGAT CTGATGACAAAGCTGGT AGAGCCATTATGTTCAAT TTGTGATGACATCAAGAAT GATTATGCTGCTCGCG TTGGACATCCCTTCTTGGG CTTTTAAAGCAATTGATAA GCTTAGACGAGGCTTTCTG TGGAAGGGAGGAAGGA GGTGAATGTGGCCATTG TCTGCTGACTTGGCCCAA GTCACTAGACCCAAAGGAC CTGGGGGCTCGGGTATT AGTGATCTCGGATCATGA GCTGGGCACTGGGGCTA GGTGGCCTTGGCTGCAA AATCTGATAGGACTAAGC CATGTCGGACTTCCAGAT TCAAACCTGCGACATTGTG CAAAGCTTGTGGCTCGG GCTGTTATCACTCATTGG GAGATGGTCTGTACTC TGTTTTGGAAGGATCGTTG GCTGAATGGCAATGTATT AAGGAGGTGGCACCTGCT GTCTTCGGTATGGTCCCA AGCGTATTGCTTCTAGAGC	HLFLRCNFAKACWNISGIPPTQNCSSRGGGLG KPKQTTT	cd01 650	506	704
-------------	-----------------	----------	---	--	---	-------------	-----	-----

[illegible]

[illegible]

GCCATCGCTACCCG
 TGTGTGGGACTGTG
 TTGAAGCCGTCAT
 GAAGCGCTGTGGG
 GAGGGCTCAGGA
 GGTGGCCGTATTC
 CCAATCAAAATTT
 CAGAAGGAAGTGC
 AGAGGGCTGCTG
 CTTACAATCCCT
 CTCCTGGCTGCTCA
 GTCTGCGACCACTG
 CTGATTTTCATCGT
 CAGAGTGCTCAGGC
 CGGGCTCTCGGGT
 CCTGTGTGGCCGG
 CAGAAGCGGCAGC
 GTCGCCCAAGGGC
 ACTTGTGAGGGG
 GCTCTACTCTGGCG
 GAGCAGCGCGTTG
 GCTTGAGCAGATG
 TTGAAAGGTAAT
 GGCCCCAGCACA
 GATTGGAAGGAA
 GTTGTGAGGATTGC
 TAATGAGCTCGGCC
 TTCTGTGTGGGGAT
 CCATCTGAATCGCT
 GTCCTCGAGGTTT
 TCGATAACTATGCC
 AAGGTTTTTCGGTGA
 CGGCTGCACTGAT
 CTTGCAAGTTAAGC
 CTCCTGGCCTCT
 TCAACTGGACTGTC
 CCAAGAAAGACTTGCA
 GGGTGCAGTGGTG

[illegible]

[illegible]

				ATGCAAGTCAGAA ATTATTTTCTAGTA GGGTACCCCTGGG CTAGAAGGCTATT CATTATGCTCAACT AA	ATAATATAATAATAATAAT AAATACA	QRRTHRNLIPIHSTLESIKTQVAQQISQHETD SHTQIEHDVEQTPEITYDDNTQPINNTDLDLI NKLDTLQETLTFTGTDPTRPRPKLHNSQ QLNNTLHCMNTHILPKLINAETTLIEIHTLIYAA AVTVTTHLGLKTLPNKIGTNTKHLKPAWQK RLEKQINEIRTELGRLTQYTKNNTSTKLMKSIR KIFRKLTHRHENHNTHIKEYIDTLKQKLAOK SURLKRYKTSARKLDNTLFTNNEKTFYRNINT KDPKQTNREIPTQQSLTSFWSALWAVPVT HNQRASWIAEEETRTLDVPEMPYTTITTEELT NTIRYTHNWKAPGVGDIQNFYWKHFTCLHD LLAKHLTHLHEPKDTPSFLTHGITFMKPKSTN TGNPSQYRPITCLPTLYKILTSCTNKIYAHIEEN NILTEQQKGCRSHMGCKEQLIIDTTLKQVH RQHRHLHAAVIDYKKAFFDSIPHSWLIQTLQLY KINTKYITFLQINIMQHWRTTLHITINNHNHAT EPIHIRRGIFQGDLSPLWFCALNPLSNKINT SNGGYDIRQNTTHYIHSHLFYMDIKLYAST QTQLQLLQIHTHFSKDCIMEFGLDKCKTLHI EKGVLKGGFTLENGDTIESMEEGEVYKYLGF QQARLTEQKTKQTLTKLYTNRLNLIKTHLN GKNTVKAINTEFAIPILTYSGVIQWTRTELES QRKTRTTLTKYQYHHPKSAIERTTLPRKEGGR GLIDPHLHQRQYNLREFFHKKQHTSQLHKAI VLADRNITPLNLQNNTPCQIDNTAHIQEKKTS WCKKELHGRPHDLQKHFDIDAKASNSWLVR GDLFTETEGFIIAQDQVINTKNYRKYLKDSL TNDTCRKCHIAPETIQHITSGCKLLAQTDYLHR HNQVANIIHQKLATKHLINTLTPYKYIPDTIL ENSTHKMYDYDRTILTDTKTIHNRDPDITLIDKQ NKVTYIIDISPDTHNIQDAFTQKISKYTELSQEI KRLWGMQKVTVPIILSSTGVIPIHTLHKSCLKTL DLHPKTYITLQKASLNTCNVRFKSLDSTDSDS MEFLVDDLSILAPLGKSDHAVLSFSFVSKTELR YPTSNKRWNFKRLNV/LALQDYLQVQV/DWDV HPQLEVDHAWDFLLHTILCATEHSVPKMPVK SYKQPPIIKNRTRRLLSRKRHCWAEYKRTDNN GAYROYKHIRNICTKAIREDRLQFOTKILDKEV	cd01 650	288	537		
	SR1	GCF_000 237925.1	Schistosom a mansoni		TGATTACCAAAATCTTTT TTCCTCTATTATCGTTTCAT ATACCTAGGTCTTGCCTG GAGGTATTGGTGATCCAC TGCTACCAGACACGGAAG						

	CR1_AC_1	GCF_000 090745.1	Anolis carolinensis		CCCCTTAAGCGAAAGCTTC TTCTATTCCGTCTCAACC ATTTGAACCATTTG	SNPKLSFYAASLRQKGTGVSQLLGPNGPTN NDSDAANLAEQYSRTFQLTHINHTDEFTCT CTGLSEVDLSADLVLRKLQHLKDTSPGDM VHSAVIREAASILATPLSVMFAHSLSRGKLPEI WKLAIHTPIFKGGRSEPSSYRPVALLSIPSKI MESLVYDGILEYLSKFFSQHGFGRKHSC MTNLLTAVDRWTTILDRKGKVDVYLDVSKAF DRVNHACLIKKLGRGIKPLIDWLSYLENRH FKVRVNTLSQAMECPSPVQGSILGPLLIY INDLPQVSSDLLLFADDVKLWREIRNHNDIL VLQEDLTRLSWADDNGLTFNTSKCKVWHLR HVADSYNLGNSPLEVSQVEKDGLVLPYDL KSYANCDKNASQANLALVTLKRIFGQDGR T FHIIFNSFIRPHLEYGNIVFPSPQLQKDKDLERI QRRATKSVRGLKFKPYEERLKSLLNLYPLEYRRL RGDLLMTYSILNTSGHPLKHLKLSHNTNLRG NTQKLETLHSRTDCRHNFYSVRVKWNLSLP TELVAQTSQESFKRKLDFLRTKDNILL	cd01 650	1008	1259
					TTGATCCCAAGAG GGCAGTTGGTTCTC GCAGAAGCAGAGG CAGCAGGAGGACA TTTTTGCTCCCTGG CTTCAGGTAGGATT TGGCTAAGATTGG CTAAGATTTAAAC TGAGTGTGGGCTT GGCTCCTGTGGTCA AAGCTAACTGTTG TGTGACTGTTGTGT TGAAAGAGAGCCA GGTACTTAAGTTGA TTGACAGCAGGCA TTGTCCCCTGTTAA TTGAGTTTCTGGGA AAGCTTTATTGGC AGTGTGAGTTTCTG CCAGAGCCTGATCC	TGAATGCGATTCTCGTT CTTAGCAGGGGGTTGGAC TAGATGGCCCATGTGGTCT CTTCCAACTCTACTATTCTA TGATTCTATGATTCTATGA TTCTATG	MQQKNIPITCTSCGMFRFFQTKLDNVCSK CKQMTLMEQRIQLEHRIKTLKDIALEFLD TAQHAADVQOPASHQHQQDHDQEPYGEIN SKVDNPQAWKEVTHRRRRTRQPQNSSA QLHLNRFELTSLTYNQETHLVEDHSFLDIT QWTTLDQCAGDSPDGDSASPQSHLCNHAN APSPIDQEQQEHPLGETSRNRPWESNGLS DVSQWIVIDECTGDVEEDNTLHLHDSLQOE HSSGDIHTVLHKRDPVNPQRQVLVWGSLL RGTEAISRDPDGMARETCCLPGAKIHHTQRLS RLKPHHPHMLIHVGTNDTARHTFQKITN DFRALGTLKLYNVHVIFSSLLPVVGHGSTR GKIVQVNNWLKRWKCQEEHFGFLDHGLLQOE DGLLASDGVHLTQVGKHLFAHRLTNLRRLN PWSTLPRGWTTGKRWGASHTSRKTSFCTQT HKPHQAHFKLPLLEGNNLSANTILPTTTGN RRKAKRRAAQTTQGGPSTESTIIPNKQFEGRS QGLTCLYNAQSMGNKQDELQLLAQHHTYD VIGITETWDDSHHWNLTIEGYNLFHRNRT KGRGGGVALYVKNYSVAEEMQDCNPQNQL		

167

	Penelope- 2_Psi	GCF_000 230535.1	Pelodiscus sinensis		CCAGAGCCTGATCC CAGAAGGGCAGTT GGTTCGCAGAG GCAGCAGGAGGCA CCTTTACTAACTAT CCTTTGCAGTACAT ACAGGAAACAAAG CAACATAACAAAT ACACAAAGAGGTG GTTTGTGCAGTTT GCACATAAAGTGC GTGCATATTACTTA ACTGTACAAAGATC CCACTTGGCCACCA CGCTCACCAAGAG	TAAACAAGTCTTTGAACTTT CTACCCATTATCAAGACA GTTTCCCAATTATCACCT GTAATACCAATTAACTACA AACATCCCACTCTCCCTAC CTCTAATATCATCAATTCA CAGACACTTACCTCTCTC CTCCCCCTCCACCCCCC GCCTCCTCTCTGTTCTGC AATGTGATTTGCTCTTTC ATATTTGTTCATTTTTTTA ATTGTATCCTTTGGTATAT ATGGTTGTGACTACTTTCT TCCACTATTGTACTGAGG AAGTGGTCTGCCCCACG AAAGCTCATCATCTAATAA ACCATCTGTAGTCTTTA AAGTGCTACATTGCTCTGC ATTTTGTCTCAACTACCCC AGACTAACACGGGTACATT TCTATC	MDIVPKGLKVNPNLRSTYFTDYSERLCLTSLKK LRNHLISILYKQKTIKNDISNLDRLISROTSTPT DTLQDFCSTRQEIVYTHFYSLQQRKDNKF-SNL LHTTGYN-SNCLNPPDNIVNLSYKLNPAESV LSRGLSFCPTPTNWIQFCGDLEAFFRRLRE YFQTTTTEHQSDLPDPYPVQHQRKNNSIWTPP DGRNVNLFYIHSFRNNAQTDIIRKQQATH NLSRAEHHAIQSLKNNLDIVIKPADKGGAVI MNKADYNQEAATRQLSNTTFYKPLPSDPTLDY QKKLHCLLKSLPTATREQUILTESPSDPRPGLFYL LPKIHKPGHPGPIISGIGTLTTGLSSWDTLK PFVNTPSYLRDTTDFLRKLNIDNLPNNTILA TMDVEALYTNIPHEDGLQAIRNTIPEDTTANL IADLCNFVLTHNYFRFENNLYLQISGTAMGTR MAPQYANIFMADLEQRFLNSRPLSPLLYLFYI DDIFMIWTHGQETLEIFHRDFNNLHPTINLSL DHSREIHFLDTTVQINNKGKLDTTLYRKPTDSY SYLHASSSHPEHTTRSIVYSQALRYNRCNSPT DRDQKLQDLQYAFINLNVPPRETCKQIERARR IPRNHLQDRPKKTNNRTPLVTYNPQLKPVQ HIINKLPLEQDTILKEALGDRPIVSYRQPNNL KMILTNNHRTYHTNTNPGTFPCNKPRCQLCP HIHSADTIIGBNQVSVYKIKNTYSCASRNIIIVAIM	IPRO0 0477	305	558
--	--------------------	---------------------	------------------------	--	---	---	--	---------------	-----	-----

[illegible]

[illegible]

	RTE-1_PM	GCF_010 993605.1	Petromyzo n marinus	GCCTGTGCGTGTGT GTGTGTGTGTGGC AAAAACCCCTA CCTCTCGCGTGC CGGACGGGCACT CTGCTGTGGACAGT GCTGTGGGGGAG GTCCAATTTGATT GGCTCTCCCTCA TGGCATATGTGCC ACCACAGGGGCTA AACGGGGGCTCTG GACGGGTGCGAGG GGCTGGTAACCCCT CCACCTTCAATTAA TTACCACGAAACCA CGAATTAACTGAA ATATGTAACATA AGCTTGTGCCGA GGCCCCGGGGGC CAGGCGTCT	CG	LSLLELNIIHQFSDKYTKYPIRNLRPKNLF QEPKVTKYKAVRASVYPRVFNKLPRIEIT KISELKRELKQWFNRKELQLAEKSTRGWNL NCWLCVLCRELWSKGVVMSCFHFFELFAR N MTSPGQTPIGGARRNLNYPGSSRNNEKPLR VATWNVRTLAEERLPLLCRELDYRLDLCGI QEVWRTGSGSCHHEGWTVLYSGQQTAKQQ GVALLNKRMGAWERGGEVWEAVSPRL WARLPINRRRGMVVVVTYAPTEDEHPWRD AGKADESDFYDLSQLAKVAPDRVIMLG DFNARYGHDAGTWSGVIGGHPDQLNN GRRLLDFCAAHGLVTNTLFRHRIHTTWSM HAGSKQEHLLDYVTDQRMRAHVLDTRTYRG ADLPDHRVLVICKMRLPFFHSGGGCTPRTPF QPKVAWSQLSDSCCKREFHRLRLGLASSPT PPDVEGKWARFRTVTHAAAMAALGTRSRPP QRPWLTEEVFRMVVEKKRQAHCHRLQNPTSE NEEAVRSRSEVRRAVRCKDGVWSRVAEE MESASVARDHKSIFNSIKKVTCSATPITIIRGK NGDPISDPQEQVCRFAEHFCEVLGEGKLSQ ENIVAQSEDDPAEAAVPQILEEVTPCEWLAE VPSVEEVLKAIRSLRPGKAAGRDDLPSEMLRE GGVCAQTALYDIITTVTTTGRVPQDWKDAL VWPLFKGDRDTCNNYRGISLLSVPGKVLARII QHLRARHAEERLAEPCQCGFRKGRSCTDAIFSV RLQQRCREFFQDHLCLCFIDLKAYDTISRPG LWVVLRAFGVPDPPLTIKDLHDGERARVKLR GRESEFPVHLGVRRQGCPLATLFIYFDHV REAFNSCTGGISWYRYNGRLIDARVSRDGS LLLNHVMYADDLVIAAHSEEELETLRLERAT KRWGLTISPTKHLPGYFVPSDTPPPQLPIG DGAVVERVESFPYLGSLMTGSSSLTAEVSLRIS RAAISFHRLRNAVWKLPGLSLRTKLQVFSATV IPSLLYACETWTPLEHLLRLETFRLACLRSIL GLTRIDCVRSQILERSGGQPIGELLRQARLR WLGHVARMPSHRMPKQLLFGRIEGAKRARH GLEKRWSDVQEDVELSGLADDWYQRCQD	cd01 650	578	856
--	----------	---------------------	------------------------	---	----	--	-------------	-----	-----

						RPLWRLVKDATGQLEAVALQQRRAEERR SQQRAERRVAKAQVQVTVGGTGGASASHLS QSTRGTWVCPVTSQORAFDTS			
Perere_S med	GCA_002 600895.1	Schmidtea mediterranea		TTTAGCTATGTTAAATCAT GTTTGTTTTTTTAAATATTAT TAAATAAATTGAAAAATTTT AAATTGCATTAAATTTTAA AAAAATTAGAAATAAGTA AAAAATTTAAAAATTTAAAA TTAATTTAAAGTGAAATTGA ATATATTTTAAAAATATTT ATTTATAAAAAATTTAAAT TAATAAAAAATAATTTGTT TAAATTAATGTTAATTCA TAATATTTTGTGTTGTTT GAGTACCTGAAG		MIAVLSKYIVRVFYWRYRLIKSLPNLFNYLFIN HAAAAPIFKRFMKVYNSLKRKELDIFSLETCL EKMIPSFYKMKLSDCLSSKDQVJHRRTLRR MLESEKIRTSIDKLTNIGLMLNNATNEFCL QVCVEWCIKNSIVQIKSKENAHIIKLINLRKIF GISKFSVENTVINLSKIVLNDDEERVLFKGPNH CIKNRDETCLCAKIESLYFHIAIRSMKMGINGNY VKSSLSASTMEYIDKNKKQNIRDFTMLKNLAK KNINICKFDKNGGIVILDDHDVYVYNKLNLDLNS PQFVKLPKGRKNGKLPPLKDEEFVQKMLNGF IEKKLIDEEVGKSIRPIGSPAKLYGLPKIHKDG VPMRPVLSMVGTAQYQYIAKFLDGLLKLPMRS EFECKDSFEFVSCIQLQKRSMNDVMVMSFDV CSLFTNVPLVETIDLCALWNEHSEHHILDR RAFRKLLEFATSNVNFLENDEWYQQIDGVA MGSPAPTMAISIFLASLEKKIASFQFEKPFVYK RYVDDIFLIFENQKHVEPFLQFMNSLHKNIIV TCETERKSSIAFLDLIIQRKDKQYETEIRKPTD TGLYTSPESECFEYKRNMMVKGLIYRSWALSS TFANSVKSVDKLVLLIKNGYSKSFLLMIKETV DKLIKNPKDCCGKSECNLSDGGCELSVDLGDG KNGYKIEPKYVVLVPYSEGFNTYKRKLGKLI NLEYKIVSNSCKVRNMMFINKSKTPVGLCSDLV YQFTCNGCNATYIGETSRHLCTRVLEHCRKLG LTNISEHNRGCKSDIGMSDFRILLRSFNYSWE RVICEALLIRSLDPKINVQSAVTTNVLVNFK	cd00 304	474	565
L1-3_MD	GCA_000 002295.1	Monodelphis domestica		TAAATTATTAAAGAACTATTT TGCCCAATTATATGCGCAAT AAATATGACAACTAGGT GAAATGGGTGAATATTTA CAAAAAATATAAATTGCCTA GATTAAACAAAAGAGGAAA TAGAATACCTTAAATAATCC CATCTCAGAAAAAAGAAA AGAAAGTGAAGGAGC		MRKKLKEIKEIDSKTILVGDNLNLPSELDSNQKINKKEVREVNEMLEKLE LIDIWRKLNDRDKKEYTFFSAVHGTYTKIDHVLGHRNIANKCRKAEINIATFS DHNVIKIINGPWKGFKNWKNLNLQNWWVYKKEIETIMDFIKEND NEETTYQNLWDATAVLRGKFISLSAYINKIEREENELGLQKKLEKEQIK NPQJTKLEILKIGEINKIESKRTIELINKTRSWYFEKTNKIDKVLVNLIKRKR KRKILTVSKMKRVISPLMKRKL			

				AAGCCACATACCCC ACCAACCCACAT CTGCTGCCAGGT TCATGAACCTAAGC CCAAGCCAAATCAG ACCCTGAGATCCTG CTGGCAATAGT GGTACAAGCCAAC CCATACCCCTTGAA ATTTCAAACTGTG GAGAAGTCCAAGG CAATCTGGGCTGA AGTGGGCTGATCC ATGTGGGCCCAGA GCAAAGCCAGGAA TCCAGTCTGGGCT TGGATAAGGACA TAATTCAGTTTG GGTGAGGTGATA GTACTCAAGGGA GCCTCACTGGGATC TTTTGCCTAAGC TCAGGGCAGAGCT CCAAGACAGGGCA ATCAAGGGGTGTC CTGATTGGATCAAG TACACAGTGAGACC AAAAAATTGAGCC TAAGCCTGGGCT AGAATTTGATCAGC CCCTGACCTGATCA AGATCAGAGTGAG ATCAAAAGCTTACA CCCAAGCCTGAGG CAAGTCTTTAAGCT ATCAGCATCTCAAG GGTCAATTGAATCA GCAGGGGAGGG	AAGCTCTAAGAAAAATC CCAGGGCCAGATGGATT CACAAGTGAATTTCTATCAA ACATTTAAAGAACTAACTAA TCCCAATACTATACAAAT ATTTGACAAAATAAGCAA AGAAGGAGTCTTACCAA TTCTTTTATGACACAAAT ATGGTACTGATTCAAAAGC CAGGCAGACCAAAAAACAG AGAAAGAAAACCTACAGAC CAATCTCTTAATGAACAT AGATGCAAAAAATCTTAA ATAAACTACTAGCAAAAA GACTACAGCAAGTTATCAC AAGGATTATTCACATGAC CAGTGGGATTATACCA GGAATGCAAGGATGGTTT AATATTAGGAAAACCATCC ACATAATTGACCATATCAA TAATCAAAACCAACAGAAAT CACATGATTATCTCAATAG ATGCAGAAAAAGCCTTTG ACAAATACACACCCCAT CCTATTGAAAAACACTAGAA AGTATAGGAATAGAAGGG CCATTCTCAAAAATAATAA ACAGTATATATTTAAACC ATCAGCAAGTATCATCTGC AATGGGACAAAGTTAGAA GCCTTCCCAATAAGATCAG GAGTGAAAGCAAGGATGCC CATTATACCTCTATTATT AATATTGACTAGAAATGC TAGCAGTAGCAATTAGAG AAGAAAAAGAAATTGAAG GGATTAAAGTAGGCAATG AGGAACTAACTATCACT	AAGCCACATACCCC ACCAACCCACAT CTGCTGCCAGGT TCATGAACCTAAGC CCAAGCCAAATCAG ACCCTGAGATCCTG CTGGCAATAGT GGTACAAGCCAAC CCATACCCCTTGAA ATTTCAAACTGTG GAGAAGTCCAAGG CAATCTGGGCTGA AGTGGGCTGATCC ATGTGGGCCCAGA GCAAAGCCAGGAA TCCAGTCTGGGCT TGGATAAGGACA TAATTCAGTTTG GGTGAGGTGATA GTACTCAAGGGA GCCTCACTGGGATC TTTTGCCTAAGC TCAGGGCAGAGCT CCAAGACAGGGCA ATCAAGGGGTGTC CTGATTGGATCAAG TACACAGTGAGACC AAAAAATTGAGCC TAAGCCTGGGCT AGAATTTGATCAGC CCCTGACCTGATCA AGATCAGAGTGAG ATCAAAAGCTTACA CCCAAGCCTGAGG CAAGTCTTTAAGCT ATCAGCATCTCAAG GGTCAATTGAATCA GCAGGGGAGGG	AAGCTCTAAGAAAAATC CCAGGGCCAGATGGATT CACAAGTGAATTTCTATCAA ACATTTAAAGAACTAACTAA TCCCAATACTATACAAAT ATTTGACAAAATAAGCAA AGAAGGAGTCTTACCAA TTCTTTTATGACACAAAT ATGGTACTGATTCAAAAGC CAGGCAGACCAAAAAACAG AGAAAGAAAACCTACAGAC CAATCTCTTAATGAACAT AGATGCAAAAAATCTTAA ATAAACTACTAGCAAAAA GACTACAGCAAGTTATCAC AAGGATTATTCACATGAC CAGTGGGATTATACCA GGAATGCAAGGATGGTTT AATATTAGGAAAACCATCC ACATAATTGACCATATCAA TAATCAAAACCAACAGAAAT CACATGATTATCTCAATAG ATGCAGAAAAAGCCTTTG ACAAATACACACCCCAT CCTATTGAAAAACACTAGAA AGTATAGGAATAGAAGGG CCATTCTCAAAAATAATAA ACAGTATATATTTAAACC ATCAGCAAGTATCATCTGC AATGGGACAAAGTTAGAA GCCTTCCCAATAAGATCAG GAGTGAAAGCAAGGATGCC CATTATACCTCTATTATT AATATTGACTAGAAATGC TAGCAGTAGCAATTAGAG AAGAAAAAGAAATTGAAG GGATTAAAGTAGGCAATG AGGAACTAACTATCACT
--	--	--	--	--	--	--	--

	<p>CTTTGCAGATGATATGATG GTATACCTTAAAGAAATCCTA GAAATCAACTAAAAGGC TAGTGAAATAATTAACA ACTTTAGCAAGTTGCAG GGTACAAGATAAACCCAC ATAATCATCAGCATTTCT ATATATTTCCAACAAAAT CAGCAGCAGGAGTTAGAA AGAGAAACTCCATTTAAAA TCACTCTAGACAATATAAA ATATTTGAATCTATCTGCC AAGACAAACACAGGAAT ATATGAACACAACTACAAA ACACTTTTCACACAATTAA AACTAGATCTAAATAATTG GAAAAACATTAATTGCTCA TGGGTAGGATGAGCTAAT ATAATAAAATGACAATCC TACCCAAATTAATCTACTT ATTCAGTGCCATACCTATC AACTACCAAAAACTTTT TTATAGAAATTAGAAAAA ATATAACAAAGTTCATCTG GAAGAACAAAAAGATCAAG AATATCAAG</p>	<p>GGCTCTCAGAGCTC TCAGCCCTCAGACC ATCAGGTAACTG GTAAAAACCCAGA AAATAAGCTGAGA ATAACATCAAGGA AGGACTGAAGTTT AAGAGGTACCCC AACTCCAGAAAA CAGAGCCTACCTAT AATTAGATAGGAA AAATGAGCAAACA ATCAAAAAAGCACCC TAACCTAAAGAAT TTTTATGGAGACA AAGAGCAAGGTAC AGACACAGAAAGG GACAGTGAAGCA AAGGAAACACATG CAAAGTCCAAAAG AAAAATATGGATTG GACACAGATTCTG GAAGAGCTCAAAA AAGACTTCAAAAAC CAATTAAGAGAGG CAGAGGAAAAAGTG GGGAAGAGAAATG AAAGTGATGCAAG AAGAAAAAGAAAGT GATGCAAGAAGAA ATGAAAAAGTGATG CAAGAAAGAAATGG GCCAATTGAAAA GGAGAACCAAAAA CTGACAGAGGAAA ATCAGGTCTTAAAA ATCAGAAATTGACCA TCTAGAACTAATG</p>				
--	---	---	--	--	--	--

ATTTCATGAGAAAT
CAAGAAACAATAA
AGCAGAATCAAG
GAATGAAAAAAC
AGAGAAACAATG
AATATCTTATTGA
AAAAACAACGTACC
TAGAAAAATAGATCT
AGGAGAGACAAT
TGAGAATTATTGGA
CTACCTGAAAGCCA
TGATCAAGAAAAA
AACCCTGGACATCA
TATTCAAGAAATT
ATCAAGAAAAACT
GCCAGAGATCTT
GAACAAGAAAAA
AAATTGAAATTGAA
AGAAATCACAGATC
ACCTCAGAAAGA
AATCCTCAATGAC
AACTTCAGGAATG
TAATAGCTAAATTC
AAGAGCCACAAG
CCAAGTCAAGAAAT
ACTTCAACAGCCA
GAAAGAAACCATT
AAATACCATGGAG
CTACAATCAGGATC
ACACAGGACCTAGT
AGCTTCTACATTAA
AGGATTCAAAAGGC
ATGGAATATGATAT
TCAGTAGAGCCAAA
GATTGGGTTTAC
AACCAGAGCTCAC
TATCCAGCAAACT
GAGTATATCTTTC

AGGGGAAAAAATG
GACATTCAATAAGA
TAGAAGATTTCCAA
GCATTCTGAGGA
ATAAGCCAGACTA
ACAAAAAATTGA
AGTCCAAACATAAG
ACAAAGAGGAAGC
CCAAAAAGGTAAA
TAAGAAAGAGAAA
ATTTAAGGACTCA
TTAAGGTCAAAATG
TTTATATGCTTACA
TGGAAAGAGGATT
TCTGTAATCTCAA
AATTACTCTTAGT
AGTATGAGGAATA
GAAGGAATTTACTC
AGAAAGAGGGTGG
AGAAGTAAGCTGA
TTATGATGATATGA
TGTTATATGATGATA
TATATGTAATGTG
ATATGATGGAATG
ATGATGATATGA
TATGATATGATGC
ATATGAATGATATG
TAAAAAATCAATCA
AGGCTGAAAGAG
AGGGTTGCACCTGA
GAGAAAAAGGGAAG
GGAGAGATAGAAAT
GGGGTAAATTATAT
ACACAATAAGAGGC
ATGAAAAATCATTA
TAGAGAGGGGAGG
ATTAAGGGTGTGTA
TGGGCAATGCTTAA

ACTTTACTCTCATT
GTAACCTGGCTCAG
AGAGGGAAAAACA
AGATACTACTCAGTG
GGTATAGAAATTCTA
TCTTACCCTACAGA
GAAGTATAGAGGGG
AATTAAGACAAGGG
AAGGGGGAGGTAA
TTTGGAGGAGGGG
AAGTGGGGGGGTG
ACAAAAAGCAAAA
TACTGGTAGGAG
GAACAGAGTAA
GGGATAGAGCAG
GATCTAAAGGGGA
TAATATGATGGAG
GGCAATACACAGTT
AGTAATCATAAAC
TGAAATGTAATGG
GATGAACCTACCCA
TAAAAATGGAAGTG
GATTAGCAGAGTGG
ATTAAAAACAGAA
TCTTACTATGT
GTTTACAGAAACA
CATTTAGGCAGG
GTGACACACACAG
ATTCAAGGTAAAA
GGCTGGAGCAGAA
TTTATTATGCATCA
TCTAAAGTAAAAAA
AAGCAGAGTAGC
AATCATGATACCAG
ACAAAGCTAAAGT
AGAAAATTGATCTGA
TTAAAAAGAGATAA
GGAAAGAAATTAC

				ATTTTGCTAAAAGG TACTATAACAATG AAGTAATATCATTA CTAAACATTTATGC ACCAAATGGTATAG CATCTAGATTTCTA A				
Poseidon-1_PM	GCF_010 993605.1	Petromyzon marinus		TGATTGATTGTCACGCTT ACAGTGGCTTTCAGACAG CATACATAAAAAATTCGT CTCCCCCTCTCTTTTCAA AAAAATAGATGGACTAGTC CTGCTGGGCAAACTGCCA AATCGTTGCGCGTGACGC AAACGTTTTTTTTAAATAA ATAAATATGCGGTTTTTCAA ATAAATAAATTTATGCAAA CCGACACACCTGATGATG ACCGTTTGAGTTATGGTCG AAACGTCGTGAGAAATTTT ATTGTTTTATATATATTTG TTATTTGTTGTTCTCTTC AACGTGACTTTACCGAAAA AACTAAGAAATATGAATTAC TGCAGTCACGAAAGCTTTA AATCAAAA		cd00 304	440	524
Poseidon-1E_PM	GCF_010 993605.1	Petromyzon marinus		TAATTAATTGTAATTTTACG ATGACCTTCAGACCGCAGC TCTTTTCGCCAGCATTCCT CTTAATACCTCTCACACCA CGGTTAATGGTATGGTTTA CAACATGGGATACTAGT GTTCACTATATAGT ACTGTGAATAGCTC		cd00 304	453	537

AGATGTAGGCC	TCCCTCTCTCTGTTCCGA	PPSNISKDERFELSQLKRRDDIWLVSADKGN			
TTGACTTCAACATA	CCATGGACGAACCTCGTCC	TVIMDKTEYNNKIQELLNDGYSYTKIKRDPNTS			
GTCATTTCTGTGAC	GCTGAACGGGGCCCGTA	LVTNTKIUSDSTALIDENKQFSLKPSAARPPRL			
CAGTCACTAGCAAA	GCTATCCGTGGTACGAAAT	YGLPKHKEDTLPPIVSTISSPTYKLARFLADIL			
TCACCCAGTTATTT	TTGAGGTAAGAAATCCAC	RPFTGKTTSAVLNSTHLVSTIKDISPNDLLVS			
CAATGCTAAATTCG	CCATTTTTTTCGCATATA	FDVSLFTKVPVPTDIAIEDLVHQGLHTDVP			
CCTAGTTATTTCAA	GCAITGTGATATATAACGA	LVRHCLTNTYFTWNGTFYKQTEGTPMGSPLS			
TGCTAAATCTCCT	CTTATGGTTGGTAGCTTAG	PVANVFMKEFETDAIQATLRLWKRVD			
GTTTGATTTGACT	CTGATTGGAGAAATAGA	DTFVIWPHGQPALDQFRDHINSIHPAIRFTM			
AAGTGCCTCGGTG	GTTTATTGATAACATGAAG	EVEKDGLPFLDLIKRRHDSMSHSVVRKAT			
AATGTGATGCTCT	GTAATTTATATTTAGTTTT	HTDAYLKPDSHHPSQKNCLIKLYHRAQCIS			
GACTTCAACGTAAC	TTATTGGTTCACTATATAG	DOEHITSEKLVHQALRANGVSDYTIKAITP			
TATTTCTGTGAGCA	TACTGTGAATAGCTCAGAT	VMKKDPLVEPLKTTITLPYIAGVTKKISKILNKHK			
GTTACAAGCTAATC	GTGTAGGCCTTGACTTCAA	IQVRENTVHKIRDLPVSKDALPDQIYPGVVK			
ACTCAGCTATTTCA	CATAGTCAATTTCTGTGACC	LKCLGSSYIGETKRHISDRIREHKADLKHGKT			
ATGCTAAATTTCTAT	AGTCACTAGCAAAATCACCC	NSSAVADHCYNSVSGSHDIDFSKTEVLCKPTGY			
GCTAAATCCATTGT	AGTTATTTCAATGCTAAAT	HERFVHEAIIHYKHRDNFNREDGFRLSNHWK			
CAAGGGTGCCTC	TCGCCTAGTATTTCAATG	DVLNHFCS			
TGCGGATGTGAT	CTAAATCTCCTGTTTGATT				
GCCTTAACCTAGCT	TGTACTAAGTGCCTCGGTG				
ATTTCAATGCTAAA	AATGTGATGCTCTGACTT				
TTCTATGCTAATTC	CAACGTAACATATTTCTGTG				
CTTTAACAAATGGGT	AGCAGTTACAAAGCTAATCA				
GCCTCTGCGGATGT	CTCAGCTATTTCAATGCTA				
GTATGCTTTAACTT	AATCTATGCTAATTTCCAT				
TTAACATAGCTATT	TGTCAAGGGGTGCCTCTG				
TCTGTGACCAGTCA	CGGATGTGTATGCTCTTAAC				
CAAGCTAATCATTC	TTAGCTATTTCAATGCTAA				
AGCCATTTCAATGC	ATCTATGCTAATTTCTTTA				
TAAATTTTATGCTA	ACAATGGGTGCCTCTGGG				
ATTCCTTTGTCAAG	GATGTGATGCTTTAACTT				
GCTCTTTGTTCTCT	TTAACATAGCTATTTCTGT				
CACACAGCTTAAAA	GACCAGTCACAAGCTAATC				
GCCGTGATCTGAAC	ATTCAGCCATTTCAATGCT				
GTGGCTCTCTCACT	AAATTTTATGCTAATTTCTT				
CCAGGCTGTCCTA	TGTCAAGGCTCTTTGTTTC				
CCAGACAGTCTCGC	CTCTCACACAGCTTAAAAAG				
TCCCTTCACTGAT	CCGTGATCTGAACGTGGC				

				GAAGCGCGCTTGT GTTGTGGCCGAAA CGTGTGAGAAATAT TATTGTATTATTA TTTTATCAATTTAT ATTTTATTCTATT TACGTGACTTTACC GAAAGAACCAAG AAG	TCTCACTCCAGCGCTGT CCTACGACAGTCTCGCT CCCTTCACCTGATGAAGGC CGCTTGTTGTGGCCGA AACGTGCGAGAAATATTAT TGTGTTATTTTATTTATCA ATTTATTTTATTTCTCAT TCTACGTGACTTTACCGAA AGAACCAAGAGATGAA TCATTGCAGTCACGAAAGC TTAAATCGAAA			
Poseidon-2B_PM	GCF_010993605.1	Petromyza marinus	T	GAAGCGCGCTTGT GTTGTGGCCGAAA CGTGTGAGAAATAT TATTGTATTATTA TTTTATCAATTTAT ATTTTATTCTATT TACGTGACTTTACC GAAAGAACCAAG AAG	TCTCACTCCAGCGCTGT CCTACGACAGTCTCGCT CCCTTCACCTGATGAAGGC CGCTTGTTGTGGCCGA AACGTGCGAGAAATATTAT TGTGTTATTTTATTTATCA ATTTATTTTATTTCTCAT TCTACGTGACTTTACCGAA AGAACCAAGAGATGAA TCATTGCAGTCACGAAAGC TTAAATCGAAA	cd00304	361	445

					TTCATCATGCTAAATTATA TGCTAATTCCTTTGTCAAG GCGCTTTGTTTCTCTCAC ACAGCTTAAAGGTCGTGA TCTGAGCGTGGCTTTTTC ACTCCAGCGCTGCTCTCCA GACAGCTGGTCCCTTCAC CTGATGAAGGCCGCTTGT GTTGTGCCCGAAACGTCG TGATTTTGTATTATTTT TTTACTACGTGACTTTAC CCAATAGAACCAAGAGT ATGGATCCTGCAGTCAGG AAAGCTTTAAATCTAGAA				
Poseidon-1B_PM	GCF_010 993605.1	Petromyzon marinus	CA		TAACCTGCTGATGTTTC CATCTACTACCGATGTT AGCAGTTATTTTCGACTCT CTGAGTTACATTCACAAAC ACAGCCAAACGTTAGCTGT GTTAGTAAATGCGGCTT GCCACTGCTCGTCCATGC CCTGGCCGAGCTTGCCCC CCTCGATCTTACATTTAC ATGCAAAATTTCTCTCTAC TAGCCCGCCCACTCGCTC CTACATGACCCCTGGCTGC TACACCATCTCTAAACCT TTATAAACAGCCTGCCTCA GCAGCTTTCCCTCTTTTTC ACCTGATGACGATTGCTTG TGTTGCGATCGAAACGTC GTGATTTTAAATTTTATTA TTTTTCTATCTACGTGACT TTACCGAAAGACCCAAAG AATATGAAT	cd00 304	1	83	
L2-24_ACar	GCF_000 090745.1	Anolis carolinensis			MGSPSPVAVNVFMKFETDAIQASTLRPSL WKRYVDDTFVWPHGLPALNQFLNHINSIHP AIKFTMEVEKDGKLPFLDLVLEKRPDGSMSGHS VYRKATHTDAYLKPDSHHHPQKNSLIKTLLHH RARCSIDHHHLASELEHVHQALKMNGYSYDT IRKAITPTVGKEPSEDPKITTLPYIAGVTKNISKI MSKHQVRFNTVHKIRDLPISVKDVIPDMQY PGVYKLSVCVCGSSYIGETKRHVSDRVRHKAD LKHGRNTSAVADHCFNAVGSHEIDFSKTKV LCKPTGYQQRFFVHEAIQIKYKRRHNFNREDGY KLSNHWKDVINQFKS	cd01 650	257	517	

				TGGCTCTGACTTTTGGCCC AATGCTACTGGCCCTGTTG TAATGGCCACACAGTGCTT TACTGTTTTAAATGGGTA TAATATTGGTTTTTAAGTG TGTTTTACGATCGTTTGT TTACTATATTTGTATTATA TGTGGCATTAACTTGCCA TTATGTAAGACCGCTCTGG GTCCCCCTGGGGGAGAAG AGCGGTATATAAATTAAT AAATAAATAAATAATAA ATAAATAA		DRARAKALRKAYSVALRAAKSKFQTARIAAK NRPSLFRVVGELINGTPQTGPDDLAENCA AFSHHFADKVAQRSELDATVDASEVARA PVCPIILMDSFLVQPDDVKILGEVRATTCAL DPCPSWLVEARGGLVWVRVINASIREGI FPSCLKQAOWKPLKKPLSDPTNNNRYRPSIN LPFLGVLERVASQLQDFLDDTFDLPSQSG FRPGHSTETALVALVDDLRELDGRGSVTLLVL DISAAFDTIDHGILLGRLLGMGLGGTVLQWL QSFELEGRSMVMKLGDTCSDPWPLTCGPVQG SILSPMLFNIMKPLGEVIRSFVGRCHLYADDT QLYYSFPASKEAPQILNQCLAAVSDWMRA NKLLNPDKTEVLLVSRKSDRGIGWQPVLDG VALPLKAQVRSIGVLLDSSLTEAQVSAVAGR AFAQLKILVRLPYLVKSDIATVVHALVTSRL DYCNALVYGLPLKTARKQLVQRSAA RLTTGA NYRERTTPLFKELHWLPVIFRAQFKVQVITYK ALNGLGPTYLSDRISPYPEARRLRSSGEALLSV PPPSQVRLVGTRERAFSAVAPRLWNLSPREIR QAPTLLSFRKQLKTMWMDQAFGPS	cd01 650	470	731
Jockey- 8_DPer	GCF_003 286085.1	Drosophila persimilis	GGCTAATGCAGCA CACAAATTTACTTG CCTGTGAGATCGCA ACGCACCCATACAT TCCAAGAAAGGCTG GCCAGGCCAAGGT ATGCCGCAATGATA GAAGAGCACAAATC CAAACAGCAACAG AAGCCTGACAGCC ATCCAAACAATACT GCTCCAAGTCAGAC GCGGAACTGGCT GCGCAATCCATTG CAACACACAATATG GAAATCAGCCGCA GTAGCGCTAATGC AGCTGTACGCAG	C	MSALPPANAGDGNGWFMKRDYQFSQQ DEHNRNIPHISLAPTREDISIQMCTIAIKIN HQTNIGAIYCPPAHEWIDTFKQVFDSLGR FIVAGDWNAKSQWWGNARSRRGAALLKCI QENDQSILATGSPTHFPNSRRHMPSAIDFAV YKGIRNLHIETSYDLGSDHNMQIIKLLRPEY GKKILLPKNANIRRFOAHLEQRIHLETSLHN GDDIDDAVAILEKNIYIAAEELASQTSSNQHTVY GPQNFDLSDSQLKAIKIRILCKINQKDRCNPN LIARTINQLNTRISKRIKLQKQKMNRLLHDID PTDRYSMQKLWRLNTTIKRQIPINLPVRRHV AEGQOAPADPWCKTTEEKAEFAQHLEARF TPKITAALEDREEINAVLQENGPDAAAQIRPV TIQELKSEIKLTKKAPGLDSIDNRVIKSLPTS AIEYLLICNSALQFGHFRKSWKCAKISMRLKP GKPSDRLESYRPISQLTGFSKILERLUILKRLFEE SFAMAIPNHQGFGRSQHSAEQQIGRVTOFIL RSYEAQRHCASVYMDIQEAFDRVWHQGLLC				

	R4-1_PH	GCA_001 587735.2	Parhvale hawaiensis	ATAAGAGTATAAAC AAATAAAACGGTTC TCTCTTC	CGACCGCCAGCG GGATAACTGGCAA ACCTGTCTCGACC ACGGCCCGTGAA TCCATCGGGCGGT ATGAGTCAGACAC AGGGGGGTGTTTA AGGTGACCTGTGTG CGAGGAATGCGC AGCAAAGCCGGA TGAGCCTTAGAACA TCGAAGGCCAACG ACAACTGCCGAAA GACTGGCGGCTAA TCCAGTCAACCTCC TGAGGTCAACCGG CTGGTCATGCTGAA TCTCAGCTTAACCA GGCGTACGAACTT ACAGTTGGAGGGT CGAGCACCCCTGAT GGCTGAAAGGAC CATCAAAGTCGAA GGTAGCCAACGAG AAGACCAACGGCT GATTGAGGCGGAA GAGTCAACTCGTTG AATGCGTTCGACA GCCTGGGGTAGAT GGAACCTCTAAGCC CTGAAAGGCAGTC CATCTTCGACAGC CTAAGGTGCCCGCC CGCCTGAGGGTTAT CAGGCCCGCCCGC	TGAGCCTTAGGTCGCGGG ATGTGACCCGGCGCAGA GTGTAGAGCTGAACATCG GTCAACGAGTCCAATTTGG GTCGTGAAATCCCTCGAT AATAATAATAATAA	MKMSHNDRTPSNGVKGTSVRLGTSLVRSPV GEAGAVRERGTHPSESVSQSDSDASVNATGE GSVREQAPLSPPGAEEATVPTQRRTRHKWSK EDRWVLWECFVASKREGPGYLRKLQKLWDER GIPGNFQASLSGQIRQCSKNLLSEERLQJA ARMEAQVASADEPARQVPTRPVTPRSP VEPARPSIPSEETPDILGAVPSEIDSDPNRSP SRGPRHLPAHNMSQSESDVTDVDPVSDQQ RSDSLEPRDLLRNSSESTPGHPNQELSDTLLS NVVPSEIDSDPNQSPRRGPRHLPAHDMSL DSMDEETPDLSQQRSDDLLELRDLRNSVE TPPKGPHSLRHLPEPKIRAAAFRVNSVIGKIHT NNITETNALIKAGADLAVRILEVQRPRTQR KKDPPWKRLEKNIAIRKHLW/ISEWRRGN LHDEEKTLLSRYRCLEVGLTNLEDTLKQRLS AKRSKVRREFARVAGFHQNLQNFNTNQKRLY QTLRGEETSSDPNAEESIRFSWDIWSKEVRH NNAEAWLHDVKEKNVAADPDLTITSQQLKK QLSKTKNWKAPGDMVQGYWIKFTSLHSRI AAQLNHCLQRGTVPMTWTTGKTVLQKDKA KGTEVSNVRPITCLPLMWKVLGTIYERVYQH LDSKLLPDEQKGCRRNTRGTDHLLVDKLLT KDARSKKNLSMAWVDYKKAQFDMVPHSWI LECLDIYGIAGNIRNLIAITMPNWKQLTSAN KHLGEVSIRRGIFQGDLSPLLFVLTMIPLSETL NKAGQGNYSRTMKLNHLLYMDDDLKYAKS KDQVEQLLNIVHQYSQDIKMQFGISKCVLNI ERGEVTASEGIIIEGTIKDIEETGYKLGVMY NTILHRTMKDSIRKEYLTRLLKSHLNGGNTI KAINTWAVPVWRYSGAGIINWTKDCTDMDIK TRKLMITYRALHPRSCVDRLYNRRGGRLIS VEDCVAEKRALSQHFRESDDPWARCLIEAK LLKETADQFKERRRLDRTNKWKSMKMSG QYLEAVQDKIVPDSWNWLLRGELKRETEGTI LAAQEQALRTYIQNKIDKRNVPSTCRICRSS	cd01 650	616	879
--	---------	---------------------	------------------------	---	--	--	---	-------------	-----	-----

				CTGAGGGTTACCA GGCAACC				DETINHVISECGVLAQKEYKRRHDKVARHLH WTLRIHNPVSERWYHEPAPVVEAVEAQI YWDKRMETDRVLHANRPDIWVKDQKESAK LIDVSIPIFDSRIIDKEAEKKEYRDLAIELQRLW QMKVDVVPVVGALGAMSKNLKTALRELKC GHLHPGTLQKSALLGTAHIRKVL	cd01 650	85	341
	Proto2- 4_CTe	GCA_000 328365.1	Capitella teleta	GATGAATTGTATAA TTGTGTTGATTATG ATCGTGATTCA				MNELIAFFKISVQNVCDQSCYSYSHQTTAQG VRSVASKLKDSDCKDNFMSNHVIHGGDTLF VHLSFLNCMISHTCDPKQLRMSVLIPIPKNR RKLSVSTNYRSIALSSIVGVFDRMILTEHVNI FSTSNLQYGFKPKHSTVQCTFVLQNVINHYS RGSPCVMLLDASKAFDRVQVVRFLYLLRERG LCARIFMLALLYNQSMCVRWKNCISDNFS CSNGIKQGGVLSPTLCYFDELLIRLKHCGSG CHLGHFTGALSADDDVALLAPTLTAAKVML NVCEQFANNFNVLFNADKSALLVCGRPIKE VNLRLNGRIQQNSTAVHLGTTIGDINSSTRNV KKAASDLYCLTNVLMCKFGHCSIDVLSALFFFF FFYCSSFYGCPLWDLRNIPILTAWRKCLKRI WKLNIIRTRKYNNLPKGNLISALLIRLSFYM QCLSSNAIVKASNYCFYGCMSYVISENVHL CSVLSSVPLRFIFDDFENVKSIQRGMVDEATI AESLALKDLCYLRDGLMFLNFETISEELLCAFN HFTI	cd01 650	465	731
	R4-1_AC	GCF_000 090745.1	Anolis carolinensi s	CCCCCCCAGGTG CCACCTTGACGTGG TGGGGGGGCTTAG CTGCTCCAATGATG ACTGAGGGCTGTG CTGGCGGTAGTGT AACTACCGGCAGG TCCAACCAAGCCAG AGAGGCCTCAGCT GAGGAGCAACAAC AAGAGCACCTAACC CATTAGCATT	TAGACACTTGGGAAGTGT TCGACTTGTGGTTTGTGA AACGAAATCCAGCATATCT ATCTTGTGTGTGTGTCAT ACAACGTCGTTGTGTGAT			MENQTQTSILARSSPRIKRTAVDAADSGHP STGLRNHQNPNEQSQKQKQKYMENRTIM KCYKSEPPRRGYQKRMHQLWKQEPDSQI TEPRADQRRFIIRNKVFSEVELEEIQKICKVNY HQTTAQTAETPATLGMVEQIEPEESVELLQ EFEPEALTSVEPPGTLTARQQELKDKIMAH AANAIRKRLPTLKTVPKRHLAPLMKDVNAALS TVQITSIEQINQYASAAVIVTEELGLLQPRQP QRKSTGPKWKVRLELIKLRSDASNLNKIM KEKLNKNDKIKQYLIRKYWLNRKIEEALEIVKE QITATARKIERYEARIIQYRQNLQFSQDQRRFY QSLNQITDVTIKPEKTAATKFWKELWENNK NYNKNAGWIKFEGFKSQNMELMETTEM ISKRVQKVNWNTSPGSDQLHGFWLKHLISLH	cd01 650	465	731

CR1-66_Aae	GCF_002204515.2	Aedes aegypti	ATCGCTGTTGCTAC CTGA	TAGATTTTATGTTTTTTT TTCATTAAAGACAAACAGAT GTCAGATGAATAACAACTT CAATAAATAAATAATAAA TAAATAAATAAATAAAT	GKMAQQFNEMLQKGSIEWLTTGRTYLIQK DPAKGAAPGNYPITCLPTMFKLLTGIIADRIQ DYLEENILPDEQGNKRKSRGTQDQLIDKM ILENCKSRKANLHMTWIDYKKAEDSLPHSWII KCLDAIGICKNVGTFIENMMEHWKTELFVGN ESYGLVNIIRRGIFQGDLSPLLIAMIPLSTILQ KTNLGYQTSKNSHKISHLMYMDDLKLYGKTE TEIQSLTNTVRIFSTDIINMEFGLDKCSTVALKK GKIIIEGINMPNGQTIKCHQPEAYKYLIGLQL DNIKHEHVKTVVSKEYTQRVRKILSKLNGGN TIKAINTWAIPIRYTAGIINWTQVELDNLDRK TRKLMTHHSLHPRSDVDRLYLPRRSGGRGLL QVKQAVKEEEHALAEYVKQSEEPALIEVKNQ KLLKTQQTKNQYKKTALQTRADSWHNKTLH GKFLDKIEGKADKEKTWLTWLTNGTLKKEG LILAAEQEAI RTNAIKAKIEKADDPKCRCKE TDETIDHILSCCKIAQTDYKQRHNYYVAQMIH WNLCLYHLPAAKNWWDHKPAKVLENEHA KILWDFRIQTDKVI LEHNTPDITVVEKNKVWII DVAIPGDSRIDEKQEKLSRYQDILKIELQLW QKPVQVVPVVMGTLGAVPKDLSRHLETIDID KITICQLQKATLLGSARIIRKYITQS	cd01650	638	885
------------	-----------------	---------------	------------------------	--	--	---------	-----	-----

	DongAa	GCF_002 204515.2	Aedes aegypti	GGTGGGAGCTCAG GCAACATAACTCCT GGCGTCCATGAA AAACCAACCATCTTG AGTAGGTGGAGC TTGAGGTAAATAT CCTGGCGTCCATG TAAACCAACCCCTCG GCGAGCAGCGGCG CTTGAACCAAGCTA TTAAGCACTGTAAT TGGAGGAAATGCC AATGCAGAACCCGT AGCTTCGGGA	TGAGAACTCATCCAGCA GACTCATTAGGATAAGTTA AACCGACCAATGAGATCC ACAGAGCTAATCCCTTT GGCATAAGATTGCCCGG GGTAGGTGAAAGTTTCCA GCATTTTGTGCTGAGAAAT GCCAAACTCTATAATAAT AATAA	FVNSKRKEVGLPRSVHLDGGQTATSEQEKCFIF AQHFQKTFHDSPATADQIAAAIEYTPRDVFD FRSFTIDEQTIQAARFKLKMESHAPGPDGIPSCI LKNCSHSLATPLAKLFSRSLQEGKEPDEWKVS SMFPVYKGDKKDVHNYRGITSLCSCSKLEIL INDALFNSCKSFISTAQHVFFPKRSVTNTLTF VSLCLTDSGGQIDITYDLKAAAFDRV/DHGIL LAKLDKLVSHSLVRWFESYLTNRLLYVIGSS QSEFTNLSGVPQGSNLGPLELLFELNELSQLL PPGCRIFYADDVKIYMIIDSTADCELLQSRVNI MANWCSSNFLTILSVNKCNIISFHRQKPIFD YRICDTVLRVDQVKDLGITLDQEMINFKPHY NDVIAKANRQLGFIFKLSDDFRDPLCLKGLYCS LVRSLFESVWWCPYQSCWIALESVQRRFL RYAFRHLWRDPLNLPAYEDRCRLLGIETLEQ RRRAQAVFAAKLIVGDIDSPPELLRQLNIYAE QVIRQRNQLVPCNRLYGMNEPLRASADAL NEAFTAFDNPLISFIRRLSL	cd01 650	530	802
--	--------	---------------------	------------------	---	--	---	-------------	-----	-----

[illegible]

[illegible]

				AGGTATTGGCGG CAAGCGCACTCTA GAGAACACAATA CCACGTAATACTGC AATAT				AALYADDLVWCSEEFVTTATYRMQLALDKL TTWTKWCVTANKEMSSATLFTLSQAEPKHL TLDKTPKFENQQYLGIFDKRLTWKQHIQN AENKARKKLINMRKLAGTQWRANEKTLRQV YLGAIHPHLEYGSSAWMTAAQTHKYTLEKVQ NQALRIITGAMRSTPIEKMEQKTSIVALQRR WECKALVQYTKATAMEDHILHTRTERSN TAR LKRSNFIKESRKIQRSIQDHLPATVNMNLNHTP EAPSWEEKNTNISIHTSIPMFTTKDDHSDLSK KSLTSLMLEERYPREAWIRIYTDGSDATDAVKR GGAGVYQFPNRRRQAEAVPTGRHCTNYRA ETEALVQAANIKNVVDPQSVVFLTDAMSV LQAINSGKLPRLRKEIEDIPCLRIMLQWIPSHC GVLGNEQAQDRMAKLGAKKEQPDNSVNYTEI KTIINTLFKAPQTPDGYHELRSQEQVIIFRLRT QHTRLRLMYKCFKIVQSPKCQCGEADQTV HILQHCKTHRALRQETWPTPLTLEEKLYEPPD ELTRTAQFMVTDGLQV	cd01 650	457	716
	JAM1	GCF_002 204515.2	Aedes aegypti	TTGGCGCGGTAAA TGGCTGGGCATGG CGTACCATTTGGTAC CTCGGTACTCTGCA GGAATAAAATAGA CCCCTTTGTGCGGT CCTTAGCCTCTTGC CCAGCAACTCCTAT CCCTACCTCCTCGT GGTACTGGCCGGG GTACGAGTAACTTT GGGGAAGATCGGG TAACCAACCCCGG TGGGAACCTTTGGTC GTATGCTGACAGG GAAGGGGGGGTTTT GCTTTTGCAACCT GGAGCGTCTGTACT CCACGTTAGGAGC GGCTCAACACAGC	TAATTGATGTAAAGCCAAA TAAAGTAAAGTAAGTA AG	MGMQRRVIGWWPINERMCKLRIGRFFN FSIINVSPHSGSTDDDKDAFYAQLEREYDRC PSHDVKIIGDLNAQVGQEEFRPTIGKFSABR LTNENGLRLIDFAASKNMAIRSTYFQHSLPYR YTWRSQQTESQIDHV/LIDGRHFSIDTDVRTY RGANIDSDHYLVMMVKLRPKLSVINNVYRRP PRYNLERLKQPDVANAYAQHLEAALPDEGEL DRAPLEDCWRTVKAINDAAESVGVYVERSS RNDWFDEECQEVLEEKNAARAAMLQHGTR QNVERYRLKRQQQTRLFRDKKRRLEEVECCQE MELLYRSQETRFYQKLNTRKGFVPRAEMIC RDKDGSILTDGREVIERWKQHYDEHLNGAEN TGTEGQDSEGDGVSTADSGNQNPAPTME VKDAIQQLKNNAAGKDGIGALIKMGPDRL AACLRILVRIWETEQLPEEWKQGVICPIYKK GDKLECNVRAITILNAAKYVLSQILFRRLSPIA NEFVGSYQAGFIDGRSTTDQIFSVRQLQKCR EYQVPTHLLFIDFKAAYDSIDRIELWKIMDEN SFPGLTRLIRATMDGVQNCVKISGEHSSFE SRRGLRQGDGLSCLLFNIALEGVMRRAGLNS					

				GTCTGTTCCCATG TCAGGGCGGCTG ATCATGTCGAGT GCCAGAGAAGGAC TCTAAGCTAAACTG CGCACTATGCCCT CCGAACATTTAGG GGGAAATGGTCTC CGGAAATCTAGGG GGTGGTGTGAGG CCCTGCAAGCCAAC CGTAAAAACACATC AGCACAGGACGT CAACGAGAGAATA CGGACCGGAACAA TCGGCAAAAGACA CAGCGACGAAAT GGACTAGCGATTG GAACTCGGTACG TGGAACCTGCAATC TCTCAACTTCATTG GAAGTACTCGCAT CTCTCCGATGTACT GAAGACCCCGGT TTCGACATCGTAGC GCTGCAGGAGGTG TGCTGGACAGGAG CATTGGTGCGAAC GTTTAGAGGTAATC ATACCATCTACCA AGCTGCGGCAACA CACGCGAGCTGGG AACAGCTTTCATAG TG	TAGATTCAACGCAATAGCA AGGAATTGTTGCCCATCAA GTATCAAGTAAGTAAGT	MGSPNFSPTTKNGERRNNERYFRQPTWQR NKDYEWKLGWNVRLNEPGRASLLARELQ KVGVGVAAIQEVWRPRSGEREFRAVDPIANT TFKNYHSGGDKAEGHGVFVIGQMKRVM	cd01 650	561	821
				RGTIFRSGQFVCFADDDMDIIGRKFTVADLF TRLKREATRVGLMVNASCTKYMLVGGTERD RTRLGSSVTIDGDTFEVWDEFVYLGSLTADN NVSREIRRIISGRAYYGLQKLRSRKIHPRTK CTMYKTLIRPVVLYGHEAWTMLEEDLQALGV FERRVLRITIFGGVQENGWRRRMNHELALQ YGEPISIVKAKAGRIRWAGHVARMPDNNPV KMFATNPVGTTRRRGAQRARWIDQVHQDL ESVGHSRGWREAAAMNRGNWRNIVGEALSR					
	RTE_Ele2	GCF_002 204515.2	Aedes aegypti						

				GGTCCTCGCAAGTC CCTACTCACGCTT CCTCGGTCTAACG ATGACAAAGACCG CCAGCTAAGGGTT GCGTACTTAGCTGG TAGTGCAACCTGG GCACTGTTGCTT CTGACATCAGCTAG AGTGAGGGGGTGC GTCCGAGCGTCTG TTCACGAGGAGGT GCGGCTCAAAACAG CGTCTGTTCTGTA TCCAGCGGCTGAG TATGAAATGTTGTA TCACGTCAACTAAC CTAAGGTGGCAGC CCCATCAACGTGAT GTAGGTAGGC CA
--	--	--	--	---

				<p>TGCGTACTAGCTG GTAGTGACGCTG GGCACTGTTGCTCT TCTGACATCAGCTA GAGTGAGGAGGTG CGTCTGAGCGTCT GTCCACCAAGGAG GTGCGGCTCAAAC AGCGTCTGTTCTGG CATCCAGCGGCTG AGTATGAAATGCT GTATCACGTCAGCT ACACCTAAGGTGG CAGCCCCACCATCG TGATGTAGGTATCG CGACCTGGTAA</p>			<p>DHYLVSKIRALSTVANTRQPLRFNIQRLS TEGVADERYRKLDERIRGYNVGENLNNLWES IHGAMSTAAREVIGTAQRPRNGWFDEECQ NVTDEKNVARSRMLVSGTRQNRERYKEARA AEKRIHRRKKRQYEEDIIEAQNSIEQNDMRR FYETVNGVRRKTAPSPVMCNDREGNLLTDKI LVAARWKEHFEQLLNGENTSASENRLIISNDG QAVEPTLDEVRRRAIKELKNCKAAGKDELPAE LFKHGSEQLYELHRTLRLRIWEEELPASWLEG LICPLYKKGHRLECANRYGITLLNSAYKIMSRV LFNRLRPLEESFVGEYQAGFREGRSTTDQMF TLRQLDKFREYNLQTHHLFIDFKAAYDSVKR NELWQIMVEHGFPAKLIRLIRATLDGSKSVR VADETSTFVTLGDLKQGDALSLLFNIALEG AIRRAGVQRSGTIVTRSHMLLGFADDIDIIGID RRAVEAFVPFQRETARIGLTVNTSKTYMIA GRQGSNRDVGSEVLGGEKFEVEEFVYLG TLVTCNDVTVREKRRRLAANRAFYGLRNQL KRSLSQTKTKFALYKTLILPVALYGHESWTLKE VDRKAFGVFERKVLRTILGGKQENGIWRRRM NHELYQVYKEVDIIKLKHGLRWRWAGHVVRM PEERQAKIIFSRPGRRLRGRPRTRWLFVAV EEDLRALNVQGDWKRQAQRAQWRRIIHS</p>	cd01 650	570	830
				<p>ACCGTAGCGCAT GACGGCTCACCAT GTAGCATGTCCGA AAGAACTTGAACT ATTGAGAACCCAGA GCTACAGGTCCAAA GCCCTGATAAAGG TGGATGGTTGTGAT GGCTATGACCGTG GTCATCATGTAAAG AACAAAGGGACGA TGCTGTATCGTGC AGCATCGAGGAGG CAGCCCCTCTAG</p>			<p>TGATGCTACGCTGTAGCCA TCAAGGTATCAAGGTAGG TA</p> <p>MHDVGSATLVRWPIDSSPTKKGSRANGLI LRQQTRQRIKNDNDWKVGSWNVRTLINEPAR VGLLARELRNVGVNVAIQEIRWPRTGEREF RAVDPIANTSFKYHIYSGGDRAGRGVGFVIG KQMKRIIRWKPVSDDICVLRMKGKFFNYSLISI YAPNDKPDMMKDEFYESLDKAYGECPKQD VKIVIGDANAQIGKEDFFRPVIGTESLSHVTN DNGLRLVTFAAARGMAISSTYFARKNIRKHT WRHPSGSDACSQIDHVLVDGRHFSDDVIDVRTF RGNIDSDHYLVVAKIRARLSSVTSSATHRTLRL FNIQRLSTDGVAQYRQQLDEKLGRVNVAG DVEDSLWDPHEAVTTTAREVIGTGQRRRND WFDEECQRTDMKNVARSRMLVAGTRQNR ERYRAARAEKRIHRRKKRQHEESVADAQE SMNRNDMRRFYATVNGARHNTVPVPAMC</p>			

	RTE_Ele2C_AAe	GCF_002 204515.2	Aedes aegypti		GGGCTGCAAAATG GTGAGTCACAAT CAGGAAGGAGCGT CCAACACAGCTCTG GTCCTCACAAAGTTC CTACCTCGCGCTTC CACGGTGCAAATG ATGACAAAAGACCG CCAGCTAAGGGTT GCGTACTTAGCTGG TAGTGACGCCTGG GCACTGTTGTCCTT CTGACATCAGCTAG AGTGAGGAGGTGC GTTCTGAGCGTCTG TACACGAGGAGGT CGGGCTCAAACAG CGTCTGTTCTGGTA TCCAGCGGCTGAG TATGAAACGCTGTA	TAGCCCATCAAGTATCAAG TAAGTA	NDREGNLLTDKTAVAARWKEHFQQLNGEN GNVARNRMNIDDDDQAVDPPTIGEVKKAIS ELKNCKAAGDEIPVELLKHGSEQLYQSIRHVL LKVWEDEELPTGWLDGLIRPIYKKGHRLECA NYRGITLLNSAYKILSRILFNRLRPLEESFVEGY QAQFREGRSTTDQMIFSLRMILDKFREYNLQT HHLFIDFKAAYSVSVRNELWQIMSENGFFPAK LURLIRATLDGSSVRIADEVSTSFVTL DGLKQ GDALSNLLFNIALEGAIRRSGVQRNGTIITRSH MLLGFAADDIDLIGIDRRSVEEAFVPLKRETARI GLTINSTKTMYMAGRDRVRHGGSVAEVFD GDVFEEVFEVLGTLVTCNDNVSVREVRRRV AAANRAFYGLRNQLRSRSLQTETKFALYKTLIL PVALYGHEAWTLKESDRKALGVFERKVLRITIL GGKLENGVWRRRMINHLYQVYKDANIINRV KYGRLOWAGHLVRMSEERIKIIFSREPGRGR RLRGRPRIRWLIVAVEEDLATLNVRGNWR SFA QDRRRWSSSTIRPAMA	cd01 650	561	821
--	---------------	---------------------	------------------	--	--	-------------------------------	---	-------------	-----	-----

				TCACGTCAGCTACA CCTAAGATGGCAG TCCATCAACGTGA TGTAAGTAGCGG ACCCCGGTAA				DQMFTLRQLDKFREYNLQTHHLFVDFKAAY DSVKRNLWQIMSEHGFPAPKLIRLATLDGS KSSVRVVEISSFVTL DGLKQGDALSLLFNI ALEGAIRRAGVQRSGTIIIRSHMLLGFADDIDI IGIDRRAVEEAFVPFKRETARIGLTINTTKKY MIAGGQRSGRVSSEMVLGGDKFEVVEEF VYGLTVTCNDVTVREVKRRIAAASRAFYGLR NQLKSRSLQTKLALYKTLILPVLYGHESWT LKEVDRRAFVGVERKVLRTILGGKLENGIWR RMINHELQVYKEVDIVKRIKHGRLRWAGHV ARMPEERQAKIIFNREPGRRLRGRPRTRW LFAVEEDLRALNVQGDWKRLAQDRAQWRR LIHSAQIHRSSEL			
JAM1C_A Ae	GCF_002 204515.2	Aedes aegypti		GGGTGTTAAATGA CTGGACAATGCGT ACCATTTGGTACTTC GCGTACCTGCAGG TATAAAATAGACCC CATTTGTGTCCTT AGCCTCTTGTCCAG TAACCTCTATCCCT ACCTCCCGTGGTG CCGCTGGGATAC GAGTAACCGTAGG GAAGATCGGGTAA CCAACCTGGTGG AACCTTGGTCGTAT GCTGACAGGGAAG GGGGGCTCTCTCT TCTGAGGGGTAG CTTATCAGAGCGTC TGTTCTCCATGTTA GGGGCGGCTCAAA ACAGCGTCTGTTCT CCATGTTAGGAGC GGCTGATCATCGTC CTAGTGCCAGCGT GGGACTCTAAACA	TGAAGGACGTAGAGCCAG CAAAAGTAAAGTAAGTA	cd01 650	550	810			
								MKSSVQEAAYNVNSDRNRHRRHRRKRTND WKLGSWNCKSLNFLGSTRILLSLLVRKFDIV ALQEVCKWGSTVHTYRDTGYTIYQSCGNRHEL GTAFIVMGMQRRVIGWWPIDNRMCKLRKIK GRFFNISINVHSPHSLASDDDDKDAFYAQLER EYDGCPSHDVKIVIGDLNAQVGGQEEFRPIIG KFSAHQLTNENGLRLIDFAASKNMAIRSTYFQ HSLPYRYTWRSPPQTESQIDHVLIDGRHFSOI IDVRTYRGANIDSDHYLVTVKVRQRLSVVNNI RYRPPRYNLERLKQPEVATEYAQSLEAALPE EGELTEAPLEDWCSSLKAINNAAEGAIGFVE ANRRNGWFDEECQTVLDEKNAARAMMLQ QGTRQNVERYKQKRRQQTHLFRDKKRRLEEL ECEEMEQLYRSQETRFYKKNLNASRKGFPVR AEMCRDKDGGILTDEREVIERWKQHYDEHL NGAEEDQDSRRNGFISTADEGDVPTPTIGE VKDAIKQLKNNKAAGKDGGIGAELIKMGPDRL ATCLHRLIARIWDTQEQLPEEWKEGIYPIYKKG DKLECENYRAITILNAAAYKVLSQLIFRRLSPAS RFVGSYQAGFVDGRSTTDQIFMLRQILQKCR EYQVPTHHLFIDFKAAYDTIDREELWKIMDEN GFPGLTRLIKATMDSVQCCVKISGALSDPFE TRKGLRQGDGLSCLLFNIALEGVMKRAFN MRGTIFNKSSQFCFADDVDIVGRTFQWVAE QYTRLKREADRVGLKVNNTSKTYLLAGGTGERD			

				GTGCTGTGCACGAT GATCTCCGGCGA GACAGGGGGTTGG TGCAGGCCTTACAA GCCAGCCGTAA	TAGATGTTAACTAAATAAA TGAAT	RARIGRRVTIDGFEFEVWDEFVYGLSLITSDNN CSREIRRIIAGSRAYYGLHKTLSRGLHFRKTC TMYKTLIRPVLYGHETWTMLEEDLQALGVF ERRVLRITFGVCENGVWRRRRMNHELALY GEPSTKVAKAGRVWAGHVVRMPDNPAA KMFVNSNPAGTRRRGAQARAFDQVEQDL GSVGRSNWRLAAMDVSRNIVAQVMS	cd01 650	553	795
	JAM1B_A Ae	GCF_002 204515.2	Aedes aegypti	GGTAAATGGCTGG GCATGGCGTACCAT TGGTACCTCGCGTA CCTGAAGGAATAA AATAGACCCCTTGG TGCGGTCTTAGCC TCCTGCCAGCAAC TCCTATCCCTACCTC CTCGCGTACTGGC CGGGTACGAGTA ACCTTAGGGAAGA TCGGGTAACCAACC CCCGGTGGGAAC TTGGTCGTATGCTG ACAGGGAAGGGGG GGTTTGCTTTTGCT TTTGCTTCTGCAA CCTTGAGCGTGTG ACTCCATGTTAGGA GCGGCTCACACA GCGTCTGTCCCA TGTCAGGGGCGGC TGATCATGTCGGA GTGCCAGAGAAGG ACTCTAAGCTAAAC TGCGCACTATGTC CTCCGAACATTAG GGGGAATGGTCCT CCGGAATCTAGG GGGTGGTGTGACG		MKNQATNNQRENTNRNRRRQRKGTSD WKLGTWNCKSLNFIGSTRILADVLKDRFGIV ALQEVCKWGSMTFRGNHTIYQSCGNTH LGTAFIVMGDMQRRVIGWWPINERMCKLRI KGRFFSINNVHSPHSGSTDDDKDAFYAQL REYDSCPSHDVKIIGDLNAQVQGEFFRPTIG KFAHRLTNENGLRLIDFAASKNMAIRSTYFQ HSLPYRTWRSPPQQTESQIDHVLIDGRHFSDI IDVRTYRGANIDSDHYLVVMVKLRPKLSVNN VRYRPPRYDLRLKQPDVAAAYAAQLEAAL PEEGELDEAPLEDCWRTVKAINDAAESNVG YVGRSRNDWFDEECQEVLEKNAARAVML QQGTQRNVRYKRRQQQTRLFREKKRLEE TECEEMEQLCRSQETRRFYQKLNASRNGFVP RAEMCRDKDGSILTDEREVIERWKQHFEHL NGAESTGNEGRDNGGNAFVSTAEQDNQPA PTLREVKDAIHQLKNNKAAAGKDGIAGELIKM GPERLAICLHRLIGTIWETEQLPEEWKEGVICP IYKKGDKLDCENFRAITILNAAKYVLSQIIFRRL SPVNEFVGSYQAGFVDGRSTTDQIFTVRQIL QKCREYQVPTPLGIMDENSPFGKLTIRLAT MEGVQNCVKVSGHSSFGSHRGLRQGDGL SCLLFNIALEGVMRRAGLNSRGTIFTRSSQFV CFADDMDIVGRTFEKVADLYTRLKREAAKVG LVVNAAKTYMLAGGAERDRARLGSSVTIDG DTFEVDVYGLSLTADNNVSREIRRIISG SRAYYGLQKLRKSHIARTKCTMYKTLIRPVV LYGHETWTMLEEDLQALGVFERRVLRITFGG VQENGWVRRRRMNHELALYGEPSIQKVAK AGRIRWAGHVARMPSDNPAMVFASDPVG			

Lian-Aa1	GCF_002 204515.2	Aedes aegypti	<p> GCCCTGCAAGCCA GCCGTAA </p> <p> TTAGTGGCTGCTTG GGCAGCTCAGGAG TTAATCATCAATAT TAACCTCTTTTCCC GAGCAGCTAGAT AGCCGCTAGTGT CGGTAGCGGTGTT TCAACTGGCTAAGA ATTAACACTACGGA CTGCTGTTCCGGT GGTAAAGTCCACC TCACAGGTGACCCC TAATTTATGTTGTG ATGCGTACCGTGCC TAAGAAATGAATGG TTAGGGGGGTCTA AATAAAACCTAAC GCAAACGGAGCCT GTGGGTGACGAG GCGCCTCCACAGT ATTGAGTCTTCTT GTGCTACCCGGAG CAATGGTGACGTT GACCTTGTGTTTCT CCGAGATAATCGG CTGCCCTTCTCAG TCTCTATCTTGAGG CTTAATAAGGGTG GGATTATGAATATG TTGACATTAAATTT AAATTATCACCTAT ATGGATTGCGATTA TGCCTTTTACACAG TGTATTCTGTGCTC TTTCGCTTTGGCG ATTTTAAATAGATA </p>	<p> TTRRGAQRARWADQVRIDLASVGQNRGWR DAAATNRVLWREIVDSVLSV </p> <p> MQNGAAQVALVQEPYFRRGNFYLGNLVDP VFATFSKLEMANSRMPRACVLNKAIVATLI SELTRDVCATVIDVSVGDLNRKVVYCSVLP HDEPSPDTDFKRVVWVCVTGGLPLIVGSDAN AHIIWSSDINLRGSSLMMEVLSSTDGLNLIG NRPFTMVSNREEVDITLCSNRISHELTNWH VSDEESLDHRYIFFEHSNVTQAQLFRNPRST NWELYELVATKFHGYSPSIENPSDLDADVT TTSYIMEAFEEACPLRSVKTTTRGTPWWNSDL TLRKQRRSWNRRRSAGSEFSKARKAYKK ALRSAERSGWKNLCTNVSSISEVSRNLKILAK SKDFQVNEIRLPNGDFTSSDEEVLECLFNTHF PGCVDIASDTDEPNVFCSYESLASARSIVTESI QWALNSFAPEKSPGADGIYPVLLQKGFEEKH VLKILLVSSFATGYIPKSWRDIIVKFIKPGGRA SYEEAKSFRPISLTKLERIIDHHIRDVYLA NMPLHVNQHAYQSGKSTVTLHKVVYDIEKA FAQQSQCLGVLDIEGAFDNVSFDALEARN HGLPTMITNWIHQMLKNRHLFSTLRQAARK LSVCGCPGGVLSPLLWNLVADTLLRQLNCC GFPTYGADDYALIVGMCISTLFDLMQSALQ VVESWCRCYGLSVNPNKTSVLTERRNRDGI RPLRIFGTINVTQDVKYVGVLDKSLSWTPIH DFRVKKACMAFGQCRRFTGKTWGLKPKYIK WIYTVVRPILAYGCLVWWQKGEVRTIQSKL GHLQRMCLMAMSGAFSTPTAALEALFDVA PLHIYLKQEALSCSYRLWVLDLEKNPNRRST HTSLFPLLVNWDKIVLAPSLDTIACNFPYRTT TQFPSREEWTSGYLERSISNNIVCYTDGSLLEG RAGAGVYSRELRLNQFYSLGRNCTVFQAEIFA LMCGVQSALQQRVMGKVIYFCSDSQAIAKAL ASANSRSKLVIACRTQIEELNSVNSVNLVWVP GHSSIAGNELADELARDGASHDFIGPEPAIPIS KWVKLQINSWAATQHKQYWNLESQRQT KLYITEPSPKVAKYLTNLSKQNCSLVRLTGH CRNLVHMANIORADSEVCSDDNDVGTSVHI </p>	cd01 650	468	725
----------	---------------------	------------------	--	--	-------------	-----	-----

				CCGATCTGTTTTT TCGTTGCTGGTCTT TTTCGTGCTGAGAT TGCAAAAAGCTTAA TCCTGCCTAGTTTG GGTAGTGGCTACG GTTAGGTTAGCTCA GATCAATCTTCAGC ATAAAGAACAGC AACGATCAATCTTT GCAGACTC	TGACTGTAAGGTAAGTA	ICNCPVFAQMRFQLLGKLLSETETEVRSINLQD ILLFLTRCGNEL			
RTE_Ele4	GCF_002 204515.2	Aedes aegypti	GGGCTGCAAAACG GTGAGTCGACAAC TAGGAAGGAGCGT TCAACACAGCTCTG GTCCTCACAAAGTTC CTACCTCACGCTTC CACGGGTCAAACG ATGACAAAGACCG CCAGCTAAGGGTT GCGTACTTAGCTGG TAGTCAGCCTGG GCACTGTTGCTCT CTGACATCAGCTAG ATTGAGGAGGTAC GTCTCGAGCTGCTG TTCACCAGGAGGT GCGGCTCAAACAG CGTCTGTTCTGGTA TCCAGGGGCTGAG CAAGAAATGCTGA ATCGCGCACAGCTA AATCCAAGGTGGT AG		MPHQDRPSVSWDVKQSWHDGAPARQEC WRRPNKPPVKNIANNIGDNTTRNNRQRPT RRNKDYDWKLGWTWCKSLGFGAGCDRIYDEL HPRNFDIVALQELCWTGQVWKSIGHRAATF YQSCGTTNELGTGFIVLGKMRQRVIGWQPIN ARMCKMRIKGRFFNYSIINVHCPHEGRPDDE KETFYAQLQIYDGCSPRDVKIVVGDMNAQV GREEMYRPVIGRNSLHAVSNDNGQRCVNFA ASRGMVVRSTFFPRKDIHKTTWRSPDNLTEN QIDHVLIDGFFSDITNVRTYRSANIDSDHYLV AVCMRSKSTVTNSRRSRTPLNLERLRDSEV AQEYAQQLEVALPTEEQLEGAATLEDGWRHIR SAIGSTATALGPAAPNQRNDWYDGECEQL KNEKNAAWARMLOHRTRANEARYKQARNR QNSIFRRKKRQHEERDREAMEELFRAKDTRK FYEKLNRSRKGFPQADMCRDCDCDGNLLTNE REVVERWRQHYDEHLNGDVASSGGGTEITL GARADDERLPPDLQEVETEIRRLKNKKAAG VDQLPSLLKYGGDALARALHWVISKIWEEE VLPPEWMEGVVCPYKKGDKLDCANYRAITIL SAAAYKVLSQLCRRRLSPIAREFVGGYQAGFMG ERATTDQIFAIRQLQKCREYNVPTHHLFDFK SAYDTIDREQLWQIMHEYGFDPDKLIRLIKATMI ERVCMCVRVSGTLLSPFESRRGLRQGDGLSC LLFNIALEGVIRRAGIDTSGTIFRKSVQLLGFAD DIDIVARNFETMADTYIRLRAEARRIGLININVS KTKYMIARGSREDTARPPRVHIDGDIEIWE	cd01 650	588	848	

CR1-10_AMi	GCF_000281125.3	Alligator mississippiensis	CT	TGACCCCGAGTTGCTCTTCCTGCCTAGTGCAGGGGCTGGACCCGATGATCTTCCAAGGTCCCTTCCAGCCTTACAATCTATGAATCTATGAAT	EFVILGSLVTADNDTSREIQRRIVAGNRAYFG LRRTLRSKVRHRTKLTYYKTLIRPVLLYGHET WTLRAEDQALGVFERKVLRTIYGGVQMED GTWRRRMNHELHQLLGEPTIVHLAKIGRLR WAGHVTRMSDTPNVKMLVLENNPTGTRRRG AQRARWVDQVEDDLRLRLRMGRWRRRAAM DRVEWRLLRTAEATQALA	cd01650	539	790
------------	-----------------	----------------------------	----	---	--	---------	-----	-----

CR1-2_Asi	GCF_000455745.1	Alligator sinensis	CCTGACCGTATTAT GAGGACTTCGAG GCTCTGGGACCA GGCTGAAGGAGAC GGGGCGCAGGTG GTTTTCCTCTCTC CTTCGGTGAGGG GGCGGGACAGAG GCATGGACCTGC ATTGTGAGGTCA ACGAGCTTACGC AGCTGTGCCGCC GGCAGGCCCTGG GTTCTTGACAACG ACCATGCTTTCGG GAGGGGGATTGC TGGGGCGGGATGG GCTTCATCTCTCC CTAATGGCAAGCG TATCTCTCTCCAG GTTGGCTGATCTCG TACGGCGAGCTTTA AACTAGACTCGCCG GGGAGGGGGGCT GCAGGAAGACGAG GAGACGCCATCAG GGCAAGGGGTGCA AAGGTAAGTAATG GGGGTCTTGCCATC AGGGCAAGGGGTG CAAAGGTAAGTAA TGGGGGTCTTCCC ATCAGGGCAAGGG GTGCAAGGTAAG TAATGGGGGTCTT GCCATCAGGGGCAA	TGAGTCCCGTTTTTGT AGGGTGAGTGGTTTGT GTGTTTTGGTGTCTC CCGCCCTGGCGGGGTC GGGCTGAAGATCTATTA GGTCCCTCCGACCCGAA ACTATGAAACTATGAAA	RFRNLIRKNFFTVRSRLWNKLPEVWQSPTL EIFKRRLDSLHAGVT MGNKQELALRIASSNPDVGLTETWWDDSS HDWAVNIRGRLVRRDREGKGGGVALYVK EHTSSASSMSGEGRAEVLWVKIQGGQGE RDLTVGVYVRPNQEEELDREFSGQLAETLKA RDVVMGDFNYPDICWEEQSAGTDRSRFL AELQDLHLIEQVHSPTRGDALLDLVLAAGDDL VRGLQVLHLGNSDHRVLEFTIQRVRSRARS KVVALDFRRANFSEMRLVGEELGSLRAEELS VQEEWSFLKETILGAQGVSIPTRSKAGRSVQK PPWLTGKVRECLAKRVAYARWKGGVISKEE YTSSARACRRAVRKAKAEIELGLASRIKDNKKS FFKYIGRMKKAPGNVGPLQDMRGNLVVASE DKADLFNQFTSVFLCRDRDFPAVIEAGREGI AFRPQVGEDRVRELLEGLDAFKSAGPDALHP RVLRELAGVIAGPLARLYERSWCLGQVPPDDW KVANVPIFKKGKREDPGNYPVSLTSILGKLF EKIIKEHICDGPASEMMLRGNHQHGFIHRSC QTNLIIFYDQVTKAMDAGVAVDWFDFSK AFDTSHPILINKLGDGIDAYTVRWIADWLK GRTQRVVVDGSYSTWGEVSGVPQGSVLGP ALNFHSDLDVGVESNLFKADDTKIWGEVS TLAGRERLQKDLRLQGWANRMRMCFNTD KCRVLHLGSRNQQTHTYTMGSSLLESTEAEERD LGVIDSKMNMGRQCEVMVGRANRTLSCH RCISSRAREVILPYATLVRPQLEYCVQFCAPH FRRDNDNMERVQRRATRMIRGQQQGRPYSE RLRDLNLSHLKRLRGDLVTYVYKLTGRNQKG LGSETLPLAPAGITRNNHKLMMERYRLDIRK NVFTVRAARSWNQLPREVVLAPTLGIFKKRLD TYLAGLI	cd01650	470	720
-----------	-----------------	--------------------	---	--	---	---------	-----	-----

201

[illegible]

[illegible]

[illegible]

	RTE-5_AC	GCF_000 002075.1	Aplysia californica	TGGCTGGAGCTAA TGCTCCTGGTAGG GCCACCCATGCCAG ACAGGTCGACGGG TAGAGGTCAGACA TAGCGTGATCCACC GGCCTCCAGGTTT CGGAGGTTACGCT CAGGCTAGCGAC CCAGACTGGTAAA AAGCATATGTTACG GAAACAGCAACAA AGAGAAACCATATT TCTGTGTGTGTTGG ACCTTCAGAGTCGC TTGTGACTCGT	TAGTAAGTAAATAAGTAA	MYGDLTVVQRIKLARLRWAGHVVRMETDD PARKVFLGRPQGORRRGRPKLRWQDQGVAS AIKAGITDWQTKARDRERFRTLROAKTAKRL MNAGESRQEAVERCKIEDLNTAKTKVIIGLW NVKTLYGTGSAQVTAEMRRRYGLHVLGISES QWGTSGWLRTATGETLIYFGREDNQHREW ATILNKGLEKLEMEWKPINSRMIMTLRLKGGH TNSIIQCYAPTNDSEDESKDYFEEELQRLTG IPRDLIVMGDLNVKVGEDNTNYERTMGQ HGMGTRNENGERLVDFCALNNLVIGGTLFP HRTIHLTNWSPNGRDKNQIDHLMIGGMILR RSLQDVKVRGADVGSYHHLVTSRIKFLKKT FTREATRTQFNIQKLSDGNVTSAFRLKLNRF QALQDLEETPTNTDEVNTQWKLIVEAYTESS KTC LGVKKKNASNEWIQQGTWNIEERREIRK RELAAKSIRIQEYAEYKEVCQKVRLVRED KREAMENLATEEAAAKGEOGNVYKIRIIC GKYHGMGPIKGIGRLLTIEKEQEARWPE HFREVLNQPPPTDPIQEATLNDISSDPPQ RHEVVKAVKSLKEGKSPGMDNLNAELFKVDP ELTTTILQPLFKTKIRKEQITEDWRTGVIVKVP KKRALGDLNNWTGITLLSVASKILAKIINRMS DAVDICLKEQARLKERRCTDQIFALRNITE QCTERQRQLVNVVDFEKFVDFGVHRESLWCI LRSYGIPGKIVEIKSFYQSFKHVVGNRTTAFE VKTGVRRQGCNMSAIMFNLAIDWVMRKTE DTSRGIRWTLTFEFDLDLDVDDLAQLSHTNQ HMKETNPLHKYERHIGLRISQKKTMMTLN VSSPTLIMENNAHLPTDKFTHLGSINKTDGG TKEDIMNRLGKERSVHRSVINIWKSEQFSIQT TLKQYNSCVLSTLLYGFECEWRKTGHDRAKQS SFHTTSLRQILRIFWPDNISNTDLLQKWRQED RTTIKRRRRLIGHVLRREPDSVVKTAFHWT PEGHRKRGRPKVTWRRKVSESEMKSWNHSHW GTLTKLGQNRQSWRSFVAARNRWRNG	cd01 650	557	811
	BovB	GCF_002 263795.1	Bos taurus	GCGCAGCGCGGCTG CGACGGCGGCACT AAGCGCGCCGAG	TGAACTGAACTGA	MKLSHARGATQDGRVMVERSDRMWSTGE GNGKPLQYSCLENPMNMRQNDRILKEELP RSVGAQYATGDQWRNNSRKNEGMEPKQK	cd01 650	610	861

AGGAGTACCCAC	QYPAVDVTGDRSKVRCCKEQYICGTWNVRS			
GTCCGAGGTCAGG	MNQGLEVVQEMARVNV/DILGISELKWGTG			
GGCAGAACCCGGG	MGEFNSDDHYIYCGQESLRNGVAIMVNK			
AGGACCCCATGCC	RVRNAVLCNLDNRMSISRFQKGKPNITVI			
GAAGGGCGGCGGC	QWYAPTSNAEEAEVERFYEDLQDLLELTPKKD			
CAAGAGGAGTTAC	VLFIIDWNAKVGSOETPGVTGKFGLGIRNE			
CCACGTCGAGGT	AGQRIEFCQENALVITNTLFFQHKRRLYTWT			
CAGGGCAGCGGC	SPDQHRNQIDYILCSQRWSSIQSAKTRPG			
CGAGAGTGCCAGA	ADCGSDHELLIAKFLRLKKVKGKTRPFRYDLN			
CTGCAGCGGCGCA	QIPYDVTVEVRNRFKGLDIDRVPDELWNEV			
GGAACGGCCGAGA	RDIVQETGIKTIPIMEKKCKKAKWLSGEALQIA			
GGAGCTACCCAC	VKREVSKGEEKERYKHLNAEFQRIARRDKKA			
GTCCGAGGTCAGG	FLSDQCKEIEENNRMGKTRDLFKKIRDTKGTF			
GGCGGCGGCCGAG	HAKMGSIKDRNGMDLTEAEDIKRWQYEYTE			
AGGAGATACCCAG	ELYKKDLHDPDNHDGVTIDLEPDILECEVKWA			
CGTCCGAGGTCAG	LESITTNKASGGDGIPVELFQILKDDAVKVLHS			
GGCGGCGCACGAG	ICQIWKTTQQWPQDWKRSVFIPIPKGNNAK			
AGGAGTTACCCCG	ECSNYRTIALISHASKVM/LKILQARLQQVNR			
CGTCCGAGGTCAG	ELPDVQAGFRKGRGTRDQJANIRWIMEKARE			
GGCGGCGGCCGCG	FQKNIFYCFIDYAKAFDCVDHKNLWKILKEM			
GAGGAGCTACCC	GIPDHLTCLRLNYAGQEATVRTGHGTTDWF			
ACGCCCTAAGCCC	QIGKGVRRQGCILSPCLFNLYAEYIMRNAGLEE			
GAGGCCAGGGGCG	AQAGIKIAGRNNINLRVADDDTTLMAESEEELK			
GCGGCCGGGAGGA	SLLMKVKESEKVGKLNIQKTKIMASGPITS			
GCAACCCACGCC	WEIDGETVETVSDFFLGSKITADGDCSHEIKR			
GAGGCCAGGGGCG	RLLLGKVMNTNLDISIFKSRDITLPTKVRLVKA			
GCGGCCGGGAGGA	MVFPVVMYGCESWTVKKAERRRIDAFELWC			
CCAAACCCACGTCC	WRRLLRPWTARRSNQSLKEISPGISLEGM			
AAGGAGCCGTGGC	MLKLQYFGHLMRRVDSLEKTLMLGGIGGR			
TGCGCGGCGCAG	RRRGRQMRWLGDITDSMDVSLSELRLVM			
GAGGGCCTAGAGG	DREAWRAAIHGVAKSRTLSD			
AGCTATCCACGTT				
GAAGGTCAGGAAG				
GGCGCGGTGAGG				
AGATACCCCTCGTC				
CAAGGTAAGGAGC				
AGTGCGTGCCTTT				
GCTGGAGCAGCCG				

[illegible]

[illegible]

211

212

AAATTAAAGAAACT
GTTGACCAAAATTAA
AAAGATTCTGGATA
CTCATAGTACAAGA
CTAGAGGAAGCTG
AACAAACGAATCAGT
GACCTGGAAGATG
ACAGAAATGGAAGA
TGAAAGAACAAAA
GAAAGAAATTGGGG
AAAAAAATTGAAA
AAATCGAAATGGA
CCTCAGGGATATG
ATAGATAATATATAA
ACGTCCAAATATAA
CCTCATTTGGTGT
GACCAAGGGGAAG
AGAAGGGTAAAGG
TCTAGGAACAGTA
TTCAAAGAAATTGT
TGGGAAAACTTC
CCAAATCTCTAA
CACCATAAATACAC
AAATCATAAATGCC
CGCGAACTCCAAA
TAGATAAATCCAA
ATAAAACCACTCCG
AGACATATTCTGAT
CACACTGTCAAATA
CGGAAGAGAAGGA
GCAAGTTCGAAA
GCAGCAAGAGAAA
AGCAAAATCAACCA
TACAAGGAAACA
GCATAAGACTAAGT
AGTGACTACTCAGC
AGCCACCAATGGAG
GCGAGAAGGCAGT

				GGCATGACATAATTT AAAAATCTGAGTAG AGAAAAATTGCCA ACCAAGAATCTTT ATCCAGCAAAGCTC TCCTTCAAAATTGA GGGAGAGCTTAAA TTTTTCACAGACAA ACAAATGCTGAGA GAATTTGCTAACAA GAGACCTGCCCTAC TTGAGATACTAAAG GGAGCCCTACAGA CAGAGAAACAAAG AAAGGAGAGAGAG ATACATGGAGAAA GGTTCAGTACTAAA GAGATTCAGTATG GGTACATTAAAGG ACATTAAATAGAGA GAGGAAAAAATA TATATCTGACAAAC ATAAACCAAAAGGA TAAG				
R2NS- 1_CSi	GCA_003 604175.1	Clonorchis sinensis	TAGTCGGCGAGCT GA	TGAATGTACATTCTTGCCA TTGAATCTCACACAGACC CAGTATGACGGACACTTGG TGCCTGATGTGCGAGTCTT GACTGTGCTAGCGCTTACC GCGCCTTGAAGAGCATTC AGCATTGTTTGTCTTTCTT CGGTGTTAGACTTTTACAC GCATGTTTCTTACCAAAA TTCTTACGTACGTTGGGAT TCATCTATCTGACTGGAA CTGTTGGTTCATGACTTC GAATGAGACATTTCTTCT TTTATCTCATATTTCTCTGT	cd01 650	431	689	

				AAACCTTTCGCATTTCATCG CTTGCACTACTTTTATGT CTGTGACCATGCTCTTCAA AAATAAACGATA	ARITLVKVPNPPTSPDQLRPIISVSSILVRCFHKV LADRWRRRLQLPSLQFAFLHRDGCLEATSLH ALLRHSSATASNLISLAFVDISKAFDSVSHDTIV RSAEAFGAPSLVRYIAQSYENAVAVFSPSEV HCHRGVRQGDPLSPLLIFIMAMDEVILGSMIP QLGYQFHDTLVDFGFAFADDWVVCASQARL KEKLEAAAVELGRAGMINARKTKAMVICGD RKHRTAVSVEPFCEAEELITPLGPTDITVYL IPFTFKGKGVFNHRQHLLKLLDEVTRAPLKPH QRMEITRNYLIPRLTYSVLGQVHRNLTKRLD NYIRQSRGWLRPKDTPISYIHAGKQHGGGLG IPLSATIPMQRRVRMEKLLSTQCRVLRNVVN DSAFKVVVDLSLPIRVHSGSCVNTKEELVAA WGDLSHNSVDGRGLRELVASPLSNRWLVFP ERVFPRIFIGIQLRCNLLRTRVRSARHGGG QTILCRGNCQPESLVHILQSCWITHDARCA RHNVRARELAKRLRLGYTVFEELRAPTSTFI KPDLIAVRERRATVIDVSIVSDGRGVTWNEK KQKYGADEFSLAISALRAIGCDVDFLVHQP IISYRGICFPQSAKAVIGLGLSKVTVSDLCLLAIV GSLRTYDTFMRGTWR	cd01 650	385	648
	L2-2_CCa	GCA_004 011595.1	Cyprinus carpio	CAAAAGGTAAAGG GGGAGGTGTGCT TCAATTTATAACAA TGTTTTCAGGATTT CTCAGAGGGCAGG CTTCAAGTAACT CGTTTGAAGTATTG GTGCTTAATAAAC ATTATCCAGAGAAA CAAATGTTAATGAT AAATCCCCTGTG	TAACTACTAATAACACTTC TAATATCCAAATCCGTTAA AGGATTTTAGGCTGCATT AATTAGTAAACCGGAAC CGGGAACACTCCCATAAC ACCCGATGTACTTGCTACA TCGTTAGAAGAATGGCAT CTACGCTCATATTAGTCTG TTTCTCTTATTCCGAGGT CACCGTAGCCACCATCC AGTCTGTATCCAAGTCAGA GGGTCACTGCAGTCACCC GGATCCAGTATGTATCCAG CCCAGATGGTGATCAGC ACCTAGAAAGGACCTCTAC AGCCCTGAAAGACACGCG AGACCAGGACAACATAGAG	MFVLATVYRPPGHHHTDFIKEAFADFTSELVLA DKVLIVGDFNIHVNDNEKDALGSADFIDILNSIGV RQHVSNGTHCRNHTLIDLILSHGIDVDGVEIM QPSDDISDHYLVLCKLHIAKTVNSTSCYKGR ITSTTKDCFVSNLPDVQSQFSISKTSQQLDDVT ETMDSLFSILLNTVAPLRLRVKENSILPWCN EHTRTLKRAARKMERSWRKTLEVFRIAWRE SNLSYRKALKTAKSDYFSSILLEENKHNPRLFN TVAKLTKNKASTSDISQHHSSNDFMNVFTS KVDTRDKIVTMQPSATVSHQTVHYRSPEED FHSFSTIGEELLYKLVKSSKPTTCMLDPIPSKLL KEVLPEVIDPLLTIINSSLLGYVPKTKFLAVIKP LIKKPQLDPKELVNYRPIINLPFLSKILEKVSS QLYSFLEKNGICEDFQSGFRPHYHSTETALLRVT NDLLSSDRGCISLLVLLDLSAAFDTIDHTILLH RLEHFVGINGSALAWFKSYLYDRHQFVAVNE EVSYSQVQYGVPGQSVLGPLLTFLYMLPLG		

				CCCCAGACAGATCCCT GTAAGACCTTGCTCAGA CGACACGGGACAAGAC CACAGAAACAGATGATT CTTCTGACAATCTGACTT TGCTGCAGCCTGGAATTG AACTGCTGTTTCGCTGG TCAGAGAGAACTGGCCC CCCGACTGAGCCTGGTTTC TCCCAAGGTTTTTTCTCCA TTCTGTACCGATGGAGTT TTGGTTCCTTGCCGCTGTC GCCTCTGGCTTGCTTAGTT GGGGACACTTCATTTACAG CGATATCGTTGACTTGATT GCAAATTATTGCACAGATA CTATTTAACTGAACTGAG CTGCATGATGACATCACTG AATCAATGATGAAGTGGC TTTAACTGTCAATTTGCATT ATTGACACACTGTTTCTCT AATTAATGTTGTTCAAGTTG CTTTGACGGCAATCTTTTT GTTTAAAGCGCTATATAAA TAAAGATGACTTGACTTGA CTTGAC	DIIRKHGVSFHCYVADDTQLYISSRPGETYQFEK LTECIVDIKNWMTSNFLLINSEKTEVLIIQPKSS ACNDLERCLSDGCSVNSSSVRNGLVLFDSN LSLESHVSSICKTAFHILKNISKLRPMLMSNA EMLIHAFMTSRDLYCNCALLGGCSARLVNKLQ LVQNAARVLTTRKYDHISPVLSLTHWLPIK HRIDFKILLITYKALNGLAPQYLNELLHYSPPR PLRSQNSGNLIIPRISKSTAGGRSFSYLAPKLW NNLPNIVREADTLLQFKSRLKTHLFNLAYT			
	L2-2_DRe	GCA_008 692375.1	Danio rerio	GCAGGAGAAGCAC TTTAGCAGCATCTA GAACAGCAGCCTG TAAGTACATTTAAG ATTTGTTTCAGTTG TTGTGTATGGTTTG AGGACTTGTTTCCA GCTGTTTGTTGAAA GTTGTAGGACATTT AAACTTGCTTTTCAG TTGTGTATAATACT AGAAGTTGTTTAGC	MCFIPVVNTNTRKTREVRRCRRNPHNLSIHVS TISQLSLVGLWNCQSAVNKADFITSIATYSDY NLMALTETWLRPEDTATHATLSANFSFHTP RQTGRGGGTGLLISKEWKFTLIPSLPTISSEF HAVTHHPFYINVVYIRPPGKLGHLFDELVDLL SSFSNFATPLLVGLDFNIYVDKPOAADFQTL ASFDLKRAPTSATHKSGNQDLIYTRHCFDQ TIVTPLQSDHFLSLNIHITPEPHTPTLVTFRR NLRSLPNRLSTIVSDSLPPSRKLTALDSNSAT NTLCSLASCLDRCLPLASRPAPASPPAPWLS DALREHRSKLRAAEIRWRTKNPAHLLTYQTL LSSFSAEVTSAKQTYRYRLKINNATNPRLLFTE	cd01 650	508	768

CACTGTTTCCTTGG TACTATAAGAGCT TGTAGCGAAGG CAGACGGGTTGG CGTCGTCGCCTCA GTTTCGGCCCTGT TTGACTCGGGAG GCGTGTCCAAAG CCAGGTAACCACTA TATAGTGAGCAG GTAGCATTAGTCG GCAGGAGAAGCAC TTAGCAGCATCTA GAACAGCAGCCTG TAAGTACATTTAAG ATTTGTTTCAGTTG TTGTGTATGGTTTG AGGACTTGTTCCA GCTGTTTGTGAAA GTTGTAGGACATTT AAACTTGCTTTCAG TTGTGTATAATACT AGAGTTGTTTAGC CACTGTTTCCTTGG TACTATAAGAGCT TGTGTAGCGAAGC CAGACGGGTTTCG CGTCGTCGCCTCA GTTTCGGCCCTGT TTTGACTCGGGAG GCGTGTCCAAAG CCAGGTAACCACTA TATAGTGAGCAG GTAGCATTAGTCG GCAGGAGAAGCAC TTAGCAGCATCTA GAACAGCAGCCTG TAAGTACATTTAAG	AAATGACTAAATGTAAAT GTAAATGT	SSLLYPPPPPPASSTLTDDFATFFCTKTAKISAQ FAAPTNTQDTPHTLTLSFQSESEVSKLV LSSHATTCPLDPIPSHLLQAISPAVPTLTHIINT SLDSGLEPTTFKQARVTPLLKPNLDHTLLENY RPVSLLPFMAKILEKVWFNQVLDLFTQNNLM DNKQSGFKKGHSTETALLSWEDLRLAKADSK SSVLLDLSAADFDTVNHQILLSTLES LGVAGTV IQWFRSYLSDRFRVSWRGEVSNLQHLNTGV PQGSVLGPLLSYITSSLGPIVQRHGFYSYHCYA DDTQLYLSFHPDDPSVPARISACLLDISHWM KDHHLQLNLAKTEMLVVSANPTLHHNFSIQ MDGATTASKMVKSLGVTIDDDQLNFSDHISR TARSCRFALYNIRKIRPELSEHAAQLLVQALVL SKLDYCNSLLAGLPANSIKPLQLLQNAARV FNEPKRAHVTPLLVRLHWLPLVAARIKFKTLMF AYKVTSGLAPSYLHSLQLIYVPSRNLRSVNERR LVPSQRGKKSRLTLNLPSWWNLPNCIR TAESLAIFKKRLKTQLFSLHFTS				
---	--------------------------------	--	--	--	--	--

[illegible]

				AAACTTGTCGAACA ATAATGAC				LEERVFIGDINGHVGREVDVDFKSIHGGHGF GTTNEGKTILDCLAYDFLIANTLFFKRNEHL ITYKNGYSTTQIDYFLIRKLNRLDCKDCKVPISE NLCTQHRLLVDFKIKGHKIKNVKQFSRIRW GHLKNINKNIFESKILDLSVHEDANIMWEK VASDIKNIADVLGVSKGAPNSKNSWWWN PNVQEKIKKKICYKNWFNFKNSDNWEIYKV ARKDSKKAICDAKTLAFSDLYNRLDTKEGERY MYKLAKQRDRSSKDLVQVKICKEDQKILFLE DDIQQRWMRYFAKLFNEDCENKIDYSDLSTL NENKDFCYRRISKTEVREALKMKGGKAVG PDDPIEVWKCLGDEGISWLTCLFNTILTKKK MPDEWRMSVLPIFKNKGDAQDCANYRGIK LMCHTLKLWERVMEHRLRRDTNVSQNQFG FMPGRSTMEAIHLLRGLMEKYRENKEDLHMI FIDLEKAYDKVPREVLWRVLEKKGNNAYIQII KDMYAGAITCVQTQGGTLKYFPISVGLHQGS ALSPYLFALVMDVLTRLHQEDVPWCMILFAD DILLVDKTRREGVEGKLELWRSTLESKGFRLRS KTEYMECNFSSNRPSGIVTLGDQVINKSTRF RYLGSIVQSDGEIDGDIISRIQVGLKWRNAS GLLCDRNVPCLKGFKYKMMVVRPAMILYGAEC WPLKEKHNTKLSVAEMRMLRWMSGFTLRD RIRNEHIREKVGAPVEDKIESRLRWFGHIK RRPSDDPVRREVLDLTVYKKGGRPKKTWL ENIRNDLSLLDNLNLTNRTQWRKRHVAD PT	cd01 650	153	443
				TTTACCAGAAATAT TGGGGGACTAAAG AATAAAATAGATG AGCTTTGGTTGT TTAGATGATATAGA AACTGATAATGTAA TAGATATACTATGC CTGTCTGAGCATCA CATTTGCTACTGTA TGGAAAATGTGAA TGTCAGTGGATATA	Locusta migratoria	GCA_000 516895.1	Daphne- 2_LMi	MCIMSEISTSNKKITWNIITRETGQPRTE DSITIKVNEKITNKKSEVENTNNYFLNVVENI GSKCSLEEARQLMEQGVPRQFDTMEIPTTSQ SEIRKLINSLSKSSHGIDGISSRVLKACSQEICR ILSHICNSSLEQGIFFPDLRYAIVKPLHKKGDTL DVNNYRPSILLTALSILEKVMYSRVASHINKN KVLTKCQFGFKGFSTENAIYIFTNEILNALDN RKSPVGIFCDLSKAFDCVNHGILLDKLYCGMI SGKVHVKWFSYLTGRVQKVDISNSHNIQKQP SPQTGELSSMGCHRVRYWDLCCSYTKTAESS NWGTIKHGVPPQGSVGLPLFLIYINDLPYIHK			

220

				AAACAGTCCTCCAG AGTTGTGCGCTCAA TTAATGATTTAACA ACTAGAAATTTCAA AGAAAGTCTACAG CAGTTAGACTGGG ATGAGGTTTATAAG GAACCTGATGCTAA TTTAAATATAACT TATTTTCATGATACA TTTGTAGGGTATT TGAAAACTGTTCC CCAAGAAATTAGTT AAATATAATTGTAA GAAACCATGCAAA AAACCATGGCTTAC TAAAGGAATAAAA ATATCTGTAAACG AAAAAGGGAACATA TATCTATCAGCAAG AAAGAGTAATAAC CCAGAAACAATCAA ATATTATAAAACT ACTGTGCTACATTA AGAAAAAGTTATTAA AAAATCCAGAAGC	TGAAGACCACAACA	MMASSWVKYSGGKIVPHSDLRAGTTQEDVV IKKNTGILRIGAWNVLTLNRVGKLENLKREM DRLLDMVGISEVRWQDEQDFWSGEYRVIN TKSVGGNAGVGLIMNKIGMRVSYEQHSE RIIVKIDTKRPTPTTVQVYMPMTSSADEDIEE MYEEIKIQQVKGDENLIVMGDWNSVVGKG REGNVWGEVGLKRNERNRGRDLVEFCAEHNIL MITNTWFKHHERRLYTWKRPQDGTGRYQIDFI MVRNFRNQVLNCKTYPGADIDSDHNLVMI NCRLKLRQKGRKVRWNLDKLEPEVVES FRGSIRQGLAETGESNTVEEWVALDRVVK AAFHDHIGRRTRPNRNPWJTDIILJIDERRKY	cd01 650	548	807
				AAACAGTCCTCCAG AGTTGTGCGCTCAA TTAATGATTTAACA ACTAGAAATTTCAA AGAAAGTCTACAG CAGTTAGACTGGG ATGAGGTTTATAAG GAACCTGATGCTAA TTTAAATATAACT TATTTTCATGATACA TTTGTAGGGTATT TGAAAACTGTTCC CCAAGAAATTAGTT AAATATAATTGTAA GAAACCATGCAAA AAACCATGGCTTAC TAAAGGAATAAAA ATATCTGTAAACG AAAAAGGGAACATA TATCTATCAGCAAG AAAGAGTAATAAC CCAGAAACAATCAA ATATTATAAAACT ACTGTGCTACATTA AGAAAAAGTTATTAA AAAATCCAGAAGC	Locusta migratoria	GCA_000 516895.1	RTE- 2_LMi		

				AGGGCAACAGTC TGGATGATTGACTG AGCTGGCCTGTAA CATCAGCTAACATG GCTTTGCCATGCAG TACTGCGAACGGC TGAAAGCAAGGGG AACTACAGTCGTT ACTTTCCGAGGA CATGCAGCTCTACT GTATGAATATATG			KNSASETGKREYRLKNEIDRCKRLAKQEWLE DKCRTVEECITRGKIDTAYRKIKEIFGDKKSSC MNIKNSDGKPVLSKEGKAERWKEYIEGLYKG NKLEGSILEMEVEEDEIGDTILREEFDRAK DLSRNKAPGVDDIPSELLISLGEPMKKLFHLV CKMYETGEIPSDFKKNVPIPPKKASADKCENY RTISLISHGCKILTRIYYRMMEKLVESDLGEDQF GFRNVGTREAILTLRILEDRLRKGPFTFLAF VDLEKAFDNVDWNTLFQILKVAGVKYRERRV IFNLYKNQTAVIRVEGHERDAEVGKGVQRQGC SLSPMLFNLYIEQAVKESKEKFGGKIKVQGEI KTLRFADDIVILSETAKDLEEQINGMDSVLKG RYKMINKSKTKVMECSVRKSGDAEGIRLGN ETIKVDEFCYLGSKITSDGRSKEDIKCRMAM GRKAFLLKRNLLTSNINLVNRKSLKAFVWSV ALYGETWTIRSSDKKMEAFEMWCVRRMM KIRWVDHVTNEEVLNRIGEKRSFWHILTKRR DRLVGHILRHQGVNLALEGSVWGKNRRGRP RDEYSKQIQKDVGCRTYSQMKRLAQDRVAW RAASNQSQAD	cd01 650	148	420
	RTE- 1B_LMi	GCA_000 516895.1	Locusta migratoria	CAAAATAGTGAAG CAAGACCTACAAAA AATCCTTGATATC ACAGAAATTTAATT AGAAAATATAAAA ATGCAGTATGAAG GCAAAATGAGTAC AGACGACTAAAAA TGAGATTGACAGG AAGTCAAAATGGCT AA	TGAAGACCACACAACAA CA	MAGVLEAKCKSVEECIKGSRVRLQENKRTFG TRKQLVEYQELDGKPVLSKEGKAERWKEYIEQ LYKGNELRDDVVGREEEVDQDEIGDTILRDEF DRALKDLGRNKAPGVDDIPSELLKSLGEPAM TRLFELVCRMVETGEIPSDFKKNVPIPKKSGA DRCENYRTISLISHGCKILTRIYYRMMEKLVEAD LGEDQFGFRNVGTREAILTLRILEDRLRKKG PTFLAFVDLEKAFDNVDWNTLFQILKSAGVKY RERRVINYLSQTAVIRLEGHEREASVKGKV RQGCSPMLFNLYMSRRPFSVDVQFVHEQA VKESKEFGIGIKVQGEIETLRFADDIVILSET AKDLEEQINGMDSVLKRAYKMNINKSKTKIM ECGLRKSQDAEGIRLGNETMKVVDEFCYLG KITCDGRSTEDIKCRMAMGRKAFLLKRNLLTS NINLSVRKSLKAFVWSVALYGETWTISSAD RKRIEAFEMWCVRRMLKIRWVDRTNEEVLN RIGEKRNFWHTLTRRRDLRVLGHILRHGIVNL				

					ALGSSVWGKNRRGRPRDAYSKQIQVDVGCR SYEEMKRLAQDRVGWRAASNQSPD				
Daphne- 1_LMi	GCA_000 516895.1	Locusta migratoria	TAACATAAACCTCT AGGACAGCAATTG ATAATATTTTATA GACAAATCAAAGG AACAAATCATATC ATAAACCAATAAT AAATGGCCTATCAG ATCATGAT	TGATGCATGCACCTCATCT ATATTAGTATTATTAT ATCAGCAAAAAAACCTA AAAAATTATGTTCTGAT TATTATAAGTGTATGAT CTGATTATTATTAGT ATTATCTATTATTCAGCT GTAAATTTTGTCTTGAC ATGTTCCACATCCACGAGG ATCTCTCAACATGGATCT ACGGAACGAATAAACTAA TCTAATCTAATCTA	MQLLVLGIDSKQIMKTAKSTYRRVWNPPIEN FRKLLKDMNWKDVYSAHGMNEKYNVFNN VSI MFENCFPLKVTRIKQSKIIPWITQGIKISC KTRKMVYLSAKSSCNDDDLTKYKEYCKILKKVI QASKQMHYEKKIATSGNQIKTIWIDVKEETG RTRKEQEQVALRVDETIVTDGHSVANLFNKY FISVTDMSGLSGSVNNALEYLKLAFNTFTYM NMSLTSPQEITSIIKSLKTHSSGDEISTKLIKA CSECLSTILSYLNCQSIITGTFPDWLKYADIKPL FKKGDKIENYRPIPLLPAFSKILEKVMYKQLL NHLTTNDILSRTQFGFLKGSIDIEKAIYTVSENV LNSLDNKLQAAGIFCDLSKAFDCVNHNILLSKL EFYGVTGNAAKWFKSYLTNRKQRVSQVTS LSQSSSDWEETCGVPQGSILGPLFLVYIND LSSVTLPEAEFVLFADDTSAIKSTSSVIERIAN DIFMINIKWFKANSLSLNFETHYMQFRTCK RFPDSICIKYEEIEEVDISKFLGLQLDNKFWS EEHTTELQRLNKSVSFIRV/LSDIGDIKITKLAY FAYFHSIMSYGIIFWGNSSSQSKVFRVQKRVI RIICGVNSRTSCRSFLKELGILTTASQYVSLMK FVINNISLFTNSSVHTYNTNRKNKNDLHKDLKA LTLVQKGVHSGTHIFNNLPSSIKNLATNKDQ FKRNLKDFLVANSFYSEFLNCTK	cd01 650	288	549	
CR1- 2B_MUn	GCF_012 275295.1	Melopsitta cus undulatus	C	TAGGTACCCTGCTCTTCG AGGGGGTTGGACTAGAT GATCTTTTGGAGTCCCTTC CAACCCTTGGGATTCTGTG ATTCTGTGA	MQIPVAPEGVRTKEEFALVDWVRDQLCN LDIYKSMGPDMHPRVLRELAEVAPRLSIIFG KSWETGEVPEDWKMANVTPVYKKGKEDP GNYPVSLTSVPGKVMIEQLILDSISRHIKDEG VIKNSQHGFMRGKSCLTNLIAFYEEVTRWRD DGRAVDVVFLDFSKAFDTVSHSLIDKLKRCGL DDQVVRWIKWNWLGRRQRVVVNGAESWR SVTSGVPQGSVLGPVLFNIFINDLDEGTECTL KFADDTKLGGVADTPEGCVAIQRLDLRLSW AGRNLMKFNKGKCRVHLGKNNPMYQYRL GVDLLERSEGERDLGVLVEGRLRELGLFSLEKR RLRGDLISVYKVKGGCDDGARLFSVSSDR	cd01 650	77	327	

[illegible]

[illegible]

				GCATTTCATGCT AAAAATAGTAGGG AGGAAATGTTTAG GCCTACTATAGGG AAGGAAAGCCTGC ACGAAGCCAGCAA TGATAACGGTATTA GGGCATAAAATTC GCGGCGGCAAAAG ATCTATAATCAAA ACTACGTGTTTTAA GCACAAGGACATA CACAAAGGCAACGT GGACATCGCCGGA CGGGGCCACACAG AACCAAATTGATCA TTTCCTCATTGAAA AAAGACGTAAGGG CTTACAGAGGGGC AGATAGCGACTCG GACCATTCTCTAGT AGTAGCCAAATTAA GAGCTAGACTAGT AGCGAATCAAAAT AGTAAGCGAGCAA ACAAGGTAGAAAG CTTCGATATTGAGA AGCTACGAGATAG AACAGAGCGAATT AGGTACCAGATAG AAATTAATAACAGG TTTCAGGCCACTTGA AGAAGCAAAACACA TCGCCGGAGGGGA			WEDNNKADLVEIGRVGVDRRGASWWGLTQ DRAAWKACVDEAMNFRVPNAT			687	
				TAGGGGAAGTTGTGAGAT ATTTCTACAGGCACCTAGA GTGAGACTAGCCAGTTAG GAGTTAGTCTTAGGATGCT TTGGTTTTTGGGTC	Nicotiana sylvestris	GCF_000 393655.1	RTE-2_NS	cd01 650	MTIKLVGECTLNVSAYAPQVGLDEIEIKRRF WEGLDDIVRSIPPERLFIGGDFNGHIGSSAG GYTEVHGGFGGERNGGGTSLDDFAKAFDLV IANSSSEPKREEHIVTYOSSVAKTQIDVLIIRRC	438	

GTGTAATTTTCTGT GTTTTAGGAAAT TCTCGAAGTTGTTG GAGACAAAGTTGGG CGACGAGCACTA GCAAAGGAGAGCA ACCTGGACGAGCG TCAGCAAGGGCG GTGGGAAGCGGTG CGTGAGCGTACAA CTACATCGAGTACA GCAATCAACACA GGTTGGTTCTCGG GAGTTGGTACCTCC CGTTCTACTGTTTC CCTTTGGTGTAG CTAAATGATGTTA CTTAGTTCACAAT AGTTCAATCATCA ACTAGTACTTGT AGTCACTGTTTCC TTCTAGTTGATTC GTTGTCGTTGTGC ACTAGCGAGTCTTC TTAGATTGTGCT AGGTGTAGTGGCT GCAGTCTGGGATG ATAGATTAGGGC ATGTCCTCGGGTG GGGCAGAGTGTGG GGCCGGGTCAGG GTGTGGGACGGGA GGCAGGGGAGGTA GAGGGGGCAAGG GAGCCTATAGGTT GAGAAATCGGGTCA TGGAACATAGGTTTC GCTAACGAGTAAG	ACTGTCAGCTACTGATGC AGGGCTTTATCTGTTGGAT ATTAGTATACCTTCCATCTT TTTCGTAATTCCTATATTC TTATATTGCTGTTATTTTGT TTTTATGTTTATGTTAT GTTATGTTTTATGTTAT TTATTAGAGTCTATTGAT AGTACTAATATATCGTCTC TTGTTGCTTCTTGAGCCG AGGGTCTCTGGAAACAG CCTCTGCCCCCTCGGGT AGGGGTAAGGTCTGCGTA CATATTACCTCCCCAGAC CCCACCTTGTGGGATTATAC TGGGTTGTTGTTGTTGT	DRRLCEDCKVIPGETLATQHLLVMDIGIMIR RKRSVRGRPRIRWGALTKDQAQLEGRLSA MGAWRSSGDANTMWSTTADCIRKAAREVL GISSGRTGGHKGDWWNNAVVQGVKEAKK AAYLRLVGSTGEEKTRANSERYKVARKEAKM AVTEAKTAFAFARLYEELNGKGKKLFLAKV RERTARDLDQVRCIKDDDDDKVLMGDDQIKR RWQTYFHKLLNEEGDQDIVLGELRNADSPHE LSYCRDIEVDEVMEAMRKMRRGRATGPDEI PVELWRCVGRAGLEWLTGLFNVIKTRNMP EERWSTMVPLYKNKGDQSCNNYRGIKLLS HTMKVWERVVEMIRVRTVSISDNQFGFMP GRSTTEAHLIRRMVEQYRDKKDLHMVFIDL EKAYDRVPREVLWSCLEVKGVPAVIRVIKD MYDGAKTRVRTVGGDSEHFPVVTGLHQGSA FSPFLFALVMDAITHHIQGEVPWCMLFADDI VLIDETRGGVNERLEVWRHALESKGFRLSRTK TEYLECKFGAEPTAGVEVRLDSQVIPKRGSF KYLGSVIQGTGEIDEDVTHRIGVGWWMKWL ATGVLCDKKVPLKKGKFYRAVVRPAMLYGT ECWPVKNSHIQMKVAEMRMILRWMCNGHT RMDKIRNEDIREKVGAVAPMEDKMKREARLR WFGHIQRRSTDAPVRRGERLAVVGTGRGRG RPKKYWGVEVIRQDMARLITEDMALDRELW RSSIKVVVG				
--	--	--	--	--	--	--

[illegible]

						WILAQSKYQKDVYLCFIDYAKAFDCVDHNK LWITLRRMGIPEHLIVLMRNLYIDQEAIVQTE QGDTAWFKIKKVCQGCILSPYLFNLIAEQII REAGLYEEHHGIRIGRLINLRYADDTLLAE SEDDLKHLMIKIEDYSLQYGLHLNIKKTKILT GPISSITINGEKIEVKDFILLGSTINTNGSSSQ DIRCIALGKSAAKDLFQVLSKDVSLTKVCL TQAMVFSIASYACESWTMINKEDRRRVDAFE LWCWRMLKVPWTAKRTNKSVEEVQPECS LEARMARHLITYFGHVVRDRDHSLEKDVMLG KVEGQRKRGRPSTRWIDTVAATMGSSJARIV RLAQDQAVFHSVHRVAMSQRRLDGT	cd01 650	572	818		
						RGATGGGAAAAAGCCGAAA GAGAAAAAGAGAA	ATCTAGCCACAGG GGTTGCAAGCCAG TTACTTCTGTGCCG GTCCAAGCCCGG ATAAATAGAGAGG GTTGCGTCAGGAA GGGCATCCGGCGT AAAAAATTGCCAAA ATAACC	Oryzias latipes	GCA_004 348135.1	RTE-2_OL	
			</								

	RTE-1_OL	GCA_004 348135.1	Oryzias latipes	<p>TGTGAACCGCCACC TTAACGTGGTGGG GGAGTTTGAGAGC TCGGGTGATCCCA GGAGCTAAGCTGT CTGGGGCTTAAGC CCCTGGTAGGGTCT CCCATGGCAGACA GGTCTGGGTGAC GAGCCAGACAAAG AGCGGTTCAGAAA CCCCCTG</p>	<p>TAAGCGGAGGAAGATGGA TGGATGGATGGA</p>	<p>QGRSVRLQGAEVKKVQEFKYLGVSTVQCDDGEC GKEVKRRVQAGWGWGRKVSGLCDRRVPA RLKGKYYKTVVRPALLYGLETVAVRQREAE MEVAEMKMLRFLSLGVTRLDRIRNEVIRGTAH VACVSDKVREARLRWFGHVQRDDSGYIGRR MLEMELPGRRARGRPKRRYMDVLTTEDMKL ANVRVEDVHDRVVRWKRMIRCGDP</p>	cd01 650	613	868
--	----------	---------------------	--------------------	---	--	--	-------------	-----	-----

[illegible]

					TATTTTAAATCATTTTTCA ATTTTTTCTTTCTTTAA AGTAAATTTACGTTGATT ATTTCTATGATGATTGTG ATTTTAAATGCAATTTTCTGT TTTGTGAAGCACCTTGAAT TACTTTGTGACGAATTGT GCTATACAAATAAACTTGC CTTGCCTTGCCTTGCCTT			NVNDRNAKELNAVLETDLIUHVSCPTHSRG HTLDLLITKGVNIYNVSDVALSDHFCVFFDL SVAPKPPGPAVVQRRQINESTRAQFVEMIN PENASGANVDEMLNSVTSSILNVLDAIAPMK VKLRHRQKAPWRNDDLVRQAQKQCRRAE RKWRKSLQVHYGMYGELYAYNQILCRTR RYFSEIIGSCSSSRVLTFTTVNRLNPTPTQLE LVCTPKCNEFAVFFNDKVQGIKKAINTSTHITI ERPPTHSLKTHFVPVTDQTVQETITRLSSSTCC LDVLPTRFLKSVLNSVLPSTQIVNMSLQSGIF PEALKTAVIKPILLKKSGLDPTVLNNYRPSINLPF LGKVLKVYVYQQLTDFLLNNSFDVFSQGFRRP HHSTETALIKVTNDIRLINTDDGKVSVLILDL AAFDTVDHGILLHRLQDQWVGISGALSWLKS YLEDKRYFVEIGSCVDYMALTCGVPQGSILG PLIFNLMLPLGLIRSNNSIYHNYADDTQIY VSLTTGEYGPVDSLSHCLQQISAWMQNNFL QLNSNKTEVIIFGPQKQREIVSSYLHSLKPS NQVRNLGVIMDSDLNFNSHIKSVTSAAFYHL KNIAKINIIVSKPDLEILHAFVSSRLDYCNGLLT GLSKRAVKQLQYIQNAARVLTRTRKYVDHISP VLRSLHWLPVQQRIDFKAALLVYKSLHGRAPK YISDMLVPYEPSRTLRTSGAGLLIPRVRTKQG ESAFQYSAAKIWNLSPEVWRQASSVFMFKSR LKTFLFSRVYD			
CR1-1_PH	GCA_001 587735.2	Parthvale hawaiensis	GCGTGGCTGCGC GTTATCAGGCGGCC CCCATTTCCGGCC ACCGGACGCCTCT TGTTTGATGAGCT GGCTTTGGTTTGG GGCTCTGGGTAGC CCCTGAGTTGGAG GAAATGCCACTCTA AAGGCGGATGAGA CCTCTTTTGGTCCA ACGGCATCCCTGA GAGGAGATGAACA	TAGCAGCGTTTGGGCTTC CCGGTGTGCAAGATGCTT CATTTGGGTTGATTACCT GTCCGTGGGAGCTCAAGC GTGAAGTTTTAATAATAAT AATAA	MLYTIIFVVGILCFFVFYCIYVITVVLCLFFLF LLLCGDVESNPGPGRGRGCRLLLYCNIIRGLHA NLAELDFVRGV/DVWCSETLVAGRHRDAEL ALAGFQSPFRRLCGSGPGFGRGM/VVVRSGFC AYRQSVHECACHEIIVRVCGRLNNYLFSLY RSPATDDSLYDCLLTAMASIQSTDPAKAFV/FV GDVNAHHRDWLGSASPTDCHGVAALDFCTL SSCVQLVRGTHIAGNCLDLVMTDVPDLMT VTVGSPIGSSDHSHLVWSLDLNQVVPVWDTR RVVFKSRANWVAITRAVRTLPWRQIHSED PVSEINDLTVSILERFVPKRTILVRSRDKPWFD DQCRLEAFEAKQAAVRAWRRSRDRTLWQTVV DRBAFAKRVFEAQBRJROBSRESIISIDHPH	cd01 650	536	795			

				GGCTGGTGCCATA GCTGGTACTCAGT CTTCGCAGTCTGC TCCTCTCTCTTCC ATCAGGTAGTACCA CCTGATGCTCTTT CCTTCTCTTCTCT CTTCTCTTTTCGGC CAAAGCCTGTACTG GTGATGGCAGGGT CTGGGCGCGTAG ATAGCTGGCCGCTT TCGTAGCTATCGTT ACTGTGCAATTGTT TTTGCTTTATT				RWSELKGSVFGAEPSPPLVPGGGIITDPL ARAELSAHFDGKQSRDVIALPHGCHPEPRLT SLAFRSGAIKVLLEGLDPYGGVDPVGMFPLFY KQLADVLPKLAIVFRRLRLGNFPCRWRTGN ITPIPKGPVSPYVANYRPITLTPILSKVFEKLIAG KLGRFAEVTGLLPAGQFAYRIGLGCDDALLSV SHLQSDALDCRSEARLVQLDFSAADFVNHR GLLYKLESGVGGRLVSIIRDFVSERTQSVSD GVLSASVGVVSGVPQGSVLGPLLVLYTSDM FSSLENTLINYADDSTLMVAPAPRLRDAVAQ SLNRDLRSISAWCSAWSMKLNASKTKSMISR SRTLVPQHPQLEIDDTLLQESSLEILGVVFDE KLTFEPHIRLVSRASKIGLLRKVNSVFGDSQ VARRCFYAEFLPVLEVCPVWASAAADTHLRLL DRLVSSASRLCADNDLVNLSHRRRVAELCMIF YKVYNNERHWLYSSLPALKVFGRETRAACGA HSFTLEAVRCRTNQFCRCFVPWSAKVWNLLP ASAFGRIGLQAFKSAVNGFLDFL	cd01 650	488	770
				TGGTAGCGCTAAGA GCACGGCAAGACA TGAAAGATGCCCT GGTCAACCACTGC GCCAGGCCATCTC CAACCGTCTCGACT TTAGTCTGCCATT GGAGTAAGATGGA ATCTGGGAAGAGA GAGTGAGGCTGAC TATGCGCACCTCTC CCTCACTTTAATCC AATTACGCGCAAG TCTTGACATCTCGA CATCTCCTAAGTCC TGTGGCGACGGGC GAGCGACGAAGCA GCAGGTGTGGATA CACTGGGAGCTGT AGCCGGGACCTG	Pelodiscus sinensis	GCF_000 230535.1	RTE-1_Psi	TAGATGTCAAGGTCCTCTT CGAATACGATGGACGAAC AACACAAC METGGGYTFFWRGRSSDERESGVGFAVKN HLVRKLASLPKGVNDRMLTMLPLPRGKOAT LISAYAPMTNPDEVKDFEYEDLDALLSSVKR TDRLLILGDFNARVGSDDHTAWDVGIGKNGIG KCNSNGLLLTCAAHDLITNTIFRLPTRKKT WMHPRSKHWHLIDYVVRKDRQDVRTKA MPSADCWTDRHLIISKLSFRIKPRRPQGNKA VMKKINVSRLRNPNTATLAEELGNLLELQL EENAEKDWERFRDIVHSSALKVLGSPSHRQKQ DWFDENDEEIKAILAEKHRLHRAHQNDPSST AKKNAFNIRRVQNKLRKMQDSWLSAKAD EIQKFSDRNDAKRFEYALRSLYGPQPSGSSPLL DSDGATLITEKALILQRWAEHFQSVLNRPSI NNEAIDRMPQADINHSLDVSPSETETLQAIQ LSSGKAPGSDAIPAEVYKGGGAVLVNKLTLF QSFWNQGSIPQELKDAIVHLYKRKGNRQVC DNHRGISLSIAGKILARVLNRLNHLIEQGLLP ETQCGRKERGTVDMIFAARQLQEKCEQN RELYTTFDLTKAFDTVREGLWRIMSKFGCP DRFILMVRQFHDGMTARVLDDGEASEATFPV			

				CACACAGGCGGCT CAGGTCTAATGGTC GTCTTCTTGCTGGT GGAAGCAACAGCT GGATTGGCGTCTA CCTGAGTGACTGA GCAGCCTCTTTAG GATTGCACTGCTCA CCTCCGTAGTAAGA GGAAGGGGCTAGA AAAGGTGCCCTAA ACATTGTTTGCTTC ACAGAACCTTGGCC AGCTTACCGCGGCT GGAGGGGATCCCA TCTCATGCGGTCGA ACAATAAAATTTTCG AACAAAATGCAAG GTGACACCACTCAC TATTGGCACATGGA ATGTACGCACGCTT CAGGACAACCTTC AGCAGACCGCCCA GAAAGAAGAACAG CCCTGGTCGCCAGA GAGCTCACAAGATT TAACATCGATATTG CGGCACCTGTGTGA AACTCGTCTAGCAA ATGAAGGTCAGCT C			TNGVKQGCVLAPTLFSIMFSAMLSDAFQNSS LGIRLRYRTDGLNLRRLQAVTKIKETVL RDL LFADDCALNAGSEQEMQASMDKFAAACDN FGLLINTKTEVMHQPAKPAQHQPVTITVKG QKLQAVDQFTYLGSTLSRSVTIDIEVNCRAK ASSAFGRLLTKVWDRRGISLATKLKVVRAWL TTLMYASETWTVYRRHARQLNHFTICLRRL LRIRWQDKVPDTEVLSRAGLPSVYTLLMKAQ TRWAGHVARMPPDHRIKQLFYGELFQGKRS HGGQKKRYKDTLKASLKSLDIDTTSWETLAQ DRSTWHTLIYQGCQTSEARRTTIAQEKRALRK ARVAKAATWVPTHVCPTCGRTFRARIGLISHL RTHHEQSI TC			499
				TAATACACATTACATTTTG ATTTAAAGAAATTACTAATC TACAACCTCATTAATCA TAACCAATGACTGATTACT AATCGATTATATAATTACA TTACCCCCCTCCGCTCT AATCACTGCACCTGAATGA	Petromyzo n marinus	GCF_010 993605.1	Poseidon- 3_PM	cd00 304	416	

				GCTTCGATTAAGC GG				<p>RADYDNKIRNLEEESTYIKIKRDPNTLTKRTTA IIANSLIPKNEQYLIKPRASKPRLYGLPKIHKIG IPLRPVSTIGSATYQLAKYITKQIQPYIGKTSFI KNSHLNVNIKDKLTPDDILVSFVDSLFTKVP VDESILKLDLTKNTGLKQDKIPKLAGHCLNNT YFLWRGDFKQKEGAAMGSPSPAIANLYM ENFEIRALASSTLKPSEWFRYVDDTLVWVPH GKDTLNDFLNHLNAQNPAIQFTMEMEVNNT LPFLDILLGHGLNGKITHSVYRKQHTHTDAYLN AESHHHPAQKSSLIKTLVHRAKTVSDGESLVR EIKHVKDALRQNGYNNNAQINRALNAKSRSQS NTTESDQIVTKIPYISGVTDKLARVLQKQNT VRFGTVHKIQKLVSCDKDALPKLSTEGVYRINC SCGSCYIGQTGRNISVRLKEHQRIKLSQTL SAVAEHACQQGHSDFDNVAVVCKTNYWQ RLRESIEIRNKNFNRSDSYRLSAAWKRIK</p>	cd00 304	260	343	
				<p>GTAAAGTGATGACGTCAAC CTGCTGCTCTATATAAAC CGGAGCAAGCGAGGGCT TTACCATCTGCGCAATTG CCCGATGATGCTGGAAGT ATTACACGCGAAACGTCG TACTCAAACTGTAATGTG GATTAACTCTCAACACAC CGGTGGGCTGAAGTAGTG AATTATTGCTCTTTGTCAA CGTGACACTTTTTTGTCTG ATTGCCGCTGTCGGGTGG TGGAGGGTTACGACACAT TTATAAATAAACGCGATCT AACCCAAAAACCTTTATTA TGAATAATATTGTCGCGA AAGCTACGTGTTAACT</p>				<p>MVDEIIAEVESAVRLPHNEAEMTRHEVAKII ENSKVPKNISSNEAKALRELNVNKDIQILAAD KGRTTVMSRADYDNKIRNLEEESTYIKIKRDP TNTLERKTTAIIANSLIPKNEQYLIKPRASKPPR LYGLPKIHKIGIPLRPVSTIGSATYQLAKYITKQI QPYIGKTSFIKNSHLNVNIKDKLTPDDILVS FDVSLFTKVPVDESILKLDLTKNTGLQKDIP KLAGHCLNNTYFLWRGDFKQKEGAAMGSP LSPAIANLYMENFEIRALASSTLKPSEWFRYVD DTLVWVPHGKDTLNDFLNHLNAQNPAIQFT MEMEVNNTLPFLDILLGHGLNGKITHSVYRK QHTHTDAYLNAESHHHPAQKSSLIKTLVHRAKT VSDGESLVREIKHVKDALRQNGYNNNAQINRA LNAKSRSQNTTESDQIVTKIPYISGVTDKLA RVLQKQNTVRFGTVHKIQKLVSCDKDALPKL STEGVYRINCSCGSCYIGQTGRNISVRLKEHQ RNIKLSQTL SAVAEHACQQGHSDFDNVAV VCKTNYWQRLRESIEIRNKNFNRSDSYR LSAAWKRIK</p>				
								<p>TAATACACATTACATTTG ATTTAAAGCTTCGTGACT GCAGTGATTCATGTTCTTC GCTCTTCGGTGGAGTCGC GTTGGGGGAAGCAATAA AATAATAGAATATATGT TTGTTGTTTGTGTTTCCC TTCGGCGTGGCTTTGCCGA GGGAGCTAAGAAATAGAA TGCACATTACAATTCCTTA TAGGGGAACACATACTA ATTAACGGATTACTAAGCT AGCATATAACTACTAACAC ACATTACAATAATTACTA ATCTACAACCTACTACTAA TCATAACCAATGACTGATT ACTAATCGATTATATAATT ACATTACCCCTCCCTCGC CTCTAATCACTGCACCTGA TATGAGTTAAGTGATGAC GTCACCTGCTGCTCTATA</p>				
						GCF_010 993605.1	Petromyzo n marinus					
						Poseidon- 3B_PM						

					TAAACCGGGAGCAAGCG AGGGCTTTACCAITCTGGC GAATGCCCGATGATGCT GGAAGTATTACCAGCGA AACGTCGTACTCAAACGT AAATGTTGTGGATTAACTC TTCAACACACCGGTGGGCT GAAGTAGTGAATTATTGT CTTTTGCAACGTGACACT TTTTTTGCTGATTGGCCGT GTCGGGTGGTGGAGGGTT ACGACACATTTATAATAA CGCGATCTAACCCAAAAC CTTTATTATGAATAATATT GGTCGCGAAAGCCTACGT GTAAA				
Rex1- 2_PM	GCF_010 993605.1	Petromyzon marinus	AGATGGCGCCGAT GATGGCAGCCTCG GCCGGTTGGTGCG CCCTGTTTGTCCA GTTTGTGTTTTG TCTTGAGAGGGGC CTGCTGTTCAACAC ACTCGTCTGGTGA GAGAGGCTGCTG TTCATCACACCCGT GCTGTGAGAGAG GCCTGCTGTGAG GGGCCTGCTGTTCA TCACACCCGTGCTG GTGAGAGAGGCCT GCTGTTCAGCACAC CCGTGTTGAGGGG CCTGCTGTTCAGCA CACCGGTGCTGGTG AGAGAGGCGCTGCT GTTCAACACACCCGG TGCTGGTGAGAGG	cd01 650	574	826	MALMMVASAGRCALFCLVLSVNSVFCQDFR SSLTRDELLNIRETTPVDLIPTFLPAEELVGILV NAALGFAQATKRRPRGKRSGALVRLRRRGIR TVLPSSFNSVRSLYNKMDELQLLMVKNKDFL SASALCFETETWCDLIDPTSLQLAGFHLVRSD RDIALSGKTKGGICFYINSGWCVDVTILQH CSPILLESIFITCKPFYSPREFASILVGVYLPCCP QVKEAQRMLADQILSVERANPDSLVLVLGDF NKGNSHELPKYQFIKCPTRGNTLDHCYTT VSRAYHAVPRAALGHSDHAMVHLIPAYRQKL KLCKPAVRTSKQWTSEAVEDLRACLDCTDW DMFTTATNSLDELTEAVTSVISFECEDCCIPTRT RVNFNNKPMWFSAKLRRLRLEKEEAFRSGDR DRFKESKNRFSKAVREAKRLYSERLKHQFSTN DAASVWVRLRQITNYKPRAPHTTNDURLAN DLNEFYCRFERQLDCPEPLPTQEAPSTNTT ETTSFQEGEVNRLFQRQNPKAAGPDSVSPA TLRHCAQDLSPVFTNIFNTSLETCHVPACKSS TIIPVKKLRPTGLNDYRPPVALTSVMKSFERL VLAHLKAITDPLLDPLQFAYRANRSVDVAN MALHFTLQHLDAPASYARILFVDFSSAFNTIIP ALLQDKLFLQNLVPPYTTCKWITDFLSDRKQCVK		

[illegible]

[illegible]

[illegible]

						GGAAATTGGTAGGTTATATT ACTTGTAGATATTACTGC ATGGTCGGAAC TAGAAGC ACAAGCATTTGCTACACT TGCAATTAACACCTGCTAAC CATGTGTATGTGACAAATA CAATTGATTGAT	CTATCAGTTAGCAT GTGGAACATTCAG GGCCTAAACTCATC AACCTTTGGACTGA AGAGTTTAGCACTG GAGTTCAACAAAA ATCTTAAAGATGTT GACGTCATCATCT GCAGGAGACATGG TGTAAGGCTGACAT TGCTACTCACTGC CCACAGGCTACAG AGAGGTCATTGTG CCATCACAGAAAA CAGCTCTGCAATA GAGGCAGAGACTC TGGAGGATTGATC ATTTGGTACAAATC CGAACTACAAAAATC TAATTGATCCCTC AAAAATTGGCAATA TCACATTTGGTTAA AACTGAAAAAAGA ACTTGACTGACAG AAAAAG	Salmo salar	GCF_000 23375.1	RTEX- 1_5Sa			GGAAATTGGTAGGTTATATT ACTTGTAGATATTACTGC ATGGTCGGAAC TAGAAGC ACAAGCATTTGCTACACT TGCAATTAACACCTGCTAAC CATGTGTATGTGACAAATA CAATTGATTGAT	TGACAGACCAATCAACCTG CACATGCTCTACTGTAT GCTTATTGTTATTGTTGAA TGATGTTATTTTGACCC TTGGTTATTGTTACTG TTGTCCCGTTGACAATATT TGATTCATCTTTTATTAT TTTTATATTGTAATATCC AAAAAAGCTTTGGCAATA TGACATTTGTTAGTCAATG CCAAATAAAGCGAATTGAA TTGAATTGAATTGAGAGA	MCSFAQYISPPQNPHITQRRSPPIRILLRGD ISPPQNPHITQRRYPPRIPIILLRGDIPPESPYY SEEIFLTEEETCHFQAQGNVLICGDTNARTG TLPDLTSTRGDSFITGHTVSNCLHLPHRNNSD STVNKNGRDLQLCRSLGLYFVNGRLRGDSL GRFTYCSPLGHSITVDYMITDIDPFLSSFTVKP LTPLDHSQITFLKRTDMEITTTTHSQPSKLYNI RNSYRWAQNSTEYQKATWQNIQTLDNF LDTTTHSKEGINLAVKNINYIFRQTAKEAQLK LIKNTKKTDDNWFDADCKIIRKRLRTLNSQK HRDPNNGELRHYCETLKYKRTLRTKKAQYN SKQLTIEESINTNFWQNWKKLKSKQEELA IQNGDIWTHFTKLTLYNTVQIDTNAEQRIHE KINGLEKAIDNQNPLDSPITDQELYYKQLQAL KFKKACGPDGILNEMKLTSAKFQLAILKFNL ILSVGYFPDIWNQGLTPIFKNGDKFDPNNYR GICVNSNLGKVCFSIINVLKSQIGFIPKHRITD HIYTLHTLIDKHVHQNNTKIYACFIDFQKAFDS IWHITGLFYKIVESGVGGKTYDIIKSMYTGNTC SIKIGKKRTEFFENQGRGLRQGCNLSPALFNIYI NELATILEKSSAPGVSLHNSEVKCLLFFADDLCL LSPTAHGLQQLDLLEQYQCTWALAVNPKNT CPREDPGVPVPLQDLCLLSPTAHGLQQLSDL LEQYQCTWALAVNPKTKIMIFQRRSRQGI RPKFSIGTKYIEYCHYNYLGLKISSTGHLNEAV NELREKARRAFYAIKKQIQIEPIKIWLKIEYVI EPIALYGVSEVWGPLAKQDFIKWDKHPHETLHA EFCIKILLHVQRKTTNNACRAELGQYPLIKTQK RAIKFWKHLKYSDDL SYHYQALQCQELSKES PLIQLVLGLSSQTCSTNTLKPQDQNIQIRINQ ITTSKONYIAYWETQAOQTQSKMQCYLALIN	cd01 650	499	800
--	--	--	--	--	--	--	---	-------------	--------------------	----------------	--	--	--	---	--	-------------	-----	-----

[illegible]

AAGCACCAATTAGC
CTGAAGCTAGCCC
GGCCACGGGCTCCT
GAGCTACACACGA
AGCCACTCTGGG
CTACAATCTCGGA
CCCCTTTACTGCC
GGTACGGGGCAGC
GAGCCCCGCCGAT
CCTCCACGACTGA
CTACCGACGTTATC
TGCCCGAGGATTCC
AACTGGCCCTCCG
TCGCGAGCTAACT
GAACGCCCACTTG
CTAACCGCGCCTA
CTAGCTATCTAGAG
CTACTTGGAAACCT
ACTAATTCCAGAC
TGGTCTATCGACGT
CACCGCAGCAAG
GGCAAAAACAGAC
TTACCCCATCGCG
ACGTCCCCAAAG
CTACTGCTTGCC
CCGGTCTTTAACT
GCTAGCTTGCTTGC
CCCGTCTGCTAAC
TGTAGCTTGCTG
CCCCGGTCTACTAA
CTGCTAGCCCCGGT
CTGCAACTGCTTGC
CTTGCTAAACCCGT
CTTGCAACTGCTAG
CTTGCCACGCCCG
TCTGCTAACTGTA
GCTTGTTAGCCCC
GGCCTACTAACTGC

[illegible]

L2-2B_SSa	GCF_000 233375.1	Salmo salar	<p>GTTCCGGAACCTCCA ATTGAGCAGAG CAGCTGAAAAATG AGCTCTACAATA TTGAAATTTACATG GTTTTAGAGTGAA GTACATCGGAAA AAGCAAAAGAAAA CTGCAAGACTGAAT AGGAGAAAAACA AGCAACAAACAAA GAAGATAGGCTTTT GAAAAATAACTATT TTTCATAAAATTGT AGTTATCTTAGCTA GCTGAATTGTTTAC CAGTTGCTAAGCA GTTGCTAGGGACTC TTTTGGAAGAACT AGCTAGCTAACGA AGAACAACAAAGA ATATCTAGTTAACA GAAGAAAAAGAAAG AGAAATCAGGAC AGAAGAAAAAGGA TATCAAAAGTGAAAC AGTTCTCTAAAGAC ACAGGAGACAATA ATACAAGAAACAA CACTTCTGTAGCTT GTCAACT</p>	<p>TAGGATAAAGTAATCCTTC TAACCCCCCCCCCTT AAAAGATTTAGATGCACTA TTGTAAAGTGGTTGTTCCA CTGGATATCATAAGGTGA ATGCACCAATTTGTAAGTC GCTCTGGATAAGAGCGTC TGCTAAATGACTTAAATGT AAATGTAATGTAA</p>	<p>MCLSPVLSSLRPHYKRTFPRGRCHSNLVVPA RTTHVEFQVSGSLWNCRSAANKAEFISAYATL QSLDFALTTETWITTDNTATPTALSSAHVFS HTPRASSQRGGTGILISPKWTFSLPLTHLSI SSEFHAIVTTSFKLINILIYRPPGSLGFINEL DALISSPEDGSPITVLGDFNLPTSTFDSFLAS FFLLSSFDLTSPSPPTTHKAGNTLDLIFTRCS STNLATPLQISDHYLVFSLLSSNTSHSAPTR MVLRRPNLRSPLATLSSLSLPSAQTFSNLS PDSASSTLLSSASDFLCLSSRPARSPPAP WLDDSLRAHRTGLRAAEKWRKRLPADLA SFHSLSTSSVSAAKATFYHSKFQASAPR KLFATFSSLLNPPPPPPSLRMTSSTILKRRL TISDPRLLSQTTPLVLLTPYVLEALCYLLPPE SSSPSPSSLSADDFVNHFEKKVDDIRSSFAS NDTAGPAHTALPCALTSFSLSPDEISRLVAG RPTTCLDPIPSLLQITISGDLPLFSLINSSLT AGYVPSVKRARVAPLKKPTLDPSDVNNYRP VLLSFLSKTLERAVLQSLCYLSQNDLLDPNQ SGFKTGHSTETALLCVTEALRTAKANLSSALIL LDLSAAFDTVNHQILLSTSELGIGAAHAWIA SYLTGRSYQVAVRESVSAPCALTTGVPQGSV LGPLLFLYTKSLGVSISHGLSYHCYADDTQLI FSFPSPDNQAANRISACLSISVWMTDHQLK LNLGKTELLFLPGKDCPFHDLAITVDNSLVSS QSAKNLGVLDNTLSFSTNIKAVTRSCRFMLY NIRVRPCLTQEAQVLIQALVSRLDYCNLL AGLPACAIKPLQIQNAAARLVFNLPKFSHT PLRLSLHWLPVEARIRYKTMVIAYGAVKGAP PYLQALIRPYTQTRALRSSTSGLLAPLPRKHSS RSAQSKLFAALAPQWNNKLPDARTAESITT FRHLKPLHFLKEVLG</p>	cd01 650	577	838
SR2	GCF_000 237925.1	Schistosoma mansoni	<p>CCCAAAATGCCTGG TACGGCCGAGAGT GGGGAGAGTCCAC TCTCCCTCTCGAAA TGCTCTCACATGGC CACGCGTATATAGC</p>	<p>AGAGTGGTTCTAACTTTCT TAAGTAAAGAGTCTTCTG GTTGTACATTTCACTCCGC CTTATCTTTTCATCCCTTCT CTTCTACTTTTCAATATTTT GTGTGGCGCATATGTATCT</p>	<p>MLPPIIMNTEAVFLKPRTPFKLASFNVRTLMIQI RQIQIGLAMSLEGLNVDCCLSETRIQDSSEVL QIRSPSVASKSLFHVRLSGDPVASSSGLAGVG VALSARAEAAIDWIPINSRLCAVRLESSIKVR RNRREKRCFLVISAYAPTDCSPDAIKDEFYHQL TVLLQKARSTDIVVLAGDLNAQVGRGLGTEESR</p>	cd01 650	549	808

CTCTGCCAGGAA GTCCTACTACTGC CTTCTCGTGGCGG GGTGTGTTTACGA AAGTGAGAGGACG AAAAGCGAATGTC CGGCGCTTTAACC GGTTGGTGACAC GGAGGGTCCACCT AGGGGAGTTGAA AACCCTGATTCCAA ACCAATGGTGACCA TGGGCTCCAGTATC CTGATGGAACGAA TGGCGAATGAACC TGTCATTGGTCACC GGTACCATTGGGA CTGCATCTCTCAC GATGCTCCACTGCC TTGTGGATCAGACC TTTAGGTCAAAGGC TCGGGGTGTGGCC CCCTAAGAAACCA CCTGCTTCGGTTTG GGCACCCGGGCAG TATCACAGCCACA CACAAATAAATGAA ATGATGAATCTAT TGACGATACTATC TCGCTCATACTAGA AGCAAGACAAATTA CATTATTATTATTAT TATTACCATTATTAT TATTATTATTATTAT TACTATTATTACTA CGTCGCTAATTAC CCCCTTTATCCCA ATTGTGTAACTTT	GGTGCCCTTTTGTACCAAT ATATATGTGTTTAAATAAA TAAATAAATAAATAAA	LGGRWGLVGRRTDNGDRLLQLCTDHNFLA STNFRHSRRCATWRPPSASQAWTQIDHIAI SYRWGCVQDCRSFWSTYLESDBHALVCANL TLFSGRRIDHHQRIDVSKLAATSVASKYRAEL ASRLATIPPKSIDHEHWLHLDAMKMASKAAC GFAKPAYKHWVSSGSLQIMEARRSTPGDR EFDHKRRLRNEIGQSLRKDREAWWUSERANE MEAAAASGNCRLFLQIRATGSKKSGVSETIY EDDGMPTNISRRLGRWAEFFEGQFNWPAA PATSTLSLCPWPVTTDPPNEAEVRKELQLLK RYKSPGDDLPALFKDGGDFLTKELTVLFAK VWEQSVPTSWNESIVVPIFKGSRCSNNY RGISLLPIASKLLASILRLFKTRERLTREEQAG FRSGRGCIDHIFTLRQMLEHRHTYRRPTIVVFL DIRAAFDSDLRTVLWDCLLKKGVPKFINILKA LYNTSGRVRAYNHLSPLFHSSSGVRQGCPIIS PFLNFAIDDILETALMDVNSGGVDMPLGER LLDEYADDIVLLCDNAQGMQNALNQLAISV RRYGMCFAPSKCKVLLQDWQDHPV/LTLDG EQIEWEKFVYLGYSIAGGGVSDEIDARIMIK ARAAYANLGHWRIRDVSLAVKGRIYNASVR AVLLYACETWPLRVEDVRRLSVFDHRCRLRIA DIQWQHVVSNAEVRHRVFGRRDDNSIGVTIL KHRLRWLGHVLRMSSQLPRRALFADSGTG WKKRRGGQCMTWCRGMKESCKGLASVGP SRLPGWGPDRDGTQWLETLSDMAQNRSQ WRSCCNLLLLST				
---	--	--	--	--	--	--

[illegible]

CTTTGGAGAGTCTCAGAA CGCAAGCTCCATGTAGA GCTGTGTCGATATGAAA AATCAAAATTGTTATATT AATAAATTTTACTATAGT ACTACTAGTGATTTTGT AAGTACGACTTATGTCGT AGACGTAGTTGAGCGGT TGTTGGCGCTGATGGGA GCTGATGGCGGATGATGG GTGCGGATGGCGGATGAT GGTGCTGATGGGTGCGG ATGGTGCGGATGGGTGC GGATCGGTGCGGATCGGT GCGGATGGGTGCGGATG GGTATGGATGTGTGTGGA TGGTATGGATAGGTACG GATAGTGTGTTCTAGA ATGAGTAGGTGTGGAA GTTGATAGGTGTGTGGAA GGTTCTAGAAGGTAGGTG TGTGGAAGGTTCTAGAAG GTAGGTGTGTGGAATTTG GTAGGTGTGTGTGTCAC AACTACGTGACATAATAAC AGTTAGGTAATGTGTGG TCCACGTGACATAATAAAG TTAGTAATAAGTGTGTCT CCGAGTAGTACTGACCA GTAAGTACGTGACATAAC AATGTTGACATAATAAGT GTGCTCCGTGTAGTGACT GACCAGTGGTCAAGTGAT GTAATCATGAGGTCACTG ACCAAAGGTCAATTACTAG GTAGCAACTGAGTGGACA AAGGTCAATCTCCAGGTA GCAACAGCGGTGCAGAG	RENVTRQFTRVKYVIAVIPNNPHWCLVFSF VFEHSYILNFQGRVLNERLVSIGESMTDYLNK FFISTAKVNFVNHQTHFSRSEDPHFFASAF IEFLERFFGNILFTKINETLKSFKLQITNEICK NLNLINFSINDEIITKYFKNLNLP SNFIFLNSAM CTAIMDDCTRYLNDFLNYDNIKAQVIFAIVAP PKFREILFIIDYNTDEYLLNPTTLLGNDNYLST CKILRNKINERESPNRVLNLGECPHDIRGSGGL VSKILFCSYMNYANDLPVGLNLRQIGNVVVS SFLPIVSEIKVKSLNENIINQNHKESENLERK NKIDQLLIKVKMNVDIVDVLQISIKTVPT HKPYLGNKERVESQINKSNLIKNFNSDMKTTV NTIINDVDVDRPKYENIITNFEHKIPLDRTW DKVFKCNRSDKKLLKYLSENSEVL FELKKADN TSPGVGDIKYRDLVMLDPEGKLLTFLNKISK KVIPDSWKSFTLLIPKPKDKDYDNVSSWRP IALLSVIYKLFASCLANRLTFWKNKNNLHIGQ KGGSRHDGCV EHNLSILSALEHSKYSKNSPLAI AWLDIKDAFGSVPHAYMWSLLRFVGVDEEF VTILERMYSDDTSTFYSCGPILTPNIAIRQGVKQ GCPISMILFALAINPVLEAVSRKCEPFMIGES AVKILAYADDIALVSKSAYDLQRTVNIJAVIAS EIGFEYRPEKCGYIELPLIETESQILINDIKIKKLA SSEFYQYLGVPVGEETDQSPYDILKVVSDTR KIADSDLLGWQKLKAYKIFLHSLVFPFRTREI KTSALAKSNANNLTVNITSQLRNCFRKMMMS LPNHSEVCYFYNSTENGANCVDLLDEYHTQ TITHFFRLFTSSCFEARKVNLDLNFVAGPRLG IINPTLQERFDWINGKEPKRHSKGKTRFQRA RIAIAFFKKEHDINIMFIVTKGQPSLITSEKRG TFILTPDLRKTTSKVLHLMALCDSYLSKWEKSC SNFIASAIKSPKINKAIYRGDISEFAWHFIHRA RTNTLSINAKPHNKGKRLCRLCHTEDETMS HIIQSKHLTLGSEHNDCDLISSHLSKNSNL IVVVDHVCSLVPESKERVLDLITDMVRKKIFL VDMKPCDITINN FALVDLLNLQKYDSLKMQI QDAKPDFKVELDTCIGSLGSGFPKTPPELLKIG VDRTHLKGLLKDCALSNISHSARRWH	
--	---	--

[illegible]

251

CAGCGAGGCTCTTAA
AATCAACTATATAAT
TTTATATCTAGACT
CTGTTT TGAGAGAA
ATTAAAGGGGAA
AATAAATAGTCCC
TTTTACGGTACCT
TAAGCCGGTCGTAA
ACAGTGTAAAAAAA
GAAGAGAAAAAAG
AAGAGGAATAGTT
CAAAAGTTAATTA
TTGTTTCTTAAA
ATAAAAATAAAATA
AAGATGGGAGGG
GAAGAACTCTTAA
TATACTGTAAAGTT
TATAATGTGTATCAA
AAGCCATTTTGGTG
GGGTACAAGCTGT
TGAAGCTTAAAGTA
TGCACGGGGTTTAT
TTATTACGTTCTGTT
ATTACGTTTTTTTTC
TTTTAAAAATTAT
ATATAGCCGTTTGA
GCCTGGGGCTAAA
GTATCATTAGTTAC
ACATTATCTGAATC
CGTTAATATGAGG
TCAAGTTATATCGT
CTGGAGTTATGATA
AATTGCTCTGTTTT
TTTTTCGGTGTGG
TTTTTTTTTTTTTT
TTTTTGTACACAGTA
AGAAGACTGTATTA
ATACTTTGTGAAA

[illegible]

				<p>ATAATAGATAAATA AGATATTTTTTAA GGTGCACTCTGCT CCCTAACTAATAGT AGAAATAAAGTAA TGTTATAAGGGTT CTCTAAACATGAC CCAAGTTAATAAT GTCTTAACCAAAA AATGTCTACATTTT TAATGTGTTTTTAT ACGCAATAAATTAT AACTAATTGAACAT CACTAAAAATTATAG ACACAAATTGGGATA AAAAGTTAGGCAC ATTTGAAAACCTTTA CAAAATAAGGAC AAATCCGACAGTG GCTTAAAAATCCAA ATTTGGACCAATTC TAGCTAAATCTTTC TCAACGGGGGCGCT GAATTACGGTCCAA ATTTAGGCCACTTA G</p>	<p>CTCAGTCATTACCAGATGC ATTAGCGCCAATCAAAAAT GTTTGGGTAAGTCCCATGT GAAGAAGGCCATGGGCTA AGGTATTATATTCTTATTA TAGCACACACTATTTTAT TGTTGTGCTACTAAGCCGG GGGTCTTCACGGAAACAG CCTCTCTACTTTAAGGGTA GGGGCAAGGCTGCGTACA TCTTACCCTCCTCAGACGC TGCTCACTGAGTATGTTTG GTTTGGTTGGTT</p>	<p>STLLEFATAHDLVIVNSCFKRKDDNLTFRSGG HATQIDYLLIRSDRFRFCYCKVPVEACASQH RLVMDLSLDGPPMEGLRVVTPRILWRNLHG SKVAEFKDMLEGGFRLEDDVDQMWIRMD LIRGVARRTLGVASGNVKEQESWWWNDD VQAKVQFKGCFLEFMSCPEGDRHIKRELFK LAKRMAKQAAEAKTKAYQDMYRRLGTKEG VNEIFKLAKARNKRSQDIGAVRIKDEGDRVL LYDGLIARWGRYSELSFNAAARGREIVFEQEN VNVSHNDVGMQSDITMAEVRVALGMMGK GKTVGLDEPIEVWQCLGEGQGVRLTALFNV IFRTSMPSEWRLSVVPIYKNKGDAQSCSN YRGIKLSHTMKLWERVIECRRIRVTISVQF GFMPGRSTIEAIYLIRCLMEKYRECRKDLHMV FIDLEKAYDSVPRVLRCLAAARGVPEVYRTI QDMYSAVGTCVTPVGDQTFPVEVGLHQG SVLSPLFIIVLDVITRDIQAPVPWCMLFADDI VLIAESSEVNVKLEQWHMSLEGYGLRLSRK TEYLWANFSEIIEHEADVAVCIAEGRVPQTNTF KYLGSIIQSGDISADVTHRISTGWLRWWAA TGVLCDKNVPLKLGKFYRAIRPSLLYGSEC WPLRKAQERRLETAEMRMLRRCGNTMTD HIPNDTFRLLGVESISKKIREGLRWYGHVR RKCNSAPVRRVEHISVRGKRKRGRPRRTWAD QLSLDMGALNLTGDMTLDKNDWRRRIRVVE TSRSGSQRL</p>	cd00 304	316	397
				<p>TACGATATATAGAC ACAGCTTTATGGAA CAACAACTACTGGT TAAAGTTTTATGCA ACGTTTCGACGTGT TTCCCCCGTCGTCC TCAGGCACATAATG AGAAAGAGTGATG GGCAAACTATATAT ACATTAATGCTTAT TATTAAGAACACAC AAGAGATAAATGG</p>	<p>TAATTATTTCATATTCAAAC AATTTCCAACCTTTAAGCAT AGTTAGTTATATGAACAAT GTAATTTCTATTAATAAAA GACTCCAATACCAGCAATA GTTGTTTATTATTAATCTCAA TTAATCTTTTTTAATTCATT TACTTTATTTCAATTATTTT ACTTAATCAGCCTGTATGT TATTTTAAGAAATTATATAC CTCTATTGTACCTGTAA AATTAATGTAATGTATCA</p>	<p>MRSTMLSNEFLT KSVINLSQKNLTDEQINLLS RGLKFSPTVKLNHDDFVNIERNGLHQLAPGG HINYLRHQVTEILAKNKIQPNNTKVRDAL DLRQDNKIALPADKGKATVIMDKKEYQETVT ELLSDAKTYRRLKNPTSIYKDRLDKKLINESS NQITNQLKFKLSSRHPSPYARATIKIHKNPI KARLUCRSRVSFYNAQHLRSLAPLGTADSF IKDSEEVTKLEQITGKMSYDVVDVLTNPV CKDVLSIIRDKISKETPLDTLTITDSIVDLIQM CMSSTYFTWGDVEIQHGLPMGSPSLPITEI YMTAFEEKMETSPSPICWFRKVDVDFVML RAEDDPSALLNAQHKRINFTMEAEEDDSIPF</p>			

TTATTGATACATT ACATTAAATTTAA CAGGTACAATAAG AGGTATATAATTCT TAAAAAACAATACA GGCTGATTAAGTA AAATAATTGAAATA AAGTAAATGAATTA AAAAAGATTAAATG AAGATTAAATAAAC AACTATTGCTGGTA TTGGAGTCTTTAT TAATAGAAATTACA TTGTTATATAACT AACTATGCTTAAAG TTGGAATTTGTTA TTAATACATTTAT GAATATGATGATAA GTTTATCATATGCA TGCATATAACGAA CACCTTCATCTCTG TTAAGGTCAGGCT GATATCTTCTGATG TGGATAGCTTCATG AAGTCTCCTGGTTT TATCTATGTGCTTA TTGATTGTCAGTAT GTTGGTGTGTCAA AGTTGACITTTATGT CCTGTTTTCTTCCC ATAGCAAGATGTTT AGCCACAGCAGAT GTTCTTCATTGCC GGGCTGTCTTGAT GGTCTTTGAGTCGA ACATCAAGGGGTC TCTTTGTTCCCTTA TATAAGAACTTCCA	ATAAACCATTTATCTCTTG TGTTCTTAAATAAAGC ATTAATGTATATAGTTT GCCCATCACTTCTCTCAT ATGTGCTGAGGACGACG GGGGAACACGTCGAAAC GTTGCATAAACTTTAACC AGTGATTGTTCCATAA AGCTGTCTATATATCGT A	LDVRIKRLGGQLKTCVVRKPTHTDQVYVNH NTPQHVKRAAITALVRRAKRICDPEYFQQLQ HIKKVFLHNNYPPAFVHQAFRSVLMKPTN QQPNPNRPSRLAIPFIGKASYQIKKILKSVGI ETVFSGNSTLKDILLRASGKLNPSNNQPKGC VYQVECSGSSYGETKRPLDVRLKDHQDKP GNEGTSVAEHLAMGRKTGHKVNFDNTNII TINKHIDKTRRLHEAIIHRRYQPDNLNRDEGVRL SHAYDKLINL				
--	--	---	--	--	--	--

CAGGAGCACTCAA
CTTGGTAGACACAT
CCCTTAGGTTTTTG
GTTGTGGAAGGA
TTCAATTACCAT
GCTCTTAGCGAAA
GGTTTAAGCGTTGA
GTTGCCAGAGAAC
ACTGTTTCAATGCC
GACTGACTTTTCA
GGATCTCTTGATT
TGGTATGATGCGCTT
TCCGATGAATGGG
ATAGCTAAGCGGG
TTGATGGTCTGTG
GGATTAGGCTGTT
GGTTGGTTGGTTT
GCCATTAAATACGGA
TCGAAAAGCTTGAT
GTACAAAGGCTGG
AGGGTAGTTGTTGT
GCAAGCAAAATACT
TTCTGATGTGTG
TAATTCTGTTGAA
AGTATCTGGTCA
CAAACTCTTTTTG
TCTCTGACCAAGC
CTGTGATGGCTGCC
CTTTGACATGTG
AGGGGTGTTGGA
TGGTAGTTCACATA
CTGATCTGATGTG
TAGTGTCTCTATAC
ACGATAGCTTCAA
CTGACCGCTCAACC
GCTTAATTCTACA
TTCAGAAAAGGTA
TAGATCATCTTCT

[illegible]

GATAACATCATCCC TGGACTACATTGG CGGCAGATGGCGT GCATCTAACACC ATGGAATACGCGT AATGGCGGAAAC CTCATCACTCGTCT CCGAGATCAGTCAA GCGTCACATAAG TAGCCCGCGTCTG ACCCACTGCCAAT CAGCGCCGCGAT CTCGAATACAATG AGGACCGATCCATT CTCGAATCGGCAA GACTCCGAAAGCTC CTCGACGCCATGG CTGCTGAGAAATGAT AACTCAGCTTTTCG GAGACGCCAACAT CGCAAATTCGCAA AAAGCCGAAGGGG AGAGTTTTCCCCCG AGACTGGCTCGACT GTCTCCAGACGGCC CGCACGCTCCTGAA TCGTCAAGCTCAGC AATAGATGATAGC GATAATTTAAGTCT TCACGCGGAGGAT CTTAATGTAAACGC TACAAATGTAATAA ACATCAATGAGAA ACAAACGCAAAATTG AAAAAACTACAGCT TCCAAGGATGAAA GTGAACTAAATGA AGGACACAATGTG	GGTAAACGTAATGGTGA CGGAAACTTTTTTAACAT TGAAATATATTGCAATAAT TTGTTTTGATATTATACCA TACTTTTGTAACTTTAAC CTAATACTATGTGTTATG CGTACAAATGTCAACTACA TGTATATATATATATATTC AATTGCGGTTTTTGTGT TGTTTGTTCACAATAGTTTT AATTATAAACTGGGAAAG TATTTAATTGTATACATTTT ATAATTACTCAATGGTGAC GGTAAACTTTTTTAACAT TGAAATATTCGCAACAATT TTATAAATATCTATAGCAT ACCTTATCTTAATTTGAAC TTCGTATACTATGAAATGT ACTATATCGCTTTTATTG CATGTCCGAATTATCTCTC ATATTATAATGTGCATTTA TTGTTGTTTTGATTGACAT TATTATTGAACCATGTATG TTGTAAAGGGCCCTGATG AAACCAGCTCATGCTGAT CGGGCTACCCCTTTTAAAG ATGTGAAATAAATAAATA AATAAATAAA	VNTEFQEYLKVNKILTERQFGFRPNHSCETAL LCMVDLWAQNVNDKGNLNGVAYIDMRKAFD AVNHSTLLMKLNKNGCTDRAIKWFRSYLGGR SQFVSIKSKSSTRITINYGVPQGSVLAPLLSF INDLPTSHSADMFLFADDDATMSVQGKLTSEI QTKLNALDEVSWTQKNKLLNTTKVMV IGSKQKQLTLNENNLNVRITTPIECVTQYKCL GVVDSNLLFSKHVEKVALQMKQLGILRLK GTENKHYLSMIYWGYYLPHAMYCCTVWTNR SLHNYETINKLHKRAAYIISGCSWSTPSHEVLL QLNWSLLEALYSRAMGIMAFKCVHHLAPALL ANKFILHDDVAQRMVTRNSHELKLESNRCNTE FFKRSFVNSAITLWNLNPIETRFSPNLNVFKSN LSH				
---	--	--	--	--	--	--

	HEROTn	GCA_000 180735.1	Tetraodon nigroviridis	ACAAATACAAAG GATATTGAACTT GTTGATGGC AGATTGGTCTGGCT AAGCCAGTGACGT CCAGGAACAGACT GGCTGACGACCAC GAATAGAGTGGTG ACAGCTTGGATAG ACAGCTGACAGCA GGGAAAGACGGCA ACCGGGCAGGAA GGGTAGCAACCC AGCTGCATCTCC GTAGGAAGAACC CAAAACCTGCTACG AAGAGCCCGAAGC AAAGATACCCCGAG GGGAGCCCGAGAG GGGGGGAGAAATGA GCTCCCAAAACGGA CGGATAAC	TGATCACTCCAGTCGGGT CGCCTGGGTGAGGGGGTC TGATGTTGAAAGACCCGA AACCCCGATGACCCGAG GTACTATCACTGACGATGT GTCCAAGACATGCATCAAT AGGTGTTATTAGAAATC	MATTQASVKPTAVATCVCGKICKNPRGLKIH QTKMGCLASVQPEQARARFSLSESREVPARAE PYGPOQPHSPEALGETQEERQGESPHSAQN LRAQVAQAPDNQHHRRVWPPASKVSEW QQLDELEGILESTAKGVDRLQTMITLTVIS FATERYGTMKRAAPEKYTKNRRAEKISQLR QELRVLKKQFKGASEDQKPGLAELRCTLRKLL LTLRAEWHRRAKRAKRAAFLANPFGFT KQLLQKRSRGHLECAKEEVDSYLHDTFSDAER ENSLGECEVLPPEPACSFNTKAPTWKEIQT VVRAARNNSAPGNGVPYLVYKRCPKLLARL WKILRVWRRGKVAHQWRWAEVWVVPKE EKSTLIEQFRTISLLNVEGKIFFSILSHRLSDFLLK NQYIDSSVQKGGIPGVPGCLEHCGVVTQLIRE AREGRSLAVLWLDLANAYGSIHKLVEML ARHHVPGPIKTLIMDYDSFHLRVTSGSVTSE WHRLEKGIITGCTSVIIFALAMNMLAKSAEPE CRGPITKSGIRPPIRAFMDLTVTTTSVPGC RWILQGLERLMTWARMRFKPGKRSRLVKA GKVTDRFYLGGTQIPSVSEKPVKSLGKMF GSKDAASIRETNDQLGHWLTLVDKSGLPK FKAWVYQHGLPRILWPLLVYEFPISTVEGLER RVSSCLRRWLGLPRSLSSNALYGNNNKLTLPF SSLAEFEMVTRAREVLQYRESKDPKVALAGIE VRTGRRWRAQEAVDQAE SRLHHKELVGAVA TGRAGLGTPTTHLSRLKGKERRDQVQLEVR ASIEEQRASQWVGLRQQGAWTRWEEAMA RKISWPELWRAEPLRIRLIQSVVDVLPSPNL FLWGKVESPCQCQGRGTLEHILSSCPKALG EGRYWRHDQVLKAIASISSAMEYSKRLPLP GRGVRFVAGEQPPQPPRAQPGLLATARD WQLRVDLKGQLKFPENIVETNLRPDIVLHSQS SKQVILLETVPWEERMEEAERKAGKVAELV EDCRRAGWRSRCLPIEVGGRGFGAGSLCKAF	cd01 650	362	627
--	--------	---------------------	---------------------------	--	--	---	-------------	-----	-----

							SLLGITGMIRRRKAICAASEAAERASRWLWIQ RDKPWTASWTQAGN				
CR1- 65_AAe	GCF_002 204515.2	Aedes aegypti	C				TGACTGTATGTTTAGCTTA AGTTATTTTATATACATTC ATTAAGACAACTTCGTGC AGATGGATTAAATCAGAA ATACAAATACAAAT		cd01 650	586	838
							NDHLSANMHRTLSHSGRTTSTMEAPTP STVESLQALISRPDPVVGRCGGVFPVLSGK YNKITDILLDPSPSSLRIYYQINIRLSKIDSF FLAISELDYDIVILSETWLDLDDCILSSQLFGSNYS VFRTDRSALNSDKGRGGGVLIAVSSRISCSVD SAPVSNQLEQLWVKIVLPNNVSGIYIPLPD KKNDIDTIQRHIESIGAVHSHLHENDMAMLF GDYNLSDVHWTSNDANDNVFVDLTRLNPA SSTLLDGFCHGLSQVNTIVNSGGRTLDLVL NDWLPNHSIHEAVEALTVDVHHPIELVVS CPLPVVYETVSDNTGDFDKKANFESLNVALRS MDWDRLDSIEDVNEAVAFFTNSVMQAVAE HVPRRPPSKPPWSNSRLRLKRRRSAAALFY RSRHSQQSKLHFNRSSESYSNAYLFAQYKS RTEQNLRSNPKQFWFVNTKRNEDEGLPTSM YLDLSADSADKCELFAMQFKRAFNDYASP ASQIDRALQDTPQDVFSDMFHVSEEVIGAI GKLKFSYNPGDGIIPALLKRCSTLLAPLTLL NHSLRQVFPACWKKSFLFPIHKKGDKRCVS NVRGITSACSKVFEIIVNDSLFESCKRYISTD QHGFPPKRSVSNLVEFTLCIRADAGKQVD AVYLDKAAFDRVDHRILLQKLKCGVAINFV DWFRSYLTERSLCVRIGESASFTNISGVPQG SNLGPLLFTLFINDVSLLLPTGNRIFYADDAKIY MIVESLDDCYRLQLHLLNLFCAWCTRNCLTSL EKQVSYGRKRSPIRFPYELSGTTLERVKCVR DLGVTLDKLTFCNCHLNDVISRANKQLGFVLK VSSGFKDPLTLKALYCALVRPILEFAAIVWCPF QNYWISRIESVQRKFVRHALKNLPWRDPENL PPYHERCRLLGIDTLENRRHVAQTMFVAKVL RNEIDSPSLADVNIYAPERTLRRRQFISLGSR NTRYGQHDPPVRFMALKFNEAYHLDFNVTTIT TLRNHFQYFRN		cd01 650	260	518
I-4_DAn	GCF_003 285975.2	Drosophila ananassae	TT				TAAAAATTAATAAATATA CTGTTAACTTAGCTAATTA TAAAAATTAATAAATAATT		cd01 650		

					AAAATACGTAGTCGAAGG CCTCGTAGCTAGCACTACA CCCCAACTGTTAGTCATA GTTTTGTAAGTGTACTAT CATATCCAATAAATAAATA AATAAATAT	RKTNALFKRAVKIRKKEATSKFTSTIQNPPTT KIWNIRFCGLNPVKQIHCITNNNVTTNDK QKIANLAQHWSQLSKDSNFSQEFTHKYNT TTLNHIPSKSAMEDKPIITKLELTIALRSLKGIPT GINRISYPMILKNSPLAIQNRILSHFNHIFENYIP QSYKTSIPIILKAQDKDTDIASYPISLNCVVG KTLDKIANRLWYLN SNKLLHPHQYGFKKG KSTSDNLLYADHIITKSLNTRKHTSLSLDFARA FDRVGIHSIVDQLEWKCSHKIINYIINFTKNRK IAVRIGSSISNFYPLENGIPQGSISVVLFLAFN KLANILTYKEIKFSAYADDFLLVLNLKNQKNT INLTHIFRKIYDWGSYGASLSQTKCQHLHICR KHRCNCTTTQNSNIKNVTSKILGLTINNRYN WSSHINSLIPSLSNKHNIICLASHKFCNCSTL LNITKSILSKIDYALHIYGYTPKNFLNKLKTTLN AALRTVLGAFRTPTTHNLLYEAN IPTIEQQRD HLTTKLTNLTTRSKDTPHLKHQPKTNKKWT STLERTIENCQLHNIPFKPTKLYTKPPWQLP QNIINTSLHNYKKTNTSSDTFQKLFKEIKETK NHTFIYTDGSKINLNTAFVTTETDVLRYGILP PSSSILTGEMIGILEAIKISTTKRGKSVICSDSLG AIEAIQNPLNLNKYPTVIRSLLTQKPKIKILWI PGHSNIKGNLADRTAREATKMPLYTPNENI TDINKMLKNILIQEKETLINKCNTWYKNINTN NTHISALSKIQQEEISRIDQIKFIRLRLGHTSLTH KHLLDPNTQRNRCYRCSSTPTNMEHILDSCP AISONKQNLKKNRPATEYLNKPNPTAENIKVIFK LKETNLVNEI	cd01 650	788	1054	
					ACATTTAAATACTAATTT TTTTATGTTTGTATGTTAGT GTGTTACAGCTGTTAAAGT TACAAACTCTACGTAGTTT GGACTCATCACTCTTTAAG TGACAGCACTATTATATT ATATT	ATGGCGGCGTAAA GATGGCAGCCTAG TTGAGCACGTGTTT TGAATAAATTTTTA ATA	Hydra vulgaris	GCF_000 004095.1	CR1- 83_HM	

262

	Daphne-17_SP	GCF_000002235.5	Strongylocypris purpuratus	<p>GAACCAATAACTG CAGTTGAAGTTAA GAGTCAAGTTCTGC TACTTGATACAAA AAGGCTTGTTGGTC ACGATAAAC</p> <p>TCCATTAAAGTTCTT CTTTACATTTTGATC ATAGTTGTTGTTGT TGTTTGCTTTTACT ACTTTGTTTAAAGT GTAGAATTCATATT CCTTTGCTTTTGG TTTTCAA</p>	<p>AAGGGTTAGGCTCGACT AGCACATGTGCTATTATCT GGCCCTTTTACCAACAA TAGTTTCTCTTCTGTATAT GTAAATTATGTATAAAATG TATGAAATGATTGCTGTT GGTAAATAAAA</p> <p>TGATATTAGCTACAGATTT TAACGGTTGTTAAATTCAT TATATTTTACATGTAATTG TTGTTTGTTTCACTTTTAT TACAGCTTAACTTTTATAT TTGATTAATTGGTCAGCCG CATACAAGGACCACCTCCT TTCTGGCTGCTACCAGCAT TAGCAATTTCTTTTGATT ATTCATGTTTGTAATTTGA TTCTGTGATTTCTGTGTTT ATCTATGTTAGTAATTGTA ATTGTTGAAACTTTTGAT ACTGCTTTTGTTTTACCCTG TATGTTTGTATTGCTTGG CAAAATAAATTGAATTGAA TTGAATTGAA</p>	cd01650	557	812
				DRHLSWKSHIREVFNQCIKKVGVLFRLRHFVP SRTLILYKAFIQPHISYGIEVWGSTYKTSLSRIL YSQKMAVRATITSEWRTPSDPLFKRLRILNVY NLHDLSTIFMFDLVNGNLPHILIEYCDIEHSY STRQKEDQLRPLKLRTTQGGQFSLSFVGSEF WNRLPVDVRLTRSNRFRNSVTSHLINKS MLCRARFVQCLNEKHVSVSTLREFLSPDPQS TNNFTQSHYTPDEFNTFVTQKSLDKDVNNKL SFINICRSLQHNFDALSVLMMSLRMQVSIIG VSETWLRDENSQYSLCNLDNVYFVGNNRRL GRGGVGIGIRKDKINFKRRTRDLDINESFLESIFI EIERFNSSILICVMYRPPSPQVDDFLMSLVNVF DVNVVEKLFYLMGDFNIDLNLNVGKVQSVN DFNLMLHSMIPLINEPTRVTRDRSSLLDNIFT NDSYIYSGILFTDISDFHPVYCICDIDIPKSSCE TDFYKRHLNENIRMYFLQLLYSTEWYSSDV NSSYNNFVFKFMKVYNLCFPVKLVRMKYKRK NKPWITNEIKLSKKCRLYKLLIKNPNVYKEV YRKIRNELNKKIEKKNYLKSFKDSTGNMKK TWDIINKAMGKKCKMTNIKEINNNDTIVTDK NDISNNEFYFDVGYNLNCKFCDFDSECIK YIGMNDKSMFFKLVTSEIEDIVKKIKNNTSA GYDDVDVKIVKKCYLLSHLCTIFNQCLEQGV FPDKMKIAKIPVFKNGVSETMSNYRPSVLT VFESKILEKCMYNRLYAFLTKCNIINDNQYGR AGHSTSCAIDFIHKVASAIDNNEVMGLFLDL SKAFDTLDHHILLVKLYQYGIRGIGLDLFSYLL HRHQIVSIANVKSNYQEIKCGVPQGSILGPLLF ILYINDLPVSRIKLSILFADDDTSFLSHQNIQSL QDIFNQEIKLINTWFSVNKLSINISKTHYMIFT NKKCDTDSIIRLCNTEIKQVSSLKFLGITDNKL KWTDHINAVCNVLSKNIGVMYKLSFPKCILK MIYNSLILPFLYYGILAWGNAHETHIRKVVLLQ KRAIRTICHTSYLAHSHPLFEKEMKILNFTDMY NYQCAIFMFQCYLQILPKSLLLKILNCNIHHY STRNAENFHLPLVHTTSYKRTVIFNGPCLWST IPRNIIBELKEYNTEKBPKKVYIMDVV				

R2NS-1_SMed	GCA_002 600895.1	Schmidtea mediterranea	<p>CCTTTTACGGTAT AATTATAACATTT ATCATAACATCT TCTGTATCAACTAT ATATATGAGTATAA ATCTGGATCTCTG AGAAGGACTCATAT ATATAAGATAATCT ATAAACTGGAAG GTAATGACGATGA TTCAGAGGTCAGT GACTATTCCTCGAG AGAGGAAGTCTAGT ACGCAGTCTGAAT CGGCAATCACTAAC TCAACAGGCTGAT GTCAGGATATGGA AAGGGTCTAATA GTGGTCTTAAACT CTCTTTGAAAGTG ACTTCTATTGGAC ACCTTTTATTGTAT AATTATAACATTA ATTATAAATCATAT CATACTTCTGTAT CAACTATATATG AGTATAAATCTGGA TCTTCTGAGAAGGA CTCATATATAGG ATAATCTATAAATC TGGCAGGATAGTA CGATGATTCAGG GTCCAGTGACTATT CCTCGAGAGAGGA ACTAGTTACGCAGT GCTGAATCGGAAA TCACTAACTCAACA GTCTGATGTCAGG</p>	<p>TGAGTAGTGGTATATTTAT AAATATAAGTAGCGAATTT ATAATATTTATCTATCTAT CTTTATATATTTCAAGTGG ATCTTTTATGACTAAATTA AAATACTTTTATATATTTT AACTTTTATCTTTATTTT CTATTTTATATCTATGAT TCTTATAATGTAATTTTAT GGGAAAGAGAGTAAACG GGGGCAATGGGTTAGT CTCAGTATGGATCCGGACC CTCCGGAATCGGATTCA ATGACTGGGGCAGTAGAT GCGGGATGGAAATTAAGA GTACTTTGGCAATAGGGTT TCTGGGACTGTGAGTGGT GAAACCGTAAAGGGGAG TAAGAAATCCGGCTATTAC CTGTACGGCTAGAGAAAT CTAGTTCAGCCTTGATGCC GTATTTTGGGAAAGTTT ATTCGCACTGCCGTGGTT GCGACCAAGTTTGATGGG AAGACACAAATAGATTTA GTTTTGGTCTTTTGCCATT ACTGAAATCTACAACGAA GCCACGGCTATAAAGAG TGGTACTCTTTGGTATAC CTTTAAGCACGGTAAGG AAGTGAATGGATGGGTT GCTCATTCGGGACTGCGC GTTGGAGCAATATATTTGA TCGGGTGATACTAAAAAG CCGGGAAATATCAATTACT AGTTTCTGGAATACTGCT GCTCACTTACAGTTTCTGG ATTACAAATTTACCCATAG</p>	<p>MAITGEGDCHLLKSLKLKTMKNKTNKLN KNGNDLDVTFPCDFCGLSFPTKIGRGVHISKA HRVERDAKLSGHASAKAYWYWSTEETRMA KIHLEESRGSNDHVLRAHMSACSTGRTLESI RLHMKAKRYVKIYTELNEKQIQSSGVENILN NISDASPLTLPSETETLTSYLLNPLKKGVSKW REQDLDFIHSVSDSEEQRLLDVYCKTVTEGY RKGTGPKKKSTKVSPHHAATTTKSKFLKKLK RELYGKVQQLWIKDRCEAAKLVLGSKVSTLIC EEDRSARQGEYWKKIFERESEPDIRPVPEVYA DPMMLVRPISEDEIKVVKDIKLTSSGPDNIST EVLREIPVLVLFNLILHLSHLPSEFKSRTTLI PKVETPKDPGDRPISVSVFVRALNKILASRV SNGKGAHAQKGFKALDGCLENTVLLNCILG DAKINKSIALSFIDMKKAFDSVNHDSILRAAT IAGYPKLLGYIKESYNGADTEVEKIKATFKRG VKQGDPLSPVLFNNIVELAIKAYESGIGYRM NGTVFPMAYADDIILLSESGELQDLDDIVLTE LRHNGLEANADKASIAKKDKGQKKWVISS DSVLAAGKTMPCMNANSMYKYLGIYFGPLGI TTERLTNYSNLNIGIAPLKPQRLMINDF LLPGLHEAVLGNNTNVGELKDLDMTRVAIRK WLHLPKDTCSSFIYTGKQGGGLGVRSLAEVQ FGKISRLSNLKSCKDTRIQAALTCEVVKKAFRQ WSKPIRVGDSIATSKVQFYANRELLSKCDG KGLREAGKIPFAHDWLRAKGLMKGRSFINSI KVRGNLFGTPVRLNRLNKSGLPIKCKMGCL MPCSLNHIAQTCPKVHGLRVERHNKVMNYI ADVARKGEWKTLEPIPIPYGATIRKPDVVIKD KKAMVLDIQVVSDVGNLNERNTLKSEVYNDK IADWWKNAGADSVFEGAITLNWRGLWASD SWETVSKIGVKNKSDAMIISLKVNLNWTARMYC AWNKG</p>	cd01 650	385	643
-------------	---------------------	------------------------	---	---	--	-------------	-----	-----

[illegible]

				CTAGTCGGACATTT CGATTATATCTTTT TATGTCATAGCTCT TTATCTTTTTTTTT TTAAATATTTTTAT AGTTTAGTCTACCT TGCTGATACCTCT CAAAATATATACGA TAGACTATCATTA TTCTGCTGCATTAC CATATAGCTTACTT AGAGGTTACTGCA GATCAATCACCTGT ACTGTATCTAGAGT ATAGTTAAGTGTAG GAGAACAGTAAAA ACAGTGGGCCAT TGAAAGGAGCGTTC ACGTCGGGTCCTGT CAGATGAAAACTGT CGAATCCTTCAACC ACACTACTCAAGTT TCTTTTAGTAACTA CCACAACCAACACC				SEVTADQVAHQLLNNGRGNPDHHPRRARLP KVPDALSSQDDPGFTRPFTMDLLAAIKAMK NNKAAGLDILCEQKQLRPGALNWLKMFN ECLGTNNIPKLWRKSKVIALKPGKDAASPKS YRPISLLCHTYKLFERLILNRIAPFVDEHLIPEQA GFRPGKSCCTGQLLNLTFIEDGYESKLITGAAF VDLSAAYDTVNHRLTRKIFELTKDVRTGLIQ TLLSNRRYFVVLNGNSRWRQKNGLPQGS VLAPLNFVNTNDQPIHPNTRSFYADDLCIAT QNQSFVKLEESLSDALAGLIPYYATNHLRANP DKTQISAFHLKNRDANHQLRISWYGKRLKHT PNPVYGLVTLDRSLTYKNHIANHTAKVGARN SILKLANTNWGTDAIRTITLALCFSAEY ASPVWSRSAHPKIDPALNSSCRAITGCLRPT KVEDIYLLCGIAPPHVRRVLVLSQKEKDKQENN PLHPLFEQQPVGKRLKSRHSFLHSVEPLNTSP HNQRMTLWSNHLQTTTPHKLSLSPKESLATGS SETWPTWLSLNLRLRTGYGRCRVLMQRWGL GDEQTACECGEQETMEHLLVCPLLPEPCTQ DDLEAVNCRAKACVDHWTGLV			
				TAACTCTAATATAAAACCT ATGGAATACATCTCACATA TAATAAACATAATATAATA AGCATAACATAATAACAAC ATAGCCGAAGGTGCTAGT AACCAGTGCTATCATTAA AAGCTAGGTTAATAGTTTT AACTCTAACCTCGCTGTAA ATAAATAATAATAA CTAGCAGCAATTA GGCTGCGGACAAA TTAAAAAGCTCAAG CAGCAAAAAACA AGACAGAAGAACC CCAATACTGACGCC	Drosophila persimilis	GCF_003 286085.1	i-1_DPe	MTLTIQWNLRGYNNYNELQIIKIQYQPHIAL QETHLFNIHNIPIPIFSLYAKNYSNFGGAAY LIHNSIESKEIHLQTDFTDATAIEVESKKFKIISA YISPSQNFESNLHNVLSNNSSSLIGDFNSW SKLWGSPPKNNARGNTVAKFLNNTNFILLNDK SPTHFTTHTNTYTHIDLSFCSPDLFQYANWKIID DLHGSDHFPILITLFFNNKQNKQNTQNNPIKIN SANWDAFKDKTDYFNTKNSPNNINKESSCL NKILLQSAHLTIPQTSIKNHPRNVPPWVWNGL QTLNNERKQAWNKLKHHITTTNIIRFKAAAK FKKEIKSAKNKLSSTFSEIQPQTKTSQTSWSNI RRFCGLNNMKNIHTIHPITLQPTSDNPSIAN SFAEYWSQSKSDSNFSSMFTNNKYSISTYQSS PPSQAQKIESDITSIELSSALSSLKGHTPGIDRI	cd01 650	491	749

				TTGAGTAGCAACAA CTGCTCAATGGAAA TAGAATCCGACTAA AACTAACAGATCAA GTAAAAACAACCA AATAACAGCACTC AAGAGATAAAAC TCAAAAAA				SYPMIANTSPAVKDRILKHYNNILNNYIPQAY KIGLVIAPGYKITSITSYRPSLNSCFAKVLDK IIAKRLWWFVTNNKHHITNQVGFGRGKSVTD NLAYDLHLITDALSHKQHSIVSLDFNKAFDKI GLHAITKQLQEWNIQPKILNYIINFMSKRKIC RVANNLSSTFPLHNGIPQGSPISVIFLIAYNKL ANLITQRNLTTAYADDFHIIVFKKNQTOEN FNLDLFQKIDWCFSFGSSLSFDCKCKHLICR KHNCCKTITRNLNIPSDVHLKILGLTFDSKYR WTRHIEDLASSLSNKINIICLSSNKFCNCTHT LISITKAIIVSKIDYGLVFGNSSITWLKIQITLN CAVRIALGAFRTTPTITNLLMEANINTVADRRD FLQAKLIKNLYFPNNSPLHKISKSLSQSRHFNN KIPSAITYSLQHCKKQNISLLIKPEIKTKLPFWKF QPNTINTELEKLDKNINTHDSIYQQHFLTINQ FNDHTFLTDDGSKTISTTSFAITTEKNTLVGLL QNFSIYSAEIIAIEAIIQIVKSMRGKYMICSDS LSSLQAINPDNFNYESKIRSLMLQLHPKIKL LWIPGHKNIKGNELADQKAKSASTMPLTYTH NNNISDINKHLKTKMLKDKLIKLEQTSNWWYQF INPNQITADKFTKRLDQNTTSRIKQTKIIRLIG HTRITHQHYLDPTKINTCTHCLNNVPVTIEHLL KDCPSSQLTDIRFKIFANNPLTFLSNPTKTEI DLIIKYLKETNLYPLI	SSF5 6672	436	801
	LOA- 2_DSer	GCA_008 042835.1	Drosophila serrata	TGGCTGGCAAAAA TCCTCCAGCTGAAC CCTCCACGAGGGA GCCGGCTAGGCAA TGGGCACTGTATTA CCGCGTACAATGGT AATGCAGTGGCGC ATGCTCTGTCCGGG GTAATGCAGTGTCT AGCTAAGGCTTCG ATTCAATTTCCTCC GCCTAAACGCCATG GTTGGAGGTGCCT GGCGGCAGGCATA	TAAGATTCTCCAGGACCAT GTGAAAAACAACATCCGG ATAAATAAGGGCCATACT GGCCTATGATTGGCCCTCT CAGGAGGGCCGTCGGT TAACCTAACCTAACCT	MSLRVLQLNLHKS KVASAELLAMEQGSADV ALVQEPWIASGNNAVAGLKSSNYNAFYTPSVS RNRTAVLVRKGLQANLMSHYSSDDLTVVLE SKKRLLVASCYMAHDRPAPPEELRSLVEPGP KDQLIVGADANAHHSVWVWSPDINDRGESL LDFILSTNLYIANVGEETHFVGP TSSNVLDITLV AGQSTMVSDWKVLERPSFSDHRYIQFQCEF DHSKPTVFRNPNTNNGKFDQIISRKLATS PGTVKSDEDEKSLTSLKELLGAYHASCPIRS KKRTRPPWWNKDLSLQRDDLREFFKLAKQA KDDIINEEYKILLRNYKKEIRRAQRSSWRNFC NIETAPETARLRKLLSKQPTIQSQLKRDDGQW TEGSEEAALTALMHGHFPGCTKITNTHPAISFW IGNLLSCRRVQSDWGTASIMVKGHRGTPQ					

269

CCAAGCTTCCTAGC TTGGCATCAACGGC CGGTTGCGCAAC AAAGTGTAGGTGA TGCTTGCCACCAAG CGGCAACGCTCGG CCGAGAGTGCCAT AAAGGACATTCCG GCGAGCAAGCGGG GACTCGTGCGCAA GGAAGCAGGCCCC CCTGGAAGGCGA AAACCAATCCAGA TTAACAAGAAGGC AGAGGACATCCTG TATCCCAGACACGG GAAGATGGCATGG GTAGCGTCTACCTG CGCCTCAAAAAGC GCCATCCTGGGGA TAAGGATGACCAC ACCCTGAAGGCAG GCGAGATGGAAA GGACCTAGGTCTG GAACCATTTTGA CGCGCCCAAGGAA CTAAACGCTGGATG AGTCGGAAGAGGA GGAAGACGGTGAC CTGACGATAGTAGT CAACCCGTCGAGT GGAGTTGAGCCCA GTACCC

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-
5 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.

Inspired by the success of retrotransposons in nature, it is further discussed here that the natural function of a retrotransposon can be recapitulated using functional parts derived from completely independent systems. For example, a functional Gene Writer™ can be made up of
10 unrelated DNA binding, reverse transcription, and endonuclease domains. This modular structure allows combining of functional domains, e.g., dCas9 (DNA binding), MMLV reverse transcriptase (reverse transcription), FokI (endonuclease). In some embodiments, multiple functional domains may arise from a single protein, e.g., Cas9 nickase (DNA binding, endonuclease), R2 retrotransposon (DNA binding, reverse transcription, endonuclease).

15 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
20 (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,
25 140, 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
30 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, a Gene Writer system is capable of producing a substitution into the target site of
5 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
10 an adenine, a thymine to a guanine, a cytosine to an adenine, a cytosine to a guanine, or a cytosine to a thymine.

Polypeptide component of Gene Writer™ gene editor system

Domains and functions:

15 In some embodiments, the Gene Writer™ polypeptide possesses the functions of DNA target site binding, template nucleic acid (e.g., RNA) binding, DNA target site cleavage, and template nucleic acid (e.g., RNA) writing, e.g., reverse transcription. In some embodiments, each functions is contained within a distinct domain. In some embodiments, a function may be attributed to two or more domains (e.g., two or more domains, together, exhibit the
20 functionality). In some embodiments, two or more domains may have the same or similar function (e.g., two or more domains each independently have DNA-binding functionality, e.g., for two different DNA sequences). In other embodiments, one or more domains may be capable of enabling one or more functions, e.g., a Cas9 domain enabling both DNA binding and target site cleavage. In some embodiments, the domains are all located within a single polypeptide. In
25 some embodiments, a first domain is in one polypeptide and a second domain is in a second polypeptide. For example, in some embodiments, the Gene Writer™ polypeptide may be split between a first polypeptide and a second polypeptide, e.g., wherein the first polypeptide comprises a reverse transcriptase (RT) domain and wherein the second polypeptide comprises a DNA-binding domain and an endonuclease domain, e.g., a nickase domain. As a further
30 example, in some embodiments, the first polypeptide and the second polypeptide each comprise a DNA binding domain (e.g., a first DNA binding domain and a second DNA binding domain).

In some embodiments, the first and second polypeptide may be brought together post-translationally via a split-intein.

Writing domain:

5 In certain aspects of the present invention, the writing domain of the Gene Writer™ system possesses reverse transcriptase activity and is also referred to as a reverse transcriptase domain (a RT domain). In some embodiments, the RT domain comprises an RT catalytic portion and an RNA-binding region (e.g., a region that binds the template RNA).

10 In certain aspects of the present invention, the writing domain is based on a reverse transcriptase domain of an APE-type or RLE-type non-LTR retrotransposon. A wild-type reverse transcriptase domain of an APE-type or RLE-type non-LTR 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
 15 heterologous reverse transcriptase from a different retrovirus, LTR-retrotransposon, or non-LTR retrotransposon. 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, or of an APE-type non-LTR retrotransposon from the R1, or
 20 Tx1 clade. In certain embodiments, a Gene Writer™ system includes a polypeptide that comprises a reverse transcriptase domain of a non-LTR retrotransposon, LTR retrotransposon, group II intron, diversity-generating element, retron, telomerase, retroplasmid, retrovirus, or an engineered polymerase listed in Table 1 or Table 3. In some embodiments, a Gene Writer™ system includes a polypeptide that comprises a reverse transcriptase domain listed in Table 2. In
 25 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 non-LTR retrotransposon, LTR retrotransposon, group II intron,
 30 diversity-generating element, retron, telomerase, retroplasmid, retrovirus, or an engineered polymerase whose DNA sequence is referenced in Table 1 or Table 3, or of a peptide comprising

an RT domain referenced in Table 2. In some embodiments, the RT domain has a sequence selected from Table 1 or 3, or a sequence of a peptide comprising an RT domain selected from Table 2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the RT domain comprising a Gene Writer

5 polypeptide has been mutated from its original amino acid sequence, e.g., has at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions. In some embodiments, the RT domain is derived from the RT of a retrovirus, e.g., HIV-1 RT, Moloney Murine Leukemia Virus (MMLV) RT, avian myeloblastosis virus (AMV) RT, Rous Sarcoma Virus (RSV) RT. In some

10 embodiments, the RT domain is derived from the RT of a Group II intron, e.g., the group II intron maturase RT from *Eubacterium rectale* (MarathonRT) (Zhao et al. *RNA* 24:2 2018), the RT domain from LtrA, the RT TGIRT (or trt). In some embodiments, the RT domain is derived from the RT of a retron, e.g., the reverse transcriptase from Ec86 (RT86). In some embodiments, the RT domain is derived from a diversity-generating retroelement, e.g., from the RT of Brt. In

15 some embodiments, the RT domain is derived from the RT of a retroplasmid, e.g., the RT from the Mauriceville plasmid. In some embodiments, the RT domain is derived from a non-LTR retrotransposon, e.g., the RT from R2Bm, the RT from R2Tg, the RT from LINE-1, the RT from Penelope or a Penelope-like element (PLE). In some embodiments, the RT domain is derived from an LTR retrotransposon, e.g., the reverse transcriptase from Ty1. In some embodiments, the

20 RT domain is derived from a telomerase, e.g., TERT. A person having ordinary skill in the art is capable of identifying reverse transcription domains based upon homology to other known reverse transcription domains using routine tools as Basic Local Alignment Search Tool (BLAST). In some embodiments, the reverse transcriptase contains the InterPro domain IPR000477. In some embodiments, the reverse transcriptase contains the pfam domain PF00078. In some embodiments, the RT contains the InterPro domain IPR013103. In some embodiments,

25 the RT contains the pfam domain PF07727. In some embodiments, the reverse transcriptase contains a conserved protein domain of the cd00304 RT_like family, e.g., cd01644 (RT_pepA17), cd01645 (RT_Rtv), cd01646 (RT_Bac_retron_I), cd01647 (RT_LTR), cd01648 (TERT), cd01650 (RT_nLTR_like), cd01651 (RT_G2_intron), cd01699 (RNA_dep_RNAP), cd01709 (RT_like_1), cd03487 (RT_Bac_retron_II), cd03714 (RT_DIRS1), cd03715

30 (RT_ZFREVEV_like). Proteins containing these domains can additionally be found by searching the domains on protein databases, such as InterPro (Mitchell et al. *Nucleic Acids Res* 47, D351-

360 (2019)), UniProt (The UniProt Consortium Nucleic Acids Res 47, D506-515 (2019)), or the conserved domain database (Lu et al. Nucleic Acids Res 48, D265-268 (2020)), or by scanning open reading frames for reverse transcriptase domains using prediction tools, for example InterProScan. The diversity of reverse transcriptases 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.

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).

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 (SEQ ID NO: 1539) or YMDD (SEQ ID NO: 1540) motif in an RT domain (e.g., in a reverse transcriptase) is replaced with YVDD (SEQ ID NO: 1541). In embodiments, replacement of the YADD (SEQ ID NO: 1539) or YMDD (SEQ ID NO: 1540) or YVDD (SEQ ID NO: 1541) 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).

In some embodiments, the reverse transcriptase domain is one selected from an element of Table 1 or Table 3.

Table 1: 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
MarathonRT	Group II intron	CBK9229 0.1	D4JMT 6	MDTSNLMEQILSSDNLNRAYLQ VVRNKGAEVDGMKYTELKEH LAKNGETIKGQLRTRKYKPQPAR RVEIPKPDGGVRNLGVPTVTDRF IQQAIAQVLTPIYEEQFHDHSYGF RPNRCAQQAILTALNIMNDGND WIVDIDLEKFFDTVNHDKLMTLI GRTIKDGDVISIVRKYLVSGIMID DEYEDSIVGTPQGGNLSPLLNI MLNELDKEMEKRGLNFVRYAD DCHIMVGSEMSANRVMRNISRFIE EKLGLKVNMTKSKVDRPSGLKY LGFGFYFDPAHQFKAAPHA VAKFKKRMKELTCRSWGVSN SYKVEKLNQLIRGWINYFKIGSMKT LCKELDSRIRYRLRMCIWKQWK TPQNQEKNLVKLGIDRNTARRV AYTGKRIAYVCNKGAVNVAISN KRLASFGGLISMLDYYIEKCVTC (SEQ ID NO: 1542)	IPR000477 , PF00078, cd01651
TGIRT, trt	Group II intron	AAT7232 9.1	Q6DKY 2	MALLERILADRNLITALKRVEAN QGAPGIGDVSTDQLRDIYRAHWS TIRAQLLAGTYRPAPVRRVGIPK GPGGTRQLGITPVVDRLIQQIALQ ELTPIFDPDFSPSSFGFRPGRNAH DAVRQAQGYIQEYGRYVVDMD LKEFFDRVNHDLIMSRVARKVD KKRVLKLIRYALQAGVMIEGVK VQTEEGTQPGGPLSPLLANILLD DLDKLEKRGKFCYRADDNCNI YVSKLRAGQVRVKQSIQRFLEKTL KLKVNEEKSVADRPWKRAFGLF SFTPERKARIRLAPRSIQRLKQRI RQLTNPWNWSISMPREIHRVNQYV GMWIGYFRLVTEPSVLQTIEGWI RRRLRLCWQLQWKVRTRIREL RALGLKETAVMEIANRTKGAWR TTKPQTLHQALGKYTWTAQGLK	IPR000477 , PF00078, cd01651

				TSLQRYFELRQG (SEQ ID NO: 1543)	
LtrA	Group II intron	AAB0650 3.1	P0A3U0	MKPTMAILERISKNSQENIDEVFT RLYRYLLRPDIYYVAYQNLYSN KGASTKGILDDTADGFSEEKIKKI IQSLKDGTYYYPQPVRRMYIAKKN SKKMRPLGIPTFTDKLIQEAVRIIL ESIYEPVFEDVSHGFRPQRSCHTA LKTIKREFGGARWFVEGDIGKCF DNIDHVTLIGLINLKIKDMKMSQ LIYKFLKAGYLENWQYHKTYSG TPQGGILSPLLANYLHELDKFVL QLKMKFDRESPERITPEYRELHN EIKRISHRLKKLEGEKAKVLE YQEKRKRLPTLPCTSQTNKVLKY VRYADDFIISVKGSKEDCQWIK QLKLFHINKLKMELSEEKTLITHS SQPARFLGYDIRVRRSGTIKRS KVKKRTLNGSVELLIPLQDKIRQ FIFDKKIAIQKKDSSWFPVHRKYL IRSTDLEIITIYNSELRGICNYGL ASNFNQLNYFAYLMEYSCLKTIA SKHKGTLSTISMFKDGSWSGI PYEIKQKGQRRYFANFSECKSPY QFTDEISQAPVLYGYARNTLENR LKAKCCELCGTSDENTSYEIHHV NKVKNLKGKEKWEMAMIAKQR KTLVVCFHCHRHVIHKHK (SEQ ID NO: 1544)	IPR000477, PF00078, cd01651
R2Bm	non-LTR retrotransposon	AAB5921 4.1	V9H052	MMASTALSLMGRCPDGDCTRGK HVTAAPMDGPRGPSSLAGTFGW GLAIPAGEPCGRVCSPATVGFFP VAKKSNKENRPEASGLPLESERT GDNPTVRGSAGADPVGQDAPG WTCQFCERTFSTNRGLGVHKRR AHPVETNTDAAPMMVKRRWHG EEIDLLARTEARLLAERGQCSGG DLFGALPGFGRTEAIKGQRRRE PYRALVQAHLARFGSQPGPSSGG CSAEPDFRRASGAEEAGEERCAE DAAAYDPSAVGQMSPDAARVLS ELLEGAGRRRACRAMRPKTAGR RNDLHDDRTASAHKTSRQKRA EYARVQELYKKCRSRAAAEVID GACGGVGHSLSEMETYWRPILE RVSDAPGPTPEALHALGRAEWH	IPR000477, PF00078, cd01650

				GGNRDYTQLWKPISVVEIKASRF DWRTSPGPDGIRSGQWRAVPVH LKAEMFNAWMARGEIPEILRQC RTVFVPKVERPGGPGEYRPISIAS IPLRHFHSILARRLLACPPDARQ RGFICADGTLENSAVLDAVLGDS RKKLRECHVAVLDFAKAFDTS HEALVELLRLRGMPEQFCGYIAH LYDTASTTLAVNNEMSSPVKVG RGVRQGDPLSPILFNVVMDLILA SLPERVGYRLEMELVSALAYAD DLVLLAGSKVGMQESISAVDCV GRQMGLRLNCRKSAVLSMIPDG HRKKHHYLTERTFNIGGKPLRQV SCVERWRYLGVDFEASGCVTLE HSISSALNNISRAPLKPQQRLEILR AHLIPRFQHGFLGNISDDRLRM LDVQIRKAVGQWLRLPADVPKA YYHAAVQDGGLAIPSVRATIPDL IVRRFGGLDSSPWSVARAAAKSD KIRKKLRWAWKQLRRFSRVDST TQRPSVRLFWREHLHASVDGRE LRESTRPTSTKWIRERCAQITGR DFVQFVHHTHINALPSRIRGSRGR RGGGESSLTCRAGCKVRETTAHI LQQCHRTHGGRILRHNKIVSFVA KAMEENKWTVELEPRLRTSVGL RKPDIIASRDGVGVIVDVQVVS QRSDELHREKRNKYGNHGELV ELVAGRLGLPKAECVRATSCTIS WRGVWSLTSYKELRSIIGLREPT LQIVPILALRGSHMNWTRFNQMT SVMGGGVG (SEQ ID NO: 1545)	
LINE-1	non- LTR retro- transposon	AAC5127 1.1	O00370	MTGSNSHITILTLNVNGLNSPIKR HRLASWIKSQDPSVCCIETHLT CRDTHRLKIKGWRKIYQANGKQ KKAGVAILVSDKTDFKPTKIKRD KEGHYIMVKGSIQEEELTILNIYA PNTGAPRFIKQVLSDLQRDLDSH TLIMGDFNTPLSILDRSTRQKVN KDTQELNSALHQTDLIDIYRTLH PKSTEYTFFSAPHHTYSKIDHIVG SKALLSKCKRTEIITNYLSDHSAI KLELRIKNLTQSRSTTWKLNLL LNDYWVHNEMKAEIKMFFETNE NKDDTTYQNLWDAFKAVCRGKFI	IPR000477 , PF00078, cd01650

				ALNAYKRKQERSKIDTLTSQLKE LEKQEQTHSKASRRQEITKIRAE KEIETQKTLQKINESRSWFFERIN KIDRPLARLIKKKREKNQIDTIKN DKGDITTDPTIEIQTIREYYKHLY ANKLENLEEMDTFLDTYTLPRLN QEEVESLNRPTGSEIVAIINSLPT KKSPGPDGFTAEFYQRYKEELVP FLLKLFQSIEKEGILPNSFYEASIIL IPKPGRDTTKKENFRPISLMNIDA KILNKILANRIQQHIKKLIHHDQV GFIPGMQGWFNIRKSINVIQHINR AKDKNHVIISIDAEKAFDKIQQPF MLKTLNKLIDGMYLKIIRAIYD KPTANIILNGQKLEAFPLKTGTRQ GCPLSPLLFNIVLEVLARAIRQEK EIKGIQLGKEEVKLSLFADDMIV YLENPIVSAQNLLKLISNFSKVSG YKINVQKSQAFLYNNNRQTESI MGELPFTIASKRIKYLGIQLTRDV KDLFKENYKPLLKEIKEDTNKW KNIPCSWVGRINIVKMAILPKVIY RFNAIPIKLPMFTFFTELEKTTLFI WNQKRARIAKSILSQKNKAGGIT LPDFKLYYKATVTKTAWYWYQ NRDIDQWNRTEPSEIMPHIYNYLI FDKPEKNKQWGKDSLLNKWCW ENWLAICRKLKLDPFLTPYTKINS RWIKDLNVKPKTIKTLEENLGITI QDIGVGKDFMSKTPKAMATKDK IDKWDLIKLSFCTAKETTIRVNR QPTTWEKIFATYSSDKGLISRIYN ELKQIYKKKTNNPIKKWAKDMN RHFSKEDIYAAKKHMKKCSSLA IREMQIKTTMRYHLTPVRMAIHK KSGNNRCWRGCGEIGTLVHCW WDCKLVQPLWKS VWRFLRDLEL EIPFDPAIPLLGIPKDYKSCCYK DTCTRMFIAALFTIAKTWNQPNC PTMIDWIKKMWHIYTMEYYAAI KNDEFISFVGTWMKLETHLSKLS QEQKTKHRIFSLIGGN (SEQ ID NO: 1546)	
Penelope	non-LTR retrot	AAL1497 9.1	Q95VB 5	MERSPEPSININGRHAVCTATNM SYAKIKTKYKDSKRTINKFQLTL VKLTKLKSSLKFLKCRKSNLIPN	IPR000477 , PF00078, cd00304

	ransp oson			<p>FIKNLTQHLTILTTDNKTHPDITR TLTRHTHFYHTKILNLLIKHKHN LLQEQTKHMQKAKTNIEQLMTT DDAKAFFESERNIENKITTTTLKKR QETKHDKLRDQRNLALADNNTQ REWFVNKTKEFPPNVVALLAKG PKFALPISKRDFPLLKYIADGEEL VQTIKEKETQESARTKFSLLVKE HKTKNNQNSRDRAILDVEQTR KLLKENINIKILSSDKGNKTVAM DEDEYKNKMTNILDLDLCA YRTL RLDPTSRLQTKNNTFVAQLFKM GLISKDERNKMTTTTAVPPRIYG LPKIHKEGTPLRPICSSIGSPSYGL CKYIIQILKNLTMDSRYNKNAV DFKDRVNNSQIREEETLVSFVV SLFPSIPIELALDTIRQWKLEEH TNIPKQLFMDIVRFCIEENRYFKY EDKIYTQLKGMPMGSPASPIAD ILMEELDKITDKLKIKPRLTKY VDDLFAITNKIDVENILKELNSFH KQIKFTMELEKDGKLPFLDSIVSR MDNTLKIKWYRKPIASGRILNFN SNHPKSMIINTALGCMNRMMKIS DTIYHKEIEHEIKELLTKNDFPPNI IKTLLKRRQIERKKPTEPAKIYKS LIYVPRLSERLTNSDCYNKQDIK VAHKPTNTLQKFFNKIKSKIPMIE KSNVVYQIPCGGDNNKNCNSVYI GTTKSKLKTRISQHKSDFKLRHQ NNIQKTALMTHCIRSNHTPNFDE TTILQQEQHYNKRHTLEMLHIIN TPTYKRLNYKTDTENCAHL YRH LLNSQTTSVTISTSKSADV (SEQ ID NO: 1547)</p>	
M- MLV RT	Retro virus	ADS4299 0.1	P03355[660- 1330]	<p>TLNIEDEHRLHETSKEPDVSLGST WLSDFPQAWAETGGMGLAVRQ APLIPLKATSTPVSIKQYPMSQE ARLGKPHIQRLLDQGILVPCQSP WNTPLL PVKKPGTNDYRPVQDL REVNKRVEDIHPTVPNPYNLLSG LPPSHQWYTVLDLKDAFFCLRLH PTSQPLFAFEWRDPEMGISGQLT WTRL PQGFKNSPTLFDEALHRDL ADFRIQHPDLILLQYVDDLLAA TSELDCQQGTRALLQTLGNLGY</p>	IPR000477 , PF00078, cd03715

				<p>RASAKKAQICQKQVKYLGILLK EGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAE MAAPLYPLTKTGTLFNWGPDQQ KAYQEIKQALLTAPALGLPDLTK PFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPP CLRMVAAIAVLTKDAGKLTMGQ PLVILAPHAVEALVKQPPDRWLS NARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHT WYTDGSSLLQEGQRKAGAAVTT ETEVIWAKALPAGTSAQRAELIA LTQALKMAEGKKLNVTDSRYA FATAHIHGEIYRRRGLLTSEGKEI KNKDEILALLKALFLPKRLSIHC PGHQKGHSAEARGNRMADQAA RKAATETPDTSTLL (SEQ ID NO: 1548)</p>	
RSV RT	Retro virus	AAC8256 1.1	P03354[709- 1567]	<p>TVALHLAIPKWKPDHTPVWIDQ WPLPEGKLVALTQLVEKELQLG HIEPSLSCWNTPVFVIRKASGSYR LLHDLRAVNAKLVPFGAVQQGA PVLSALPRGWPLMVLDLKDCFFS IPLAEQDREAFATLPSVNNQAP ARRFQWKVLPQGMTCSPTICQL VVGQVLEPLRLKHPSLCMLHYM DDLLEASSHDGLEAAGEEVIST LERAGFTISPDKVQREPGVQYLG YKLGSTYVAPVGLVAEPRIATLW DVQKLVGSLQWLRPALGIPRL MGPFYEQLRGSDPNEAREWNLD MKMAWREIVRLSTTAALERWDP ALPLEGAVARCEQGAIGVLGQG LSTHPRPCLWLFSTQPTKAFTAW LEVLTLLITKLRAVVRTFGKEV DILLPACFREDLPLPEGILLALK GFAGKIRSSDTPSIFDIARPLHVSL KVRVTDHPVPGPTVFTDASSSTH KGVVVWREGPRWEIKEIADLGA SVQQLEARAVAMALLLWPTTPT NVVTDSAFVAKMLLKMGEQGV PSTAAAFILEDALSQRSAMAAVL HVRSHSEVPGFFTEGNDVADSQA TFQAYPLREAKDLHTALHIGPRA</p>	IPR000477 , PF00078, cd01645

				LSKACNISMQQAREVVQTCPHC NSAPALEAGVNPRGLGPLQIWQT DFTLEPRMAPRSWLAVTVDTASS AIVVTQHGRVTSVAVQHHWATA IAVLGRPKAIKTDNGSCFTSKSTR EWLARWGIAHTTGIPGNSQQQA MVERANRLLKDRIRVLAEGDGF MKRIPTSKQGELLAKAMYALNH FERGENTKTPIQKHWRPTVLTEG PPVKIRIETGEWEKGWNVLVWG RGYAAVKNRDSDKVIWVPSRKV KPDITQKDEVTKKDEASPLFAG (SEQ ID NO: 1549)	
AMV RT	Retro virus	HW60668 0.1	-	TVALHLAIPKWKPNHTPVWIDQ WPLPEGKLVALTQLVEKELQLG HIEPSLSCWNTPVFVIRKASGSYR LLHDLRAVNAKLVPFGAVQQGA PVLSALPRGWPLMVLDLKDCFFS IPLAEQDREAFATLPSVNNQAP ARRFQWKVLPQGMTCSPICQLI VGQILEPLRLKHPSLRMLHYMD DLLLAASSHDGLEAAGEEVISTL ERAGFTISPDKVQREPGVQYLYG KLGSTYVAPVGLVAEPRIATLWD VQKLVGSLQWLRPALGIPRLM GPFYEQLRGSDPNEAREWNLD KMAWREIVQLSTTAALERWDPA LPLEGAVARCEQGAIGVLGQGLS THPRPCLWLFSTQPTKAFTAWLE VLTLLITKLRAVARTFGKEVDIL LLPACFREDLPLPEGILLALRGFA GKIRSSDTPSIFDIARPLHVSLKV RVTDPHPVPGPTVFTDASSSTHKG VVVWREGPRWEIKEIADLGASV QQLEARAVAMALLWPTTPTNV VTDSAFVAKMLLKMGQEGVPST AAAFILEDALSQRSAMAAVLHV RSHSEVPGFFTEGNDVADSQATF QAY (SEQ ID NO: 1550)	IPR000477 , PF00078, cd01645
HIV RT	Retro virus	AAB5025 9.1	P04585[588- 1147]	PISPIETVPVKLKPGMDGPKVKQ WPLTEEKIKALVEICTEMEKEGKI SKIGPENPYNTPVFAIKKKDSTK WRKLVDFRELNKRTQDFWEVQL GIPHPAGLKKKKSVTVLVDVGD YFSVPLDEDFRKYTAFTIPSINNE TPGIRYQYNVLPQGWKGSPAIFQ	IPR000477 , PF00078, cd01645

				SSMTKILEPFRKQNPDIVIYQYM DDL YVGS DLEIGQHRTKIEELRQ HLLRWGLTTPDKKHQKEPPFLW MGYELHPDKWTVQPIVLPEKDS WTVNDIQKL VGKLNWASQIYPGI KVRQLCKLLRGTKALTEVIPLTE EAELELAENREILKEPVHGVYYD PSKDLIAEIQQGQGWYQIYQ EPFKNLKTGKYARMRGAHTNDV KQLTEAVQKITTESIVIWGKTPKF KLPIQKETWETWWTEYWQATWI PEWEFVNTPLVLWYQLEKEPI VGAETFYVDGAANRETKLGKAG YVTNRGRQKVVTLTDTTNQKTE LQAIYLALQDSGLEVNIVTDSQY ALGIIQAQPDQSESELVNQIEQLI KKEKVYLAWVPAHKGIGGNEQV DKLVSAGIRKVL (SEQ ID NO: 1551)	
Ty1	LTR retrotransposon	AAA6693 8.1	Q07163- 1[1218- 1755]	AVKAVKSIKPIRTTLRYDEAITYN KDIKEKEYIEAYHKEVNQLLK MKTWDTDEYYDRKEIDPKRVIN SMFIFNKKRDGTHKARFVARGDI QHPDTYDSGMQSNTVHHYALM TSLSLALDNNYYITQLDISSAYLY ADIKEELYIRPPPHLGMNDKLRL KKSLEYGLKQSGANWYETIKSYLI QQCGMEEVRGWSCVFKNSQVTI CLFVDDMVLF SKNLNSNKRIIEK LKMQYDTKIINLGESDEEIYDIL GLEIKYQRGKYMKLGMENSLTE KIPKLVNPLNPKGRKLSAPGQPG LYIDQDELEIDEDEYKEKVHEMQ KLIGLASVVGKFRFDLLYYINT LAQHILFPSRQVLDMTYELIQFM WDRDKQLIWHKNKPTDPDNKL VAISDASYGNQPYYSQIGNIYL LNGKVIGGKSTKASLTCTSTTEA EIHAISESPLLNNLSYLIQELNK KPIIKGLLTDSRSTISIIKSTNEEF RNRFFGTKAMRLRDEVSGNNLY VYYIETKKNIADVMTKPLPIKTF KLLTNKWIH (SEQ ID NO: 1552)	IPR013103 , PF07727
Brt	Diversity- gener	NP_95867 5.1	Q775D8	MGKRHRNLIDQITTWENLLDAY RKTSHGKRRTWGYLEFKEYDLA NLLALQAELKAGNYERGPYREF	IPR000477 , PF00078, cd01646

	ating retro elemen t			LVYEPKPRLLISALEFKDRLVQHA LCNIVAPIFEAGLLPYTYACRPDK GTHAGVCHVQAE LRRTATHFL KSDFS KFFPSIDRAALYAMIDKKI HCAATRLLRVVLPDEGVGPIG SLTSQLFANVYGGAVDRLLHDE LKQRHWARYMDDIVVLGDDPEE LRAVFYRLRDFASERLGLKISHW QVAPVSRGINFLGYRIWPTHKLL RKSSVKRAKRKVANFIKHGEDES LQRFLASWSGHAQWADTHNLFT WMEEQYGIACH (SEQ ID NO: 1553)	
RT86	Retro n	AAA6147 1.1	P23070	MKSAEYLNTRLRNLGLPVMNN LHDMSKATRISVETLRLLIYTADF RYRIYTVEKKGPEKRMRTIYQPS RELKALQGWVLRNLDKLSSSPF SIGFEKHQSILNNATPHIGANFILN IDLEDFPSLTANKVFGVFHSLGY NRLISSVLTKICCYKNLLPQGAPS SPKLANLICKLDYRIQGYAGSR GLIYTRYADDLTLSAQSMKKVV KARDFLFSIIPSEGLVINSKKTCS GPRSQRKVTGLVISQEKVGIGRE KYKEIRAKIHIFCGKSSEIEHVR GWLSFILSVDSKSHRRLITYISKL EKKY GKNPLNKAKT (SEQ ID NO: 1554)	IPR000477 , PF00078, cd03487
TERT	Telo meras e	AAG2328 9.1	O14746	MPRAPRCRAVRSLLRSHYREVL LATFVRRLLGPQGWRLVQRGDPA AFRALVAQCLVCVPWDARPPPA APSRQVSCLELVARVLQRLCE RGAKNVLAFGFALLDGARGGPP EAFTTSVRSYLPNTVTDALRGSG AWGLLLRRVGDDVLVHLLARCA LFVLVAPSCAYQVCGPPLYQLGA ATQARPPPHASGPRRRLGCERA WNHVSREAGVPLGLPAPGARRR GGSASRSLPLPKRPRRGAAPEPE RTPVGQGSWAHPGRTRGPSDRG FCVVSPARPAEEATSLEGALSGT RHSHPSVGRQHHAGPPSTSRPPR PWDTPCPPVYAETKHFLYSSGDK EQLRPSFLLSSLRPSLTGARRLVE TIFLGSRPWMPGTPrRLPRLPQR YWQMRPLFLELLGNHAQCPYGV	IPR000477 , PF00078, cd01648

				LLKTHCPLRAAVTPAAGVCARE KPQGSVAAPEEEDTDPRLVQLL RQHSSPWQVYGFVRACLRLVP PGLWGSRHNERRFLRNTKKFISL GKHAKLSQLQELTWKMSVRDCA WLRSPGVGCVPAAEHRLREEIL AKFLHWLMSVYVVELLRSFFYV TETTFQKNRLFFYRKSVWSKLQS IGIRQHLKRVQLRELSEAEVRQH REARPALLTSRLRFIPKPDGLRPI VNMDYVVGARTFRREKRAERLT SRVKALFSVLNYERARRPGLLGA SVLGLDDIHRWRTFVLRVRAQ DPPPELYFVKVDVTGAYDTIPQD RLTEVIASIIKPQNTYCVRRYAVV QKAAHGHVRKAFKSHVSTLTDL QPYMRQFVAHLQETSPLRDAVVI EQSSSLNEASSGLFDVFLRFMCH HAVRIRGKSYVQCQGIQGSILST LLCSLCYGDMENKLFAGIRRDGL LLRLVDDFLLVTPHLTHAKTFLR TLVRGVPEYGCVVNLKRTVVNF PVEDEALGGTAFVQMPAHGLFP WCGLLLDTRTLEVQSDYSSYAR TSIRASLTFNRGFKAGRNMRRKL FGVRLRKCHSLFLDLQVNSLQTV CTNIYKILLQAYRFHACVLQLPF HQQVWKNPTFFLRVISDTASLCY SILKAKNAGMSLGAKGAAGPLPS EAVQWLCHQAFLKLTRHRVTY VPLLGSLRTAQTQLSRKLPGTTL TALEAAANPALPSDFKTILD (SEQ ID NO: 1555)	
Mauriceville RT	Retro plasmid	NC_001570.1	Q36578	MPNHRPLNCVSYLGENHELSQL HGMFGLLKRSNPQTGGILGWLN TGPNGFVKYMMNLMGHARDKG DAKEYWRLGRSLMKNEAFQVQ AFNHVCKHWYLDYKPHKIAKLL KEVREMVEIQPVCIDYKRVYIPK ANGKQRPLGVPTVPWRVYLHM WNVLLVWYRIPEQDNQHAYFPK RGVFTAWRALWPKLDSQNIYEF DLKNFFPSVDLAYLKDKLMESGI PQDISEYLTVLNRSVLVLTSEDKI PEPHRDVIFNSDGTNPNNLPKDV QGRILKDPDFVEILRRRGFTDIAT	cd00304

				NGVPQGASTSCGLATYNVKELF KRYDELIMYADDGILCRQDPSTP DFSVEEAGVVQEPKSGWIKQN GEFKKS VKFLGLEFIPANIPPLGE GEVKDYPRLRGATRNGSKMELS TELQFLCYLSYKLRKVLRDLIQ VLGYPVPLLRYSRLAEAINELS PKRITIGQFITSSFEFTAWSPLKR MGFFFSSPAGPTILSSIFNNSTNLQ EPSDSRLLYRKGSWVNIRFAAYL YSKLSEEKHGLVPKFLEKLREINF ALDKVDVTEIDSKLSRLMKFSVS AAYDEVGTLALKSLFKFRNSERE SIKASFKQLRENGKIAEFSEARRL WFEILKLIRLDLFNASSLACDDL SHLQDRRSIKKWGSSDVLYLKSQ RLMRTNKKQLQLDFEKKKNSLK KKLIKRRAKELRDTFKGKENKEA (SEQ ID NO: 1556)	
RTX	Engineered polymerase	QFN4900 0.1	-	MILDTDYITEDGKPVIRIFKKENG EFKIEYDRTFEPYLYALLKDDSAI EEVKKITAERHGTVVTVKRVEK VQKKFLGRPVEVWKLYFTHPQD VPAIMDKIREHPAVIDIYEYDIPF AIRYLIDKGLVPMEGDEELKLLA FDIETLYHEGEEFAEGPILMISYA DEEGARVITWKNVDLPYVDVVS TEREMIKRFLRVVKEKDPDVLIT YNGDNFDFAYLKKRCEKLGINF ALGRDGSEPKIQRMGDRFAVEV KGRIHFDLYPVIRRTINLPTYTLE AVYEAVFGQPKEKVYAEIITA WETGENLERVARYSMEDAKVTY ELGKEFLPMEAQLSRLIGQSLWD VRSSTGNLVEWFLLRKAYERNE LAPNKPDEKELARRHQSHEGGYI KEPERGLWENIVYLDFRSLYPSIII THNVSPDTLNREGCKEYDVAPQ VGHRFCKDFPGFIPSLLGDLLEER QKIKKRMKATIDPIERKLLDYRQ RAIKILANSLYGYYGYARARWY CKECAESVIAWGREYLTMTIKEI EEKYGFKVIYSDDTGFFATIPGA DAETVKKKAMEFLKYINAKLPG ALELEYEGFYKRGLFVTKKKYA VIDEEGKITTRGLEIVRRDWSEIA	IPR006134 , PF00136, cd05536

				KETQARVLEALLKGDGVEKAVR IVKEVTEKLSKYEVPPEKLVHK QITRDLKDYKATGPHVAVAKRL AARGVKIRPGTVISYIVLKGSGRI VDRAIPFDEFDPTKHKYDAEYYI EKQVLPAPERILRAFGYRKEDLR YQKTRQVGLSARLKP KGTLEGSS HHHHHH (SEQ ID NO: 1557)	
--	--	--	--	--	--

Table 2: InterPro descriptions of signatures present in reverse transcriptases in

Table 1.

Signature	Database	Short Name	Description
cd00304	CDD	<i>RT_like</i>	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	<i>RT_Rtv</i>	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	<i>RT_Bac_retron_I</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	<i>TERT</i>	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	<i>RT_nL TR_like</i>	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	<i>RT_G2 _intron</i>	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	<i>RT_Bac _retron _II</i>	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	<i>RT_ZF REV_li ke</i>	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	<i>POLBc _B3</i>	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	<i>DNA_p olB_Ko dl_like _exo</i>	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	<i>RVT_1</i>	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	<i>DNA_pol_B</i>	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	<i>RVT_2</i>	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	<i>RT_domain</i>	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 retroviral elements, such as the retroviruses and bacterial retrons. The enzyme catalysing polymerisation is an RNA-directed DNA-polymerase, or reverse transcriptase (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 polypeptide 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 (SEQ ID NO: 1558), 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].

Table 3 (below) shows 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 sequences 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.

Table 3. Exemplary Retrotransposon Sequences

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	GTCTAGTTACAACCTGGGC ATCGCTGCAGAGATCGCA CCTCCTCGTGGTCCCGCTG GTAGCCCTTCGAAGGGTG ACTAAGTCGATCTCGCC CAGGTACGAGCCGTTGG GACTACCAGTCCAACGTA ACTCCTGCCTAAATTCGGT GAAACAAATTCCTCGGTAA AAAGCCCC (SEQ ID NO: 1140)	TTCAGTTATTAGATGC TTAGTTTTGTACCTTTCT TGTGTTGTTAGGATTTG ATAGTGTAGTATTTTAT ATTTTGTACGATTGCAT AATGTTCTTTTTATACAG TTCTGTTTTAATAAATA GACGATAGCTAGAGACG TTAGGGCAGCCACAAGC CAGTTAGGTAGCGGATA GTAGGTAGGAACAGACT TTTACTATTTTCATAACGC GTCAATTACCACCTGATT TGGACCAATTCACGGGAT TTGTCCAAGGTGGACGG GCCACCTTTACTTAACCC GGAAGGGAACATATAT AATTATGTGTGTTCCGAT AAA (SEQ ID NO: 1263)	MASCPKPPVVSAGAMSLGSLTHSVLAIERGPNLSANGSDFGGG GLGLPLRLRVSVGTQTSRSDWVLDVSWSHPGPTSKSQQVDLSLFP KHRVDLLSKNDQVDLVAQLPSKFPNLAENDLALLNLEFYRSDLHV YECVFAAHWEGLSGLPEVVEQLAPQPCVGTELHSSLPRDSELFVPEE GSSEKEDAPKTPPTPGKHGLEQTGEEKVMVTVDPKNPPCCCGT RVNSVLNIEHLKLVSHGKRGVCFRCACGKENSNSYHSVCHFPCRG PETEKAPAGEWICEVNCNDFTTKIGLQGHKRLAHPAVRNQERIVASQ PKETSNRGAHKRCWTKEEEELLRLLEAQFEKNKINKLIAEHITTKAK QISDKRLLSRKPAEPEEPGTCCHTRRAAASLRTEPMISHAQAE DRDNGPGRPLPGRAAGGRTMDEIRRHDPKNGNGQQRPTKQSEE QLQAYYKTLERLSAGALNTFPRAFQVMMEGRDKLVINQTAQDCF GCLESIRQTATRDKKDVTREKHPKPFQKWMKDRAIKKGNLYRF QRLFYLDRGKLAILEDIECLSCDPLSEIYSVFKTRWETTSFSLGDF KTYGKADNTAFRELITAKEIEKNVQEMSKGAPGPDGTLGDVVKMD PEFRTMEIFNLWLTGKIPDMVRGCRVLPKSSKPDRLKINNWRP ITGISILLRFSRIVTARLSKACPLNPRQGFIRAAAGCSENKLLQTIWS AKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNM YENISTYTTKRNTHTDKIQRVGVKQGDPMSPLLFNAMDPLCKLEE SGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETCNLTLGL KTQGGKCHGFYIKPTKDSYNTINDCAAWTINGTPLNMIIDPGESEKYL LQFDPWIGIARSGSLKDFWLQRIQIDQAPLKPQKTDLKTYTIPRLIYI ADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSTRDGGGIGIT KLAGLPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWQAG GDRENIPIIWEAPPSSEPPNNVSTNSEWEAPTQDKDFKPCNWRKN EFKKWTKLASQGRGIVNFERDKISNHWIYQYRRIPHRKLLTALQLRAN VYPTREFLARGRQYIKACRHCADIESCAHIIGNCPVTQDARIKRH NYICELLLEEAKKDWVVFKEPHIRDSNKELYKPDLFVKDARALVVDV TVRYEAAKSSLEAAAEKVRKYKHLETEVRHLTNAKDVTFVGPPLGAR

R2	R2-1_Gfo	.	Geospiza fortis	AGACTTAAGTGAGTTGG TTACAACTGGGCATAGCT GCAGAGACCGCGCCTCT CGCGGCCCGCTGGTAAG CCCTTAACAGGGTGACTAA (SEQ ID NO: 1141)	GGTAGATAATCTTTGTAT AGTGGGGGGGATCTCA TGTACCGGTTTCTTTTAT TTGATTTTCAATAAACA GACGGTAGCTAGGTTCC CAAGGCAGCCACAAGCC AAAGATAGGTAGGGTGC TCATAGTGAGTAGGGAC AGTGCCTTTTGATTACAA ACGGGTCAATACCATCTG ACACGGATACCTTACCG GACTTGTGATGATCTCC AGACTTGTCCAAGGTGG ACGGGCCACTTTACTTA ACCGGAAAAGGAACAT ATATTAATTATGTGTTCC GGAAAA (SEQ ID NO: 1264)	GKWHQDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFASRA RKSVM (SEQ ID NO: 1016) VGLCPSPGVDGTHQPNDSFQNGFETNFSVQVVARLVTRNLAPRSVRG NGFESGMATHPVPADESGHSDPLVGRSCGQPARLTRQSVGTQTS RDDILPSKTKLTENELDLVNFSLYRSDLQGFVQEGHFHFSVNRVLE GFPEVVEQPAQPAVGDDLNTSLPPDNNICVLEKSGSEAVEDGTPEV AHPVPETQGESPNIVMTLPNKNPPCPCRCRVLHSLVALIEHLKGS HGKRAFCRCVKGRENFNYHSTVCHIAKCKGPKVKEKAPVGEWICEV CGRDFTTKIGLQGHKRLAHLPLVRNQERIDASQPKETSNRGAHKRCW TKEEEEMLIKLEVQFEFHNRINKLAEHLTTKTSKQSDKRLLPRKQLT DISKGVAGQKVLDPGLSHQPLQGVWDNGLGGGHLPGPAAEGRTIE PLGHLDKDNHGREIADQHKAGRLQAHYRKIRKRLSEGMSINFPEV FEQLDCQEAQPLINQAQDCFCGLDSASQIRKALKRQNTQKQDQGD QPKRPAQKWMKKRAVKGHFLRFQKLFHLDRGKLAKIILDDVECLSC DIPPEIYSVFKARWETPGQFAGLGDFFINRANKNAFRDLITAKEILK NVREMTKSAPGPDGIALGDIRKMDPEYTRTAEFLNLWLTSGEIPD MVRGCRVLPKSSKPERLKDINNWRPITIGSILLRLEFSRIITARLTACAP LNPRQRSFISAAGCSENLKLTIRTAKNEHRPLGVVVDIAKAFDTV SHQHIIHLVLRQRRVDPHIIGLVKNMYKDISTVITTKNTYTDKIQVVG VKQGDPLSPLFLNAMDPLCKLEEKGHGFHGRGQSKITAMAFADDLV LLSDWEDMNAIKILETCDLTGLTKGQKCHGYIKPTKDSYTVNN CAAWTINGTPLNMINPGESEKYLQFDPWVVGIAKTSLEKLDWFLE RIDRAPLKFQKLDILKTYTIPRLTYADHSEMIKAGALEALDRITSAVK DWLHLPSTCDAILYTSMDGGLGVTKLVGLIPSVQARRLHRIAQSP ETMKDFLEKAQMEKMYEKLWVQAGGKRMRMPSIWEALPEVPSID TATTSEWEAPNPKSKYRPNWRRKEFKWTKLIAQGWGIRCFKGD KISNNWIRHYRIPHRKLLTAIQLRASVYPTREFLARGREDNCVSCRH CEAAEESCAHIIGMCPVVRDARIKRNRICERLMEEAGAKRDWTVFQE PHIRDVTKELYKPDLYFVEGLALVDVTRFESTKTTLEAAAAEKVNY KHLETEVRNLTKAKDVIFMGFPLGARGQWYNKNFELDLTGLPSR QDIIAKTLSTDALISSVDIIMFASRRRQHA (SEQ ID NO: 1386) NKFLGSRVAYCLKPGPPVSDRGEFGSGLTTHPEPESESGHDPVTPN PGPSLGAGEGAQPLILRVSVGTQTCEDFITSRPTKLPGLIESELGPLV KFSLEVYRDLKGDVQFEGHFPDNWGVLEGFPEVYEQLAPQPNNG DELNHLSPDREGDVLKDSSEKAAPEALPSVQRARSEQLPDNIV KVTVPDKNPPCCGVRNLNSVLAHLEHLKSGHRRVCFRCACKGR NFNHSTVCHYAKCKGQIERPPVGEWICEVGRDFTTKIGLGQHKR HMHAMVRNQERIDASQPKETSNRGAHKRCWTKEEELIMKLEVQF
R2	R2-1_ZA	.	Zonotrichia albicollis	CGACTTGAGAAGGTCTGG TTACAACTGGGCATAGCT GCAGAGATCGGCCTCCT CGTGGCCCGCTGGTAAG CCCTTAACAGGGTGACTAA GTGATCTCTGCCCGGATC CAGGAGCCGCTGGGTTTC	GTAGTACATTGCACCTT CTGTAACCTGCCTGGGT GTGGGATGGGCTCGG GGTGGGTTATGGGGT ATATATGTGGGATATTCT GGTGGGAATGTCCATTCA CTGTATGCCTATCTTTT	

				<p>ACCA GCCCAGCGATTCTT CCAAATTCGGTGA (SEQ ID NO: 1142)</p>	<p>ATAAAAAGACGGTAGCT AGGTTCCGGAAGCAGCC ACAAAGCAATAGCCAGTT AGGTAGCTCATAGTGGG TAGGTGACAGGAACCTTT GACTCAGAACGCGTCCAT TAACATCTAGAACGGACC AACTTCGACATGCACC GATTAAACCGGATTGTCC AAGGTGACGGGCCACC TTTACTTAACCGGAAAG GGAACATATAGTTATA TGTGTTCTGTAATA (SEQ ID NO: 1265)</p>	<p>ENHKNINKLIAEQITTKAQISDKRRMILLKGRGTTGNLETEPGMS HOSQAKVNDGLGDDHLPGPVVDKGTIGKPGQHLDTDNSHQITA GKKGGGLQARYRRRIMKRLAAGTINIFPKVKELINDEQEARLINQT TEDCFGLLDACQIRALTREKGSQEERPKQYQKWMKKRAIKRGDY LRFQRLFLDRGLARILNDNTESLSCDISPSEIYVFKARWETPGHFG LGDFEIKANNKAFRDFITAEIKENVREMSKGSAPGPDGIALGDIK KMDPGYSRTAEFLNLWLTAGDIPDMVRCRTVLIPKSTTPERLKDIN NWRPITIGSILLRFSRIITARMTKACPLNPRQGFISAPGCSENKLLQ SIIRTAKEHKLPGVIFVDIAKAFDTVSHQHHIHLVQQRRVDPHIVGLV NNMYKDISTYVTTKNTHTDKIQRVGVKQGDPLSLFLNAMDPLL CKLEESGKGFHRGQSSITAMAFADDLVLLSDSWENMKENIKILETCFN LTGLTQKQKCHGFYKPTKDSYTINNCPAWTTINGTPLNMINPGESE KYLGLQDPWTVGAKYDLSLTKLWLESIDRAPLPLQKLDILKTYTIPR LYLADHSEMKALEALDQIRTAVKDWLHLPSCCTDAILYVSTRD GGLGVTKLAGLIPSVQARRLHRIAQSPDETMKDFLEKAQMEKMYEKL WVQAGGKKGMPSIWEALPMTVPPTNTGNLSEWEPNPKSKYPKP CDWRRKELKWTLESQGRGVKNFRNDTISNDWIQYRRIRPHRKLIT AIQLRANVYPTREFLAGRGDNVVKFCHRCADLETGCHIGFCPVTK DARIKHNRICDRLECEAAKREWVVKFEPHLRDAATTELFKPDVIFVKE DRALVVDVTVRYESAKTTLAAAMEKVDKYKHLAEVKELTNAKDVV FMGFPLGARGKFKYGNFNLLETGLPKTRQLSVAKTSLTYALMSSVDI VHMFASRSRPNV (SEQ ID NO: 1387)</p>
R2	R2Dr	AB097 126	Danio rerio	<p>AATCCCCCTACCCAATCC CCCCGTCGTGACCTCCAGG CCAGGAATCAGGAGCGTA CGACAGTGGCCATCCGGC AATGACAATAGCGTGACT AACGACAATGAGTCAGAT CCATGACCTTGGAGTGG GTTAACCTCCGCTCTTTA AAAAA (SEQ ID NO: 1143)</p>	<p>AAATCCCAGGGGATAC AGCAAGAAGGTATCGGA TCTAATAAGGTTGAGCGA GGAGAGGTTGAGATCC TTTGGGGGGGTCGGGC TAAGTCCCTCTCGGGT CTCCACCGGTGACGCTC TACCCCTCCCTCCTCGCTC GTAGAACCCCAACGGTGA ACACGGTTGGCAGGATG AAGTGACGTGAGGGGTA AGACATGCGTACGTGAG CGCGCATTTTGTCTGTCT CTGGACTGGGTTTCGTCC CCCTCACAAACCATCACTT ACACTATAGGGGCACAG</p>	<p>MESTAKGKSYWMARRPVEGATGSLGRVPFVTRDPKRKPEAKRTL HGLGLRECSVWLTRLIEGRRGRDHTPSGWNQRMGNPNDSESVPEPN GPIPSNPIPTGTQALPEPMADGEQEHGPGVVVTLPLRDLNCPCLCGGS ASTAVKVRHLAFRHGTVPVRFSCSCGKTSPGCHSVLCHIPKCRGPT GEPPEKVKCEGCSRTFTGTRRACSIHEMHVHSEIRNRKRIAQDRQEK GTSTDGEGRAGVERADAGEGSPGEGIPPKRRRARTPREPSEPPANP PILSPQDLPFGGLRDLREVASGWVRAARDGGTVIDSVLAALWDG NDRLPELVDAAATQRTLQGLPAGRLARRPATFVAPNRRGRWRRLK LLAKRRAYHDCQIRFRKDPARLAANILDGKSETSCPINEQAIHEHFRNK WANPSPFGLGGRGTENRANNAHLGPISKSEVQTSLRNASNASTP GPDGVGKRDISNDWPECTLTQLFNMMWFTGVIPSRLLKSRVLLP KSSDPGAEMEIGNWRPITIGSMVLRLFRVINTRLTEACPLHPRQRGF RRSPGCSENLEVLCLLHSEKFSQLAVFVDFAAQAFDTVSHHEMILS VLEQMNVDPMHVNLIREIYNTSCTSVELGRKEGDPVVRVGVKQGD PLSPLLFNALDPLIQSLERTGKGCEAGHKVTALAFADDLALVAGSW EGMAHNLAIVDEFCLTTGLTVQPKKCHSFMRVPCRGAFVNDICPP</p>

R2	R2-1_GA	.	Gastero steus aculeat us	<p>CATATTGGGGTCTCAGGA GGAGACACAGGGTCTGT GCGCTCCGGTAAACGGT ACCGAGTCGGTTAAGCA TCGTTTGGGCCCCCTCCA CGTGGTGGTCCGCGGTAA CACCAATAGGGTGGCTAA GAGCCCCAGTAATTTCCCC</p>	<p>CGCGGACCTTTAACGAG AACCGAACCGCAACTTCT CGCATCGCTCCCGGACT TTCCCCCTCGTTCAGCC GAGGGATGCCAAAAGGC ATGAAAGGTAAGTACCAT ACCGTCCGCAAACTCT CTTCTGACTCGGTCTCT GTTGGTTTTCTAGAGTAA CAACGAGGTGGAGGAGA GGGACATGGCAGGGACT CCCATTCGTGCCAGCGGG TGGGACAGATCGAAGG AACGGTTCGAGGGCGTA ACAGACGAGAGGGAATC CGGTACACATTGATGCC ATGCCTAAATAGGCGAG GTTTGATTTCTACTTTGT GGGTTCAGTATAGTCGG AGCATATGGTCGGTTGTC CCGTTGTTTTCACGCGCG GCAAGCGACTATCATGAT AAAGTAGAATGGGAGAC GGGCTCCCTGACAAACCC GGAAAGCGCCCCCGG TGGTTCGTAGCAGCTGAC GGATCACGCTCGAAGAA AAATGAGTGAGAGGGGA CGCGCAACCAC (SEQ. ID NO: 1266)</p>	<p>MLRGGVTPPAGGAGAVGPGMASPGGCSVRFSPGRRLLGHRGT GLSPSVWRLKRLSVLRRWSGPGLLGADGAGGGAASPRGTQVL GSGARRWLHGSGSSPSAARGLRRLTVRLKRLSGGLSPKACRDA EEGSSSPGFRNPKGLGGRGLTPLGSRFFCRLTVSLNRWRGSLVKLN ASSRASGRRTPVKPACDSRAGRSEHAEGGVSAAPMVLRSRRLT FSVDGDSNSGDRARSGSVSAARPGHLLVDGESASSRSGPAGDARLA GPSTRSRKGCCLPPVDFENPKRTRLMAKMTNGNPTSHVPCAPCS NGHEGGGRVAVIEGRLELSGSRISGIQPALPVETSFVGQSTGRGAD</p>
----	---------	---	-----------------------------------	--	---	--

GAATTGCTCTCCCCCGC GCGGGGGACCCCTT TAGTGTGGAGCGGTGC GCCTCCGCTTGGGGT TCGACGGCGTGAGCCTC GTCCCTTAAGTTCAGACG GTCCGGCTTCTGCCGG CCAAACCCCGTGCAGCG TTCTCCATGTTGGATCGG CACCCAGCCCGGGTGCC ATGCGAGTTCAGACATTT GTTTATGTATCTCTCGGT GGTTGACTTGCTAAGCTCA TTTCTCTCTCACTGCGTC CCCCAGGTGCTGATCGG TTGAAGAGGATTCGTCGTT GACTCGCGGTGAATTT GGGATTGTATTACAGG TAGGTATAGAGGCGGTGC GG (SEQ ID NO: 1144)	GTGGGCTCGACAGGTA GGGGTTACATGACTCCGT GCTGCTCAGCAGACCCGC GCCTCTGACACCGGTA GGGCTACTTGAACAAGC GACGCCCTGGTGATGTC CGTATCCTAACCTGGTTT GGAAAGCCGATACCCG CAATGCCGCCACAGGT GTCGCGCACCCACCGG ATGACGTATGGGCCCG GGGACCTCATGGATAC TCCACTGGACTTGACAA TCTGCTGCTACTGGATGC AGCGACGTTGGTGACAT AAGCAATCGCTAAGTCG GGTAGGGGAGGTGGG GACCTCGCAGCGCTGT AGGAACGGGTGTATGGG CTCCGGCAGCGCTCGTCA CTCCCATACACACAGGG GCTGCATCTGTTGGCCG GTGCTAGTTGGTTCTGGA AGCCCGCCGGCTGGT TCGAGAGCAGGGGTGC GCCAGGGTAGGTTTGG TATATCTGGGTCGGTGC GATACCTATCGATGGCA GCGAGGCCCGCTCGTG ACGCGCTGTGGAGCT GGAGCCGCTGGGTAT GAACAGTCTTGGCGATG TGCGTAGCTAGATAGT ACCGTGGTTGTGGCG TGGTGTGACCAATGTT GTCTGTGTGCACATAGG CCAAGGGTTACGTGGT GGCAGTCAGAGCACCC	GDANANSSPPNLGSGVMVPAVRDGTPLGRPGEDHSREACGG NTPLWMLSEDFRCDYCPREFGTAGRSLHMRRAHAEYDAGFCW GERLSEFAATRLWSTEEKLAVFCERGVSPSECRAIAASLGAKTH HQVRSKRLVFEAIRRELLEVAATERLEKSARRKQPAVPAPVHGV RGVLRGLLGRVPRGGTGTSTARSIVRRDDCRQGAVASASLNLRL GRKATGRSRRRLVGRPRMDVRRSVRMRMRFLYRLARLGWAK LAMFLDQMGASCVPPLVEVS AVFRERWSIVRAFLGLGQFGGT ADNAGFKLIDPAEVRALHQSINKNRSSPGDGTKVALSKWDPEGIKL AHMYSTWLVSAKIPKVFKKCRTLLPKTGDVSLHGDVGQWRPITIASL VRLYSLRILTERMTVACPSHPRQRGFASPGCSENLMLEGCMSLSKA NGSLAWFVDFAKAFDTVSHEHLLSVLVQKGLDQHMVELIKDSYEN SVTKVHCQEGCSTDIAMKVGVKQGDMSPLFLNLDPLLIQLEREG RGFPVNGKSITAMAFADDDLAIVSDSWEGMRANLDILVDFCELTGMR TQPSKCHGFLIEKSGSRYSKVNRCPEWLLNDTALHVMVGPKEKYLGV QVNPWTGIFAEDTVAKLRQWVAISKTLPLRDKVSLLCQFAVPRVIF VADHCLMSAKALTEMDSIRQAVKRWLHLARCTTNGLLYSRKSGG LGIPKLSMIVPAMQARRLLGSRSKDETVMFLTTDDHVAFERAW LRAGGSPDEVPELGPDLVEGSPAEGNADPVSTVRPRKRIVPCDW/RQ VEFDRWAGQLVQKGIRTFEADKSNCWLYDYPPNKLKPGDFTAAV QLRANVYPTRELAGRRTDIDVCCRHCGEAPETCWHLALCPKVKR CRIQRHHKVCQVLVAEAEHRHGWVEVEREKRWMLPSGECVAPDLICW LDELALIVDVTVRYEFDEESLERARIECKRYRPLIPVIRASRVQTKKVTV YGFPLGARGKWPAKNELLADLGLSKARTSRFAKLLSRRVLLHSLDVM RTFMR (SEQ ID NO: 1389)
--	---	--

				<p>GCACCTGGAAGTGATTG CCCCGGATCCCGGCTCT CTGTGAAGAGCTACCTTG AGGAAAGGTGTTCCGCT GGAACCTCAAGACCTACA GTAGGGGATATCAACTG GCTTTAGGTGCTGTGAT TCCGGAACCAAGGCGAG GCGAGTACTTAGACA TGTCCAAAAGCCCCGGG AACGTTCGGGGGCTG CTTGGGTCTTGGACCA CATCCGTAAACGATGGA TCTCGGTCGGCGCTCGG GAGAACTCCCGCATGAA CGCTGATTGCATGTGAGA ACGCCCCACGGCGCG GGGCAGGGCTCCCCCT GGGTGAAGGCTCGGGG GGGTCAGGCTCCGCTCT AAAAG (SEQ ID NO: 1267)</p>	<p>MMASTALSMGRCPDGDCTRGKHVTAAPMDGPRGPSSLAGTFGW GLAIPAGEPCGRVCSPTVGFPPVAKSKENRPEASGLPLESERTGD NPTVRSGADVPVQDAPGWTCQFCERTSTNRGLGVHKRAHPV ETNTDAAPMMVKRRWHGEEIDLLARTEARLLAERGQCSGGDLFGAL PGFRTLEAIKQRRRREPYPYALVQAHILARFGSQPGSSGGCSAEPDF RRASGAEAEVEERCAEDAAAYDPSAVGQMSPDAARVISELEGAGR RRACRAMRPKTAGRRNDLHDDRTASAHKTSRQKRAEYARVQELYK KCRSRAAAEVIDGACGGVGHSLSEEMETYWPIERVSADAPGPTPEAL HALGRAEWHGGNRDYLWKPISVEEIKASRFDWRTSPGPDGIRSG QWRAPVPVHLKAEMFNAMMARGEIPEILRQCRTVFVPKVERPGPG EYRPIIASIPLRHFHSILARRLLACPPDARQRFICADGTLENSAVLD AVLGDSTRKLRCHVAVLDFAKAFDTVSHEALVELLRGRMPEQFCG YIAHYDTASTTLAVNNEMSSPVKVGVRQGDPLSPILFNVMIDLIL ASLPERVGYRLEMELVSALAYADDLVL LAGSKVGMQESISAVDCVGK QMGLRLNCRKSAVLSMIPDGHKKHHYLTERTFNIGGKPLRQVSCVE RWRYLGVDFEASGCVTLEHSSSALNNISRAPLKPQQRLEILRAHLIPR</p>
R2	R2_B M	AB076 841	Bomby x mori	<p>GGGCGATACGATAATTTT AATTTCCGATTGAAATCC AGTCGTCTTAATCTGGTGA CCAGTGGCGGGTCACCA GTATAGTGACAGGACGT GAATGGCTCCGAGGCTGG CGGAGTCACTACTATAAG TGTGAGAGACGATGTCCT GTGCCAAGTATACGTCCAA CCCTAACGGGTTAAGTGA AATTAGTTGCTCATAACAG GGACGGGTACCTGTTTG CTCGTGGCTGGCTATCGA ATGGACGGGACCAATACA CCCCCCTGTTAGTAATGGG GTAAAGAGAGAGCGGCTCTG</p>	<p>GCCTTGACAGTAGTCCA GCGGTAAGGGGTAGAT CAGGCCCGTCTGTTCTT CCCCCGAGCTCGCTCCCT TGGCTTCCCTTATATTAA CATCAGAAACAGACATTA AACATCTACTGATCCAAT TTCGCCGGCGTACGGCCA CGATCGGGAGGGTGGA ATCTCGGGATCTTCCGA TCCTAATCCATGATGATT ACGACTGATGACTAAA GACGATGGCATGATGAT CCGGCGATGAAAA (SEQ ID NO: 1268)</p>

R2	R8Hm-A	.		<p>AAACTATGGCGAAATCA CGAGCCCACTCTACCC ATAACCTGCACGTGGTACC GCCGCACATTGACCGATAC GGGAGGAGGGGCGACAC TTGAATCACGTAGTCTTGG TGAGCCATTGCGGACT ACAGCCCTCTAGTGCC GCCTTAGAACGCAACGGG GCAATAGGTGGGCGGG GCGCTAGCGGGGGGAG TAATCTCCCTGTTGGCGT GCACCGCACTGCTCCCACT GGGGCAGTGTATCCGG AAACAGGTGGGCGGGG GCCACGAGGGGAGCAA TCCCTCTG (SEQ ID NO: 1145)</p>	<p>TAATGCCAAAAGTTGCT TGGGCTAAATGATACGTA CGCTAGAAAAGCGACTT GCTGCACGGATGACGGT TCATCAGAGCCGATATG TGATGCAAGCGGCA GGGAGAACTCACTAGTGT AGCTGTTCTTCCATTAC GACTACGCGGTTAACGT GGCAGATAGATTTACAC CAGGAAATAATACGTGA AGGGTTCCACCATATCT GGAGTTAGATCTATGAG GGAAACATTTGTAATAAG TCAGTCTGTAACCTGGC GCCGCTGTTGAGTCAAT TAACTATGCAACTACTCA TTAAGTTATCGACTTTGA TATGGCATGGGGTGATTG CGCGTTATATCAAGTCA</p>	<p>FQHGFVLGNISDDRMLDVQIRKAVGQWLRLPADVPKAYYHAAV QDGLAIPSVRATIPDLIVRRFGLDSSPWSVARAAKSKIRKKLRW AWKQLRFRSRVDSITQRPSVRLFWREHLHASVDGRELRESTRPTST KWIRERCAQITGRDFVQFVHTHINALPSRIRGSRGRRGGESSLTCRA GCKVRETTAHLQQCHRTGGGRILRHINKIVSFVAKAMEENKWTVLE PRLRTSVGLRKPDIIASRDGVVVDVQVVGQSRSDLHREKRNKYG NHGELVELVAGRLGLPKAECVRATCTISWRGVWVSLTSYKELRSIIGLR EPTLQIVPILALRGSHMNWTRFNQMTSMVMGGGVG (SEQ ID NO: 1390)</p>	<p>MNLLIVTSIKESDVPSSGKGGVAVNNITAGASGKDTCVIIHPGTDGI WCCTCEVIEHNSGKDLKRLHAKRHPSVTISGYKCNLCPFVSEQLSVG THLYRCRGVKEVVKREFACASCSSDTSGLQVHMQRKHAEWIND QLKEKTEFAWTDRELRLAEKELTTPSFRYNKIFYAALGTSRTYDAVRK IRYNDRYKSAIAEMRSQIAADAAAQERDVERGLVSAHSDRGKEML PVWETKSDIQVNNDIKKDIELTPNSRQKQTNLALARPVIEVEDLGR QDVQKYLASLRQDDYTSPAERSIFAYCREETNWSATKRQVLIKISRTTR GLRQPKKVRPFEEPEGFKPNRMKRWKRVFLQECYREKRAETVSKI LDGTFIDEPEEEIRPELEEVRMYIDRLKRTQLDITTKIVQTDDEVFCLQ SYGRITIGEVDRDALGASKDSASGPDGLLLQDVRRLPLLLCNIFNMW YLHGIPVEENRCRTILLYKSGDRHLASNRPVTIGNMLNRLYAKIWDK RIRKNVRLHVRQKAFIPVDGCFENVKTIQCVLQSVYRKRKLEHNVVFIDL AKAFDTVLHDSIRKALWRKGVPSGVVVDLSLVAGAVTSISVGKTKT RSICINSGVKQGCPLSPLLFNLIIDELAERIEATGCGLDLDGHVLSMAF ADDYVLLAKDSVMENELIRVCSTFFKEKGLSNPGKQCSLRVLPVKEK KRSMKVLVRPHRWRIKDDQDVDPMSMTYDSLGKYLGVSDPTGKIAL PIEWNWMTKLKECKLKPEQVKILKEVVCSSRVNVLVLMSECGISEL RSWTRFVRNWAKNIIHLPTWCSSDVIHSIKGLGIPDVSKGIVIQMR ASEKMTSEDGIVRVVGARLVQKNRVLWEKAGFEGIELKAARRHCEV ERLNNIGNITNGVALKTIAAVSSVNRNWMIEDNLKSGNKILVWKAM</p>
----	--------	---	--	---	---	--	---

				AACATGATGATTGCAATG AGAAACTACCACGCTTGG TCAGGTTTGTAGGAGA ACATCTCATTCAAGCCTC CCGGATGTCGGCACCCG CTGACATCTTCTGGCTTA TGAAATTTTCAATTAATT TTGTAAGTCATGGCGG CTTGAAAGC (SEQ ID NO: 1269)	AGAIPTKINLSRGVADQTLKKRRRCGLTAETDGHILAGCHTSSDAYSK RHNMCDKLAKELKNGPNRRVWRERTCTSTGRRYRPDIIVKDDSD KITVIDMTCPYEKSEGLIQCESAKVYEPKLDKYWTRLEGANGIV AEKVELMGLAIGTIGTMRSTLRKLCELKSGRIVRRLQMIACNNSAQII KGHLSTRATRRNLR (SEQ ID NO: 1391)
R2	R8Hm- B	.	Hydra vulgaris	CTTGGGTCACCTGACACAT TTTTCGGTAGCCATAGTTT TTTGAGAGGAAGAGTGGA AGTTTTTCCATGAGTCGTC TCTCGTATAAACTGTGGTA AATCCGGCCATCCAGCCTC TACGCGGCGCAACTAGAA ACTTGGATCAGTGATCAA GGCTAATGGATGACGGGA CTCCATGGATAAGGAGAT ATAAGATCTTATTGAAAC GCATCTTAAGGGGTTATG GGGCTAACACCCCTTAAT TCTGGTGACATTTATTGA CCGTT (SEQ ID NO: 1147)	MSNRITIGDVPVSGKGLTVNKTAGADGAEACWVHPGAKGIWSS PACLRKFTIGKELRAHLAQIHLAPSAVRYRCNKCPEYEGDVQLSVGTH LRYCKGIAGVVEEKQFACAIKNSFSDTFSGLQVHKQRKHVVEWNEQ LKEKTEFAWTDRELRELAKEVTIPFSVNTETFAVLDTTRTKDAVRKI RYTDYKSLAEVRAQVNAVEEAPQASDESQITLLVNTGRGAELQP AVINITSIELVTDVNEVEMVTSNSTNEEQINAPVEPAVIEADLGRQ DAKLYLASLRQSDCTNASDRWTLAYCRGEVDWCKTKSRLFVSRHA RGLRQPRVENWEPEGRPNRLKWRKYSFLOSCYRTKKKETS KILDGTFKDTPEEIRPELEEVQRVYVDRLEVRTQLDTRTVHIDERFD LVSYGRTIREVQDAISAKSDASGGPDGLLLQDVKKASPRQLCIIFNM WYLHGIPVWENRCRTILLHKGGEKHLTSNYPVTIGNMLNRYAKIW DRIRKNIQLHVRQKAFVPLDGCENVKTIQILQSYRRSRREHNWVF VDLAKAFDTILHDSIEKALLRKGIPRSVIKVVDSLYAGAVTSITVGKTKT RPICNSGVKQGCPLSPLLFNVIDELAERLEATCGGLDLEGHVSSMA FADDYVLLAKDSVEMNVLMMVNCNTFFEEKGLAVNPAKCSLRVLPV KGKRSMKVLTTRTHRWKINNQDVEIPSMTYESVGKYLGMIDPAG KIALPIEEWKLWLTRELCKLPDQKVVLKEVVCARANVVLMSGC GICELRWSRFVRGWVKSIIHFPWCNSEWMHSSKGLGIPDVVSGI VIQRMRAAEKMAKSTDGVRVVGARIVQTNRVLWKRAGLAGIELD AARKECEVRKVNKIGNQTNNGGALKTIAESSVSRHWLLEKNIRPGNKIL VWKAMAGVPTKINLSRGVADQTLKKRRRCGLTAETDCHILAGCPTS RDAYSKRHNLLCDKLAKELRLNGGPRRVRWRERMLCSGNRRYKPD IVVKDDGVITVIDMACPEKESHERLSQCEDAKVAKYEPLRLDRSWTQ ELEGNNGRSANEISWGIAGVGAIGTITRKTQRILSKLKLAKVGRPLQIIA CNESAQIRRHLSGSRLNLR (SEQ ID NO: 1392)
				ATGCCGAGGTAGTTGG GATAATGATGCACAAAGCT CGTAAGGCGACTTGCTGC ACGTATGCCGCTAAACGC TTAGCTCGATGAGTGCT GTCAAGACGGTCGGGAG TATGATCAGTGGAGCTG ACTTCCAGACAACTCAC GCGGATTCGGTGCGGT GGATACAAACACTGGTAT AACATATGAAGGGTTCCA TCTAGTACAGGGATAACG ATCCATGGGAGCAAACTA ATTAGTTGGAGGTAATCC AACGCCGCTGTGAGTCA GTTTTTAAACCGCCAGTCA ACTCTGTAGGTTATCGG TCTTCGGCAGACCTTGGGA CCGCCTAGCCCGGCCA ACAGTTTGTGCTGCACTA ACATGATGATTTCGCGAGA GAAACCCACGCTTTGTCA CTTATGTGAGGATAAAAT CTCTTGCCATATGATCCT TTGAAGGGAAACACGCGCT TTGAGCTTGTCTCGCGTT GGCACCTTAGTCTGTAA TATTTCTTGATATTAIGG	

R2	R9Av	GQ39 8057	Adineta vaga	<p>GAATAGTTTGCATGGT AGGTGATGCGCCTCTG TGCTCTCTTTGCTGGAT ATAGTTTGACGATTTTGA CCAGGTATCTGTTCTTGT GAGTTCAGCACCAGTTTG AACAGGCTTAGCGATAGA CCTTCGAACCTTGAAACACT GTTGTGAAGCTGGCTGGG CCCTGCAGATTTTCTCGA TTAGAACGTGAGTGTAC GTCAGAAATGACCCACCA GTGGTAGTTCTACGTTGC CCTGGAAGGAGAAAAGT TGAGCTAAATCGCACGG CCTAGTTGTTTATCAATA GGCAGGTGAGGAACCTCT TCTATGTACCCTGACTAAA GTAACACTTGTGCGCTGG GTTTGCTCCCTCGCATTT GACTTATCTGATCGCACTA CCACCAAAACGAAACATA AACTTAGCTCGTGGTATCA GTCCACAGCGTGTGCAGT CGGATTCAGGGGAGCGTG TTAGTGACAAGCAGGATA ATATTACATAGTTAATGT TAAGGCGTTCAACATTCCT TATCCAAATTGGAAGAGTT GACTGTGAAGTTTGTCTATG AAGACATTGGACAA (SEQ ID NO: 1148)</p>	<p>ACGAAAAAGGTAGTATG GTTGCA (SEQ ID NO: 1270)</p> <p>ACTAGTCTCTCTCTCTAT TAGTCAGTCTAATTAATT TTCTTACATCTACATCT AGTTCATTATTAATTTG GTATGATCAGTGCTATCT CTGCTACACTCAATGCTT AATCGTATGTTATTGACA GTCTGACACTTGATTACT CTTACGACATATGCACTG TTGCTTCAGAGAAACCA CTGTTATATAGTGAAGT TCCTCAGTTTCTGTTGAT ATATTCTCTTTCTATTCTC GCTTCTCTTTTCTACTGT GTTCTTTTATCAGTTTTT TGTGGAATAATGAGAA TAAATAAAGT (SEQ ID NO: 1271)</p>	<p>MNLPREHAVSVHNINKFNVLQCLSKSYDTINSVKAHYVACRRQKN ASSTTAVPTNVNNINQLAINTNQVSRNP LQCVECLMKQVDFYAKDT KALVTHMRTKHAAYEESKVVATRRVAVWSPDEQILAELEVKLKIQ KGQLLSRLWEYNKCAKDSKPSRSDAIRTRRQHDYKLLRLSLQSQ QPPVGSESDSDISSNNNPLTTTHNVTPDPSSNVLLIQLKRESVDS IVKTNLNTNMLNAASAFINQNNMDPLELSMRGIEEDVKAIRDK ELQPTRNVPSSTTSRKPTRNAKRLKSKYGYQHYLYNNKKLVAEI LDGETSGAKPPPMNLVEDYVYRNWRSSTIDSPVNNIKTVNSDSIFAP ISRDEIKLALSNKTKDSAAAGPDVATKEAKAIIDNLVYVYNIWLVGQGIP EQLKNTLILIPKGNDSLKLNWRPITISSILRVYNNRLLAYRMINKIFKT NDKQVGFKPVNGCGINISWLSLLKHARLNKNISYACLVDSKAFDS VSHQSVRALTMNGAPSLVLLKIMDQYTNVNTVITCSGSISKNISSG VKQGDPLSLLFNVLIDELFDVIKQYGYTIDNIGTINARCFADDLTIS SSRMGMINKLELTTKFFKERGLNVNPSKMSIGMSKGYGKKSKIES EPLFSITDAQIPMLGYDKTTRYLGVNFTSIGAIDAKRIKKDLQDLDKL EHLKKAQCKMDLLRTYMPFRMFQLIHTELYPKLLIKMDILIRKLAKRI LHLPISSEFFVLPFKEGGQLTSLKEAVGLAKIKLHKIMSSNDPMLC YLIESQRIVEHFMDKLDGSLTLNEMNNIKECFMKEKRISFAQKIH GVGFESSPLTNQWINGEIKMTTITKTYINSIKLRTNLTETRTTSG LNIITCRRCHVADESLMHVLQCCSTKGLRYSRHHKICAKVANKLVM NGYGVFREKSYPPDPNNSGSLRPDIIVKNGHVIVLDVTVVVEVTGAT FINAYQTKINKYNAINVQIEQMFNCVNGELHGLVIGSRGSIHHSQHLI WHQMGFSSIELKYVAIGCMEDSLRIMSTFSKAIT (SEQ ID NO: 1393)</p>
R2	R2OI	LC349 444	Oryzias latipes	<p>CGACAGGGGACACAGAG CCTGCCCAAGTACCGCTCC CGAGGAGCGGGAACG</p>	<p>GGGGGACAGCTGGGAGT CTCGCATGATTACAAAT CTTGCGCTGCACTCGGAT</p>	<p>MGTDTVVVGQDYPGSLSKRVPARLVAGPMLRERSCHAHVFRAGHM WNWRTSLPSGRWDQPALEKSRVLTRSVATATDPEITSYPGKSVSTST QVQEEEDWCSRESGWISPGLAPEEPSVSEITASMVATMRVATEEVV</p>

				<p>GGGGGTGACTATCCCT GGGTCCGGCAGAGCG CTGGTCTACGGACCGG GTGGCTGTGGCAGGCTG CTCCTCAGGCCAGTTGATT AGTTACGATGGCTGTA CCTCCAGTGGTCCCGTG GTAAAGACTTGTGGGTA AATCAGCCGCCACCATC TGGGATATGGTTGACCGT CTAACCCAGTACTCAGGT CACAAACAA (SEQ ID NO: 1149)</p>	<p>GTGCTCCCGTGACGGAC ACATTAATCCGGAAGCG AGTGGTACTCGCCTCAA G (SEQ ID NO: 1272)</p>	<p>LEPQEQVVTILPEHGRNVPPGLAEQDTASPIEVSLLPDLAENCPLC GVPSGGLRLGKHFHVRHAGVPVTECRKCAWRSPNSHSISCHVPKC RGRMPSPGDPGACDLCEARFATEVGVQAQHRHVPVWVWVRL ERRARGGGIATKLWSVAEVEVETLIREHGDGATYQLIADELGRG KTAEQVRSKKRLRIDTASNPDDAEVEERLESIAVRSSSRPPSLVA TRREAVARGESEGEIRAIALIRDQNPCLISASDIISKLGRV DGPKRPPVREQTQKGVWRRLARRKREYEAQYLYSRDQARLAA QILDGAASQECALPDQVYGAFREKWEVTVGQFHGLGEFTGARAD NWEFSPILAAEVKENLMRMANGTAPGPDRISSKALLDWDPRGEQL ARLYTTWLIGGVIPRVFKECRTKLLPKSSDPVELQDIGWRPVTIGSM VTRLFSRLTLMRLTRACPINPRQGFALSSSGCAENLLIFDEIVRRSRD GGPLAVFVDFARAFDSISHEHILCVLEEGGLDRHVIGLIRNSVYDCVT RVGCEGMPPIQMKVGVKQGDPMSPLLFNAMDPLIHKLETAGT GLKWGDLIATLAFADDLVLVSDSEGMGRSLGILEKFCQLTGLRVQP RKCHGFFMDKGVVNGCGTWEICGSPHIMIPGESVRYLGVQVGP RGVMEPLIPTVHTWIERISEAPLKPSQRMVLSNFALPRIYQADLG KVTVTKLAQIDGIVKAVKKWLHLSPTCNGLLYSNRDGGGLLLE RLIPSVTRKRIYRMSRSPDIWTRMTSHSVSKSDWEMLVWQAGGER GSAPVMGAVEAAPTDVERSPPDYDWRREENLAWSALRVQGVGAD QFRGDRSSSWIAEPASVGFQAQHWLAALRAGVYPTREFLARGK EKSGAACRRCPARLESCHILGQCFVQANRIARHNKVCVLLATEAER FGWTVIREFRLEDAAGGLKIPDLVCKKADTVLVDVTVRYEMDGETLK RAASEKVHYLPVGGQITDKVGGRCFKVMGFPVGARGKWPPASNNT VLAELGVPAGRMRTFARLVSRRTLLYSLLDILDFMREPAGRGTRVALI PAATGAAN (SEQ ID NO: 1394)</p>
R2	R2_LP	Af015 814	Limulus polyph emus	<p>TGGGAGGAGACCCAACT ATCCTAGGATGGGCGGA ACCGACCATATGAGCCATA TTAACATTGCCACACTAT CCTCTGGAGGTACCTCTC GTGGTACGGCTGGATATA GGTAAATCCTGTAAACCAA TCCTCCAACCCGTGAAGGA GAACACTAAACCCATATA GTGGCCTCGCAACCACTA TATGTCCAACGGCAGGAG AAGCTATCTCCGGATGG GAAGGAAACCCCTAAACC</p>	<p>ATTTGTCTCTTCCCAA TGATGTCTACTAGCAGC TGCCGAAGCTAGATAGAT TGAGGAATCTGCGTAATC TGTAATGATTACGCTCA TGGGCATCTATCGGTAGC GTGCAACCTGACGTTAAA TTGGGTAATAAGAAATAT CGA (SEQ ID NO: 1273)</p>	<p>GIDGYMFGYARASGTSVSIQSSMTEGETNERATPRASDSSSVSIQS SCVTEGECLPPTDNCNPSVENQLPCVTEGRFERVGSLVTVLRFKVA CDLCSKEFLTVSKFAVHOANFNHSETQACCTYCGKSDGNHHSIACHV PKCPWRRVTVAANLSNCLDCLNDSFKTSGLSQHKKHHPKCSRNA ERILSLGVRTPSARPRQVWVSEETRTLREVEVYSGQKNINVLGAGH LPGKTSKQVSDKRRDLHRISSNVHGTPTTQSRGDPVEQVEEELD WEGMHPFPDPDSKFCYSLDQLDQKGLTEPVWQEIIEIVAEWVEN LAHVQSSWNHERTTKQVPENNTPARPPFKRLHRVERKFRQRMV DLQRKLAEEILDGREAVTCNLKKEIKDHYDQVYGVNSDRVSLDDCP RPPGANNNTDLLKPTFTPEVMSLQGMKNGAPGDKITLPLQLKRLK NGIHVSLANVFNLWQFSGRIPCECMKSNRSVLPKGSNLRDVRNWR PITISSIVLRILYTRILARRLERAVQINPRQGFVPOAGCRDNIFLLQSAM RRAKRGTLALGLLDLSKAFDVTGKHLLTSLEFVAVHPHFVRIEDM</p>

NeS L	NeSL- 1	Z8205 8	Caenor habditis elegans	<p>GTGATGGGAACCTACCGG CCCATCAGCTATTGGGTA CCGGTAGGACTTGCAA CCTACCTGTATTGCA TTATAGGGAACCGTGG GCCTATATCAGAGTAGAC CGTTATTAAATATGGGTG AAAATATTAAACAGTAAAA GCTATGGTTGGCGTCCGT GTGGTCCAGGGCGGCG GCCAAACCGAGCTACTTG GCACCAACTGGGGATGGT AGCTCCGAGCGATTCCCT GCGACGTGGGACCGATC GACGATGGAGTCCAAACA TCCGGAATAGAGGAATTG AGAAATACCTATTCCACCA CCGGCTACATACCCAAG GTGAACCGGTGCAACTA GAGTACAACCTATCTGT GCGGTAGGTGCCGAACA CTCAGGTGACGGGCTTGT TTATTGATGTCTCCCTACG AGACACGAATTGTGACAA ATCCACTCCGGTGGACAAT TACCCGATCTATGAACCTG TTACCGATATTAGACAAGA AAATAAGAACTGACAAC GCCTAGAGCTTCAGGCAG CATGCTGTAAATATCCAG TCATCGAGCGTGACTGAG GCGGAAATTGATAATAAC TCTGAAACTGA (SEQ. ID NO.: 1150)</p>	<p>CCTCCAGGGCAGCGCGC ACGCCAAAAGTCTCTGGCA TAACTCTGCAAAATAACAT CAAACGTCATCAACTCC</p>	<p>YSGCSTFRVGSQSTRPVLIRGVKQGDPMSPILFNIALDPLLRQLEE ESRGFMFREGQAPVSSLAYADDMALLAKDHASLQSMGLTVDKFCSG NGLGLNIAKSAGLLIRGANKTFTVNDPCSWLVNGETLPMIGPEQTYR YLGASICPWTGINS GPVKPTLEKWI/ANITESPLKPHQRVDILCKYALPR LFYQLELGTILNFKEKELDSMV/KQAVKRWCHLPACTADGLLYSRHRD GGLAVVKLESILVPCLIKIKTNLRVHSTDPISSLAESDGLVGAIEGIAQK AGLPIPTDQSRGTYHSNWRDMERRSWERLALHGQGVVELFKGSRSA NHWLP RPVGMKPHHWVKCLAMRANVYPTKRGLSRGNLSKNKDSA KCRGCTSMRETLCHLSGQCPKLSMRIRRHKNICEHLIAEASFKGWK VLQEPTLVTDNGERRRPDLIFHRDDKAVVVDVTVRYEISKDTLREAYA SKVRRYGCLTEQIKDLTGATSWFHGPFMGARGAWFPESDVMADL NIRSKEYEEFLCRRTILYTLDLLWKSNNQYLERLAP (SEQ. ID NO.: 1395)</p>	<p>MLRRKGRHRMVMVNSVKWQPSAHAEI/GTGKSWAPQRSQASEH GWQSNAMFDPNRIILFARDWS/LNQSTHLQNQRSGSLGIRPGQV RNNMVGGGPHRAGDPKRRVELVSIQGSSEVTVRTYPSDEIFCYSKSC DIKTAGYGPEDLKLTRHIKNEHGLKARWAYQCGCLNEKSDPSVSE</p>
----------	------------	------------	-------------------------------	--	---	---	--

TCCCTCTTCTGGGTAATT CTTCGGCGTCCGGGTC TCTCCCTCGTCTTTTTTA AACTTTTCTTCTCATCCAC TCTTTTGCTCCTTTTACTA ACTCTTGACTCTATAGTC TTTTCTATCCCCATCCGC CGTTGGGCAAGTTTATTT ACTTTGTTAAATCCATATT TATCTCTCTACCCGTCAC GAAAGCGTCTCTCTTCTCAA ACGCTTTTCTGACTTTTTC TTATATTTTCAATTACATAT TTTTCTGTTTATACTAACC TAACCTCCATTGTCAATTA CTAACTAACTTGACAACG GATTTCG (SEQ ID NO: 1151)	ACAACTCTCCACTCTCTT CAAGTCTTCTCGTGCTT CCAAACACCAATGGTGA AAGTCTCTTCACTTTTCC CTCCAAATCTCTCCCAT GTGGGAAGTCTCTTCT TGTAAGCTCTCCGAGGC TGCAAGAGCAGAAAGAAA TTCTCTTCTGACAAGGT CAGAAGGAAGTCTGTTC TTGAGCGTCCATCCCGG GCGTCATAGAGAGATC AGATGCACCTTCTAGCAG GAGCTAGAAGGCTGCC CTGTCTGAGATCCCCAC GGGGTCAATAGACGGG AGGGCTGCTGGCTTCT CTTTTAAAGAGGAAGCAC CAATCCGGAGATCCTTAG GGGTCAAGGATTAAAA GGCAGCAGGTCCAATTCT CCTCACTGACTTCGGTCA GAGAGAGTCCGCTT GGAGACCTCCCGGGGA GGTTGCTGAAGAGCGG AAGCTCTTCTAGCAAGA GCTAGAGGGAGTTCCA GTCTGAAACCTTGCGG TTGATGATGGAATGGA GAGTACTTCGGTACTGCT CGTTGCTCTCTGCGTTT TACTGCCGAGGGCCGGA TTTGCTCGAATCGGAAA GGTCTCAATCGACCATC AAGATGACGGCTTATCTA AGGTCCGAAAGCAGTTG GGAGAGTAACGTGTTCTC CTACCTTCAAGTTGAAT	GHKWM EAHM VAVHQSSAEKRIKSYQKCTGARVAEQLOAAAPSLTV PGKHKSGRDAKDSMTPTKDDDPKTRIVQTRSVVKKSTQKTAEPD EGSRGPKYAFIQSVKARKSALLCELSSPKMNPPTNLTKEGNS RELAKEEAPSEGIDDIIDLDESEESP RRKRFTWCLDHESREAWL DDTAIFWYSILCRGSTKYSALDPCLWSMYVKGSRYLDRLESSITYFF PICEEDHWTLVLKDNYSYANSLHQEP RPVPDRDFINDSKRARKEFKV QVPLQRDSFNGCVHICLMTNSMAGGKWHSEEDVRNFRRLKKTLCQ EEGELYSVNSLIGIPFQAPTEQMDYKETRCKRSVASVLTQSPPAKRP DCKPDNNIFVPTKCAAEAGNPEQEKGRNEESP EINT EHVIVAGKPANN ISPRCRSTSEMLFEMVKATTSGRSSSLGTMQDEFIRTSIAEAVPLM SIKLPMELPRKILPPIPRKPTQTNGGQKGQKQRPVTKGPDTLNAKV RNWNNQLESYAMEGRSFQRLWLT EVLTSIQKAAAGDEGIVDIIC KRNPLEVAKGEMCTQ TENKRKTNNAAARIADPIQSSKGAGDVKAS YWKERARTYNRIGSKEELCKPIDQLEDFKKSTSRNVQESIMKEKSS KIPALKIGNWMEKFFIGKEVAFALRKTDTAQGADGLRYHHLQWFD PSGELLAKVYNECORHRKIPKHWEAEETILLFKNGDQSKPENWWRPISL MPVIVLYSSLWNRRIRAVPNVLSKQCRGFQEREGCNESLAILRTAID VAKGRRNLAVAWLDLTNAFGSIPHELIEYALTAYGFPQM VVDVVK DMYQGASMRVKNATEKSDRIPIMSGVKQGDPI SPTLFNICLETVIRR HLESANGHQCLKTRIKVLAFADDMAILT DSPDQLQRELSKLDNDCTPL NLIFPAK CASLVIQGVRSASIKLGNAIRCIDENTTYKYLGVQTGS AARISAMDLEKVTKELEVVKSDLT PPQKLDCLKTFTLSKLTVMYGN S IPLITEIKMFANIVIRGVKVMHRIPVRGSPLEYIHL PVKGGGLGVACPK TTCMITFLVSTLKLWSDDEYIKTLFTSLAEVVVKESKSTVTMDDIA DYLNV EERINRSEFGYNSITRLRDMVRNLAITGDSPLYRLKMMVKNKG IALLVQATSESMERIYTEEDAKLQRLKQDVNKALKHFRFNTTKVVK S KVVRVQQHPASNRFTVKGGNLSLACHRFVHKARLNLACNWNNDY KSKSKVCRRCGKDL ETQWHILQNCPFGFSKKITERHDAVLHKVKTUE SGGKKNWTKIDEELPGFSRLRPDCLKSPDEKQJILADVACPYEHGV EAMERSWQAKIDKYETGAHLRKSGLTKLTVLPIIGSLGSWWKPTGD SLKELGKGSVINSAPICLVCATVLEHSKNTYWNHIFGEAYIPNPMRNG HAKPAGNGWKKERLQKAPVRPTN (SEQ ID NO: 1396)
--	---	---

GGTCGTTTACTGTTGG
GATAGCTGACTTGATGCT
AGTACGCTCATCTGTGG
ATGACGCTCCCAAGCAG
TCAAGTAGACTTGAAAG
GTCCCTCGCCTTAGTTA
GCTCTTAGACCTTAGGG
TCGCCATGTTGTGGACG
GGTATGCTTGCCGGAGC
CGAGTCGTGTTCTTAGA
ACCAACTCGACGAGGC
GAAAGCTTGCACAAGTTA
GCACAATTGTTGGTAGGG
CCCTAGAGAAATGAGTC
CGTAGGGGGTTACGCCCT
TGGCGAAAGTAGGACCA
ATTGGCAITGACGGGTG
CTTCGGCACTAGGCCAAA
GGCGCCACCACACTGTCC
AATCTCTAAAAGTTACAC
ATTTCATCGAAGAACTACC
GGAACCAACCACACATGT
GTTGAAACCCTACACGGTG
GAAGGGAAGCGAAAGCT
TCGTGGAACGAAAGAAGA
ACGGTAGGTTCCCTTC
TTGATGGCTGTGAGGCTT
AGGATGACGGGGAAGGC
CGTAGGCCCTAGGCGG
GTAACCTGGCCACACGCT
AGTTGATCTTCGGATCAC
GACAGCCCTGGCTAAGA
GGAACTCGTGTGAGT
GTGAAGATGGCGGGT
AGGGGTTAAGCCGTGT
GACAGACCAACCGACTGC
AGTACAAAATTCAGTGAT
TATGCGGGTGGACCAAT

					CTGTTGGGGGTGTTCC CTACTGACCCGCAA TATGGTATGACGATCCT CGGATCTAAATTCATAA TGGCCACCACAACCATTA AACCTCCTAGCAGCTGG TGGTCCCGATAATTCGGG TTCTGCCACTACTGCGA CCGAGGCTCGCC (SEQ. ID NO: 1274)				MSLQRAKNARGDPGRNLCADYRDLKDLNKKQHSHTFFVPSDLRG SSLVACRCGTPCSAGTGLSRHQSRVCGLTAPRIIRNRNVGNSTNTSR PPSNTAASPVPSPSPERPSPQPAEVVASLEPLESEAEVLEVAQVDAE TVDTLEGTRAPESVPRSAEEGSTRVRELNMTPAPEEHRGEESHT NPTAPAGLENVAVSSTLPSPGTLPSLLPSQECANERFLYLAHLVPRSKP LPNNLVTFMDAAERCALAYIAQPSDSTLLAFALPKVGLTQALAPEQ PLRPSTFLKQPHIPWPEQPAPARRPPSNIRPDITTKQVIKLVENGRIGA AERVLEEDASVAELDQGVIDQLITKHPKGPSCFPGNAVGPPTGKAPDI DTIQKALDSFKPDTAPGVSGWSVPLLKTAAREPVKQFLQLLCAAIAN NTAPGRSMRLTSRLIPLKDDGSIRPIAVGELIYRLCAKALISHFQPDFL LPFQLGVKSIGGVEPIVRLTERVLEGSAGAEFSFLASLDASNAFNRVDR AEMAAAVKTHAPTLWRTCKWAYGDSDDLVCBGDKILQSSQGVRRQGD PFGPLFFITLRLPTLNALSQLSPSTQALAYLDDIYLFNSDSQVLSKTTQ FLADKQHIKLINEKKCLISFDEIRQEGFKMLGTMVGGKEKRAEFLEG RIRKEMAKVGKLDLPHQHALLLRFCIQQLRLHQLRSRSDDLVDL WERLDTMLWEEVKRMRMQREDTAEFEALGRSLTKLIPARLGGGL LSFKDVAPLAYRSAAEASDTLLDNLGLSSPEEPPTPIQRTCAELWE SQQEAILHNLGDTERRKRLTENASRLGRSWLSVPIYQLRLSNVEIASG LHDLTLVSSIPVCRFCGSDSPGHDELCLARNPWTQRRHNAINRVI YQHLKQIQGATVEIEPHTLGQQRNDLVRGSSALAFDYDLKVYSLG DRDARSTVTPCAPNGKLADFCCLDRVNWLDKVGVVSKNAPKVTG GVFKPIILSTGGLMSRSTADEWKDWRDAMPVGGFEKMEKRIGVELV KARARTLV (SEQ. ID NO: 1397)
CRE	Cnl1	.	Cryptococcus neoformans	CCCTCTTAATACCCCATAA CACATAACAACCCCTAAT CAAGTTCTCTGCACCTTA AACACCAACAAC (SEQ. ID NO: 1152)	TGAGGAAGAGGAGGTG GATTATTTTTTCTTTCTTT AATAAGTTGTTTATTAA GTAGTTTCTTTCAATCGG GCAACCCACACGACAAC CAATAAATTAAACAACGA AAAATGCAACCTCTATAA CCC (SEQ. ID NO: 1275)				MSQPNISSAETPLSQLPTVPPTPSPNSLSLPTVRDLLCPRSSHVY SSIPSSCLHSFTMLLIKTVRAASATMTPTESHRAFIHLHILPIAVLRFSFR GETGWSRSTGQHIALRQIRRRASSGRHWAALWHEALAAHQVDLD YRTHSRRYQASATSRHRIGRAMRLAADAQGRAMSALKAKPLPDL HAAATRDLTALHPPASPVPQPLSPTDLPVPVPEITEGQVLRARALNP
CRE	CRE-12_CCr	.	Chondrus crispus	ACGCCCCCTATCCATTCT GCCAGCCTCCCATCGGCTC GCCGTCTCCGCAACCCCTC TTCCTCGGCTGTACCAGTT CCGCTCCACAACTCCCT	TAAGTCTTGACGCCTGC CCCGTGATACAGCATCGG TACCCCTAGCATTTGAAT AAAAAA (SEQ. ID NO: 1276)				

CRE	CRE-13_cri	.	Chondrus crispus	CNCCAGCCAMCGATCCCG CCGCCACTCGCMGCCCGG CCGTCTCGACCGCCACCTC CCCGASGCCCGCAGCCCATC (SEQ ID NO: 1154)	TAATTCACCTTCATATCTG CTAGTGTCTCTGTAAGCG CACCCCTCATGCATTGAT AAAATTACCCCCCA (SEQ ID NO: 1277)	CGCCACA (SEQ ID NO: 1153)		<p>TSAAGPDHLSRILQLLARTTISPEAGVTGLSALTNLVRLRARGDIPDR TAPLAAATLIQPRPHKIRPIAVGOALRRVTVKVLLPPIAQDTRDHL PEQLANSVSGMDAIVHDTRMLMHRHGRNPDYIMVSVDAFNAFN TFSRQSLDLRLQTPSLARFLNLIYGRTPDLVLPSSPRELMIKSEQGT QQGDPASMLLSLAIQPLRLLTRECRDLNRWYADDGTLVGPSEVI KALRIIIRDDGPGQGFHNINCKRAYWPTVMPEKISELLIFLHVECG EGGVALLGAPLGTDFAVRRLHLMNKVQSCHASLSLLEIPDARTRFLH HRVTGSVKVEHVFRLLTPPHLSLPAATKDEQQIAAYSRINDVAVSTS MATQIGLPRFLGGHGFPLSPFIHASVAAASLIEAAPVRVKGPHNPSEF YRRMARRHIVHVLGALNPEVTRGILGTHSPGLGPEPEALLSRPERVH HTLIQAMQGATSRLYWEHTAWDLDPLPRNHSAASVRRRARNLSLR APGAASFLCSHPSTSRVPSAVWSCMLRRHLDTPVYCDIRPLICSHC CKPMIDARGDHAAICRHGFGVVRHNTVRNLLARHAFRAAGLCCDL EVPSLLPNTANRPADILVQPAPPPSGALPDRPTAYDVTVRSPYCRSTM SLAAKGLAGAAEAADLDKLRVHSRTVDAFHLQDPDPLLDWHFVP LAFDTLGATSSRTMAVLEYLAHRIANRTYSSYGTAKIRLLQRISFAVWS SLASATLSRMPYHGAALSQAQV (SEQ ID NO: 1398)</p> <p>MAXXPISPPGAPAPLRYMLQCPPPLPKXXXVPVPHMSSPRXRL PHRXMRGPPSTPPRDMHMRPHGTGPHSHRXCGRPXHCTHASXQ PRAXHXLQXPLRSPPHPHVSPLLCXGLPTMTQPRMKRALSXS AKAPPTKRPSASQGAASSHDXRTPPPXPPRPPVFPPTLDQHXF ALSXAYPHXPRRPPSPRXLRHSFPFRFGXQTFESSIPGRLHSTVLLI RLVRAATAANTPETTLXSCFTCSRLPFFERPSAXLAGGPRAVNFM LSACYGERVXDESGSYXXHCITSHPRPRLRYSRQHSRNHHTXFL HLHLLPTAXLREAFRGEXGWRSSRQLHALRLXIRRACTGREWGLX XEALDAHSXRTWQHTHARRSPVSPSARAARMLASQAQYGR AMRTFTNPPLADLNDPATMERLQALHPTPTVPVPLPPSAQPRPE VTXEAVXRAVRRLNPNSAAGPDRMSPKLLHLLAHTPISPEAGVTGLS ALTNLVSRLAGSLPPCTIPLASAAATLLPQPRPGKIRPIAIGQALRLV TKXLLPAIDDCRDHLAPEQXANGIPNGIDAIVHDARMLVRRHNDP HYXMVSIDASNAFNFSRQQVLDLQPLTRAPSLRYLDMVYARAPSL VLPSXPTILHSRXGSGQDPASMLLSLALQPLTLISRECLXMINR WYADDGTIGRIDEVXXKALDIITKEGPRFQFELNPSKTRVFWPSRQXD LLSPLMTVGPLRVIDEQVLLGAPIGSPSMAQYIREXLTCKTALA HLDHIPEARMRFLHHRVSASACRLQLHRLVPPDFAXPFAQQFDRD QLXAYXRFNSVTMSPRIVPKYGCXFTXATASPHWHLPYTPXTLLASSI PLQHGKVPFTFPPSLISVLHEARCGSFFEYLSXNPHTSRCDYVAKN RAQIRLXFSGGGHGLTSLASTIHASYAASLIDTAPARLQGPFFAVSQY</p>
-----	------------	---	------------------	--	--	---------------------------	--	---

CRE	CRE- 1_Ac s	.	Acanth amoeb a castella nii	TAACCTAACCTCTCCCT CGGCCCTCTACCTAAAG CGCCTAATCGACCGCG ACGCCCTAATCGCTACCT CTACGCCCTAATCGACTT GGGCCAAAGCGACTTC CCCGCCGATTTCTCTCT GCCTTTTCTTTCTCTCA AGCAGCGCCTTTTACTT TGCCGCCGTTCTTTTTTC TTTTCTTTGCACTTCGCT TCTACTTCACACTCTCTCT CCTCTCTCTGACCGCGC GGCCTCGAGCGACTTGCT GCAGGGCTCCCGGCTC CCCCACGGCCTGCTACT CCCGCTTTCTAGACGCCC CGGCTCTGCTCTCAGTCTC CGCATCGAAGCGGTAGT CGGGTACGTGCTCAAGT GACTAAGCCTCTTTTCAG CCTCGGCGCTCTCTCAATC CGCCTCAGTCTTAGCCTTT CAAGTTGCTCGATTACGCT CTCGAATCGCTCTCTCT CAGTCTCAGTCTCAGTCTC AATCTCGATCTTGCCTTCG CCTTCGCTCGACGCGCTTC CTCTCGWAATCGCTGCGCA CTACGTGCCAGCTTTTTCG	TAAGCCGCGCAGCAGG ACGCCAGGACGACCCAG GACGACGCGCAGCGCGA CCACCTAGCACCGCACGC GCCAGACATATGTCGC GCGCTGACAGCGGCT AGGTGAGCCGACGCGA CCGTTCTGAGCCTCAGTC GGCTTGAGCCCCGGCTT CCCAAGCCTACCGGG CGGCTCTTTTCGCCCTG GTTTTCGCCGCTGTTTT TTCCTCCCCCTTTTCCCC CCTTTTCACTTTGACTTA GTTTTCTCTCGCGCGC GCAGCTGTTGCCGCGCA TAGTGTAAATATGTTTAA AAAACGTGTAATAATA ACTGTTTAAACCCTAACCC TAACCCCTAA (SEQ ID NO: 1278)	QRFARGPLRVVLRNLPFVQPAHFMSMTEXDLGCLCPXALLARPERIHT FLLOAQYSAASSWQXPLWESFPNPGDHSASLRKRVYNSLLAP GATSELTAPHAATSRVHNATWSTMLRRHLDAVNTDSPLRCXCHCS KPMIDARGDHAXIXSHGFGTTLHRHNTVRNVLRAQLFRVAGLAYSLEV PFLIPNTAARPADILVQPPAPGLPDPXPTAVDVVTICSPFERRGMLYH AARHGGAAADAASVRKXKALERTIRXALLIEDDNPPPLDWHFQPLS FDALGAPSQSTVHVIEDHAKLMALRNSCTIATAKSRIQQRLSFAIWSS AAAAILSRPLTHAADISYPIEV (SEQ ID NO: 1399) MATTTISRPSSSSSSSARSASASTSASVASIPRLFRDGRFHCHLAHC QTRTSTWQDLSAHLTRMHDGDVPRDVAAACGIVQCLHEGCRKWF RGAAGLASHRGKARHAPPAPRAALAAVAAPRADSRGRTAPTSPV APPXAGPPRAAPRAAPSPLPCPPALPHPPPSASPTSSVTSPSPPTT PPSQSPDLFSGFANAPTTPSPPTXSPSPAGSPIAARRFVLPVATPY PAPAPRANRPKLSVARFPVKARAGAIPEASSVTPQDRAVSRRD AAAAPSSAPGLGLADEHEDDDTYGGDTIALTAPHAPRETRAPEFEA CFLEEEAPATAGDLPYARAFACPSARLQEIIPRLKSAWQAAKTIA EALDCHTAGDTQGYNAHLRLFIELPARGLAVPTNCRGAARTKLQRE RLDIAAGRIPIAPDPPCDAPGADALRGFPVSGTTAGDVSNDDSG GVHDPAAATASARQAKRLVEQGLSSRALRALERGEPAVASADTLGRL EALHPNPTRDGLWPGAPKAIIPRVTAHLAQVAKELPRGSAPGS GWTFFELVQAAIDRQPTGTVAFFLIDMAQRLRGLTHWRGLLTASRL VALKPDGGVRPIAVGEALYRIGRLV/LKADRVMSADATQVYGRH QYGVAYPGGVEAPVHARELHDSGQLRAVSLDWRNFAFNSLDRVH TALLIADRAPALARLYEWSYREDSVLVLPRAFEGAGLPASLLSQAGVR QGDV/LGPLFFAIGAAPVLEIDAIPYVTPRAYLDDIFVTIPHGVTDAAT KAAVAATFATAEREGAAAGLNRCKSAVMAADAELPPHAAAGAR EDVESCAPVREGKILGAPVGPSPAFVAKSLDGIIRAIKGLDILVADAE PLQHLVL/LRQCVAQIPTFWARVPDAGPALAVWDTALLRRTGALV GLDVRGSLQADIALRPLVRLGGLRSMKMDTAPRAFFVASILFAAALA NTRRSELTCASATARRLRAALPELARTDACNDDEAAWRRSIARGVFPD VDKLTGTLQRLV/LQGMADSKSAHRTRRQVPLFAAVFEEDAATPGSG AWLAAIPSDPTLVLPDAELAEAVRIKLLTTTANAAGVCPACHKTGIDP SHAYTCVLSHLRTHADVWVRVRELACKTEKPVREHVLAIIPVAPT NNNGDEGDSPTTADNADGHAVATKRRPETRASARAAAAAT AAAAAIINDNSLSDDDDDDDHDDNCHGEERGEGERNVTCPGHYT ATPFAADDTLNSDEDNEDNAHEDDDDEGKDDNDDDVVNNCNSS SSDGEggDDLDYEYSDQSQSVTRSDAAATGESPNPERPTTPTALLRA DLWLPTSTAVDMVMAACRRRAKAFDRAVSRKAAYGPAVADG
-----	-------------------	---	---	--	--	---

CRE	Cre- 1_Fcy	.			TGCTTGCTTCTCGTGCGA CCCGACCGTTTGCAAGC CCTCGCTTCGTACCCGC TCTCGTAGCCGTTCTCATC GCTGAAGCGTTCTACGCG CTGGCAGCAAGCCTCGGC CCTAGCTTGAGCGCCGCC GGTGGCCGCTCGCCAAC (SEQ ID NO: 1155)			SIKWPFWSPFGVLSRPAKFLKRAMGDTTAAKQAKARLRAVAA VRGTARLSYAWGACAALVGGN (SEQ ID NO: 1400)
					ATCAATCTAATACTGAAGG CAATACCAAACTCAACCCG AAATCAAAATCGTTAGAAT CAATATACGACCCCGCTG CTGTACATGTCACGCGG ATCTCGTTGTAAGGAAGAT TGCAGCTGTAAAAAAGTT GGACTTCTTTGTTCTTCT GTGAAGAAGTTGATTGTG GCTGTTCAAATCTTTTCAT AATAAAGAATTA (SEQ ID NO: 1156)	TAGCACCACTATCTATTC ATATCCACACTGACCA CCTCCACCTTCACAACTCC ACTCTCAATCCCTGACT ACTAAGAAATATTTTCATG GTGTTACATTGGAGGT ATCCCAACACCAAGCACA AAATGAACCACTAAACC TCTCATCTATCCACGGG GACCACCTTGAGCAGAA CACCATTATATTACAAC CTTAGCTAGATTAAAGAT AATTATTTAGTACATATT TATACTATTAACAAAAA AAAAAA (SEQ ID NO: 1279)		MAPLWNAATSSPPSPVPLTNDKKDSTLPTATSKNLSKNNNNKNN NTNRINIKNDNTDGSNKINLKPAAVKITNPYKKNKKKKNN AGKSNPTNQNPSSPLSDNDDDTSSNITINRRLKFGTDDLAPPN PPSNTNTIGTATAATAATATTTATAATATTTTTTTTTNTTGD NLASNNNNNNNSGNSNNNTNNNTDNGNSNNRPPRVTV DPRSDLPGAELSAANKMLDEVYGDHVDNPGSHLSGLSSSQDLW QGYFRLIPHNQSLYDCPKGLGKDITNEYSNLFAMNGKCNMEKL LVFPVVLQRRHGVTKNADVKRRLSLTAWKEGKYLVEDTHRDL IAKOSKARGDTTPAHRKAYSSKLMRGHLQSAVNYTDRGGGILYP YDVDEKSGHTVSRVLQDKHPSMRDPGPTAMPAYESVPELPLEITAD TVEIVAGLSGGAGLSGVDSIQLKLLHHGQASQRLRNVCAGFRW LANEHPWASRAMLANRLIALDKMPGIRPVGIGDTWRRFFAKVL AVMSYATDCGSDQLCAGLRAGVDGAIHGLSAMWREMESEENT GFVLIDADNAFNEVSRINMLWTIRHEWPAGARFAFNCYRHHSLV RNPGGKPTFFSKEGVTQGDPFAMIAYGVALLPLIRKLKELNVLVQS WYADDASAAAGKFEILRLFQDLRMGPDGYFPNASKSILITHPDNV VAAHHFFNETHGLGFKISTGSRFLGGFIGDTTSRDEYVSTKIADWIHG TKELAAVARLKYPHAAVTGITKCLQHKWSFTQRVIPGIDDLFQPLEDE LTNNLLPALFGDPPSTMDKRLTLALPVKHAGLALPNPVTSSATNYK NSTLMSSHLLAVQKINFSIQDHRDTCOSSLSASRELQRTENDSSLT NLLAALPAAAGQPSPTTIRAKRAGETGLWLTTPNHINGNILGCGDEFID AIRLYQKVPHNLPKACDGCASFVGHALQCKSGGLIIRRHDELNLE LASLAKMALRESAIRAEINPSASIMDSPTTITAITDNGDRGLLIKGF WDNGMDAIDVITDIDAKSYRTDRPKVLOSQEKEKKKKYLDQCLL QRRFTFPVVSVVDGLIGYEASVNLKLSKRLADKWNKPYSVTCGIVRS RSIACARASNQCLRSRIPFKTMSRQIQWEDGAGAGLYRIVR (SEQ ID NO: 1401)

CRE	Cre- 1_HM	.	Hydra vulgaris	<p>TTTCTAATGTTACGTGATA TGATATGGTTAGTTCATGG TTAGTTTATGTTTATGCTTA GTTATGGAAAAATCGTTTA TTTATGGCACAAATATGTT TGCTGTTTTAAATTTATGT AACGTGTCATTTGATGTA TATCTTGAACCTTTTAAATC TGAATTTTACTTGGTTTA ATACGTTTATATATCTTC GATTGAGCAATTTATCCTA TCAAAGCAATTTATCCTTC GATTCGAGCAATTTATCCT TCGATTCGAGCAATTTATC CTTCGATTGAGCAATTTAT CCTATCAAAATTTAGCATAT ATACTGCAATTTTCAAATA ATCTACGAAATAAGTTCAC TTACTGAAATCATTAAGT AAAAAGAGAAAGGAAGA AAAAATAAAATAAAAG TAGTAAATCCTTTCAATAC AATAATCATTTCTATTATA AATTTAAAGGAATTTTG GTTTGTACTAAATCATGC GTTTATATTTACCGAAGA AGGGGCTGCTATATTTT GTTTGAAGTTGTTATCTT AAAACTTTAACTTGTT CAACCAACGTAACATTA GTTCTGCTGTTCTGCTCAAT TATCTACAATATAAAATTT ATCAATCTTTTTCGTTACG GTAAACAATAAACAAATA AATAACTATAGTTATTTTA TTGTTTACCGCATATTTGTT AACTATAGTTAAACAAAGT ATTTGTTTATGGAACATTA</p>	<p>TAACTTGATTTTTAAATT GTTTTATTAGTTT (SEQ. ID NO: 1280)</p>	<p>MMNVSICKRCDRSFTTLKGLNIHKGQCKIFVSNNTKQINNVMNNELT TPNKNKVEINTILNCDSEISVEHYSTNTPYLPKINICESIIDPNDYLWGH MPFSELLNHVNTIYDEIVFYHKNLFKVPSPGKGGMFIEELTFWLKQFN NRTKLNGIAMKCFMIVPSLMLQKPSIRSKAKHEAECLVRRITLWRNG NFSELMREIRYIQSKINTSKKRTTFEDSRIFAKLMMEGKVAALKVLD RESSGILQCSSEVLKELKSKHPDETVPQDNCLLYGPLQNTPECLFDSID EISIFNSALQTKSAGSPGMDADLYRRVLCSCFGPSCKTREEIATFT KNIATKSYQPDIVQPIACRLPLDKNPGIRPIGIGEVLRIVGKTSHHHC QKEIKEAAGPLQTCAGHGAGAEAAIHAMQKIFHQEDTDGVLIDAR NAFNCLNRSVALHNIQITCPILAMYLNTYRKPAKLIYGGETIFSKEGT TQGDPLAMPWYSLSTVTIINTLKLVPDVQVWLADDDATAAGKLQSL KKWYKCLEDEVGGLYGYVYNQSKCWLVKSDNQAEAEAKLIFGNSINITT QGKRHLGAALGSEAYKKVYCEDLVSKWSEKLNILCEIATTQPPAAYS AFIKGYRSKFTYFLRTIEAFENFVTPVEKILSEKLLPVLFGTDCSIIKENRD LLALNPSEGGGLICNLITEAKEQHTASKKITNLHIKSIDQSDVMKEKD DFGKTFSEIKTKTNMDSKKKKEEVKKIHAGLPENLKLVEQACDKGA SSWLNTLPIKEQLDLNKEEFKDALLRLRYNVPLANLPSCACGEKFE LHAMSKKKGGFVCNRRHDNRDLTLVCLNKVCTDVQAEPLIPLTNEK FNFKTANTNDEARLDIKAKGFWRKGETAFFDVRVTHVNSKSKKQPT KHIFRRHEDAKKREYLERVLEVEHGTFTPLIFGTNGGFGDECKRFTALL AQKLSKMGERYGAVINWLRLSLMEITRASLLCLRGSRTPFRRHYNT DDVGLENVQCGLI (SEQ. ID NO: 1402)</p>
-----	--------------	---	-------------------	---	--	--

CRE	CRE-1_LSa	.	Lactuca sativa	CCAGTATCTCTGTTAAGG TAAACAACAAACATAGA CGGCATCTCTTTTAAAGT AATTAAGTATACGGCTAAT AATAAAAATATACAGCTAA TAATAAAATCTTCA (SEQ ID NO: 1157)	TGAATATATTTATATAT TAAAAAAA (SEQ ID NO: 1281)	MASSTSSDCLCFRSHCCPNGEVSGKIGXRMISHIKRHHLLTEDR KCVLREALSSDVGLFMAVEETLKAFQWQWCMGKMTLHALSRYCHH PDGRVXFVTGADGSSRYVIGILKPSTKESVTNALGGLVFDVGLDRVF KEPITTVKSIPIHSCRLAFSQALKTALYKVIAQPGSVDAWICILLPRCTL QVFRPNRQECRSGNRKSLQSSILKSLDTWKGEDGIRKLQNMILD NPEVGAMQGGGILQKESTSSNTNIRQCLRKVADGHFTAAYKVLC SGVAPYNGDTIKALEDKHPRPPSPMPSPPISEPLVADFDCVFGCIKS FKGTSCGRDGLRAQHXLDAALCGEGSAIATDLIRAITSWNLWLAGRC PTILAEFVASAPLTPLIKPDNGIRPIAVGTWRRLVSKVAMKGVGKEM AKYLNDFQGVGVSGGAELVLSANRVLSEHHADGSLAMLTVDFSN AFNLVDRSALLHEVKRMCPISLWVNFYQGAARLYIGDQHIWSATG VQQGDPGLPLFALVHPLVHKIRDCKLLHAWYLDGTGTVIGDSEE VARVLNIIRVNGPGLGLELNIKTEIFWPCDGRKLRADLFTDGRPS LGVKLLGGAVSRDAGFISGLAMKRAVNAVDMGLLPQLCDPQSELL LRSCMIGIAKLFGLRTCPQVHIEEAALFFDKLRRSIEDMVVCGGPF GDIQWRLASLPREFGGLGLYSAYEVSSYAFVSAQSWALQDHLRDS GICGMSDYLCAMTRLRDTIPGDCSGFTNKDTPPKSQKALACALFS KIVKDMEVDFDMTVRQKAVFECLRAPHAQDFLLTIPIDGLGQHMSP VEYRTILRYRLMIPLFIDEICPCVKRACLDTFGEHAVHCRELPGFKYRH DVVRDLFDACRRAGISAKKEAPVNFITDPQDGRSTLRPADILVFGW VGGKHACVDLTGVSPLVGLRSGGFTAGHAALKAAACKVAKHENACI ENQHVFPFAFDTFGLAPEAVEVLLNRVQRMHNSNVISPRSTDVVF RISFAIQKGLAAQLVARLPISIDMY (SEQ ID NO: 1403)
CRE	Cre-1_MB	.	Monosiga brevicollis	CATCTTGGCGTGAACCAC GTTGTCAGACAAAATCTGC AACCCGCTCTTTTGGGCC CGCGTTTTGGCGGCCCT CGCTCCACCGCTGTCGCT CGCTTGCTCGCTTGTTCG CCGCGGAC (SEQ ID NO: 1159)	TAGGTAGGCACCGTCTCG GGGGTCCCTCTGTGGGG ATCCCTGTGTGCACCTGT CGCTCCCTAGGTGTTCC TCGTTGTGCTTTTGTG GCTTGACTTGATTTTGT TTAAATTTTGTCTTAATTT TTGCTGTATTTGTGTGTT	MATESGGEDSWTVRGAKRPSAESPPSNITTSQTHRSKHTKHG SARHDRNHVFPDPMITTLRPHARHSVPTARASHVPTSPSPAAGATE SSARAWPAAEPVTRTSNGGGEQHPHIGNTSNASRPTPTPSPSRSA QVAAMPAAATATSSAPMTEDLSASVPSENGEGEQSPSESTGQT HHIPNTSDFLTSSDESPPRSTALRAPTIAPPAHGDGDTNG SATPEPLVQSPTPAQMVLPVPSGTQQTSDPSPSPASPATILPAI SHPVHSEHANSAPLGEVSESETHNTAGEHSESEQDVLSDPAPPIAA NVLDAQRKVLKTSGHRQLLACPFGLCKCKGPRLDKAWVNVHLRE

CRE	CRE-2_HM a	.			ATTTTGCTGAATTTTGT AAGTCTTTGTATGATG TCTTTGTCTCTGGTGG GTTGTTTCTCAATCCG ACGTTGTGCTCTGTTGA TGTGAGCGTGCCGTGGT GTTCTTTGTGTTGCTGT GTGATGGCTTGAGTTGT GATGTGACTGCCTTTT TGGGTGCTTGTTGTTGA AATGGCGTATCTCTGGT TATACTGGTGGTTTTGT ACGATTTTGTTCTATGT GCGTGATTCTTCGCGCTT GTACTTCTGGCATGATA GAAGCCAAATGAATGTGC TTGTTCTCTGTTGTTT TGCGTGCCGTCGTGATT TGATGTCGGGGTTGCAC AGCTTTGCTTTCAGCTCT GAGGTTCAACACCTAAT TTA (SEQ ID NO: 1282)	HPYDEQATDLVKQVMEAKLVAQCNKCHLFFEAAGISQHRSGANIL KRATEALFHAAGHDLEIMRGAWPQQCVGSRISVCCELLKLARHPLIM QRSRYPSNATETKLMAATLSQLYWSAVHSDYTAEEEREMCWALIALP SMILLSAPSTALSTIDLRNMFHDLRLWLVTGQLGRVVVDAMRKAVARK QSRGQLNAGAGAHNPDAVDQSLRSILVRDPLDADEAWANHVNTNR LNRGQIAKAFDADKARAVIGNSEVQAVRDLLVPPGLTPYIASTPASTS TLAPATAVSSPTVSFTKGELPKALAAATKGVTDPYVSGELLASIRIKE HFSQVLGPRQGSTDPATPSGDGAPQGPPTATGGPQVALNKIFHHI ANNTVPESIRHALCSINYTILEKANGKFRPVGTDSIFNKVVRALLEQQ QPHIAHLLQASPELAVGVKDGISAAVGMAGFELQACESTPGWTMLS LDFKSAFNVTDRARLHEIVADKVPGLLRAFERHYARPTTHCIVDKFFKV IDIDVGQIVQGNELSPFFALYSCVGLDATTDYRCKVIKYLDIVL MGPAEDVAADVEIVKARASAGHLQPSKSRFYMPRHHSASITAIS VLPDAVRETANTGMTVLGTPIGRREWMKKQLNDKAKHAGKLNDM LTTGVSQALLTAMQVYPSLINHLTYLPPSLTSGLSELNRACKDTFVK AFFAKVNSAPAGAEHDTVLEQLLEARLTRANTGGFGLHDLVERG PVAYVNCMAKLATRYPRVYDRILLEDASRAADFEAHVQRAGFQMAT VKDAAATQRPAAEIIALRSKAAALDDLMAKCALDQQAAYLASREWGVSTV LTMGRDKLRLSDTTFAIAVWSMMGFGHELINVKPTDKCPLCSSKT PQPRLTREHLLTCRPIKRHNALRDEMGRLLRYATLSHVWVEKSGYNA NGQSCRIDLHCRNPFPGGALPALPDLDGIDVTVRTAQPPTTSQACIK VGAALRAAEKEKRDYTGTFNHHGKTLVPAAMTTTGGFASFFVDLLGQ LARCAEARGVYQPLDEAFVPRWKGRAALVHQMNADHIQRHFG GVCLRSS (SEQ ID NO: 1404)
		Hydra vulgaris		AATTTAAAAAATAAATC GTTTATTTATGGCATAATA CTGTTTGTAAATTTTGA ATTGTCGCAACAAGCTGAG TTAAATGAAGAGCTGAA ATTTAAGATCTGAGCTTTT CAATCAGAGTTTTTTACCC TAAACACATTAAATTTTATC ATAACAAAAATCGTTCTAA TATTATTAACCTTAAAGAA ATTGTTCTTATTAACAAT CTTATTTCAAGTGTTCACA GACGAAGGGTTTTACTAG ATTTTATTTTTCACATT	TGAGCTCTTATAAATTTA TATTATAGCATTTTGT TTA (SEQ ID NO: 1283)	

	MoTe R1	JQ747 487	Magna porthe oryzae	<p>TGAATTTGTTTATTATAAA ACTGTAAACTAGTGTGCAA CCAACCGTAAAAAATAGT TAGCTGTTACCCAAAATA TTATTCAGTATGAAAATAT TTAATCTCTTTATTTCGAG TAAACAATAAATATCTAG TTAAACAAAATATTCTTA ATAATAAAAAACAAAAAC TTTTCTTAACAAGTACA (SEQ ID NO: 1160)</p> <p>CCCGAACCCGAACCCAAA CCCAACCCCAACCCAAAC CCAAACCCAAACCCAAAC CAAAACCGGAGGTTCC AAGTCGCCTAAACCCGAA GGGTTAGGATATTATTC GTTATTAGAATTGGATAA TTATTTACCCCTGTTGGAC AGGGGGTTCAGGGGTT AAATTAAGGTTTTTATTA TTATGCGCGGTTATTGG TTTACCCCCCAAAATTA TAAAGCGGTTCCATCCT CTTAGGAAAAGCGAAGCT TTCTCTGTAAAGTCGT AGACTTTTACTATAAAAGT CGCTAGACTTTTATACCAA TCCTTTAACAAAAAGCGTA GCTTTTGTGCCAATCTA TTAAAAAAGCGGAGCTT TTTTTAACCTTTTCTTTTT TTTTTTTTCTTTTTTTTTT TTTTTTCTTTTTTTTTTT TTTTTTTATATATATTAT TATTATTATTAGCGGT GGGCTATTATGCGCTT AATTGTGCGGGGCTATTT</p>	<p>TAATAGGTAACGTCCTCTA TTTTTGCTTTGGTTTTGT TTTTATCTTTGTTTTGTTT TGTTTTCGTTTTTGTTT TGTTTCGTTTTTGTTTT TTTTTGTTTTGTTTTG TTTTGCTTTGTTTTGT TTTTATCTTTATTTTGTTT TTGTTTTACTTTTGTTTAA TTGTTTTATATTACCTT TTGATTTTTCTATTTTC CCACCTTATTATTATAAC CCCAACCTACTAATATTT TTCTTTTTCTTTTCTTT TTACGGTTTTATTTTCCG TTGTTTTTCTATTTATT TGACGACAAAACCTTA GCAATAAGCTTAGAATA TAATAAAGCGCAATTA AA (SEQ ID NO: 1284)</p>	<p>ETQQAFAIAFTKFRSKFTYFRTIPKFEQYLAPVDEILSHLLPTLFGK DTPFDHIRKLTLPTRDGGGLIPIVVEAPHQFLSSVKLTNKLVOQIID QDKILTKNSSGNVLEDEKILTTDLRKHREKIIAIVDSMQPDSMLRNI QQTRSECASTWLNALPLENQGFVNKEEERDALCLRYNFDLKNIPRIC ECGEFNVTHALSCKGGFISSRHNDIRNLFTLLKRVINQVSEPHLIP LDNENFYHTANKSNOARLDIKANGFWRNQGTAFDFVRVTHVNSM SNKNLIDIAIFRKHKEKREYGERVEVEHGSLLPLVFGTNGGGMGK ECHREVRRLAEKLAQNEKYSVVMTWLRTKLSFEILRSTILCLGRSRT PWTCKNDFEIGVDFKMDALEARI (SEQ ID NO: 1405)</p>
CRE				<p>MVCPCTCNGVADYNDHIRKKHPDERYTALQLQPLGLTPCPICKTACK NDLGVKTHLSKIHKISGASKISTQPRIRTEENTDNTNSVPTSSFNPLPEI QTLTPGLNNSRWADNPRKRRADPTSPTRGRNTRPRRFSYTDIDLND EPADNPRANNPRVNNPRVNEPSSPNLSISEFHTPTGLPLTNSNI SLKDQDKITGPILQKPLQKIEYSKIPPEHHLHARQAKIFADAANRIA KNFIQSPTKTLFNLLILPRFIGGLINGKVTIMQNFPSQIPPIKIDFPS EKTDSDPVLNAKLEKGYIGRAAKAIIDTPVAPETPESLNILREKHPI GQNNPNTKSPISGRQITEKAILLAISSIGREKAPGLSGWTRSLDDAAI KIPTQNDVIPALRLTDMIRQGTAPGRELLCASRLIGLSKPDGGVRPIA VGDLLYKIAKAILNTLWSPNCLLPYQLGVNSIGGVEPAIFTLEEAIMG PNINGIKSITSLDKNAFNSVRAAIASSVAKYAPTFRSTCWAYNQPS ILITENGSVLASAQGIQGDPLPLFLSFRPTLETIQKSLPYTIAAYL DDVYILSKTPVKDKIAKIEKSPFTLSAKTTETDIDLKTNGLKTLGSFIG PTELRKEFLQNKIQNFESSINALKLPKQYGLLILKSTQLLRHLRLTN SQDLWELWEKTDKLIADFVINLTVTKRKRPTDFVTPITLPIKDGGF GLLRHNGIAQDIYFAAKDLTTEIRHKIQRISNDFPQNQSPTATEILHLH NGVLADCKNGLTNAQLNALTENASYLGRKW/LNLIPIQKSNRLTDWE MAEAVRLRLAPVKPLTHPCNHCGRNTNINHEDVCKGAVRKYTARH DQINRSFVNSLSRPEIDVEIEPDLNNENNENNANTTENPTSPNG QNDTGCLFTTPIRSGRTRNGQLRADFAVINGVSKYYVDVQIVAINK DSGNTNPLNTLADAAANNKRRYQFLDPFHPHIIISAGGLMEKDTAQ YKQIKLIGPVAHAHWLDTSLILLRSRTTAAISIAKNRPRA (SEQ ID NO: 1406)</p>		

					ATGCGCTTTAAATTTGTGGG GGGTATTAAATGCGCTTTA ACTTTACAAATTTTATTAT GCGCTTTAAATGCTGCGG GCGTTTAAATGCGCTTAA TTTACAAATTTCAITTAATG CGCTTTAACTTTTATATTTA CTAATGCGTTATTATATA ATTGCTATTATATCGTTG CTATTATTATTGCTATT ATTATCGTTATTATTATGC AATTTATTATATAAACCTT CGTTTGCCCTCGAATTAT CCGTTTCTTTTCCATCCCA TCGCGGTTTTCGTAAGCT TTGGTTTTCGTAGGATTTG CTTCGTAGGCTTTGCITT CGTAGGCTTTCGTCAGCTT TTACCTGCTTTTATTTTTC TTTTTCTTTTATTCCCCC CCTTTTTTTACCTGGTTTA TTAGCGGTTTACCTGCTTT TATTACCTGGTTCCCTTTA CCTGTTTATTAGCGGTTT ACCTGCTTTTATTACCTGG TTCCCTTTTACCTACTTTAT AAGCGGTTTACCTGCTTTT ATTACCTGGTTCCCTTTA CCTGTTTATTAGCGGTTT ACCTGCTTTTATTACCTGG TTCCCTTTTACCTGTTTAT TAGCGGTTTACCAGCTTTT ATTACCTGGTTCCCTTTA CCTACTTTATTAGCGGTTT ACCGTTTCTATTAGTGGG CATTTATTTCCCGTTTAT TAGCAGTTAAATTTACCTT TTTAGGTTATTACCTGC

HER O	HERO- 2_Bf			TTTTATTACAGGGCAGCC CTGTTTTTACTAGCAGTTA AATTTACCTTTTTAAGGTT ATTACCTGCTTTTATTAC AGGGCAGCCCTGTTTTAC CAGCAGTTAAATTTACCTT TTTAAGGTTATTACCTGC TTTTATTACAACCCCTTTAT TTTTCTCTATTACGGGTA TTTATTACCTGTTTTATTG GAATTCACCCGTTGGACG GC (SEQ ID NO: 1161)	TGATTAAAGACCCGAAAC ACCAATGACCCCGGTT CATCACTGATGATGTC CCTGTCGCACTACCAGA GTGTATTCTAGAG (SEQ ID NO: 1285)	MNAVCGKVKCKNQRLRIHQTKMACLRRVQAEHRSAGAVTTVEP VLSASAPGQTEEDQGPAPHSARNLRATPAPQGRKSDHHRVKWP AANSKEWSQFDEDVDMILESVSRSSTGTDQKLQSMCTVIMSMGAERF GTGQRKPTDTMKPNRREVRIQLRQELKSLRRSFKASTSGEERAALA ELTHHLREKLRTLRRAEWHKKKGERARKRSFITNPFGTFRLLGQK RSGNLTCPVEEINLHLSNTFSDASRDVLDGCPCLLVTSPEPEVHFDISE PTLKEVRETVAARSSSAPGSGVYVYKHCPRLVWRLWKVW RRGKVAADWRAEGVWIPKEEESKVDQFRLISLSVEGKIFFKIVAQ RLIKYLLDNQYIDTSVQKGGVPGVPGCLEHTGVTQLIREAKENRGDL AVLWDLANAYGSIPIHKLIVETALTRHHVPESIQNLILDYSNFWLRAG SSTATSAWQRLEKGIITGCTISVPLFALAMNIMIVKGAEGCRGPVSR GTRQPPIRAFMDDLTVMTATVPVCRWLLQGLERLITWARMSPKPAK SRLVLKKGKVAERFRFTLGGTQIPTVSEKPVKSLGVFNSSLKDTASV QQTRSDLTWLEGIDKTGLPGSFKAWMFQHGVLPRVLWPLLVVEVP MTMVEQLERTISFLRWLGLPRSLNIALYGRSTKLQLPLSGLTEEFK VTRAREVLNVRDSDSKVSAGIHVTRGRKWKAQEA/DQAEARLRH SVLVGSVAVGRAGLSCPKPRYDKVSGKEKRLLIQDEIRAGEEEDRRC RMVGMRKQGAWTRWEHADSRTVWPELCRAEPSRIKFLISSVDV LPSPANLHVWGLAETPSCQLCQRRGTLEHILSCCPKALGEGRYWRH DQVLRVLADTVSNAIQSSRSQPPKKSIVFRAGEKTRQPTQSAGL LSTARDWQLLVLDGRQLKPEHIVATSLRPDMVLVSESTRQVVLLELT VPWEERISEANERKRAYAELVVSQSNWGRARCVPEVGCGRGA GQSLAVVLKLLGVGRFLRKSIRDILEAAEKASRWLWFRERGEWPKPH GHRSGNDQPRLRPGEVW (SEQ ID NO: 1407)
HER O	HERO- 2_DR			TTCAAGCCTGGCGCAGCC AGTGACTCTAGGAATAG	TGATCAACCCCGCTGG GTCACTGGGTGAGAGT	MTHANEQTNNKIYVTCICGKLNKHNHWGLKHQARMKLEQESKVQR TGPEPGTQEPEPEATHRAKSLHVPPEQTPSEWQQRIKWPPASKG

HER O	HERO- 3_BF			<p>ACTAGGTGGCAACCAAGA ATAGTTTGGTCGACTACTG GAGACAGTTGACGGCA CGGAAAGACGGCATTGG GACAGTATGGGTTAGCAC CCAGCCTGTGCTTTCTGT GAGAGAAACCCAAACAA GCTACGGAAGCCCCACA GAGATATACCCACAGGAG ATCCCGAGAGGGGGGA GGATGAGATCTCCAATCG GACGGATCAAGGTTA (SEQ ID NO: 1163)</p>	<p>GTATGATGTTGAGAGAC CCGAAACACTCAATGATC CCAGGATACATCACTGAT GATGTGTCCTCAATGAT CCATGAGATGTTTCTTGC ATAA (SEQ ID NO: 1286)</p>	<p>SEWLQFDEDSNIIQAIKAGDADSRILKMTITIIIFSVALERFGCIEKGKT KPTPTYIMNRRATQIHHLRQELSLKLYKKATDEEKQPLAELKNILRK KLMILRAEWHRRRRERARARAAFTITPFGFTKQLLGDGRSGRLEC SIEEVNRIEETVSDPLREQUELEPNKALISPTTPAREFSLGPSLKEVKEII KASRSASTPGSPGIPVLVVKRCPGLLLHLWKILKVIWQGRVRAEQWR CAEGVWIPKEENSKNINQRIISLSVEGKVFISRRLLTEFLENNVID PSVQGGIPGAPGCLTEHTGVVTLIREAHENRGDLVLWLDLANAY GSIPLKVELALHRHHVPSKIDLIDYNNFKMRVTSGETSSWHRI GKGIITGCTISVILFALAMNMVKSAAVEECRGPLTKSGVRQPPIRAYM DDLTIITTTVPGRSRWILQGLERLIJAWARMSFKPSKSRSMVLKKGKV DKFHFSISGVIPTITEQPVKSLGLFDSLSKDSAAIQSKKELGAWLAK VDKSLPGRFKAWIYQHSILPRVLWPLLIYAVPMSTVESLERKISGFLR KWGLPRSLTSAALYGTSTNTLQLPFSGLTTEEFMVVTRTREALQYRDSRD GKVSACIEVTRGRKWNAGKAVEAESRLQKQKALVGTATGRAGLG YFPTLVSVQVGKERHLLQGEVSRASVEEERSRVVGLRQQGAWTR WNTLQRRITWANILQADFQVRFLVQAVYDVLPSNLHVWGKNE TPSCLCSGRGSLHLLSSCPKALADGRYRWRHDQVLIKAIASLASAI NTSKNHRAPRKAVHFHFKAGEKPRALPQLTTGLLHKASDWQLEVDLGK QLRPHHIAATRLRPDIJASEASRQJILELTP/PWEERIEANERKRAKY QELVEECRERGRWRTYYPEIGRCGFAGRSCLKVLSRLGITGVAKKRAI RSEAAEKATRWLWIKRADPWMTAVGTQVGT (SEQ ID NO: 1408)</p>
				<p>CTGACCAGCAGACGGGAA GCCGCGACCAACTAGTCT CCGCAATATTGCACACAG GGCGACCCTATGGAGCTG ATTGAGTCAAAATTCCTCT GAGATATACCGATAACTAT CTACAGAACTGCACAGTT AGTTTGGAAAGAGCTTTTC TACTGAAAGACAGCAAAA TCCGCCACTTTAGACGAGC GTCAAGACTGCCCTCCCA TAACCAAT (SEQ ID NO: 1164)</p>	<p>TAGAAACCCACAAGGCT GAGAAATGTAGAGCATC TGATGGACAATATTGAT GATTGAAATGTTGTGATT TTAGATCAAAATTAGAAA TATGAAACCCGAACATAA CTAAATATAATGTTTTTT TAAAGTAATGATAAGCAA TACCCACATTGTGCAATA CTATCTATGTTATGTCCTT TGTCCTCCCTGTCATGTTT GGTCAATAATGACCATCG TGTCCTGGCTCCGTGTA CCTTCTTTACTATGAATA AAGAAATGATTTTACTAC (SEQ ID NO: 1287)</p>	<p>MALPAVRSGPASTWTLTLVIVAAGTGDGFMSEFKLPLSTDTWSGY NNDVKTLGLPLHHELATNEMSPKLAGEGFSDIMCDFMASKPEFSHTT EESHSEGYISHPEQSLAQVKRLKNLKKAFRADATPEDRKAFRDAIK TYSFMKRQQKRKETTSAAHQKEKEYHKNFWKAGKCAKQGLDIPPV KPAFSVYVANEYKKNYSHPTRVDNFNKLWFPPLPVEEQLPANSFDM SPVRPKDIKAVLSKRCATSAPGPDGIMYGHLKHLHPACHLFLSLSKLL ESGDPTSWSGNVSLIHKDGPAAENFRMCLTSCVSKIFHQLSER WAKYMTENDLIDPETQKAFLTGNGCEHVQV/MREILAHAKNRRT VHITWFDLADAFSGVEHELIIYQMERNGFPPIITTYIKNLYSRLKGVK GPGWESDPFPGRGVFGQDNLSPILFTVFPILQHLKGVEQQHGYN LNDKHVTLFPADDFCLITTNKRQHKLITQISSNTKSMNLKLPKRC KMSIVSGKPSISFTIDGDPVKTDDAPEKFLGGYITFSLTKETYDILA KTIIETVENKSAIRNEKLYRMYMEYAFPSWYMLMVDHLDLTDQLQ KLDSTHTKAIKTWLRMQPSATNAILNTRGLNFKSISDLYLEAHALAYS RSVLKADEKVKHALQAKLDRESQWTRKMQKMGWGICGCHTIHQQAIIH VAKDSEWTSVRKHVKQVQVDMRDHVDWTKHQENLLQQQGMQLQLL EEKCDLTWRSAMYNLPRGILSFASVRASIDALPTLCNLTTWGWGRNTD</p>

HER O	HERO Dr	.				<p>KKLCGNRETLHHVLNHCVALQQGRYTRHNSVLKHIHTDIESIDTS RINATYADIQGYTTNGGTIPVHTIPTTKQPDLIYLPQKTLHIHELTV FEKNIKSHDRKVNKYSTLAADLETAGISATLTCEVSGRGLVTPENKT RLRTLKIVKAKPKTLFTDISRIAMLSYAIWNSRHEPYWESETLL (SEQ ID NO: 1409)</p> <p>MTTHRAEVTTSGKTQEEPGEATHSAQSLVSPTPAAGRSPATQSC QVTAAHNSPQSQVAVTRSDCVPLAQPRIQWPPQSSKKAEWLQ FDKDVNQILEVTGKGVDDQLSTMTTLVINIAAERFVTPKPTPSTY TPSHRVKEIKRLKELLLKQYKAAAGEVERAGLEDLRLGILKQLVNL RAEYHRRRRERARKRAAFANPKLTQLLGQKRTGKLTCSKEAINN HLKATYSDPNREQPLGPCGALLTPPEPTSEFNMKEPCRSEVEEVRR ARSSAPGSPGVYKYNKCPKLLHRLWKALKVIWRRGKIAQPWRY AEGVYIPKEEKENIDQFVLSVESKIFFSIVAKRLSNFLSNKYIDTS MQKGGIPVPGCLEHTGVVTLQIREAREGRGLAVLWLDLTNAYGSI PHKLVEALEKHVQKVKDLIIDYYSKFSRLVSSGQLTSDWHQLEVG IITGCTISVTLFALAMINMMVMAAETECRGLPSKSGVRQPIRAFMD LTVTTTSPGARWILQGLERLVAVARMSFKPAKSRSLVLRGKVRDE FRRLGQHQIPSVTERPVKSLGKAFNCSLNDRDSIRETSTAMEAWLK AVDKSLPGRFKAWVYQHGLPRLWPLLIYEVPMTVVEGFEQKVSS YLRWLGLPRSLNIALYNTNKLPLFGSVREEFIVARTREHLQYSGS RDAKVSAGIVIRTRKWRAAEAQEAETRLKHKAILGAVAQGRAG LGLAATRYDSASGREERQRLVQEEVRASVEEERTSRVAMRQQA WMKWEQAMERNVTWKDIWTWNPRLRIRLIQGVYDVLPSPSNLYI WGRVETPACPLCSKPGTLEHILSSCSKALGEGRYRWRHDQVLSIAEA ISKGKDSRYROATAKVIQFIKEGQRPERTAKNCSAGLLSTARDWVMT VDLERQLKIPPHITQSTLRPDILVSEATKQLILLELTPWEEEREEAE RKRKYQELVEQCRANGWRTRCMPVEVSGRFGFASYTLSKAYGTGLGI TGTNRRLSNNVEAAEKASRWLWLRGEQWGG (SEQ ID NO: 1410)</p>
						<p>TAGCATGCCACTTGGACA CAGCGCGGGTCTGATC AGCTCGGTGGGTGCG CTGGAGGAGGGTGTCTG TTGCAAGACCCGAAACAC CCTGTGAGCCAGGAAA CAACACTGATGATGTCTC CAAGGTTGTGCATCAGG AGATGTTTCTGTAAC (SEQ ID NO: 1288)</p>
						<p>AAAGCAGTAGAG (SEQ ID NO: 1165)</p>
						<p>Danio rerio</p>
HER O	HERO r	.				<p>AGACTAGGTGACAACCAA GAACAGTTWGGTCGACTA CTGGAAGACAGTTGGCA GCTCGGAAAGACGGCAC CGGACAGTATGGGTTAG CACCCAGCCTGTATCTTT CGCGAAGAACCCAAA CAAGCTACGGAAGCCCT ACAGAGAAACACCCCCAG</p>
						<p>Takifug u rubripe s</p>
						<p>TGATCACCCCGGTGGGT CGCTGGCGGAGGGTGT ATGATGCTGTGAGACCC GAAACACCTATGAACCC AGGATACATCTCGACGAT GTGTCCAGTGCATCCAG GAGATGTAKCTTTAAGT (SEQ ID NO: 1289)</p>
						<p>MTPAMEMTTTVCISLCKNQRLKIHQARMKCLEREVEVQRTGP GPGETQEEPGQEATHRSQSLHVPEPPNPNRVVQQRIKWPPANRR SEWLQFDEDSNIIQATAKGDVDSRLQALSTIIVSYGSERFRIEKGNT ETTSVTMNRFSFKIHLRKLRLTKKQFKRAXDGDQKALKELYNILRKK LKTLLRAEWHRRRRERARKRAAFIANPRFSKQLLGDGKRSGRLECSR EEVNRFLQNTMSDPLRGQDLGNPRLALISPAAPPSEFKLAEPSEVEE VIKAARSASSPGSPGVPLYVKRCPILRLHWKALKVIWRRGRVADQ WRCAEGLWIPKEEDSKNNINQRTISLISVEGKVFVFSRRLTEFLKN NYIDTSVQKGGIPVPGCLEHNGVVVTQLIREAHESKGLAVLWLDLT</p>

					GAGATCCGAGAGGGGG GGAGGATGAGATCTCAA TCGGACGGACCTAACGTT A (SEQ ID NO: 1166)			NAYGSIPIHKLVELALHLHHVPSKIKDLILDYNNFRLRVTSGSVTSDW HRLKEGIITGCTISVILFVILAMNMMVKAEEVECRGPLSRSGVRQPPIR AYMDDLTVTTTSPGCRWLQGLERLILWARMFSPKTKSRSMVLKK GKVDKFRFSISGTVIPSITEQPVKSGLKLFDSLLKDTAAIQKSTEEELGG WLTVDKSGLPGRFKAWIYQYISILPRVLWPLLVYAVPVTTVESFERKIS SFLRWLGLPRSLNSAALYGTSTNTLQLPFSGLTEEFKAVARTREALQYR DSRDKVSSAGIEVKTGRKWAKEAVVAESRLRQKALVGAVATGRT GLGYFPKTQVSHARGKERNHLLQEEVRAGVEEERVGRVGLRQQGA WTRWESALQRKVTVWSNIMQADFHVRVFLVAAYDALPSPANLHA WGKSETPTCSLCSGRGSLHLLSSCPKSLADGRYRWRHDQVLKAVAE SIALAISTXKHHHAPKKAISFIKAGERPRAGPQITGLLHTAXDWQLHV DLGKQLIFPQHIATTSRDPDIIIEASEASHLIMLELTPWEEIEEANER KRAKYQELVEECRGRGWRTFYEPVIEVCGRGFAGRSCKAFGLRGVTG TAKKRAIKXAEAAERATRWXWLKRADPWVATGTQAGS (SEQ ID NO: 1411)
HER O	HEROT n	.	Tetraodon nigrovir idis	AGATTGGTCTGGCTAAGC CAGTGACGTCCAGGAACA GACTGGCTGACGACCACG AATAGAGTGGTGACAGCT TGGATAGACAGCTGACAG CAGGAAAGACGGCAACC GGGCAGGAAGGGCTAG CAACCCAGCTGCATCTTC CGTGAGGAAGAACCCAAA ACTTGCTACGAAGAGCCC GAAGCAAGATACCCCA GGGAGCCCGAGAGGGG GGGAGAAATGAGCTCCCA AACGGACGGATAAC (SEQ ID NO: 1167)	TGATCACTCCCGAGTGGG TCGCTGGGTGAGGGGG TCTGATGTTGAAAGACCC GAAACCCCGATGACCCC AGGTACTATCACTGACGA TGTGTCCAAGACATGCAT CAATAGGTGATTAGAA ATC (SEQ ID NO: 1290)		MATTOASVKPTAVATVCVGKICKNPRGLKHQTKMGCLASVQPEQR ARFSLSESREVPARAEPYGPQPHSPHEALGETQEERQESPHSAQNL RAQAQAQPDNPQHRRYKVPWPPASKVSEWQQLDELEGILESTAKG GVDRKLTQMTTLVISFATERYGTMEKRAAPEKYTKNRAEKISQLRQ ELRVLKQFKGASEDQKPLAELRCTLKLLTLRRAEWHRRRAKERA KKRAAFLANPFGFTKLLGQKRSALHLECAKEEVDVSLHDTFSDAEREN SLGECRVLISPEPACSFNTKAPTWKEIQTVRAARNNSAPGNVGP YLVYKRCPKLLARLWKILRVWRRGKVAHQWRAEGVWVPKEEKST LIEQFRTISLNVGKIFISILSHRSLDFLLKNQYIDSSVQKGGIPGVPGC LEHCGVVTQLIREAREGRSLAVLWDLANAYGSIPIHKLVEMLARH HVPPIKTLMIDYYDSFHLRVTSVTSVSEWHREKGIITGCTISVIFAL AMNMLAKSAEPCRGPIITKSGIRQPIRAFMDLTVTTTSPGCRWI LQGLERLMTWARMFKPGKRSRLVKAGKVTDRFRFYLGGTQIPSVS EKPVKSLGKMFDSGLKDAASIRETNDQLGHWLTLVDKSGLPKFKA WVYQHGLPRILWPLLVYEFPISTVEGLERRVSSCLRRWLGLPRSLSSN ALYGNNNKLTLPFSSLAEEFMVTRAREVLQYRESKDPKVALAGIEVRT GRRWRAQEAVDQAESRLHHKELVGAVATGRAGLGTPTTTHLSRLKG KERRDQVQLEVRASIEEQASQWVGLRQQGAWTRWEEAMARKIS WPELWRAEPLRIRFIQSVYDVLPSPSNLFVWGVKVESPCPLCQGRGT LEHILSSCPKALGEGRYRWRHDQVLKAEISSAMEYSKRLLPLPGRG VRFVRAGEQPPPPRAQPGGLATARDWQLRVDLGKQLKFPENIVET NLRPDIVLHSQSSKQVILLETLPWEEERMEEAYERKAGKYAELVEDCR	

NeS L	LIN9_S M		Schmid tea mediter ranea	AAACGACATCATGAACGC TTGGCCGCAACAATCCAGT TATCCCTGCGGTAAACATTG TGGAACCTATAAGACAAG TACTAAAGAAGAAATTAG AAAAATTAGAAGAAAAA TTGAAAAATAATTTATTAT AAAAATTTAAAAATTTAAAT AAATTTAAAAATTTAAAT TAAATTTAAATGAAGATAA AAATTTATTTAATCCAATA AATAATCAAGAAAAATCAA GAAA (SEQ ID NO: 1168)	TAAATGGCAAAAAGAT ATTTCAAGATGAATTTGTG GACTCATCTAAAAAATGA CCACCTTGAGTCCAATA TGCTAGCTATCATGGTT GCTGATGGAACAGATAA GGCACCTGATAGCTAACT TTTCACTGTGAATATCTTC AGATATTACAGTGACAC GAAAGGACACCACTAGT AAAAACCACTAGTTTTT CTGACACCTCTTGCTACA AACTCTGTAATAATCAAA AGGATCGATAGCGCGG CTTTCACGGTCTGTATTC GTACTGAAAATCAAGATC AAGGAAGCTTTTCCCTT TTAGTCAACACCAAGTTT CTGCTAGTTGAGCTTC CCTTGGACATCTCGGTT ACCATTTGACAGATGTAC CGCCCCAGTCAAACTCCC CACCTGACACTGTCTCA AAACAGTTCAATTGCATC CGAAGATCGCAATTTTT CACTAAAAATAAATTAACA AAAGTTAATTATACTGCT TCATTGAGTAAGTAGAAA AACAAATC (SEQ ID NO: 1291)	RAGWRSRCLPIEVGGRFAGKSLCKAFSLGLITGMRRRKAICAASEAA ERASRWLWIQRDKPWTSASWTOAGN (SEQ ID NO: 1412) MMDSRQLNTPKIRKYQNPKMTNDIMKSYNAYV/SDVTPQETTQT THLNVNDIDNETTQKQPLTKSGPKSPKPIAVSYKFKDATFIWDTTPQT NPPRDCTKLIDKTRPKRTIFKSAFQSLYKLSNETSEVETKFLMATH KYRFKDENSRLAYRINRYVMETANEFKETETEDMARFAKFFTFPENW LKLKPYSTATETSPADRIK/VQKLVLDLTCRYPFKTQEEQTSVANFLHFF TQRSIGISRDYKQKPIPFMARKNTRPETTTSTMTVTSPTQNR/LPMVI ITPLEEPKSEHRRPEKRGASNDTIV/LSDEEPL/LKRRLLPTRKSKNPTGA GNVPTTECTDEVKFLNNEYQIECKEKGKVVWENVRNGLNHLRQKH DFPNRTDVMVSCVRCEVPKGAECVNIHKNHKDDKEESEAGSLVA NTQDIPNESSLSQAALIEVYLRNLIKMKENQERNIQYLEPSTANFLNRN LRAFQNVKIEKLGWEQVLIHWNKCHWIVYLANCDSKTSVILDS DNQMTLQQRNCNIKAKFKELEGTFEETVLTGLERKVPQPPNFDC GIWVQISDFLKDPRIDYHTPDSKRIRKEIGELILEEMKNPASKIKPN KEIQLSKFRLLQINVNDVFWFAAEVQKSLPKIRTKRDGKLNKLSCS YQIQLRFLGLAPKRAVKEIYQETSTADLETRVLNEHFKKDESTMIKECKI KNGNHYQDWITKAQIDNKEILEALKNSTDSAPGEDNIPLRQWIIWN NDGVLFDMFNYIKRTHDIPDMWKNYTTLLIKPGQSQNIPANWR PISILPSTYRIFMKVLNKRVLWANRGENELISKWQKAVDKANGCDEHSY VIOALIEKANRSYKNEQCHLAFDLADAFGSIPQVWHTLKNMGM DEETINLLKEYIKDCSTKYKCGKNESEKIKITGVRQGCPLSMTLSLSCI QYLIQGIAEKKKGATGAEVCILAYADDLVIVANTAKDMQMLLTIE NLAKQADLIFKPAKCGYRDRDKKSMKIYGEISIVDEKNVVTYLG VRIGDTKKDLNVRFEEVKKKTAIFKSLRSDQKLEAYNIFCQSKFVVI LQGEDIAKTIETYDEEIKMIKEDILKQDKSPFTDFVIYSPREKGGIGI TKIIDEQTIQTINRTAKLLNSSHRAIRAIYEEUQVANLRGEKEINTIEEA LKWLEGTNKKYKNSNAKTTWITRVREAFQTLKHHKIKVRFVPKNCI GYKIKCDTQEKVELDNSKLSLHWMIMEAYYKEWK/KALCQCYISL KTSEFMEWKMPRGLPDPDWRF/LTKV/KANMLDVNM/KQANQGGRL GSTKCRKCEDKESASHVINHCASGNWSRVEKHNVQVQNELAKELTKR NISFEKDSIPKETKESL/RPDLVIR/LKD/KIMIVDIKCPFEESAIESARNKNI DKYRELAKIEIAQTGLQTTVT/VFVCSLGTWVDKRNELLRQMGIRYE ESKEMRINMIQKAIHGSRTYD/HHRNFNNG (SEQ ID NO: 1413) MPLXISDCVHLVSAEGDTMNGRSTCGPLSRSSSVVSRSSPSPSPVP HPSPSIGPDTGLSAGIIGTSRGCSLWLP/PEVDNALSQWL/RKGLERDHEV LV/CGFEEAAKPLSLSKARLLRKTFRNTGVVRHILEFDGLR/LVHTNCNETE CVLSTLXSSXAAVEVVRISLKCEPREPCEPKCVLSILCSDKIVXISFECETRE
NeS L	NeSL- 1_CBr e		Caenor habditis brenne ri	AAGGACGCTGGTTTAAGG CCGAAATTCGTTCTGTTCTTT TCTGGCGGCTTCTGCTTGA	TAAACCCACACGAGAMC TACGACGCCATAAGATCA GGCATGTACGGATGTGA ATGAGACTGATGAACGG	

			<p>GCTTGTTTCCGATCCT (SEQ ID NO: 1169)</p>	<p>AATGAGCAGCGTGCCCAT AGATCGGGTATKAAAGA WCAGAGAGCATCCCTAM CATCGGGAAACACGAG TTACTGCTTCACTGAM CTCGCTAAGCTCTCATAA TGACCGAACTTGTTCGCA ACTGCCTCTAACCGGGC GGGTGTGAGAGGGAG GTCGCTTGAGCGGAC GCAATGAGGGATGTGTG CAGGTTCCCTCTTGAG ATCGAAAGTCTAAAGT ACTAGACCGAAAGATCG AGGACGCGCGGATGCG CGGAGGCACACGCGG GTAAACAGCCAGATAAC CTAGTAGATCTTCGGATC TCGTCGCGCTGGAGATAT GTGGAACCTGGGAAAG GAGAAAGTTGTTGTGG GCTGGCAAGAGTGAAGT TTGAATGTGAACACCGT CATGCAACCACTAAACCA GTGCGGATGCGGGTGA GTCATCACAGGAAATGT TTCTGTTGCTTGACTTATC AGTGTGTTGATATCGCCT CAGGCACAAGTATGAAG GCCCCACCCACATAAC TCCCTAGCAACTGGTAGT CCAGCAAGCGCTGGTWC TTGCTACTATTGGCCCC AGGCTCGCCC (SEQ ID NO: 1292)</p>	<p>PPFFXDRKREP IPFFERMYPDRDPIPSFICWWMYDLRQRMTPGTL XNPLSXENKDSWGRPAVKNEIRSMRSYLEENVKENLNLRLRGG GEGKMIRKLVAEKSDTEAVCRILYPLDDRYECFVDCGCTTSTMGYG SSDLKYMTHIKKEHGVQVWTEYCSLCKNQAPFMGGAAKSWTA HMAKHTETVTKLKPSISTTAKVAALKDEIAVSLPKRQVRVLRDPD EVKEKAKPTLASTREEVKRNALNMAPLVELSSQNLTGAERPEETS EAMRLEECRTPEKIALEGKIQTRTVTKLSALKESMEKRTREEKVGKP SLAPIHEEVKKTARSLAPLVEPSTFTHLTGASRLQAVRDAFSKANKD AAAKRRSLAKPARLSEIMNTFTKETVNETKEPVNDTDESATIQQPQ VRVRFNTWCLDHETTREAWLTGEVDWFMGKVTEKKDQYRVFDS LVWSMYFHGVYVLDLMDPLTYFLPICEHDHWVLLVDEKGIWY GDSKGAEPCEIAKFIETKRERRMFPVPVPLQRDGVNCGVHICLMV KSIVNGEPWYTEEVKVFRRNVKRGLEFEGFELYSERIVYVGGDSIKV NDEHDDVVFLSEETNNTFTIEQAEADPAEEDAQHLESVPVKLME LKIPKIEIKKEIRKPKQKQIEKKRVPTGKPDDELVRVRLWLEREVQSY FDSGKRFQRLWLDVLTAAIHKATAGDEQAERIEKRSPLEVEEGE MSTQTEPKRERKEKESGCEMKASHKEMYFKNRSKAFNVIIKDSKQ CEIPIETLQKFFEGTTAETNVPAEVLKEMGSRPLKLEALDWMANFIE SEVSDAMKTKDTPAGVDGLRYHHLKWFDPYKMLTLLYNECKNHR KIPSHWKEAETILYKGGDETRPDNWRPISLIMPTIYKLYSSLWNRIRS VGGVMSKQCGFQEREGCNESIGLRTAIDVAKGRRNLSVAWLDLT NAFGSVPHELIKSTLESYGFPEMVTIEMDMYRGASIRIKSNEKSEQI VIKSGVKQGDPIPTLNFNMCLENVIRRHLDSDASGHRCIKTKVKVLFA DDMAILAENRDQLQTELNLKDKCESLNLIFPKVKCASLIERGMVVK NAEVLIRGKPIRNLIDENGSKYLGVHTGIATRVSTMQLLESVTKEMD LVNQSGMAPFQKLDCLKTFVLPKLTMYANAIPKLTCLKVFANLTMR MVKEIHEIPKGSPLYVYQLPPSQGGGLGVACPKITALITFLVNMVKKL WSSDSYRKLRYDYLDEVAETETGMEEMTKEDIKYLSDGVPIDKKAF GYNTFTRVDVNSLTIXGAPLHKLKIVERDGDFAILVQATKEGMEK IFTCAQEKLLQQLLKAENVNTALAHREFFTEKPKVSAMVSMRQYPSN AFVNGKNVSIHVSHWIKARLNLHCFNTYGENKSKVCRRCRGD VETQLHILOXCEYGLPKLINERHDAVLHVVRNLIRKGSKKDWKLKIDET VSSCNQLRPDIYMCSPDGKIVIMADVTCPYESGMQAMQESWNRK VTKYEGGFHFXKMGKFTVLPIVVGSLGTWVWKPTTNSLVQLGIEKX TIIRVPIELCSMTMEYSKDVVWNHIFGDTFRKPPMRFGVEKPKGNS WKKEGSEPKGAASSD (SEQ ID NO: 1414)</p>
--	--	--	---	--	--

NeS L	NeS- 1_Cla p	.	Caenor habditis japonica	<p> GCGCCCGGTTACATTGT CGGGCCACCTTCTCTTG GAGTAGAGTACAGTCTAC TAATTTTTGATAAGCTAG TCGGGTCCGAACCACTAG AGTTTGCTTGAAATGCGT CAAACGACATTTTAGAAC TCGCCCAAAAGTTGGGCC CGACCCCAACAAATGG GACCTCTTGACGATTTTC CCTGAAATCGGAGGATG GAATGGTCCCTATTCTTG TAAATAGKACTGTGCAATA CCCCTTCGTCTGTGGG GAACAGATGACACGTGAC GTCATCGGTGACGTCA CGTTTTCCGTCCTGCGG GAGCCCAATCGAGCAA TTTTGTCTTTTGTAGTGTC TGGAACGCTTGAAACCCC AGACAAATCAGGCCAGT CGTCGGAATAATTTCTTTT GAAATTTTTGGCGCCTGC GAAAAAATTTTTTAACCG CCACAAACCCCGGAGG CGCGGTAGGGATATCG ATGTCATCGACTCGTCGGT GATCTTTGATTTTCTCTG CGTCTCCTATTTTGGAAAC GTCTCGACCAAAAACCG GGCTGGCAACCCACCGA ATCCGGATGTCGGAGGGA TTTGGCAAGAAATGTTGG AAATAACGAAATTTCTGTTA TTTTCAGCACAATTGTCAA ACCGGCAAGAAACTGGA TGGACAAGACACACAATTT ACCGGAAATTTGTCTGTT </p>	<p> TAAAGCCAAAAGCCAC GGAGCATCGGAAGAA AAATGGAAAGGACTGA AAACGAGACTGAAAAAT CCCAAACAAAACAAATCC AAAACAACTGAAAAAA AAAAAAACAAAACAA AAACGGACAGACACTG GAAACAGTGTGAGGCAA AGTCGCGATTACTGT TCCAGGCCTTAAAGTCC CGAAATGGCGCAAAACA ACCTGAATCTATCTGAAA GTGCTCAAAACCGCAC AACTCGGAGAAAATCAG GGACAAATGCTTCAACG AACGGCTGGACAGGT ACCCCTCTGAAACCGC GAGTTGAGGATGGACG GGAAGCCCGAGGCTT ATGGCGGTAACTCGGT TGGTGTGCTAGTAGATG ATTTATCCGACAGCCC CAACTAAGAGGAATCCTG GGAAGGAAACCTGAA AAAGTTTTACAGGCTG GTAATAGTTTACGACAAT TGTAGTACTGTCTTG AACCAACAAACAGTG GTTCTCGGGTAGATCAA ACTATAATTTGTGTTTT CTTTTACTTGACCCGGC AACATTATACCACGTC CACAAGGACGAATTCATA ATGGCCCTCCCTAAATA AACTCCCTAGCAACTGGT GGTCGGCGAAGCCGGT TCTTGCCACTATTGCGCC </p>	<p> WRRPAPQTKNSSLHLGHEVKRIARLPGIFEFHAKPKNSSLHLG HGVRKXARLPGIFEFHAKXKNSSLHLGHEVKRIARLPGIFEFHAKP KNSLHLGHEVRNRSLPGIFGYQKQKNSSLHLGHEVRRIARLK PGILEFHAKNRKSLKVTFLSDLXAHAGALACRSFLASTLKEHCRQK SFKPVGFLHLKNSINEVASLRNVKXFFESGKPIGGMASFSRTKI TFFKLKLFVLSAENPIIRQKTQNKASXVQJARGGHLSDCLPSQK MAGVLRFLSVQSTLSHRPFDTLRSDDDRGRRTIKLQFFIKENLVT PXVARDVKILXQTKNNSGNSDSETKNFKNKVRQNGPLIGGNN HKKIGENQITRLEIESKDDNVLRILYPTNDWYKYSQWQCHKS LVGYGAHDLKYLTDHIKSTHKKVSVSYQCSICDAEAGTGTGAARW ITAHMPKVHGEATHRIKQNSEKTTNVKTANSIQEMALSQPKNGP KKVVMATSTTPEKKISELSKIQTREVAQKLSALKESAKNQGNKTK NVKSLKTIAENTNETKISARSLINYLKPEDVLNHIPEKPKASAKXG LQELTGAQLQETRRFRMAGNRDSIARRESLSGLKISNFKIELKNAP EKTLLKPAVTQKQNTSQNVSSSTVWKENKTGNDVTIDTETVKKRI NTWCLDHESTENAWMADDIIFWYQKQIEISLNDKKFKVIDPLIWTY RIYGEVQDELVGFEKYFFPICENGHWVLLIIDDKRWWYSDSLADKPI EVIEDLINKLNTQGFNQTPVKQKDGFCGVHCLVAKSVITENFW YTEKDVNDFRKTVKLWLFSEGFELYSEPYKQIQKNISVNSEKNQISD NEKNWGDKTQTVNESTIKERDEDFILRPHISVGVALKTEDEKNQKA ENLKAPQKLAIRLRLKLTCLKTAVKKGPEETERAAIPNLMALIKLT PKVPEVRNPEKGENYKXSPNKKRQIPTGPKDELVKVREWFIEIQ FQAYFEDGKSFQRLWXTGLLTAIHKASAGDEQAQVGIKRCPPLEIE EGEMATQETKQPKNQKSTKGANSSSIREAYENRARTFNKIIGKD DKCEIPIEKIEFFENTSTNTVPTETLARITSDLPKLEIGSWIEEFREKE VAEALKTKDTAPGVDGLRHHLSWFDPKXKLLTKLYNECREHKKIPG HWKEAETVLLYKGGDETAENWRPISLIMPTICKLYSSLWNKRIKSVT GVLKQCRGFQEREGNESIARLTAIEAAKGTCKSLIAWLDLTNAFG SVPHESEIATLIAYFGPMVTEVIKDMYNGASIRVKTNEKSKQILIKS GVKQGDPISTPLENICLESVIXRHLKSADGHKIXSNIKLLAFADDMAIL SDSKTLQELQKMDDDCTPLNIFPAKCAJLIEWGVKQKQDKIKL KGQFIRSLAEQDQTYKVLGVQGTGIETRVSAMQMLMKTYSSELDKNCAL AXWQKLDVAVTFVLPKMTMYANTVPKLSLSEKAFANITMRAIKVMQ NIPVKGSPLVYQLPIGKGLGVACPKTTALITYLVTMMLKWLSTDDYI RKLHTDYLKMWAIKETKTKEVTDLASYLSDDKTVCKKAVGYNSTFR VREICKTSLKNKGALLSQLKIAKDGKLAIVQAXKDGKTKIFTHDVKT LQKXLLKEINEALLHRTTEKRYKSEVRVWVQEPQCNFSVRDGGKVS IGHRFVHKARLNLLACNVTWQDAATKQCRRCGYEKETQWHILSS </p>
----------	--------------------	---	--------------------------------	--	--	--

NeS L	NeSL- 1_CRe m			ACGTCGAATTTCCCAATTT TGAAAAAATTCCTCGTTCC ACTGGTCGGACGCGGAGG TCAGACGATCTGCACGTCT GAAACCCAAAAATCTTCGG ATTTATGCAGTAG (SEQ ID NO: 1170)	CCAGGCTCGCCC (SEQ ID NO: 1293)	CPKSMGGKITERHDSV/LKTVKEMIQTGS/LKNWKLKDHELPGSTRLR PDIYLRSPNGSEILGDVTIPYEHGIEAMQTAWQKQKIEEYEGFKYLRST GKLLTIVPVVVGALGSWWKPTTDSLVSLGIDKNTVYKRAIPEICSTVLEYS KNIYWNHIFGDSYQKVPMPFFGGEKPKGQSWKKVKPPEGKTASNHE PPG (SEQ ID NO: 1415)
	Caenor habditis remane i			CGCGAACCAAGTCAT (SEQ ID NO: 1171)	TAGCCGATCGTAAAGA AACCGAGCCGTAAACAAC AAGCAAAGTAAACAAAA GAAAAATCAATAAAAAAG GAAGTTGACCTCAGAC CCGAGGAGGGAAGAGA GACACCSAGAAAAAGAG AGACGAGAGAAAAAGGA GAGACACCTCTCATAAGG AGAGTAGGTCAATCCA AATGTAACAGAAAAAA CCAGTGGGAGGGAAGA AAGACTGATTACCCAC TAAAAATGAATTTGAAAC AGAATTTGGAAGAGAAA AGAGAAAGGGAACCTA AAGAAAAATAGTTCTTG CCAAAAATCTGTAGAGGA ATACTTTGTCAAAACATG ATAGAAACCAAGTAATCTG GTACGAAAGACAAGTAA GACCTGAACCTGACAAGA AGGAAGTCAGAAAAAGAA TACCGCTCACAAAGCCTG TGATCGATTCTCTTACCTA CTGAACCTGTTCTCTGG CCTCGTAACCGGCTAAAG GGAGAGGAATGTTAAT TGGAGATAGACATAAAG ATAGGTGGAGTGAAGGT CCTGTTCTTGAACCTAGG	MTVIFDRGIGERGQMAVCSLHRYFSFSPPIPPVYVNNSGFGENCGG TDKSLPVIEVVVREVKNWSENILVVECLIMVKSERVVVKRQNLK VIQNLARINSTLFSNLGNQJFCVVPRIKDSTNKEQYKQKXKFHVSF RSIKSQPPYLRGGGDVMEDEIRGIRKLEPEAQDSSKPLICRVLYPT QGYMYKCFYPCCKGHSNGSTDLSRLKHKHMVDKHFTNIEFAYKCATC MFLTTGKSATALKSIKAMASHHKVTMEPGKSLVQKLNARLEEAAP SLMPNRNSKVIQLTPEKSISELEKKQTRSAKQLSTLKEQAQKKEE VKIAEVKKREPLSIIPESNVRRSLAAGLEQCINPEQSVQAQRIKREEY AKASREAAKRRSSLAMKPARLPDKENETITLQETKKIDDPIDLEKECI LTTVLQVPRNQFNSWCLEHETTIDAWLTDE/HMYMCTITENRYF MAIDPVLWPVYVRNGAEDLLRTSCPGTFFFPICESNHWVLLVIEHD VWYLDPKGEEPKNVEILLESKMRKQYEFPPSQORDNVNCGVH VCLMAKIVDECGYNWYSEEDVRSFRTNMKDKLSKGYELCPPEYNR QNLLKTEKQKEVILEEMIDSFVEDDMTFTVHRSDHGDDEVEHLKT IEQEPENEISEIENVEGSDSVIPKLMEMRVQTPPVINEKRGKKRVS KEKPRKQKEKEQKVPPTGKPDDELVKRVVWFEKEFKSVVEDGKSFQRL EWXTDVLTAIQKASAGDEKAVELIEKRCPPLEXEEGEMCTQTEKKKK PKSGKNGGQESMKSLSMASYSENRAKTYNRIIGKHSKQCEIPAKVQ KFFEGTTAETNPVKETLKEMCSRLPKVEVGTWIEGEFSESEVTEALKKT KDTAPGVDGLRYHHLKWFDPPELKMILSQYNECREHRKIPKHWKEAE TILLYKGDESXDNWRPISLMPTIYKLYSSLWNRIRAVKGVMSKQ RGFEREGCNEISIGILRTAIDVAKGKRNIAMWLDLTNAFGSVPHL IKETLESYGFPEIVDVVEDMYRDSIRVTTREKSDQIMIKSGVKQG DPISTPLFNMCLSVIRRHLDRSVGHGRCLKTKIKLAFADDMAVLAES SEQLQKELTAMIDADCSALNLLFKPAKASLILEKIGVNRNLNEVLRGK PIRNLMENTYKYLGVQTGTETRVSIMDHITEVSREIDLVMMSQLAM HQKLDILKAFILPKMTYMYQNTTPKLSLKV/FANLVMRSVKEFHNPL KGSPLYVQLPVKGKGLGVACPKNTALLTLVTIMKKLWSSDSYIRKL YTDYLEEVAKEIGKFEVNLNDLAEFLSDERAVIDSKLFGFNAFTRVREV VRSICKNKDPSLHSLKIEREGKLAISVQATEESIEKIFTDEQEKLMYLL KGELNTALQHRFFFTQKVFKESEVMRVVQQHPQSNFVRNGGKMSFS

					AGGAATGTGAAAGAGC AGAAAGCCGCGAGGCTT TAGACGGGTAACTCAGTC AGTTGCTAGTGGTCTTCG GATCCAACGGCTTCGGAC ATAGTGAGGAACCTGG GTACGGAGAGAAATGG AAAAGAGATAGGGCGGG CAAAGGCTAAGTTCATAC ACTGTCATGCAACCACTA AACCAGTGGGATCTGCG GGTGAATCACTTTGAAA AGAAGTGAATGGACGTG CTGATGCTGACTTTAAA GAAGTCTGAAATTAAAA AACAGATATAAAGCCCC TCACTATAAATCCACAG CAACAGGTGTCGGCGG AGGCCGGTCTTGCCACC ATTGCACCCAGGCTCGT C (SEQ ID NO: 1294)	AQRVHPGRLNQLPCNVNTWAKGRTKLCRRCAKNENETQSHILQVC DYSIGNIIKERHDAVLYKFRILKRGSKGHWLERTDRTVPNTGSQKLP DLYLESPDGKXVILADVTPYERGIEGMQKAWNEKINKYTDGYKEIFR RQGKSLVLPVVGSLGTWWKPTESLIKLGVKETTIVRRIPETCGMV AEYSKNCYWRHIYGEKYVQTPMINGGKPEGNDWKCKEGIEVPKV TN (SEQ ID NO: 1416)
NeS L	NeSL- 1_TV	Tricho monas vaginali s	GGGTGAGTAGTCTAGTGG T (SEQ ID NO: 1172)		MIPVLGTGGPEKLPQSVYVCGNTAITDSFTPTAKTILKPEEQNLIDIVL KNIAALNPENYSDILRSLSKMEFLRDYPKEIENYVISEKLFQSQSIASLPIS LLVASMFSPEDRDLSTEPFHCNADGNCNFCNDCERMVHEHREHNT DPMINTFETTEDTFRITAIKIDKTGIEELNPLKYRCSYCDLFEAEDH AIHMISHLTEKLSPDISFFNDILRLYKTIDKPTVQNLFPETQVAIFDTLE ETNRFRLIVGREAIETIEEAFPPSPGTDKRPSTIITDTCQLRFVPCMDPEP PKGDLGILTLLRDFSAHNIPIKSLNNKELIADKIDIDYSPDFVEGALANA EEHDTTNSQNNNGRYNSAEKLTFLIQCEDYLTNIKTLEDLERYTTIK DYRVNKEVIAEDTPIFYFLVEEGKLPKPLRCPLSEYEGHEDKAFESLR KLCDFHFKGEIAKTSFDPKVHTIDIWVEFLAQAYGTGTfVYKDENGID LDTHVFKCPYADCSYTNDRSKLMDHMKTKHAKNVYIERYGFFWG IVIEGVNRPKGIVYPTLKDIKEHACRKCPEAGCNTYVTELSDIKEHLKKK HKSTTAGVDGEIAHTDATYCWITKEELDHALHAERARERAEQVDNTPV QQIINADNNEENENQDNGNNEEADALDPNNNTTETETEAEVHAV IINPPATEEEEEVAIAEARNRIPELQQAEEGCVTPKMTSLVRLKLLGG GELFNKLTPLATRYAATNGTNEADKIDVLTLCNAALREMIYTNH SESKFMTAENGEDTAPPPRISEFTRDRIQKAANEIKGTILIKVVKHISHA	

NeS	NeSL-2_CBr e	.	Caenorhabditis brenneri			<p>AAGTAGCTCGCTTCGCTC GCTCACTAAAACCTTTCG TCGCTCGCTTCGCTCGCT CGTCTTAACCCCTTCCGA ATAAACACTTACAAATCC CGGCTCGCCCATTTTTT (SEQ ID NO: 1295)</p>	<p>RCLKDSTRDEHNKFMVEMIAKINDRDNKFEQVNIIEIFQGPISDQSI LNVNTEDNNEFIKKMDYINRILGTPQDASPYARKKKLQACFADNPTKT LRNIIADKVPQQSLKPSEYLDYYPQWANEAGEYENFLHHDYALPE RYGQVFANDELDFTMTNESKIEVIRNKNHLSAHLGDPNSVYMLFPV SAAKFSLIFRSIIISGHIPDCWKLKSTVMLFKKDDPSLAKNWRPIGITS CTYRIFMTLVNKAQMIPMFHAMQKGFVRGATLSEHIAVANEVLCQ STRTQSEMFTQIDFTNAFTGTVPHQLIFDSLEAKKVPDSIINLLKLYK GARTAIYTRHAHSEIVPVRRGVQGCPLSPILFNCCDPLLYAVQRRHF EDGYRFQDKAGQYSIAQAYADDVLVISPTHEGMQRILNTVDEFQKI AKLVAPQKCVTLAKTSTAIQPFRIQPDPIKITSMDNITYLGIPISGTKT SRFAAATGILEKVKQIRVVFAHLALSKIALRVFLPQLDFYMFHNV FRVNDLKATDQMIRGLIDKEAPTSNIPVFFYMPKNKGFGGLVKLELR QPQLVLTFARLWLSQQAETKAFHHTMAQEEKFRKVVEDQENGFL GIKMEKIVQKNERSKRTNCFITQAAKAAKLEVRKEWDKGGIQV RGVGENATDWYRSKHIGQIPLIGRVIQQRQYEEFKKDETHSHTECEP AALAESHDIMKRPQAVPNLYSAAIALRTNTAPTANMFHNPEVL ANCLCGCQCTLFHTLMCRNRFSLYKWRHNIIICDDIYQFIHDHYP GVTIKCSARITSDGYQTTGPELDVTVDLLPDLVVYDEANKMIKIEVT CPYGTDDNVGNSLDAAYDKKVNKYSLAEQTERLFWNTTLLSIIVSS LGVIPLRTKLDALRISPADHIQLKRLSMHAIASACIVFEKVPEFFGM RCRPLPGRVTAPNAAIPNNNNNNNDTDHGQENQQATSEEQPTNN GNAQEDNGQGEQINNSTEQTISVDQIIEEDAENNAEQALDQDPDED EFLN (SEQ ID NO: 1417)</p>
NeSL	NeSL-2_CBr e	.	Caenorhabditis brenneri			<p>TAAACCGGCTCCTCTGGG AGGAGGTATGTCAGAGG ACATTCTCCGTGGCGGA TGGAGGAGTAGGGTAA CGACCGTCATTCTGGAT GCCTAAACCAACCAATC TGTCAGGCAAAAGTGCC CCAAAAGCACACGCGTG GATCGGTTTGGATGCCG ACTGAGCCAGAGGGCAA AGTCGAGGCGGTAGG CTCCCGGCGGTTGTCCG TCATAGTCAGTGTGCGC CTACACCAACTGCTATG ACACACAAGGACAACCC</p>	<p>MRYHXSXPAXRTSDNXWRSIXKDVRRPDPSTIEEKSRYNRSIGIPDS LXRSSAVRSXSXSPSPGPDVRLXNSPSLDDRRRLVDCETTLGSYRE WTDKPMMGKMTYAAVTKRAPRPQTGGARLSTNLADEMEIKYRD TNDIRLVIDLPNPHLIKPCCKSCISARGRGANALKYMKRHIADAHHL NADFYVKSCRCQEHENVCAGKWNHNLKRVHGYTLEDAVSTAKP STRQKIANAFNDSAPIDARKTSDVPEKKSREAGLEKFLAPTSEDRE KTPPSTRKSSSESEASIQSTIQTLESDDLTVQEIINISSEDEMEDEPP KRRVNVWALIHENGKDAWIDSLMVIFLESRRARGYESCIIDPLNFIG TDMSYLTIVRRRMEEGYKKIIFLCANDHWTLVITGTSTATFYDPMG NEPTETVKMIDELDLEMLQAPSNSPQRDSWNCGVFMKMAEAY IKDTQWDLTDVDTDFTRSLTELKAKNFIAEDIQTYRPPSRKALT RNSQSPVWVCHKCSRATPIQDVSRMEVEEAPVLPTPEEPQEWTF VGKNRKRGVTSRTPNTPSPEAKRPAPFPVPLKPSANRWHFPEEETEK MEVSSADEVNSTPPKPPKIPNLLAMKIASPVPLKRGNPSKKHKGH MMINTARKGPTKKEMPKGEPANLVKIRSWFDEQLKMYKDEGSNLQ</p>

<p>TAGGAGTAGGTAACAAW CTGTGTATGGCTTCAAAA GCATGCACAAACWCCTGT CAAAAAGTAWTTCCCATC MTGTAAATAGCTCAACGA CWKGAGMCCAATGAT (SEQ ID NO: 1173)</p>	<p>AAATAAATAAGCCAAG GCGCGTAGTTCGAG CTAACAAAGCTCCCGAGA GGATGGTGCACAGGG CACATCTGGGGAACG ACCGATCTTCGGATGC CCAAACACGCCAATCTG TCAGGCAACGTGCCCAA AAGCACACGTGCGGAGC GGTTGGATGCCGACTGA GCCAGAGGGCAAGTCG TAGGCGGTAGGCTCCC GGGGGTTCTCCGTCGTA GTCACTGGTGTGCTTACA CCTAACTGCTATGACAAG CGTATAGGAGGCCCGGA AAAAAAGCAAGGCGG CGTTAGCTGAGAGCTAAC AAGCTTCTGTTGGTGG GTGCCAGAGGGCACCAT CCTGGTGGTGGATGGG GGAGCTTGGAAACGTC CCGATCCTCGGATGCC AGACCACCGCAATCTGTC AAGCACCGTGTCTCCAAA AGCACACGCGCGGTTG GTTTGGATGCCGACTGA GCCAGAGGGCAAGTCG TAGGCGGTAGGCTCCC GGCGGCTCTCCGTCTA GTCAGTGGTGTGCTTCA CCCAACTGCTATGACATG CGTACAGGAGGCCCGGA AAAAAAGCAAGGCGG CGTTAGCATAGGCTAAC AAGCTTCTGTTGGATGG GTGCCAGAGGGCACCAT CCTGGTGGTGGATGGG</p>	<p>RLTWLSDSLTAAIGKAFNGNKYIVDQIIKRNPPPLVEKGAMSTQTSRK RDEFKPRERMAQEPNEPLRIQYAKNRQKTFKIGKQSEQQINIEVTE QHFRKTLAPVSENAIKTVCSIKKVLMPKTIEDPISSVEVKSILT VKV DTPSGTDGVKYSNLRWFDEGERLAKLFEECRKHREIPSHWKEAETIL LPKDCSDEEKKPENWRPIALMATIYKLYSAVWSRRISGVQGVSPCQ RGFQSLDGCNESIGILRMCIDTASVLNRNLSCSWLDTLNAFGSVPHL IRRSLEFGYPPQSVIQVTDYMYKGATMKVKTADQKTQSIKIEAGVKQG DPISTPLFNICLEGIIRMHQMREKGYDCVGHKVRCLAFADDLAILTNN KDEMQEVIDKLDADCRSVLIFKPRKCSLTIVRGAVDKYAKIRINGDA IRTMADRDYRYLGVKTVGGGRASETEALIQVVKELQKVHETDLAPH QKLDILKTLPLRLQLHLYRNATPKLSELREFENVVMKSVKRYHNIPKIG SPVEYQIPVKKGGGLVLSPLRLTCLTFTLSTLCKLWSDDPFISSHKDAL SRITVKAMGLTTQSATIKETCEYLNRKAVTKGGYSLFCRMNLSRLTS VIQGAPLKSMEFIPVNNIEIGIAVQATKDSEIKVFTKADSLKMSKLKDL VRSAMILKRFLEEKSVKSRVTQVLQHPQSNRFRVDRGNCSIAAQR VHPARLNLLSCNANTYDVNHPKGRRCQADFESQQHILQNCYSLA GGITQRHDRVNMNRILQEIENGRKAHYKIMVDMETGATREPDIME ERDGEVLLADVTVPYENGVOAVERAWDKKIEKYKHFLDYRKIGKK ATILPLVVGSLGTWYWPDTSHSLKMLGSDGQIRNVIPEICQIALESSKNI YWKHILGDSYKTVLEGLFCQRNNKEVRFEGKGEKHVSRFQPLKCEK VRTMKSTKEGRSRNAKGNWRRRSKSDGRSVSKGRYWRDPS NKPPHSKMTQSALAKR (SEQ ID NO: 1418)</p>
--	---	--

NeS L	NeSL- 2_CRe m	.	Caenor habditis remane i	CCAACTCTCATCGTATTAA CCTACGGTATTCACTCCTA GTGAGTGTAATAAAGGTT AATTACGTTTTTCTCTTTGCM AGAGAAAAAGAAAAATTCG AATCCTTTTTGTGTAATC	GGGAGCTTGGGAACGTC CCGATCGTTCGGATGCCC AACACCGCAATCTGCCA GGCAACGTGCTTCGGAW GGTCA TTGGTTCTAGACT TGTAATAGACCATTTGCC GGAAGACACACGCGCG GTTGGTTGGATGCCGACC GAGCCTAGAGGGTGCAA ACCTGAAGGGCGAGGTC GAAGCCGTGAGGCTCC CGCGGGAAACTCCGTC ATAGTTAGTGGTGTGCCT ACACCCGACGACTATGAC ACATAGGAGGAATCCTG ATCTGATATGATCATGTA TATAGGGAGGGCGAAGG TAAATAGTCAGKGTCAAA GTCCACGTGGCAGCTACT CCCCAGCATAGTAGTGAT GCGAGTGGAWCCAACTT TGACACTGATGTTCCCTG AGCTGACCCATCTGCAC AAATCCAACAGTGTATGA TGGCCACACACTGAGG ACGAGTATCACTTGTGAT ACTCAGAGGTGTCCCCA TGATCAACCAATATCACA GCTAGCGGACCTACCGT GAGGTAGACCCCGCGCG CTGTAGCAGGCTCGCCTC (SEQ ID NO: 1296)	MTNVYLKPVNDNQTNKTDGNSRNTMSNSQCEMTWKPVARTYAQ AASNTPADDKTVTVLGGCKYNLLKLGNTPQTSKRSPKPSRGGARISV YTLTDELEITHREEGKITFAIDLPNKNNILCPLCRECTQTRGRGSSTKH MKLHVKEKHQLDATFIYKSCMCNEYEPKCKGKWKIQTHLQKVHNY KYDESAIVVPVPPNTRQQIANELNNAAPFVDIRKPKAAAEEKKTEN GALLKFLTKSNKDDQVKSPSPXDIPDAESPEKETQALTIDPKGNNSPSKS
----------	---------------------	---	-----------------------------------	--	--	--

ACAACTGACAGACCT ATCGAATTCCTTTGTTTC GTATATAGGAATAGTCACT CTGGACCACGAAGTGAC AGTTGCGCGGACTTCC AGAGTGGAGAGAAAGG TGTGAAGAGAGAGGTCT AGAAACACTTCGGCTGTCT AGGACCACTTCCTGAGTG GAAAGAGGAAGGTCTAGA AACACTTCGGCTGTCTAGG ACCAGTTCGTGAGATCTCT CGTGAGAGTGGAAACA GTCAGCTGAGGCTACTGT ATTTCTTGATAGCCCGCC CCCAATCCCTCCCGCC CCCTCGACAGATTTTCTG TTTGACCTCTGGAATTTG CGAGGAGTGGCGGAGAA TTTCGAATTCCTCGCGGT TTTCTGAAATTTTCCAGA AGATTCGAGCGGAGATC TTCGAGAAAGTGAGCTGA ATTCGCGCGAATTTTCCG CGATTTTCAAATATCGAT TTTTGTCGGAATTTTAT TTCTGCAAAATTTGATTG AGTTCACGCGGGAGAGAA GGAATTTGTTGAAAGGG TATTGATTTTGTGGCGGA GGAACCTCCCACTGAATCA ATAACTCTCAAAGGAGAA CTCATCGAACAACTCGGG TGACCTGAATCTTGGGCG AAATTTTCGATTGACACA AGATAAMACAAATTAATG TKGAAATAAATCAGAAC AACTGTCAAAAGAGAG	GGCGTAGGGAGGAGTT GGGTAGCGACCCGGAAG TATGGATGCCCAACCACT GCAATCTGATCTGGCATT GTGTTTCGATGGTCTCT GTCTAGATCTGAAATA GAGCTCTGGCTGAAGA ACACACGCGGACCCGG TTGGATGCCGACTCGATC TGGAGGTTGCAAACTCG AAAGGAAAGTTGAAGG CCGTGAGGCTCCTGGCG GGAACCTCGTCATAGTC AGTGGTGGCAACACCC GACGACTATGACATAGTT GGAGGAATCCTGATCTG ATAATAATCATTTGTTCA ATAAGGAGGGGGATG GTAATACCCAGGGTCC GAAACCATCAAAGCAGCT ACTGACCAAGCATAGTAGT GATGAACACATAGACCT GGGTTTCCCTGAACTCGA CCCATCTGCACAAACCCA CTTTGTACAAATGAACCA AACTGATGAAGAGTTTAA TGATTTCTTACATCACAG CTAGCGGACCTACCGTGA GGTAGACTCCCGCCGCTG TAGCAGGCTCGCCATTG (SEQ ID NO: 1297)	SIRSSQSSASSVCQEIQEITISEDEPKGARPKPGINVWSLNETGKDA YIDTIMMAFLKMRVENCDSVNIIDPLNYQFPAVDLVPLIQRNLED GKKRVFPICADEHWTLLTSINGIAAFDPTGSRMSSYIEELVNLGLII PKEQDEQPRQRDSYNGVFMKMAEAFIQDTEWEMEVEEDVK FRNLLEELKPNVEIFAEEKIYNSPGKSAQSRPTSSQCAV/CPTCSR SATPMDVGNMVEVPVQPPQQTPKSREPEQDEGWKVGKARKR GVWTERSPNISPEAKQFTGTPEKVVSPGKFHPLVGETEEMEVTCDS PTKEPTTEPKVTPSLPAMKIASPEVTKQTSKKKKKKKKQXKKKAQ PPKGEPTKKAQPKGEPAKLEQVQRTWFDKQMSYQEQGNSIQTLTW IADSLTAAIFKANSNGKYLVDKITARCPPPLNNEGEMATQTSRRTEAV KPKDRFVKESNEPLRIQYAKNRAKTFNVIIGHSARCEIDINVVENHFR QTLKAQPVTEEALNTVCSGIIKAKVDPSIEGPISSGEVKAILAKIKDTSP GTDGVYSDLKWFDEGERLALLFDECRQHGIKIPSHWKEAETVLLPK DCTEERKKPENWRPISLIMATVYKLYSSVWNRRISSVKGVISDCQRG FQADGCGNESIGILRMCIDTATVLRNLSCSWDLTNAFGSVPHLIR SLAAGFPESVINIISDMYNGSSMRVKTAEQKTQINIMIEAGVKQGDPI SPTLNICLEGIIRRHQTRKTGYNCVGNDRCLAFADDLAILTNNQDE MQDVLNQLDKDCRSVALIFPKKASLTIKKGSVDQYARIKIHGMPIR TMSDGDYKYLGVQGTNGGRASESESLTQIAAELQMVHDTDLAPN QKLDVLKAFILPRLQHMVYRNA TPKLTKEFFENTVMKSVMYHNIPIK GSPLEYVQIPVKNGGGLGVMSRFTCLITFLASTLFLKWSDDDEYSSIHKK ALSRITAKVMGLTKQKATLQEQCEYLNTKAITKGGYSLFSRMNEAIR TSLVNLGAPLKMFIQFENGIEALEVQASENSQIKVFSKADSMKLVTK LKDVKSAMLKNFLENKKVSVQVQLQHHPSQSNKFVNDGKNXSIS QKFVHPARLSQLVCNGNSYSKDLPKNCRWCGYECESQAHLQHCTYS LSSGITQRHDRVLRILXEVIKGRKNNDYDIMVDTEPGPTRERPDII MIQKDGPEVLLADVTVPYENGVAIEAAWDMKMEKYSHFIDYFARL GKRAVILPLVGSGLGTWPDTSNRLMLGLSDGQIRNLPIDISMIALES SKQIYWRHIFGDSYRIVSDLYCRKQDQEQIRFGDEPMENVQVSDRFQ FKTREREKSEEEKRRSKKGTWGRGSKKQTDSDRSQSGKSNQNG FQSVGQGVSR (SEQ ID NO: 1419)
---	---	--

NeS _L	NeS _L -4_CRe _m	chrUn	Caenorhabditis remanei	<p>ACAAAAAGTATTGATTAAAC AACATC (SEQ ID NO: 1174)</p> <p>CCCTTTTCTATCGTATTAACTACGATAACCGCTCATTTGAGTGATAAAAAAGGTTCCCCCTCTCGCTGCGCTTACCCAGGCATCTCTGCGCTCTGGAAAGCGGAGGGTCAACTTGCGGGTCTGTGGATTCTTTCTATCCACGCCCATATTCTGTGCGAAAGCCATCTAGATCAGCGGGGATTTTCTATCCCATTCAGGCGATCGCTCAAGGCTGTTTATCGACACTCCTCTTGACAAAGTATTATTCTTGA</p> <p>CAAAATCTATTTCTCTTTATCGATTTCTCTTATTATCGATTCTGTGAAAT (SEQ ID NO: 1175)</p>	<p>TAACTGCCTTGGAAAGCCACCAAGCCAAAAGTCTGCAACTGATTTGAATAATGTATAAAGTAACATGGAACCAATGCCGATAGGTAGGGCGGAGAAAAATGACCTAGAAAAACAAAAATGCCAAGCCCGGATTTCGAAGACCTATAGGAAGTCAAGTGAATAGAGAAAAATCAAAACAAATCTCACCTATTCACAAAGGACTTACTGTCGAGTATCAGGCAAGCGCGCCACCAACAGCAAAATTAAGTCCCGATAGGTAGGGCGTGAGAAAATGACCTACAACTCCAAACCCGAGCCACGGAAATCGAAAGACCTATAGGAAGTCAAGTGAATTTGATGGAATAACAAAAACCAAAATTTCTCCATTACAAGGACTTACTGGTCGAGTAGAGCAAGCCAAAAATATCAAGTATGACGCAAAAAATGGGTAACTTTGGGCGCATCCAATCAACGGATACCTCTGCGTATCAGGCAAAAGTCGCCACCAACTGTACTACTCCGAATAAAACCAAGAAACATGATTTCCCACTCCGTTAAAT</p>	<p>MTKTEWSWRHRSRVSIVIVKIDTSDYANVRVHVAAADLSNEDGHTSHNNILIPIMKPSVDRFCQIQYPPRGYVYPHPQSQKHDAKPSRHWNFEAQPPYHNNHNGRRGRSAKPSGRPRPKPILQIESLAAHPQJPGDTASAVLYSDVNVNENKSGQKPPQSGSHRSRGRPTKPSVPVGEAEQETNSRIAPPIVKFKHDKHGWTTVQGSSSHSGRPVPSVPVVS EANRQQLQEGDFPLTTSESSQEEIKVPNYQRVSPILPSEEDSKLPTKSNYRAPKGRSRNKKYKQKPPQKQKPYQQLPYQPKVNNAPTDRMAPEQLKGGGGKTAHNDIEEMEIEEDTDEKIQVKIKIVNKLTPHHFVCMMTYPTDNIRCFVKGCTATSGGGWGAEDLKYLTVHIRQEHKIVEWYECGICGDLGGAGKHISKWKIPHMIRKKNRDAPTNFKMGSRSSGPKITELLEESAPSCSNPRKTLNQKTAITQVTPKLTGYQTRSVTKALSVLKESRQKELEVLREEKANAKQKSLHPFFTKAPHIDGVKPTVRRELSKMITPGGEHKGTKIPMVHTKRGILQIKINRAKAKAKPMHLDSTIEASQLDVTITDDDEDDNMTPMRRFNTWCLDHTTQEAWLTDVINWYLDKLCFNGEQYMLVPLVWLIYKMGGMAGVEQRFKSKKTCLEPCEADHWILLVFDENLCYANSLGSPNGQVKNFIQQLNRKLCSEFEVPLQKDSVNCVHVCLIAKSIVNGQFWYDDSDVTRFTNAKAA LKAQGYELFEAPKQIENPDSSHREDIKENSMEMCESLMIVATPQRS EAPMELVDTEPSDLESKSDRVVEDCITALSDVSEPRMTPEKSETPEVPVVEERDLDPKLESPKSDRVVEDCITDLSVSEQRMTPEKCEP EAPLVVECELERLPKDLPVTDRTVVAIPEAVKLEEKSEVIPRLMELS YTVPPSPVVEYTPYTHHTKPKVATCQMGKKRKPVTGKPDLELI QIVRQWFEKEFNDYVTEGRNFRQRLWLTNLLTAAIQKASAGDEETIE KIRKCPPEVRENEMSTQTSQRQKPTTTNQKRSRNTTQSDTQAN TYWRNRAKTYNQIGQDFKQCDIPALIEEFYKKTTSVTNVQPETLVKY TSRLPRLDIGKWIEDPTEQEVFGALKTKDTPAGTDGLRYHLQWF DPCKMLSSYNECQHHLKIPAQWKEAETILLFKSGDESKPDNWRPIS LMPTIYKLYSLWNRRIRTVKIMSKCQRCQREGECNESIGILRSAID VAKGRSHLSVAWLDTNFAGSPVHELIESTLSAYGFPEMVHVHVKD MYKDASIRVKNRTEKSEQIMIKSGVKKQGDPISTPLNMCLETVIRRHLL KESSGHKCIDTRIKLLAFADDMAVLAESKEQLKELTEMEDCTPLNL IFKPAKCAASLIEFGKVRTHHEQIMLRPIRNLNDDGTGYKLVGHTGAD ARTSEELISVTKVEDLVNRSALTTPPQKLDCLCTFTLTKMTMYANAI PKLTLSAFANVMVRGVKIIHYPVRGSPLEYIQIPTGKGLGVPCPRI TALITFLVSTMKKLWSDDEYIRKLYNSYLLKKVVEAETGIVEVSTKDLAEY</p>
------------------	--------------------------------------	-------	------------------------	---	--	---

LSNKVPSRKHEFGVNCYSRIREVCNGLALNQAAPLYKLEFIEQDNELA VVVQPTTEESKERIFTKDHVKKLQSLLKASVNDALLHRLFTTKPVKSEVV QVLQQHPQSNSEVRMGGKVSISVHVW/IHRSRLNQLTCNYNIFDPKQ PKNCRRCGYKNETQWHILQDCTYGWAKLIRERHDAVHHKVVVTMIC AGAKKNWGRKIDQELPGFTSLRPDICTSPDGKEVIFADVCPVYSRTR NIEFAWKEKIRKYTEGYSHLVAQGIKVTVLP/IAIGSLGTWWTPTNESLY QLGISKDIRSAIPLLCSTVMSEYKNAYWNHIYGNSTSVPLRYGHQK PDGDDWKKELSCEPVLALQQ (SEQ ID NO: 1420)	GCATCTCAACCAAGCTAA AGCGGTAAGGTTATCAT GTCAAAAAGGTGAGCTAC AGCAACCTAAAGCCCGA AAGGTAGGGCCGTATAA AAAGACTACACCCCTCCA AGACCTAAACCCACGAAC TCGAACGACCTACAGGA AGTCCGTGAATGGAGAG AAATATCTCACCAAACTCT CTTCATTACAAAAGGCT AACTGGTCAAGTAGAGC ACAAAGCTAAGCCTCCAAG CACGAAGTGATATGGGT AATTAGGCAACCAATCA ACGGATACCTCCGTGTAT CAGCAAAAGTCGCCACA AACACTGTACTACTCCGT TACTCCAAACACATGGA TCTCCTTCTCTACCAAAA AGCTTTATAACCAAGCTA ACGGTGGAAAGGACATC ATGTCACGAGGAGTAGC TACAGTAACCTCTCTTT GAGACTGCAAAAGTCGAG GATGGATTGGGAAGGCC GCGAGGCAAAAGCGGG TAACTCGCCAGACGCTA GTGATCTTCGGATCCGAC AGCCCTGGCCTTAGAGG AACCTGGGATAAGGAG CACGACGGGAAGGATGT TCGCAAGGATTTCCCTT CCCATTAGTCAGGGCTGG CAGTTGGTAATATAGCCT TTCTACACACCACCGTCTT GCACCCACTAAACCAAGTG GGATATGCGGGTGGACT					
--	--	--	--	--	--	--

NeS	NeSL-4_SM	.	Schmid tea mediter ranea	<p>CAATGTAGAAAGGTGTTCC CCACTGCCTGACTCGCCA ACTTTATATGCTTGTCAC CATAATGGCCCTCACTA TAAACTCCCTAGCAACTG GTGTCGCGGCAAGCCG GTTCTGCGCACTATTGCG CCCAGGCTCGCC (SEQ ID NO: 1298)</p> <p>TGAGTGTGCTACGAGGC AGCGCTGGTAATTGCATC GGCGTTGCAGATTTGTGT ACGATAGATAAAACCA ATAGTAATAAATGCTGAG CCTAGCTCGCATATCTAA GCCGAAAGGCAGCATAT ATATGAGACAAATTTAAAA AAAAA (SEQ ID NO: 1299)</p>	<p>CAATGTAGAAAGGTGTTCC CCACTGCCTGACTCGCCA ACTTTATATGCTTGTCAC CATAATGGCCCTCACTA TAAACTCCCTAGCAACTG GTGTCGCGGCAAGCCG GTTCTGCGCACTATTGCG CCCAGGCTCGCC (SEQ ID NO: 1298)</p> <p>TGAGTGTGCTACGAGGC AGCGCTGGTAATTGCATC GGCGTTGCAGATTTGTGT ACGATAGATAAAACCA ATAGTAATAAATGCTGAG CCTAGCTCGCATATCTAA GCCGAAAGGCAGCATAT ATATGAGACAAATTTAAAA AAAAA (SEQ ID NO: 1299)</p>	<p>MNVLDATIKSIGMNTKETTPNSQLRVETTPCTSTTTIMHASCNTTST ISYSLPSAVSLPESPASSITITDDNCDIETPYP/LPQTNGDLSEILKDIE ANKDITMSNKNVLDSCDSGDDRDIIENDRESMDLFSQSLLNTN QSDERREKNLTENAPTEITTEKSYFDIISKASDNTTSKKLLNVKNELTAG LPPMPVVTNTAKFIRNVREDIADPTLYRLDSRGKLGCRQYKKPGCG DIAVYDYEAIVEHAFFHTIPENQNVDCQPCPKKGDVHTVLIKY ADIFNHIEAHSHVWQTAITDNMKTLYRLTKENXFYCSVRNNKKNKCK KAFNLESNMMDITEHMKTHGTGSDXNLLNICYCGIWKPFTELIAHIK TEHLQYINSIPNKENIHNTTIVSPNLFAGILASGETQNPIDEIIPRD LPENLAFNRNIENELWSQHLVKAYIFSAYVKTSTIFINPYTCNALIQCN YKTFEFPKDFAKWNEVLPIHNNTSSWSFFFLNKKRVAMIIDPTA DSDHTLHFEATDILRTILNVQNIFFDLNPLTEVEYVPCHEANLSAFX VCHFLCLMSDLPIDIPDIDHMKETMRPIIRKYNCAKFPESDVNRVY LIEDLYQLNLDITICEEILCEIERINGRLNPKRYFKESKPKTDIHLQKKKS AELLVRLKQISQKTEIGKIWENDVDVDRPPMARFLKTFASQDCP VSNTSSINLPYYMDTDTDXCTDCENLSHIMKNLDSSAPGMDLITGGD WKKISPKHELITACNLCILRNKVCPEKWKLFRTVLILKPGKMSEFRAN SWRPLAIMDTAYRIFTLLNNRLLQWIRNGNLISPNQKAIPIPDGCAE HNATLHFAIDRAKRCTELHIVWLDIADXFGLPHDLIWTLANMGL KNETLILIKELYKDVKTIFDCQGTISEPVITKGVKQGCPLSMTLFLCSI DYILKSILTNYPFLHDLNISILAYADDLVLSDSYLEIKKSLESTVELAFA NLKFKPSKSGYLSINNVSNDILKHLNYNEEIPITISENNKYRYLGVDFSYK RNQDVGRLGSALALTRSLFKSYLHPAQKLNAYKTFHSLKLFSLRNCVI GHRILDCDRNRVTQGREKQLGDFQDEIKALLKTMIGDFQAXNNVPY THCKLGLGSLTSAIDYLIQSTIGITRLFHSSNLSFRKMLITELAHSGGK NFEAGLKWLNCEVNAKFPNTSFFVKFQKSAALAKRKCICVNLKFVED NFSLEMTYKRTSVNHNQNLSTLSKELHDFVGLYYAEQXCQMRVQG HIATAIGDSITAKYLIASDIINDAQYFLVLRARNLNLNINAYRLKYN GTKRCLHLDDEETQAHXFNHCRAPKNARRVKHENVLSIVAFLEKIGF</p>
-----	-----------	---	--------------------------	--	--	---

NeS L	R5	AY216 701	Girardi a tigrina		<p>CTTTTATTAAACAATAAAAT CGCAGTAAGTAATAATTCCA CTGTTATTAAATTTAAAAAC AAAAATTCCTTTAAAAATGC CTCTCTTTTTCAGTAATAAC ACCTTTTCTTGCTTTTATTTA CTATTTCTTGCTACTGTA CAAATCGAGCACAGTTATT GCAAATAGGACATAGAAA TTCTTTTAAAGTAAATTTA AATCCATGAGAAATAAAA TAAATCCTTTTGATTCAA AGTTTCTATGTTGCTTTCTA ATAGAATGGTGAAGCAT TAATGGGCTTTGATTTTTA TAAATTAAATATATTTAAT CTATTAAATTAATATGTTTT TATTAATTATTAATTTTTAT AGTGGGGGAAATTAATA TACTTGATCCCAAGAAATCA ACTGATGATGAAGAATAT GTTATTTCAAATACATAC AAGAAGCTGGAAAAACA AATCAATCGCTACA (SEQ ID NO: 1176)</p>	<p>TGATCGGTGTTTGTGT CGTATGATTGTTCCGTG TGTGCTATATTTTCTTT TTTATACTTTCAATTACCT CGTTGAATGTTATAACT TCATATGGAATATATGTA ATTTAGTTTGTGTTTGTTA GTTTAGTTTGTGTTTGT AGTTTAGTTAGTTTAGTT AGTTTAGTTAGT (SEQ ID NO: 1300)</p>	<p>TTGRNLQWQWCSRSIQQSNYSFKLSSTEVGELVEQSPAPLQSPQFSN NYNNLNNNNLYSLNTFNQSNLCLVNIIEFFPTQHLGLGDVNSGCI NYMNNYNNFDNINLNSVLSYNNYNHSLASPYTTNITEHADINM HVQEVNMQQDNNTQHAITQQVSLQATSLQHTLDEMIVQFNATVR LKKHKVAKIIFRGHNHRKDLPTLPAREQYKTKPKLAIREVLHRKTTATS SPSENAIKAFFSSYRPAELFTGQELLESSWFPVHPEDDFEERIPGRDQ IAKYIKFASKSAAGLDWITYEDIKLGDPSEILQPIFEYIVQNNICPSEK ASRTIMIPKPKSDYSDPSSWRPITTSAYVRLLMKYLTWELYNWILLN QMLSRSQKSLGKFEGCHDHNAMLNMLIQDVRRQTNPSNPINKNR LYIVFLDTNFAFGSVPLDTLMVVPQRFGLGTSALTLLKNLYLDNYTNVT CGESKIENVKLNKGVKQGCPLMLLFNIFINIIRAIIEAMPDVHGYPLG DMDIRILAYADIALISDSHKDLQEMVYKAEYIGRILGLLFNPSKCALM DIPHDKKRTPPILVNGEMIKCVGKADPYKYLGTFRSWFRKLDIKELLQ</p>
----------	----	--------------	-------------------------	--	--	--	--

				<p> MMMDETKLITESNLHPHQKHAYETFIHSQLPFHLRHSRIPESDFITNR KTNKTTNNSNDSEKSIQKAYDPESGQLFNTFALPSGCAKDFYITKD AGGPQLTSGLDEYLIQSIMYIFRLLGSEDPTLNSAIKHDLSHLNLKGFV NINFSAISIFNSNFTDRTDHFSLSRTEWARLQLARKKLSTLAIQTN VCLINGHLVLTLSLENNVLLIDSKEKGDVKKIHASLMGFLRLAHLRLQK HGWSKLLFSATTHEILNKRILNGHVYPYKIWYFIHRARLGLLPTKLFSV SNLCRCGGKKETMSHALVNCPPMMQTLINERHDALEISLVQILSSKF QGTVIRQKTYVNELRPDITMESDTQYVLVEVKCPEDTKMSFELRTQQ TTDKYNIIEILEDVHPGKEVRLVTFIVGTLSWGPQNSDFLRDLGFSK DEIDQVKTRIMLQNIINSCEQWKRFVQYAPTITPGPIPDASEDDQG TSDNGPTAATVQGPVIGDEEEELQYDSGLDESSDDPDPDAELLFT IDIEQYLSNVITD (SEQ ID NO: 1422) </p>
				<p> AAAGAGCAAAAAATAAGAA AAAATCTAAGAAAAACAAC TGAAAAATAAAATGAAAT TCCTATTCAAAAGAGATAA GATAAGAAAAAGAGAGTCT AAAAATTAATACCGAAAAA CTAACTGAAAAATATTACTA CTTCTGAAATACCACTTGA AATTGCTCCTCCATACCTT TACCTTCAGCAAGTACCTC GGGTTCTCAACAACCGGC CAATCCTCCAGAGACGCT ACTCTAAGTGATACGGATC TCTTCCTTACACAGGATGA TCCCGATAGTCTTATCTTT CTGGAAGTACTCAACCAAC CTTTGTTGACCTCAACCT TCACAGCAATCGGAACCTC CTTCAATACTGACAGCCA AAGATTTGAGCGGGTGA AACACCCAAAAATCATACT TCTTACAGGGATGACCTTT TCTACTCTACAGTCCTTCA CTACAACCTCAGATACAGGT TACGGTATAAGTGTGAC AATGGGGAGCAGAGGTTT CGAATTCCTTGCTAGGAATC TTGTCAGGAAAAACCAAGG ATAAGTCCCTCTTTATAT GCTGGACAAGTAATTAGA CACACAGTCTCTTCAATC ACTTCAACCCAGGCATACTA CGCCAATAATAACTGAT AGTAAAGGTAATCTAATTG AGTTTTCTGTGATAAGCC TTTTCAAAAGTATACCGACT GACCCAAAACTGAACTA GAGCAAAATTAGGAGAGAG </p>

NeS L	Utopia - 1B_CP B	.	Chryse mys picta bellii	AGACAACATCTAGTTGATA GAGCTCTTAGACATAATCA GTTACGGGAAACTTATATT TTAAATAAACTTAATAATA ATAATGGGGGGGGTGGC GAACATTTGAAAAAGGAAA AAGATCAAAAGTCAATACG GATGATGCTCTCCAGCAAT GATGGAGACAGAAAAACAT AG (SEQ ID NO: 1177)	TAACCGAGACCGCCGACC AGGGAATAACCCACTTC CTCCCTGACGAAACCAAG GGACGACCCACCCCATG TACTATTGCTACACCG ATACTGACTTGGACTCT TATACATTCCATGGGTGG CGTACCCGAGCCCACTTA TCCACTGACACTTTAAAA ACTCTTGACCCCAATCT GGGTCTATGCCGGTTATG CGATATGTATGATCTCT TCATCCTTGCAACCGATA CCTGTAATCCCTCATAAC CCAAAGCCTGACCCACAGT GTACAGTACCTCCCTCTT AACTCGTGTATATTAAAT TTTAAACATTAACTTTAAT AAAAATTTTAAA (SEQ ID NO: 1301)	MESPXIFEKIDAAALXIYSAAXLXXNSLSLSPXXAXXAXXAAASSTPQ KTQXKPIPTTLGASRKRRTTXKDEXIXWXKAPVDTXGXRSTRTT ALRDLTSRXNIXXALQEEEDPRRTPPXSRDQDAERRPAAPEKAATRG APPTIQDQDADRCAPGRDATGGAPRRPRTRMLXAAPLGRMPPEEP PPTTRDQDADRRPAAPERDAPEGTTSSTPDPTTYHPVRRRAAPRG THSXAXDLDAARCPGQRDIVASESTPPGATSPQASLPDXEESPAE SAGTTEVRPTEGEAGEDDCYLQYPXPTGLLLCPFCPLXHGVTLGAL SKHVRKAHNKRIAFRCRCDAPFETQKCKXHXATCKGPLTTAKVNP TDTLRVPTPTDGPASAPQAPSPQXVRGDDQPTTEGSAATPASRTD DATKRTSPASRIPTLDPAVRGITATSQVSDLTRCLSLIKTIRHNTDTRR XSAPPQVTSRCPAVGATSTAPQAARRDPANGGASRSPQIPRPDPAP GRPNTSSKVTQRSDRQKPHAPRTPQPDITTRRTTIPSAKSHDRA PTKPNVTGVSRTPLPPGRSSAASETPRAAPPHPHQDPRLKTHTLNTAPQ SEGQQGHRLSPQHLNPRQRNRNDGREQRVATPWQSAWMEELA KAEDFETDFTLMDRLTAELSAEITARRREPQEAASRATRRFPAPTRNNT AREGRRGDVGRYDPAASRIQKLYRTNRTKAMREILDGTSSVCAIQ PERLYSYFKDVFDEHAQTNLQRPCELLPLRINLTEDLERDFSPQEVQ ARLMRTKNTAPGKDGIYHLLKKRDPGCLVLAIFTCKQFHRVPRS WKSMVTVLHKKGERDDPGNW/RPISLCTIYKLYASCLAARITDWSV CGGAVSSVQKGFMSCEGCEYHNFLLQTAIQEAARRSKRQCAVAWLDL TNAFGSIPIHHIFATLGEFGMPETFIQILRDLYKDCITTTIRATDGETDAI PIRRGVKQGCPLSPIIFNLAMEPLRAISSGPTGFDLHGKISILAYADDL ALVADSSSESLLQMLDVTSSQAAEWMGLRFPNPKKASLHVHVGARAL VRPSRFIQGEPMASLEEVEVYQHLGTTPTGVRVQRTPEDTIAEILRDA AQIDSSLLAPWQKINALNTFLIPRISFVLRGSAAVAKVPLNKADSTIRQL VKKWLYLPQRASTDIYISHRQGGANVPRMGDLCDVAVMTAFRLL TCPDPTVRS/AQEAVRDVVRKRIARAPSEQDIATYLSGSLEAEFGREG GDLSSLWSRARNASRRLLKRGICGCKWKCCEERRELGLVPRIKTPDHTI
----------	---------------------------	---	----------------------------------	---	--	--

						VTPTARAMLERTLKDAIRCHYAENLKRKPDQGGKVFEVSSKWDASNH FLPGGSFTRFADWRFVHRARLNCVPLNGAIRHGNDRKCRKCGYAN ETLPHVLCGCKQHSGAWHRHNAIQNRLVKAIPPSLGKITLDSAIPGT DSRLRPDIWVTDAEKKVLMVDVTVPFENRSPAFHEARAKALKYTPL AETLRAQGYEVQIHALIVGALSGWDPHNEPVLACGVGRRYARLMR QLMVSOTIRWSRDYITEHITGHRQYHTE (SEQ ID NO: 1423)
NeS L	Utopia - 1_ACa	.	Acanth amoeb a castella nii	CCCGTCAAGGGTGCTCCA CGAGATCCCTGTGCTAGC CGACCGGTTTTACCACCCC ACCCGCGCCGACAAACA CGGACCTGTCTCCGACG AGGACCCACGCACG (SEQ ID NO: 1179)	TAAACACCATGTATGGTG AACACACCTCTCTCGAT CTTGATTCTGTGATTGG ACATCAGAGTTCCTGCGA AGGATACACTCTGCCAA TCTGTGGGTGTGATAA ATCCACACCTTCAACA (SEQ ID NO: 1302)	MAAKSVACPHDGCANKYSEASLRRIKHNKHATDEEGDTSWSCPH CHRFSTARGLSVHIGKSHRQAPPEPTRPPAPAPADPLDPPGPT VTPSRDDEDEEDDDPVEIADLSCPHCAQALPSAHLANHLRACK DHRVPAPGAPRSGPPSSRYWTAVEHHRVVEAMARFADHPDLLARA AAHIGTRYKQVDSHRTKVIAEREGRPVRTLDPTMDWRMRPYCAS TTARWLAEQGRSPVAPRSPCEPHAPPPAAALLYIPATPPAPTAPV APPKLAPPAESTVPATPDGNPEAPAPFSPAGPTPKALPPPPSRRN LRPHLVKDAWQGVADAVAPAAASRLRTPLAHLSTEQWATEAALA GLEATLHHAARSAAEAVPTRCASAREDAERQLREARKTREIFGKAAAL YAAGKDPATATIERIPPEVRLHLPDGSAAEWPARAAAARRVIRRAVARA DRLKRMGILDSRDQLRLFNANQKAVRQILAPSTKAPRCQLDPA VEEAYIQTAKPPIDPSPWKNVQWPRPPTAADDDGGSPFSAEVR AQLRLPNGSAPGIDGIPYEAYKRTKLDATLAHVFEVRLNARLPAR WDVARTVLLYKKDPNDTGNWRPISLQVTIYKIFTAALSRLISWAGK HNTFSASQKGLPAEGCHEHAFVLSVLDARRHKQNVYLAWYDLR NAFGSVSHDLIAWCAAMIGLPRYLDAIGAIVRHSALFVQVGDQETT GVIPMRGVKQGCPLSPLFNLCVEPALRCLRRTTGKFGTSTVEG QAYADDLTAAPSAYHAARQVATIEEWANWAGVSVVQALSADAP AGKCAALAINFEGGLMHSIDPALKVQGAIPAMSRNNVRYLGVHV GLTDALGQANELLEKASRDARTICASGLEPWQKVVAIKTILSLPFF HNGKIQRGRCQDFDRELRENLAALRLPVCTTNAFFHSRVASGGLGI LPIAEQVYLAHVFKLTSPDLIRAIARHQLAEVTHARHTTPVQD GEASFFGWLMRGQEVASTTPSGDVSSIWFAAAGAYSRMGWSVR DALHPTLVGPGVQFEGRFQRANV/PALRASAFSRHAEVWSALRTQ GRAAAVQHAVHPATHHWWHNSAGLITKEYRFAIKRGLLPTRAAP HHRNGPTACRACSYARETANHVHLCPCATKAIEVIAHNRICRALAQ AAEASWTSVLIEDVPIPGVDSPLRPDIYCSRPQGQCAIEAVSYEDAFN ASMEGRAKQKTDKYAGLAATVEEQLRLQTRHAAFFVGFSGVVLPA VTATATSIDLPKKTWNVLKRCVAASIKGSYTAWRERRRSTP (SEQ ID NO: 1424)
NeS	Utopia -1_AEc	.	Acromy rmx	GGTGACAAACGGATGCAT CATACGTGTACCGGAGCA	TAAATTATTTTGTCTTTGT CTTGCCCCCCCCCTTTTAA	VCSVRGCRREDRSRRFYKFFPLNFVKVPKTVIGSAFQKSSVSARSQN HSRSTRVPKTRQPRTSNTIGRYTAASANNYLTVIITGNTYVFAQWICY

			echinati or	TACGGGCTGTCACGGGG CTGCATGCGCATCTAGCT CGGAGATTTTATTATTA TTTATTAATTTATTTATTA TTTCATCGAGTGTGAGTGT CGCGTTTTCGCGAAGC GATTTGTTAAGTGATAC GCGCGGTTTCATAGGTT AG (SEQ. ID NO: 1180)	ACCAAGCAGGAGAGAGT GGCCCAATGCCCACTAT TATATATTAACTATTACT GTGATATTATTATTGA CTGTGGCGGGGCCCTC TCGCTGTTTATTATTA TATATTTTACTCGGTA CTTTTGTACTACTATT TTCTTTTATTATTAGTA TGCTATTTTATCTCTTC TTTGTCTATTCTTCTCT TTTTCTTCTTCTTCTT TCTTCTTTTATCTCTT TATTTATCTTTTCTTCTT CTGTGTGGGGCCCTGAC CGTCCGAGTGTGAATGCC GCGAAAAACAATATTATG TTTTATCGAGTGTGCAT GTGCGTATATATTATC TATTTATTTTATTATTT ATTATAATTTATTGCCGC GCGCCTCTCCGGGACT TTATTGTTGACAATACT GTGATATTTTCTGCKCA GGCTGGGGGGGCTTGC CCCGAGCCCTTAGTTT AATTGCTATGCGGGG GGGCTTTTGTCCCGCA AATGTATATATATA TTAGCGCGGGCTTAGC CGCTTTTGTGTTATTACC CCAGAGGGGAATTGTCC CTCTGGGGAATAAAT GATTGGAAAAATAAAGT GAGCTAA (SEQ. ID NO: 1303)	RECTWLSKFVNFLLTIGYFQLRLVVIYEGPVILDTFSNCGSLFMRG QXSKALLVRLNRSALAMADPQVHYIDYLPPLRVKVCFCGAEAGKV KGEYDPPHLAKHLKCHPGDTLNYKSCIDRLGTGKYPLRDVKAHYA ECHVSPAVDAAGPSTRGSLGECGAGQPTASRAAKATTRLAETVGGT DKRRAATSGSRQLTLPFAATPSPSTAAGEARAPRSXSTPTSRSPVA AVTAGPMSRSTTSTTARSKTVAKAAPNTTTTTARRSGEAAATR KPPTATVSKRVLVETVRLPVDDIQRAGVQNAAKPARAPSRPQR TSPAGGPRTTGAKEKCGEGAYKLPANSGNPISTRRATSRATSPVEKS EGTARRERSVPHPPKGIIDILSSSEEGTPYQGGVGRRLRRKKVT GPPKMTPREGVVTRARRSTAPVEKSALDARLTALDRTSSRATGNP TSQIAGGLYTRGQPERTPPRLPSPTTRGSPSGSLGEIRTPISPATS LPATLTCTVTTTTCGSPITSTGTGGVGRLLTPPSLPQTNILPTIGEEGT SPCVAVTTHPRPTGEDADCEAPQVSDHQSGISEPRRDTDTHLA CDVATGNAHLHGDDDLWKPHNIHGLHRWRGEADNTXEPPE HPPDHRGGRNVTRGGRHHPSLGGRSTSPILPRPTPEPERGQEE RLEGAAQPPTTPVVEGDQWQWTVSVRRRRRQQLNDTSPSN SESPTAGPSRSPRIAPLSALIAASTSRHETSLLNCTNGNICMDRTTP RNILPVXAERRRETSPDRVEGDIGYGAGKVS AEHPAPVNVGVM SRGRATASSVPPRANRGEGRQHHSRRRPDAPVGQPSRDHPAPAT VARQRRERVAARDALLDRAKDVATIADLEAFASVAFFGEDASAT GAAARADRSVRREAGARRGVKGERPEREGRPGSAPADPGA SGEARGDWVREAKRLQALYRANRRKAVREVLRQGPADQCQPKRQV QEYFERLYSGGDLGAGVEAERPDPSPREVSALVGLAEREVDRL RMNNSAPGPDGVSVDLRGADRGARLLTALYNICLRLEAVPASWK TSNTVLHKKGDRGMLENWRPLALGDTVPKLFAALLADRLTDWAVT RGKCSAQKGLRDEGCYEHNFVLEQVLTAKRSKRQAVVAWLDLS NAFGSIPHATIRRALIRSAPRGLIAWDSMYDGCCTTRVTAEGHTAPI PIRSGVRQGCPLSPIFNLAIDSVVRAAEXNDGYSLHGNWTSALAYA DDIALLAQTPEGMERMLASVEAAASVGLRFPNPAKCATLHVAGNG GRVLPSTFQIQGETINP LAQGESYTHLGVPTGFSVDQTPYAAVGDIVS DLRAVDRSLAPWQKIEMFTFILSLDFLLRGARVFKGPLTAVDLNIR RHVKSUWLNLPQASAEVGYMPPRWGGCGLPLSLADVLTVAHAY RMLTVRDGAVRELAWESLGRVVGRRIGHAPSCEDIASFLSGSLDGR MRGGGEASLWSSARNAALRQSERLSLRWVWEATEEMTELCRGP GAAIKIPPEARQGVNRLRSVAEHAHYASRLSKPDQGVKFEVSSSRV SNHFIRGGSTFRADWRFIHKARLDVLPNGARRWEANDKRCRCG EVSETLPHVLCGCIHSAIQLRHDAVLHRLWKATRLPGVVVNRV EGVSDDELGALRPDLVVRHEPSKSVICDVTVPFENRWTAFEDARARKI
--	--	--	----------------	---	---	---

NeS L	Utopia - 1_AMi	.		Alligato r mississi piensi s	<p>TGCTGGAAGACGGAGAA CCGCTTCTTTTCCCTGCG CCTGGCCTGGTATTGCAGT ACCTCCAGGATTAGCGCA ACTAGTCGGCAGACTGT CGGAATACAGCAATAGAA AGWGAGCTGACTAGCAG CTTGCTTTCTTCTCTCCGGT GCAGCATGGGTTCTCGTC AGTCMTGACGGGCTAGG GAAGCGGTGCTGCCAGT ACGTCCGAAAGAGTGCAG GTTGCGCAAGCGACCGCG CCTACTAGGTGAGTAGCC AAGGTCTTACAGTTACAC GGACCCGAWAACGCGAA AACCCCAACTCGGGCTAGT AGCGGAAGACTGGGTCC CCCCMGGTCAGAGTAGG CGAACGCCWKGCTCAGA GGACGGAACGCGGAAAC ACCCCAAGTCCCAAGGA CGCCTGATCCACTGACAA GAACGCTCAGGCGACGCC AGGAGACCCAGCTAGG GTGACCGCGACTGCAG GTCCGGAGGACCTCCCA GGAGGGTGGACCAAGCA ACCCAAGTTGGCGACGAA CCCTGACGACCCCCACG ATGTCAGGACCCCGACAG GCGCGGTGGACCACTGA CCATCGACCGACCCCGAGA GGCAGAGAGACTCTCAGA GCCCGGAACCCCGGCTGA</p>	<p>TGAACCCCTCTGCGAC CAGATGACCTTCACTTC GAGAGATTCTTCAGAA TGGACGACCCGCTCCAC CCGAAGAGGACCCCGC GATGAGACTCTATGGA CTGAGACACTTTTCTTC GAACCACTTCTCCACCA TTGCGGACCATTTGAACG GGTTGTGTATCTATC TTCCTTCTCTCAGCGTC GCGAACCCCTCCCTCCC CTTCCCTCCCTCCCTCCC CCACCCCGGGCTTAGTT GGCTAACATTGTATCTCC TGTAACCTAGTTGGGTT CCCTCCTCACCCCATCCC TCTATTGTTAGTCCCTCGC TCGGCGCTCTGTATTTTC CCTACCGCTTTGTCATC TTTTTGGATTCACAATCC TAAACATCTACTAAATAA AGTCAATC (SEQ ID NO: 1304)</p>	<p>AKYSP LAEELQRRGYRVVVFVVGALGSWDPRNEAVLRLRVGNQ YAAMMRRLIVSDTIRWSRDYVHVSGTRQYLAPSRPSGDLATPPRA VRRRWLAERSAQDAARRGSDSVSA (SEQ ID NO: 1425) CHHAGLRPGTNRTRRPDQTAPLPDRGHPMPNRRGSRSRPEEPS RREPXPACQGLRVWSPQQRMPTWQTLWLELSRATTFKAFE ASVARLTELAAAARPGQPRGNNRATRRDHLRQQRPRQRQYD PAAASRIQKLYRANRPKAVREILEGPSAFCCQPRETLFNYSRVFNPPA EAAAPRPAVEALTPVPAEGFEFADFTPQVEARLKRTRDTAPGRDG IRYSLKKRDPGCLVSVLFNRCREFRRTPTTWKRAMTVLIHKKGDPT DPGNWRPIALCSTVAKLYASCLAARITDWAUVTGGAVSRSQKGFMT EGCYEHNFTLQMALDNARRTRKQCAVAWLDSNFAFSGVPHRHIFT LRELGPDGVIDLVRELYHGCTTTRATDGETAEIPIRSGVRQGCPLPI IFNLAMEPLLRVAVAGGGGLDLYGQKLSLAYADDLVLLAPDATQLQ QMLDVTSEAAARWMLRFNVAKASLHIDGRQKSRVLDSTLIQGQA MRHLRDGEAYCHLGTPTGHRAKQTPEETINGVQDAHKLDSSLAP WQKIDAVNTFLPRVAFVLRGSAVPTPLKKADAEIRLLKKWLHLPL RASNEVLHIPYRQGGANVPRMGDLCDIAVVTAFRLITCPDXTVSIA ASALEETARKRIGRQPTRRDLATFLSGSEGEFSRDGGDFASLWSRAR NATRRLGKRIGCAWTWTEERRELGVSLQAPAPHADRVTVTPRTRTELE RFLKDAVRNKYAGDLRAKPDQGVDFVTSKWDSSNHFMPSGSFTRF ADWRFHLRARLNCPLINGAVRFGRDKRCRCRGVVAETLPHVLCSC KPHARAWQLCHNAVQDRLVRAIPAAAAGEISVNRTPGCEQMRPDI VITNEEAKKVIVDVITPFENRRQAFTDARARKREKYAPLADILRGRGY DVTVDALIVGTLGAWDPNSVSLHACRVSRRYAKLMRCLMVSDTIR WSRDIYVEHITGHRQYTDPTTRTAAGPDPEGTA (SEQ ID NO: 1426)</p>
----------	----------------------	---	--	--	---	--	--

					CGAGAGCCGCTCCGGC GGAGACCCCGGAGCTG AGGATGCCCCCGGATGA CGCGGAGCGCCCCGAGC GACAGCGGACCCCTCCGG ACCCACGCGCCCTCGGT GACGATGGGGCCCCGA ACGAGACGACCCCCGGA CCCCGGGTCCCGAGGA CGCCCCCGGAGGGTCTC CCACGCTGGTGAGGAG CCCCGACCCCCCGACAC CGACCCCCCGGACG ACCCAGGCGAGGCGTAG ACATGACAGCACTCAGTT CCTCCCTTCCCCCTCCCG GCGAAGCTGTTCTGCCCG ACCTGCCACCCGCAAGAC AGTACAGGTCGCACGGCG ACATGAACAAGCACCTAC GGCGCTCCACCAGCTGC GCCTAGCCTTCTACTGCGC CCTCTGCGGCACCGAGTA CGAGGCCCTGAAGTCCT GAAGAACCAACAGAGGG ATGCGAGGGCCACGGAGC CGAGAGGAGACCCGGCAC GCTGGTGAGTCCGCTGC CCCGGCCCGCGACCCA GGCCGCGGTGCGAAGGCC CGCAGACTGGCCACCCC GCCGCAACCCACCGGA CCAGACCTCCAGGGACCA CCCGACGGAGAGACCTGC CCCAGTGA (SEQ ID NO: 1181)				
--	--	--	--	--	--	--	--	--	--

NeS L	Utopia - 1_CM y	.	Cheloni a mydas	CTCTTCTTATGAATACTTG CAACACTGCACTGAAGAT GGATTCTCGGCTGCTATT TTTGAAAAAAGTATGCTGC TTTGAAGGTGATTCTGCT GCTGCTACCTTGAAGGA AATTCTCTGCTGCTCTG AGACATCCCGAGCTGCACC GTGTACCACCAACCACT GCTGCTGCTCCACAGAAG GTTTCTCGGACA (SEQ ID NO: 1182)	TGAGCGGTACGACATC GTGCATCAACTATGAGAA AGGACTGAGAGACTTT TTCCATTGGACCATATGA ACTGGAACCATAAACTCA CTGAACATTAATCTCAC CAATGAGGGTAAATCC ATCTCATCTGCTATCC ACTCATTACTCCACACC TGAACATAGCCATTATAT GAACAACATACCCCATAT TCTCAATGCTGCTACTTGG ACCGTTAACCTTTTACCC CCAATCGGGGATATTGCA GATTATGATTCTCTTACG CCACCGGATCCTAAACCG AATTTCGACCCCTTGAT AATCTGTACCTTATCCCT GATAACCAAGAACTCTTA TGCTTAAACTCTGTACCG TTTTTTTTTATTTCACAT CATCTTAATAAAATTATTA AA (SEQ ID NO: 1305)	MTTKKVLGASTTQTSTKGNKSGCKDPLRDAVPGRSWILRPACRD ITTRNIPAPQQQPPMESPTTLQDALARPSPTPAAAQVADAG GALAALHTIKRIGISVDWTSISPKXQRTSASPACASSETTQRDXR LLDARPAGPLDTPRPHQDEPASDTADAAGTLLQGNEDTIYLYPLA ADMUCIPCSPPQSFLHLLGVVTRHLKRCHSKRVAFSCALCSLPFETQKQ CKMHQVACRCKLKGTTQSPAPAPPAARRPAAPEPQRRKXTSQA VKPAPVARPAERDAIEKVPAASGNITQVLSRRPVSPSHVAKXISM LRLSASPPVQHVVPVPRISAPPRIARDPVAGRASAAAPQATLTPA AGGASTTPQIALRPTAGGASAMPQTLLPXPRRPDWRNQPRSHK APGLHRQTDQHGQVHSAGHCLREISRSSNRLGSSHSAAATHRR GGVPATPEPDRVSPITTSNAXIPPEIPPHQHPTEGNPDPRRQADHTA GSEPAPDEVEXEQPMVMVRAATPWQTAWTEELQAAASFFDFLL VDRLTRELSAEIAPRRSSNQENAPPAHRTAPNHNHTTRGARSRDAS RRYDPAASRIQKLYRANRSKAMREILDGSPYCTIPSERLYSYFKDVF DRIARNDARPECLRPLPRVDEAGVLETDXTPEVMARLSKTKNTAP GKDGIPYLLKKRDPGCLVLATLFCQCKFCRTSPSSWKKAMTVLVYKK GERDDPSNWVRPISLCSTMVLYASCLASRITWVSGGAISSIQKGF SCEGCEHNFVLQTTIETARRARRQCAVAVLDLANAFSGMPHHHIF ATLQEFMPENFLRVREYVEGCTTIRSEGETAEIPIRSVGVKQGCPL SPIIFNLAMEPLLRANSNGTDGFLHGERVSVLAYADDLVLTADDPESL QGMLDATSRADWMLGRFNAKKCATLHIDGSKRDSVQTTGFIQ EPIVLAEGQAYQHLGTPGTFRVROTPTDITQELQDAAKIDASLLAP WQKINALNTFLIPRISFVLGRSAVAKVPLNKADKIVRQLVKKWLFPLQ RASNELVIAHRHGGANVPRMGDLCDIAVITHAFRLTCTPDAMVRNI AANALHDA TKRIGRAPSNQDIATFLSGSLDGEFGRDGRDIA SLWSR ARNATRLGRKIGRCWEWCEERQELGVLPQIRSNNDNTIVTPSARG MLERTLKAIIHSLYVETLKRKPDQGKAFELTSKWWDASQPLPRRGLRH PFRRLAVHPPCPAQLRPAQRSRPPPREQALQEVRLRLQDPAPRPVQ LQAPLOSAAAQCHPEPPGESHTAPGGGRRELHPRYPASGTPA NHFLAGGGFTRFADWRFIHRARLNCVPLNGAVRHGNRDKRCKG YSNETLPHVLCSCCKPHSRAWQLRHNAIQNLRLVKAIPRLGEVAVNCA IPGTDSQLRPDVVVTDEAQKKIILVDVTVSFENRTPAFREARARKLEKY APLADTLRAKGYEVQMDALIVGALGAWDPNCNERNVLRTCGIGRRYAR LMRRLMVSDTIRWSRDYIEHITGHRQYQEV (SEQ ID NO: 1427)
NeS L	Utopia - 1_CPB	.	Chryse mys picta bellii	TTTTTCTGATGCTTGACT GCAAAACACCCATCCAGAA GATGGAATCTCTGTCAGC CATTTTTGAAAAAATTGAT	TGAGCCAGAGTGACATC GTTCTCCACTACGAGAA AGGGACCAAGTGACCTT CTCCGTTGGATCATATGA	MTQDDQADCCPAGKDIYETGAPPMTHLQDQDADRCPAAPERDAPEG TTSSTDQPKTYHPAVRRRAARRGMHLRAQDLDAARCPSGQQRDNV ASESSAPPRTATSPQASLPDPEESPGESAGTTEIRPTTEGEAGEEDRIYL QXPLPTGLLLCPCFLPVHGVQTLAAALSKHVRKTYNKRIAFRCSCDLPF

NeS L	Utopia - 1_DYa k	.			<p>GCTGCTTTAAGATATACT CCATTCTCTWKTITGKAA GAAACTCTTTTCAGCTT CAGCTATTCTGTATCGGC TGCTGCTTCTCTGCTCC CAGAAAGCTCAGMAAAA CCTATCTGAAGACCWCC CTTGTGCTCAGCGAAG ACCGGASCACCTGCAAG AACCAAAACATTAGGAGC TGGCTGAAGAAACCCCC GTGGATACCTCWGCAGG GAGACCTGGSTCCAGMAG GACAKCTCTCGGACCTC MCATCSAGGAGCAAGAAT ATCTCAACAGCTTTCAGG AGGGGACCCCGGAGAA CCCTGCCGCTTCCAGAA CCAGGATGCTGATCGCCG CCCCACCGGGAAGGATGC CACCGCAGGAGCCCCC A (SEQ ID NO: 1183)</p>	<p>ACTGGAACCATAAACTCC CTGAACATTAATCTCAC CAAATGAGGGTCAATCCA TCCTCATCATATATCCAC TCATTATAMTCCACACCC GAACACGCACTCTATG AACTTACACCTCATATC TCAATGCTGTACTTTGA CCCATCAACCTTTTACCCC CAATCGGGGATATTGCA GATTATGATTCTCATG CCACCTGATCTTAAACCA AACTTGCACCCCTCGATA ATCTGTATGTTATCCCTG ATAACAGAAACTCTAT GCTCAAACTCTGTTCACT ATTTTTTTAACTCATCT TAATAAAATTTTTAAATCT GTT (SEQ ID NO: 1306)</p>	<p>ETQKKCKFHQACTCRGPPTTAKVNPTDILRVPTLTPTDDLASAPQASP ESQIRGDDQPTGEGSVTPASRTDDATKRTSPVSRPTLDPAVRGTTAT SQVNNLTRSLDLKTRHNTDTRRCSAPPQVTSRPAVGATSVPPQA ARRDPANGGASRSPQIPQDPAPGRPNPTSSKVTQRASDRQKPHAPP RTHQPDAAARRRTTIPSASKHDPATKPTSGASRTPLPGRSSAASET PRAALPTTPGPPQDPPEHRSTVRGTTTRPQTVPAAPEPAETTQQUEER RPRARVATPWQSAWMEELAKAEDEFNFTLMDRLTAELSAEITARR REPQEAARATRRFAPSRNNTAREGRGDVGRYDPAASRIQKLYR MNRTKAMREILDGTSYCAIQPERLYSYFKDVFDEHAQTNLRRPECLS PLPRIDLTEDLERDFSPQEQVQARLSRTKNTAPGDKIRGYPLKKRDPGC LVLAIFNKKQFHRVPRSWKSMTVLIHKKGRDDPGNWRPISLCS TIYKVASCLAARITDWSVCGGAVSSVQKGFMSCEGCEYEHNFLLQTAI QEARSKRQCAVAWLDLTNAGFSIPHHIFATLGEFGMPETFIQLRD LYKDCITTIATDGETDAIPIRGVKQGCPLSPIIFNLAMEPLIRASSGP TGFDLHGKLSILAYADDVL TADDPESLQGM LDATSRATDWMGLR FNAKCATLHIDGSKRDSVQTTGFIQGEPIVLAEGQAYQHLGTPT GFRVRQTPEDTQIELQDAAKIDASLAPWQKINALNTFLIPRISFTLRG SAVAKVPLNKADKIIRKLVKKWFLPQRASNELVYIAHRHGGANVPR MGDLCDVAVITHAFRLTCTDATVRNIAANALRDATERIGRAPSNQ DIATFLSGSLDGEFRDGRDIASLWSRTRNATRRLGKRIGRCRWEWCE ERQELGIRVPQIRSDNTIVPTPTARGILLERTLKAIRSLVETLKRKPDQ GKAFELTSKWDASNHFLDGGGFTFRADWRFIHRARLNCVPLNGAVR HGNRDKRCRKCQYPNETLPHVLCSCCKPHSRAWQLRHNAIQNRLVKA IAPRLGEISVNTIAGTDSQLRDPVVVTTDEAQKKIILVDVTVSFENRTP AFREARARKLEKYAPLADTLRAKGYEQVMDALIVGALGAWDPCNER VLRTCGIGRRYARLMRRLMVSDAIRWSRDIYIEHITGHRQYQEA (SEQ ID NO: 1428)</p>	<p>YAPGYEAAQSPCGREPPRDHRRPRDACGSSHSPCLTTPRLLPET VSAEPCDDSQRTRYASPHKQARTLHDAEPRDASREHAPSCAEPCH RCQWTHWKDCPHSTNTTDDGPEGTDRCADTITSPATAACPQRSPCP LGSSNGCDETAPEKRQPAADLVHTAPFAVLVRAGFADLVRAGPFA DHHQDDDDLPHRSGSLGPLCSKQDKPKTHQHRHSGQAGNQHTD IPRAAPSRAAICLMAAAATREDLLRAATSLSEMAAANQPTRSPTG GGEPTSQRRGPQALADAARKIQQIYRTNIPRAMRKVLRTLTAVFS ACLRTHGVDPDLCKKSTVLIHKKDRDLSNWRPLSMGDTIPKLFAA VMADRLTAFLTNGGRLSEEQKQKFLQHEGCEHNFVLGVLEESRRQ GKDLVMGWLDLSNAFGSIPHATIMDAVAGMGPISRTIHHQATGA ATTAKTIDGMSEIEPIEAGVRQGCPSPIFNIAIERVLRIKTIKTVNAGYL</p>
----------	---------------------------	---	--	--	--	---	--	--

NeS L	Utopia - 1_Gav	.	Gavialis gangeti cus	<div>TTGCTACCCCTTCGTGGG ACCATTCGGTGCTCCGTA TGCATGGATCGTCCGGG ATGCATCCCACTAGGTCGC TGGCGGAATACGGCACAT ACGCTGCGGCATATCAGC ACATAACGGCGGCCACCC ACAAAGCTGTATTACATAC CGTTCCGGGCTGTGGC GCTGA (SEQ ID NO: 1184)</div>	<div>AAACATTTACTGGCCATA TGGCTTTTTTTTAA (SEQ ID NO: 1307)</div>	<div>LYGSRISPLAYADDLVLIASSPEEMRSLIRAAADDAIEAGLHFNPKKCA TLHLTGKSSRRRAVQTGFLVRGTPIPAMTEGDAYEYLGIPLGLKKNQT PRAAMEAIVGDIADKIDSLAPWQIKIDAARTFVAPKLDIVLRSGATLR APLRHLDTVIKKHKKWLYLPQRASAEVWVYPLKKGGAGILPSSILADV LTIQAQHRMVSCPGEVVSRIASEGLREAVKRKINREPSGDEMAHFLS GSTLSGETASFGDAGWSVRIMATKRAQAVHLGVRWAWRGGELLV ESRGQRNRPVATDSNSRSLIQRILRCAAQDEFLLINKPDQQKGVAKL STLTPVSNAFIRDGSTRFADWRFIHRARLGVLPINGAIRWVGSQDKR CRVCGYQLESVPHVLCHCMHHSNAMQQRHNAVMDRLAKAGSRIG TPRVNCRVEGVAEDMAALRPDLVWRDERSRKIVDVTVPFENGAE AFDNARGEKEEKYRPLAEALRAMGYQVKLEAFIVGALGSWDPKNER VLKTLGVSRYFAGLMVRRLMVADTIRWSRDIWEHVSQIRQFTLPSCGA PSN (SEQ ID NO: 1429)</div>	<div>TGAACCGCCCCCTCCG CGCCACGAGACCTTAC TTCACCTCGAGAGGATT TTCGACCACGAGACCC CGCTCACCCGAAGAGC ACCCCGCATGAGACTC TATACGGAAGGAGGCACT TCCTCGAACCACTTCTC CACCATTGCGGACCATG TAACGGGTGTGTGTAT CTATCTCTTCTCTCTCA CGCTCGGAACCCCTCC CCACCCCAACCCCGG GCTTAGTGGCTAACATT GTATCTCTGTAACTAG TCGCGTCCCTCTCTACC CCATCCCTCTATTGTTAG TCCCTCGCTCGGCGGATC TGATTTCCTCTATCGGCTT TGTCATCTTTTCTGGAT TCCCGACTAAACATTT ACTAATAAAGTCAATCT GTTCTTT (SEQ ID NO: 1308)</div>	<div>MSGPRQAAADPRPSTDPRRQRDSQSPERLITRAASRRRTDPEDAP RTTAEHPERRRTDPDRGPSATTAGPERRRRPDPGGPEDDPEGLPT LVEEPTPTTDPDGRPRRCRRGSAHVPLPPPCFAAVPDLPPAK AVQVAQRHEQTPTLPPAAPSLLLPLRHRVGRGEAPEEPQGMGP PRGREETRHAGEVRRPTTAAARRPAAPATPDQTSQDRPTE RPAPATPPRSAPRDPDPDTPRDPGPPGPPGPDAPDPRIPEPP GEPEPPGALQLPSVPGSPAETSAAQRMPTRQOALWLEELSRATAF EAFEASVARLTELSSAARGQPRRGADNGPTTRDRHRPQQRPR RQRYDPAASRIQKLYRANRPKAAREILEGPSAFQVPRETLNVFSR VENPFAEAAAPRPAETVETLTPVPAEGFEFAFTPREARLKRTRDTA PGRDGIYGLLKKRDGPKGLSVLFINRCREFRRTPAAWKRAMTVLIH KKGDPDGNWRPIALCSTVAKLYASCLRAKATDVAVTGGAVSRSQ KGFMTGECYEHNFTLQMALDNARCTKQACAWLDISNAFGSV HRRIFGTRELGLPDGVIDLVELYHGCTTTVRATDGETAEIPIRSV QGCPLPIIFNLAMEPLLRAVAGGPGGLDLYGQKLSLAYADDLVLA PDATLQQLMLDVTSEARWMGLRFNVAKCASLHIDGRQKSRVDS TLTIQGGAMRHLRDGEAYCHLGTPTGHRAKQTPTEETINGIVQDAHL DSSLAPWQIKDAANTFLIPRAVFLRGSVAPKTLPKADAIEIRLLKK WLHLPLRASNEVLHIPYRQGGANVRMGDLCDIAVWTHAFLRLTCDP ATVSI/AASALEETARKRIARQPTGRDLATLSLSEGEFRDGGDFAS LWSRARNATRLCKRIGCAWWTWTEECRELGVSLQPAHADRVTVTP RTRTFLERLKDVAVRNKGADLRAPKDPQGVDFVTSQWSDASHFMP SGSFTFRFADWRFHLRHLRNLNCLPNGAVRFGHRDKRCRCRGYAAETLP HVLCSCKPHARAWQLRHNAVQDRLVRAIPAAAGSEISVNRTPGCGES QMRPDIVITNEEAKKVIVDVITPIFENRRQAFDTDARARKREKYAPLAD</div>
----------	----------------------	---	----------------------------	--	--	--	--	--

NeS L	Utopia -1_LV	AGCV0 13581 06	Lytechinus variegatus	GCTAGGCGAGACGCGCGA CCACGGTCGCGAGGAC CCTCCAGGAGGTGGAC CAGCGAACCCGAGTCGGC GACGAACCCCGACGCACC CCCCCGCG (SEQ ID NO: 1185)	ATCTACTATC (SEQ ID NO: 1186)	TGAATAGCATTATATTG TGTTCCAAACAACATACT CATTATTATCTAAACAT TTTTTTTCTGTTCTGAC AATCTACGTAAGTCTGC TAACCACTGGCATGATG AAATAAGATAAAATCCCC TTACACATTAATTTCTTGT CACATCAATGCTTTGT CAAGCAATGTCCTACAT AATATCTCGATGTCACCC CAATTAATTTTACATCCTT CGGTAACCTTTATACCGT TGGATCAACATATATGAT TTGTAAACTGTTATTTCT GAGTTTTTCTATGCTAAT AAA (SEQ ID NO: 1309)	TLRGRGYDVTVDALIVGTLGAWDPSNESVLRACRVSRRYAKLMRCL MVSDTIRWSRDIYVEHITGHRQYSDPTRRAAAGPDPEGTA (SEQ ID NO: 1430)
NeS L	Utopia -1_NVit	.	Nasonia vitripennis	CCATTCTTCGTACGGGTT TTCGTGCGCGCATAGCCG GGTGGAGACTCGCGCGG GGGAGGTCATATCTCACC		TAGTGGGGCCATAACAC CTAGGCCCCACAGTGTG GCGATGTCATGTGTGTC GTCCTTACTTATTATTATA	MSCPREGSDHLGPDPEPALHQQGSDIRVTSSRLGSRGKSSRQPPSSR HQVPASEASATAQQTAAANECQVCGSSFATSSGLRRHMARLHRAAS ADPEGAAPASITEIFDYPLPSRWKCSACSENFNQQTILKRHQTRHHP ATTFAFRCCSRSEFDSARRAANHWQVHKKERSQLSGTEPQASS QARVSMASHPPPLPNTSWAELASNPAPISFVWESPKNRPSVEEFG SSLPDVTMMSSQPPQVQSSVPALTPLSPAATASSPPGAARQLT PPTQINTPVTQRARLQPEADVPELPSPVTEHPVSDAQHWVDVSS ASDWSEFEAVCDQFVIHAVAVSRPNLARPQQQDRQSGDHPPRQQ RGQHRPTFDVREASRIQKLYRTSKKRAIRHILKEKSPFSGESDVLDF REVSAKEVDEEAVGKLASSLFDVPQDDSDATSLSLPSAKEIGARLSR MTNSAPGKDRLEYRHIRRADGFSFISEAIFNKCLAEGRIPAPWKTASTI LLHKAGPTDDPANFRPILQSCLYKLFMAVLADRLTKWACENQYLS EQKSARPCGCEGFEHFLSAAKDCRRNQKTCIGWDLRNAFGSIPH PVKIVLSSLGVPDSLVTLLMDAYNGASTFTLTGGQTDTPIRSGVKQ GCPMSPIFNLAELIIRAVKKNASDNHGLVTVQGKNSILAYADDLVL LSRDTEGLQSLQVAGSSASTLQMQFKPKQCATLTLDCRGTNVRQS AHHIQGAAPSLTTEERYRYLGVPIGLPRLTSLQESSRKLSSDIETISSLL APWQKLDAIKTFVPIPIQLYTLRATEYLSKDLKPLRAAIKHVKICHLVPR SSNAFVFASRPSGGGLAFVDPGVDAADILVVTQAVRTLASDDDTVRAVA LGQLTSVHRTVHSAPSDDCIDKFLSGSSEGPLANSNGSQASSLWS RTRAASRLKIRIVGASSGDIKVESGGRAIPSKKVTAGLRSDHHEMS EKLRLPDQGVKVARALSLDSFANATSWLTSGSFIRFCDWRFHRLN CLPTNAAVRRWKQNANTKCRCDHQLETLPHIINNCRPNMVPPIRR HNSIQERLVKAIHYGDIYQDQHVPGDPNPREPDTIVVEGNKVTIVDI TIPFDNGPDALSTAANAKVMKYDTRLQELASRGMDVEVHAFVIGSL GSWHGDNERNVLRGLISRRYRTLMMRLCCIDAIGSRDIEVIEHVTGHR QY (SEQ ID NO: 1431)
NeS L	Utopia -1_NVit	.	Nasonia vitripennis	CCATTCTTCGTACGGGTT TTCGTGCGCGCATAGCCG GGTGGAGACTCGCGCGG GGGAGGTCATATCTCACC		SGXTGREVKCTVNVLMQEQPHITKAIREGDFVILLPQSDDETLCCPLC VGRGRYSGKTRVECLNRHVKEVHPDLTTTFCRWGCFFAAPGDKKYP RKIVTQHCATCVPVESSAPSGRVDGERRVNTRRLGIAAATEASPVRR TRRNLASPPVEQNIQSAAPEPARVPQHPHPEIVALGESADDEVFRSP	

				ACCATCTCGAGCTTTTG CTGCACCTGA (SEQ ID NO: 1187)	TTTGCTCTGCTGAGCTGT TTAATCTATTGATGAT GTGTTGTGTGCTGCTCC CTCGTAGCAGTTTTATTC CGTCCAACACAGAGTCC ACCAATATTAAATAAGC ATGGCTGAAGCAGGCC AAGCGCGTGTCTAACC CCGTTTAGGGGAAGTTA CTTAACCTAAAAATACAA CTTTCC (SEQ ID NO: 1310)	VNSPRDWRAAAPQQAASSPXAVPGITAATPSNTRTNGSAXSIL AEHIPAPPTTTEANGRADIPRSGRAPPGXQAARRRATTEQRR IVGLLEAATREQLEEAATQAMILFARLTGRPEPRNAIRPGXRQHP AQGDVQAQAPDRIXEAKKLRLYRTSKKRAVQKILAGPXMNCQIDK NTITAFVELAARRDGGEDWPDVDFREEPTAASGEALCTPTREEVFR RLKGRNNTSPGDGITYRDLAKAXPGAHVLAALYNXIWRIEATPALW GVSNTLIYKKGDAIDISNWRPISLGDTPVKLFAAILADRIKRWAVA NGRYSASQKGFEGECYHFNVLQEAIREAKGGRKELVAVWLDLAS AFTSVPHSSILQALEGHGLPSKARNIISLYTGMTTRFHTAEGPTDPII QSGVRQGCPLSPDVFNLTLEVLRLEIQTGEGYTIERRISHLAYADD VAILADSPAGMRRLFAAERGAVGLTFNPAKCATLHIAGRGEEAV RPTFSVQGTPTVRALASGEAYEHLGIPTGYQVRQTPINTLRDLADIGS IDRSLAXWQKLDVAVGTFLRLDFTMQGAHIDKGFLEADKIIKAA KSWLSLPQRASAEVLFLPPSQGGGGLTVAHSYKMLYSSDVTSTIAG STLRTVSERLKKNASNIDARFLSGDLDLPRSTSPSTFWTKVRSALRI KTKLGLRWSWCQGGEVLLMACDPRAPGTRVSPQTKHLVTTSLRRC LNRHYAESLLAKKQKQKVFVTRRSQSNHFLRSGSFRFCWRFH RARLDVPLNAAKRWQGMKRCRCGSDLETLPVLSHCGPUSA ARQKRNNIQDLRLVKAASRCPGTISVNQTVVGVGRGPDAAALRPDIV RDDVNRRTIVDVAVPFENRLEAFDGVREAKIAYTPARQLTDSGYT VTVEAFVWVGALGAWDPNERNVLSLSISRYAILMRRLMVSDTIRWS RDIYVHVSGIRQYRE (SEQ ID NO: 1432)
NeS L	Utopia - 1_PCa	.	Phytophthora capsici	GCCGCGCGTGAGTAGC CATGTTGGCCACACCGCC CAAGTCTCCGCGCAGCT GCGACTGCTGCTGCTATC GAGTCGCACTAGTCGAG CTGCACAAGCACACCTCC TGCCGGACGCGCGCGGT GGAGCACACGCGCGCGC TGAGCCGTACCAAGACCA AGGNTTCCAGGCTCGGC GCGGTGGMGCTCCAGC CAGGACGCGTTCAGCAG CAGGTCGCTGCGGCGAG CTCGACGCTGCACGGGA CGGACGCGGCGGCA CCAAGACCACGACGACCA	TAGACGGCACAGTTCTG GCCACGTAGGCCGAAA GGCCCCACCCATGTAG GGAACCGCCCTCGGGAA ATCCATTCCGGTGTCGA CTGAAGAGATGCTCCTTC GCCTTGACGAGGTACA TCTCGACAGTCGAATTC AACTCGCAACATATCCGA TACAGTTACAAACACAG TTAGATATCAGATAGGAA CCTTCTTTAGGAAGCTA ACGGGTACACTGGATGG TAAATACACATACATTTT (SEQ ID NO: 1311)	MVVSRTARLEATPAPRWDPPLPRRVIASRIADRLVPATAPCRSALNA AFSPSRDVTFTQEDRQLEPLTRHVDEETKDESLPGRAPT/LDEE SKNDATAGEWLLRFDGACRANPGPGGAGAVLFPNGAASWTCSH FMPGATETNTAEYTAFLGARAADHAGATLLRVQGSQVLVRQVK GIYGAKSTRLLRLRDVRAELARVGQFSLHHIDRDNAHADRLANRA LDMKSTLVECATHPGRNACTTTLTSAAESPASPPVPGARDTPMA DAGEERLADVDDGEVYAAWRLPGPEVPERRRRLRLRLQSLDELEAAS EMVERLGAALSAKITDAEDWASAEYITALPYMLYDKLQSYQAAPRG PQPVLTRSPRGDRPASSEPNSSTTGGVASEHQPRRRRRRRRK GRRRRNPRRSREGREGAGHQQHKKHPRPRPRETQHHRHLDEA LDELHALERTDPHNRAPIAKARRVGRISAINQQLLRHKFDTEKAC VDGILSTARAEARAARATPSPASGAPTTTVPAGVITNDDGTCPIP SDKLWRHFDVANTPRLDFAEAPGSAFAAAMDHLPAATRLDLLK EAPSTDEIETQLQHVKAASSPGLDGVGVYKRFITQLLPVLRFAAFRC WLYKKVPQSWKLVGVRLLHKKGPREDPANWRPCLQQAQIYKLYTGIL ARRLTRWMDANDRHAPGQKGFRAVNGCGEHNFLAATLIDNARRK

				<p>HRPLVEVWYDFRNAFGSVFALLWDSLQRLGVPPDYVDMCKGLYN QASFVVGNAVDSGTAPVEQVRGVFGCPLSQLFNAISPILYALRR LPDTGVQLSSVDRPGASAYADDLKIFSGTKAGITQQHELVAFLRWT GMQANPAKCRSMGVRRNTNGAVEADNVHLEDDTPIPSMTHMQS YTYLIGDGFHVRRRVELAPKLTILKHDVTALVESGLAPWQVVKAV KVLYPRVEYALHRLRPDDQHLIESFDLHRRGLRHLRLPKNATNEFF YAPVSRGGLGLPLVELHAALQIAHGWQTLHSPDPAIRRVAREQLYQI ADARHLDDKHWPHRREELCELLNGELGTSAHAPKRRNGDIGSL WVDVRKNLKTFGKIVATAPANQETGVPAPQLQLRVPHHAEWLDHG NVLRHVKLHIKNIHWWQWCAISDQGKTARVHGGVGSFLLTRPRGM WESDYRFVAAARINVVDTVNTLSRRRLRAHDCRYPCACRWKETLAH VLNHCPTMIDAVGRHDDALKEIHTLRASSGDRRELVRNQTVPL PGPPLRPDIQVYNHDKRTVAVVDIAVAFDEQPSDEPSSGLAKAVQI KKAKYAGIKEHLENQGWKVHLSAIVYGSLSVAASNHKVYTEHLGLL KRDAKRLDRQLSSACIQSSRRW/NFHCAKHRAHQHEHQAPPSQATR GRRVTETGGNPSRTDRR (SEQ ID NO: 1433)</p>
				<p>AGCGTCGCTGCCCTGCC ATGGACGTGACCAAGGT GCTCGGTGGCGACAGCG GATCAMCACCGCTGTCG CCGCCACCGAGCTCCGA GTCGCGGCAAGCGCGC CGCTGAACGACGGCGAC GACGAAGACATCCGCGAG CTGGCCGAGCTTCTGCTG CCGACGAGGAGGAGGCC GACGACCAAAACCGCG CCAGGTTACCCGCGACCA CGGCTATCCGGCTCCGT CCTCGCTGTGACGCGCAC AACCGCAGCGCTTCAACT GCACGTTGTGCGGTGACA CGGCTGCCAGCTTGGCTGC TCTTACGGGACACAGGGA CTCTCGGACCGGCGCGT GACCTTCTGGACAGGTTT TCGGCGGTTGCGCGTGC GGCAAACCTTTTGGCTCGA GGCTGGCGCGAGCAAGAC ACGCACAAACGTGCGCCA GCCTCAGCACCACTGGT CGCGGTTTGCAGCAGAGG AGGAGCATCAAGCCACAC TGTCGTGCGAGCCAAAC CACCGTGGCCACGCGGT CACCGCGAAACCCCTG CTCCACCATCAAGCCTCGG AACTCACTGTGCCCCCCC CACGTGTGAGTCCCCCGA CGTCGATGTGAGCTGCA CAGTCCGCCACAGGAAGA TCAGCACGAGGAGCGCAC CCANCAACCGGAGAGCAC GCAACACCAACCTCTGAG</p>

NeS L	Utopia -1_PI			GCAACCCGCTGGGGTTCTG CCGCTCGCGCCAGCTCG TTGCCTCAGGATTGCTCA GCGACTCGGCGAGCTGGA ACCTCCAGCTGGGCCC ACCATTAACCCCGCGCG (SEQ ID NO: 1188)	TAAGCTGGTTCATCATGAC CGACAGGCGCACTACCA GGTAGGGAACCGCCCTT AAAAAACCCAGGAAGAC ACAAACCCCTCCACTTA GTGACATACATATTTAG CCTAGATTCAGTTACGG AGAGGTTACTAACTGGTA AATACGAACACATATTCT GTTCTAATCAGTGTGAAA ACTGGTTTCGCCTTTTG GCGGACTTTTCACTCGC ATTTTGGCAATCGTCT GCGGCTAGCTTGCTAGC GGCGACGAGCGGTCTC CGGGGGGTTTCACTTTC CCCCGAGGCCCACTAC ACCGATCTTCTACACTT TTCTAATCGCCTCCGTCT TCGGTCTTCGGTTGTCGG GCTTTTCTTTTGTGACCA ATCAGAGCGCGCCATGC GCCTCTTCTGGCCAATCA GAGACCGGCGCCTGCTCT CGGACGCGAGGCTCC ACGGCAGCCAATCGAG TCTCGGAGCGGACGCGTC TTTCTATAGCGCAGCTGA CGAGCGGATCTGGCGG CCCCGATTGTCGCGACT TTGCGCCAATCAGCGACG	MLADPAALAAGLARPPPPAPQDPSAPFAGPAGQNPRAAAPAR VEVHTVAPPGRAGGMLPDPGLVEEPIQATYAHDAQAQFEALCPYV AESMAVLVQHRRSAHRGTRFKDIFTSGCCSLVFYARIVAASHAVAC ARRNQRAVPPAPTVPAPTRPEATPQPTYLAAAMTAAAAASDVT VAAATNMQSAVPAAAKTTGLVPELEPALPQRASCHAGKRRRLN ADEAVTPCTPTARVSPQTEVAMAPHADPQDDTVLQREAAEPQDP AATPGAQVQRVEDTTAAQDDTVQQDHDADTAQVSPRRTPTRWG PRPSTQEPSMTGEPATLAARRPLTPAATGTRATRWGPGCHRAIG AAAIARLVLTGLSTEPAQRRQPPPPQPEPPSQPEPLAAAAATAADIA ATVAADIAAANAAMVDVGGPAADETWLLRFDGACRRNPGGG AGAAIFAPSGAVVWTCSHYMPSETNNTAEYALLGVQSAVHHG ASHLEVEGSSSLVIAQVKGTTACRNARLQRLNRVRHALRSVDTHKL RHIDROANAHADRLANRALDQRTSSECGTHGSCMDSCLAAPTALA AQETPPAAPSTSATPAEGNAMDDIAEIAARDEGETFPVLPIGPGS APERQRLRLQLSDEERDAADALQELADTMASKIEDADSWTSGE GYISSIPERIREVLQPVATAPPPQGRSRRQRRRPPRVTRNQREHRLD EALDDMAATQATPRDQRAVRRARRVGRVRRASMAQQEILRHEFA KDESKVAKILKTASTETAEEDEHPETCPIDAAATLHAFTGVNAPRTD FDYDATSGREFRAAMSDLPATVEIDAFDAELTIDEVEDQLTAAKTS SPGHDGIDYIGYSRFAAQLVPLLHAVFQFCWRHRRVRLWKGIVRL IHKKGDPQPTNWRPICLQPTIKYLSGLLAHLRLSRWLEGNDRPLPMA QKGFRAFNGCHEHFMATLLDQTRRQHRKLYQVWYDLRNFAGSL PQQLMWRLRLHLGVDGFDRCRDYRDSAFVANAADGATDPVR QEVGVVQGCPLSPLLFVAALVPLVRLLEKLDGVGPLADGVPRCTTA YADDLKFSDSAAAGIRKCHDTVAGFLAWTGLRANPGKASLAVTTN ARGNPTRSSMRLEVHDAAITTLHESRYLGVGDGYDHRHRLQL EPKLKQLKREAVALLTSLAPWQVVRALKVYVYKVEYALRHLRPLQS QLQAFDRWSKGLRHLLSLPRSAETVLYAPTSSGGLGLQPLVLEHRL QLAHAWQMLHSDKPAIQAVARAQACQVVRKRYRLQEDHWRGRD DELVRSLNSELAASPHAEVLRNGDIASLWSDVQRWLRIRHLRFEHC DETEAHGPLSRVPHHNKWLTHKTVLRHVKLHLKIRHQTRWKGMMV
----------	-----------------	--	--	--	---	--

NeS L	Utopia - 1_PMi	.	Patiria miniata	<p>CTGATGTGGATACCTTGG AATTACTCAACCGTGTGCG AGCTTTTGTCTTTTGGCG CCAACACCTCATGGATACC ATGCTTGTGCTGGAGCG ACGTTACAAGCGTGAGGG CGCCCTCCATGCCGGACA GCTGGTCTGTGCC (SEQ. ID NO: 1190)</p>	<p>ACGAGGGGCAGGGGTT TACACTTTTGCCCGCGTTT CGACTTCAACTCAGGCC AAATGGCGATTGAC CTCCACGCGCGTGCCAC TGCTCGGCACCGCGCGC GATTACGCGGTGCAAC TTCGGCACGTGTCAAC ACATGCAGCGCCCATTCG ACGCCAAGCGGCATCGC GGGACGACGCTCGGCC GCCAAGCGCAGCCCG CCCTTCAGCAGCACTC GCGCCGTTTGGCGGATC GCCATCAAGACGTGCGA GAGCCAGCGGGGTGCG GCAAATATACTTACTCT AAGTATGCCCGAATCCCT GCCCTCTCAGGCTGAACG CGGCCCATACTTGATCT AAGTATGGGAGGATCCC TGGCCTCTCAGGCTGTAC GCGAGACCC (SEQ. ID NO: 1312)</p>	<p>DQGKTVRHGGVGAKFMTTGAAGLSDDDDYRFGVKGRNLNQ/DTNSVL KRKRLRAHTTCRDPCTSSAETLAHVLNHCESNMDAIRQRHDDALEQI GSKIRGALDRAKSPTELRNLQTVPEYTGAAALRPDIVLNVAAKTMVIA DLAVTFEDQAARARHSSLQLSHDHKTLKYQPIVAELQHKGWVRVQTA AIVYGTGVSQPSNFKAYTEKFKLHKREARQLDLQSSHCIRASHRIW GWHCRQHRDRQRSGTASRASRGSGGAPRRTSQAPARR (SEQ. ID NO: 1434)</p>	<p>MCLKSFSTSGLRRHIMARLHRQSPDASTPTMTVEVPYLPKVVWP CVVCRENFYHNQTLKRQKNFHPNVDLTTVYQCSVCGQEFVTGRKA SFHFVHRRMSASAIPTPSAMPSSPMDLIRGLVGEPLPPSPARTPPPL ARYISAPRSSFPWNPPSPRPTLPRLPTPPRSPSPRPSPTPP PVTLLTAPVTEPAVPVALTTAQVTEPSAPAVHTAAPVTLNAPVTEPA TPATDPAVTVTRLHSPVTHISCSISFTASHAPYCAAPTSVVYACSPRR RQCSSTIAAVCNSEASSGNPCLLALPVHRHLLPDTSPQRPGLLHHHPG IPPHRPGRPHCHGSHLHRHDHRRARRPGRQHLLPRSPPPRSPSP SRSPSPRSPSPRPTPPRSPPTPRSPPTPRSPPPRPPILPPRSPDSTH RSLTSHARSPRATPIPVQPRHLRPTPAHPGVNDNAVPPSQSAID VWLAELSRSADEFEDVCDRFVEFAAAEGRNNRPARPAHQPPRD RGNQGRPQRP RP RHP RGLGPEFDAQEASRLQKLYRTHSKKRAIRTLT GSDVRYSGYRGPHPALMSDAAATEVLTSGFLSEVFSAREVDTDIAT DTSLLFPNSAQARESQQDILLRPVTQREVSLRGLRMSNSAPGKDRLEY</p>
----------	----------------------	---	--------------------	---	--	---	--

[illegible]

NeS L	Utopia -1_PS	.			GGAGTTGAGCTGCAGT TGCTGGTCTGGTTTAA CCCCATTGTTTTCTTT AACGGGGCGGTATGC CTTGACCGCGCTCTCAG CTTTCTCTAGAGCAAC CTTCGTTCAATCCACTTT TAGTTTTCTCAGTTCTTGG CCATTCGGTTGGTTAAT TTTTATTATACTTAATTT TATGTTTACATTTTCTGGT TGGAGACCATTTTAGCTT GTTTTAATAGCTTTCTTC TTTAATTAATACCCTCTGC CATTGAGGGTTTTTATTA CTATTAATTTTGTACTC TTTGAACTGTTTGATTG AATATTTAATAAACCCAC (SEQ ID NO: 1313)	GGAGTGTGAGCTGCAGT TGCTGGTCTGGTTTAA CCCCATTGTTTTCTTT AACGGGGCGGTATGC CTTGACCGCGCTCTCAG CTTTCTCTAGAGCAAC CTTCGTTCAATCCACTTT TAGTTTTCTCAGTTCTTGG CCATTCGGTTGGTTAAT TTTTATTATACTTAATTT TATGTTTACATTTTCTGGT TGGAGACCATTTTAGCTT GTTTTAATAGCTTTCTTC TTTAATTAATACCCTCTGC CATTGAGGGTTTTTATTA CTATTAATTTTGTACTC TTTGAACTGTTTGATTG AATATTTAATAAACCCAC (SEQ ID NO: 1313)	MDVDGGPAMPPEWVLRFDGACRRNP GPGGAGAAFKPCGTVAW TCSHYMPNNSSTNTAEY TALLGVQSAVHHGASHLEIEGDSHLVVA QVKGTFACRNPRQLRNVRHALRAVTS LTKHIDRKANAHADRLA NRALDLKRS LAECGEHQGAMESCLHMNPAARQREQP APPARPAC APTRAESADSHDEIDAEIAARDGGGEAFTLP IGPGTAPARQPRRLR QLTEDEQEAASALQMAEELACKIEDADSWTS GGGYSAIPSRIRQL LQPTAAQPHPRPPLQQQRQRPRTVTRTQREHRLDEALDEMAAVQ QERPTSRSAVRRARRRVGRIRASMRQQQLRHDFARNESKCVEDILRA ASAEATAEEHPETCPIDSGTLHEHTAVNSPRINFLPDEACGALFREA MADVGTGPQERRSALTDELTMDEVEDQLMQAATNSSPGHDGVGYDI YKFAAQLVPLHAAFCQSWRHHRVPALWKVGFVRLHHKGGPNDP ANWRPICLQTAIKLYSGLLARRLSAYLEANGLLLMAQKGFAYNGC HEHNFVATLLDQTRRMRRRLYQVWYDLRNFAGSVHQDMLWYVL RLLGVERAFVERCDDIYEDSYFVVGNAADGATEPVRQEVGVYQGCPL SPLLFIALVPLRLALEKLDGVGVVALADGVPCPTTAYADDLKVFSDSAA GITRCHAVVEKLEWTVLQANPGKCAFLAVTRNARGNPAHDKDMK LSLHDEEVSSIKLHDSRYRLGVGGDFHVRHRLQLEPKLQKQIKREAVA LMQSGLAPWQVVKALKTVYKPYEYALRHLRPLQSLQSGFDRVAK GLRHLRLPRSATNEVLYAPTSSGGGLQPLVEMHRLAQIAHAWQM
					AGACGAGCAACGCGCTGG GGCCCAAGACCTGGCAGG AACGACACTGCCAGGCT GTGAACGACGAGCAGCT GCTAACCGGCGAGCGAC CGGCTCTGGGCTCCGCT GCACCGGTTACCGGTGCA ACACGGTGGGTCCACGT CACGGCGGATCGGGGCA GCGGCTGTAGCTCGCTG CTCACAGGCTACCCACGG CACCAGCACGACTACGC GCCGGCCTGCTTCGGCTC GGCGTGCCAGACCCCGC CCGCTCCCCCGGGCCAC GACGACAGCCCCCGATGC G (SEQ ID NO: 1191)	AGACGAGCAACGCGCG GAGGCTCGACAGCTGGA CCTTACGCTGTCGAGCCA CTGCATCCGCGCCAGCCA CCGACCTGGGGCTGGT ACTGCCGCGCCACCGC GAAGGACAAACGAGCGG CAACGCTCGCGAGCGC CGCGTGGTCTGGGGGG GGCCGCGGCGCACATC GCAGGCTCGGCGACCGC GGTAAGCTGGTCAITTTGA CCAAAGGGCACTACCCA GGTAGGGAACCGCCCTT CAAAAACCCAGGAAGAC ACAAACACCTCCCTTTA GTGACATGCATATTTAG CCTACATTTACGTACCG AGAGGTTACTAAGTGGTA	Phytop hthora sojae

NeS L	Utopia -1_PU	ADOS 01001 321	Pythiu m ultimu m	<p>GCGGTGTACGGCACAAC GCCGCGCTCTCGAGTGC ACGTTGTGCGGCACACC GCGCGGATCTCGCGCG CTCCAGCAGCATCGCGC TCGCGCACCGCAGCGTC CGCTTTGTGGATCACTCC ACAGCGGATGCGCGTGCG GCGTGAGCTTCCACTCGC GTGCGGCGCAACCAAGC ACGCGCGGAATGTCAG AGAGCGCTTCTCGTGCG CCGCGCGCGCGCGACTG CAGCGGCCAACACCGCAG GTATGTCTCTCGGCGCGA CGAAGCGCACCACTCGC GTCCGTGCGCGCTTGCA GATGTTGCATCCCGCTT TTGGCAACATTTGCCGCT TGCGTTTTCGACCGCGC AGACGCATCAAGCGCCAC CGTGATCGCAGACGCGC CATGCAGCACAGTGCTGT GCCCTGCTGCGCGCCAA TCCCTCGGCGTGCGCAC GTCCCCCGTGGCGCGC GCCACACGACACCATCG CGCTGCGGATTGGTGA AACGCGCGCGCTGAACG ACGACGGCGACAACGAAA</p>	<p>AACACGAACACACAT (SEQ ID NO: 1314)</p>	<p>LHSDPAIREVARAQVWQVARKRHRLEEHWRERDDELVRCFLNSE LAASPHAEALRRHGDIGSLWSDVQRWIRIYHLSLVVQDDRNGLDPL GLRVPHHAKWLDHKSVLRHVHLKIRHQTRWKGADQGGKTVRAH GGVGAKFMSTWAGLSDDDYRFGVKARLNQIDTNVLKRRLRSHKT CRDPTCSSAETLAHVLNHCESNMDAIRQRHDDALEQIGSKIRNALKR GKSTAEIRLNQTPYETGAALRPDIVLIVAAKMMV/ADLAVTFEEHA AGARHSSLQLSHDKTLKYQPIVAELQLKGWQVQTAIYVYVGLSVQ PSNSTPTRKS (SEQ ID NO: 1436)</p> <p>MGTQSARERGAPSHSHTLGPRTPPRPACSKHGELESAAGGRDG QCSDDGAERERDAERDIRANERDCNGDGDGDDADSDSDRNDARR RSRRPRATATTTTAPTITTTTTTTSATTSATTPATDSSPWVLRFDGAC RRNPGGAGAALEFPGGAVWTVSHYLPGETNNTAEYSAMILG VRSIHGATRLRVEGDHIALSQVRGTFACTNRRRLKLRNRVQAAL RELGDYRLVHIDRQANAHADRLANRALDLRKTVDGCGPHATTTDAC VQPAELIAPTARLSSSSSSSSSSDEPMGLEEPAADDETDADEADI AMRDGGEIFPTLQIGPGSAPAAQPRLLRQLSDDESEAAARTLEHFA NDMAKIDADDWRSGEGYSAIPVRLRELLAPYAVPIRPPRNASSR PPRQSRPPRPVTRHQREHRLDEALDAAAQQRSTSTDDQSRINA RRVGRIRTAQAQSDLSQFATNERACVESILRAAKPDGTEPQASAG TCPIDRATLHAHFAGVNTPREFDFDADGADFRALDLVLPDQAA DAFADELSIGEVEDQLDRVVASSSPGLDGVGVDFKFRRLQLLPLHA AYQCCWRHRRVPATWKVGLVRLHKKGDRAEPNNWRPICIQAAY KVYGLLARRLSRWLEANEFTTAQKGFREFNGCHEHNEFVASSLDQ TRRLHKLAVWYDLRNFAGSMPQPLMWVRLARLGVDTAFQRC DIYADSFVGNAAADGATDPVQEVVYVQGCPLIFLIPILRAL QRLPGVGLADGVPCCTAYADDLKVFSASAAGIQQCHGTVARFL RWTGLRANASKCALLPVTITARGNPAIDDTLQLELHGDAIRLTQSS YAYLGVGDGFDHVQHRVQLAPKLAELKRDAVALRSLAPWQV/LKA IKVLYPRIYALRHLRPLQSQLEGFDRAVAKGFRHLRLPANATNELL YAPVSSGGLGLPLVELHKAQIAHGWQMLHSDAAVQAIARAQVR QWQKRYTLDADHWQGRDDELVLINSELAASPHATIKRRNGDIG SLWSDVQRHLTLQRLTRETPTADAPSPNGLLHLRVPHHRKWLS HKTVLRHMKLHRLCHKHKQWMSDQGRVTRAHGGAGSHFVSRG VGLWDADYRFALQARLNQIDTNSTLKRRTQRTNATCRAPNCSRTET LAHVLNHCETNMDVIRQRHGDGALEQIGAAINAAIKRRRTDTEVRLN QTVPEFNGPAWRPDIQVRDARSMTMVIADLITFEDQPNQDQSSSSL QHSREHKIAKYQPIAAALERQGWWRVHTSAIVYGLSRVSHPSNFTVYTE</p>
----------	-----------------	----------------------	----------------------------	---	--	--

				ACAGCAGCGCGCGCAGC CCGACATCGAGATGCGG CCGACGACACCGACGAC CAGCGCGACCAACCCCG CGACAGTGGGCTGCAA CGCCGCGCGCACAGCAC CAGCAACAGGATGCCG CGACGCGCGCGCGCC ACAGCG (SEQ ID NO: 1192)			LLGLLKRDARRLNTT LSCHCIRSSRRVWNWV HCGQHARQHQRCQE GRAHSGGNQRAEGGTATT (SEQ ID NO: 1437)
NeS L	Utopia -1_SM	.	Strigam ia maritim a	GGAGTGTTCTTTTCGGAG ACGCCGCTACTTTAGAG GAGAGAAATCCCACGGGC ATCCTCATTTGATCTGATC CATCGAGTATCTGCGAATA GTGCGGCACTCTTTTTCG CATGATCCCGGGGCTC ATGGTAAAAAGTTTGTG GCACGGCTTAGTTGACGC CCCTCTCCACGCTCACTCG GCCTGCATCGATCGACTGC TCTCTCTCTTCTCCCTCC CTCCTCTTACGCTCTCTTC ACGTACCCCTTCCCATCC CGCCCTCGGCTTTTGCCA AAGATCTGTGTCTCTCCA AAGCACCCATCTACCATTT GCTCGAGTTCGCAATGGT CGAAGCTGCCACGCCACTC GTCTACTCTGCTCTCTGA CCCCCTTCCCCCTCTCTCC TCTGCGCCCTCGCCTCTC GGGTCTTCCCATCTCTCC TCCTCCCATCCACGTGCTC CCTCCTCTCGTACGCTGAT CGTTGCGGCTCGAC (SEQ ID NO: 1193)	TGATGGGAGAGTGAGGA ATCTTCTCCACTGTGCAA AACCATACAGTCAGAAAG ATGCTAACTACTAGTTTG ATACCTGTGCCCTCTGC AATGTCCGCGTGTGTA CCCAAGCCCCGCTGGCAT TGAGACACATTAGGCTCT CGTCCCCCGTATACTCT CATAATTCGTGACGCT AATCTACCTACCCCTCC CTTTGACCACTCACCCAA CCATGTGTAGCTGTGC TGATCCCGGGCGG TTATTCACTGGTTATCATA TCATTCTAAAATGATCTTT GATCTCTCAATTAAACAAC TAACCTATTTCTGTTGTT TCATTGTTTCACTCGTAA GAGGAAGTTCATTGTGC GATAAATCAA (SEQ ID NO: 1316)	MATVRLKYPYPEGILCGPCAANTNAPQTRPYSDKSLAKHLKLYHK ATLVVECRHCGHEESDLRKMKKHISTNHVAAAAAPTVPRLGPTAP PPRVILRPRFIPRPTSPSSSSSSASSPASSRSVSLPPASPPVSSASS PAARSGRNPDSQGTAPVTPIGTVRNSPAGSPALSYSTASPIASTITTP RHLSPASALSAGPSLGAAPVPSPTAAATVPPAPATVPAVMAATVP FVAATTVPVSGSSTVPQRPAGPRRPPPEPIDDWIGRIARVSSSLPELDA VSRLEDEVVKRRPPDPNARPASLHPTRRPPPTPRRCHGTGTS VSLAALTSSCIREDHRGSLPLCTCGWDSPWPLSPSPSSHCPSSNPCPS SSTPPSLLGPPRSHLTLWRGSGSTTSPHRHCSRARYHLLWRLPYPS SPRLPYPSALARYPSVPPALDDPLPSLSTIGSEGSPECHLCRNWTPSQ GCSRMKWSRDAPLTPTDPPHSIRHAALPLHLLALVPAMGPAELQSL WRRSRPRAFAKITEGASPFCAIPVGTVGHFLLQVHQATAHLPTVPPL PPLRPPSSDPLVTPISPGEVLDRLRRATDTAGPDTIYSEWRAIDPT GRLLSLFKVQTFGAPTRWKESTTLIHKGGDHTAMSSWRPIALLST VAKIYGSILSHRLTTWAVQNGRLSLSQKGLPFRGCLDQNYLVQSCLO DARRNKKTLSLAFDLKNAFGSIPHLTIRHSLEWLGLAPSSIDILEASFL GSSTRVRTETGLTPPISLDTGWQGAFLPILFNLAIPLRTVPSAHSG FSLHGHVSVVAYADDLAILAPSTPALQSDLAISGMADWAGLSFNP AKCATVLTGKDNSTRDLSLQGSVPSPISDGDAYKHLGVPTGTTTFS GTDIAIKMTTDLQADHSDLAIPAQKLDALRTFIMPQLSFHLSHGSP KAPLTQLDKKIKRAAKHWLFPQRASNEILYMSHLHGGQSLPLSVLA DIGQVTHAVALLQSRDPAVDLALRTCREVASRAKKTVNGPELAQY LSGSTDGIYCTPSLWTTARAATRLSLTPLTWTSPLPSGVFELS INGSPISPFVQSTLTNAIRHNHLSLTIARQDQNSYRTSHDPDPSNY WVKGGDFLRCDFWRFIHRARLNLVNGARRWDANSIKTCRRCGA PNETLAHV/LNVCPVGLPEMKKRHDIAHARIKKAIRPSPHTVVVHHDR VPGCGPLRPDLIRISERDKSVAIVDIHV/PFDNGTDAVERAHETKRAKYE	

NeS L	Utopia -1_SP	AAGI0 21405 37	Strongy locentr otus purpur atus	AAGGCTCAACACGAGCTG CCAAACCAAGCTGCCAGCTT AGGCACCAACCAAGCTGC CAGTCCAGGCTCCAACCA GGCTGCCACCAATCTGCC AGCTTAGGCTCCAACCA GCTGCCACCTAGCTGCCA GCTTAGGCACCAACCAAG CTGCCAGCGAGGATCCA ACCAAGGCTGCCACCTAGC TGCCAGCTTAGGCTCCAAC CAAGCCGACGACCCAGGC TCCAACCAAGTTGCCACCC GAAAGCTGCCAGTTAG GCACCAACCAAGCTGCCA GCCAGGCTCCAACCAAG CTGCCACTCAGGCTCCAA CCAGGCTTCAACCCGAAGC ATTACCCAGGCTGCCCGTT TAGGCACAAACCAAGCTG CCAGCCGAGGCTCCAAC AAGCAACCAAGCCAGGTT CCAAACAAGCTGCCAGCC GAGGCTCCAAGTAAGCTG CCAGCCGAGGCTCCAAC AGGCTGCCACTCGAGGCT CCAACCAAGGCTTCAACCCG AAGCATTACCCAGGCTGCC CGTTTAGGCACAAACCA GCTGCCAGCCGAGGCTCC AACCAGCAACCAAGCCCA GGTTCCAATCAAGCTGCCA GCCAGGCTCCAACCAAG CTGCCAGCCGAGGCTCCA ACCAAGGCTGCCACCAAG CTGCCAACTTAGGCTCCAA	TGACGAAATGTTCAATAT ATGACATTCTAATGTTCA TGTAATTTTGTGTTGCTG ACAAAGCTAGATGAATAT ATTCCCTGCTCTACTACAT ACATCTCGATGCGCCTT GCTAGACGATCCATTGGT AACCATGATATCAATGGA TTAACACATGATCTGTAA ACAAATATGATTTACAC AATGATTTATGTTTCAA TAAATCTGTTCTTTCACT TTAATACATGAACATGA CTGTGCTTCTCCAACCTG GAGACCTACATAATTTGT TAAATGATATAAATATT TCGAAGATGAAATATTA TTAATAA (SEQ. ID NO: 1317)	LIRHYEHQGYRVTFDSLVTALGRSLWRGSEAAALQALQISSQYSKLLR KLLVADAIHGSRNVYAHMTGMVM (SEQ. ID NO: 1438) MSHSITEVFDYPLPSRWKCTVCLENFFNQTLKRHQARHHQTTSLFY VFRCSACAEFDSARKASHNWSHKKRPILSQPAVNEIPSSGLDPSP PRSRPVEVIGSSFFDDVSMISEPSTPSLQMDPEVWHPSPSISFSP MHLSPQTQSPASPIQIGVEVFNSSSLQMDPEVQPPSPSISFSPMHL PTQASPIQIGVEVFNSSSLQMDPEVQPPSPSISFSPMHLSPQTQ VSPQIGVEVFTSSPLQMDPEVQPPSPSISFSPMHLSTLQDPSP VDIDVPAAEVPLDIEIPSPDRHPAAEVLPLDIEIPSPDRHPVAEVP LPDIEIPSPDRHPQSPRRPVMMEQPVHTPPADTQQANGPQHWV TVLANATNWEDFGRVCVEFANHAVEAARSQDAPQVRPAARQRP RRPTRPRQPTFDVREASRLQKLYKRSKKRAVRHILRDDAPSFSGSEQ LLDYFEIYAPPEIDENRAQQLAESLFDLEEKESAAALMSPISQSEIS TRLSRMSNAPGKDRLEYRHRQADGACRVTHIMFNRCLEHRIPSA WKEATTILIHKSGTTDDPANFRPIALQSCLYKLFMGILSDRMTQWAC NHNLLSPEQKSARPCGCGCHEHTLLSSVIKDKRNQKNTANIAWDLR NAFGSIPHQAIAVLTITIGAPVSLVMLLKDTYTGAFTSFLSTGETDPI QIQSGVKQGCNMSAILNLTILIRAVKKKATDDGLGVVHGQRLSI MAYADDLVLMSTPEGLDAILSVASEQAETLRLAFKPTKCASLSLSCR HGTSVLPREVTVQGHLMIPALDEEQVRYLGVPGFLPRFTNLKDLGKL KGNETIASLLAPWQKLDKFTVQPGLSFVLAADYKLSDLRLSKSAI TTNVKKICQLPLRAANAYIFAAKESGGLAFIDPNVDADIQVITQAVRVL SSDDEVQTIATSQLKSVVHRTIHAVPTEEDIDNLYSGSNEGGLANS NSGQASLWSRTRSAARRLHLTLRATSGTVVNVQADIDHTRDILP ASITRGLRIQRTTNAEKLKSLPDQGVARSLSNDFANGSSWHATG KFIRFCDWRFIHRARLNCLPTNVATKRWKANANGKNGHQETLPH VLNHLCPNMVPIRRRHDNIQRLVTAIRHGDVFNQHVPGDPNPRE RPDITVIEGNKVTVIDISVPFNGPNACTTAAQAKVEKYSALRQALRD MGRDVEVHGFVIGALGTWHQGNERALGRGLGSRWYRTLMRKLCI DAIQASRDIWVEHVTGHRQYE (SEQ. ID NO: 1439)
----------	-----------------	----------------------	--	---	--	---

				CCAAGCTGCAGCCAGG CTCCAACAGGTTGCCACC CGAATCTCTCCAGTTTAG GCACCAACCAAGTGCCA GCCTAGGCTCCAACCAGG CTGCCACTCGAGGCTCCAA CCAGGCTTTCACCCGAAGC ATCACCAGGCTGCCCTTT TAGGCACAAATCAAGCTG CCAGCCGAGGCTCCAACC AAGCAACCAAGCCAGGTT CCAATGAAGCTACCAACC GAGGCTCCAAACCAAGCTG TCAGCCGAGGCTCCAACC AGGCTGCCCAACCAAGCTG CCAGCTTAGGTACCAACCA AGTGCCAGCCCAAGGCTC CAACCAGGTTGCCACCCG AACTCTGCCAGTTTAGGC ACCAACCAAGTGCCAGC CGAGGCTCCAACCAAGGCT GCCACTCGAGGCTCCAACC AGGCTTTCACCCCAAGCAT CAACCAGGCTGCCCGTTTA GGCACAACCAAGCTGCC AGCCGAGGCTCCAACCAG GCCTCAACCAAGCTACCA GCCGAAGCTTGCCAGAT TAGGCACCAACCAAGCTG CCACCGTTAGAGGCCAG AGCCACCAATCCATAAC GTGTGAGATATGTGAAGC TTCCTTCCACACTCGGCC GGTCTCCGTCGCCACACG GCCAGGCTTCATCGACCA CTGCTGAACACACTGACG ACAGC (SEQ ID NO: 1194)

NeS L	Utopia -1_TSP	.	Trichin ella spiralis	<p>TTTCTGGTATGAATCCCAA GCGGATTCGTTACGAAATT TGCATAAGTTTTTGAAAAA ATAGGCATTGGTCGAGT GCTCGCACCAACCATTTGTC GCGGGTCGCTCTGATATT GCACTACATTGAGGAACG GCCTATTCCTTCGGGGAA TTGTGTTTTAGGAATTGGA ATCGGTTTGGTTACGATCG GTCGAGTGGTTCGTGAGA TCGAGTGACAGCCGGGTG GCAGGGACA (SEQ ID NO: 1195)</p>	<p>TGATCCGTCCCGAACCAA CGGAACCACTTGC CGCA TGACTCGATTTGCTTTT GCTCTTTTGTATTAAAT TTGCTATTAAACAATTCAG TTTGTTAACTGTTTGTAT TCATTTGAAGATCCAAAT AAAC (SEQ ID NO: 1318)</p>	<p>MCSAKTPALSGRRRGKEVNVGEQIVRVERRRGRSSTSATDLGTRM VTRGRKKLMEASVREAGHHGGESASTVDVNVESKKITGKTARRNR RAPSGDGKRRSCGAEQCGAVCGNAVADSEASRPTPNVSKSGRD KCGQPTIKASTSPPKRKPTTSSSPRTCLSKRGASRPSKIPSTPTSTSG GSGKQRLVSPLLRTEKLPDLLEVLRQTEEQVTVRATFPIAQAVVCPILG CEKPYAVRPDGGQFAHQTLTRHFMRVHNCHSVQWHYRCRNCNTD FLPADHRYPLRVNVNTHVSCVSRWEITRKLGESEDLHGVRCDLQDVV GVSKRAVGLHRRRHANENIMQNTGTAAQIEALSQVGEIRVAGDYS QFKFGRVQRQVAVPTQRDGLDEEEVHEAEVEEVPASRTILGEPSTA TAIGAEISATGPVRADTAQQMICRIGQWCVWPQDYHSIPAPQC WTDLTMDLMIEQIVLQRPDGGVSMSCSAVSAAIHHEISAEFAA QVMSSHDASLYCIPIVNVNHWQMIVLDVAERVVHYVCSLREHNTV VLSLSLVELSGKHTGCTSWKIETHDGAQVQTNADFDCGPFSCFLKH LLHGIDMNFGRDRESAALRTDLKFMIDAVSTPVVPATDKLKKPDGSA TQLTQFQKFLSASDNWQSPVDLQAVVDEVESVSGHNEPSSRN TSRLQKSPGKGGKQVRRRSVAVTRDPAWLKSASAVQKAFNSAPAR TVNAILRRPNACPSFTATQVADHYFNLRPAVTSLAPEVIDILPPATDH SMLVAELSESEVWEKMQKAPNSAPGADRTIRMVMRMADPGAMILT RFYRACLLRWVPLQWKQSVCKLLYKDGDKERLANWRPIALEPVLQ RVLSAVASRVTNWARANGLISLEAQKGFQADGTSEHNFVMEVAI QEARTNAQLAISWLDISNAFGTVSHQLLSLLERYGLDPTFTSFQNL YKDATIVKGANGTHVTARWSVGRVQGDPCSGILFCLFVEPLLRSLV PSLPCEAETTAVNVLGQPTALAYADDIALFAPSIGVMQQQLCKIQG MASAMGFRFNPKKCASLYLNRAVVNAAFTTISGEEIPALVHGDTRY LGVAAGLGKQPTPFSLLENLREAEILFRSKLAPWQKMDAYRTYVLP LTFQLMIAKFNNIKSAGQYDRAILRLVKRCFQLPVTSTDFIRAPRQC GGLGVPSLRELYATAKVSRAKMLWSPCRVSVSSLAASQLQRVASAYF AKRLDVEAADLSTFMNAARSTPLDRSGYPTCLWMDARKQM/SYLT KVAGDCYFLVGEAGTSFFIRNGLGQTVSVLSPLRKNKVMVSLGGAI QTRHLDALWLCRQKGTASCIVDRSSRFITGRYTSFAAMRFALP ARLDLPCRARSSMRSYQNCRCGYDRETLPHILQHCRQFSAPAYQA RHDVQGRLETVMRRFPNLRVNRALPEIGSNKRPLDVVDEEKLRLV ILLDIAVFENTAAAFVDARTKWAHYEKEILAYRLRGYSVTYDAIVVG ALGTWDPKNDAILKRIGVVSRYLRLMKVLVSEMELHSSRIYRKHL GLRDLLPDTGTKRPPVGTETDPPGGDLRQKKRNTISARASGGKCLE RRFTSPVGTSPQRGELQCQPCGPRRPPALAGIAPNPPSLQPRKPPPR QHQQPVTKSTAH (SEQ ID NO: 1440)</p>
----------	------------------	---	-----------------------------	--	---	---

NeS L	Utopia - 2_CM y	.	Cheloni a mydas	GGAGAAACTGCACSTTCT AGAAGACTCACTGCTATC CTGAGGAAGACTACCGCT TTGGAGATGGATTCTACTG CCGCTGCTTCAAAAGAA ATCTTCATGCTGCTTCGGA GGCTCCAGGACAGASTGA GAAGATGATCGCTCCCTC GCCGATTCACAGAAACCT TCAACTGCCGCTCGGACC GCTGCTGCTCCACGGAGC GCCATCGGGGAACAGCC TCTAAGAGACCCTCAGCGC TTCAGGACATCGCAGAT GAGCAGCATCGATTGGAA AAGCAGCGCTCCCTGAA GAAAACCGTGGAGATGC TGCTGAGGGAWATCTTG CACTCGAGGACGGCCTC CCAGGACATCGCAGCCAG GACAAACATCATCTCGCT GCTCTCAGGASAAAGGAT GCCMSAAGAACTTMTCCC ACCTCTCMACTGCCCG GATCCTSATGCTGCTGCTC GTCCTGCTGCTGCTGGAAG MKACCCCAACTGGCGAGA CCTCAGAAACCCAGM WGGACCKCAACATCTCAC TAGATGCTGCTCCAGCCG AAGCTCTCCATGCACTCT CCAGAGCAACSAGAACCC ATCCAGMGAATCCGCTGA TATGACTGAAGCCMATCC AACAGAGGGAGAAAGGAA AGGAGAAATGACTGCATCT ATCTCCAGTATCCCTCCC TAGGACACGCTCCTCTKC	TGAGCTGGAGTGCCGAT GAGAAGCGCACTCGGGA AAGTAAGTAAATATCTT CCTCATGGATTGATTCTT TAAATGGACAACTACCT AATCTCAATTAAGG GACAATCTCACTCATTTG ATATAATTTGCTTCCACA ACCAATCTCTGTACAAC TTTTCATGAGTGATGATC CCGAGTACTTGGATTCTA ATATCTAACTGTATTGT TAAATCTATCACCTAAAT TTGGTTATTGCTGATTA TGACTCTATGTATCATAT GACTTTTAAAAACAACT TTGATTTGTGGATAATC TAAGCACTATACCCAGAT GTACAGACACTCTTTCC CAACTATGTATTATTT TTTTAACATTAGCTTTAAT AAAAATTTTTAA (SEQ ID NO: 1319)	MLQLRLPTQLRLHPSQLPQSHSTKRWSNIYEERARAPKDTTIERV SAASKPKLDPAKRRIGAPLQMQGNSISRQLSASSQVYQHNAVRR VSAPHTGNFTGSCRRKLALPHRPSETQPAESYLHCWTTQRDPTLA DQHGLQDHPHTGLPEGSHCHIKDRQNSSTEGRRSGNSQARRIPKR ASTAASKTVLPKRTSAALKSVREDTALVLEDPKAWSSQHREGXRQQA NPTAVFQPEPAIEEQPXVRAATPWQAAWMEELARTASFXDFDLV DRLTKDLSAEIVSGRKGTEQNTPTAHQNNQNNMREARRNISRCDY PAAASRIQKLYRSNRPKAMREILDGPSSYCAIPSERFLYFKGVFDRVA QNDMQRPECLXPXPRVDYAEDEXQDFTSWEVEARLTKTNTAPGK DGIRYNFLKRDPCGLVLTAFNCKQFRRTSPSSWKSMMLVYKKG KQDNPNTRRPSISCTMYKLYASCLAARITDWSVNGGAISSIQKGF SCKGCVENFVLTQTAHMMARAWRQCAIAWLDLANAFGSMPHQHI FDMREFGMPENFLQVRELYEGCTTTTCSMEGETPEIPRSQVKKQGC PLSPVFNLAEMPELIRAISSGLGGFDLYDNRVNIAYADDLVLIADNPES LQQMLDITSQAANWMLRNFARKCASLHIDGSRDSDVQATSFQIQ GEPMIFLEDGQAYQHLGTPTGFRVQQTPEDTIAELRDVARIDSSLLA PWQKINALTFILPRISFVLRGSAMVKVPLNKADNTIRQLVKKWMFL PQRASNELVYSHRQGGANVPRMGDLCDVAVITHAFLRLTCDAMV RNIAESALQDQAVKKRIARTPSNQDVATYLSGSLEGEFGRDGGDFASL WTRARNATRRLEKRIGCHWVWCEERQELGVLPQVKNTHITIPR ARTMLERTLKDAIRCQYVENLKRKPDQKAFVETCKWDASNHLPG GSFTRFADWRFIHRARLNCVPLNGAV/RHGNRDKRCKCGYANETLP HVLCSCKPHSRAWQLRHNAIQDRLRAIPPPVGKAVNSAIPGTDQ LRPDVITNEDRKKIIMVDVTVPFENRTPAFHDARARKEKYAPLAETL RAKGYYQVQTHALIVGALGAWDPSNERVLRECGIGQRYARLMRQLM VSDAIRWSRDIYIEHITGHRQYQEG (SEQ ID NO: 1441)
----------	--------------------------	---	-----------------------	--	--	--

[illegible]

NeS L	Utopia -2_PI	.		<p>CCGCTTCGTGGATCACTT CCACAGCGGGTGCACGTG TGGCATCGGCTTCAGTCG CGCGGGCAGCTACGCGA CACGCTCAAGCCTGTGCA GACACGACACAGCCACC GTAGCTGCCTCGCGGAC CCGCMNCCASTGCSSCS GNGCCGMAGCGASGAG GAGGACNCGCACCCCC GGTCTCTSTCGGCGAG CATCTGCAGCTGCAGCAG CCGCATCAAGCGCCACTCC AGCTGCAGATACTGCCACC ACGCAGAGCGCCGTGCC ATCGTACTCTGGGCCCC AGTACGCCCCCAGTGCT GGAGCCACCTCCAGAGCT CCGAGTTTCGGCAAACG CCGMCGCCTCAACACGCC GATCGACTGCAGCCGCT GGACGTGGACGCGCTG (SEQ ID NO: 1197)</p>	<p>TAGCGGAAACACGCGC CAAGACGCGCACAGGG CCCCACCGGTAGGGA ACCGCCTAGAAACCCAT TTCGGTGTGCTGACTCGAA GGCCTTACCTATTTTTCC TTAGACATTCAATTAGGT AGCGACCAAAATTACAAAT TTGGTAAACGAGTAAGCCA AATGGTAAATACACAAAC TTTTCTGTTCTAATCAGTG TGAAAACCTGGTTTCGCC TTTTGGCGGACTTTTCA CTCGCATTTTGGGCAAT CGTCTGGGCTAGCTTGC</p>	<p>SELAASPFADCLRRNGDIGSLWTDVQRTLRLHLSLTAQDDRRGQD PLALRVPHHTKWLDHKTIVLRHVKLHMKIRHQTRWKGVLVDQKTVR VHGLGAKFVSTGAGLSDDAYRFGVKARLNQVDTNVAVLKRKLRS KTCRDP TCSSAETLAHALNHCSNMDAIRQRHDDALEQIGSKIRGAL ERAKSTTELRLNQTVPEYTGAAALRPDIVLRNVAACKMVIADLAVTFED HAAGARHSSLQLSHDHKTLKYQPIVAELRVQGWQVQTAIVYGSLG SVQPSNFKTYTEKLIHKREARQLDLQLSSHCIAQASHRIWGWHCRRH REGQSGNTSRASRGSGGTPrRTSQVRARR (SEQ ID NO: 1442)</p>	<p>MILADPAALAAGLARAPPPSAPQDPSPAFAPGAGQNPRAAAPAR VEVHTVAPPGRAGGMPLDPGLVDSSPAAATAATPAPVAATAATTAR AAARVAVEHHAHAEPNQEHLPMARVLVEPMQVDECCSDRSTLTA DDGSGDDVAAPSSMLSDNVAAPMDVDVDSGTSCPPTLQQLQRPRL HVGSKRRRLDADDGEEAHQLQEEEEEAGIHAPALRLSAAASAQPSVLA VYTHNASRFDCTLCAYTAGSFASLLTHRNSRHRRTAFDRFSAGCAGG VPFASRLAAARHAQACASLSSAPSAEASSAAGTSPTADGADSTVSA VAHAEPGLPHHNDTELTASPPVLVSSSDVEVQATKTEATDNRWGAPL PRVLVASRIAGRLAQVPPRWGPPLPRTTIAARIATRLAATPAPRW PPLPRSLVSRIAARLLPALPDAPACEEAKDSDTMDWAPTWTNET KESEPHDEAPQVDEETIDDDADGEWLLRFDGACRANPGPGGAGAA LFKPSGPVWVTCSHYDPSTTATNNTAEYTHALLGARAADHGVTKLR IEGDSLVIQQVRGIFATSTRILARNVKLELARVGSFSLHHIDRQA NGHADRLANAGLDRRRRTKLECSHPDGRGCTNTSVATAAPTAPAAP LPSARPPASTAAPSPDDDDHSDQGDIDDDGEVVAAMWICISPDVPHRRP</p>
----------	-----------------	---	--	---	--	--	--

[illegible]

NeS L	Utopia -2_PR	.	Phytophthora ramorum	TCAAGCCCGCCGCAAG CCAGCTGCGGCTGTGCG GCCCTCAGCAGCAGCA GTCGAGCTGAACCTACA GCCCTCCTGCTGATCGG CCGCCGTGGAGCCCGCG CGCGCGCGAGCCCGCCC AAGAACAGCACCCCGAG CTAGCGCGCGTGGAGC CC (SEQ ID NO: 1199)	GATCCCTGGCCTCTCAGG CTGTACGCGAGACCCGTA CGCCGAATCCCTGGCC TCTCAGCCTGTACGCGG GC (SEQ ID NO: 1321)	MSGDVSSDGSRTTASGDGDDGAGSSDAAGDVGVAMVDVQ GARRQPPWQR/GGKRRRLNDVDEDETRAEELLEEEDEAGDHA PAPRLSASARPASVLSVYAHNAQRFQCTCTYTAASFASKRHRDSR HRTAFLDRFSAGCAGAPFASRLAAAHACASLNRSLVAATPA AGELSPTAGAAATVKAATVTPDSRQDPPELAASPLASSPDVAVQ AADMQAPTRWDPLPRTLVAIVASRLTDLTPRWGPPLPRATWS RIARLEAAPTRWGPPLPRVVSASRIERLAPPELAADDETKDGEED QSFTPEVAARSXGEGEDANGEWLLRFDGACRANPGPGGAGALFK PSGPVWVTCSHYMPSSETNNTAEYALLGMRAAADHGATRVHV EGDSTLVIQVIRGIFATRSTRRLRGLKSVKAEARMMEHVTLLHIDRQ ANGHADRLANAALDRRKTKLECLHFDGQGGCSSTAATTAVPSWPD RPPSSTAAPTSAEPDETEQDIDDGEVYAAAMCIGPDSIPERRPLR LRQSETEEEEAGAIERLAATLAGKITDASDWATAEGYITALPYTLVD KLQFAQHRHQRPQRHQQRDPPLGTHDGHGQPSRSRRRR RRAKDLRRRPPRITRHHREHRLDEALDRLRAVEHASPDRPAVARA RRVGRVNSAIAQQQLRHKFKDEKACVDGILAAARASGLATPSAS ASRHPVPVSTAADDGSCIPISDELHAFITAVNTAGTFEPMAVGA PFRSAVAILPAATSQPELLSDPTDDIEDQLQARGSSPGLDGVG YDIYKAFQAQLPALHAFAFACWRHKQVQPSWKVGVVRLFKKGER TEPANWRPICLQQAIIYKLYAGVILARLRLRWLDANGRHADTQGFRA MNGCGEHNFLAATLVDQARRKRRELHVWVDFANAFGSVPHDL WEALERQGVSPFIACCRLYADAVFTVNAADGTTAPIALRVGVFQ GCPLSPHLFTAAPLHALLKRLPDTGVQLSRVDCLGASAYADDLKIFS GTEGGTKRQHALVADFLRWGMRRANPAKCTMSVQRDRTRGVKA CNGLQDGAIPALMTMSAYAYLIGDGFHVRRIELAPKIQELKH DATALQSGIAPWQVWKAIVKLYPRVEYALRLRPFHQLEGFDR HLVRGLRLHLLRPANATTAFFAPVSRGGLPLTELHAALQVAHG WQMLNSKDPAIRRIARVQLRIADARHIDAQAQWQDREELAQLLL NSQLGASTGAPKRRNGDGLSLWVDVQRHLRLSLKLETAPACAEATG TAAAMLQLRVPHDKWLDHKTVLRHVKLHYKNKHWAARWAAAMXD QGKTARTHGGAGSGFLTRPRGMWEADYRFAVAARLNQLDTHSVLK RRRLXHDRCRQPGCTQGDDAGARAQPLRRHHGRGPRPPRRPQA HRARAARVAGRPGPRRAPGQPDGAVARRPRATARPVAVQPHQE
----------	-----------------	---	----------------------	--	---	---

				<p>TGCGCTGCCGCCAGTGCG ACGGCGTTGATCCAGCGT CATCTACAGCGCTTCTC CCGTATCGACAGCAACCG ATCTCAACGGCGATGATA GCCAACCTCCAGACCCGC GTGCCAACATCCACAGCT TCCAGCCATCCATTCGACA ACGACAACGACAACGACA ACGACGACGACCCGCTGC TCGACGACGTGCTC (SEQ ID NO: 1200)</p>				
NeS L	Utopia -3_pi	.	Phytophthora infestans	<p>GTACGGCCGAATCCCTGG CCTCTCAGCCTGTACGCGG GGCTATACTTGGTCTAAGT (SEQ ID NO: 1201)</p>	<p>TAGACGGCAAGTTCTG GCCGCGTAGGCCGAAA GGCCCGCCCGAGGTAG GGTAACGCCCTCGGAA AACCATCTCGTGTGTTGG CTGTTTTCAACAGTCGA ACTTCACTCGGAGCATA TCAGATACACTACTCAC ACATTTAGATATCAGATA GGGAACCTTTATTAGGG AGATAACGGGTACACCG GATGGTAAATATACAAA CCTTCTCTGTTCTAATCAG TGTGAAACTGTTTTCG CCTTTGGCGGACTTTTC ACTCGCATTTTGGGCAA TCGTCTCGGCTAGCTTG CTAGCGGGGACGAGCG GTCTCGGGGGCGTTTTC CTTCCCGCGAGGCCA ACTACCGCATCTTCTCT ACACTTTCTAATTCGCT CCGTCTCGGCTCTCGGT TGTGGGCTTTTCTTTT TGACCAATCAGAGCGCG CCATGCGACTCTTCTGCG</p>	<p>MPRSLASEPVHSSASRLPSQAPPTSRSGAAHTPPSLQARPLASADA GLAATALASPDPPYDVAPPGRAGRLPDSVSPGATLSAATARALAV RHWPLELSDSDDEDAQDPHAAAPGPADVASVLAAPGRAGSML ADPAALAAGLARAPPSPAPQDPSPALPAGPAGQNPRAAAPARVEV HTVAPPGRAGGMLPDPGLVDSSPAAATAATPAPVAATATTARVA VEHHAEPNQEHLPMARVLVPMQVDECCSDRSTLTADDGSGD DVAAPSSMLNDVAAPMDVDSGTSCPTLQQPLQRPRLHVGSKR RRLNADDGEEAHQLEEEEGIHAPALRLSASAQSPASVLAVYTHNA SRFDCTLCAYTAGSFASLKHNSRHRRTAFDLRFSGACGACGVPFASR LAAARHAQACASLSSAPLAEASSAAGASHTVDGADSTVSAAGHAEP DLPRHNATELTASPLVSSTDVQATETETETENRWGTPLPRVLVAS RIAGRLAQVPPRWGPPLPRTTIAGRIATRLAATPAPRWSPLPRSLV ASRIAGRLLPALPDAPACEDEAKDSEMDWEASEPHVEAPGPVDEE TIDDADGEWLLRFDGACRANPGPGGAGALFKPSGPVVMWTCSHVD PSTTATNTAEYTAALLGARAADHGVTKLRVEGDSTLVIQVVGIFA TRSTRLRALRNKVKLELARGVGSFLHHIDRQANGHADRLANAGLDRR RTQLECSVHPDGRGCTNTSVATAAPTASAAPSTTRPPTTAAPFHS DQGHIDEDERRADIDDEIYAPMTLGPDEVPAARRPRLRLQSLDEE LEAAGAIVERLSASLSAKITDAEDWGTAEVITLPHLLYDKLLPYSRTA PRHQRPSPSRNQDHPQRRDQPNRVDEQQAESQGGEDQR QQQPPTRRRRRRKRRRRRQRPQPGQSTRASQSSRQRRPRPP RVTRHHREHRIIDEALDELHLELERRARPQDRSAIDKARRRVGRVRAIN QHLRHRFDTDEKACVADILEKAHARAATAQAAGAAATSTGGAAT APTQQAATSALGDADDGTCPILADELWQYFTGTNTPRWFEFNPATPV GEAFRTAMARLPATRLRELLTEAPTADEIETQLQHVGRSSSPGLDGI GYDVVQRFAQLPLVLTASFKRCWTAKMVPQSWQVGVVRLLYKKG</p>		

CAATCAGAGACCGGGCC CTGTCTCGGACAGCA GGCTCCAGCGCCAGCC AATCAAGTCTCGGAGC GACGGCTCTTCTATAGC GCAGCTGACGAGGCCGA TCTGGCGGCCCCGATTG GTCCGACTTCGGCCAAT CAGCGACGAGAGGGGG CAGGGGTTTACACTTTTG CCCCGGTTTCGGCTTCAA CTTCAGGCCAAAATGGC GATTGGACCTCCACGC GCCGTGCCACTGCTCGGC ACCGCGGCGGATTTCAGC GGGTGCAACTTCGGGCA CGTGTGCAACACATGCA GGCCCATTCACGCCAA GCGGCATCGCGGACGA CGCTCGGCGGCTCAAGC GCAGCCCGCCCTTCCAG CACGACCTCGCGCGTTT GGCGATCGCCATCAAG ACGTGCGAGAGCCAGGC GGGGTCGGGCAAAATAT ACTTACTCTAAGTATGCC CGAATCCCTGCCCTCTCA GGCTGTACGGGCCCCAT ACTTGACCTAAGTATGGG AGGATCCCTGGCCTCTCA GGCTGTACGCGRGGACC AAGTACAGCCGAATCCCT GGCCTCTCAGCCTGTACG CGGGGCTATCTTGGTCT AAGTATGCCCCGGTCTGCT GGCCTCTGAGCCCCGTACG C (SEQ ID NO: 1324)					AHDDPANWRPCLQQAIIKLYTGLARRLVRWLVDNDRHAPGQKG FRAVNGGGEHNFLAATLIDQARRRRSLYEVWYDFRNFAGSVPFQLL WDSLQRLGAPADFDIMCKGLYHQAAAFVIGNAADGPTAAIRQQVGV FQGCPLSPQLFNVAISPLFLARRLPETGVQLSGDDRVGVSAYADDLK TFSSTKAGATKQHELVAALAWTGMKANAAKCSMGVRRNSNGAT EADNLDLALDGTPIPSVTHMQSYTYLIGIGDGFHVHRIELAPKLKTL KQDITLALLESGLAPWQVVKAVKYLPRVEYALRHLRPEDQLLESFDL HLRAGLRHLRLPKNANNDFYSPVSRGGGLLPLVELHAALQIAHG WQMLNSTDPATRIAREQLHQIADARHRLDKAHWKERGDLCQLF LNLDLGTSAHAPKRRNCIDIGSLWVDVRKNLQAFGLKLETAPADAES GTPALPLQLRVPHHEKWLTHRDVLRHVQHLKNKHWRWCAFDQ QGRTHAHGGVGSFTRPRGMWESDYRFAVAARLNMVDTSATLA RRRLRAHDCRYPGRCRWKESLEHVLNHCPTGMDAVRGRHDGVLREI EHALRAPSGARRELVRNQTVPGLPGPALRPDIQVYNHDQRTVAVVD LAVAFDRQDRDDPETSGLAKAAAEKKAKYTGIGRHLERQGWKVHLS ALVYGSLSGVAPNNYKVYTEHLGLLKRDAKRLDRLTSLVACIQSSRIW NLHCAKHRARQHQTSPQSRGRRTTETGAPSRDTRR (SEQ ID NO: 1446)
--	--	--	--	--	---

NeS L	Utopia -3_PR	.	Phytophthora ramorum	TGCGCGGACACCAAGAC GCGCAGCAACCAACAGAC CGTGACGCTGCGGGTGG AAAGCGCGCGCGCTGAA CGCTGGTGACGATGAAGA CCAGCGAGAGCTGGCGGA GCTCTGCTGTGGACGA GGACAAGGCTGGCGCGGA ACACCGCGCTCAGGCT GCCACGGCGCAGCGCTCA TCCGGCTCCGCTCCTCC GTGTACGCGCAGCTGCA ACTCGCTTCACTGACGCG TGTGCAGCTACAGCGCTG CCAGCTCGCTTCGCTCAA GCGCCACGCTCGTCTCG GCACGACGCGCGGCTT CCTGCACAACTTCTGGCG GGCTGCGCTGCGGCGACG CCCTCGCATCGAGTTGG CCGACGCGACACGCGC AAGCTGCGCCAACTCT GCACCACTCGGCGAGGA CTTCGACGCGCAGCAAGG CATCAAGCCCCACTGCTGC CGGAGGCGACCCACCGT CCGTGCACTGGTCAACCG CGCGCCGACCTGCCCG CCAGTATCCCTCGAGCTC GTTGCGTCCCCCGCGAGC CGAGCTCCACCAACGTTGC A (SEQ ID NO: 1202)	TAGACGGCAACACTCTG GCCGCGTGGCGCGAAA GGGCCACCCACGCTAG GGAACCGCCCTCGGGAA ACCCAGTCTGGTGTTCGG CCAAAGAAATGCACCA CCACGGCGGAGGTGCAT TTCGACAGTCGAACTTCA ACCGCCACATATCGGAT ATAGTTACAGCTCTAGTT AGACATCGGATAGGAAC TTCTTAGAAAAATTAACGG GTATACGGATGGTAAAT AAAAATAAACTTC (SEQ ID NO: 1325)	MQQDVDAEQARNRWGPPRLPRVLVASRVAARLGEVPPRWGPPL PRGVVVSIAARLEAVPVRGGPPPLPRSFVATRIADRLAPPSPDLSLD EEMKESEPDPDTHSADESDTADADAVMEFAFVSDPPTATPREW RLQFDGACRGGPNPGGAGALLVNTPEGAVVWTFGSHVMPGAKETNN SAEYVALLIGARAAADHGARQLRIEGDLSLVIRQVKGlyATKSTRQL RNAVRHELARVQHSILHIDRQGNAFADLANRALDLKSDKVECKE HPVAGACTTCMSPSAGPATPTPTTADIEADAGSDDELADIDD GEVAPMRLEPGVPIPTRSRRLRLQLTDDMEAAAGEVVERLSAGLSA KIADADDWETAEGYITALPYMLYDKLQYQTVQVRHGTARSPAPHPQR RDVQGGQVETHREPRHETIGQDPQGPSPTRRRRRGRKRRQR HPRRTNCGGGRRQQRKQRPPTGTRHHREHRIEADIEDELHALER ARQARPAIAKARRRVGIIRSAIDQQLLRHFRDTAEKCEVDGILAAAR TARDARTTVRAAAATGTTATPETAVTSGTEQQDDNGTCTPISEVLW RHFDVNTPQRDFDPEAPEGAARFSAAMARLPAAATRFMELKEEPST DGIEVQLQHASSTSSPLDGVGYVYKRFASQLLPVLAFAFKCCTWTH KQVPQSWKLGVVRLLYKKGREDPANWRPICQQAQIYKIYTGVLARR LTRWQDANDRHAPGQKFRPVNGCGEHNFLAAMILIDHARRKHRP LYEVWYDFRNFAGSVPLGLLWDALERTGVPAEYIAAVQGLYDHAAF MVGNAVGGSTAPILQVRGVFGQCPSPPLFSAAISPLHALQLLSSG VQLSGDDRPGVSAYADDLKTFSGTGKAGVTEQHELVAFLWRTGMA ANPAKCRSMGVRNNGNGAIEADHLELALDDTPIPTLTHLSQSYTLGIG DGFHVRRRVALAPKLKQDATALESGLAPWQVVKAVKGYLYP RVEYALRHLRPDDQLLESFDLHRLRGLRHLRLPKSANNDFFVAPVSR GGLGLPLVELHAALQIAHGWQMINSPDPAIRRIAREQLHQVADAR HRLDKDHWKQRGDELCELLNNGELGTSAHAPPKRRNGDGLSWVDV RKNLKAFLGLKATAPADPESGAPAKPLQLCVPHAEWLDRHNVLRH VKQHMKNKRWRAWCSHVDQGRARAHGGVSGFLTRPRGMWE SDYRFVAARLMLDVTNVLARRRLRAHDCRHPGCRWKETLAHVL NHCPTGMDSIRGRHDDALKEIERTLHASSGDRQGRTELRTNQTVPGL AGPALRPDLQVYNHDQRTVAVVDLAIADFDEQPRDDPESGLAKAAA EKKAKYAGIKRHLERQGWKVLHLSALVYSGSVAPSNYKVYTEHLGLL KRDAKRLDQLSVACIQSSRRWNLHCAQHARARHQDQDPAPRGR VTETGTPSRDTR (SEQ ID NO: 1447)	NeS L	Utopia -4_PI	Phytophthora infestans	AGCTCGGCTCGCGGCTG CCTTCCAGGCGCGCGCG ACTTCGCGCTCTGGCGCG GCCACAGCGCGCGCGCG AGCTCCAGCGCGCGCGCG	TAAACGGGTCACTTGACC GACAGGCGCACCCACG GTAGGGAACCGCCCTTTA AAACCCAGGAAGACACA AACACCTCCACATAGTG	MLADPAALAGLARAPPPSPAPDPSAFPAGPAGQNPRAAAPAR VEVHTVAPPGRAGGMLPDPGLVEEPIQATYAHDAQFEFCALCPYV AESMAVLVQHRRSAHRGTRFKDIFTSGCCQCSLIFYARIVAAASHAVAC ARRNQAVPPAPTVPVAPTRPEATPQPTGYLAAAMTAAAAASDVT VAAATNMQSAVPAAKTTGLQLVPPLEPALPQRASCHAGKRRRLN
----------	-----------------	---	----------------------	--	---	--	----------	-----------------	------------------------	---	--	--

TTGGCTTTGCGACGCA GGGCTCGCGGACGGCC CTGCCAGCCCCAAGACC CCCCCTACGATGTGGCACC ACCGGCGAGGCGGCCGG CAGGCTGCCGACTCGGT ATCGCGGGTGCTACACTC TCAGCCGCTACAGCTCGG GCCTTGGCGGTCGCTATT GGCCCTTGGAGCTCGACG GCGACAGCAGCGACGACG AGGACGCTCAAGACCCCC ACGCCGCGCCCCAAGAAC CCCCAGAGAGCTCGCGA GTGTGCTTGGCCCAACCCG CAGGCGAGGACG (SEQ ID NO: 1203)	ACATACATAATTTAGCCT AGATTTCAGTTACGGAGA GGTTACTAACTGGTACAT AAAATTACACATTCGTT CTAAATCAGTGTGAAAACT GGTTTTCGCTTTTGGCG GACTTTTTCACTCGCATT TTGGGCAATCGTCTGCGG CTAGCTTGTAGCGCGG GACGAGCGGTCTCGGG GGCGTACACCTTTCCGCC GCGAGGCCAACTACACC GATCTTCTACACTTTTC TAATTCGCTTCTGCTTCG GTCTCGGCTGTCGATT TTTTCTTTTACCAATC AGAGTGCAGCATGCGAC TCTTCTGGCCAATCAGAG ACCGGCCCTGCTCCTCGG ACAGCGAGGCTCCACG GCCAGCCAATCAAGTCTC GGCAGCGACGCGTCTTTC TATAGCGCAGCTGACGA GGCCGATCTGGCGGCC CCGATTGGTCCGACTTC GGCCAATCAGCGACGAC GAGGGGCGAGGAGTTTA CACTTTTCCCCGTTTCG ACATCAACTTCAGGCCAA AATGGCGATTTCACCT CCACGCGCGTGCACCTG CTCGGCAACGCGCGCGA TTCAGCGGGTGCAACTTC GGGACGCTGTGCAACAC ATGACGCGCCCATTCGAC GCCAAGCGGCATCGCG GACGACGCTCGGCGC TCAAGCGCAGCCCCGCC	ADAEVTPCTPTARVSPQTEVAMAPHDAPQDDTVLQREAAEPQDP AATGQAQVQVEDTAAQDDTVQDHDADTAQVSPPTPTRW GPRSTQEPSPMTGEPAAQTAAARRPLTPAATQTRATRWGPGCHRAI GAAAIARLVTLPTPAQPPPPPPPPPEPLQPEPQAAAAATVAADI AATVAADIAAAAAAAMDVGGPAADETWLLRFDGACRNP GAGAAAFAPSGAVWTCSHFMPSETNNTAEYVALLGQASAVHH GAKRINIEGDSHLISQVRGAFACNNKRLSLRNRVQASRLQDLWYR LQHIDRKANQHADRANRALDLRRTVTECGPHAEIRNRCFTQPL VEPGETHCVPGSDEVLAANTAMEDATAVPTEDDEAEVAARDGGEV FPTIAIGPDSAPAKQPLRLKLDEDDFDDAAAAAVTRVSEELASKIVDA GDWTSGEYISAIPELRLAALRPFALPTQPARQPREPRMQPPRRP PRVTRDHLHLEHRLDEALDTMENVRQSTPQNKAVRRARRRVGRLRSA MDRTRLRKFAHRECEVAILRRASTEAAANPSQEKCPIDRATLHEY FTATSTQRTPFDDYSAKSGTEFRTELEVMTSPSHETSALTAEPTLDEID QLAHVKAGSSPGHDGVDYVRRFQVQLPLLLHAAFRFCWRHRRV PALWKVGFVRLHHKGGDPQPPNNWRPICLQTAIYKLYSGLLARRLSK FLEANELLPMQKGFRAFNGCHEHNFVATLLDQTRRMHRRLYQV WYDLRNATFGLPQQLMWGLVRLQGVTEEFVARCSTGVEDSYFVVG ASDGATEPVREVGQVQGCPLPLFITALVPLRLALENQDGVGVPLA DGVPCATAYADDIKVFCDSATGIQRCHALVTRFLEWTGLQANPAK AFLPVTRSQHSNPTDRDIELRIHGEAIALGIESVRYLGVGDGFDH VRHRLQLEPKLQIKREAVALLHSELVPWQILKALKVYVYKVEYALRH LRPLKSQLOGFDSAIVRGLRHLRLPENSHDGLFFSPTSAGGLGLSLV ELHEALQVAHAWQMLHSDPAIRAIARTQVGVQVARKRKLVEEHW RGREDDLAQRFNTELAASPHATETRRNGDGLSLWVNDVDRDTLQTLGL KFAAGDEEEAPGLLQLRVPHHTKWLHSTVLRHVKLHMKLRRMDT WKSQVSGTTRVREHGGVGSRFITAGLSDAEYFAIAPRAHLIDTN STLKRRLRANDTCRAPGCSYTEPPAHILNKCSPNMDAIRKRHDDAL ERIALRKRKVEKSGRLEVAINKTVPEYDGAALRPDIVLRNTEKRAII ADLAITHENQPTDATSSALQQSRDNKITKYQTVAAAMMRAGWRV RVTGIVYGLSVLPNSFKVYVTELLALLKRDARRLNRLSSHCIRASARI WSAHCRRHRRQRSGNASRASRGSGGAPRRTSQASARR (SEQ ID NO: 1448)
---	--	--

NeS L	Utopia -4_PR	.	Phytop hthora ramoru m	<p>CCAGGAATCACCCCGCC GCCCCAAGCGCTCCGCC GCCGAGCCAGCTGCGGC TGCTGCCGCCCACTGGTG ACGGCAGTCGAGCTGCC CCTGCCACTCTCCTGTTG GCCGCGCCCGTGGAGC CCGCGCGCGCGCCGAGC CGCCCCAAGAACAGCAC CCCAGCTAGCGCGCGCG TGGAGCCC (SEQ ID NO: 1204)</p>	<p>TTCCAGCACGACCTCGCG CCGTTTGGCGGATCGCCA TCAAGACGTGCGAGAGC CAGCGGGGTGCGGCAA AATATACTTACTCTAAGT ATGCCGAATCCCTGCCC TCTCAGGCTGTACGCGGC CCCATCTTGATCTAAGT ATGGGAGGATCCCTGGC CTCTCAGGCTGTACGCGA GACCCGTACGGCCGAAT CCCTGGCTTCTCAGCCTG TACGCGGGGCTATACTTG GTCTAAGTATGCCCGGT CGCTGGCCTCTGAGCCCG TACACA (SEQ ID NO: 1326)</p>	<p>MSGDVSSDGSRTTDSAGDGGDAGSSDAAGDVGVVAMDV/DQ GARRQPPWQVRVGGKRRRLNDVDDDEDTRELAELLEEEDEAGDHA PAPRLSASARPASVLSVYAHNAQRFQCTCTYTAASFASLKRHRDSR HRRTAFLDRFSAGCAGVFPASRLAAANHAHACDSLNRTFVSAAAP AAGELSPTAGAANATVKAATVTPDSPRQDPPKLAATPPLASSALVVD PDHAEQQAERWGPPLPRTLAVGRVAAARLSEVPAPRWGPPLPRGV VAFRIGHRVLPPEMTSDEETKDDSSVQDGDQDYPVAAAMDVDSDGM SGEWLRFDGACRANPGPGGAGALSQPDGSSVWWTCSHYMPSSSE TNNTAEYALLLGTAAADHGTTTTLRVEGDSLVIQQVRGIFATRSVT LRHLRDQVKLELARVGRFSLHHIDRQANAHADRLANRALDLRRTVSE CGVHPDNGCTPTAIDDRPLAPTQQPPDAPPPPAADIEMEDPDDE DLADIDDGEVYAAAMRVGNATPQRRRRGRSGTAKKRRRQRPVVT RHHREHRLDEALDDLHAVERSTPSDRTTVRRRRRVGRVNSAIEQQR LRHFRDTEKACACVTDILAKACATREAAARTASGGDPAGPATPAAGS ADDGTCPLIGEELWRRFFDSVNTPRQEFAPDAPVGAARFASALARLPA TSCKELLTAAPSAGEVEDQLQHVGRASSPGLDGVGYDVYQHFAAQL LPALTAFAKACWTAKRVPSQSWKLGVRLLHKKGARED PANWRPICL QQAIYKLTGLLARLVRWLDANDRHAPGQKGFRAVNGCGGEHNFILA ATLVDQARRKRRTLFVWYDFERNAFGSVFPFALLWDALARLGVDDY VTMCKGLYESAAFFVGNAGIDGTTDPIALRVGVFGCPLSPQLFNAAIS PLLALQRLPATGVQLSGDDDCPGASAYADDLKIFSGTEDGKIKRQHALV</p>
----------	-----------------	---	---------------------------------	--	--	--

NeS L	Utopia -5_P1	.	Phytop hthora infesta ns	CTCAAGCCTAGCAGCGGC TACCGAGCTACTCCAGCT CCGGTAGCTGCTACTGCTA CAACTGCTCGCGTGTCTGC TCGCGTCCGCGTGGAGCA CCAGCGCACGCTGAACC GAACCAAGAACATCTACC G (SEQ ID NO: 1205)	TGACGCACCGTGACATA GTGCGGCACGTGAAGAT GCACATGAAGCTCCGACA CTGGCCAAAGTGGCGGG CCATGCGCGACCAAGGC AAGACAGCTCGTGACAT GGTGGGTTGGTAGTGG CTTCTCACAGCGCGCG AGGCGTGTGGGAAGCCG ACTACCGGTTCCGGGTG GCCGGCGCTTAAACCA GGTAGACACGCACAGTG TCCTAAGCGCGCGCGCC TCCGAGCACATGACAGG TGCAGACACCCAGGATG CACGCGCTCCGAGACGCT GGGCGATGTGCTTAACCA CTGCGACGGAACCATGG ACGCAGTCCGTGGCCGG CCATGACGCCGCACTCAA GATTATTGAGCGTGCCT CCTCGCATGTCGGCCGA CCAGCAGGACCGTGCTG AGCTCCGCGTGAACCGAG ACCGTCCGCTCACTCGCC	ADFLRWTGMAANPNKCMTMSVQRDGRGVKLTDDQLDLQDLAGTPIPA LSMSASYTYLIGDGFHVRRLVELAPALKQLKDDATTLQSGLAPW QVVKAVKTYLYPRVEYALRHLRPFQQQLEGFDRHLARGLRLHLLPG NATAECFYAPVSRGGLGLPLTELHAALQVAHWQLNSKDPAIRRI ARVQLRQIADARHRIDSRAWEGRDEEICEILLNSQLGTSPDAPPKRR NGDIGSLWVDVQRHLRTLGLKFATAPACADAGSAATTLQLRVPHHD KWLDHRTVLRHVKLHVKKHRHWKSWAAMRDQKGTARAHGGAGSG FLTRPRGMWEADYRFVAARLNQLDTHSVLKRRLRAHDHCRQPG CSRAETLAHVLNHCAGTMDAVRGRHDDALKHIERALHASSPGGQD RVELRVNQTVPSLAGPALRPDLQLYNHTKMKMVAVWDLAVAFEEQAS DDPESSALARIAAHKRAKYAGVKRHLERQGWKVLHLSALVYVGSILGAVP AGNHKVLTEHLGLLKRDAKRLDRQLSVACIQSSRWNLHCSQHRAR QHQAPEGGSRAAETGGTPRTGRR (SEQ ID NO: 1449)
					MARVLEPMQVDECSSCDRSTLTADGSGDDVAAPSSLSNDVAA PMDVDSGTDCPPALQPPQPRALHVGSKRRLDADDEEEARQLQ EEEEAGIHAPALRLSAASAQPASVLAVYTHNASRFDCTLYAYTAGSFA SLKTHNSRHRRTAFLDRFSAGCAGCVPFASRLAAARHAQACASLSS APLAEASSAAGASHTVDEADSTVSAAGHTEPDLPRHNATELTASPP LVSPDVEVQAPETEATENRWGTLPRVLVASRIAGRLAQVPPPRW GPPLRTTIAGRIATRLAATAPAPWDPLPRSLVSVRIAARLLPALDA PACEEAKDSDTMDWAPTWTNEETKDEPHDEAPGQVDEETIDDA DGEWLLRFDGACRANPGGAGAAKLFPSGVPVWVWTCSHYDPSITTA TNNTAEYALLIGARAAADHGVTKLVEGDSLTIVQVGRGIFATRSTR LRALRNKVKLELARVGSFSLHHIDRQANGHADRLANAGLDRRTKLE CSVHPDGRGCTNTSVATAAPAPAAPLPAPPPATTAAPSHDDHDS VQGDIDGGEVVAAMCIGPDAPVPHRRPRLRLHRLTDEESEEAGDVVE RLAASLAAKIADAPDWETAEGYITALYALYDKLPYSQAQPPPSQ QQQQQQPRQQRQQRQRRRGRGGGQRRQRTRRRRPPR VTRHREHREIDEALDDLHAIESRRPQDRTAISKARRRVGRISALDQH QLRHREFTDEKVCVDAILAGARASQGATTAPPSATTDPPAPMDDSR CPIPGDDLWRFFDSVNTPRRSFDEAPDGAAFREAMACLPAAATRAQ ELLTEAPTVEVEDQIQHARASSPGLDGVYDIYKQFAAHLPALHT AFVCCWNHHRVPOSWKLGVRLLHKKGRDQDPANWRPICLQQTIV KLYAGILSRFRVRWLDANARHAEAAQKGFAMNGCGEHNFLAATLV DHARRKRKELHVWVYDLANAFSGVSPHDLLETWELARQGVPPTEVDC CRGIYSDAFTIGNAADGTTAPIRLVRGVFGQCPLSPHLFTAAIAPLLH ALKRLPVTGVQLTGVDRPGAAYADDLKTSSVDGKIKRQHELVAFL RWTGMAANLSKCSAMSVQRDRSGVLKLTGDLCKLINDAIEIPALSMT	

					GGCCCCGCGTACGGCC CGACCTTCAGCTCTACAA CCACACCAAGAAGACGG TGCGGCTGGTCGACCTG GCCGTGGCGTTGAGGAG CAGCGAGTGACGACGC GAGTAGCTCGGCACTGTC CCGGATCGCCAACCAAA GCGAGCCAAGTACGACC ACATCAAGCTACACCTCG AGCGCCAAGGATGGAAG GTACACCTCTCGGCACTC GTGTACGGGTGCGTTGG GGCGGTCGCTAGTGCCA ACTACCAGGTGTACACCA CACACCTGGGGCTACTCA AGCGCGATGCAAAAGCGG CTGGACCGGCAGCTGTCT GCCTAATGCATCCAGTCC AGCCGCCGATCTAGAAAT CTACACTGCAGCCAGCAC CGGACTGCGCAACACCAG GCCAGGCCCAAGCCGAGG ACCAAGAGGCGAGCCGGG CGACGGAGACCGGGGG GACTCCGTCGCCAGACAAG CCGCCGCTAGGCGGAAA CCAGGCCCAAGACGGCC GACAGGCCCCACCCAG GTAGGGAACCGCCCTAG AAACCATTTTCGGTGCTC GACTCGCAAGCCTTACCT ATATTTAGACGTAGCGA CCAAATTACAAATTTGGT AACGAGTAAGCCAAATG GTAATACACAAACTTT (SEQ ID NO: 1328)
					ASYAVLIGDGFHVRRRLELAPMMKQLKHDATALMQSGLASWQV VKAVKVLYPRIEYALRHLPFKQQLAEFDEHLRRGLRHLLRPTNATS AFFSAPTSRGGGLLPLTELHAAQLQJAHGWQILNSPDGATQRIAREQL REIPDARHRLDTAHWRNRDAELCELLNSQLGQSSHAPKRRNCDIG SLWIDIRRLQGLTLGLKFETAPGRRSHQPARPAIAAFACRTTTS (SEQ ID NO: 1450)

NeS L	Utopia -5_PR	.	Phytop hthora ramoru m	CCAGGCATCACCCCGCGC CCCCAAGCGCTCGCCGC CGAGCGAGCTCGCGCTG CTGCCGCCCACTGGTGAC GGCAGTCGCAGCTGCCCC TGCCACCTCTCTGTTGGC CGCGCCCGCTGGAGCCC CGCGCGCGCGCGAGCCG CCCCAAGAAACAGCACCC CCAGCTAGCGCGCGCGCG GAGCCC (SEQ ID NO: 1206)	TGATCCCGCAACTACA AGGTGCTTACTGACACC TTGGCTGCTCAAGCGTG ATCGAAGCGCTGGAC CGGCAGCTGTCGGTGGC GTGCATCCAGTCCAGCCG CCGATCTGGAACCTGCA CTGGCGCAGCATCGAG CACGGCAGCACCAGGGC CAAGCGCCAAAGGGCAG TCGGCGCGCGAGACCG GGGGACTCCGCCACAG AGCGCGCGCGCTAAGC GGACATCGGCGCCGTAG CGGCCGACAGGCCACA CCCATGTAGGGAACCGC CCTCTAAACCCGCCCGT ACATTATGCTCCGACAC TATGAGGTGCAACCGGT ACACAAGTTACACACCAC ATAGCGACTACCGGTAT TTACTACTGGAAGCCAA GGATTACCGGTCGGTA ATACACATAACTTT (SEQ ID NO: 1329)	MSGDVSSDSSGSRRTDVS VDQARRQRPWPQRVGGKRR AQAPALRLFAASHPASVLS RDSRHRVSFVDFKFSAGCA APAAAGDLSPATGANATAS VAIQADAAEQTAWTRWDPL SRTLVASRIAARLDAQTSR PPLRTVIAADRLPELAADE ADVMTDHDGEWLLRFDGAC SRYPSSATNTAEYALLGARA RGIFAARSTRMRALRNQVQ RALDLRTVIEGICHGVDGCT EVDVVDVCGVCGGDRGTG AVPARPRLRLKLTDEEQEAG GYTALPYLLYDKLLPYSQGP RSAARRRQRHRRARRPRV DRRSIRARRRVGRVNSAVE STTVARTAVGEPNSGAAED APVGAPFRAALHLPAAATE DGVGYDVYKAFSTQLLPVL KGPREDPANWRPICIQQVY KGFRTVNGCGEHNFLASTL ELLWEVLQRMGVPAPFVFC VGVFQGCPLSPHLFTAISPL DLKIFSGTAGGVKQRHALV RGLVKACDLEQLDGLDGA LMQLKDDATALLQSLGAPC EGFNHLVCGLRHLRLPLV RRPAPPRPRLEGAGGENM VWGRRAPPRTRPOARDRAG PHGPAAGEAAHEEQALAA RPMGDRPLPCSGGAQPAQ GTRAESLRHHHGRGPRPR PCPRSPAPRCGPTSSSTT WRASPTSEPTPASSGTSSA NO: 1451)
----------	-----------------	---	---------------------------------	---	---	---

NeS L	YUREC i	Ciona intestin alis	<p>ATCAACCCACTACTACACC CTCTACAAAGAACCACCTA CAGCAAGATCTGGCGAC CAACTACACCTGCAGCCTC CCGTCGACTCTTACACACC TGCGCACCTCCCTCCGACT CCAAGCGGACAGACGTGG CACCACAGCACTCCAAG GGTCCACAGCCAGAAAG TGCAACCCCACTGCTAGG AAGAGCTAGGCCATTGCC CGCTGTGGCCTTGTGCT GGCTGTCTGCCGTGATTG GCAGTTAGGGAGTGTGC GCTGAGCACGATAGGGAG GCTGCTGAAGCCGAAGAA CCGCGCCGGGACTTTA GGACTTTAACTAAATCTG TAACACTAAATTTGAATA CTGGATTACTATTGGATTA CTACTGGACAATAACCTGA TCTACAACCAATTTGGA TTACAGAACCAAAATTACAA TCTCAACTACAAACAAAA GTGAGTCCCGCGCGGGT TTCTATTATACCAATTACT GTTTTTACAATTTTAAAAAT TTATAGTATTTTACTAATTT TTCCGCCCGCTAGCACTT AAATTGGCACCCCGCCCC CGSTCCAAAAAATAAGATA TAACCCCTCCACCACTAAA CCCGGGTCACCCCTCAGA ACTAAATTGGAACCGGATC TTTTCCCTCTACAAATTC CGGTTTCAGTTGGCTTAAT TACCTAAAAACCGTCTTTA TTTTATCCCTTTATTGCCCC</p>	<p>TAGCAAAAGATTGGTG TTCTGGGAGCTGCAGT CCCGCAGAGTAGCGGA GATCGTAGCTCTACACC GCTTCACGATTTGACGAG TTCTCTTGTAGTCGATCCC CAAGGCAACGCACTCCA AATTTAACACCGCATGG CGACGCCACTGATCTCCC AGTCATCCGTGCTCCGAG GGCACTTAAGGGGAACC GATTCACTGATGCTGCAA GTTGGGGGGGCGCACT CACCAATGGAAGATGAG GGAGCCCAACGATGCC CGAACTACCGCACCGTCC CCAGCACCGTGTCTTGT GAGTAATTAATTTTCACTA ATCCAGGACGGGGCCA TGAACCTGAATGCTGCTG CACGCTGCCCGCGCTAGC CTCGGCCACAAATAAATA GTCAAGCCGGGCGGTAC TACCAATTGATTCGGATT TGGCGGGGCCCATCACG CGCCTCGCGCCCAACACA CAAAAACACTTTTAAACT CTTCGGTTCCCAACCCAC TACAACAAAGCGAGCGG CCCTCTTAGATCCAATTT TAAAAATTTAAATCAGTG CACTCAACTTTTTTACGTTG TTGTGTTTTTGTATTTT TCCACAGATTTTGTATTT TTATATATATTTATATA CACAACTATTTTATAC ACTACCTTGCACGTGCC ACTTTTGTAAATTATCACC</p>	<p>MATSSSSSGNVQTEVRCVYHGKDLLLECPVAHCPSIHPVTATITK HLKHHTPQEQITTKNLTITYTSCQCSFSTGLTQHHSKHYYTKCGV GAVQEGNKGRFCPCACGTWRWALLCKARHFNHNVHFEYDTPPIAASF GTPYKLKRKFTINKALTYSCPLNQLLCLPLWSSCSTILNKPLSVQQ ETAHGDGSGQSVVPTQLRQLRARCHCNGNPIGKGHWASQCGKR PLSPKGRSSPTPANLTLHFLNVLPLQPSQSSSQSSLDLPTACKA RVPIPSFLRGCEVTFIIPSVNFYRPLSYPLRMFWRNRTSSGHCSLH RVVRGVRERLVPHPRSKSPARTPLELCEFRLAGVFPDPGKVASLRP VPAPLTLCLSPVAGPMISCEHSDHSPSVRSSPSPINSPASVSSVEAHL SDLLDKVSSGELRPLSPTLPSSGFFGLPPTPPRPTPSAEKASPSGLS YLPCEVVKIASARPSASQSVGCDADRTGPSLNPNYQTSPPSPSF SPVRRPKFPRSGAKVNSKPPGVRPRRAKPIEGTESASPDVDVTIS SSVQEPCTPENRTPEFFYERKWLVLNLIHIEREGSNFFQFNRLDLEYWT QLLSGQKGGRAKRSYNRGAANQAMKNRDSGRKDFDPRPVAGH SSGGGTGELSRPYPKARLRADFWRDMKGTVRKLLDGSNGERRCG IPLDIIERKFRQVSMGPWIDHRRYAAAGASPSLVTOAETDVAITSEVE AVLSGLNVQSAPGSDGLSYRWFKGLDPSGRLLSCLFEIVRRHGRIPGA WPTCSVILLCKDAQGDVDVGNWRPTITICRTLYKLYAAVARIQTW AKQGGVLSRLQKGFMPVEGVFHVFMLDVLSDAKLRRKNLLAVFL DVRAFSGVSRHECLLVLRHFDAPHYLVELVRDIYTGATCRVRSVGE TGDIPCDRGVRQGYPLSGILFNLVTEVLPLGSLAGNDGYRMACLGKGL TQVLAAYADDLVVTENRDQMLRQLGVCEEFGRWAGLAFNRKCGGL IGWRTLRRGRRVALEDPLLLNGVEIPLLRPGEHYKYLGAAMTGVMVSV RTGSQLIKDFARLQRLFTSLTPHQKIALKRLPLSLFHLRVRPIARS ELIALDRVRRECLRVAFRLTKPSQAVFHTPTDMDGLGVPSVCSSESL TIAQGFKVLTPDGTVSATASARVKLYAAKFGGLTEAGPSDWARYLS GDDVNGNSTRKPGANLPSGLWTRVRCASRQLGAVVRVCPENGITV RVNSVITSRRRKLIRSHDCSNQWKEQWQMHPNQEKTAAAH MAYADANRWVKQPSVMEPHYFFALRNLPTVRSAIYSRDQH PDILCRRCGASVESLPHVLNHCPPNMSIILGRHNLVLEVLNVDKQ FKEISVDRTPVPEHMSSETGEALRPDIVARRNDGWSVVVDVACPDFDQA NFDEAAKRKLLKYDKLCCNIAASTGKPVCHSIWVSGLSGLAEGLSTSL RALGITDFARSKLVACHQG (SEQ ID NO: 1452)</p>
----------	------------	---------------------------	--	---	---

	<p>TTAAACCAAAATTATTCAAA TTGAAATCCGGCCAAATTG CCTTGCTGAACCAACATTT TTGTTTTCTTATAAAGT AAATTTTTCAACAACCCAA AATCATAAAATFAAACTG TTATCTGATTAGTCTAAAT TCTGTTATTGACAAAACC TCCATACCAACCTTAATCT TATCATATCTTCACCCAG ACTGCTCAATCTAGTTCTT TTTTCAATCAATAAGGAGG TGTTTTCTTTGGTCTGTGT TTTATAGTTTACCGTTATA ATAGTCTGTTGCTCTATAG TTTTCGTGGTGATTCTT GGTCTACTATAGTTGGTGC CTCTAAAACATTATTATAC CGAAGAGTGCTAGCGATT TCTATTATTAACACCCA CCAGTACTGCTGTGCATCC ACGACGACACCGCCTGC CATCAGTACAGTGGCTT GCTGCTTGTGTTTTGTTTT GCAGACCACTGGACGA TAG (SEQ ID NO: 1207)</p>	<p>TTTACCTTTTATGCGCT CGCTGAGCTCCTTGCACG GATGACCAGACAAACTTT TATAAAATTATAACATTG TTTTAATTGCGGGGT ATCAGTGGCGCCCTTG CGCCGTGAAAGCCCTTT TCACCTTAATTCGCCCTT AGTCCAAATTTTCCCAT TGCTGTAAAAGTGGTT GCCGTGACTCGAACTCG TCCTTTTACGCTCTCT GTTACTGAACTATGACCA GCTTGGCTGTGAAGTCG GCTTAAATTCGCCGCTT CACATTATATTTTTTGT GGTAATCTGTTTTTTTC ATATATAAACCATTCGG CTTAAATCACCAATCCC CATCATCCACCTGGTCA CCCATTGAAACATCTCTT TTAGTCAATTATTTTTCA GATTACCCCTCCAAAT GTCATATAGTTTAAACACC CCGTTGCCAAGTTGGTC TTTTCCCATCCCTTCTG TCCTCCGGTAATCCCAT ACATTTTATTCTTTAAT CGCCCCATAAATAGTGT AATCTTAAGACGAAGTCC CCAGAGGCCGGTCCCTC CCTCTTTGTCATGGACC GGCCAGCCCCCTGTTA CCACAACACCCACCAT TTATCTTCTTCTTTTA TTAGTATTTATTTATATG CCGCCAACTGTGTGTC GTGGCGCAGGGGGTTC</p>

					CCTCTGTCGGCCAGTGG ACGACCGTCTAAAAACA GGCAGGCGAGCAGGCA CTTATACCCGGTACTC CGGTACCGGAGGCTGG TTTTCAGCGCACTGTACG TGATGGCCAAATTTATTC ATTGCAATTTATCCGGT CGTGGTGTTCGTGGAT GCATTATAAAAGATGAA ATTCC (SEQ ID NO: 1330)	MPVSTGAETDITSSLPIPASSIVSPNYTLPDSSSTCLICFAIFPTHNLLSH ATAIHISCPTPVQDGSQMSCVLCAAAAFSSNRGLTQHRRHISEY NELQRQIAVQPTSRWSPFDDASLLSIANHEAHRFTKNDLCQHISTIL TRRTAEAVKRRLHLQWSRSPATITSSNNHTIDIPNTEARYIFPVDL DEHPPLSDATTPNASTHPLPELLVILTPSPTRLQNISESQTSHESENKN SMHTPTYACDPDETILGATPSSTIPSCFHSYQDPLAEQRGKLLRASAS LQSSCTRIRSSLLAFQLNESTLMDEEHVSTFLNSHAEEFVPRTWTPTS RPKHPSHAPANVSRKRRKIEYAHIQRLFHHRPKDASNTVLDGRWR NPYVANHSMPDFECFWTTVFTKTNPSDSREITPIPMTPSLIDPLPS DVTWALKEMHGTAGGIDRLTSYDLMRFGKNGLAGYLNMLLALAYLP TNLSTARVTFVPKSSSPVSPEDFRPISVAPVATRCLHKILAKRWMPFLP QERLQAFNLNRDGCFAEAVNLLHSVIRHVHTRHTGASFALLDISRAFDT VSHDSIRAARKYGAPPELLCRYLNYYRRSTSCVNRTELHPTCGVKQG DPLSPLLFIMVLDEVLGLDPMTHLTVDGESLNYIAYADDLVVFPAPNA ELLQRKLDRIISILLHEAGWSVNPEKSRITLDLISGGHSHKITALSQTEFTIA GMRIPPLSAADTFDYLGKFNFKGRCPPVAHIDLNNYLTEISCAPLKPQ QRMKILKDNLLPRLYPLTLGIVHLKTLKSMDBNIHTAIRKWLRLPSDT PLAYFHSVPAAGGILHLSSVPFHRKRRLTLLSSPNRLLHKLPTSTPT LASYSHLSQLPVRIGHETVTSREASNSWVRRLHSSCDGKGLLAPLST ESHAWLRYPSQIFPSVYINAVKRLGGLSTTKVRRSRGGRVTNGLNCR GGCAHETIHHILQHICALTHDIRCKRHNELCNLVAKKLRQKHFLQE PCIPLEKTYCKPDIIRDSIAVVLDTVSDGNTHASRLKISKYGNERT VASIKRFLTSSGYIITSVRQTPVLTFRGILDRESSQSLRRLCFSSRDLDG LCLSAIQGSIKIYNTYMRGT (SEQ ID NO: 1453)			
					ATCTCACGTTTTAATTTATT TTTGAACTACTGCAGTCTG AGTGCTTCTAACGACCCGA AGGCTCAGAAACTACCCA CTTCTTGAACGTGCTACTTT TGCTGTTTATCCACAACAA CAGTTGTGATTCTATTCTC CANATAATCTTGTGCTTT TGTCACATTTATCTATAC CAACTGTACCACCTACTTC TTCATCTCACGTTTTAATTC TGGTCTTATTTTCTCATCAT TAGTCACGGAGAGGGCCT ATGAACGGTCCGTGACGC GAAATTTCTATCCGCGATT CGACCTCTCCTGCTAGTGG TCCCCGAAGTACGGTTCTCT CTGGCCTGTCAGTTGTGTT AAACTATATAATAACG (SEQ ID NO: 1208)	Schisto soma manson i	BN000 800	PERER E-9	R2
					TAACGGCTGAACGAATA GCCCTTCACTCTTAGA CATTCCCCCACTGTTGT GCTTATCTTCATGCTCTTG TGTTAATTGACTGCTCTCT TCTGGGTTGACGTCTGAT TGCTCTCTCTCTTTCCAT ATTGCTGTCTGCCCCG TTACTTCCAATAGTTGTC ATATTATGCTTTGTTTAC TTGCCATGCTAACGACA ATTACTTATCTACCTTAG TTTGCTCTTGGTTTCGA TTGCCATCATATGTTTCATG GCGGAATCTGATGTTTAT AATGACTATTCCTATTAC CACCCTACAACTACTAT TATTATTTTCACTACTAT AACATTATTATAAACATT ATTACTATTATTATTATTA CTATTATTACTTCTACAAT TAATATTATGGCTACTCC TCTCAGCACACCAATAAA ATATCAATCAACATCTC AATTATATCCACCTATTAA ACTCTCTCTATTTCCCTCTG				

R2	R2-1_BTe	.	Bom s terrestr is	<p>TCAATAGTACTCGGGG CAGCGGGATATTGGTCT TGCCTTGCCCAAGTCACAC TCCTACTCTCTGTGTAC CGCGGTAAACACGCGCAC GTCCACATCAGCGAGGG CGTACTCCCCGGATGTG GCGGCGGTGGCTAAACG GAGTGTGGCGACGAAGGA GCGAAAGACTAACAACTA TAACGGTCTCCGTAACGG CTACTTGGAGCGGTGAAT AATGGAGCCTATATTAAAC CCTGGAACCTGTTCTCTCG TTCTGTTGACGACTGGAAC GGCAAAGGACATGATTG GATAACAATTGGAACCTTA ATATCGAATTCACAATTAA (SEQ ID NO: 1209)</p>	<p>AGTTATAAACTTACAATT CAGTCTAACCGAATATCT CTCTTTACAAATCTTAAG TATGTAATTTTGTGCCAA ACCCATTTGGGTCTGTAC AATTTGATACTTAAAAAT AAATGTTATTAGCC (SEQ ID NO: 1331)</p> <p>TGAAACACGATAACGAT TATGAATCAAAATAAGAA AGTAAACATCCAGAAAT TGCTACGTCTATTGTT ATCTATTATTGTTTA (SEQ ID NO: 1332)</p>	<p>IAKFDNNTNSASDAAPLSPGGAVADLSASEGTTDNDQAMSPAMSLX TVPLVGNRVACPCXCEKEANLFFNLISDLDRHLTQHHPDAPXWSCISL DCAKCFPKLHGARCHIPKCGGASSQARTGEFQCEACPMFSGSRGLS THERHAHPAVRNIRKRGADPPPEENTKSWKVEVARLKLGLWEIFKDH KYPNKEISFLTCTVDDXXYQRKKNLNLIGXESQEAATSLATEGGCDLV SSGNASFGSPVGRNENEELIHEWKLKNEINKPTEVPPIKKEVYNRL MLIWEEHQDDRDSLTESLDHFRTALYELINKINKNQDCLKTKRAAKT KSPKNRNSRKRFSYARCOELFHECPRRLADAVVNDQAYLEPARQ PPGSEEVRLGLEYKLWGVQSTVVPAPVTRVPKLSLSEIFPIAAEDVG ERIGKIRKAAAGPDGLQRDHLTIPGLPIIMAKIYNILVYCSYFPSAWKE NRTLLIPKINPCSLVENWRPITISPILGRIFSSIIDGRIRRGTVLNMROK GFTSENGCKINIELLSALNYSKRNSGGIFTIVDISKVFTVPHAAALKPC LAKKGPALIVDLIDEMYKNVKTITKDGGEIMEIRRGVKQGDPLSP LLFNLCLEPLEEIEEQASGINVSEHRKVS/LAFADDIVLLGADAREAQ HQINVLTDYLSLMMNLSIEKQCTFEVWAKKDTWFIKEPGLKIGNQI MPTVDPDEAFKYLGAIGPWKGVCVIVPELLSVWVRKRLSLKPG QKVELLTXYIFPRYIYHLLVSPSDTV/LKLDSVROEVKTILHLPSTAT GFFYTPKACGGIGIPREHIIKLTLSAIAKIANSIDPAVAGLIDDAIKKL KQTANSLRINWPASLEDEIEKARKLRKEHISQWADLKCCQGGVPDFI KNKTGNLWLEDHSLKPSRLIDALRLRTNTFTGTRSVLARADKNIDVTC RRCRAQPETLGHILGLCQHTKGLRIKRHDEVKSLEGLRKSNNNEVF VEPTIKAGGSLFKPDLV/KNGERVLDVTVRYENKNYLALAEKEIEK YRPLRALKEIFNAKGGEILPVVLGSRGTITPNTKVLKRLGIANNIDIKTI LLNVLRSSELNCFIDD (SEQ ID NO: 1454)</p> <p>VPPGAEARGRYHHPRXEXARQGPSPXRVFLVXLIPDSNPPCPCGDH VXXSVLALHCEVGHXWAXVQYQCTHCGLCHIPRCQGRVXEXTGK DXXCEPCPAFDEKAGLSQHKRHTVTVXSXERVAGXLLRAXLRHGCWS VEEETLRLDAMFXGARNINQLIAAEXVSKMXKQISDKWRXLXLP QTTXGXAESASVXXESMTPEMAEQSPAXPPGKIRKIFTGQDGH GGXAWENQEDFWTRWARRWLKRGQXLSDKVQEVLGXWVEGQ</p>
R2	R2-1_Crp	.	Crocod ylus porosus	<p>AGGCGTCTCTTTAAGGG CAACGGTCTGGTTACGCG GTCGACGAGCTTGCKCC AGGTACCTCCWCGTGGTT CCGCGGGTGCCGMAG MCCAGGSGTGTGCGGTAG</p>	<p>TGAATCCCACTCTGGGGA CCCCCAAAATTAGAAAA ACCAAAACAGTTGTGTT TAAGTGTCTTCTGTTTG TCCCTTTGGCTTCACCTCC AAGTTGCGATCCCCCATC</p>	

CTCGATCTGGCACAGTA MGGCCAGGGGAGTKCTTC CTTGCTGCGWGGTGCCCA CAAGCGTGTGGCAGSMC MCCTGCTTCWTCGCAAM AATMASAGKGTSMTCAGT AGTCGGCCCCGCGCTAG CCAAAACGTTCGCCACSC AGTTACAGATGGTGCTG TGCTGACCKGKGCCTCGT GGGCTCGGWGGTGCAAG GGWGGCGGCCCATGGC TGGGCCAGACSTGGGCK TGGAGCCCGCTCCAMCC CAGAGTCCCCCTCAAMG CCGCCAGGKAGMAMCA KCCGGGGAGKGMACCGG CCCCGGCGTGA (SEQ. ID NO: 1210)					TCCCCTGCGCTGCTTTCT GAATGACCAGTGGTGTT GAGGCTGGTGTACCTC GGTCACCTCAAGGCCAA GTGCCCTGGCCCCGAGTA GGACCAGGTGGCCAGC TTGCTGGGCACTCGTCAC CGCCAGGGCAAAATGGA AGGGATCTATCTGACCA CTACCAGGCTAAGTGTG GTGCTGCCCTAGCCTGCCG TAAGGTCAAGCGCCTGC TGCCGCTCGGTAKCAGT CCTCGTTCACTCGTCCCTC CTAGTACCCTCCGCTCT GCGCTTCTGCTATCCACT GTGCCGGW/GATGCCG AGGWTGSWMCAWCCTC GACACCTCAGGGCCAA GCGCCTTGCCMCAGGT AGGACTGGCACCTGCC AGGGGCCAGACACTGC CTGGGGCAAGGAAAGG GAGCGTCCCTGACCGTT ACCGGCTTGATGCTGT GGCTAGCCCGCGCTATGT CAAGCACTCCACAGCTGC TCGAGTTCGCTGGCTCA CCTTCATCCWCCTAGTG CCTTCGCTCTGCGCTA TTTTGTCCTCGACTCGTA CCTCCCACTCTGCGCT ACMGCTATCCTGMAAG GACCAAGTGGCAGGGG SGTTGCCCCCGCCCCG AGGAGCTCGGCTATC CGTGGCKTCWTCCTCC ACCGTCTTGCGCGCTAG	PRXAWVDXVSLVLTFLGVPPGPQAPSKKGPXEGGKPTSWMNK RAIKWGTFLRYQHFLGANRLAAXILDGAXRNQXTLLLEEVXQYXG KWEAEPFEGLRFGXRDVDSFAFALITXEEAVKHMVMXMAXXA PGPDKLTLRDLRRADPEGDALAEFLSLWXITGVVDPRLKEXQXVLIPK AVDFEKLRLQGNWRPITIXSIVQLXSRVLTARLTAACPIXPHQQGFIS APXCAENLKXPELIXKVKXDRRPLGVAFVDXARAFDSVSHDXISWVL KAKGVQDQHNLIEDSYQKVTMRVQVFGSTPPIKXGVKQGDPMIS PLLFNIAMDPLIXKLTVRQGVKVGASLTTLAFADGLXLLXDSWEG MQHNITTSXTPGRACNTTSRHPRLQLQPHGPTSATXKMXGVLLSX MRLLYGEQLRGLDXRPPXHDAXARRADTISGLEXRSLGWDXQTRF GYATXLLAREDDGDCXAQTNAEALCWXSXFPFGCAXRXPYANXGWV ASEALDSMSRRXVKEWFHLPACTDXLLXSRHRDGGGLLRARXXLA AXVRRPIRVATSSDEVTRKVSACGISDEVERLXLAGGDXSNVPRFE DPXAPKXXVQGPHEAAQETPRVWRTQAIPWPSNWRAEEHKKWA QLSCQGERVELFCNDPVSNGWINSRGQLAERLWIMALKRSNIYPTR EFLGRGQAGTNIGRCHTTPRETILGHILGICPAMQEARILRHNLCKI LAAEGKNCEWTVYEPHILHNAAGELKPDLI FVRDGTALVDITVWY EGGPATLLSTTAEKATKYLDINTQIQLTGAEQVTFYFGPIGARGKWH ADNWRVLSLGLNSRKERVTRLLSWRALLGSVDMMVNFVSKHRQE NLLDEHCTPAEQVVSYYAS (SEQ. ID NO: 1455)
---	--	--	--	--	---	---

					AGGGTACCTCMGAGAC CGCGCAACACGACCTT GACGTTAGACAGTAGG GTGMAACAKCCCTGCTG CAGGCSTGAAGGGCCAA ACGCTGTGCCATGAGA GGGAACCTTGAAGACC GGSCAGTCAGCCAGTTA GTCAGTTGGGGAAACA ATCCAGCTGCAGGCCCA MMAGGGCTGTCAGGTG AGGGGTATCCCAWCC ACCCCGCGCCGACTA CGGAGGCAKGAAGTCCC TAGTGACTTCKGACCCCC ACGTCTTGTGCCGGAG AGGGAACTTTGAAGAT CGGGCAAGCCGCACTT GATAGTTAGCCAGTKGG TGAAACAATCCCAGCTGC GGTCCGAAAGGGCCGA CTWCCAGGCGAGGGGG GCCTGCGGAAAMCCCCC TCCATGGTACGGAGKTCT GGCGTCCTMACCGACWC CTTGCCACCAACGCTTTG TGCCGGGAGAGGGGAAC CTTGAAGATCGGGSCAA GCTGCACTTGATGTTAG TCAGTCGGGTGAAATAAT CCCAGCWGCCCGCTG TGACTGCTAAGM/CWGG TCCCCAAGGGCATGAG GCATSTGCGCTGAGCCGG SAGGGGTGACACMCGGC GATCGGCGCAGCACAKA STGAAGGGGAGGCACTTG CTGAGACTGCTTCTGAGG

				CCCAGACTTGGGGTGG TGCAGCCTTGTCTGGGT ATGGTACAGCACCTACT GCTCCCTTTGGKAGCAG AATTGCTCCGACCTCTT ACCCACCGAGTCTGCGC CTGTTCGCTATCCTGC ATCTCGATCCACCTCGC TGCTCCCCGCTGCGCTG CTTTCTCTCAAGTGGGT TAAATCTTGTCAATGATTA CCTCCACGTTTCGCTC AAGGGCAATGCCCAAMA TGACGGGATCGCTGGT GCATGGCAGTCATGAGA CCATCCGGACCTCCGGT GGTCGCTATAGTCATTT KTGTTGCATGGGGCATSC TGAGTCACTTAACCGAAA GACTCWAATAAATACTCAA AAGAGKAMCCTCTGSG GTTCCGGTAAA (SEQ ID NO: 1333)	FERRNSWGYRPLEPRSVGTESNNNSPRSNITTTSATSRPGDQPREAI AVVNLAGEIPCAVCGRLFNTRRGLGVHMSHQHKDELDTQRQREDV KLRWSEEEAWMMARKEVELEASGNLRFNPKKLAIEVTHRSSEAIKCF RKRGYKAKLEQIRGQSTPTPEALDSTTSQPRPSLLERNHQVSSSEAAQP INPSEEQSNWEIMRILQGYRVPVCECSPRWRAQVLQTVDRQAQVQKE TTLQCLSNVLEVFPLPNEPHTIGRSNLRPRTRRQLRQYEYQVQRR WDKNTGRCKISLLDGTDESVMNPQEIPEYWKQVMTNPSTCSCDN TRFRMEHSLTVWSAITPRDLRENKLSSAPGPDGTPRTARSVPLGI MLRIMNLILWCGKIPFSTRLARTIFIPKTVTANRPQDFRPIVPSVLVR QLNAVLASRLASKVNWDPQRQGFPTDGCADNATLVDLILREHKKR WKSCYLATVDVSKAFDSVSHQAIKTLQAYGAPTNFVSFIEEQYKGGG TSLNGAGWSSEVFIPARGVKQGDPLSPLLFNLIIDRLRSYPREIGAKV GNTMTSAAAFADDLV/LFAETPMGLQTLDDTTVGFLASVGLSLNADKC FTVSIKQAKQKCTVVERRSFCVGERECPSLKRTEEWKYLGIIFTADG RARYSPADDLGPKLLRLTRAPLKPQQKLIFALRTVLIPQLYHQLTLGSV
R2	R2- 1_DWi	.	Drosop hila willisto ni	GAAGCTGGGTCCGGATGAG CGCAGAAAGGGGTGTTCTT TGGAACACTGTAATTCATA AGTCGTAAGTCTGATCAA GTCGACTCGAAACCTCCTC GTGGTGTTCCTGGGTGCT GTTGAGTTCCTAGTCTCTA GGTCTTTTCAGTAGCTAA (SEQ ID NO: 1211)	TAGATGTACTAACCTCTA GCTTTTCTTACTTTTGC CTGCTACCTTGGCATTAC ATCTAAAAGGTACAAAC ATCGATTGTCATAAAGA GGTGGTTTGTAGACGTAG GCGCTGTGGGACTTCATT GTCCGGGTGATGCAGTG AATCGTGCATACGAGATT GTCCAGTAGTTGGTTGCT CGTATCTTTAGAAGATTT CCTTCTCGCGGATCAAA AAAAAAAAAAAAAAAAAAAA AAA (SEQ ID NO: 1334)

R2	R2-1_Gav	.	Gavialis gangeti cus	AGGCATCTCTTAAAGGGT AATGGTCTGGTTACATGGT CATAGCAGGTTTGTGTC GGTACCTCCAGTGGTTCC CGCGGGTSCAMAGCCC CAGGGCTGCGGTAGCTC GATCCTGGTACAGTACGG CCAAAGGAGTTCTCCTTG CTGTCGGGTGCCTCGCAA GCACTGGCAGCCCCAATC GCTTCATTGCGAAAAACAC AAAGCTCCTAAGGGGATG ATCAGCTAGTCAGTTCTGC CGCTAGCCAAACTGTTTG CCACCCAGTTACAGATAGC GTCTGTGCTGA (SEQ ID NO: 1212)	TGAGAAGTTGCGAGTTCT TATGCAAGTTGAATACCA CTCTKGACCCCAAAA AWWAAACCCCAACA GTTGTGTTAAGTGTGT CTTGTCGTCCTTTGGCT TCACCTCSAAGTTGCGAT CCCCCATCTCCCTGCG CTGCTTTTCAAGACGGCC GGTGGTGTGAGGCTGG CGCACCTCGTCACTC CAAGCCCAAGTCCCTG GCCCGAGTAGGACTGA GTGCCCCAGCTCGCTGG GCACCGTCACTCTGG GGCAATGGAAGGGATC TGTCTGACCACTACCG GCTAAGTGTGGTGGCGC CTAGCCTGCGGTAAAGTC AAGCGCCTGCTGCCACT CAGGTATCAGTCTCGTT CACTTGTCCTCTAGTA CCCTGCTGCTGCTCTT TGCTATCCACTATGGCCA GTGATGTTGAGGTTGGT GCATCCTTGGTCACTCC AGGGCCAAAGCGCCTGG CCACAGGTAGGACCTGG CAGTCCAGGGGGCC AGACACTGCTGTGGCA	MIGVLRKCDRLVRQFVRRWLDLDPVPVAYFHAPHTCGGLGIPSI IAPMLRLKLSNIKWPHLEQSEVASSFIDDELQRRARDRLKAENVQLCS RPEIDSYFANRLYMSVDCGLREAGHYGPQHGWVSQPTRLLTGKEY LHGVKLRLNALPSKSRTRRGRHELERRCRAGCDAPETTNHILQKCYRT HGRVVARHNSVNAVVRGLERKGCWVHVEPSLQCDGSLNKPDLVGI RQNHIVIDVQVTDGHSIDQAHQRKVERYDRADIRSQMRFFGAT GEIEFHSVTLNWRGIWSGQSVKRLIAKDLLIAEDTKLISVRVNGGVT SFYFMYCAGYTRS (SEQ ID NO: 1456) PAAPRAWGAVEAGWPGRTRVAPESPSPSEAAARAAPAGEGH GPGHESPSVQRPEADTTAPGVSAPTREGPEPSTRVLRLPDSNPCC PICRDHVGKPSALALHCVESHAWADVQYQCTHCKKVSANKHSILCHI PCCQGRVPEWTGKDWAPECPASFNNKVGSLQHKRHVHPVTRNVE RVAGLSRAGLRPQTRRGCSVEEETLTCLDAMFRGARNINQLIAA EMVTKMPKQISDKRRQLGCLPEQTTLLGGDAESTSVVEESMTPEME TQSPINPGKIRKILAQRRARWLKKGQGLSKVREVLGAWVEGQPRI HAWVDSVSLDLTLFLGVPSPQRPAPNKKRPKEGGKPTSWMNKCA VKWGTFLRYHLFGANRKLVAIVLDGADRNQCTLLLEEVFOAYREK WGLEEVLRAYRGKWEVESSFEGLRFGVRRDADNFAKALITPEEVV KHMAMASKSAPGDKLTLDLRRADPEGDALAEFLSLWLTGTVP DGLKECRSVLIPKTVDRKLGQGNWRPITIGSVILRFSRLTARLAA ACPINPRQGFIAAPGCAENKVLLELLRKRKRDRQPLGWVFDLARA FDSVSHDHISWVWKAKGVDEHVNLIJEDSYQKVITTRVQVFNVTPIIS IKTGKQGDPMSPLLFIAMDPILAKLETGQGVKVGSAASLTTLAFA DDLVLSDSWEGMLKNISILEDNLTGLRVQPKCQGFNLPTCDSF TVNNCEAWKIAGREITMLPGESTRYLGLNVGPWVGIDKPDLTQL SSWLERIGTAPLKPQKLSLLVQYAIPLNYQADYAGIGRVALEALDS MNRKKEWFHLPACTSDGLLHSHRHDGGLPLRLAKAIPAEQVRR LIRVATSSDEVTRKVSACGIDEVERLWLRAGGDMSSVPRFEDPEAP RSPGVQGPCEAAQEIIPSVVRKLAIPRPSNWRKSKHWAQLSCQGE GMELFNDPVSNGWNNSRQGLAEHLQIVALKLSNIYPTREFLGRSQ ASTNVGCRHCTHPHETLGHILGICPAVQEAIRIHNKLCILAAEKKC EWTVYVELQLLNAAGELCKPDLIFVRDGTXLNVNVTGYEGGPAXLLS TAAEKATKYLDXNAQIELTGAEQVTFYGFPIGARGKWHADNXXRVLS ELGNSRKRVARLLXWRALLGSVDMVNIIFASKHRQENLSDXALSP S (SEQ ID NO: 1457)
----	----------	---	-------------------------	--	--	--

					AGGAAAGGAGCGTC CTGACCGTTACAGGCT TGAGATGGTGGCTA GCCACCATATGTCAAGC ACTCCACAGCTGCTTGAG TTTGTGGCTTCACTTCA TCCACCTAGTGCTTCT GCCTGCACTATTTTCAT CCCAACTCGTACCTCCC ATCTGCGCTCCTGCTA TCCCTGCAAGGACCAAGT AGCAGGGGGGTTTCATC CCCCTACTGCAGGAGAC TCAGCATATCCATGACTT CTTGCTCCACCGTCTGT GGCGTAGAGGGGTACC TCAGAGACCGGCACAC ATGACCTTGACGGTTAGA CAGTAGGGTCAAAACAC CCTGCTGCAGGCCCAAA GGCCAACAGCTGTGCC ACGAGAGGGGAACCTTG AAGACTGGGGCAGTCTG ACCATGCTGGTTAGTCAG TTGGGTAAATAATCCA GCTGCAGGCCCAAAAGG GCTGACAGTCAGGTGAG GGGTATCTCCATCTGCT CCCCACTGCCAACTACGG AGGCATGAAGTCCGTAG TGACTTCTGACCCCCACG TCTTGCCCATGAGAAGG GACCTTGAAATTGGG ACAAACCGCACTTGAAG TTACTAGCCGGGTGAAA ATAAGTCCAGTTGCGG GCCCTCGGGCTGACA GTCAGGTGAGGAGGCT

R2	R2-1_S	.	ixodes scapularis	<p>ATGACGGAGATCGTTGG TGCATGGTAGTCACGAG ACCATCGGACCCCTCCAG TGGTCGCTATAGTCATTT TGTGTTGCATGGGGCAT GCTGAGTCACTTAACCGA AAGACTGTAATAACTCA AAAGAGGTACCTCCGG GGTTCGGTAAA (SEQ ID NO: 1335)</p>	<p>MQCSTRADAPRFARVGEVGEVGSNGTDAQLWYGCTGCDEA FSSLRLRIHAAQKHGNDGLLRLPAGRPRKRRVGKSTTAGASDRV TTDVPAPVPESGGLPLGPPSLPGCSDLPGLPGGWSASPGPLS WPPSLDAGPLPGPSRVSPSPSPGKPTGPPSLDAGPLPGPSRVSP GPSRSPGKPPGTPEPLPGSPGRRGVSPPGQSGSRTPDSSSAGAGHF VCPQCSRAFSSKIGMSQHQKHAHLEYNAGINITRTKARWDPEETYL LARLEATLNPDHKNINQTLHAALPRGSCRTLESIAHRKQAAVRDLVT SLRSARSESAQHVDPDRPLETPEQTPANPQRDSKQAVIEALQSLGR APPGSFQGARLWDJARQATRGTNILPLNLSYLRDVTLPKTRKKPA VRPARSRKQKQEVARTQDLFRKKQSDCARAVLDGPTSSSVPGTG AFLQTWREIMTGPSPALEAPPLPTRGEVDLFFPATAQEIQSAEIAVNS AAGPDGFSARLLIKSVPAALLLRV/MVNNLLVRRVPAALRDARTTFIPKV PDAVDPQFRPTVASVLQRLHRLAKRALEAIPNFRQRAFQPVVG CAENIWLSTALNEARTRRRPLHMASVDLTAFDRVTTDAILRGARR AGLSGEFIGYLKELYTTSRTLLQFQGESLLVEPTTGVRQGDPLSPILFNL VLDEYLSLDPDISFVSGDLRLDAMAFADDLIVFASTPAGLQDRDLAL VEFFDPRGLRVNKKSFSLSQPGRDKKVKVCDQIFTIGGTPLPASK VATPWRYLGMTFTPQGSINKGTSEQDLDLLTRTSKAPLKPPQQLVVL RNYLLPRLYHRLV/LGPWSAALLKMDTTIRGAIRRWMDLPHDTPLG FHAPVTEGGIGINSRASIPAMV/LQRLDGLHFTSTHPGAEVAIQLPFLT GLHRAEAAAQYQGRLLSKADVHRMWSARLHSGCDGRPLRESKR VPAHRWAAEGTRLLSGRDFISITKLKINALPTLRTSRGQHKDIQCR AGCQAVESLGHV/LQACHRGHRRIRHNDIARYVCGRLTQIGWAVK WEPHYSVAGRTLKPDIVAHRAETVVLDAQVVGTSMRGLGFHHAQK KEKYSLPDLLHQVCEGRDAAVSTITLNFRRGVWAPESAQDLKSLGLT DNDLKLLTVRCLQGGACQCFRLHRRMTTVKATGDDEANALPAHSGLP PTQLGGRTLGPSTAHNQASRTT (SEQ ID NO: 1458)</p>
R2	R2-1_S	.	ixodes scapularis	<p>GTTCCAAAGGAAGGCACT CCTTTGGTTCTGTGATGAGA TGTTTCATGGTGTGCTGCTA GCTGGAGAAATCCGACTC ACACCTGCACGTGGTCCCT GCCGCTGCCAGTATGCC GAGGAAACGGGTGCAACT TAATCCGTGGATACTGGTA GCAACGTGAGCAACGGTA CGGTCTTCGCGGACCACC CTGGGCGTTTCGGGTTGCC AGCCCGTTTCGCCCGAAATA TCTTGGCCCTGAACTAAA AGAAAA (SEQ ID NO: 1213)</p>	<p>TAGTGTGACGGAGTCCTC AAGCCCCACAAAGTGCTC GCCAGGTGGCAGGAAAG GGCAACTACTGGTGAGC GACCAAGCAAGGCGGA GCCAAGACCAAGCTGGA GCCAAGAGCAACTCCAG GAGCAGGGGTGGATAT CAAGAGCAACCCCAAG GACACAGACCACGGCA ACTACTGTTGAGCGCCCA AGACAGGGGTGGATATT AAGAACAGCCCCACAAA GTGTTACCTATATTAACA ATAAAGTTGAAGCCTCAA CCACGCATTGCGGGTTAG ATGGCGTGGCTTGCCCC GCCGCCATGATGAGCTG GAACCTCCACCTGGTGG GCCGCACGAGACCACCG GCTCTTCTACTAAGGCC GGTCTCCGTGACTGCGGT TGGGATAAACTCCAAGCA CTGAGCGGTAAAAAAA AAAAAAA AA (SEQ ID NO: 1336)</p>

R2	R2-1_MLe	.	Mnemi opsis leidy	<p>TGGGGGCCCTTGACTT GCTCCCTGGGCGAGACA CCAGTGAAGAGATCCT CAAGACAGGACAGAGAGA CAGGCACCAACCTTCG AACCTGAGGAGACACCA GATCTGTTACCCATTCC TGTAACCATGGTTGCTCTC CGTGCGCTGTTAAGGAAA CCAGCCTAGTACCCTCG GGAAGGTTGCGTAGTAC TTAGCAGTGTCCGATCA AACCTCTACCGTCTCTCT ACGATGAAAGTTTCTCC GACTGGAACCTTGAGGAC TGGCTAGCCAGCTAGCCT GTAGAGCAATGCGTATAC GATGCTTGCGGACAATG GCGACCGCTGCTTAGCAG ACGGAGGTTAGTGAAGG GCGACTGCTGTTTCATAGT CACGTGAGTCTCTAGAA ACTGCACCGATGCTGCCCT CTGTTCCAGGAGAAAGGAC AGTGGAGGACTAAATCGT AGCGGAGCGGTGTT (SEQ ID NO: 1214)</p>	<p>TAGTGACGAGAAAAAGTC GCTTTATCCTTACATAAC AGTGTGATAGTACCTCTT ATCACAACCTGCTCTGGCG AACAAAATGCGAGAGGG ATAGTACTCGTCCAACC AGAAGAACCCACACAG ATGCCAACTTGCTGTAA CCACGTGAGCGGAAAA CATCTCCGAGTATAGTA TGATGGAAGGATACAGG ATGTGGACCGCTCTAGGC GGGGAGCGGTGTAGACG GCGACCTTACCATAGCTG CGACTGCTGCGGATCCG GAACGGCGCTTTCTCACT TACCTCAAGCACCGTGTC GGATCGCGGGGTGGGCG CGGGATTCCTGCTGGGA AGTTGCTGCACCCAAGT CGTCAGCTAACTGTCGGT ACCTATACGCACCCAGAG CCATGCTCTTACCAGGCC ACTTGACGCGCTACACC GGTCTGTAGGAGGTAT CTCAGCGACTTGGACAAA ATGGATCTTGACTGCCTG AGGGCGCTTATCGAGTG ACCCAGAGTAAGCTGGT AGGAAGAATCTTGCACTT GAGAGGCTAGTAGGGC CAGCCTCTGCTCGTCAAC TGTCCTAGGCGTTAACGT GTGCTCCCTATGAGGTA ACCGTGAAGTATTACTTG TTTCCCTGTAGAGACCAT ATAACTAGCTGAGAAAC GCCTGCGGGTTAATATA</p>	<p>MSNTSHSKLNKMDNKLKLTETPSGVRADSIITRVRTSSNRGEHSNG VTYPRCEQGVAPLDTHGGICDAPQVTVPATETDKQKCEYCEFTYL KPRQIGHTMRKRHPQEWNDIKRTKFLSEKQRKWLDEDPELLCIGQE EYLVSSIGKQKGKINQYIQTKEPTLSTDAIKSQKRRFSEYSEKRSR ELQPCNTSSDPEELNEAVTENSPLSFDPDLDRDVVKISSKHGDQILL VQEHNGRYQEANTLAKAIFELKSGFPNLKTGDHRPGKQQTARKV GKKVRGSGKKLSPSKQNRRELVAIQKQWRTKRSKVINQILTGNL NKEQSYHTPDQLAQFWSTLFGVSPRDDRPNINHRRSVPELDKPLSV EEVEAALKAKDAATGIDGVPIHLKHLGSAALTILYNGLVVTGSIIDP WKRARTILIPKSNPPASPGDYRPISSYFYRIYTSISKRLASAVSLDDR QKGFIEDGIRDNLSDTLINETKAGSKSLFMFMVMDVKKAFDSVSHY AIARSLWAGVPDGMRSVADLYQDCTTDCGRSVKVTRGVKQGD LSSTLNLVIEMVMSNVPERLGIQFQGHRLFLAFADDLVLLTRGPTA NQKLVSLVHEQLARVGLHLPKCKSIAMADPKRKTTFVDQGSVLI GGEPVSSLGPQEWYKYLGIKLGSGGMPQGIYRDQLADLLAKTDSAPL KPQQLYILRSHILPKFNHRLMFERVTCQTLEGLDKLIRTHVRKWLKLP KDTGPAFYADKSGGLGLITLRYRVPPLKLRHKKMADSPDPVIRLIP NAEPTISLARWTKMCSLYGKQKHSELSKIIRDYVWTMCDGKGLR TEVPPDTAKTISLLEFEDRTPKPGQLIGAVRLNTLGTARNNRKAK GYSPEANICDKCPGNRQATLGHISQTCPTHGRRVKRHKIVNRIAK ALKERGSVKNILTEPHLRHDKLPLRKPDLIVHTEKSVEIIDVQVADQG ISRHEDEQKKIVKYDQVGYKRAAYKMLGIDYGSIPCNVSAFTITWR GNLAPHSLKLASRLQFSPVLKYVADSLVDTWGAFLIWGKTS (SEQ ID NO: 1459)</p>
----	----------	---	-------------------------	--	---	--

R2	R2-1_PM	.	Petromyzon marinus	<p>CTATTAAATGGGATGAAGA AGGGGACACGAGTTTGT GTGTGATCCAGTTTCCAT GGTGCATGCAGGAGTGGT GGTTAAATGGCGAGACT CTACAGGGCTTCCATGGCT ACACGGGATGCAAGGCAT CAGACATTTTGGCACAGG CAATCCTTTTGGTCTCTAC CGCAATCATGTCTTAGACC TCAGTAGCGACCACTACAA CCACAGTGGTGACTGCTG TTGAGTGAAGGACGACTG AGCGCTGGATAACAACTTT CTTGCGTGGCCCAACATCG AAGCAACCACTTCGGAGC TGGCACAAAGGCAAGAGGG CAGCCCAAGGTGTGAATC ATCTCAACTTCACTGCAGG AAGAAATGCTGTGCAAGG</p>	<p>CAGTACCGCCGTCCTGC ATCCCGAGTTGTGACCA GCGTACAGCCATGTACCA CGGCACCTCCAGCATTTCT CTGGTCTAAGAATGTATC AGCTGCCCGCCGAAAG ATAGAGCGACCCCGCCTC TATCACTAGAGTAAGCAG GCACGCAATATATGCA GGACGCTTGGCTCAGTA GCTCTGGTACTAGTATA GCATGGCTGAACGCCCT TTAAGTTGGTGGCCAG AGACTCTCCGAGACTCC CAGTCTGGAGAGGGTA CACCTTTACCAGAACTTT CTGTGGTCGGTTG (SEQ ID NO: 1337)</p>	<p>TAATTTAAGGTAAATCG TGGGATGTTTTGATGGC AATCTGCCTAGTCGCGC CTTCCATTTTGGGTAGGC AGCAGACCCATCTATATA ACAACTACTTTGCTTTTC ATAGGGGTACCCGACCT ACCAACTTTCGGGGAGT AAAAAGAAA (SEQ ID NO: 1338)</p>	<p>MNERLTDELTFEFLSDMFLWDYPCTDQNKCYPCNLVFLDHRTWSS HMARVHPHANKTYKRICNRTADSIHKIASHYGRCTCKSLIGKTNATTT TIDETLFSCLHCSRGTFTKGLGVHTRRTHPTHEAILQQNTPGRKVR WGEVEEIMAHKEAQQKDEDINMNQLIQNSVMPHRTLEAIGKRR NIKYKELVRTLKETTYKVENQCLVNLVLPPTTSEITTTTSEGDDQPAIRAEK EQSPAAEDLQVINDLKSQNFHNAQALLLNHVEKFLNRSKPIKRKD HVNQEIENRHRQSKQTKYRRYQYLYHTNKKALLDEITSDRSGPSI YPTESIRGTFVTLFESNPPDNIPSKLNQDQSCIDIVKAITDELIKTLAI MKDKSPGQDNITLSDLRITLPIKYLILNLIYQDIPQIWQHRTRLPIK TKEELEKPSNWRPITISSIVIRLLHKILSYRLGQQLKLNRYQKAFLPVDG CFENSALLHFIIHNARQKHENTQIVSIDLSKAFDSVSHESIIRALNRFNL SKESITYLTNIVKCNLTDIVFGTMRNINILKRGVKQGDPLSPLLFNMI MDELNDLPTYGVNVGNQKVNMMFADDLILFAETECGMNKLLDI TKFLDDRHLKININKNSLRFIKYGKQKTFVATTSSYFINNEPINPVS YVKGFKYLGIEFDPGRKRSICNLLAAMNLKLTAPLKPCKKVVYLININ LIPRIHQVLVGVTKGLLMSLDSEIRKTVKLLRLPHDTPDSFFYTSVSN GGMGRNLCDVSVALSIINRHNKLTSSDDLVRALSQQSYTIATLKQAHII AGSKFPSKSLNQNKWSNKLQYTTDGRGLVYCQSQSTENNSWITGNH RTIKSVNYIDMVKLRINALPTKSRNCRGTLETQKCRFKCRSINNQISEET LAHLQKCDRSHYSRIARHDSLVQFLATAAAQKLNWEVIKEPTLPSDTN</p>
----	---------	---	--------------------	---	---	---	---

				<p>ATGAGTGTGAACGACACC AACGGGATTGTGCTGAC CAGGAGGTGCCAACCAAA TTTGAATGGATTGACTTGG GGCCTGGTTTCTCTGCGT GTATTGCACGGAAAAACA AGTGGCTACACGTGTGGC CGTCGTGTCCTGGGGTTTC GCAACACAACCTCCACAAG ATCGACAACCTATGAGGAT GACAAATGTACTTAAAGAA CAAGAGACTGACGCCAA AGGGATTAAACCGCCA AATCGTACATTGGGTCTCA CTACAAATTTTTTACGTGT ATTATTTTCTAAGTGCT GTACTGCCATTCTTCGCT GCTTTTCTGCATTAAATG CATATCGTATGCAATAAG CGAAATTAACCAACCCGTG CAACTATATGCAGATGTTA CAGCTGAGCCCTCTATCAT ACCGGTGTACTAATCTGGT ATGGTGTGGCATGCTAT GCTTGCCTAACGACCTTGG CTGATTGGTTTCAGTCGGCT GATGGTGGGTTTCAGGCGA AACATTTGTATATTGGTTT AATCAACCGAAACACTA AAATTTTGAACACAGTTT CCATTACACCCAGTTGTATT GCTAGAAAGTGCAAAATCGA AGGAGTCAATTTTGACCG ACGATTAGCTGCCGATGT GCGGTGAAAAAGCTGATC ACAAATAGCATACACTTGG GCCGACAAACCCCGTGTGC TATAACGTAAGTCGCGA</p>		<p>KAKPDLILVRDShVLIVDVA/PWESRSLAHAYDFKVKKYATDKKMQA YLKTIYPEKEIRTEALISARGGWCALNNMVTkKVGLSAWVKLALIKV MEGSVKIWRWSKSG (SEQ ID NO: 1460)</p>
--	--	--	--	---	--	---

R2	R2-1_SM	.	Schmid tea mediter ranea	<p>ATTATAAAGAAAAACAAAC CGGACGGACTACTCGGTG ACGAACATAACATCGCTC (SEQ ID NO: 1215)</p> <p>CAGTGCTATTCGAATGTCA ATGTGAAGAAATTCAACTA AGCTCTGTTAACGGCGG GAGTAACATGACTCTCTT AAGGAATTAAGAATTACC TGCKTAKTAAAAARTGAAA TCMGTGTTTCATWGCAG TGGTATTTGACACCTTCCC GCGGTGCTAGTCGTTTAA AACTAAGTTACAACCCAG AGGGCGTCTGACGGAC TGSAAAAGCAATTGAGRG CMTGAAGAGAGGCTCTTA TTGTACGAATCTCTCAAC GATCGAAGTCTGGACCGA TATGAGAACTAATACATTA GTTGACAGGTGAAAAATA CTGTTGATTACTTAGTTCT CAGTCATGTGGTATATTGC CAGTCAATTACTACATWA ATATTAGTGTGGCTCTCAA AGGAACACGATTGRTCGG CAGTCCAATGCGCGACTG GCGGGCTGTTGTTTGCAT TTGTTACCGGCTACTTGAA AAGGTTATATATAGCAGA CGCTTAAAGCGCGACTGT AATTACATYTCATTGCC AGTATTTGCTTTTGTGAG ATTAGCAAAATTCATAT TTTGTAAATTACCTTAAC GGTTAAACGATCCCAAT TGCTTGAATTTATATAAA GTAATTCAGGTAATAATTA</p>	<p>GGCCACGCGCTCGTCTCT TGTTTATCACTAGTGGA TCAACCTTCGACTCCCG GAAGTGGGAGTGGCG GAAGAAAGGCCAGAGGA TGCTCTGAAACCATATAT TTATTTATAGAGTTTAC TTCATCTATTTACGTATT TCAGTATGAAATAGTA AAGTTCTGACTCGATGA GTTGGGGCAACCATTG GGGTCCTGAAGAGAGG CTCTCACTGTAAAAATC TCTTCGTGCTGTTTATTC CTAGGCACNTGCTGCATT ATGAAGCGGWGAAAGT AAAGTTAGAGCTGAGAG ATAGGTACTTGTGCAATT ATGAAGCGGAGAAAGGC CTCGAATAAATAGGGTGT TAGAGTTATTGATGGAG AGTATACTAGTAAGCTTA AGCTGCGCTCGCGCGG TGCCCAAAATATACTTA ATGAGAGCAATAACTCAA GGNGAGTTTAAATCATAT GCGCATGCGGCACCAAG GTGCTGAATGGCATCGAT TAAACCTCTCTGTTGTA GAAGCAGGTCAATAATG GAGRGGGCAACCACTG AAACTTATGAGCCAGAA GAAGCTTAACACAWAA GTTTAGGCAATTACTGA</p>	<p>MKKVLNNETKLPGSNLTFMCGFCDREFDTARGVHESRGLHVER DAAVQSRVAVVSKYYVSNEEDVALAKMQLXHADIKSEXLEAMY LALGKTRTREAIEQHIRKSLRYKGVLEQRKILLETARGNVQRNNVGP ASNATKNLQRFLESLPLGTRREERLDRIIRSNSIESQRELIHYCNDM CQDFVLDCQXNPINAIRRNPKRLSKQLKRAKFSALQRLWIRDRK AAAQLVLKDKLSDLSLNKEDSKDLGSYWQQVFERESLDRRPIQV ENEELNSPVEKEVEWAVKNIKKSTAAAGPDGLTALALKIPYSELVKLF NIILLVGFLPDVLKNSRTILPEVDNPQGGGDRPISINSLTRTLNKILA KRVSEDFGQKGFQKSVKVDGLENLATVESILADARMKKNKLAUVF LDMKAFDSVNHESIVRAGEIKGPKLLMTYVKECYNDAATTNAVAGVT AKFNRGVKQGDPLSPALFNNDLAIERVSMTGIGYNNMGKKYSVA YADDLVLFGESREGIQLALTALLEELKLNGLTPNPAKSASLTFERSGPH WFASTDTVTALGDQIPAMGNIETKYVLGKFNCSGVVKGSLPGIYTKK LELSKAPLPQQRAMLTDLPGLVHQAQVFGQTNAGDRLSLDKRTR RAVRSWCHLPSTSTAFIHAKAKDGMGIPISIRAEVQFGKDRFGKL PNVKDESKVLADNAHIKKMKLEKLGVIPIKGVRCNKLEFYKNMR EELIKSNDGIGLKEASLVPASNTWLKLSDLHMSGRFTVGGCLKTRGNL MATVTRTSRGGQNPGLNCKKQYQYQSLNHIVQKCPWVKGLRIKR HDEVVKYVEETKKAGWSATMEPIPFEGSHRKPDLVLVRGDLGKW DIQVSDHCGLDKNSCKIGKYDNDIIRNVYRGLGPSRVEVAATLNV RGVWSRDSFNLIKRLGTMTEMDAKIISMRVLASTAKMFKCKKVLPEV CRTKTADCDGYGPEETSARPCHELNLKESST* (SEQ ID NO: 1461)</p>
----	---------	---	--------------------------	--	--	---

				ACGAGTTAACTGTTAGT TAACACTACCATGTAGTT GTTTATAAGCAAATATC AGGTTTCAGTCTATATAC TAAAAAGTATTTTTTGATA CCGTGGTATATAGGCAAC TAGTTAGGAAATAGTAA GGGATGACGCATTGTCTC TCTTTATGGTACTGAGGA AACTTATCGACTCGCGAG GGATGAAACCCGTAAAC CGATCGATYTAGCCTATA AGTACCAGCGACAGTTAA ACCATCTTACGCGAGGG GTAAACCTGAGGACCG ATTATGGTATAACTTCTC AAGATTAGCACAAAATGC GAGTGCAACTTGAGGAG GAGGATTGAGTGTTAAT TCATAATGTACTAATCTA ATTAAACTGTGACGGGA ATTGCAGCTTCGGCTGTA ATTACTTTGAGGCCTATC ACGGATTGTAAGGAACA TATTGACACCGTAAGTCT AACGTGTTCCCGATTCC AACCAGGTCATATGAAG GGCTGCCCTTGATAAGGC GGATTGACCCAATTCCT CATATGAGAGGCTTATC CAGCCTCCCGTAGTACC GTGAGGTTTTCCCGCTC GACGGAACAATGTTGC AGGGTAAATTAAGTACATC GGGCTATATMCGCATAT TTAACGTTTTTA (SEQ ID NO: 1339)	CATATCTGGCTGATCCTGC CAGTAGTCATTTTACTTCC GCCGCGCTATAAAACAGTT TAAAACTGAATAGGAAT CAAAAAGAACATGGCAAG CGACTATATGTAACGGG CATTCAACATTCCTATTA CATATGGTGGTGCCTGGG GTCTGTTTTATATAATGGG TACCCGGGAAGTGGATCT GTATCACCAGTCATGGTGC CATATCTTTKGATAAAGAT ACAGTTTAAAACTGCGATG ATACTAATAGAGATCCTCT TAGACCTTCGTAAAGAAAT GGGGATTGATGACATTAG CATTGGAAGAATTAAATCT CCAAGGAAATGGAGTAAC TTCAATGAAGTCCCACAAC CCGTTGAAGGGCTGGGT TCGAGTATCGAGAGAAAA CTCTAAATTCCTTCGGTT MTGTCCAACGGAGGGGA CATTACTGTAAATATCCT CTAAAAACAAC (SEQ ID NO: 1216)		
--	--	--	--	---	---	--	--

R2	R2- 1_SP	.	Strongy locentr otus purpur atus	TCTCGGACGCGTCTTCT GCCTGATGAAGTCAGTC AGGTAATAGACTTAGAAG GTTGATGAGCGTCTCTC CTGGACCGGGGTGAGG ATGTGTTCTACTGAATCAT CGTTCGGTTGTGAGGTCC GCTGCAATAGGCCTTGG GGTGTCTACCTCGGGT CGCTACTCTTTTGAGCTA GTCGATCCAGGTGAGAGT CGGGAAGCCCACTTAGG TGGGCCAGCTAAGCAGAT CACCCCCAGCAGCAGACA GTGCGCTGTAATATAGCCT GTTGAGAGTGCACCCATT ATACTGTTATAAAGTATA AATGGTCTTAGCTAACCT ATCCAATGATTTTATTGTT ATGTATAATACTGACCACCT TTAGTCTAGTGTGATGTA GGATCCGAGACAATAGCC TATGTCTCTAGTCTAGCAA TAACACAGAGGGATAAT CCACCTCCCTACTCCAGCC ACCACCTTTACCTTTCTTAT CTTCCATATAGATGCCAAC CAGACGGGTGATGGGCA GTGAGAAACAAGAAATGGG GATAAATTTGGAATTTGACA TTCTTTCTAAAGCAGACC TGAAAGGGTTAAGAGTCAC AAGCGGAGTCGGGACTGT CTTTATAACAGACAATAT TTTCTTATTTTACCACCTAT ATGGGGATATTTCCTCAACT CGTAATATAGGGCCCACT GTCAAGTGGCTGGTATAA	TAAACCTTCCTCCCGG GCCCTCCTCAGTACTAA GACAACTTCCACGTAAT AATCATATATTTGTATAC CATGTATTAATCTAGGTA ACAAATTGAAGTAACATT GAACCGTATTCATACTTT AATGAGTACAAATAGTGA AGGGAATGTATATTACA TACCTCGAGATAGAGGTT TTTGTACCTTAAGGGTTC GTGAGAAATCCATACTGCA TAAAGGGGTTAACTTGTA ACTCAACCCAGGGAGA CAAGCCGACAAATTCGGC ATTACTATGTAGGGTCA TAAAGTTTAAAGGACCCA TTGTATATAACTTGTAAA TGTACAGAGAAATTCGAGT TGAGTTCGAAGAATAAAC AAACAAAAAGAAAAACA AAGAGAAATCGTTCGTCC ATAGAAAGCGAAAAAGC GCAAAAAGGCGTCTCTAA CCTCACGGTCTAACCCAG AAACATTTAAGGGGGA AGGAGCTATTACTCCAA TCCAAAACACGTTCCCT AGCCAAAAGCGCGCTTC AATGAGGGAGCCGATAT TGGAGGACAGAGTTTCA CAAAAGCCACTGTAAAGT AGGCCATTTCTGCCAAG GACATGCGCAAAAGGAA GCAGATTATCAAAACAGT CAACAAGCACAAACATT GGGGGATGGGATAACC CCGGAAGAGAGGGATC	MENSAWEGTSSAEGRTTVEDSPSSDDFVSNVGFVAKADPTVWE EANMSEDNITIEDPPSSDDFANNVFKVTRADPTAWEEASTETE DLPSSNFIDNVETQIDMAGPTAWEDADSDNEDIDEETPNINNNL AIVRGRADAYACSCERNFISLKAIGTHLKETHNKKVVFECACQHTF VKAHGLACHVPKCKEDATPMLNRLHGCCEGLAFNRRGLSQHE RHRHPSACLSRRSRDLGIARCKSLRNRDIIWTNDEIRLLQLMIOY EHAKINIKIAEHFNHNAKQVMHKRRSLREKDMALGAPHDAPPPL AEEPIEVVEGARELEQAPVDVILPDEALTVDNRGGGGSPVLTGG ESTRDMEEENGDTDRSPSPREERAGSTPWERGWQPRVDRGRGEYK GYGGERGRHTVLSQRGESGVDP LTPGGVVVEDYDSDYLEDYFPGW DEDEHMHIGRLDLSDESEGEAAVSPRLGSFGDLLEEVNAMEGKD NLSEALAEITGLVLHEGHRVEYKEKMNINVKQMATEILAHGANKGN PKRRKEAVAAKRNPGTRLDRQRDNQAKAKAKEKRFSETQTQYKK NPHRLVEKLLDGDERCSVSLVQRTYMNRFRESKEVDIGAVDP ETVEDNQGIQVPIKAEIERAISTTKKSAPGPDGVTDALKAYNCQ LYLLIMYNTWLMGKVPSEAKTYRSILIPKQGGDPMDDNNFRPLTLA NVISRLYSKILTRRLDGA VSVCPQRQFTHKASIEDNTLIRELIMKSKR NKECLAVLLDLAKAFDTVSHDLIIKALRRHRVHEHLISVIMDLVEGGT TSFTTDEGTTCPAIRSEVKQSDLSPLFNLAIDPLLATLEQRKGVEI GGHTFVSLAYADDTALVSSSHLDMTANLDITVEYL NATGSLNVRKC QGFLLPINKSFLVNEAESWVVEREAIWPVEPGDTAKYLVGVQVGPW SRPWPSIQPVIKRLTAYCESIDKAAKPRQRIHILTTYIAPRIAEI AEGG YSTLVDCRGGIQTIRIREVDMTIRNVYRKWLELPACLSNFSLYTRRGE GGLGLVSYFYDVYPTERMRLKVRVCDSEDPVIAGAAAASGLRERAAKIS AQTGLPVVPKPGAHNAWRKVQKKWKAQPTQGGKGVSCYQHRLG NKWLGAPSLTENDYIWAIKRLTNLPTREAMMGRGIGRNQVCECRHC HTTIETMGHISGYCQMVLDIRLIRHNRIKALIKAAATATGLRVTEEPRI VGTGKNYLPDLIFSAGAGEPCVYVDPVTVVWDDDPKNLREAWRGK VRKYTPPIPAVEAMLHPSSVQJFGFVCGARGTWCPMNDDIKIVGLK NSGISRTLQIVLCLDTRIMVKAFMAR (SEQ ID NO: 1462)
----	-------------	---	--	---	--	--

				GTTACCTGTGGGTAACA AATAAATCAACAAGA (SEQ ID NO: 1217)	TTTAGATAGTGATGGA AGGGTGGGACGTTGAA TCCAACCATGCCGTTTT ATGTTCCCGATAAAGAAG GATAAGGTCACCTCAGCC TGACACACAAGTGGGG TAAAGAAGCTCCGCTCGT ACGGACTCCAATAGAA (SEQ ID NO: 1340)	MSGKRIVMSGCEKICQNHCLKRRWAWISGPKGETSPPRKRGTC ENVSFQDKSHASDPDLKAPEAREADAGSVAPQWVGEIKTPSLTSRD GVSEVLPQPVHAEGVSPASDSKD KATKITLLISLPVCDLRCRCGERP LETVGKAVRHFVAHPTVSVVFKQCKEKSXNSHSISCHIPCKCKGMT ETRTDVEGDHGDHCQEKFTTAMGLTQHKRHRHIVQYCKEKEGEM TARRKGEVAVKWSEWESEVARLSGLAGLKMNRRIADSLGTGKT AEQVRQRRMRPEKVRCDKPEAKSKNLKMLSPSATPTPQTGL KGFLGELNGVATKGEVQIGGVTSLRGVEQDSALLNTSALEQLRLG GRAGSANPLSLQRETTLPSEERRTKQGEYRRVQKMFRRSNEKKIAK YILDNGDGEAASPLEIALAFKSRWEEVETFHGLGQFYSRGEADGV VFRSLISMSEVCENLGAIKNNTAAGPDGITKPALLEWDPTGAKLAIF SIWLTSGTLPGFKKCRTLLIPKTDPIILLTQVAGWRPLTIGSVVLRLYS RILTHRLERACPINPRQGFSSPGCSENLMILGLIKRSWAKGERLAV VLVDFAFADSVSHSHILEILRQGLDEHIIIGVGDSTVDVTTTTVSGE QSPPIDMRVGVKQDPMSPLLFNALDPMIDTLERYGLGYRMGEQ QITALAFADDLVLVSDSWEGMACNIRILEEFCLRTGLRIQPRKCHGFLI QKIQARSVNLCKPWIVCGEELHMVGPESVSYLGMKVSPWHGIM EPDPVERLCNWISSIGRSP LKPSQKVRMLNVVAAPRMTYQADHGGI GPVNLVLDGMIRKAVKWWLHLPCTCDGLLYSRCQDGGGLGIVKLAC QIPSIQARRVRLVHSEKATRVVTRTVEAEYRGMWMLRAGGSEAG LPPLIEDREEGAVQCTDTAGSVKPNPVPIDWRRAEFLKWQNLTAQG VGVQVFGGDKNSNHWMANPETLSKERHYIAGLQLRANVVPYTR LSRGRDLPKVCRCQAGTESCAHILGQCQPAVKDSRIRRHKLCDLLA SEAESAGWTVIREMCCRTRAGALRRPDLVFKTGFALVVDVTVRYE MAYDTLMGAAAEKVARYPTIPVAMTLKARRVKVFGFPLGARGK WPGSNDRLKAMGVGGGRRKQLAKLFSRRALLYSLDVLDRDFYRAEG ETGDLDDDESVDHDL (SEQ ID NO: 1463)
R2	R2-1_Ssa	AGKD 01072 455	Salmo salar	AATCTTTAACCCCGACTC TTGGGGTTCTTACGACTCT GTATGAGAACAGTCGAA GAGAGGCGCTACCAATC CAAGTATATGTCCCAAGA GGGCTGGACAGGGTGG AAGATTGACCTCGCAT CTGGGCGAGGAAGGAAT GGAGAGAAGTCGAAGAA GGCTGTAGAGAAGGGGC TCTCTAGATCTAACCTG TATGAGCCCGTAAACG GTGACCCAGTAGCGAAT AAAGGAGGCAGGTGACAA TAGAGGGCAGGGCCGACT TCCAGGTTTACATTGTTG TACTTGTCAACATAAAGAG GTGTCTCAATAGTTGAAT CAACAAGGGAGAGGAATA CCGACCTGCTCCCTGGGG CGGGGTAAGTGTCTTAG CCCGGTTCCCGCAAGTTT CCTTTGCTGGGATGTGCC TGACTGGCTCATCCCTT TCCCATTAGGCACGGCTA GATGACGACCGATGGGC GGGTGTAGGTCGCTAC CGAAGGGGACTGGGGT GTCCGGTGAACAGGACT	TAGATCCGTTTGTATGA TTGAGGAGCCTGCGCG AGTGATGAGCGCTCC AACTATTGAACCATATG ATTCCGAGGCTCGGCA GACGCTAGATGCTGCC ACAATTGAACGACCCCT AGCTTGCTAGGAGATCCA TAGGAAGCTGGCTATGG GGCGTCATGACGGTTGA AGTTCCTCATAGCGTGC TTGGAGGGGACGACAA TGACGAGTCATGACGTAC CGAGAGAACCCCAACC AGTTGGGGGAGAGAGC CAGCAAGAGCGGAGATG CTTGATACCAAGCTAG CAGAGAGAGGTTGAAG AGGATGACTACTGGCT CAGAGTCATCTCACCTA AAAGCGGTGGGGCATC GGTTGAACACCTACCCAT ACCGGATGGGAGGTGG TAGGCCGAAAAAGAA GGAAGATGGTGAGTAA GTTGAGAGCGGTTGCTC GGGAAGTTATGTTGTAT AACTCCATTAAAGCCGGT GGGCATGGTGGCGATAA	

				TCCCAAAATGGTCTCACAT TTTTAAGCGGCTTGAGTAT CGCCAGTATCCTCGCGG GCACTGGGAACCCAGTCA ACCGCTCTGTGCCCGGC GCAGGCGGGGTTTAAATG TCTCCCGGCTTCACCGGC GCTTCGGCAGCAGCAG AGGAGCACCGGAGGCC CCATGAACCTAAACCAACC TATCTTGAATATGGCTC TCGTTCGGGTGAAGGCA GTGGGAAGAGAGGGCT GCCTCAGATAAACACCTA GTCAATAGCCAGTCGGGA AAAATGTGAATGTTAGG ACAGGGAGGTAAGGGAG GCGGCTTTTCGTAAAGCT CCTTCAACCCCTACCTGTA GTCCACCTATTGAGGTGT TGACAACATGCAAGATGA CCTGCCTCGTTACGGGTG CGTATCATTGCTACAGGTC GTGTCCCGCTTCTAAGAG GATAGTAAGGAGAGGTTA TAGGGAGGTCTGTGTTAGG GCTTCTCAACCCCTCTCT ATGGATTCTTACAGGA GTGGATCGAGAAAGTCCCG GACGTAATACACCCTGGA GGTAAGGGAGTGGCCTTC TAGTAGGGCTGCTTCAACC CCTCTGATGGGAGTGTAC CGGGAACCTCGACTTGT AGCAGAGTTAGTATGGG AGCAGAAAGGGGAGCCG TAAATGGGCTTCTTTTAC CTGCTTACATAATACCTGT	TGAAAACTATAAAACAA TAAAAAGAAAGACCAAA AAAATGTTCTGTTATGAT GCCTTACATGTCTGGG AGACCCATAAGGGTCTC CCTTATACCTTCACTGGG AAACCCATAAGGGTATC CCCCTATATTACTGGGA GACCCATAAGGGTCTCC CCCTATAGATGTAGAGCG TAAAGGGTCTCCAAAGTA CCGGCCGATATGGCCTTA TGGCAAACTCTGGTGGTA GGGACAAGGAGGTAAAG GCAGTGCCAAACCCCTACT TGATCGGGACCATCCAG GGAATGCCATCTCCTCCGC GAAGGTGATGTGGTGAG GTAAAGGGGGAGCCCGT CTTGAGTTTCCCAACC CCTACCCACAGGTGAGA GGAGAGAAGAGGAATC TGTCCCAACGGGAGGA GGGTGAGGTGTAAGGGG GAGACCTTCTAGTAGGGT CTTCTCAGTCGCCTGACG TCCTGACTGTGGGTGG ATCAGTACCCTACAGGTG AGACCGGTGAGGTAAGG GTGTGGCCCTCTTGAGG GCTGCGCCAAACCCCTACT CGAGGTAACTCTGAGGGA GTGGTGAATGGCGGCA TGTAGTCTGGGACTTG ATTGCGAGGGTTTAAATGA GAGTGGCCCTGCTGAGAG CAACACTTGTGGTGCTTA AAGCGGGGCGGCCCATG
--	--	--	--	--	---

				GGTGCATGTATCTAGGTCT TGGCGGAGAGTACTGAG AGACAAGTTGAGACCCC AAGATTGGGTCTCCCTAGC CTCTATAGCTGCGACTCTT AGCGGGATATGGAGTAA CATGTACCAAGGGAGTGA ATAATAAAGGAATTGACG GGGTACAAGTGACTGTTG GCCGAATCCTAGCCACTT GATGACCAAGGATATATT ACAGACTGAGCGGAATC TAGAGCGTGAGAAATAAA GTGAGAAATTGGTTGACCG AATACAGAGGAAG (SEQ ID NO: 1218)	ACCAACGTGAGATAGGA CACTGCACAGTCAGCCA TGAGGTTCTGGAGGAT GATGCGATGAGGTGGGG GCCTCATCAGCCCCTCT GGCAGGGCGTCGGCCAG GGAACTAAATGTCTCTA GCATGTCAGTGCAGTGA GGTAAGGGGAGAGCACT CTAGTAGGGCTCTTCCAA CCCTACCTGTAGGTAC CTGGTCCAGGTGTCGATG ATGTGAAAAACAAGAGCT ACTTTGGTACCGGTCTGT TGCAAAAAGGGTTCTGCA GAGGACGACGGCTATCC CTATCGGGAGGGAATAG TCGGTCCCAAGGTAGTGG AAATGGGGCTTCCACT GAGCATGAAAATGTGGT AGAGGTTGCGTCCAACCC AATGATTTGCAGCAGAGC TCTTGGACACGAAGTCTG TATAGTCCCATGCAGGCA GCCAACCCAGAGAAATGGT GGCAAGACCCAGCTCC GTATGGGAGGGGAGGG CCAAGATATACGGAACG GCTGCTAAAGCGTTCTGC CGGTGTCAGTCTAATCAC AGACAGCTGTGACGAAA CAAAGTATGGGTCCGAC ATGCTTGGTCAGCTCTTA GCCGCAAGGCTTAAATC GAACGAGCCCGCCGAG AGTGAACTTAAACGGG GATGGAATGTGCTAGC GGTTACGTACTACCAGG

					GCTCAGGTTGCCTGAGC CGAGGCTCTACACGTCAT GGTGGAGTTCTCCAC GCTCGTGAGGCATGTA GTGGATGGCATGTGGC GGACCATCAGCTGGCACT ACGAGCCTCGGGCTTGC CCGAGTGGGACCTCA CACATTGAGGTGTGCTT GTCCCCCTACGTTGAA GACTTGAGCGGAGAAT ACTCATAGCCCCACGGC AAGGGACACAAACCGG AGGCTTGTGCCGACGA GCCGTGGACTCCTATAGA CAGCCGGGATATCACTG GGCAGCTCATACTGAA GAAATCGATGAACCGG GCCTACCGAGCAAAATG CACTCTAATCGCCTTGT GGCGACTGTGGCCCCC TCATGCGAGTGAGGAAT ATCATAACTGCAATGGT TCAAAAAGTGATTCCTAT GGCTCGTCGGGGAGGGC TGACTGGGGCAAGCAAA TGATTGAAAGGGGAAGA ACCTTTTCACTGTTTCT TGCCAAGCCCGGTTGATG GTGCGCTAGTAATTGC GACGGGAAATGCGGTT TAAGTCTCCGAAGTAGTG CGTAGCACCGGATGTCG ACCGGGGTAAAAGCCC TTCGTAAAGTCCCTGGGG AGGTCAGTCCTGGGGCT ACTGATGCGCAGTATGTA ATTGCGAGAAATAGGGCC

R2	R2-1_TCa _s	.	Tribolium castaneum	<p>ATCGATACCGCCTGCGTG ACTCGACTGGGTTTCCAC TTGAGGATATCCGACCGT AGCGTGCACCTCTTGTA GTTGGCCCGAAACGGC TGTGTTCCCTCACGTATG TGAGGAACTCAACAAT GTGAGTGGGTAACGGC GGGACGAACTATGGCTC TCGT (SEQ ID NO: 1341)</p>	<p>MSRRPGKSNEPPVRSRAMGLTTLSTGKTSNSGAQGPSTAPMQNM AGGFVCDGCRSYALKTSLARHKKECGKNNAECRWCGTRFNFLAGTR QHERKAHFVQYQSDLAKALPQPESELMKIAIVEARSXNGIFYKEMM ASTGLTHQQVRSRREKPEYKGFLEARRSLAQTNIRAGSISASTXAG SLESASPKAGCSSASPGPTTTSRAPTKGVXRSSNSARIVEAQVHT RAPPNITGETEVALRESRRTPRLGXNPSRPCGISPLMAIAIDEDSVLG GLRVQAGPSPTAVHSVEAFPGTSSMTMETDRVHNKSGIDPILEHN GTRQVREESSTREDPVEQWSPNYKTPVTMPNITTTADAXTSYNR TPQTLPGNRRRRSRLPPVQRKSASDXLESVDLGPWAVFLQDQVD AGSLGNDLADLVRVALTKSDRGVNDVAVNRYLAQRAESLRIRRG SKGRKSKTGRHYGQTTSGSGQRAALFKKHQDLFLKNRRGLAETILSG KEDFGPRPEPPTVSVEEFGIFESPDPDNEFEVRATGVEDPPPLTS PWTKRTLRSPCYWRRRPPPTYTMDIKAARAGWQISAPGSDQIPVA AVKTMSELELAILFNILFRNVQPSAWGVRLTTLVPKDGDLRNPANW RPITISSALQRLHRVLAARLSKLVLSSSQSGFTEIDGTLANALILHEYL QYRRQTGRTYXVSLDVRAKAFDTVSHCSVSRAALGRFGIPSVIREYILAT FGAQTTIKCGSVTTRPIRMLRGVRQGDPLSPVLFNLVMDLELEKVNE KYEGSLQSGERCAIMAFADDLILADRDQDVPAMFDDVSTFLERRG MSVNPACRALIAGAVSGRSVVRTGSSYKIHNTPIPNVDALDAFKYLG LEFGHKGVERPTIHNLVWLNLRAPLKPQDKCLFIQYVVPRLLYG MQNPQVTSRVLREADRLIRRHUKTYHLNVHTPDSLIIHASVSDGGLGI MELKAIPRIFLGRVLKLLNKNKDSVLSVLSQSNRVRTLMGKLSMAG EVPESTFWRNQIASGPLSKGLEQAAEDSASRLWISEKPSGWSGRDH VRVQLRTGNLPTKAIPSPVPGQRRCRHGGCACDESISHVLQMCPLTH ADRIIRRHDEVVKVARHCTSRGWTVVEPHIRSRCGRLEKPKDLAVHQ PGGAIVADVQVSWDSESLTPYERKRAKYDVPQFHQAQHAHPG KALTFAPVIVGARGIWPRINNDRAALQIPPVVRRACVNSVVKWGSS IHATFMIRSVWANRLNPRPLRA (SEQ ID NO: 1464)</p>
R2	R2-1_TCa _s	.	Tribolium castaneum	<p>AGTCATAGAGCCAGAACCC TCCTCGTGGTCCCGCTGGG CACAGGGATTAAATTTTCT GTGGCAAAATTTGACTGGC TTCAGAGAGCGTTTTTCGA AGTGGWCTGTGTGACTGC GTTCCCCCTTAGTTGCTA TWTCGGCTKMGATTAACA TCTCACCTCGACGTWTAA GATCATT (SEQ ID NO: 1219)</p>	<p>GTGAAAGAGTGGCAGTT GTGGAGACTTCTTTCGGA CGTCGGGTCCGATTTCG GACGCCAGGTACCTCC ACCGCTGGGTTCAAAAC TAGGCGAACATCTGCCG ATACCCCTTTAGGTCA AGGACACATGTCTCTGC ACGAGATTAAACCCA (SEQ ID NO: 1342)</p>

R2	R2-1_TGu _t	.	Tinamu s guttatu s	CTGGGACCGTGGTTACA ACCGGGCTTAGCTGCAG AGACAGTACTCCCGTG GTTCCCGCGGACCCCGTA ACATCGGGTGACTGAATC TGCTCTGCCCGGGAGTA GTTCTCTTGCCCTATTG TGTCCTCTGCCCTATTG ACCAGCGGTGCGCGGCTG CTCAATAGTATTCTAGGCG TGAAATATAGCGATAGTCC TAGTGTTGTCTTACTGGG CCATAGCCCTTGCTTCAG GGGTCAATTCGCGAAGTCT CTCAGGAGAACTGGGGGT GGTGTCTTCTGGGTATAG CTAAACCCCTAGACTGTG TCCGATCC (SEQ ID NO: 1220)	TAGGGGCTTGGCATTTC TCATTGCTGCTCTGAA AGGATATGGGTCTCGG TCGCGTGGTAGGCAGAC CCATTGCTCGAGTAGGG GGCTTGGCAGTNTCCATT GCCTGCGCCGAAAGGA CGTGGTCACTGCTGTG TCTGCTACACCTCTCTA GACTTGTAACTCTAGTC TGTCACCAAGATCAAAAT TCTTACACAGACGACCG AGCTTGCTAGTCTTCT GTACCCGAGAAATTTGC TCTTGCTCTCTTTGGCTG TGTCTGGACGTGGACT ATTCCATCTGTCCTCCAA TGCCGCGTCCAAATTATAC CGGATTTGACAAAGCGG ACGCCCGCTTTATAAGC CGGAAAGGTGCTCTGT AAAATTGCAAGGTTCAAT AAATAG (SEQ ID NO: 1343)	MGSWIVFVSATQTGEFPVDTARRAPVPVTSYPESECHXPLPTFC NSDVTWGGVRPEPVDCLGDLPEXYDALPGVAGPREXVGGSPGEG VRSPIASPGTAVQHDGSPILVPGAEEAEVSTPVVQPDHPACP CCGTRVVKTALSEHLRRAHGRKRVLFQCSRCGRMNEKHHSIACHFP KCRGPPVEEGLPGAPEWCECEGQKFNKTSGLSKHRSVHPLTRNV ERIEAARPKGKGRAHKGWTEAEVAQULIELEGRFNQRFINKLIAE HLPKSAQISDKRRQLAAATKTSPEKRVTSSTSGESSPEVEKEGK REYRRRVGEWLACGSLXDQTSFKILEDVESGESEIVTGPLEELASFAR GKLAARVRHHRKHPEAEVPAAREEQRMKRRVGRRLYLRFQRLF ALDRKLAGIILDVESIKCLPMEEVADVFRRRWEVAVPFTGSGSFR SLGADNGAFKPMISAKEMVKNVXEMSRSSAXGPDGLSLRDLMKID PQGRMAELFNLWLAGRVDPQVKAGRTVLIPKADPGKIGNIDNW RPITGSVXLRMFSRLSARLRACPINRRQRGFIAAPGCSENLKLLQAL IKSAKRDRHTLGVVFDLAKAFDSVNHQHFQVLVQKGVGDGHIILR DLYTNAGTYLESQSORSGFILRGVKQGDPLSPILFNLALDPLLCRLD RGLGKYGDQQLXSIAFADDLALLSDSWEGMQQSIRWEEFCQRTGL RVQAPKCHGFLRPTKESYINDCPWTIADMQLMDIDPGSESKYLG LGIDPWIGLSRPELSEVLRWVKNIGGAPLKLQKVDILRSYALPRLFI ADHAGLSATCLHSLDSIRSASVKGWLHLPSTCDIIVSYKDGGLPL RLASLIPNVQARLRVIAQSEDDVRSVVLQEGIEEKRWISAGGRP EKVPSVTGEFPVMEQAADAEALSEWERRAPRTYPIPCWKREME NWTNLKSQGHGIRNFENDRISNDWLLHYGRIPHRKLITAIQLRANVY PTREFLARGLEGAPRGCRHCPAEWESCHIIIGYCPAVQEARIKRHN DICGVLAEEARKLGWVIFIEPHLRDNTNELFKPDLVLVKGSCAKVVDV TIRYESGLTTLSDAAAEKARKYQHLAGEVRALTSATTVDFLGFIGARG KWVVGNNGLLSDLGFTSRVVRIARALS KALLSSVDIIHIFASRARQA QTSE (SEQ ID NO: 1465)
R2	R2-1_TSP	.	Trichin ella spiralis	CTCCTGACTAACCTGATT CGTCGCGCGCGCGGCTT TTCTTTTGCCTCTCCGCTCG TCGAAATTTGCTGTAGTTG ATTGCTTTTCTTTGCGGTT TCTTCTACTTTGCGAGTTT TTCTGCATTGCCACG (SEQ ID NO: 1221)	TGAGGTTTTGTTTTCTTT TTTCTCTTTACCATTCCTG TTCCATTGTTGTTATTGCT TTAATCTGTATTTTACC GCCGGCAATTCCATTGTT ATTATTACTGTTACTGTTA TTATTGTTACTATTGTTT TACTTTTACTTACTACTGT TATTATACTTTAATTCGTT AACTTACGTTATTGTTAC CACTACTTACTTTGCTCTC	MSNRLANTAAGGVPEKTSGLTDIPGQPSSSGEKRAISYGPFGCNS CSFTSTWLSLELHFKSVHNIRDFVFLCSCKKSWPSINSVASHYPRCK GSVKAADVPTSLANTCTTCGSSFGFTSLQLLHRKRAHPDVFAAASCK KTKARWNSDEFTLLARLEAGLDPAKCNINQV/LAERLMEYNTIRGVEM IKGQRRKDQYKALVRQLRSNSETQQCVGLAGMSDNSVPANDTSSSV ASEVSITYPEYGAVMSCDLIKEATGMAIVDINELQNLKAFLSGRKL MKFHGARETAQKKMANPRVAKFRQLFRSNRRKLASHIFDKASLE QFGGSDIEADHLEKFLSRPRLESYSYVSGDKSIGVAHPILAEVELEL KASRPTAVGPDGIALEDIKLNTYDASLFWNLWLAGDLPASVKASRTI FLPKSDGTGTTDISNCRPITIASAMVRLFSRIITRRLAARLEINVRQKAFRP EMINGVFENSAILYALIKDAKVRSREICVTLDLAKAFDTVPHSRILRAL

				TCGAAACGTCGTTGTT GTTTCTTTGGACCGGT TTAGAGAAATCGACGCA CAGCGAACTGGACCGC TTAAGCCAGAAATAGTAA AGTAACAA (SEQ ID NO: 1344)	RKNNVDPESVDLSKMLTGTYYAEIKGLQGLPIRNGVRQGDPLSPLL FSLFDEIGRLQACGPAYDFHGEKICILAFADDLTLVADSAAGMKILLK AACDFLEESGMSLNAEKRTLCTIPRSRKTFFVNPAAKFISDWKTGI SSEIPSLCATDTRFLGHTFDGEGKIHIDTEEIRSMKSVKSAPLKPEQK VALIRSHLLPRLQLFSTAEADSRKAWLIDSIRGCVKEILHSVKAGMCT DIFYPSRDGGMGFTSLGFEFSLSRQKALAKMAGSDPLSKRVAEFFIE RWNIARDPKVIEAARRVYQKKRYQRFQTYQSGGWNEFSGNTIGNA WLTNGRARGRNFIAMVAFRSNTAA TRAEINLRGRPGTKECRFCKSAT ETLAHICQRCAPANHGLVIQRHDAVVTLGEVARKEGYQVMIEPKVST PVGALKPDLLLIKADTA FIVDVGIAWEGGRPLKLVNKMCKDKYKTAIP AILETFHVGHAEITYGVILSRGCWLKSNDAKALASIGLNI TRKMKEHLS WLTLEIIFITQISRIYNSFMKK (SEQ ID NO: 1466)	
R2	R2- 1_Tur	scaffol d_6	Tetryny chus urticae	CTCTCTATTTTAAACATATT CGATGTACTCGTACATTGA ATATGCTTTTATTTTTTTC AAAGTTTTTTGGGTGCATA CCCTGGAAAAATCTGAGA TGATATAATCTCCCATCAG CTTTGGCTGAAACGTTGGC TAAGTTTTGTAGGTTGTTT GCCCTCTACTACTAGTCG CAAATGGTATTGCTAACA GTTGTTAAATTGTTACATT TACAAGTCTCTATCCAGTGC CTCCTCGTGGCGCTACCCG GTAAACACTTAGAGTAATCT GAGTGGCTAAACTGGAAAG GGCGGAAATGCAACACAG GCGGTTGGTAGATGCTTTC GGCATTTTGCCAAAAATCC ACGGCTTTTAGCCCAACA ACATCAGGGTGATGGACC CGCCAGCTTGTGGTCAGG ATCCCATCCATGAATAAAG CATGGCTCTGCTCTGGTG CATCCTCAACGGGATCGG CTTCGGCTGGATGTAAGTC	TGATGTATCCCTTCAATA TATTGTAATCCTCATTCGT CCCTATCCTTTCATTGAT AAAAGAAACTTTGTTGCT CCTTTAAATAGTTGGTCC CTCCTGCTCCCTTTCTGGA ACCTGGTTGTATCGATTA TTGAAAGTTGCAATAAAC GGATTTAA (SEQ ID NO: 1345)	MCILGLTSHSREGGLSRGSSQLTKVPQNEEDNGTTQLKAGSADSF PRPSGDLNPEEPLSIDICPVCFRQMKSYLGVVRVHMQMHLSEYNASI PDPVWSHTRWSDEEAAQLAFTAEKIEVDKLLPRGKINKFLELLPGR TLESIKSHRKQSHKDLVRKYKFEVDTLAADNDDDTIICQDNGDIFN DPVIGATDSOSETETVADPAEFKTFIELADDP TKPVVAKLRNLKDKP KSEILGSDILVRILRTHLHGLPVEDELQYLEVYFTGKIKQRRSKTQTALS KKQIKQRDYGRILQELYSRKRRCANEILNPTSMGGFGHQELSEFWT KTFGPDEQPTLGEVEIIPKENCWWDIFSPISSEDEIKASYPSIGKAAGPD NFSAYQLRKVPVWHLECLYNIFATYKDIPSRDKDAKTILIPKKDNEASP GDFRPITLSSIITRHFHKILATRVNNFVRHPMQRGFIQSDGCLENTALI QTVIREAKVRRKQVHITFCDVRAKAFDSVRYDSIIAAIAKKGAPGSFIMY LSNLVRGNKTTLLTAGGETRITPTRGVRQGDPLSPILFNCVMDQILTAL PSRTGFTLSAGDESVNVNCLAFADDIILSKTKNGHQELLDVDTQRIKE NGLDLNPDKCCSLSPHSKTKIKVVRADVFVNGVKVRSMSIGDSTC YLGVSINVTGQVAPVKMYQALCEKLD SAAIKPHQRLYLKHFVITKMF HPLILSTIAAHKIKNLDLISRYVRKWLHLPHDCSGSMIHAKVSDGGL GVPLLFRTIADLVRRKEKIQVHENPIFRILAKLSTVSELENCCKIASKT TDIQUESTKEMLATYDGLSLKEARAVPEVHKVWVDSYDKRYKFAGRDF VQVIOARFNALPTRSRVWRGRGADEKSLRCRAGCNARETLNHVSQS CFRTHRVRTARHDKILDFICERLDVVGKVVREKPI SFPGKKLPDLIVE NTDQALVLDLQVGNSELP LDERGKNKVIKYN CSEMQLYKRKKKTL AVKALT LHYKGLMAPETSNI LRSFGFKSKDLEK MAYMALFGTVAAW GIFNRSTETMRSVANWPRPEEL (SEQ ID NO: 1467)

R2	R2-2_DWi	.	Drosophila willistoni	TTGCGGAGGC (SEQ ID NO: 1222)	GAAGCTGGGAAGCTGGGT CGGATGAGCGCAGAAGG GGTGTCTTCGGAGCACT GTAATTCATAAGTCGTAAG TCTGATCAAGTCGACTCGG AACCTCTCGTGGTGTTTC CTGGGTGCTGTGAGTTCC TAGTCTCTAGGTTCTCTCC AGTAGCTAA (SEQ ID NO: 1223)	TAGATGTAACCTCTA GTTTCTCTACTTTTGCC TGCTACCTTGGCATTACA TCTAAAAGGTACAAACA TCGATTGGCAAAAAGA GGTGGTTTAGTACATAG CGCTGTGGGACTTCATT GTCCGATGATGCAGCG AATCGTGCATACGAGATT GTCCAGTAGTTGGTTGCT CGTATCTTTAGAAGATTT CCTTCTCGCGGATCAAA ANAAAAAAGAAAAA AAA (SEQ ID NO: 1346)	FERSNSWGYQNLEPSNVQDMNTVPRINNTTTTTPATSRPGDQPR EIAVWNLAIEIPCAVCGRLNTRRGVGHMSHQHDELDTQRQRE DVKLRWSEEEAWMMARKEVELEASGNLRFNKKLAIEVTHRSEAI KCFKRGEYKAKIEQIRGQSTPTPEALDSITSQPRPSLERNHQVSSE AQINPSEEQSNWEIMRIQLQGYRVECPSPRWRAQLTIVDRAQAV GKETTQLCSNLYLEVFLPNEPHITGRSNLRRPTRRQLRQGEYAQV QRRWDKNTGRCKSLDGTDESVMNPQMEIMEPWKQVMTNPSTCS CENTFERMEHSLETVWSAITPRDLRENKLSAPGPDGITPTARSV PLGIMLRIMNLILWCGKIPFSTRARTFIPKVTANRPQDFRPTVPSV LVRQLNAVILASRLASKVNWDPQRGFLPTDGCADNATLVDLIREH HKRWKSYLATVDVSKAFDLVSHQAIKTIQAYGAPT/NFVSFIEEQYK GGTSLNGAGWSSEVFIPARGVKQGDPLSLFNLIDRLLSYPREIG AKVGNMTISAAAFADDLVIFAETPMGQTLDDTTGLFASVGLSINA DKCFTVSIKQAKQKQCTVVERRSFCVGERECPSLKRTEEWKYLGRFT ADGRAQYSPADDLGPKLLRLTRAPLKPQOKLFAHRTVLIPQLYHQLTL GSMVIGVLGKCDRLVRQFVRRWLDLPLDVPVAYFHAPHTCGGLGIPS IRWIAPMLRLKRLSNIKWPHLEQSEVASSFIDDELQARDRLKAENVQ RCSRPEIDSYFANRLYMSVDGCGLREAGHYGPQHGWSQPTRLITG KEYLHGVKLRLNALPSKSTRTRGRHELERRCRAGCDAPETTNHILQKC YRTHGRRVARHNSVNAVKRLERKGCVAHVPSLQCDGSLNKPDL VGIRQNHIVYIDVQVTDGHSLDQAHQRKVERYDRADIRSQMRRFF GVTGEIEFHSVTLNWRGIWSGQSVKRLIAKDLIAEDTKLISRAVNG GVTSEKYMFCAGYTRS (SEQ ID NO: 1468)
R2	R2-2_PM	.	Petromyzon marinus	CGGTGCGTTCCTTGGGTA AGGAACACGAGTCTTAGT GGCCTTGACCTCCACGTG GTCCCGTGGTAACATCAT CTCTTGATGATGGCTAACA AGGCTAATGCACCCATTCC ATCTCTATCTCGCATGGA GGCCGCTATGCGTGATTA CTAG (SEQ ID NO: 1224)	TGAGAAATAGAGGTGC TAACCTCCTGGGCTGA CCAAACCCAGAACACATC ACTGCCAAGATGATTTC CCGCAGCACGTTGCTTTT CTCTCGATACCCGAGA ATGTTCTGCGGAACACTAG CAGTCTATGAACAGTCAC AGACAACCTCTGATCCAA G (SEQ ID NO: 1347)	RPQTKLMTDKLKFSSQLARGLAKRAMD GARVGDPPITVRPTETDL CNTEGSWGRPMKLLFVSSTQTQNE DALWASDVAKPMASRSALK MTSIPSMTHFNSSLEKEEMINYDFEQIKSLVESDSDSDFTDEEDV EESFLDISAEEPVLGKFPIDTKGTTTVLPSLEYICVICQHMKGASELV AHFNKIHRIPLVFKACDKTNSNHRSIACHAPCGGKILTEESLPM VCECCOARFATLSGLSQHKRHAHPVTRNEERIKDGKTSQRGVHRS CWSLKEVEQLALLELQFGQKKNINKIAEALGTKNQVSDKRRDLK KTPAMSDSLHFSSRPLETSPPPNVTTGTSSILAQAERLTNENSGTL EKPAEMAIKAWLNGEGQHDLAVETATMLCPMRLVKNKGRSKP ENDIIPRILPTRSWMKKRAEKGSMKHQKLFKNRSLASLVLDTGT ERHECRIPNADVRFYCEKWEKVLFPNGLGQFKSSGVANNEYEPLIS VEEVQTAIRAIKPTSAAGPDGLTRAAICAAADPEGRTLTALFNAMMIT GIIPKELKKNRTILIPKVMDDEKLELGNWRPITITGSMILRFSRIMTAR	

R2	R2- 2_SMe d	.	Schmid tea mediter ranea	AGTCATAGGGTGAAGTGC AATTCTGACACGATGACCG AGCTGTGTGAGTTTGCAGC TAGTCGTAAAGACTCGAT CAGTCGCCCAAGTGAAGT GGCCGGGTATTCGAGCA CTAGAGCCACTGGTATCAA GAGCAGAGATACGCGAGT GGAAGTTGAGTACGACTA CCTTACGCGGGTCTCTCTG ATAACCCACAGTGGAGTGT GGGAACATAAATGTGTGCT CAGCGTTCCTACTTTCTC GTAGGGTAAAGGGTATGA TAACCCAGAGAAATATCCCA TGGGAGATATCCATGGAA AAAGCACCACGTTAGACA ATCCGATGGTCTAACTCGG CTCCGAGGGGCTAACTAT CCCAAAGGGCTTAA (SEQ ID NO: 1225)	TAGAAGGGGAAACAAAGG AAAAACGAAATGACTGG AAACTATGAGGATATA GCTGAAAGCCGCAAGGA AGGCTAAGTCTGAAACC GATCTACATCTTCGATCC CAAGAGGAACGTGTGGGT TAAGCTTGAGCCGACGG AAAAAGCGAATGCTGTT AGCAGCAGGATACAGT CACCTCCTCGTGGTATT GGCGGGCAATGCTCACT AAATTAAGTGTGAGTAGC TGAGAACTGTATGTGTAT CATGAAAAA (SEQ ID NO: 1348)	LARACPLNPRQGFIAASGCSENKVLQDLMRHAKKLHRLPVLAVMFD IAKAFDSVSHAILWVLRHKKVDEHVGIQNAAYDRCTTSFKSNGEST REISIRGVKQGDPMSPLLFNLAMDPLECTLESHGVGYSDTDHVTAL AFADDLVLVSESWVGMAANLAILESCGLGLEVQARKCQGFMSIPT KDSYTVNNCDPWTKNKDVHMIQPDSTKYLGKICPWTGIIRSDLIH VQLKTRISKIDEAPLKPTQKVELLNAVALPRLYPADHSDCKQSLRLV DQEIKA/VKWLHL/PASTCDGLLYARARDGGLAILKLENAIPSVQVRR LQRIANSSDAIARNIASSQGVVEEYRSLWVRAGGDEAIPTFFLRGSES KEPVYPRCDWRKRRESRRCEKPVQGRGIVNFAQDRISNAWLGP GFKQCFIAALQLRANIYPTRESINRGRDGASRCRCSARLESLSHILG QCPAVQKFKDCATQTEAEMVPHGPAGNALPGWSLRTFLVNVPO KNSIARRNKISDILADEAARLWVWVYKEPRTSEAGELRKPALVFAK GEEALVIDTVRFELSRKTSSEASHQVAYYTPPCDQVVKVLTKASNVT FFGQVVGARGKVAP (SEQ. ID NO: 1469) RKELVTIKNLFEESGATAPAPVPLEVAVEVHQSSVPEITDETTTQEG SYSEPIHRCENCGREFRTRAGVQQHRRKAHTNEFMEKEKAAPT LRWTDDEKEILIESEIKIEGSLKEHEINKILASRMMPGRSQDG/AKIRQ KQEHKAEIQRRLHGTVTNETRGNRTSEITEPIRSLPINTKTWSEDEM KRMLAEVVKLRTNEKDINKLAEIFPNRTMGSIKSRKTKDKDYQDLV KLTMTQISENPDNETDFNTSNTENNSTDAEKEVKNYLMILLTINEEE WILTSTLKEAATLALQGGKTEASEKINEYASKTLFPLGLKTNQTRKREK ISKRETRQYAEIQLYKKNISSAAEKAINGKWSIKPEEYHNNKDLIK AWKPILEAPPFSDCRPIENIKEMDYALMEISTAEIFLAIRAMGKTAPGP DGIKYSKLKKNIOQMAILFNTCLLTSLPLKARTILIPKQENPGILDYR PLTIASVWTRVHSLAKKLDNNAQLSQKQKFRKCDGVAENIVILETIL TNSRSEKRPLCMFVLDLRAKAFDSVGHESIIRGAKRVGVPPMILEYISS YQNASTNLFGELNSRRGVRRQGDPLSPILFNFVIDEALENLNRNIGYLL KEEKVSCLAFAADDIVLIAETKGLLENHIEKLEKNGAGLELNASKCATL MVMKNGEKESTYISTAIKIKENDIPTMKATETTYKYLGLQM/GKARE QNAEVIETEGLENITRAPLKPQQRIHLRDLFPLRLIHLKVLGRVAKSL KRIDQIRKKVRNWLHLPKDTTAAFIHADAGDGLGVPALHTIPLIK RERITNLKSNDPVTKECLRMETKQVLGKWSRPTKIGETLATNKSQ KEAFRQMLITLDGKGLKDHHTETHTHKWIRRGENTMTGQK/ITAVKI RGNLVATKSNRSGRPEQEKLECAQCGRPDSLGHILQGCWRTHGM RVERHNNICRIKAIMKGESEWEEPRLQTNELGRKPDLLICHKGKII CDAQVADSSNCSLESENQRKIDYKKDSVSEARKLIGRVDEIIIMA VTFNWRGAISKTSIRDLDMLLDIKSKEVIKMSRKIIRDNSIMVEMHNR RTEKRR (SEQ. ID NO: 1470)
----	-------------------	---	-----------------------------------	---	---	---

R2	R2- 2_TCa s	.	Tribolium castaneum	<p>TGGAAGACCCGCCCATG AGGCTTGAGAGTGTGAT CCTGATCACACTGAAAAG TTATGCTGAGTACGTCGTG AGAGTCGGTAACTGTCCC AGGATGGTCTGGGATAGG CTAAACCTCAGCAGGGGA AAGTTGTAGGGGCTGCC ACCCCTACACTTTATAGA TATGGCATTGCATACCTCA AATAGAGCCTCGGACTTG GAGGAGCATGTTCCCT CCTCCTGCTAGACCTG GAACCAACGCTCTGACA ACCCATTGGACCTACGG GAGCGGACCATGCTCATG GACATGGATTCCGAAGAC GAAGCGGGGAACACGG ACCCCGCCCGATAATGCT CACTTAACGTCAGGCGAA CCCATCGAAATCATCTTGA TGTTACCCCTTCAAAGCAG GTCATGCGGCATATGCTC AATGCCGGAAGGGTGGC TCCCGGGCGTAACGGACA TCTATCGCTGATGGAASCT AGCTCGCTCTTGAAGAAC GGCGACGGGGACTTGG AATCGTGGTGTGTTCTG ATGTAAGTCTGAAATTAT GGCGTGATGCCCGCCCG CCCGACCGGAGGGACTTA GAACCCCTTCCGCGAGG GTCCTGTCTAGGTCCW CCATCTCCGTAAACGAGT TGGAGGAAACCGCAGACG GGG (SEQ ID NO: 1226)</p>	<p>TGATGCTCTTTGGTTTAA CCATCTGTGGGGCATC GGTCTACGGTTTCTCTC GGGTTTCTATTGTTTTC CTAAACCCGACAAGGAG CCCTTTGGCCCTCTCCTT AAACCTCTCTTCTATCC TSTTAGTCCATTCCGGCTA AAATGATGAAGACCGAG GAGTGTCACTCTCTTGGC GGGGTTAACCCGTCCTAA GTGTAATGTGACCTCGC CATTCGGGCTCTGATA (SEQ ID NO: 1349)</p>	<p>MKSRFRIGDCAAGSSRRGVRITGKAGREGRAASPHLSRYLAGS VSGNVSPVPGPGLGAGAPAFAGARNADGGGQANPCPYCARSTT ANGRGLHIRRAHPDEANNAIDIERIHARWSHEETAMMARLEAGAIQ RGGVRFMNFQFLVPRMPGRTLEAVKSKRRDATYKALVQRFLOAQIN LPELRDGDAPRQPDQENPEPPSFDGAIRGAVADLVGGVDWQR LGFQDRLCNIARRACDGGVSGQLLGLWLRDVFVKRVSTRGDS DLVDGALVSRRTARTREYARVQELYRKEPKACARILGDRREGANR APNRDPAFIDFWRGVSEASAEVEGWAEVSDHGEARRVWDPIV EEVGRSVRNGAAPGPDGIAVSVMWKNLPPEAAALLFNVLGRCIPA ELTRTRTVFIPKTDAPRTPADYRPIASVVARHFHRVLSARVQRIPDF TKYQRFSLGVDGIADNLSVLDTMLTMSRRCKHLHAAALDVSKAFD TVSHFAIVRACRSIFGSAETVLEEGRRHFVQVRXGVQXDPLSPLF NLVLDRLKRLSTDVGFRLTDATKVTALAFADDVLCATTARGLQTNL DVLEAELRAGLLNPNKQALSIVASGRDHVKLVTKPTFKVGQNTI HQVDASSIWKYLGIQFRGSGMCGGSEGVAAGLKRIKTCAPLKPQR MHLRVFFLPKFYHAWTFGRINAGVLRRLDVVVRTSVRTLRLPHDI PVGYFHPTKSGGLGIPQLSRFIPFLRKFRDLGRSAVDYVRECAFTD IADRKIRWCRERLSGIVDQVAGGRDALDAYWTAQLHQSVDGRALRE SASVASSTQWLRCSTRAIPASDWLHYTAVHIGALPSRVTSRGRGG QDVSCRGCLLDETPAHCIQVCHRTGGRLHDAIAKRISADLMEL GWIVTREVSFRITAGVFRPDMVAVKEGTVILDVQVSPAPTLDEAH RRKVAKYRDRADLARYLAEEAARGRAPPANIRFASATISWRGVWS AESVGLRELGLSARHFDRTTMCALCGSWRNWVRFNASTASRMGR GRGDASPRRHNQQ (SEQ ID NO: 1471)</p>
----	-------------------	---	------------------------	--	--	--

R2	R2-7_MR	.	Megachile rotundata	TCTAGTTAGCAAGCGGCC CCCTCTAA (SEQ ID NO: 1227)	TAAATGTCGAACCGAATT TTGGGTAACGTGCACCCC ACCATCCTTAATCGGCAG CACGCAATAAAGCCGTG GGCAGTGGTTTTAGTGG GTAGTCATTAGGAGTCCC ACAGTACCCAGCGAACAT CTTAGTGGGTCGCGTAA ACGCATTTCCACTGCCTA TCCTCGGGAAAAAAA AAAAA AAAAA AAAAA AAA (SEQ ID NO: 1350)	SGPATSTFGTETSKRICEPTSGALGCRPGAVVQWQIHKHKKRIVGW PLHGLGPTSLKLRHPRMLQAKRIVPVLAEMLMQCLCARHVSGRSPQKSA WAFLEERTSACGVDAMFCVCTCQSFATKIGLVHVRRAHVEVAN AAISVERVKDRWSEEEERRIMAAVEVRGVLGSAFNEYIMSHLQTSRT LESVKGTRKNPKYKELVATLEEARTSVREESPRSAVNDATQSPGPS DTRSLRTEHLFTSEFPEHREILIGDLEGVDFRAELLSIAEQQLQG DEVAESLTRLWGEVFKPENQQQVQRKRRQRKAPVSGQLPKWRE RRDYAAMQTLFHRNPSLAAGRVLDGKNESRPPDLPEMTAFWEPI TEQSAEHRVGPASEKSELCSVWGPVEKELLSSVPLDTAVGPDGV TARQWRAVLPAVRALLYNIIKRGSPASMLESTRVFLPKQHSVNPA DFRPISASVVVRQLHKILAMRLRRTNLVDERQCMDGCAENITVL ASLLDDARHGLKELHLVSLDCAKAFDSVSHHAIDATLKECGLPAGFVQ YISRTYSDSSTRLEVRNRSEPIKTRNGVRQGDPLSTLIFCLCFDRVART LSPHIGYDLNNTISTLLYADDAFLVSTTAPGMNILLRSVEESAGEVGL SFNTSKSALSIPSGKEKMKVGTTPFTKTSQGFITQIPTSQEWRYLG VDFQYSGPKKASRSKIELERISKAPLKQQRLIILRVYLLPRYYHHLVLS RTTLGHLRGLDLQVRAAVRWLSLPRDIPIAYFHTTAKEGGLGPAFE TSIPCLMLARLSMETSTCKAARAQVGFVWQKRIHWATAALTNG EALTCKADVDRWWASRLHKSDGRELRECSGVSSSTWVNSALNIT GRDVYQYHHVRINSLPRTIRTSRGRVREGMEVTCRAGCQVTETAH VIQSHRTHGGRIILRHNAVKV/LASGLRDKGWEVREEPKLRTROGLR KPDIVAIDGVARVIDAQVVS GSGPLDEAHETKRYKYSNDGVDTAAL ARECNIAPSNVAYSSCTISWRGVWSPRSAADLLQVGLSKLLGFITLR VLRGSHLNWTRWNKMTTMRVHHQRTGIG (SEQ ID NO: 1472)
R2	R2Amel	.	Apis mellifera	TGTAATCAAAATGCCTCGC TATTTTAGTAGCGGTAGCG CTCCGCCGCGCAGGAAC CATTGACGCCGCGTAGT GTGGGTGATTTTATATCCA ACCAATCACGTCAACTACG ATCATTTGTAATCACCGAC GGTACTTGGTAGGGGTAC CACATGGGCATTTCTGCTC ATCCACAACGCCGCTCC ATCATGGCAACAATTTAA ATATATATAATTTCTTAAG GTTTGACCGTATTCATATA TATATATATATTTAATAT	TGATCGTTAAAAAGTAAAA ATCTATTTATTTATTTA TTCCTATATTATAACACAT TATTTATTTATTTACTAT TGTTTTAAAGATGACGAA GCCGCAAGGCCAATCCA AATTTAACAAGAACGCA GACTACTGGTCGACATTA AAAAAGACGAAGCAGCTG CCAGCTGATAAACACAG AGCCGTCTCGGCTTTA CACCGAGCGGTGCAAGT CCTGACGTACTATTGTAC GTCTAGGGCGCGGGCA	MSSNEEGASDTGAPGPGVPVADVASADGRATYDDHGMSTDYEQK TIELPNGQIQQLWCHIEGRNQRFLEQSQYLKHKTQHPKSGEIIWRC AACQKEFEKLHGCRCCHLPCKCKRKEAKGVAKFCDSCSEESFLTQRLS MHLEHRHPAIRNLKRTQGTSGRGNTRPINRASVWSKEETDLILNERY KHLKQPNVALKEYFPDKTLKQJSDKRRLLPVQEPEDVAITDETGPPTS DSSESIYESATEDEGGDMMQQTAPNDSWKPEFIQSRTNHLEEEDSL RKVEEAIERMAMNEGVEQEVGTLEQFVDSLTSQPTTERKGRSRKS QKTTTKRKTTHNNRKKFLYAKHQLYKKSPPRLELALSSESSGREVV NLP EADSVGPLYKSLWGQPEKTHRNQPMCNNIDMSEIWTPIALES LVEKFKKIKSDTAAGADQIKFHRLKKKGALHVFAKLCNLMLHRIYPA QWKTNRTTLIPKPGKSAEEVENWRPITGSLGRIYSAMIDKLRSKIK QHIRQKGFQEDGCKNNIALLSSALT KMKEDSGTIIIDISKAFTVPH GEISQSLMNMKGVSPICEYIQKMYGCKTKYICRDKKTLIPVDILRGVKQ GDPLSPILNLIIDPIIGTDELTEGIKLENENISVLAFADDLVLAKDKET

				<p>ADKQNRLINEVLDLKMVSAEKCCTTFEIKRQNKTVFLGDPQLTLGQ QRIPYADPEAAIKYLGTFNPNWRLGCKTSIKEIIDAARTVKQLKPHQ KINLIRTYLLPRYIHKLVANPPPLGTLDLIDKELKTIKEILHLHPSTTDGLI YTDKSHGGLGIQRVANIVKLAKLHSLMTRSEDNAVKIALNGQEGM VKRYATSIGLQWPCGIEEIEETRKKLRADTNKWKTLISQGGQIKEFFG DKTGNAWLYNPEMLRPSRYLDALKLRTNTYGTKAALHRAKRDIDINC RRCGVQVETLGHILGLCTHTKNKRIKRHDEICDLIAKNVSKEYVIFREPE VEVNGDRRKPDMMVIKDHKVVYVDVTVRYENNDLSINKAYKEKENKY KETAEMRRDLKAKESRVLPPVIGSRGAVPRATIENTIENKVLGLQTKHALT ASLIALRSSIEMANEFLDYDHTT (SEQ ID NO: 1473)</p>
			<p>GATTCTACCGTGTAAGT CTGGGGCGACGCCTCG CGAGGCACTCCCTGGAC AACGTACGCTAAAGCGTA CGGCTAAGTGGCCTCCC GAAAGGTCCTCCGTTCT AATTTTCCGAGCCCGCG GGCAGATCTCGTGGCAG TGACGCTAGAAAGTTAA GTCCGCGGACATATAAAA TTACAGCCTTAAATAATG AACCCACGAAGGAGGT ATCCTCGAAATTCGCCA CGATCCTTCTGATCGTAG GCGCAAAACA (SEQ ID NO: 1351)</p>	<p>TACAACCATAAATCTTATA TCGAGCCTTCTATTGGTC TCAAAAGCAATACGTTGTC AGATCTTGAGAACATCAG GAGTGAGCGGTGCGCTGT GGATCCGTGCTTTGTGC GCGGCGACAAACCAATAC GCTGCTGCTGTCGCAAA GCAATACGCTGCTGATTCT CGGATCGGGGTGTCGACG GTCACGCAAAAGCGATACG CTGGTGGGTTTCAAAAC AATACGGCGCTGGTGCTA AAAAGCATTATGCCGCTAA CGGCTGGATTGTCGATCG CCGCTCGGGGGCTAGTG GCGCACCAGAGAGGTGC GACGGCAAGCATTGGTT CTGTGCGAAGCGGAGTTC TTGAGAGTAATGTTGCT GGGGCAACAAAGGCAAC ATATAGCCTCTTATGCCTC AAGTCGTAGTTCGTACCTC CACGTGGTCCCGCTGGAA TGCCTATCGACTCTCCCC GGAGGATCATAGAGTTCCG AAACCGGCTACGGCGAGG CAAGGGCGGTGAGGTGCA CACCGATGGGAGCAGCG ACCCCACCTACCTTAGCT AAGAGAGCAGGCGATCCG CCAACTGTCAGCACGAAAT AAATAATCATATGTATAC GAGGAGAAATTTACAACG GGTACCTTGTCGCCGAACC GCCTGAGGTATCACCTAC AGGTGTTAAATGAATCT GATAGCTGGCGGATCGTC</p>

R2	R2B_N Vi			<p>GACCTCTTTGATGGCTCT GCGCAACGACTGGAAAG AATAGGAACGGAAGTCTA ATGGAAGGAAAGTGTGG GAGCACTATAAATTCCTAA AGAAAGAAAGAAAGAA AAAAATAAAAAACCAA ATTAA (SEQ. ID NO: 1228)</p>	<p>TGACTGAACAAAAACGTG TTGCTCTTCTTGTCTAAA ACTATTATTGAAATAA GGGAGGCTAACTGCCT GCAAGTTGAACGCGAAA GTTAGACCTTCCACCTA AAGCCAAAAGTGATCG GGGAATGAATCCGCGG TGACCCAGAGTTGGGT AAACCTTGAACGTTGG AGAACGGAAGAGAGTC CCGCCACGAGCATCGA GTGCTGCGGCGCCGAA TGAAACGATCGCGGAT GGTGCAAGTCGTAGGAC GGGCACGACCTAAGCC TCTGTCACGCGCGCGAA GCCAGGAATCACCATGCA AAGGTGTGAACGCGGCG GGATACCTCCACGGGTT TCCCTGGGCATCGCGCGA CGCATGGCCAAAAGTCCG CTTCTCAGCTACAAAAC AAAAATGGTATGAGACTT GTTAACACTAATTTTCC GAGCTAGCAGGCTCCCT TGCAACGCTTATGAATC TGGAAAAAGGACACAAAG TGGAAAAAGCGCTGATG GTGGACAAAAGTCAGTT</p>	<p>TFAPTHMVRSGPCRKTRKPSDYRESLIMDSGNNVASEPRGAVDV TSAAPIGAELNAEPCEGRNQRREAAALSAQTRRRNXARRARNAQQAD EPGDDEEIEHGLTIRTXEPMEIVAIKNPQACPKCLGGGTQLLCM GSWELSRHINKEHPSVDVTWVGACQRRCTTLRSWSCHV/LHCKGR QEPKDLPFKEHCSLSFDSQIGLSQHERVHVHPEVRNDKRAAEANKPK GKSGRRPSIWDEDELLIRESEYHGARNINEKIAEHFPDRTGRQVSD ARRRDYALRGGPGQPAEGVEAIEEVDEGEIPEGEELVATDGAA LESGPPENGSGAPAEQVNAAPALESSSQDRECSPAVGSDEQIEDSSD DDEFSDALGEISLPEPLSVERTTISPPRRDDWKGPWRWEICNASEEAG SYANWVTGLQELVRNNALSEIGLSLYDQLIQIMRHPSDDDNEQDRL QLNARGPPRRGHRKNRRRRRLTAADKRFAFARCDLWNNNP/KL AELVIANDLSILQRRQAPGRTEQTLYNELWGRVGPNIAPRRTEPI PVSRIFTPTPQEIIMGRIRRIKNDASAAGPDGVTKDDLGRGVSIALSKL FNSILLAGYYPKAWRENRTLLPKPEKDPADVKNWRPITISSMVSRYV SGLLDQVRVAVIKQCDRQKQFTEENGCFNSIQLLDDAVSNACKAGG VITILDVSKAFDTPHAVIQGCLEKKGIPETVAAYISSMYRDCSTAIRTR SGDVKIGMKRGVKGQDPLSLFNLVLEPLERLQETSGVEIEGMNLS CAAFADDIVCFANTAPEAGRQLRMVADYLGRLDMSLSVSKJAVEV PHRKTWYTKNPGLEVNGNAVPSISPSETFKYLGAKVSPWKGLLEGFE SDAFREVISRVQRLPLKPMQKVLDLLQMYIFPRYTYGLITSPPAKAVLKT IDRIIRTRIKELHLPESVSSFLYPRKQGGGLLEKVMVLIALLRNGL RARQSHDPVTRAAMNSNAADRLKSYADALRLHWPLTTKELDTYKY QLRLSYAQKWAQKQWQGGQVEEFAQDPVGNISWLQRYDLLPASRYI DAIKLRNTYPTRALMKIDGRVDSRCRCQSGSETLGHILGRCRYTKD KRISRHNEKDLLKARLAKNHQVMDEPQITVGRQRFKPDVVKTNEG RVHVIDTVRYEHRTYLDEGRTEKIGYRQLSTLRDLHNSNAEEVPIV IGSRGAIPRETRKALSGLGIGKSDWLTLSIALRSSLEIVNAFMDD (SEQ ID NO: 1474)</p>
----	-------------	--	--	---	--	---

				<p>TAGAGCAAAAGCGGTAGG CAGACTTTGTCAAAGTAG TTCTGCCGCAAGCAATGC GTGTGGAGATCTTCGCCG GTGAAAGCAATACGTGTG GGCGAACTTAG (SEQ ID NO: 1229)</p>	<p>GAGACTTGATATCAGTTG TTTTGACTAAGAAATTTAT TATCGTTGACTTTTAAAT ATTTTATTATTGACTGTTA ATATACTGACTTGGGACC AAGTCATCTCTGTTACCC GGTACCGGTTCTGTGTCAT CAAAACGGAAAGTCCGTC CCACGTAATGTGGTAGAC GCAGGAG (SEQ ID NO: 1352)</p>				
R2	R2Ci-B	AB097 122	Ciona intestin alis	<p>CGACGGTGAACCACTTG TCGCGGTGTGAAGACTTT AGTGCTCGAACCAAGAAA TAGCTTGTGTGCTGCTCT CTGGCGGTGCACATACT TCTTAACCTCCCGAGGCCA TGCCGGCGGGGCTTTAG CCCCGGCAGGTTTTACCA TGCCGGACGGGTTGAGA GGTAGAGCCAACTAAG AGTTCACCAGCAGACTTCG CACGCGGTGGCCACTGG CCGAAGTTTAAACAACAG GGCGGCATCTCCCAACT CAATATATGGTGTAAAGTG AACCCTGCCG (SEQ ID NO: 1230)</p>	<p>TGACAGTAATATGAAAAAC ATCACATCTGACCGGCAC AGAATCACCATGCCGTAA TGACCCCAACTAAGGATT CCAAATGGGTAAAAAAA AAAAAAAATAAAAAAAA AAAAAAAATAAAAAAAA (SEQ ID NO: 1353)</p>		<p>MGEWPVWSWLSLTVLVEKWRPFTILQPYMPGQLRDVVLPRKTSY LMDKNIYENTTSPGGGPLCGEKTNRSDVIIPPGFAPSTDTASNTLGE NVDASATTSSANPLSQEPGWCECSKLFKSQRLRVHQRSKHPELYH SQNQPLPRSKARWSDEEMVIFAREEIANRKIRFINQHLHKVFPHTLE SIKLRGKNVRYARIMADLEAEMTSQPEAATSLCTETSENLAASNVL QTRGWAENLVENIDTAHLANLGPFSQFEPGKPSSTKEAINTYEYNDW ISKWLPSSAAHRRERANPPSTKLINARATRLQYSRIQNLYKLNRSACA QEVLSGAWKVQSGELNLKEVQFWEKMFRESAKDRRKPKPTGEVL WGLMEPLTIAEVGSLKSTTPSAPGPKLTLTDGVKRIPIAELVSHYNL WLYAGYQPEGLREGITTLPIKIGTRDPAKLRPTVSSFCIRFHRCLAQ RMETSLPLGERQKAFKVDGICHNIWLSRLHNSKDNKLNITFLDV RKAFDSHKSGLGAAARLGLPPLITYISNLYPNCSTKLKVNKGISKPIEV RRGVRQGDPLSPLFNAMVMDWALSELDP RVGVQIGEQRIINHILAFAD DILVSSTKIGMVSSINTLSRHLAKSGLEISAGKEGKSASMAIVVDGKKK MWTVDPLPRFKVNSQIPALSITQQYKYLGINIDAQGARNDAAARILTE GLAELSRAPLKPQQLRVLLRVHLLPKLQHGLVLSSCAKRALTYLDKSVR SAIRRWLTPKDTPTAFYHAKACDGGGLGTRLEHTPIILKRNRMIMKLT LSEDPVIMELVKLTFTNLLHKYSNVKLLNSWPVTDKDSLARAESML HTSVDRGLSNCSDVPRQSDWVTNGASLLSGRDFGAIKVRGNLLPT KVSAAARGRQREITCDCRRPESLGHILQTCPTWTGPRISRHDSLLKRV RNQACLKNWPTPIEPSIPTNIGLRRPDLVLAKGNIAFLVDATVADNA NMQLQHEAKVEKYNNSDIKEWIKVHCPGVDEVRVTSLTANWRGCL YGGASFLTEDLGLPKAELSLSAKINEKGYVLWCAHYRGRTARLWNR LRS (SEQ ID NO: 1475)</p>		<p>WVTSRRPRVYVGPQKKASDNGNDGAAAAEAETPNPGGPDRAADD EGDVKFWECEFPQCDRFFMTRSGRGLHHKKGHPDWNDRNLGKQ HRKEIWSEEEERLLAKKEAELISGARFINVELRDFTRSLDAIKGQRKR</p>
R2	R2C_N Gi	.	Nasoni a giraulti	<p>CGGGTTCCTCCCGACTTCGG CTTGCGGTGCTCGGGC TCACTGCTTTTGTGGAGT</p>	<p>TAGCGGACTGGACTGTCT GGAGGAGTGTTTAACTC GGGTCTCATGGGAACCC</p>				

			CATGGTTACATGGTGACCC TGGTTCCTCGACCCCGC TGAACATATCTGGGAG GCCATGATTGGGTAAACGA TAAAGGTCCTGGTCGTGC CTCCTGAGATAGGCTGAA TGGGTCACTAAGTGGCAC CTAA (SEQ ID NO: 1231)	GACAACTGTTGTTATCTTG TATGACAAATCATAAAAA AAAAAAAAAAAAAAAAAAAA AAAAAA (SEQ ID NO: 1354)	PDYKILVEKFVRELRVGRIRQGVASRSQQARAMAVAGAPAAATSSGAP PVATPPPSGRVLRSQVVEAPAMEIPVASEGSSGDELFEDEVPR LSDLPDRFTIYFAGLEIPGTEDIYAHRLHTICLMTTWRTKEEVRLGL FLKDLFSGKSQERPTNLPDRNRIERRRGEYKCCQDLWRNKST CVQRILKEDLSQGECLPRELMEFPWNATFTQNPGTAPVLPPTTEVYS SVWEPIRPNKGNYPQNTAAGIDGLTVGDLKGVSRMLARIFNL MWCGKLEHLCASTRILLPKPGAKVPGEFRPTVTSVLIRTFHKVLAE RLKVPLDPRQGRFESDGCENVMLLDMTIRYHHERRRKMFLALL DMAKAFDSVSFESMREVLTTKGIPTFIEYFMTLEDSTVLQHGNNW QSGKIHTCGVKQGDPLSPPIFNFIMDEM LKRLPKEIGVNL DGLFVNA MAFADDLSLVANTEQGLQILIDEATSFGLCGLRANPNKCVTLAIIKIP KEKTAIDPSSHFRIGNAVIPSLKRTDEWVYLGIFKFNNGRLUSDAKPKL IKDELLTKAPLPQQRLWALKVIVIPGILYRGTLGSGTAGYLRSLDCVIR AVRRWLRLPGDCPNGYFHAADVADGGLGVHPYRYKAMVDRLARLR KLEKSAYITGPEAARYLQRQVSAENRLRDGANRIMSDASMLREFLRE LLYKSFGRPLENSSKVPQGHRWVEEPTFLSGADYMNICIRARIAAL PTAARCARGRLDKHCRAGCNGVETLNHLVQFCHRTHTGRIGRHD VWYVVGGLKRGYAVKEPKVLQDVVYKPDVMTATKEGKTLLDAQ VLGDQDMRLAHEDKLRYGAPEFKIRSETGSATIKSLSVTLWSRG LWGPDSVKGLLEEGVILKDLKILSTRVLIGALAGWRRFRNERTSMATS GRREVTTRMVRWRKRRERVGVG (SEQ ID NO: 1476)
R2	R2La	JN937 617	Lepidur us arcticus	GGGTAGCAATTGATCGA TTCCCGCTTCTCGTGCGG CTACCCCTGGGTAACTAT GAGGAATTGATCACACCG TAGCGAACGTCATCAGTCA CAGCTGCACGAATCCAGA TAGAAATATAACAGACGA GTAATCTTTTAAAGCTC GTCAGTAATCTTCCCGAG (SEQ ID NO: 1232)	TGATAATCGCTCCATCCT GCAACTAATTATGAATGC AAATCTGTTAAGTGACAT TAGTGATACTTACTGAT ACTTACCCTGGTATTTATT TGACCTACTTACCCTG GTATCTACTGACATATA TTTATCTAACCACTACCT ATGATGACTCCCGCGGA AACTCTCACTTACTTATT ACCCACTTGGTCTTTATT TTCTCGTTCCTTATTACTT TGTTCTTTGGGTAGGG TTCTCTGGTTTTTGGAAC GGCTTCTCTAGCCGGAAT TTTGCTGATGATCTTTCG TTGTGCTCTGAAATATA

R2	R2LcA	.	Lepidur us cuesii	TTTGGGTTAGCAATTGATC GATTCCGCCTCCTCGTGG CGCTACCCGGGATAGCC TCAAAGAAATTTGACGGT AAAGCAAAGAGGAATTGA TCACCCAAGGCAGTACATC GGCCTTCTCGCAGGAGCT CTGATAAGATATTAGTGA GTTATTCTGTTGAAGCTCG CTATTTCAATCCCTG (SEQ ID NO: 1233)	CGACCCAGGCTTGCGTCA TTTAGGCTCTGGGAAA (SEQ ID NO: 1355)	SCRAGCLDVQVQPTIETINHIAQVCPRTHGARIKRHRDLVQLSLNLR KNPKRNVLVEYNFRTVAGIRKPDIVIEDTRAVILDVQVVGDDSSNLEM EYLEKSRKSYNDANFINALQKLYPTVTNLTTHAVTFNNRGLIAKSTVAA LRMLGVPPRCIMILCVISLEKTLLEVWRMFNQSTASARK (SEQ ID NO: 1477)
R2	R2LcB	JN937 619	Lepidur us cuesii	TTTTGGGTAGCAATTGAT CGATTCCGCCTCCTCGTG GCGCTACCTCGGATAAC CTCAAAGAAATTTGACGGT AAAGCTAAGAGGAATTGA TCACACCTGACGAATATC ATCAGTCACAGCTGCACG AATCCAGATAGATATAA CAGGCGAGTAATCTTTTC GAAGCTCGTCAGTAATCTT CCAG (SEQ ID NO: 1234)	TAGTGCTTGAGTGATGCC TATCCTTTCTTTGATTAAC TCTTACCATATACTTACCA GTTCTTACCCGTACTTACC CTGTATACTTACTGTGT GCGTACCTGTGTACTTGT CCTTTAGCCGCTTGTGT TTTTACCATTTGACTTAC CTTGTGTTGCTCCGAT ACTTACCTTGTATTGCT TGTAATTCTGCATGATAT TTATTGTTAGGTTCTCG ATGCTTACCTGATTGTG CCCTCATCATCTTAGTT TCGTTCTATTTCACTCCAT TATGGAGTTCCTGTTGT TTTTGTTGAGGTACAGC ACCTTTAAGCTGGAATT GAGTGAGTTTATGTACTT TGGATGTTGTATAAAC TACCCGGAGGCAT (SEQ ID NO: 1356)	MSEERSPKQTASKRGAAVEKTMMSGTYVCTLCGRSEKSVGLSLHT NRMHPEAYNKLKAEKPV/LKARWSEEEVFLAQKEALSFIGIKF MNIHLKIFPERELEGIQGRKNPTYKAQVVSLLAEIRESKANDSSSSSS SSSSCDASLGSINWLEFLALPKTSNQFQEGRLDRISDALRGVDVLE NLDAYLLEVFAPMAQNPCPKPPPAKNSRERRDRYSRVQNFYKK NRSACINSILDGNTSRQNVIPGLTKFWTETFEKNSPPDDEAPQFVA DEPRDMYKWTTFYEMSQDVLDSSTAPGVDFSAKLRSMSPRVLNK ILNLLSENLPNSFKMHKTVLIPKIDDPKSPGDFRPTITSPVLARLNKIL AARLSKLPISQRQKAFPLVDGCGENIFLLDYLRSSKSSKSVAMAVL DVKKAFFDSVSHSILRALNEAKPINFNFVRNSYDGGCTTKLTCGGTSF PDSVRMNRGVKQGDPLSPVLFNLIJDSAIRKLPDSIGYVIRDGLKINCL AYADDLILVASSRAGLKTLLNINVAEHLISRLGLDINAACHGLSIIASGKA KTTYVSAADSLDGGPIKNGVLDWTYLYGIPFSLHGRAEKVSPDLT NLLNKLQKAPLQKQLYAVRNFVPRALHGLIUSKTNLKEINTLDRAI RVFLRTLILPKDTPGLFFHSPKSGGLGITCFRTSVLKCRLQRIARMRS SCDGVIAVAESDIFADEYAKLRDLIRINGVLDTTESIKRYWAQRLHS SVDGKTLAYMDYFPQNLWMSSEDKVSQSVYFADCVKLRINAIPTR VRSRGRPNKEMCCRAKCFDSQRMPAFESLNHITQVCPRTHGSRQ RHDKIAKFLKLNLCPSRSVLYEPHFVTV/DGLRKPDIIDDDSHMVVL DVQVSDSANLEKEFECKAKKYANDVALRSAMLIKYPFIKSFVAAT YNNRGLIAKSSVQVLRQLGLSPRSIMVSIILCLETLETWRIFNQSTMN AH (SEQ ID NO: 1478)
R2	R2LcB	JN937 619	Lepidur us cuesii	TTTTGGGTAGCAATTGAT CGATTCCGCCTCCTCGTG GCGCTACCTCGGATAAC CTCAAAGAAATTTGACGGT AAAGCTAAGAGGAATTGA TCACACCTGACGAATATC ATCAGTCACAGCTGCACG AATCCAGATAGATATAA CAGGCGAGTAATCTTTTC GAAGCTCGTCAGTAATCTT CCAG (SEQ ID NO: 1234)	TGATAATCGCTCCATCCT GCAACTAATTATGAATGC AAATCTGTTAAGTGACAT TAGTGATACTTACTGAT ACTTACCTGGTATTATT TGACCTATCTTACCCTG GTATCTACCTGACATATA TTTATCTAACCACTACCT ATGATGACTCCCGCGGA AATCTCACTTACCTTATT ACCCACTTGGTCTTTTATT	MSGKSKPRTVSSGSSSQTPPSGNAACDTCGKCFMKPVGSLHMSK VHPTQYHARLEKNQPKAKKFRWTDLDLYFLAKKEALLLGGIKFMN KELAEFFPEKSV/DQIKQRRSETYKQVWSIHSELLQLQAVADSPPSR IPAKEYSAWLDLFLALPKTKNKFSEDKLDQIRTAQEGTVPVLDLIL REVILQPTROGERQAKPLPPKSSREKRDREYARVQNFYRNKTACV NAILDGNKKCENKIPDIDEFWKAIFESQSPDPAEPVSVVDEEPKNIW SWISFFEMNRNVPDTPSPGPDGVTARMLSIPARVLNKLNLLEFIE DLPAVFKCHRTVLIPKVDNPAALPGEFRPITISSIIVRQLNKIIARVSEGV PINPRQKAFRQIDGCAENVFLDLFILDAKTKISLSLATVDIKKAFDSV SHHSIFRAIRGARPENLVNVIQNSVSGCTTQISVGSGISTTKILMNRG VKQGDPLSPVLFNLVINEIRKLPAISIGYPINSELINCIAYADDLILVANT

					<p>REGKLLNLLNEELPKRGLEINASKCGLSLTALGKLGKTHLCTSDQLD LHGTLLKNTLAEESWVYLVGFHIGRSKSFDPDEALLNKLQKSPKL QQKLFALRVYLLPRLHGLVLSRVAIGELKIMDKLLKHLRWLRLPKDT PLGFFYPVKLGGIGIKLNRNV/LKCRKQRIERMLVSPDDVRLVAES EIFLKETDKLDLLTINGMCLDXRNVPRTGKNNKFWSERLYTSFDGKT LAYSEYFTQGGGWIREDKILQPAHVFAECIKLRINALPTKSRVAHGRRP TKDRSCRAGCLDVQKVPKPAIETINHIAQVCPRTHGARIKRHRDLRVQFLS LNLKPNKRNVLVEYNFRFTVAGIRKPDIIIVIEDTRAILDVQVVGGSSN LEMEYLEKSRKYSNDATLSMRINALQKLYPTVTSLTTHAVTFNNRGLI AKSTVAALRMILGVPPRCIMILCVISLEKTLVWVRMFNQSTASARK (SEQ ID NO: 1479)</p>
R2	R2Nve c-A	.	Nemat ostella vectens is	<p>GGTTGGGGCCTTCTCGTG GCGGAGTCGTGAGTAAGG GGTATAGGGGTAAAGGAC ACCACGGACCGAGAACGG TTACCGCTCAAGGCGAGT GGTGAAGGAGATAAAATC GTAAACGGCCGCTCCGACC TGCTCTGAACTAATGCC AACCAACTGACTGTGGGG CTAACCTCCCGAGAGTCAG G (SEQ ID NO: 1235)</p>	<p>TGATGGTGGGTTACTCGC CTCTGTGTAAACAGGCAAA TGAAAGCTGCGCAAGCA GTCGATGAGCCAAAGCC GCACAGCCCGCACTGG GGTACAGGGCAGCCCTG GGCTATGCCCGAAGTCTT TTGACAGGTCCAAACTTC ACCTTGCCGCCAGTAGG GCACAGGGCCAGGTGC AGGTGGGCGCTTGTCT ATTTGGTTTCGCTTAATT TTGTTAAATTTTCCCGTG CGCCACCTTTTAAACCC TTTTAACACCAATTTT TACAACCTCTACTCAAT AATCCAAATGAATAAAG CGGCATAACAGGTGAAC AC (SEQ ID NO: 1358)</p>
					<p>MPVSTGAETDITSSLIPIPASSIVSPNYTLPDSSSTCLICFIPPTHILLSH ATAIHHSICPTTPVQDGSQQMSCVLCAAAFSSNRGLTQHIRHRHISEY</p>
R2	R2Sm- A	.	Schisto soma	<p>ATGTTTTTAATTTATTTTGA ACTACTACTGTCTGAGTGC</p>	<p>TAGCCCCCTTCACTCTTA GACATTCCTCCCACTGTTG</p>

R2	R2Tc	EU854 578	Triops cancri formis	<p>TTCTTACAACCTGAAGGCT CAGAAACTACCCACTTTT GCTGTTATCCACAACAAC AGTTGTGAATCTATTCTCC AAATATTCCTTGTGCTTTT GTCAACATTATCTATACC AACTGTACACCTACTTCT TCATCTCAGGTTTAAATCT GGTCTAATTTCTCATCAT AGTCACGGAGAGGGCCTA TGAAACGGTCCGTGACGCG AAATTCAATCCAGGAATTC GTCCTCTCTGCTAGTGGT CCCCGAAATACGGTTCCTC TGGCCTGTCAGTTGTGTTA AAACTATATAATAACG (SEQ ID NO: 1236)</p>	<p>TTGCTTATCTCATGTTTT TGTTGTAATGACTGCTC TCTTCTGGGTGATGCT GATTGCTCTCTCTCTTC CATATTGCTGCTCTCC GCTTACTTCCAATAGTTG TCATATTATGCTTTGTT ACTTGCCATGCTAACGA CAATTACTTTATCTACCT AGTTGGTCTCTTGGTTT GGTTGCTTTCATGTTTC ATGGCGGAATCTGATGTT TATAATGACTATCTCTACT ACCACCATACAACTATT ATTATTACACTATTATTA ACATTATTATCTCTCTAC AATTAGTATTATGGCTAC TCCTTTCAGCACACCAAT AAAATCTCAATCAAAACAT CTCATTATTAAACTCTCT ATTCCCTTCTGTTATAAA CTTACAATTGAGTTTAAAC CGAATATCTCTCTTTTACA AATCTTAAGTATGTAATT TTGTGCCAAGCCCATTTG GGTCTGTACAATTGATA CTTAAAAATAAATGTTAT (SEQ ID NO: 1359)</p>	<p>TTTGTGGTCTGGCATTGA TCGTTTCCGCTCCTCGTG GCGCCAGACTGGGTAAAGC TGATTTATCAGTGAGCTAA GAGAAACGATCACCAGCAG GAGTCCATCTACCTCGCT GCACGTGATTTCATTGTC ACTTGGCGAGTTATCCCTT GTGGAGCTCGCTCGTTAG</p>	<p>TTACTAGAAATTATCTTTTC</p>	<p>MSQKRRPEKAVPDEGATAHDVAQDPKSKSCVCGETFKGPASVTMH MVKKHPVEFNEKMAKPPVKVVRWSEEEIFQLARTEAELTLQGVRF INVELQKIFPAREIEGKQKRLAKYKELVKDQLDEIGRAPNPPEQEIGE DVSPFKAWLELLALPKTPNDFLEHKLNIIVQALKEDVNSDQVFN LNSYLKILPEPSGRKSPVGEIHDGDPGSAKTSVTKAPKPAATVSSSRKK RRDAEFARIQRLYRKNRSCINTILDGNTREHEAPKNMEGFWRERFER ESPDPPDDIFILEEASDIWKYISFYEMCNLYPPSTAPGPDGFSSK DLRRMTPRVNLKLNLLLRDLPQLKSHRTVLIPKTDLPKPGDFRPI TISNLRHLNKLANRSHLIPINERQKAFLLPDGCAENIFTLDFILHHA RTKIKSLMAILDISKAFDSVSHSIFRALREARCPGIFKFIENCYGGCF</p>
				<p>NELIRQIAVQPTSRISWSPFDDASLLSIANHEAHRFTKNDLYQHISTV LTRTAEAVKRRLLHLQWSRSPTAITSSNNHHTTDPINTEARYIFPVD LDEHPLSDATTPDASTHPLPELLVLPLSPTRLQNISESQTSNHSNR NSMHTPTPYACDSDESJLVTPSPSTPSCFHSYRDP LAEQRSKLLRASAS LLQSSCTRIRSSLLAFLLQNASTLMDEEHVSTFLNSHGEEVFRTWTPS RPKHPSHAPANVSRKKRRKIEYAHIQTLFHRPKDAANTVLDGRWR NPYVANHSMIPDFDCFWTTVFTKTNPSDREITPIIPMTPLIDPILPS DVTWALKEMHGTAGGIDRLTSYDLMRFGKNGLAGYLNMLLALAYLP TNLSTARVTFVPKSSPVSPEDFRISVAPVATRCCLKILAKRWMPLFP QERLQFAFLNRDGCFAEAVNLLHSVIRHVHTRHAGASFALLDISRAFDT VSHDSIIRAAKRYGAPELLCRYLNNYRRSTSCVNRTELHPTCTGVKQOG DPLSPFLIMVLDLEGLDPMTHLTVDGESLNYIAYADDLVVFAPNA ELLQRKLDRIISLLHEAGWSINPEKSRITLDISGGHKSITALSQTEFTIAG MRIPPLSAADTFDYLIGKSNFKGRCPCVAHIDLNNYLTEISCAPLKPQQ RMKILKDNLLPRLLYPLTLGIVHLKTLKSMDRNIHTAIRKWRLPSDTP LAYFHSVPAAGGLGLHLSVSPFHRKRLETLSSPNRLLHKLPTSPTL ASYSHLSQLPVRIGHETVTSREASNSWVRRLHSSCDGKGLLAPLSTE SHAWLYRYPQSIFPSVINYAVKRGGLLSTKVRSRGGRVTNGLNCRG GCAHHETIHHILQHICALTHDIRCKRHNELCNLVAKKLRRQKIHFLQEP CIPLEKTYCKPFIIRDSIAVVLVTVSDDGNTHASRLLKISKYGNERTV ASIKRFLTSSGYITSVRQTPVLTFRGILERASSQSRLRCLFSSRDLDGLCL SAIQGSIKIYNTVMRGTRQLNE (SEQ ID NO: 1481)</p>				

				CTTTTGAG (SEQ ID NO: 1237)		<p>GTTCTCGTCTAATTGCTT TTTCGGGTAGGAACGG CTACCTAAAGCTGGAAGT GGGAAGTGTTTCAATGT ACTTTGTGATTATAGAAA TATATGACCCGAGGTGCA TTGTTGGCATTTCTCG AGAAA (SEQ ID NO: 1360)</p>	<p>TKLFCGGVKYPSEVMNRGVKQGDPLSPVLNVLIDGLIRQPSALGF NVSDQVKVSCIAYADDLIIATTRAGLKTLLDNTSYLAKRGLSLNPK CSALSVASGKQKLVYASSEHFDLAGQKMRNLNVGDSWRYLGIQFS HLGRAEKVTPDLTCLNRLQKAPLQKQKLYALRYLIPRLIHGLTSLKTN LGELKTLDKIRKYIRAWLHLPDDTPMGYFYTPLKAGGLGLPSLRVL NNRLRILMKASQDIIVRTIAESETLGVIRKLDLSDIDGTILDSVKI HSFWAERLYSSYDGKCLNSANFPNGNKWIGEDSLNQSHIFADCLK LRINALPTRSRTARGRLKDKPCRAGCRNSDGVKVIETLNHITQVCER THGARVKRHRDLVDFAVKGLQRHVRVLKEPHYKTVNGVRKPDIVIK IPDHTYICDFQVWSDTSCLELEFRKKALKYAEKGLCDQLTRDHPGELS FTAITFNTRGLIAKSSVTALRKLGMPPRSIMTLQKICMEGSLIWRIFN QTTAMARN (SEQ ID NO: 1482)</p>
R2	R2_DAn	.	Drosophila ananas sae	<p>AGAATATGGATTGATTGT GCAGAGGGGTGCTATAC CGTAACTCGTAAGCCATGC AATCAGATCAAGTCGACTC AAAACCTCCTCGTGGTATT CTCTGGGTGCCAGTATTTA CTGGTAGCTGA (SEQ ID NO: 1238)</p>	<p>TAGCCAATGCACGGGTTC CAGATTAAAGCTTGCTGCC GAAGCATACCATCAAAAT CGGCATAAAATTCGCTTA ATAAGGAGGTGGTTTT AGTAGTAGGGCTCCCG GGACTTGCTCGGGATG AATCGTGCATCGGTATAA TTGGGATCGATAACAAAT ACCAACTAAGTTATTACT AATATATCGAAATACATA AATATCCCGTCCTTACGT ATCTTTGAAGATTTCAT CCTCAGCGAACAAAAA AAAAAAA (SEQ ID NO: 1361)</p>	<p>FERRKDPWGYRPPGTLLKQIGATENNEPRNLNRFVREGESTASSLESTQ FGTSAEVNLAGRVPCTICEMTFSSKRGGLGVHMSHRHKDDLDQAQLR VDKKARWSEETLMMARKEVELAASGVRFNLKKLAEIFTHRSADAIS SYRKSEYKAKLEIQRGQSVPTPEAEIINTQRRPSNEQNRRVPRSE GGPIAPTEQTNNEILRVLQGLAPVCLPRWRAEVLQNIVDNAQVSG QETTLQSSYLMEIFPRNNEPHILTRPRTEPRNMQRQRQYARVQ RNWDKHPGRCKISLEEDDESVMNPQEVMEPYWRRVMTQPSSSI KRDMFNMEHSLERVWSAVNQDRLRATKVKLSSSPGPDGITPKTARS VPEGIMLRIMNLILWCGNLPYSIRLARTIPKKTATANQPDYRPIVSP SVIVRQLNAILASRLSAINWDTQRQGLPTDGCADNTTIVDLVREH HKRFKSCYIGTLDVSKAFDAVAHEAVYNTLASYGAPKGFYNYLRKAYE GGTMLAGNWVSEAFIPARGVKQGDPLSPILNVLIDRLRLSLPSEI GAKVGNAMTNAAAFADDIVLFAETPMGLQLLDTTVCFLSSVGLTLN TDKCFVSIKGOAKQKCTTVERRSFLIGGRECPSLKRTDEWKYLGKFT AEGRARYDPAEDLGPKLLRLTRAPLKPQKQLFALRTVLPQLYHKLTLG SVTIGVLKFKDLVRYTARKWLGLPVDVPVSVFFHAPHKSGGLGLPSLR WTAPMLRLKRLSNIKWPHLERSEVASSFVEEEMRRARDRLQAGSEEL LTRSQVDSYLANRLHMSVDGCGLREAEERFAPQHGWSVQPTRLLTGK EYTDGILRINALPSRSTRTRGRHELERRCRAGCDAPETTNHILQQCYR THGRIARHNGVWNFLKRGLERRGCWVHVEPSLQGETGLNKPDLVAI RQNRIVIDTQIVTDGHSLDQAHQRKVKYDTPDRTNLRNRRSFGAFDI EFHSATVNWIRGWSGQSVKRLIASDLLSSGDSNISVIRVISGGGLWSW RQFMVLSGYTRDWT (SEQ ID NO: 1483)</p>	
R2	R2_DM	X51967	Drosophila	<p>TTGGGGATCATGGGGTAT TTGAGAGCAGAGGGGA GTAATCTTCTGTAATTCGT</p>	<p>TAGCTAAATCGTTTGGTT CAAAACATTGTCTTGTCTG TCTTGGCATAACATCAAT</p>	<p>MTTRPSVDIFPEDQVEPNAAATLSRVPCTVCGRSFNSKRGGLGVHMRS RHPDELDEERRRVDIKARWSDEEKWMMMARKVELTANGCKHINKQ LAVVFANRSVEAIKLRQRGDYKEKIEIRGQSALAPEVANLTIIRRRPS</p>	

			melano gaster	AAGTCATATCATATGATGT GCGGAAGGGGAATTTTAC TCTGTAACACAAGTCTC TCTTTACTCAAGTCGACT CAAAACCTCCTCGTGGTG GTCCCGGTAATGCTAAACT CGTTAGCAGCTAAATTTGA GCGGAAAAAATTTTCCGAT GGGCTGGTTCACAGAGG AAATTTATTCATATTGAA CTACAAGCACAATAACG AGCCTGGATACCTTTACA CAATCTG (SEQ ID NO: 1239)	AAAGGCATAAACATCGC AAAAATAGTTATAATT AAATGGCTAGGAGATG GTTTTAGTACGTAGCGGT TGCGGAACCTCGGTTTCT ATAGAGCAATGAATCGT GCATGTAGGAAAACTG ACCACACAGGTGTTGGC AGACCTAGTATCTTTCGA AGATTTCCATACCTCCGC GATCAAAAAAATAAAAA AAAAAAA (SEQ ID NO: 1362)	RSEQDHQVTTSETTPITPFEQSNREILRTLRYGSPVECHSKWRAQELQ TIIDRAHLEGKETTLCQLSLLGIFPAQGVHRHTLTRPRRPNRRRESRR QQYAVVQRNWDKHKGRCKSLNGTDESVMPSQIMVYVWREV TQPSPPSSCSGEVIQMDHSLERVWSAITEQDLRASRVSSSSPGPDGIT PKSAREVPSGIMLRIMNLILWCGNLPHSIRLARTVFIKTVTAKRPQDF RPIVPSVLVRQLNALATRLNSSINWDRQRGLPTDGCADNATVD LVLRHSHKFRSCYANLVDKAFDSLASHASYDTL RAYGAPKGF/DYV QNTYEGGGTSLNGDQWSSEFVPARGVKQGDPLSPILFNLVMDRLL RTLPEIGAKVGNATNAAFADDLVLFAETRMGLQVLLDKTDLFSIV GLKLNADKCFVGIKQKQKCTVLEAQSFVYVGSSEIPLSKRTDEWKY LGINFTATGRVRCNPAEDIGPKLRLTKAPLKPPQQLFAIRTVLIPQLY HKLALGSAIVGLRKTDLIRYVRRWLNLPLDVPIAFVHAPPKSGGL GIPSLRWVAPMLRLRRLSNIKWPHLTQNEVASSFLEAEKQARDRLL AEQNELLSRAPIEKYWANKLYLSVDGSLREGGHYGPQHGWVSQPT RLLTGKEYMDGIRLINALPTKSRTRTGRHELERQCRAGCDAPETTIN IMQKCYRSHGRVARNVNCVNRKIRGLEERGCWVIVPSLQCESGLN KPDILVALRQNHIDVIDQIVTDGHSMDDAHQRKINRYDRDPDIRTEL RRFEAAGDIEFHSATLNRGWSGQSVKRLIAKGLSKYDSHISVQV MRGSLGCFKFMVLSGFSRDWT (SEQ ID NO: 1484)
R2	R2_DP e	.	Drosop hila persimil is	AAGATATGGATCTGAATA ATAGCGTAGAAGGGGAGT CATTCCGTAAITTCGTAAAT CGTAAAAATCAGATCAAG TTGATTCAAGACCTCCTCG TGGTATCTCTGGATGCTA TTAGACTGA (SEQ ID NO: 1240)	TAGCTATACACTATGTT GGAGAGAAGACGCTTGC TACCTAGCAAAATGTGA AATTAGGTATAACATCG TGGTTGTAAACTTGAGG TGGTTTTTGTAGTACGTAT GCGTGATTACTTCGTAAT CATGAATCGTGCATGCTA GTGGGTTTGGCCTCCAC TAGTATCTTTGAAGATTT TCCTTCTCAGCGATCAA AAAAAA (SEQ ID NO: 1363)	SSFGVLNLTNSETVLWGCQPLGQFSLIGTNNMQNTTPRIINTSLTNQ IPTVSSLGAQSEHSAQVNPNSGVQCTICSSFRSGSLGVHMSRRHK DEFDQLRLRTDRKAQWSEELSMMAKEIELAANGERYLNKKLAEVF TNRSDAIKKCRQERYTKIEQLKGQAVPLPEALESETIQRPSIRER DLLVTPPNTLTGTPTELSNSELAVLQGYPPVVCNDQWRVEVLQSV GAQASGKEITLQRLSTVLMFVPSQNDRIQTRPPRRPNRRRQGRQ QYALTQRNWDKHKGRCAIKAILDGTGATMPSQGIMGSWWRQVM TQTSPTYSGTNTTFRTEHPLEGVWSPITLGLDVRHVSILTSPGPDGI TPRTVRSIPSGVMLRIMNLILWCGKLPVSIRQARTIFIPKVGNASRPQ DFRPITVQSVMVRLNAILASRLTSSVDWDPQRQGLPTDGCADNTTI VDLILRDHKKRCKSLYATLIDISKAFDSVSHAASVATLTAYGAPKEVD VWQNSYEVCGTTLNGDQWRSEEFIPARGVRQGDPLSPIIFNLIDQLLR SYNPEIGATIGDHTTNAFAAFADDLVLFAETRLGLQTMLDTTVDLSSV GLTLNSDKCFVTGKQKQKCTVVPETFRIGSRSCPAKLRKTDEWKYL GITFTAQGRTRYSPADDLGPKLLRLTRSPKPKQKLFALRTVLIPQLYH KLTILGSMIGVLRKCDILVRSTVRKWLGLPLDVSTAFFHAPHTYGGGL IPSVRWVAPMLRMKRLSNIKWAHLAQSEAAASSFLTDELNKARGRTL AGLNLTSRTEIETVWANRLYMSVDGRGLREAGLFRPQHGWVWCQPT RLLTGQDYRNSIKLRINALPSRSRTTRGRNELERQCRAGCDAPETTIN

R2	R2_DP s		Drosop hila pseudo obscura	CAATTGGAAGATATGGG TCTGAATAATAGCGTAGA AGGGGAGTATCCGTAA TTCGTAATCGTAAAAATC AGATCAAGTTGATTCAAG ACCTCCTCGTGGTATCTTC TGGATGCTATTAGACTGA (SEQ ID NO: 1241)	TAGCTATACACTATGTT GGAGAGAAGACGCTTGC TACCTAGGCATAATGTGA AATTAGGTATAAACATCG TGGTTGTAACACTTGAGG TGGTTTTTTAGTACGTAT GCGTGATTACTTCGTAAT CATGAATCGTCATGCTA GTGGGGTTTGGCCTCCAC TAGTATCTTTGAAGATTT TCCTTCTCAGCGATCAA AAAAAAAAAAAAAAAAAA AAAAA (SEQ ID NO: 1364)	ILQNCYRTHGRRVARHNCVNNLRIEEKGHTVHVPSLQLETSYSK PDLVCIRDNHACVIDAQIITDGLFDDVHHRKVEKYKRPVEVISALRREF GVSGNVEVLSATLNWGIWSNQSVRRLLIAKGLISSGDSNVISARVVT GGLYCFRQFMVLAGYTRDWT (SEQ ID NO: 1485) SSFGLIVTLNSETVLWGCQPLGQFSLGTNMQNTTPRIINTSLTNQ IPTVSSLLGAQSEHSAQVNPNSGYQCTICESSFSKSLGVLHMSRRHK DEFDQLRLTRDKAQWSEELKEQLKGQAVPLPEALESETIQRPSIRER TNRSDAIKCRQERYKTKIEQLKGQAVPLPEALESETIQRPSIRER DLLVTPNTLTGTPTELSNREILAVLQGYPPVVCNDQWRVEVLQSIDV GAQASGKEITLQRLSTYLMIEVFPNQNDRIQTRPRRPNRRQGRQ QYALTQRNWDKHKGRCKAILDGTGEGTATMPSQIMGSYWRQVM TQTSPTYSGNTTFRTEHPLEGVWSPITLGDRLVHRVSLTKSPGPDGI TPRTVRSIPSGVMLRIMNLILWCGKLPVSIQARTIFIPKVGNASRPQ DFRPITVQSVMVRIILNAILASRLTSSVDWDPRQRGLPTDGCADNTTI VDLILRDHHRKCKSLYIATLIDSKAFDSVSHAASVATLTAYGAPKEFVD YVQNSYEVCGTTLNGDGRSEEFIPARGVRQDPLSPIIFNLIDQLLR SYNPEIGATIGDHTTNAAFADDIVLFAETRLGLQTMLDTTVDFLSSV GLTLNSDKCFTVGIKGQPKQCTVVPETFRIGRSRCPALKRTDEWKYL GITFTAQGRTRYSPADDLGP KLLRLTRSPKPKQKLFALRTVLIPQLYH KLTLSVMIGVLKCDILVRSTVRKWLGLPLDVSTAFFHAPHYGGIGI PSVRWVAPMLRMKRLSNIKWAHLAQSEAASSF LTDELNKARGRTLA GLNELTSRSEIETYWANRLYMSVDGRGLREAGLFRPQHWVWCQPTR LLTGQDYRNGIKLRINALPSRSRTTRGRNELERQCRAGCDAPETTNHI LQNCYRTHGRRVARHNCVNVNKLRIEEKGHTVHVPSLQLETSSVK PDLVCIRDNHACVIDAQIITDGLFDDVHHRKVEKYKRPVEVISALRREF GVSGNVEVLSATLNWGIWSNQSVRRLLIAKGLISSGDSNVISARVVT GGLYCFRQFMVLAGYTRDWT (SEQ ID NO: 1486)
R2	R2_DS e		Drosop hila sechelli a	GGGATCAGGGGTAATTGC GAGCAGAGGGGAGTATT TTTCTGTAATTCGTAAGTC ATATCATATGGTGTGCGG AAGGGGAATTTTACTCTGT AACTCACAAGTCTCTCCTT TACTCAAGTCGACTCAAAA CCTCCTCGTGGTGGTCCCC GGTAATGCTAAACTTGT AGCAGCTAA (SEQ ID NO: 1242)	TAGCTAAACGTTTGGTT CAAAACATTTGCTTGCTG TCTTGGCATAACATCAAT AAAGGCATAAACATCGC AAATAATGGTAATATATA AATTGGCTATGAGGATG GTTTTAGTCGTAGGCGT TGCGGAACCTCGGTTCAG ATAGAGCAA TGAATCGT GCATGCTAGGAAACTGA AGTGTTGACAGACCTAGT	FERQNFSDGLVPQRKFHIGTTNRNNEPRSNLRLNMLTRPSVDIFPED QYEPNAAATLSRVPCVCGRSFNKRKGLGVHMRSHRHPDELDEERRR VDIKARWSEEEKWMMARKEVELTANGHKHINKQLAVFANRSVEAI KKLRQRGDYKEKIEQIRQSALVPEVANLTI RRRPSRSEQNHQVTTSE TTPITPEQSNREILTRLGYSPEVECHSKWRAQELQITIDRAELEGGKETT LOCLSLYLLGIFPAQGVHRTLTPRRPRRNRRESRRQYQAVVQRNWD KHKGRCKISLNGTDESVMPSQEVMPVWREVMTQSPSPSSCREVI QMDHSLERVWSAITEHDLRASRLSSSPGPDGPTPKTAREVPSGIML RIMNLILWCGNLPHSIARLARTVFPKTVTAKRPGRFPIPSVPSVLVRQL NALIATRLNSSINWDPQRGRGLPTDGCADNATIVDLVRHSHKHFRS CYIANLDVSKAFDSLSHASIYDTLRAYGAPKGFVDYVQNTVEGGGTSL

					ATCTTTCGATAGATTCC ATACCTCGCGATCAAAA AAAAAAAAAAAAAAAA (SEQ ID NO: 1365)	NGDWSSEEFVARGVKQGDPLSPILNLMMDRLRLNLPSEIGARVG NAITNAAFADDLVFAETRMGLQVLDRDTLDFSLVGLKNADKCF VGIKQPKQKCTVLEAQSFVGSREIPSLKRTDEWKYLGINTATGRV RCNPAEDIGPKLQRLTKAPLKPQORMFALRTVLIPQLYHKLALGSVAI GILRKTDKLIYYVRRWNLPLDPIAFIHAPPKSGGLGIPSLRWVAPM LRLRLSNIKWPHLTQNEVASSFLEAEKQARADRLAEQNELLSRPAIE KYWANKLYSVDGSLREAGHWGPQHGWNQPTRLLTGKEYIDGI RLRINALPTKSRTRTRHELERQCRAGCDAPETTNHIMQCYRSHGR RIARHNCVNRKRGLEERGCVVPEPSLQCESGLNKPDVALRQNH DVIDIQVTDGHSMDDAHQRKINRYDRPDIRTELRRRFEAAGDIEFHS ATLNWRGIWSGQSVKRLIAKGLLSKYDSHISVQVMRGSGLGCKQFM YLSGFSRDWT (SEQ ID NO: 1487)
R2	R2_DS i	Drosop hila simulan s	GGGATCTGGGGTAATTGC GAGCAGAGGGGAGTATT TTTCTGTAATTCGTAAGTC ATATCATATGGTGGCGG AAGGGAAATTTTACTCTGT AACTCACAAGTCTCTCCTT TACTCAAGTCGACTCAAAA CCTCCTCGTGGTGGTCCCC GGTAATGCTAA (SEQ ID NO: 1243)	TAGCTAAACGGTTTGGTT CAAAACATTTGCTTGCTG TCTTGGCATAACATCAAT AAAGGCATAAACATCGC AAAATAATGGTTATATAT AAATGGCTATGAGGATG GTTTTAGTAGTAGGCGT TGCGGAACCTCGGTTTCAG ATAGAGCAATGAATCGT GCATGCTAGGAAAACTG ACCACACGCGAGTGTGGC AGCCTAGTATCTTTGCA TAGATTTCCATACCTCCG CGATCAAAAAAAAAAAAA AAAAAAAAAAAA (SEQ ID NO: 1366)	TCLAANLSGKNFSDGLVTQRKFTHTGTTNTNNEPRISLHNLMTTRPSV DIFPEDQYEPNAAATLSRVPCVTCGRSFNSKRGGLGVHMRSRHPDELD EERRVDIKARWSEEEKWMMMARKEVELTANGHKHMKNLQALVYFA NRSVEAIKKLRQGDYKEKIEIRGQSALVPEVANLTIIRRPSSREQN HQVTTSETPTPFQESNREILRTLRGVSPVECHSKWRAQELQTIIDRA ELEGETTLQCLSLYLLGIFPAQGVHRTLTRPPRRPNRRSRQQYA VVQRNWDKHKRCIKSLNGTDESVMPSQEVMPVYVREVMQTQS PSSCSGEVIQMDHSLERVWSAITEHDLRASRISLSSSPGPDGITPKSAR EVPSGIMLRIMNLILWCGNLPHSIRLARTVFPKTVTAKRQPDFRPSIV PSVLVRQLNAILATRLNSSINWDPQRGFLPTDGCADNATIVDLVLR HSHKHFRSCYIANLDVSKAFDSLHSHASVDTLRAYGAPKGFVDYVQNT YEGGTSLNGDWSSEEFVARGVKQGDPLSPILNLMMDRLRLNLP SEIGAKVGNATNAAAFADDLVFAETRMGLQVLDRDTLDFSLVGLK LNADKCTFTGIGKQPKQKCTVLEAQSFYVGSREIPSLKRTDEWKYLG NFTATGRVRCNPAEDIGPKLQRLTKAPLKPQORMFALRTVLIPQLYH KLALGSVAIGVLKRTDKLIYYVRRWNLPLDPIAFIHAPPKSGGLGIP SLRWVAPMLRLRLSNIKWPHLTQNEVASSFLEAEKQARADRLAEQ NELLSRPAIEKYWANKLYSVDGSLREAGHWGPQHGWNQPTRL LTGKEYIDGIRLRINALPTKSRTRTRHELERQCRAGCDAPETTNHIM QKCYRSHGRVARHNCVNRKRGLEERGCVVPEPSLQCESGLNKP DLVALRQDHIDVIDIQVTDGHSMDDAHQRKINRYDRPDIRTELRRRF EAAGDIEFHSATLNWRGIWSGQSVKRLIAKGLLSKYDSHISVQVMR GSLGCKQFMYSLSGFSRDWT (SEQ ID NO: 1488)	
R2	R2_DY a	Drosop hila yakuba	GGGAACATGGGGTAAAG GTGAGTAGAGGGGGAGT ATTTTITACTCTGCAACT	TAGCTTAAACGGTTTGGT TCACATACATCTGCTGCG TGCCTTGGCACAATATCA	FERRIFPKGLVPLTKDNHIGTTNLQNEPRIFTNDLLTRPSVDHVPEDQ YEPNAAATLSRVPCVTCGRSFNSKRGGLGVHMRSRHPDELDEERRV DIKARWSEEEKWMMMARKEVELMANGCKHINKQLAVYFANRSVEAI	

				CATAAGTCTTGCTTACT CAAGTCGACTCAAAACCTC CTCGTGGTGTTCGCGTA ATGTTAACTTGTTAGCA GCTAA (SEQ ID NO: 1244)	AAAAGGCATAAACATCG CACATATGGTTATTTAC GGCTATGAGGATGTTTT AGTACGTAGCGTTGCG GAACTTCGGTTCGGATG AGCAATGAATCGTGCATG CTAGGAACGACCAATA ACAGCAGCCCTAGTATCT TTCGAAGATTTCCATACC TTTGGCATCAAAAAAAA AAAAAAA (SEQ ID NO: 1367)	KKLRQRGDYKEKIEIRGQSALAPEVANLTIRRRPSRSEQDHQVPTSE ASPTPLEQSNREILRTLGRVSPVCPCKWRAQELQTIIDRAEFEGKETT LQCLSLYQGFVQGVHRTLTPRRPRNRRESRRQQYAVIQRNWD KHKGRCKSLNGTDESVMPSREFMEPYWREVMQTQPSPPSCNGEVI RTDHSLETVWSAITEQDLRASRVLSSSPGDPGVPTKAREVPSPGIML RIMNILLWCNLPHSIRLARTIFIKTIVTAKRPQDFRPIPSVLVRQLN AILATRLTSSIDWPRQRFSDTGGCADNATIDVLVLRSHKFKSCYI ANLDSKAFDSLHAAIYGTIRAYGAPKGFVDYVQKTYEGGGISLNGE GWCSEEFVARGVKQGDPLSPILNVIDRLRALPSEIGTKVGNAMI NAAAFADDLVLAETRMGLQLTLDKTVDFLSTVGLKNADKCTVGIK GQPKQCTVLEAQSFVCSGREIPTLKRDEWKYLGHIHTFASGRVRCN PAEDIGPKLRLEAPLKPQQLFALRTLIPQLYHKLGLSVTIGVLRK TDKLIRFVRRWLNLPDSVPIAFVHAPKCGGLGIPSLRWVAPMLRLR RLSNIKWPHLVQSEASSFIEAEKQARGLIAEQNELLSRAIEKYWA NRLYLSVDGGGLREAGHYGPQHWVYVQPTRLTGTKEYLDGIRLINA LPTKSRITRGRHELERCRAGCDAPETTNHIMQKCYRSHGRVARH NCVNRIRGLEERGCVIAEPSLQCESGLNKPDLVLRQNHIDVIDV QVTDGHSMDAEHQKINRYDRDPDIRTELRRRFEAGDIEFHSATLN WRGIWSGQSVKRLIAKGLLSKYDSHISVQVMRGLSGCFRQFMVLSG FSRDWT (SEQ ID NO: 1489)
R2	R2_KF	GU949 558	Kaloter mes flavicoll is	GAGAGGTCACTGTGCTG ATCAGCTGGACACCTAGCT GACTGCGTGATGCTGC ACGCCGCGCGCCTCCT CGTGTACCGCTGTTAAC GCAGGGTCACTGAGACTA CCTGCGGCTAAAGGCGCG CGCAGGGATTGTAAGTC CACCACCTCCACGTGGTTC CCGCGGGCAACGGCATA GTTCACTGCGCGGCAAGC GTGCGGTTCCCTCGATCCT CCTCTGATTAGGGATAGG AGGGGGCGGTGGCTCCC GCCGACGACCTGGCAAC ACCTTGGACACGCTAAGAT AATAGGCCGCTCCTCCGGG	TAATGTCCCTTTGGCTT GCCCCACCTGCTTAAAG GAATGGCAGGAAAGAG AGTGATCCGTGCCATAGA AATATGTTATCCGGGGC AAGTCACTAGCAATATGG GACTTCTCCGGGTCCGTG CGGTCTTCCAAACATGAG CTGGACGTAGTCCACTCT ATGACTTGAACGATACGG GGCGGTATCTCCCCCGGA AGAGGTGCGCTAGGCGA CTTAGAA (SEQ ID NO: 1368)	MEVPLTSSLSGSGATQAPGTPPELLGEHTVERPGLDQGHSGYGLMDDV ELPVRLPFFGLPCGCRLLTSEETISSHHRRVHPDARTRWVVCYGCDS PFMTYRAIKCHLPKCSGRKVTGDIHNGCTKRFESQRGSLHKKRRA HPGLRNEEMLEPPVRAERRNAHKSIWSIDEIRILEQYEAAYVGDH INMKIAHLPFKTNKQVSNYRNDRRKSRATDASQQLGPNNGNR GINPSGQSSPLFEGSDAEGDEDFNVLPPTLGGLEPAGQVHSLSEG ETSPLVGEADPCFMGGTSPSAGEASGSLTLLGPDTPADGYSVLRKDLQ LSVQTSPLAVGSVGTESVQFERGVLSGCTPPEFLHPEQFAHCANND PVLNASEEQVHAPLGEANDLPDNNHPSSELGVDPEDPTCSPATEQV QPSSEEEADDPFAQKAWRRRRVASYALKIETGVLPQVDDLLRLRD GDTQSKVTCEEVEVLSLTRLTGATPKRVEGRTKWTYKSRTNHE ARKRIMYARCQDLYRRPQRPVERAVGYQAEESLDNDQDERPSHGA FETFTGLWGKSGQCQNTMPPGVPRHTGHVLEVPKDIYSRLKIK KDYAPGPDGVTCLKVQSMGAYPSLAKVYNLMLTYFSSCWKEHK TSLIPKDRGSPMDVSNWRPITIGSLSRITYGLIERLLRTVSDIHQRQV GFMPVNGCAANLIFDECIRQAKKEGTIVGSLIDVAKAFDTPHEAIL RALSSQGVDEHTMAHIRMYSGIRTRINGKSDIPLVRGVKQGDPLS PMLFMVMIDPLRDLQRKGFRIGGHEIGALAFADDDIVLADSIDGAQ

				CCGGCGAATCATAAGCA CACA (SEQ ID NO: 1245)		DHVDQVGRYMNKLMGMITNPKSSFLITAMRKTWICRDPGLSIGET KYPGARSSALKYLGVNNTLSEGLSGALIDKLMQAVNRARGALAKPL QKVNILIRIIPKELYGIILGGPSLTRLHAADKCVRMVAVKEILHLHPSTT DHVLYARKKDGGMGIPRLAHLVRLASRLSGALLASGDVAVQAAAGM AGDLEGRCKKVANDRLNWPVTLRDVVVRASNKFKSQESKDWERLAS QGHGVKDFRNDRLGNCWLYDPTVLSSRYTDALRLTNTFGVNVAL RRADKLEVNCRCHGKPETLGHVLGECVAGKGMRIQRHDKMAAF VATKCEEGYQTTREQLFSEQGLKPDLVVIDGERALIVDVTVRFESG NALSRGASEKIEKYQPLADYFVSQGAVERANVLPVVGSRGAIQTATL KSLATLGLDVERVGKYLAICAVASSVEIACMHLDYT (SEQ ID NO: 1490)
R2	R2_RL	GU949 555	Reticuli termes lucifugu s	CTCACTGCTGTCAATGCCAT TGTTAATGCGTTGGGTGAT GGCGGTGATGGGAGAC GAGTTACAGCAGAGCTGG CTTCACGGGGGGGCTGT AGTTGCCGAACCACTCT GCTTCTGAGTGCCTCCACG TGGCCTCGCTGGAAACGT CGGAGCTGCTTACGGCTA ACCCGTAGACAACCTCGG CGGCCAACTCAGAAAC GGCCTTACTACCGGGTC ATCCCTCGCGGGCGCT AAGACCTTACTGTATGTAC TACGGACTACAGTTGAGT AGCGGGGAGAGCACTTG ATGTGAGGGTAGCAACTT GTTGCTACTTCGACTGCT CCCTGAAGATTCCGAAG GGGTGCCGCAAGTCCATG GGGCGGTCTCTGCTGTGG GAGTTATCTTGCAACCGG GAAGCCCTCAGTGATAA ATACTCCAAATGGCTTAG CCACCTCAGCTGGAGAGA GGGAATTAGGGGTTTAC TCCTAAGGGCCGATGCCTC	TGAGTTAGATATATATGA CGACTGATCTTCCCTAA CTAAAAATATAATTGGAA TGTGCATTACTTACTATT ATTGATTTATTGACATCC GTAGTAGTCTTTAGTG GATTAGATCGTTAAACCA TTCTGTGACGTCGCTGAT GTCTATGTACCACTGG CAGAACTGCTGACTAAGA ACAAATAATTTTAAACA ATAGACAGCACTAAGAA ATTCTGCAAACTGGAACG TGCCCACTGATGACCA ATATACCGGAAAGGGAA AATGACACATCCCCCTT AAA (SEQ ID NO: 1369)	MMADYNNNSVDHAEDNTRLIFARDAVLARVCGPFDNLECGLCGVLL TSLQGVREHCHRAHNLDLTFQCTKCDKGFSSYRGICCHFKCIGARI SVSEGPLSCSECEFEFSKRALSTHERHMHGIRNAKRLKDFNPRGG GGKTIHGNTKWTEEEVQLLVLSKRFEYKSKINKEISLITSKTCQJSD KRRYLNHNGNGGLAAEALEFCEDSHPEVTESDGAVLSEIMDEEC HQSSVTMRSSIVHGDIGREVQKGKELVRIPPDNSVMGNCVLLRLKAT DKRSDLDLKDRELIDIEKATKANRESADGRVQSVAGNEIDPDTFQ WKELLGQVGRPRVDENSELFDLDDKLTKESSDSPVWVNDNCEIV SDLCQVLCKKYEYELGRQHVRKGRHRGHIHKREKRECQKIFRKSPPR KLAEYLRDKDLSHISKDASTPQIEQYYSQVWGEPELLESNTIEEKLP SSLFDCLPPITPEEVEGRHRIKRPSSAPGLDGVKRIHLVGKGLTLVLKLY NILLFTGYPECWKRNRNRTVPIKIGKDLSEVGGWRPLTIGSLARMYS AFLERRIRRVTSLSQSGFTNIQGGCHVNLTKIEGIRQAKVKNNGGVIV SVDIEKAFDTIPHSVIFSRLASQGVPLLRKISNMVKDVTVIEGQCIP KRGVKQGDPLSPLFNIAIDPVLRSLEEFQGGPLGNSAIKILAFADDIIL GASSAGQAQMVMDMLGIGLITSCGLGVSHRKCFCGQVKNKNTWTI VDPMITLNGSSLPFSGPEDRLPYLGVDPINPWRKSRVDAGQRLISAA KRGSQLSKPQQKINLITAFLLPKFLYLIEDPPSPAYLKSIDHDLRQYKN ILHLPNCVSTAFMYSKPRDGGGLPRLSCLVPLAHLKAGIKLSLQDLSL VREITTSDFRVRTMGSAHSLWASWSLTLDIYKLSALKRRKAKAW ESCVSQGGAAQFRGDSIGNNWLHNPGRYRPGQYIEALKLRANLTG VRVNLKRSQVYNVPIITCRFKDIPETAQAHVLGLCPKTKGMRIQRHDSIV NRVRDKLTKSPVALMHEQNFTVEEGQVFKPDIVITLGEVGYVIDTV RYDDRDIKDAVSEKIRKYEALKGYLKDLYPQLNKVEVLPVLFVSGRGAV PGSTVHNMMGLLGTREMVMVHSRKVIADSLIISNFLEVY (SEQ ID NO: 1491)

				CAATCGGATTTCCTCCGGGC ACGGCTATTATCAGCCG GCTAAGTTCGGACCTCAT CTACTGATGGATACTCTC TCTCCCCCGCTCGTGA TTAGTTGGGAATGCACGC GGCGGGTGATTCTGACA AGCCAGCCCAAAAGTTCT TACCACGAGTACCCGG GGTGCACCCGAACATAA TATTTCCCGCATCGGT CGCTGTGGGGGATGTCG GGCCTGGTGGCCGGAA CCGTCAGCCGCTCTTGAG TGTCTGCGTAACCCGA GGCGGTGACGTCGAAG GCTAGGCTAGCACGGTCC AGACCCCTGAACGCTGG GGAGACGGGCCACACC TGAGTAGGAGTGGGTCT TTACCTCATTGAGGTGC TCCTCGTCTTTGAGGGG CGAAGGACGAGGAATGCA TCCGTCCTCCAGGATGCT GGTGGTTCCGCTGCTG GGCTCATGCTGAAACGCG CAGGTGCCGATTTGTCGA AGAGACATATGGGGTAA CCCATAGAACCTAGGCAG GAGTAATCCCTTAGCTGG GGGGTCAGCTGGTGGGT CCTGTCAATTTATCCTCCCC TCCATGCAGAGCCGGTC CGAGGTTAGAGCGGACT CATTTCTCCTTTTATATG TC (SEQ ID NO: 1246)

413

R2	RaR2	FJ461304	Rhynchosciana	<div>GTGGGGGATGTCGGGG CCTGTGGCCAGGAACCG TCAGCCGCTTTAGTGT CCTGCGTAACCCGGAGGC GGTGACGTCGAAGGGCTA GGCTAGCACGGTCCAGAC CCCTGAACGCTGGGAG ACGGGCCACCACTGTG TAGGAGTGGGTCCTTACC TCATTTGAGGTGTCCTC GGTCTAGTAGGGGCGAA GGACGAGGATGCATCCG TCCCTCAGGATGCTGGTG GTTTCCGTCTGGTGGCTC ATGCTGAACCCGCAGGT GCCGATTTGTCGAAGGG ACATATGGGTAAACCAT AGAACCTAGGCAGGAGTA ATCCCTAGCTGGGGGG TCAGCTGGGTCTCTGTC AATTATCTCCCTCCAT GCCAGACCGGTCGAGG TTAGAGACGGACTCATTTT CTCCTTTTATATGTC (SEQ ID NO: 1247)</div>	<div>TAGAAATGTGTCGATAA GGTGTGAATAGAAGGGT TCACCAAGAGGGAGACC TAGTTTGGACCTCAGAAA TGGGGTCATAGGAGTGA TAGGTTGTAAGCCGTTG GGGAATCCGGCTACACA TGGTATCTCAGGAGCCAT TCATGCGCTGATCTCAT AAGCGGTAATAAAGTGT GAAACAGATCTGATAAT GCCGTGCTACCAAAATGAT GTAACGAGGGCGGAAATA</div>	<div>MSNVNETSGDNPRMIATQTTGSLSSGPNQHTCELCCRTFTGTRA GLGQHVRKTHPIESNQSNVERKKRRWSPPEIRRMANMEAOATINN IKHLTQYLATYLPQRTLNAIKGRRRDAEYKELVTGIANLRNSNSTQQT NQVCNESEMSQSKILQSIRESVRDLRSRRNKYAKALQELGEAALCGK MLNEEQLIHCISMFNTAKCPGRFRKTATSHGNTKQQRQRYAR VQLYKMNKVAAMVLEETDKIQIKLPDHPMFKFWESEFEKEGEG MPERMPKDLKESPDKAIWDPVTEEEVRKAKVANNTAAGPDGQIPK SWNRISLYKTLIYNLLLYEKPVPHKLKVSRTVFIKKKDGSSDPGEFRP LTICSVLRGFNKILVQRLVSLYKYDERQATAYLPIDGVGTNIHVLAAILN DSNTKLSELHVALLDITKAFNRLHHTSIKSLVGKGFPGPYGFTFIRRMYT GLQTMIMQFEHGCKMTQVNRGVYQGDPLSGPIFLAIEKGLQALDKE VGYDIGDVRVNAGAYADDTDLVAGTRGLGLQDNINRFSSTIKQVGLV NDRKSNMTI SILVPSGKEKKMKVETGKGPENINIVPIKEI SINDFWRVIG</div>
----	------	----------	---------------	--	---	--

AAAAAGATGCGCCTAGCA TGACTGATAATTTTCACT GAAATAACGTTGAACCTTA TCTGTGTCATGTGCACAAC ACTATGGTGCTGATCAAG CACCATAGTGGTGACCC TGCTAATGTATTAGTAGAA CGTGTCCAGCGGATAATG CACACACGGCTCCGGC CCATCGCTTTTITGAGAT TCCCTAGAAACTTCAGTGT GTGCGACAACCTGTATAACC CATAAGGGATGGACAAAAG GTTATACTAGGGGTAAA AACCCTAATCGGCTAATGG CAAAATGGGATGTAGAAAT GCCAAAGATTACTCGCACC GAATAATGGTGCCGAAA AGCGGGTAATCGAATGAA ATGGTAATGCTGTAGCGG AAACATGATCACATTCTGT GACAGTAAACCATTAGAC CTAGGGGGAACTATGATT AACAAAGATACCAGCTTACA TGGAAGCAATGAAATGTA AGTCACAGTAGTGATAAG TGGTGAAGAGTCTTGTA TCACCGTAACTAGGCCAG GTTCTGGGGATGCCATGA ACTTAGGGGGAGTATGGT TAGCAGATCTACCAGCTAA CACTATTACTGATAATATG TAAGCCGCAGTAGCGCTA AGTGGGTACAGATTTC AACCAACCGTAACTAAGTTC TGTTTCGATGGACTAGGG GGAATCATGATTAAACAAG ATACCAACTACATGGAAG	AAATTAATCTGGGCGTT CTGCGGAATGACTACTAA ATATAGCGATGCTATATA TACAAACGACTGATGGTA ACACCGGCCTTA (SEQ ID NO: 1371)	ISYTNEGPERLSLTIEQDLERLT LSKTTAKGLKRTDRQIRQYVRRW LQYWIPLLRVNRVNKITESQSVLA GNPTMLAYRTYWEKELYAKVDGKD ISGEDYLNHYHKLRTNSVPTKVRT QQCHRTGGGRTL RHDRIVGLLQHE LVLIKDDTAHIVDVQVARSCLINES KVMYEACTISYKGIWDKQSVMSMR WKRFNMITSVRS (SEQ ID NO: 1493)
--	--	--

R2	YURE- 2_Cis	.	Ciona savignyi	TTATGAATGTAAAGTCACA GTAGTGATAAGTGGTGAA GAGTCTTGTAATCACCGTA ACTAGGCCAGATGTAGTC AAAGCACATGTTTAGGGG GAACAAGGTTAAACATTGG TAAAGACCAATGCAACCT CCGTAACTAAACGTGAAG GCACAAACTAAAAAGTCC AGTGGATGACAGGTGAGG TCATCACTGGACCCAAATG TTTTAAGCTCATATAACA ACAGGTGAAAAATCCAG CATTATTGCCTGATTGA GTAGCTTCACACTATTCC AAAGCCGAACCTATCTGG GTTTTTCTGAAAGGCCGT ATAGGGCATATGTCGAGA AATAAGTCAAGGTGAGG TAGTGTGGCCTGTACCCA GGGTAAAGGTACTATACG GTCGAGTGGCTAGTAGG CCTAACTAGCCACTGAGTC ACTATAATGACTAGTT (SEQ.ID NO: 1248)	TAAGATCCGCGGCTGTG GCGCCGAACGAGCACCT GCCATTCTTCTGTAGG GACTTTTACCCCTCACTC CCCCCAATAGTTTTTTTC GTTTTTCTGTTTTTTCACC CCACCCCACTCGCCTCT GGGTGCACATCCACAC GTAGGGACCTGTTTATAT TATTGCCCTTTATATGTA CCACTTTTAAATATATTT TTGTACCCCAAGATGTC TTTTCGCCAAAAA	MAGHKITMSEGLLEAVRVYGGVRNVSECPVDCCTKTFSQANNLIR HLNFGNTKHAHNFTYFTCEKCKIQHSNTKHNINHYKQCCATG GGPSCETGQYFCACEQAGLGNLESALRHFQSSHPEFNLPPRSQFSK SHPNSTYLSLKPDKHLMKILYSGPLTPGQLVCPKICLRSSAARLFHDVS KLKRLMLVDHNRVLVYETTCGKCLRPVDTSKNMRKTTSHFEKCSGES FISSPSIPQKTYKLDLPSTSTPPRKSPLQPYKPIRTFNPLTKSSQSK SDNPKPTFPFSPRTLERSASWPALESEVDP LKLEKHPSLPCALDKC PPSPRIKPSTLVPPCHTANNPKPTSPSPSTLPLPRPIRPSKPLEDWL TVRSVGPDRVILNIGPRPRGPAAGSRTTSPSPSTAPAKRVAANPIAA PLSGEPGATLDCGQTGRKVQPPKKRPTESAGSLPPPAEPATDILLTGRE GLARLVEEYHLSGDFGAFCDRLERWTALSTNNRRPKRRGRYNRGAA ARATNRGRDDRDQDDQDGGPGPVTGCRPQRYKRAALRSA FGRDMKATVRRIIDGERGDARCEIDPKTIEGRFRDELSPVREGPECSL
----	----------------	---	-------------------	--	--	---

			PPWMAEAQAQAGEHAPSNDSDQPDAYDGPITALEVEMVLSTLNVGSA PGSDGLSYGFWRALDPKGLVLSLSEFVCRERRVPGPWKSSRVTLICK DAEGDLDDLGNWRPISICQTVYKIYAVALRRQLSWALDGGVISRSQ KGFMPFEGVYEHVFLDSVADARATRRSLAVCWLDLRFNAFGVDH TTIVEALSREGAPAGLVLEMISDIYTGSGRIIRTRAGFTPDIPVGRGVRQ GCPLSGIIFNLVMEVLLRGVEANNACGYRLSCAGGASVRVLAYADDV ALVGSRAEXKIQLGVCFERFAAWAGFSFNKKCAAMVLKHQRGGR RLDSAPLRLCGEEVAILGPDSEFYKYLGAHTGYGRQTGGQLVDRVER QVRLFTSLPTQKLSALKRIVLPAMSFHLRVPCAEGLHRLDNTV RRCVKTALRLPKGSCRAFFHTSPDAGGLGITSVVAECDILTVQAFKM LSSPDHLVSLVAKGRGMHAAARMGRSETASACAMADYLSGDSVMG HXSWKTYRMPADLWTATRAASRRLSLRFSPQPGEGFLESGTFKIA PRRRSLTRRLHHRQNLWWRNQWAAALPNQKTVAAHSAAYAASNN WVKGPSLAPQALFFGLKARLNQMPTRSVKACYSRAPNYDKSCRRC GAEVETLPHVLNHCPSMKSLERHDSVLAELAAIPRGTFASVDVDR TSREHFRRVGEALRPDIVARRHDSVVVADVTCPFESCASALDTAAA RKIEYDQLCANLRQLYRKPVESHALVWGLSGSWGRNTNNTALAALGI RGAVRSRLAKQLVNLVSEGSNHWLWWSGGIPKDLVR (SEQ ID NO: 1494)
CCCTCGAACCCAGCAGCA AGGTACGTGTACCACCTC TCCACAGAGCCAAAGGCCA GAGTGGATAGAGCAGCGG CCTCTACAACCAAGTTCTC CACTCCGACGACAAAACA CCTGCTCTGTGGCCAGAG TCCTGCCAAGAAGAAATCT TCGACCCCAACACCGTCTC GGCAGGGCTCCAAGCAAC GACATCAGCCACGAGTGC CCACCGAGTGGAGCGGG CGAACCGAGGACCGCTG TCGCCAAGAACAATAACAT CACCGCTGCAGCAGCTAC CAAAACAAGGTAGTCTCC TACCTACCTACAAACTCG TCTGAATAGACGCCCGCGC GTGGGAGCTAACCTTGGT AGCTAGCTGCAGTGCCCT GCGACAGCGGCTCGAAG AGCTGCAGAGCGTCAGCC TGCTTCGACCTCGCTTGT CTCATTTCAACCTACTCCG TCCTGTGGAATCAAAAGA GCCCACTAACAATTACAT TCATAAAATCTAGCAAGAA CGAAGAGAAGCGACCAAC TTTTAATCCATAACTTTAG ATCTTTTATTATTACTGT TTTTAAGCCCTAAGCATAT TGCTTTTTTTTAGATCTTAT AATTATAAAATAGATTCA AAGTTAACACCAACCGGC CGCTACAGAGCATTTTATT TAATCAATTTGTACCGAC CTCCTGCTGCTTTTAC TTTCTCAGACTACTACCC	AATTTTGTATCACATTT TATATTTGTAAACACA GATTTTATAAACTTGC ACTATTTTATAAACTT CGCACTTATTTAAATGA ATCGCATCTTTTATATA CACCAACAAACAGGAT GTGAGCTCAGGGGAAC CAATCTGCGTCCCTCT AGCGGGGAGGGGCG CCCACTACCCACGCT CCTTGTAGCACCACAG GGACTCCCTCCGAGCC CCTGACCCCTCAACTTTC TTATTTTAAAAA AATCATATATTGATCTT GACGACGGGGTACAT TCAGCCCCAAACCCCA CCCACTACCCACGAG TGCCGGGGCATTGAAGA GCTCCGCACAATTAGCA CTTAGCTTATTATT TTTGTCAACATTTTGT TTTCAAAATTTTCACCCC TCACCCCAACCTAATAG GTCCCTCGGCTTGGCC CCTTTTCTGCTCGAGA AGCGTCACATCGCCCCAC TGACCACGACCTTCCCG ACATTGGAGTCTTGGCG TCTCCAGGTCGAAACAG TCCAAGTGATAGCACCT AATGCTCGACTTGTTCG GCCTGGCGCCGAGGA TTCCAGAACGACCATTC TTCTAAATAATTTATAT TTCAGAAATAAACTATAT ATATATCGTTGGCGGGAC		

R4	Dong	.	Bomby x mori	GGATACAACCCCTTGAAA CGAGAGGA (SEQ ID NO: 1249)	TTGTCCCGCTCGATACC GAGTGTGCAGAGCGGC AAATAAAGAAAGAACCC GACGTCGCTCTGCAGCCA AGGACCAACCAAACTCAA GCCAGCACGTCGACAAC CAACATCCTCAAGTCGGC GGTTGCTGGAACTCA TAACATCTTCAANATAAA TTATCACCTGTGCAGCA GGAGCGCTGCTTTTAAA ACTACTGTAGTGGCTC ATGATAATATTGCTCC TTTTTGGCCCGTGATAA CTTAGTNGATCGGAATAA AATCAGTTGAATCA (SEQ ID NO: 1372)	TAAAACTAGCATAATTA TTAACTCATACTAATGT ATATTACTTGGCCAAAAG CCCGTATATACAGTTCCA CCGCTCTGTCGACAGAC TGAACTGAGAAAGGGGA AACATATGAAATAATAA TAA (SEQ ID NO: 1373)	MLRRGRIFLPASTKAGKTRGRMKWSREVNLFMRTYVYVTKLETDLTI YRKLIHEHSLKYPNVISQQRIISDQRAIERNKLLSQETLDRLEEVRK QLEDEQTNVENEKLNSETSYHEVTLTPQTLTKKQQHTNIJSTQT SHSSTQTESITLLENEVDILNTNPTEGATQTEVKDKFETNLTMVSG MDPKARPLPKLYSSKLNELIRNFNDILVDYISPDQLSDVHLYCT AVTISEQLKYKIIAIEGNARHKNFPPWQQRLKADIAKLADIGKLTQ YNNRSKVVQSVQEQIFKNTKIHTSHENGNGKSKQEFDLTKQKLALK AHLKRYNNSQKRKNENTIFLTNEKLFYRNLIKPKTDNRNSNIDIPTAE QLEMYWARLWENSAKHNDAKAWITEEKERWDTIEEMQFDDVTEE EITITARLHNWKS PGIDKIHNFVKLIKLIHKTAKNLTDIISGNQSIPE FIATGYMIPKGFSEASQYRPITCLPTIYKILTTVITKINSIHENIL AEEQKGRRHMGCKEQLIIDSTIMKHATTNRLNHLCTYIDYKKAFFD SIPHSWLIQVLEIYKINPIISFLRNIMTHWQTTLLKNPPNFVTRQIAI KKGYYQGDLSPLWFCLALNPLSHQLHNDRAGYRIKQQDNTETIISHLI YMDIKLYAKNDKEMKKLIDTTTIFSNDISMQFGLDKCKTVHIKGV QPGDYITDNTITAMEPSDLVKYLGQQLKGLDHITIKQSLTSEYKRR INAICKTLGKHLIKALNTYAIPILTYSGIIKWSKTIDIEQIERITRTLTK HNNLHPKSAIERLTIKRQDGGGRMIDIWHLWRQIHSLKTFEYKSDL SEIHRAIAQNNDNNVTPNLNKKQELIDNTENLRNRNPQKDMEEWKK KALHGRPHDLSQSHIDSKASNMW/LKTSGLFPETEGFLAIQDDQVINT KNRYKIIDPTIRDDKCRKCNTPETIQIHITGACSTLTQTDTYTHRN
----	------	---	-----------------	---	--	--	--

R4	DongA G	AB097 127	Anophe les gambia e	CTCGGTCATAGAGGCTC ACCAACGTCGACTATGG (SEQ ID NO: 1250)	GAAGGCTAACCAACAATA (SEQ ID NO: 1251)	TAACATCCGGTGCAAACT CATTAAACATTAGAAAAAG AGAGAGGAGAAATGAGA ATGAGATTCACTCCTT TGGCATTTGAATAGCCCG GGGTAGGTGAAAAAGTTC CCAGCATATTGCTGAGAA GTGACAAAAATTCGGATAA TAATAATAATAATAATA TAATAATAATAATAATA TAATAATAATAATATGCA TAATAATA (SEQ ID NO: 1374)	QLANIIHQALAKHLKIQNTNTPYNYKPTVLENDCKLYYDRAILTD RTHYNRPDITLQDKNNKVYIIDIAVPNTNHNQKTFTEKMTKYTELKE EIVRIWKQKAYIVPIIISTTGVVPNHHNSLKLLDLKDNIFISLQKAAIIN TCRIVRFKFMQLEENQTYTYTQ (SEQ ID NO: 1495) METRSMRKRTTLLPEEGAPGTGAGPGTGDRAISIQRIEDEMVMQERSFS QRALPVPRTQNRNGSPINHQGNAASANAVADRQQLSLLAGGRRQ RIMWTREMNHYVIRCYVYVYTRMETDMGPRVKMLGMFENDRPFRA HQLDSLKYIRQRAILPEELEFKILEVRREFGEEAGWRESSRISARLNT IDQNTSRASEDRDLDEPTAPGLSDVQIQHMATAVTQFHGTDPLSRH RLPKLHYSYRLKTAVSINQDVLPLQYLDVSGSIEDQLIVYSAAVAVVRT LWLRTYPQGDSEGRPCSKAEPAWMMRRLENRINATRTKIGRMQEY QRGNSMKVVRQIAEMVKPELRDLTDANITEVLDIHLQRLSALAKRL RRYAECRKEQNRMFNINEREFYNNWIRNDKPNFREGLPDIGDTQF WANLCEKPVQHNSSEGMRLADERFDSDGIEDMPVLVNAQDIREAT QYTRNGAAPGPDFVYVNFYKLTITHEQIAACFNTVLEDSRKLKPFIT GGVTYFLPKDQNTKNPAKYRPLTCLSNLKNVLSVVITQVKVDHCDTN NVMTEEQTRRNKTQCKDQVIDAVIVGQAQKQNRNLDMAIDY KKAYDSVPHSYLLKVLQYKVDGNVILMQHAMGMWSTSLHVTDG KVLRSRLNIRRGIFQGDFTSLWFLAMNPLSRTLNOQCNFGYLLK SEESTRITHTFFMDDLLKLAETVQKMHLLKNVQGSNDIKMEFGIG KCRSIIHHRGQVLDADSFRAEQEIRHMMVQGETYKFLGLQLRGIH YAVIKKELQDKFLHRVCSILKSFLSVGNKVKAINTFAVALLTYSFGVMK WSNTDLEALERTIRVSTKHQMRHPKASVERVILPRKIGGVGIDIQAL CISQIHQLRSYFVESQNRHELRYTVYKADHGLSALHQAQQDYQLNCNI KTVDGKGA TWKQKELHGHHTHQLNLEHIDKVSSTWLVRCDLFCET EGFMVAIQDRVIATWNYRRCILREDEVEDRCRKNCSGGESIEHV/AGC PVLAGSAYLDRHNDVAKIVHQQLALRHKLVERFLPCYRYLPDPVQEN DCIKLYWDREIITDIILIRANRPDILVYEKKRKRATIDIDIAVTLDHNVQTT FSTKVMKYHDLAEELKQTWYLEDIRIVPVIISATGIVPMALLRSLDELEL QRELPRIQKAVILRTCSTLRRFLNPNY (SEQ ID NO: 1496)
R4	R4- 1_BX	CADV0 10081 75	Bursap helench us xylophil us	GGGATCCTGGGTTCTCTACT ACCTCGCTCCACCTCCTCG CGATGGATCTCTGGGGAAG TCTCCGGACTGAGCTAAG AGAGCGGTTAAAGTAGAGG GTGACGGCTAAAGTACCTG CCAAGTTGCGGTGAGCGG GAACATCTACTCTTCGGAG	TAAGAAAAAGCATGAAAT AATAAGAAATCAGATAA GAATAACAAGAAATACTAA TAAGTATATCATGTAAC ATGACAAAAAGAACGCA CCAAATAAGAACATGCTTG AGTGGCCAGCTCTGCAG GCAAAAGTCGAATTTGG	MTCNNAVFPADGNPAGTADNFAIRFSPSEPPGSGIRPSEPLDG RTGIGDVEHAQAGNGGFLVDVLEYKEAHRYGSKCEFCYVQTKGTCS KPRTDALWKLCEILFLHHAYTANQNKSIELAESAFRRAGITRRSKATIA KRWSLIQRGKGDYKEYWDEYEFKRYECNPTPIVRKRNRRLAAGLQ SPSSVPNGYEFERKRTCTETPLDTKASSLPILCNLLTIGVGVENVEENMS VECTPEKLSGTANSSVPGAEGYERRHNNVNNVNEPAAGCPQDVPVA NNLIDSPITNDRLAEAFKAQLDRAERSYMMRRRLPRLKNLSPDERMWI GTTVERLLETVSEPVCEQWRLANAGLYAAIRSIAMRPLDAAREAH	

				AGAGGGGAAGCTCTATGG CGCGTTAGAAAGTTGG ACTACGGCAACGCCAGG AGATGGGAAGGTTTCATC AGGTGATACTAGTTCGCTA CTGTCTTCGATGTATCCG GAACATACTCGCAAGTT GTGAACATGTGAAAGTC TGGATCCAATCCAAGACCA CGGGCGCAATTAAAGG TGTGAAGCAGCTTGCTGG TGATCACACGGTGGTAC CTACACGGCGGGGGAAC ATCTTTAATGCCGAGATGA CCGCGCAATAATAGGGA GCCGGCAAAAGCCGAGCG TAAGTGAAGGATACAG AAATTGCTCAAAAGTATA CCAACCGCCACTTATAAA AGACCGAGAAAAGGGTAC CACTAGAGCTATTATAA ACATCT (SEQ ID NO: 1252)				AACAGCCGTAATGGAA GACTGCAACAAACGTG GGGTAGCAGGCAATATG TAACTATGACAGACCAAA ACTCCGAACTCTGGTAA TGAGCCCGTGCCCCCAA GCATGTGCTCGTTTCA TGATAGTGAACAGTTC TCTTTAACCGTGATGAT TACGCCCTGCTTAAATG GCAGGTGCCACCAATAC CGAACACTGTTGAGGTA TGGTGGTCCGAATGTGA AGCTGGAGTACAATTT GGTACGAGACCAACGAGC GCCCGCATCTAAGTGAT GACGCATGCTCGGAAC AATGAAGACGGCTGGCA AACATTACGAGTCGCAA A (SEQ ID NO: 1375)		KTWLLNMKMKTERKLRQIQIGWVETTRTKNEARTERQEIVRKVAKL RRERPEMDLDSVSVHLKRKLELLKGRIVQRTAERLRDRTREAAAGPYG KTAIRGQGFAPNVKDATQYWSGLAQSPSGQKCSAILSQDWKELVE CNLSLPDQMEPLVVQGISRASPWKSPGPDGIFNYYWRQDFIVDWL KQLMLDSLRTGHYPWKLSGRVTLVKDGDPTKAENYRPTCLNGCF KMINVWSEVLKRVNTIALPIEQMALRRKVVWACVESQIWDQIKQR KLSDRQTKVAVWDFSKAYDSLNDHDAIKFVIGVLKPTGINNYLLDS MQNWSTHLELKSSGKVVGRGPIKRGVLLQGDLSLPTLFVVVTSIIVR HIKTISSDIQMYMDIKLYGDKQETLTRLIKELQTVSNKLGCMNLKK CAILGDDLPEEINGIEHLKESYKYLGVQPREITQVRATMAALEKKILTEV DTSGLAAELSYRQIRSRVNSKIAPLVRFVQSMVLVTPRDVLKYNRLG GIDVEIRRLVYKIRYKSNVRLYLDRKVGIGFVNLCRIMVEAVAA RAVYCRAPSFNEQDFLAEQNTSPITAAQTILDKCGINIELSTSLGDV KKIVRNHYHELWLTAWKNTGLYKRWENDHVDIKRSSLWINRGNLSA NNARIGIQDNSIFCRGFGVGNKCDTKYCRLCGDGIESVSHIVTGCTPT HRTNLYIERHDCVARNVAYLAIRYGPVPHYTRQVKTIEKNGDQSV LYWNYKFPCTRALEACRPDIVLIDKYSKRTHIEVAVSWRGLQEMVD RKVYKYTVNGEYADGSSRGWNIVRELNDQYGFPEVYTVLIGAGGE ILPCTVKDVERLTGGAATDNLIERMERSAVLGSCKRIKRLHAL (SEQ ID NO: 1497)
R4	R4- 1_HG	Hetero dera glycines	.	TGGCGTACTCGGAACCT CCGGGAGCCTGGTAGGA GTTGGCTACAGGTCGCG AAAGTCCTAGGTGCTGC ACGGGTTGCGCTAATCCG AGGGCGCTGGGTACCTT CCCATCGGCCAAAACGTC TGGGCTTCTAGCCGCG GGTCCAGTATTTCTGTGA GCCTGACAGTTCTTCCCGG ATATGGCGAAAGATTACA GGGCGGTATTTCTGTA CCTAAAAWGGTCGGCC GAATGGCACGGACAGCT CACTTCGGAGTGAGCTCG	TGAGGACTCANAAATTGA CAATACACCTCAGA (SEQ ID NO: 1376)		MISCLERETLTQMALFRARSDKTPHAGIPAPDEVREGGCGQNR NPAAPRGKAAAIQRQNGITIPXACAQSGLVRTQRVQQWSAVEESAL KDVVVRNTDDRGILINWAKGVLPEWQRLCQLNPTMYMARSSPSLN KWASLRTHVGPCKPSKEGSPQDLSVKIQPARLAHDTVAELPQ RTVPCGTDGHGVIDSDETETALAEVSRSPFGEREPDLIGATERITRKR LRNAVDRVPPRRKRVPTSPSRKEQDLVPEVDGPAPTDLVTHPTTES EPEMLDPLSLVQLVLPQLGRAMGWAAEEMELGNVMDVELKREF NREVRVGRTPPDQMYKRGAGPPLPKREPERVALLEQLIAARVER GINRGLDWFLNLNVAFAAARVLSRRERVETLADRLHINDSATLSEVS RRRAKAERKLRCAREQPMWMSRRIRXLGVRVERLQLADLVQRRIAGR GNRSYEGPRRRRLRPSLSRVTEAPVNPPLNGNEVYTFWHSLSWAQ SLRANTDDCQLREFKNQLSAAARHTDLTSVGTSSLVQMFSAAALRKM KGKAPGPDGIRAAWWGVFRRIAPVYATWVVRVIRGAEPVANWICN GLTVLLPKSSDNADPSNYRPTCLNTCYKLTAVIAQITASYVDVLGGL PRQQUALRKGVWGTSVSLMIDALTVAARRAKRPLGVCWFDFFKKA			

R4	R4- 1_PH	.		GGGGCATCCGTGTTTAC CCCGTCACCCAGCCGA AGCTGCATAGCGAGCCC GAAGGGGAATGGCCATGG AGACTCCAGCCTCACCTTG TAACTCGAACTAAAGTCCA GGCCCTTTCTGGTGTGG CTCGCACTGGTAGGAAC ACGACAGTCTCGTGAATC CCACAGCACCAGAAAGCC CTAGCCCTCGTAGGCGAA GGCTGCGTTGCTGTTTCA GAAGTGTAGAGWAGTGG GGTTCGGATGGCCGAGTG TACCAACCTGCTTGTAAAC AGGTAGCATAGAGTAATA TGCTAGTAAGCAGGCGAC GAGAAAGGCCGTAAAG GCTCCGGCTATGAAGGAC CTTGCGACACCGCGTGT CTCCACGTGCGGATCTT GAAGCCAGAGTCTTGAC TGCGCGCAGGATGGAGCC TGTGCAACTCCTCCCTCGC TGATCGCAGGAGTGGAGG ATCACCACTCTTTTACCTT GCTAGCTTGGGGTACCAC CTTGAGCTGGGGCCGCCC TTGCTAGCTTGGGGTACG ACCTTTGAGCTGGGGCT GGCCTTACTAGCTTGGCGC GCCACCTTGGAGCTGGG GTGGCGCAAGATCACTTG TATACGGTCTAACCAATAC ATTTGAAAAGCGATCAAA GCGAA (SEQ ID NO: 1253)	GGGGCATCCGTGTTTAC CCCGTCACCCAGCCGA AGCTGCATAGCGAGCCC GAAGGGGAATGGCCATGG AGACTCCAGCCTCACCTTG TAACTCGAACTAAAGTCCA GGCCCTTTCTGGTGTGG CTCGCACTGGTAGGAAC ACGACAGTCTCGTGAATC CCACAGCACCAGAAAGCC CTAGCCCTCGTAGGCGAA GGCTGCGTTGCTGTTTCA GAAGTGTAGAGWAGTGG GGTTCGGATGGCCGAGTG TACCAACCTGCTTGTAAAC AGGTAGCATAGAGTAATA TGCTAGTAAGCAGGCGAC GAGAAAGGCCGTAAAG GCTCCGGCTATGAAGGAC CTTGCGACACCGCGTGT CTCCACGTGCGGATCTT GAAGCCAGAGTCTTGAC TGCGCGCAGGATGGAGCC TGTGCAACTCCTCCCTCGC TGATCGCAGGAGTGGAGG ATCACCACTCTTTTACCTT GCTAGCTTGGGGTACCAC CTTGAGCTGGGGCCGCCC TTGCTAGCTTGGGGTACG ACCTTTGAGCTGGGGCT GGCCTTACTAGCTTGGCGC GCCACCTTGGAGCTGGG GTGGCGCAAGATCACTTG TATACGGTCTAACCAATAC ATTTGAAAAGCGATCAAA GCGAA (SEQ ID NO: 1253)	TGAGCCTTAGGTCGCGG GATGTGACCGGCGGCCA	<p>FDSVPHNLIRVILGILPPVILSVIVMDQWATRLKIGGKVMPTKI PVRTGVFQGDTLSPILFCLSVWPIFALDQFPQYQFRCANHLQQGFS VGHVFMDDLKCYCPDREVLTAIIQQVQKASALGLTIHYKKSALWD QDGGKSGKAVLGVPLVGTGYLGMHERFIMVSKDSLESVRGKFMG RLKTLWTSKLTFGQAMLTGKSCXCMPPVRYVLQNLFLPKSEFNQTRLV LREWDRQIRDLLEDNIRQVFRSKTELYVSREEGWGLPSMEDALEE EVTKLAILVARQETEPFRVCEALERKRCPTPLSLGLQLKDWGVGVE LQGRILLNGNTVGPSPQATRLTGELVLRREAERLSRWRSKVKPGCG MTGGAWRDVPGIDVHLSNRWLKGAALPTVVNSLAIRANTVILRG SGGGYTKGTLRCRCGCGNTGETRRHIVSACSLGRQKGAASRRHDNV CRILVRAICHKLNIIEPPNSANFPVWVLEGSGAKMWIDFPVWPHKIR HTRPDIVLFEWNGVRRLSVIEVAVSDVANMQTHIRKSHRYGTNS TEPFVAGVTPTYRNDCLAAQLRAKFAQQVDVPIIVGTTGETLDGEF GRIKGLPMLTKLQMPRLWSEIQRAVILGSRVILVEHLALPKGGA (SEQ ID NO: 1498)</p>	MKMSHNRDTPSNGVKGTSVRLGTSLVRSVPVGEAGAVRERGTHPSES VSQSDASVNAATGEGSVREQAPLSPPGAEAEATVPTQRRTRHKWSRE
----	-------------	---	--	--	--	--	---	--

			hawaie nsis	ACCAACGCGCCGTGAATCC ATCGGGCGGTATGAGTCT GACACAGGGGGGTGTTTA AGGTGACCTGTTGCGAG GAAATGCGCAGCAAAAGC CGGATGAGCCTTAGAACCA TCGAAGGCCAACGACAAC CTGCCGAAAGACTGGCGGA CTAATCCAGTCAACCTCT GTAGGTCACCGGCTGGTC ATGCTGAATCTCAGCTTAA CCAGGCGTACGAACCTTAC AGTTGGAGGGTCGAGCAC CCCTGATGGCTGAAAAGG ACCATCAAAAGTCGAAGGT AGCCAACGAGAAGACCAA CGGCTGATTAGGCGGAA GAGTCAACTCGTTGAATG CGTTGCACAGCTTGGGGT AGATGGAACCTCTAAGCC CTGAAAGGCAGTCCATCTT CGCAGACGCTAAGGTGCC CCGCCGCTGAGGGTTAT CAGGCCCGCGCGCTGAG GGTTACAGGGCAACC (SEQ ID NO: 1254)	GAGTGTAGAGCTGAACA TCGCTCAACCGATCCAAT TTGGGTCGTGAAATCCCC TCGATAATAATAATA A (SEQ ID NO: 1377)	DRVLWECFVASKREGPGVLRKLQKLWDERGIPGNFPQASLSGQIR QICSKNLLSEERLIQAARMEAQVSPADPARQVTPRVPTRPPRSP VEPARRSPSEETPDLAGVPSEIDSADPNRSPSRGRPHLPAHNMWSQS ESEDVTDPDVSDQQRSDLEPRDLNRSSVESTPGHPNQELSDTLLS NVVPEIDSDDPNQSPRRGRPHLPAHDMSLSDSMDEETEPDLSDDQ RSDLLELDLLNRSSVETTPKGHPSLRHLPEPKIRAAAFRVNSVIGKHT NNITETNALIKAGADLAVRILEVQPRPQRTQKKDPWPWKHREKNIAE IRKHLSEWSEWRRGNLHDEEKKTLLESRYRCLEVLNLEDTLKQRLSA KRKSVRRFEARVAGFHQNLQFNNTQKRLYQTLRGEETSSDSPNAEES IRFWSDIWSKEVRHNNTAEWLHDVKEKNVAADPDLTITSQQLKKQL SKTKNWKAPGPDVMVQGYWIKFTTSLHSRIAQNLHCLQRGTVPWTW MTTGTVLIQDKAKGTEVSNRPITCLPLMWKVLGIIVERYVQVHLD SKKLLPDEQKGCRRNTRGTDHLLVDKLLTKDARSKKNLSMAWVD YKKAFTMVPHSWILECLDIYGIAGNIRNLATTPNWKTLTQSANKH LGEVSIKRGIFQGDLSPLFLVLTMIPLSETLNKAGQGYNYSRTMKLNH LLYMDDLKLYAKSKDQVEQLLNIVHQYSDIKMQFGVSKCGVLNIE GEVTASEGITIEEGTKDIEEAGYKVLGVMYNTILHRTMKDSIRKEYLT RLRLILKSHLNGGNTIKAINTWAVPVVYSAGIINWTKDCTMDIKT RKLMTIYRALHPRSCVDRLYNRREGGRGLISVEDCEAEKRALSQHF RESDDPWARCLVEAKLLKETETADQFKERRRLDRTNKWMSKMSG QYLEAVQDKIVPDSWNWLLRGLKRETEGTILAAEQEALTRYIQNKI DKRNPVSTCRICRSSDETINHVISECGVLAQKEYKRRHDKVARHLHW TLRIHNPVSERWYEHPEPVPVENEAVQIYWDKRMETDRVLHANR PDIVVKDKQESAKLIDISIPFDSRIVDKAEKKEKYRDLAIELQLRWQ MKVDVVPVVGALGAMSKNLKLTALRELKCGHLHPGTLQKSALLGTA HIIRKVL (SEQ ID NO: 1499)
R4	R4- 2_BX	CADV0 10090 48	Bursap helench us xylophil us	GAGGATCCTGGGTTCTCTA CTACCTGCTCATCTCCTC GCGATGGATCCTTGGGGA AGTCTCGGACTGAGCTA AGAGAGCGTTAAAGTAGA GGGTGGCGGCGTAGTGAC TTCCAAGTTGCGATGGGG CGGAACATCTACTCTCTG AGAGAGGGAAGCCCTAT GGCGGCGGTAGAAAGGT GGGCTACGGCAACACTTG CCATGATCAGATTCGATCA	TAAACAAGTGAATAAAAA CCACCCATGCGTAGTAAA CCGATCAATTATCTAGCA AAATCGCAGGTCAGAA ACCAAAGAACCCGCCCA GAGGAATAGGACCACAG CTGAAACTCTCAGATACG CCAAAGGCTCTAATAAAA CGTCGTTAAGTAAAGCAT CGTTAAGTACAAAACAAA AGCACTGTAACGCGCAG GCCCCCTCTTTGCCAAAC	MVFNNCKPKHLCPAIRPTGQQTNGSGSEGTAETAGPSRPVSEDA AQVPVLFEEGEYRAHRDKTCPCYEVWLWIGARSSKARSDSWPLCQLY LMKRNDLRLGRQTRYPLLESSLRAAGIARTKFAIKCIRNVLDRVVPN GPYSEHWKIYRANSGEVPQGATITGKRSARVAGLPSPSQSGHHTKR IQAGTGIEETTIVTETNTTPEVSHHRDPCGEPETAANVDKVTTELTE DGSETRGTANVANGGVSVDPRGRQSSSQNRGNITTNPELVGM WEDMFGVQLDGMARITTEPRPLPKLHLSEPERLWIRAKLEQAWLQ CVSYDVEQW/LNANAVLYAARSPVAASRCPCKEAREAKTWDNKKK DEAKLRRLIGRISSVHSMIPKGDRTPREKKLVKNITKLNTHYPDMDW GGLNHFVKVLSQLKEKISVRVAEHKRVNRAAGQYKGSVAGSAGL APDVVSATAYWSGLAQPGPKFKASSPIFQTKDDVAKNLNTEPVL LYPIIKECIRKPSWKAPGPDGIYNYVWQEFVAQWIQTLVKRTLIDIG

R4	R4-2_HG	ABLA01000389	Heterodera glycines	<p>AAATTAGCCTCTGGGGCT GGCAACCTTACAACGGAT TGTAACCTGAACATATGCTA TGCAAAATGAAAAATAAAA AATGGGGGCTTTACAACTCT AAGGTGTTGGCAGATCAC GAAAACTGCCCCTGCATG AGGGTGAGATAAATCCT ACGGAACAGCCCCTGCTG GACCAACCAAAATCATCCA CAAAATTGAGGGTTTCTT GGTAGTTCCCTTGGCACG TCTTGTTTCATAAGCCAGA ATAAAAACGATACCATACA GACATGAGGCTGGGTACC TGCCAGCCGCGACACGGA AAACCGGTAGGTGCAACC GGAGACAGCTAGGGGAA AGAAAATAGTAAAGTGTG GAAAACAAAACAGGTAGC CCCTGGCTAGAGGGGAATG GGACATTGTCGATTAGG TTGCCCTTGACCAATGAAA GCCAACCGGGTTTACATTA TCGGTCTACCAAGGGCAA TGACCAGAGAAACTGCGA CTATGCCGTAAACCTTCGC AGATTGCCGATGAGAACCC ATCGATCGTAAGTCGAAG CCAAGCGGATTGACCAGT GGAGGGTTATCCCGACAA CAGC (SEQ ID NO: 1255)</p>	<p>TCCGGTAATCCTCGTAGT ACGGTCTTTTCCCGCT TCAGGAATCAGAACGCTG CCAGATCTTGTCCGACAT GGCCGTAGGGTTTGGG AGTACCAGCTGGGGCG GAGAGCTACTGGGTA CACTGCATAATCGGTCT CAGAAAACACCTATGGT TTATTATTCTGTCTCCAT CTGCAGGGTAGCTTTTCG TTGGCCATAGGAGCCT AGGGCAAGAGTGCATG TAGTCTTCAACGGCCATG CCAGGAAACTTGTGAG AGGTGAGGGATAAAGTAG CATCAGATAATATCAGTC ATGAAATTAGTAAACACC AACGTTACCGCTCGTTGG CAAAACACCGACTAACGA TGCTAGTTAGAAAGATC GGGCTCTCCCAAAGTTAG GTGCTTGCAACCGAAGCC GATCCGCTTACCCACAG CTCTGCCAGCGTT (SEQ ID NO: 1378)</p>	<p>RFPTALMCGRTVLLFKSGDKSMPQNYPITCLNGCFKITNAVLTKVIL QRVQDTCALPREQMALPKVWSCMEALQRLDQALQSEIGDDCKTA WIDFSKAYDSLHDALRFVIQTIALPAGMEEYLLKSLDSWRTQLVSLD AGKVSCKPYIKRGVLQGDLSLALFVLTTPSPIVAHLQRTCTGRIQL YMDIKLYGKTESDLCMLIKETQRVANKLGLNLKCALFGSIKQSI AGFDPLGDRTYKYLIGIPQRDVAIDIKQAYDELKAKTVQITGETMACDY LTTRQVINRLNSKIPPVFRFVTQSALCSAPMTRGLYNKITELDNVSRAE LRKVIYKATNVSRFYLATKEGGFGYASLQQVFVEAVVSRAIYCLRAP LCDIREFILSKFDPVKVARIALARSKIDMDIERMDMASATRTIRQHYQ AKWKTLLFQSKLYQKWVQHKIDIPNSSRWLQRGESPRNCRIAVAV QDNTLLCRGFVSGKDPNKQCRLCNAGIETASHIVTECSTHRVHMYIE RHDSVARNIYAVLAKNCGFWIPHYSQKIPTVKITKSVELYWNYPFCT QALEACRPDIVLIDRAKKRILVVEVAVSVYTRLEQMTQRKLYKYGVNG EYQADGETRGNWICRELQKYNMRIDLCIVVIGACGEILPCMVKIEIK ISKVSGRQLLERCQSAVLTGTVTRRHLAN (SEQ ID NO: 1500)</p>	<p>MRKSLFQHIPELSSHIAISVPAARNYPKMCSLAQSGSTLSHNGKV AMHRCPSDDCAGKDPQQRGSQKGNLRSVRWTPSEKAVFEVWSRL EQHAMLNGSEARGTCAITRSQFLIHWDDGERESRSLSDGVPEYPMRT ERAYYERVRLRQRGWQWDCANECLVIGQCAEPCRPKNVVAIKADK GMKRSVLKGLSLPHVMGEINQVSVQVEVPLPSVPASVPQVEGES KGFTEPSNKPSPLEGNPAEEGLRPERVNVVPHGIIISDSERKDLKDRF</p>
----	---------	--------------	---------------------	--	---	---	---

				<p>GCTAGCACTGGAGGGG GGGGTCCAAAGAATACG TGAAGCGGACCCTGTGTT TCTCTGACCTCATACGGGT ATATCCCGATGATATCTAT ACAGTGTTCAITTTTCATTTC TTAACCTGCTTTTCTCTCAA GAGAAATATGACCACATTC GTTTGCCATTTCCGAAGGTG CGACTTTTGAGGGAGTTCT CAGGGACATCAAGCTGTT CATGGACGGCTTGACGGC TCGTGGCGGCACGACGCC GCGCGGGCCAGTTAGGGC TGGTGCCAAATGAGGCAGC GAAGAAGTCTCTGAACG CCAAAAGCAGCGGGAGAA GAGAGAAAGCGGGGA CGGGTTAGAGATAATAGC ATGATTGACGGCGAAGCA ATACTTTGACTTCCAGTGA CGGAGGACGGCGGCAAA AAGCAAAAAGGGGACGA AACCAGTGGCATTGCCAA AGGGCCGAAACCATTTGA GCTGGACGCTGATGGCC GTCTGCCGCGCAAAAGGA GGGGGAGCCCTCCGTC TGGGAAGAAGGTGCGGC GCCATTGTCCAACAAACC GGAATAGATCTTGATGAT CTCAGCAATITGATCATATT GATCTCTTTTCCAGTTGTT GCGATAAATATCGTGCAT TATTTCTCGATTTCTCCAA AGCTTAACCTCTTTCCTAA TACCTACTCATGTATACGT TAGTAGGCATGTTTTATGC</p>	<p>GGAATGCCCGCTAGAG TATATAAGTCCGGAGC ATTGAAGCACAAAGTGAG AACCTGGAAGTATGGT GGTGAGCTGTAACCTC GAATTCCTCTATGGGTG CTTGCGCCGCTAGGTCAT TTGTGATGTAGGGATG GAAGGAATCTCGATAGC GCAAAATCGGATACCGC ACTGGCGATCATGCTGCC AGTGGCCCTGCTGAGTT GGCGGGTTAGCAACCC GTCCAGGGACGCAGATG CCAAAAGCCTGGTGACAC TCTAAAGGGAGTCTACG GCGACGAGAACATCCCT AAGTC (SEQ ID NO: 1379)</p>	<p>WSAYKTAKRSGVFRPALKIEPNRVNRAQWEVLDSCVWEVLKRETSN GYRGCVLRHINAVVAAGYVLAEGNKERRQVIRROSAEWLLRQKSEI NNIRRHIGWITDELRRRTGKNPTRSQLSNFAWLQRYQVGLKPVRE TRDLVQRERLVSRLRAQDNSSMDREERVRKRMPLRRKLEELPG DSKLDTKQARTFWASLIGERKEFGKIPLENWAEVRSKVTDDGGQFA SDHVDQTVWKKILGKARPLKAPGPDGIPNLLWKRLPSANQALFKWL MGKRRQLSVPSWLTGGRVLLPKGGDPVDPANYRPIACLNTQYKLV TGMVTAWVSEHLTTYSILPIEQRAMVSGTWGGCTHAMVIDRAITSYA EATGLPLVGVDFDAKAFDSVSQPWIRYALKVAGVHKRIRCLIGILMK CWSVRVEVFKSGRVLRSAPLAVKNGVLQGDLSPLFLCLSAVAVSSAV GSLDFEVTIPGRGVMQQNHLYFMDDFKGFAPSEASLTRMLVLTLE RTASALGLKINKRKCALVHPREREETGSDIPVLGLRDTYKYLGIERF GIVFEDAWDRVTKMFERMRTLCTEHTFGEIRAAFASTIAPVARYL FLNVIVGGPSWSETLTKAKDMDLRIRLLWERRDNEPGWRFKHCSA DRLYLVRQYGGGLGFVSVDLTSESIYCWAVYVQCRPELELARELFTLN RSARSGIKQSIAGKARKVFRSYALLSKNSAQRVSDLDGDASPGFRVGE MIFMEPTRGARAIVKILRKENDSRRLAAMKGRPMGGRVVSLELDQ VHSYHWILIRARIGRRFRDCIAAQEQQLKARELMCPHINAKAKWCR RCGDGRVETEQLHSGCAWSRTGTMIDRHNHGVVRQVHTALCRKYG LPVSSHVVPLHAVIENEHAKILYDVALHTSPAGVLPREDGGTSYTLRS TRPDMVIFDKKARTILIVEISVPWRENILVKQELIKWRKYAINSMIEPLE LAEAEIPGNLKHALLGAYGTSFTPVKVVPVWVWSCGEVLPNITKRLEL GIPKRGPSLLESIQRAAIIIGSGHVIRAHLSVPRSESET (SEQ ID NO: 1501)</p>
--	--	--	--	--	--	---

R4	R4- 2_SRa	CACX0 10020 06	Strongy loides ratti	AGGTAAGATAGACCTGG TGCCCGGGCGACTTGG GATGGGATGATGGTTGG CGGAGAGTTCTGATGACG CCGTAGTTCGGGAGGAAC AGTTCCTACTAGTGCCAGC CTGGGCTAGTGGAGCTGG TCTGGCGAGTGCTTTGCTC GTCAGATAAGGGGTGGTT GGGGTGGTACCTCTGGAA TGATGTGTCGGATAAGCA CTCTGCCTTATAAAGCTGT CGTTCTGCCCAGGTTTCTC CTAACAGGTTGAACCTGT AATGAGCCTTTGGGTATCT GGTCGGGGTCTGCGCGGA GATAGCTCGTGAGGAGCT GTTATTACATTATTGGTTG AACGCTTTGGGTGCTACT GCACAGATAAAT (SEQ ID NO: 1256)	TAATAGATTATATGTG CCACTTTACCTCAAAGAT ATATTAATACTAACTTAAT TATAATTTATTAATGAATTA TAATATAATAAGTTTTAA AATAAAA (SEQ ID NO: 1380)	MQKFSVPKDSQFLVDSILNKHICSTKNKVKDVKRRSIKTLICAAGLT LRKLVCGKLGNNKYNKINQLWKKERKINCIEDLKHUETNKRHHNFG KRLRNAKVSPSEMLKDYNNKLRKYKNEITACLDEHKKAILRTKFLTPSI KIISNIQNHNDDEEAELPKKEEFVKYKELFTNKDGDDEKETHPLDNWLK KFSKTLIVDWTINDKEILEALKYCGNFKAPGSDMMVMKVCKWFKSAQ NYLRWIKSTWYGEYTIKKDTNAVTFMWIKRDGKPKNDVKSYRPIS CLNCDFKLLNKLANKIYESIEKILPINQMAVINKKHGTCEALLLYKSLVQ SMKFRRTKDVKEIWCSDWIDFSKCYDSISHKCLKMIQSIKAPPIHKLIL DGIDSWNISICNGKNISKTKIPVKSGLQGVASSLYFVLLTGEISYALN KEEQVPIETITPSNTLKNHISFIDDYQLYATSQKKVEKLTIKLREIAEEM NLKLNPKQCGIYGTDDLKGLKRLMLKESLINFPYTSEYKYLGLVENSIDLK DINIQLFKDKILSKYSTIFESRLTHQKRKVFNSTIPSCAAYYLGNLITNK CSIQELLNECKFDQMVRNQLVNQNIKKLQVSNRIYLPKEYNSGLN EIEIEVAANIIRKACYIKKRETLRGVDKLYIAMSNGHRNTLSDALYITKK YSNFQINWNIMGVMVKDQNNILLDAKKIENIKEKRRNLWLEHWKKG NMTYANEAIKKEFHLPDLNIDSKYLMILCYAGSEEQIYNGHVS LVNQS SPSSRLCRKCNKLEETSYHVASVCEFFHKKNLHLMRHNSAVYHIITELCR
----	--------------	----------------------	----------------------------	---	---	---

IMKVCTLRYPEASGIKSGNMKIAAGVKYTFGTAKIYHNKPDVWYT		GTTATAACGATGAAAGAA				
PEVIYVIEVSISSLKNKASQMKMKITARYAVNSTKKLENFAALNNLKG		GGGAAAAAAGATAACG				
ENFVEILSHKANFKRVHFMPLVFCFTGEIPKETMKYLEKLNFSNEKI		TACATACTAGAGATGTCGA				
ASPIARYTGRTLKAHFTN (SEQ ID NO: 1502)		TACCACATATCTAGAACTT				
		GAAGATGCAAGAAAGAAT				
		AAAGTATAAAAAATATTGT				
		GTGAGCTCCATGGTAAAA				
		ATAACGAATGATAATGTTG				
		ATTCGATAGCTCGTGATTT				
		TAACTCTACTAAATTTAATG				
		GAAAGAAAAAGAAAGGTGT				
		AAGATAAAGTTTGGATCAT				
		TTGTATTTGGCTGTTATGG				
		AGAAATATGCTTGACTGAA				
		TCTGCTCTAAGAACACAGA				
		AAATCCTGATTGAATTGG				
		GATTCTCCAAAAAGAGA				
		TGGA TAGCTTGATCAAG				
		AGTGTCTTACAGTTAAAT				
		GAATGAGACAGCAAGAAT				
		TATTATGAAACATCTAGAA				
		GGCGAAAAAGAAATCAAGAT				
		GAAATCAGCCATTAAATAC				
		TTTTATTACCCCAATTATTT				
		ATTGCGTTAACTTGATTAT				
		ATGATATTTATTGTATATTT				
		ATTGTTTAAITTTAGTTTCAT				
		AGTACAAATAAAAGAAATA				
		TACTACGATCTTTACTTTT				
		AAGTTCTACGAGAACGTT				
		GTTTTAGAAATATTTAATA				
		ATATTTCTACTACAATTAA				
		GTAATTAGAGAAATCAA				
		CGAAAGCAGCTAAACTTA				
		CTCGCAAAATTCGTTGATC				
		GAGGCTGGAATGGCACCA				
		ACTAATATACTAACCAACA				
		ATAGAAAAAAGAAAGA				

				ACTTCAGTGGACTTCAAAT AAAAATAAAATTATTGGTAG AATTGTATGATAAACTGA GAAAAATTTGACTAAACA AAAAGATTAGAGCAAATA TGTCATCATTTTCTAATCA TACAAATAAGCGATGAT GACGAAATTGAGAGAAAT TGAAGAGAAAAAGGG AAAATTGTGAAATTAAAT GAGGAAATGAGAAA AACCTGAGGAAAAATTAA ATTTGACAACTATGAAGA AGCAAAATTGAAAAGAGG AATTAAGTGTAAAAAAGA AGTAAACCAATTGTTATT GAAACAAGGATTAGGAA TTTTTGAACCGAAAAAT TATTTCTCAAGTTTGCAA CACAATCAATTACCAAGA AAGAAAGGTAAATCAACC AGA (SEQ ID NO: 1257)	TGAGACCAACCATGCGCA GAGTATCCGAATCAGTGA AAGTCCAAAGTTTCAGGAC AGAAACGTCAGATAAGT CCAAGAGAAACGAGAAA ACAAGTTCAAGTATGCAA GAGTTAATCAATAAGAG AGTACCGTAAATGTATGA CCCCCCTTTGCCAAGT CGACAACTGTCATGCAG GTGTCTCTTTTCAACCG CCATATGGACCAAAACGCT ATCCAGCTCGCTCAGAA GAGCCTTAGGGCTGGGG AGTACCACATGTGGCGG AAACTGAATCTGGATGCG				MISRQADRPVEGQPVTAMSFHNLEPNLNLYPENLRPTGSQDANRG VADIAEVTGPSGLVTNEEAARAPPLVEGEYKRAHCGGGKCHYCRV LWIGARSSKARTDSWNLCLEILINKCMELGNVRRYSPLESSLKEAGI NRTRHAIVKRLAVMRDRFVDNAPYSEHWRLYNACAEENRAVVVPM DSATTVKRTARQAGLESQIGVAGKRVHEAETGTDINAVIETNTT PLEDIDLSPETPEGLAELPSTVEIMELTEDGSRSGTANDADGGVSID PLNRPSSSQESRNVPEQVDPDDELVWESLYGAQLRGAMRTTDRPR LPKLTFSAAEQWLWKSKEKARLECVSIEQQWLRSASVLYATIKTV AACRPYNKAREAHKVWLENKRAEEKRVRIIGRIETVRTMPKGRTD KQIRLARKINRLKRVSFPEMDWHGFLNHFKAFLDLKLLISVRVAEHE RKISRKIAGTYGKSVSQSGFTPDVVAATTFWVSLAQPGPKFKKNSL IFQTWKDSVVENMNTPEVLLHPLIIECMNKPSFKATGPDGFIENSW RQGFIANWVKSIIQRTIQTGEFPASLMCGRTVLLYKNGDTAKPENYR PITCLNGCFKMTNNAVITKVIVQRVQDTCALPGEQMALPKVWACME AQLRDQALQSEIGNDCKTAWIDFSKAYDSLHDHAIRFVIETLALPDG MEKYLKLSLESWTKLVLSNRGKVATGRPYKIKRGVLQGDLSLPAIFVI
R4	R4-3_BX	CADV01008832	Bursaphelenchus xylophilus	GGGTCCTCGGTTCTTACTA CCGTGCTCCACCTCTCGC GATGACCTGGGGTAGG CCTCGGGCTGAGCTAAG CAGAGCATTAAAGTAAAG AGTGACGGCGAGTTGCT TCCAAGTTGCGGTGGGC GGAACATCTACTTCTGA GAGAGGGGAAGCCCTATG GCGGCGTTAGAAAGGTTG GACTGCGGCAACCTAGC CATGATCAGATTCGATCAA AATAGCCTCTGGGGCTGG CGACCTATAACGGATTGT AAACTGAACATGCTAACC TGTCAGTAAAGACAGAAC					

R4	R4-4_BX	.	Bursap helench us xylophil us	GGGGGCTTTGCAATCTAA GGTGTGGCAGACCATA AACTGCCCTTTGATGAG GGTGAGATAACATCTCTAC GGAACAGCCCTGCTGAT CCAAACAAATCATCCACA AATCGGAGGGTTTTCTTG GTAGTTTCCCTTGGACGT CTTGTTTCAAAGCCAGAA TATAAACAAATACCGAAGT ACAGAATGGCTGGAACCT AACAGCCATGGCACAGAA AACGGGTGGTGACACAC GGAGACAGCAGGGGAAA GATTGCAAAGTGCCGTAA ATAAAGATGGTAGCTCTCT GACCTGAGGGAATGGGAC ACTGTCCGATTAAATGCCTT ATCCCGAAGACGGTCAAG GTCTATCTTATCGGTCTC CAAAAGGGCTCTGGCCATA GAAGCTGCGAGGAAGCCG TAACCCCTACGCAGATTGCC GCCGTGC (SEQ ID NO: 1258)	GGGGGCTTTGCAATCTAA GGTGTGGCAGACCATA AACTGCCCTTTGATGAG GGTGAGATAACATCTCTAC GGAACAGCCCTGCTGAT CCAAACAAATCATCCACA AATCGGAGGGTTTTCTTG GTAGTTTCCCTTGGACGT CTTGTTTCAAAGCCAGAA TATAAACAAATACCGAAGT ACAGAATGGCTGGAACCT AACAGCCATGGCACAGAA AACGGGTGGTGACACAC GGAGACAGCAGGGGAAA GATTGCAAAGTGCCGTAA ATAAAGATGGTAGCTCTCT GACCTGAGGGAATGGGAC ACTGTCCGATTAAATGCCTT ATCCCGAAGACGGTCAAG GTCTATCTTATCGGTCTC CAAAAGGGCTCTGGCCATA GAAGCTGCGAGGAAGCCG TAACCCCTACGCAGATTGCC GCCGTGC (SEQ ID NO: 1258)	ATGCATACCGGGAGCGC AGCGAAATCACTTAACGC TGGTGACCTCCTGCTAT CGTAGTACTTCTCTAGATA GATGAGTAGGGTGGGCT AAAGGTAGTCTCTTCAA TGGCAATACCGAGGGGT CTCGTGACAGGTGAGGG ATAACTAGTATCAGATAA TATCAGTCAATGAAATTAG CAACAACCAACGTCAACG TCGTTGGCAAAACACCGA CTAACGATGCTAGTTAGA AAGATCGGGTCTTCCCA AAGTTAGGTCTTGACAC GAAGCCGATCCGCTCTAC CCACAGCTCTGCCACGCG TT (SEQ ID NO: 1381)	ATSPVSHLKRVCPSGRIQLYMDIKLYGKSETELMLIKEVQKVANKL GLQMNLLKCSYTGAGLTESIAGFDPLGDRAYKYLGVQQRSAADTNLA FGELEGKVIRESIETMACEVLTMRQVVTRLSNVGPIVRFVAQSVLTS QAKVSWYNKISDLDSKIRAKLAQTGLRYKSKSNVARLYLSKSKNGIGLV NVQQVLVEALVSRAIYCLRAPSLVEIREHILTAEFDPVGAARTVLRRSRI QLEIERVEMASASAIKTNQARWMTKFTQSKLYQKWVHHIDILAN SNLWLERGEISPOQNARIAVAAQDNTLLCRGFGVGNRESEKQCRMEN MGIECTSHILTECSYHRAHYHIERHDSVARNIYAVLAKDHGLWIPHYS QPVSSTKTPTCELYWNYKFPCTRALEACRPDIVLIDRAKRTLIVEVAV SVVTRLKQMVSRKVYKGVNGEKGADGESRGWNMIRELSEVYNMK VNLCAWVIGASGEVLPCTVKAIQSISSKTSSRQLLERCQRSAVLGSTRV VKRHIAEFH (SEQ ID NO: 1503)	MEILWEDRLKIEDRYGVTLQRSASSLKNQYKVLRLGLPDSGLPWA GVQVNDTGQVWVVDHAEAAATLRGSSPAADVGEAEPPVPLPAAE VWESAADAAPDPQSEIVADQGVETRPVENPPANSRETETETEPVEVEP YLEGQYKFFVKILGKSMWRKPIKYPRRVPETLWRQANELIERSIRQG EVSQSLNCMVYAAAGCAVKSSLDKKDQEAKRRESEWYACRKAIEKAL ERYLNFIDLELKRRSASRPLTSRQRQNLGLVLTIKYGRARVRSVRLSEL QAMLRDALVGIKCMAKRSADKKRQKGFVPIQRYLEPSSAEPRLS DTVRAYVNDIVGSSQQSDSTIQDWSSNLSVPSQELNASKIMGW WRAAVSKSPKNAAGPDGIPGVLWKRFRSASEWVCTWLVRLLQKR RIITPRWLSVGRVLLPKGPLEDPANYPRIACLNTVYKLTSWEMAV REQIQACPGLVPVEQJANRKGWVGCTHASIVDRMITGASREGKGGG FPDLRVLYDCKKAFDSNRDHFVAF/LRVANVNVKVVHLLHTLSQQ WCVRYELRRNNRVERSSPLRVKRGLLQGDTLSPTWFCCLCMAPISASI
----	---------	---	---	--	--	--	--	---

				<p> AAGTCGTAGCCTCTTGCG TGTAGTCAAAGGAATACA TTGCCGTCGGTGACAGG GCTATACCGCGGCGACTAC GGACTTGTTATTACGTAG TGCAAGCTCGTTTAAGACG AATGTGAAAAGAAAGGTGT GATTACTAAGCGTTATGA GTCGGGTTATCTGGAAC TCCGCCCGCGCAATGG CTTCAAGGGTCAAACTGCT AACATTTTAAACCAATATT TGGCAGCAGCAGACTATG ATGTCAGAGTGCCGGGG TCTATCGATATAAGACTGA AAGCGATAGACGTGGAGT GAAAGGATTCCCTTTGTTA AGAGAATCGGTAGAAATTC ACTTTTACTTATTCAAGA ACTTAACAGCAACAAGCA CTCGCGAGGATTACGCC CCAGATTGCGTGCGCGGT ACTTCACCTGCTTTCTTCCA CTTTCGGAATCTGGCATCC TGGCTTTCAGTGTGGTGAT GGCGGCTTGAGTTTCTT GAGTCGTGCGAGTGCCTC ATCTGGGACGTCGGACC GATTGATGGAGTCTGCAG TGGACGAGGACTTGATGG ACCGTAACCATAGTATATC CCTCATGCGTCTTCTCGAC TCGAGGGGGTAGCTTGCA CTACCCACCCCTCTCTCTC CGATTGGGATTTAGACCTA GCCCTCTGGTGTGTCTCGA CCCGCGATATCAGATTCTCT GAATCGACTGTGAGAAAT </p>	<p> GAGCGCAACTTGCTCT TTGGGAGGTTCTCCGG GTTCTTAGGGTTGGCGT ACCATGCTTCGTGTGAGA CCTAGGCCGCTTGACCT AGAGTTAAGCTGTTGCAT GTTTGAGGATGCCTTAGG CGGACGCCATG (SEQ ID NO: 1382) </p>	<p> KTLPNPGPTLRPNMGNRNRGQVAIQVSHVFYFYMDDLKVYCPRVADQ RRMEQNIPLFGEIGLSINASKSAAAAGRYVESELPVLGTDEYKY LGIESGFVWNEVAALDRMQAVLLNRVEAILSVEKHTVQRRDAIRAK AIPGGAYILGHIILSDLDPRGAAERMRLDIEIRRLVKSAGILHDKCSTA RIHLSCEQGGGLAWPSMERAYVAVASVYLLTSQDETISRARDYFVS GRLSNKFVYKHLTSIVDSLGLSVLPDPNGLPTGQPSVLARTIARAD AKLEAQWKETLLTYQRAGRVERADPTVVDHANSYHWLRKAWINEK AYQHAVSVMEGTLLGVNPHGVLTMCRAKAPSASIAHIITGCAELR KSHMKVRHDGVTRWLYNALTEVDGSLPKFHYTQQIPAEIMRGERLTV RYDSDIVTPNKPRHNRPDLVWFDSTRKVIVVEVSVTWLSVLQKQYDN KLNRYAVNSNHEFSIPYPPGVNLANEIRVLPQFTGGVKVFPMIISP TGEVHMQFVPHLAELLENPNIPRIEIKQRSVVLGTDYIIRSYFAM (SEQ ID NO: 1504) </p>
--	--	--	--	--	---	---

		GTCTACGCGCAAAGATCG ACCATTGCCACCCGGCTA TGTGGATCGGGCTTTGA CTGCTTATCTCCGGCTTA ATCGCTTGAGGAAAGGGG GGTGTCGCCCGAAAGGT TAGCGGATCATCCGATTCC TACCGTAAAAACGTAGTTG AAGTAGGATTGAACCTGA GTACCAAGTGAAAAGAGT GGTATAATGCTCCATGGT ATACCCTAGGGATAACCGT GTCGAGCGCACTGCTCAA TACCCTGATTGTTAGTGT AGGTATGCTAGTGGCTG CCTAGGCAGATTGCTCTAT TCACTACGTAAAATCTGGG TGAGTACCATAAGAACCTC CTGTAGGGCCAGGAGTCA AACTAGTCCAGGTTCCGT ATGCTTCGGTATACTCTCT CACGGGCTAGTCACCTAA GAGTTAAGAACCCGCTTTT TCTGCACTGTGAATAGAAA AAGAAAGGGCGGGAGAA ATACGGGCGAGGTAAGCA CGTCGGAACGGGGTGTGC CCAACTCGTACCGTACGCC CGGACAAGGACCATCTTTC GTACCCCGTTCACCCGGCA CAAATCCGATTGTTCTCTCC CGAGAGGCGTCGGCGGA GGTGGCTAACGCCGTCTC TCCAACCATCGGTTTGGGT TTAAGTAACGCCCCAGTG GCCGGTAGCCAAACTGAT GGAGGCAGAAATGCCGGT CCGTTGCTAAATGCGGGA				
--	--	--	--	--	--	--

R4	R4_AL	U2944 5	Ascaris lumbric oides	ACGGACCAAAACGCCGGT GTGGTGATACAGGTGGA AGAGTCTTTGGTCTACGC AAGAAAAGACCAAGGCTCA AGTACGAATACGACCGTCT TCGTGAACGACGCCGTCG AGACAAAGTTCGAAATCCT ACGGTCGATAAGGCC (SEQ ID NO: 1259)	TAGTCGCTAAGGGGTCC GGAATGGTCCGGTCTCT GCGCTACCGGTTCTGGT AGCAGTTCAAGCGCTCA ATCGCTGCTGTAGGC AGTCCATCTGTGGAAGTC GCGCTCTTGATACAGATG TGGACGGATGGAAGCAG ATGATAGAGCGGTGAC GGCCTACTAGCCAAACG C (SEQ ID NO: 1383)	MPCTNSFFERGTPPEPHREPISGTDSSSLGMGTTHRSPLNDDEVIN GPKGHESDPVHVVRAPRTLPHRRLPIGVNNGEASQLRQDSAIAE EAQLESTENHDGRRPPLRGGRKLWSEKEIATLRLCEAYGNRQVCW KEVQKFADFHEERTVAALATKWGALKRPRAPMVGAPPTPDHDPE RGPAGEGGGTTSQENVPTDDPIPANGPTGKESDVRPAVACRTE PEEQLMESDVRPPAVVRLADPEQHTMKSGVKPVALDGSADLEERP EKDIEQMGVDFEGEPFRFAFRKAFYGFRAWVNSFDREPVRVRD CPKFYAYADYLIATIGSSKALGNQSRIGRLNGLVYAAARTIHQFWRE EVGHRQGEKGWYTKATREDLQMLISMMESELARRKEKRPGA KELENIHKLVARLGRSTSGIVRRLEMTQRQLKLEDRLSHEQEKRK RLRQFAETPSLKLTKGAKDRGDTMTVMKSVMDFWRPPIGRRVTS NPDQLQLRDLWRDEQKKAYPADLDLEKADLEEKYEGAIRRIQPWKA PGPDGLHAHWWKALPSAKRLGELVDWLTTGKVTTGWMCMCRGT ILIPKGDGDPNRYRPTCLNTCYKVLTSVMNSVILSHLSRGEALPMN QRAMRKREWGCTHAMVLDRAVMMDAMAQKKSLSVAWLDYRK AYDSVSHEYIRWAINSVNIPRSVQLTKRLMSDWETRETFESTQCRPKLR SDKMKVLINGIFQGDSLSPTLFVLCAPIISYALNKGVGQCQSSGSWSA GYGFEIGHQFYMDDKLKYARTPAMLDSDQIQVSVSEAMGLHLNLS KCAKAHYAPHGAGGAQAEVGAEGSRKGEIPLGLRSTYKYLVEQR LLPMEVALKEFEKFMIDRAETIFASELTWGQMATAYNITAIAGLRVY YSNTNGASPKLLEALKRAATLDRIDLRRHKCRFRNSFVERLYIPREC GGYGLKSVEDTLRESILATWSYATNPHLAGQQYFFERLAARGKRT MADGVKILLDGVPEQVDLKRRTVTVDGIVFEDPTKLHRYLVGKLKA RTEARRRWKEASLAGRLVNDTIDMLSLWMMKGFVSARNLRDA LAVQESLLTRACPALKKGGQEVCRCCHAAPETAHEHITSACRYWLP SLVVERHDSVARNLYVICCRYGITPVHYSNRVSPLENSQCRVLWN MDMQTRTPMKHRKPDIVVFDLKRKILMFEVSHASGLLKQREIKI NRYTVNSEELPDEITTPYPPGNLAADLAATYGWQVEFAPVVVGTGC
----	-------	------------	-----------------------------	--	--	---

R4	R4_H mel	.	Helicon ius melpo mene	ATAAATAATAATAATA ATAATAATAAGCCCTAA AATCCAACCATACGTCGA GTCGAACATCTGATTCTCG TGGGGGCGGACACGTG AAAATAA (SEQ ID NO: 1261)	TAACTATTGTCAGAATT CCTTACTAGTAATAATA TTATCGCTGAAAAATCTCC ACCAATAATTGCTTGGC TATATGCTGCAATTTT GGTTAACGTACCCCAATG ATTTGGGAGAACAAAA TGGTAAACATATAATAAT AATAATTATTAATAT (SEQ ID NO: 1384)	EHVPAVKEDLQRTLDLPHQVEALLERISRS (SEQ ID NO: 1505)
R4	Rex6	.	Takifug u rubripe s	TTCTATGCGCTTATCGCA CTGGATAGGCCAGTGGTT TACGCCGCTGACTTTGGTG CGGAAGGTTGTGCGGTTCG AATCAGGCGGACCCCTTA GGCAAGGCTCTTACGCA	TAGAGGACCGAGTCTG AAGGAAGGAGGCACCGC CCAGAGGGCGAGGAAG AGATTTTTTTATATATA TATATATATATA (SEQ ID NO: 1385)	MSGTDRVIPARTSPGSTRASGVGEPDPVKLATGTRHSWSRAE NWLMECYGNSPSEGYMQRMWEKWLNRPTSSLTQKLLAQCS NIRNKKLSQLEIDARRCASPTVICYKGEGPGRQVSXGVSSPPNI AGYKAPMTDGLGTRAADLERIMKMSWGNSTSLPRLTHKVPDQSL EDMNTALSTIPTTTTETNQLMYAAATVILQMLGYKMSMNSKEQ MAPWRRLEAKIMATRREVSLLTSLSRGVNLRTXPKKYNKLSTEAL ETAKQRLTALATKRYTREVEARRINKVSTNPYKVVYSQWQGNKVM

[illegible]

Table 30: Exemplary monomeric retroviral reverse transcriptases and their RT domain signatures

Name	Accession	Organism	Sequence	RT Signatures
Q4VFZ2_9 GAM R - residues only	Q4VFZ2	Porcine endogenous retrovirus	MGATGQQQYPWTTTRRTVDLGVGRVT HSFLVIPECAPLLGRDLLTKMGAQISF EQGKPEVSANNKPITVLTQLDDEYRL YSPLVKPDQNIQFWLEQFPQAWAETA GMGLAKQVPPQVIQLKASATPVSVRQ YPLSKEAQEGIRPHVQRLIQQGILVPVQ SPWNTPLLPRKPGTNDYRPVQDLRE VNKRVDIHTVPNPYNLLCALPPQRS WYTVLDLKDAFFCLRLHPTSQPLFAFE WRDPGTGRTGQLTWTRLPPQGFKNSPTI FDEALHRDLANFRIQHPQVTLLQYVDD LLAGATKQDCLEGTKALLLESLDLGY RASAKKAQICRREVTYLGYSRLRDGQR WLTEARKKTVVQIPAPTTAKQVREFLG TAGFCRLWIPGFATLAAPLYPLTKEKG EFSWAPEHQKAFDAIKKALLSAPALAL PDVTKPFTLYVDERKGVARGVLTQTL GPWRRPVAYLSKKLDPVAGWPVCLK AIAAVAILVKDADKLTGQNITVIAPH ALENIVRQPPDRWMTNARMTHYQSLL LTERVTFAPPAALNPATLLPEETDEPVT HDCHQLLIEETGVRKDLTDIPLTGEVLT WFTDGSSYVVEGKRMAGAAVVDGTR TIWASSLPEGTSAQKAELMALTQALRL AEGKSINIYTDSTRYAFATAHVHGAIYK QRGLTSAGREIKNKEEILSLLEALHLP KRLAIIHCPGHQKAKDPISRGNMADR VAKQAAQGVNLLPMIETPKAPEPGRQ YTLEDWQEIKKIDQFSETPEGTCYTS GKEILPHKEGLEYYVQIHLRLTHLGTKH LQQLVRTSPYHVLRLPGVADSVVKHC VPCQLVNANPSRIPPGKRLRGSHPGAH WEVDFTEVKPAKYGNKYLLVFVDTF GWVEAYPTKKETSTVVAKKILEEIFPR FGIPKVGSDNGPAFVAQVSQGLAKILG IDWKLHCAYRPQSSGQVERMNRTIKET LTKLTAETGVNDWIALLPFVLFVRNT PGQFGLTPYELLYGGPPPLVEIASVHSA DVLLSQPLFSRLKALEWVRQRAWRQL REAYSGGGDLQIPHRFQVGDSVYVRR HRAGNLETRWKGPYHVLLTTPTAVKV	IPR043502, SSF56672, IPR000477, PF00078, cd03715

			EGISTWIIHASHVKPAPPPDSGWKAEKT ENPLKLRLHRVVPYSVNNFSS (SEQ ID NO: 1559)	
--	--	--	--	--

POL_ SFV1 - residu es only	P23074	Simian foamy virus type 1	MDPLQLLQPLEAEIKGTKLKAHWDSG ATITCVPEAFLEDERPIQTMLIKTIHGEK QQDVYYLTFTKVQGRKVEAEVLASPYD YILLNPSDVPWLMKKPLQLTVLVPLHE YQERLLQQTALPKEQKELLQKLFLKY DALWQHWENQVGHRRIKPHNIATGTL APRPQKQYPINPKAKPSIQIVDDLLKQ GVLIQQNSTMNTVPVYPVPKPDGKWRM VLDYREVNKTIPLIAAQNHQSAGILSSI YRGKYKTTLDLTNGFWAHPITPESYW LTAFTWQGGKYCWTRLPPQGFLNSPAL FTADVVDLLKEIPNVQAYVDDIYISHD DPQEHLEQLEKIFSILLNAGYVVSLLKKS EIAQREVEFLGFNITKEGRGLTDTFKQK LLNITPPKDLKQLQSILGLLNFARNFIPN YSELVKPLYTIVANANGKFISWTDNS NQLQHIISVLNQADNLEERNPETRLIK VNSSPSAGYIRYYNEGSKRPIMYVNYIF SKAEAKFTQTEKLLTMMHKGLIKAMD LAMGQEILVYSPIVSMTKIQRTPPERK ALPVRWITWMTYLEDPRIQFHYDKSLP ELQQIPNVTEDEVIAKTKHPSEFAMVFY TDGSAIKHPDVNKS SHSAGMGIAQVQFI PEYKIVHQWSIPLGDHTAQLAEIAAVE FACKKALKISGPVLIVTDSFYVAESAN KELPYWKSNGFLNNKKKPLRHVSKW KSIAECLQLKPDIIIMHEKGHQPMPTL HTEGNNLADKLATQGSYVVHCNTTPS LDAELDQLLQGHYPPGYPKQYKYTLE ENKLIVERPNGIRIVPPKADREKIISTAH NIAHTGRDATFLKVSSKYWWPNLRKD VVKSIQCKQCLVTNATNLTSPILRPV KPLKPFDFYIDYIGPLPPSNGYLHVLV VVDSTMGFWLYPTKAPSTSATVKAL NMLTSIAIPKVLHSDQGAAFTSSTFAD WAKEKGIQLEFSTPYHPQSSGKVERKN SDIKRLLTKLLIGRPAKWYDLLPVVQL ALNNSYSPSSKYTPHQLLFGVDSNTPF ANSDTLDLSREEELSLLQEIRSSLHQPT SPPASSRSWSPSVGQLVQERVARPASL RPRWHKPTAILEVNVNPRTVIILDHLGN RRTVSVDNLKLTAYQDNGTSNDSGTM ALMEEDESSTSST (SEQ ID NO: 1560)	IPR043502, SSF56672, IPR000477, PF00078
---	--------	------------------------------------	--	--

POL_ MPM V - residu es only	P07572	Mason- Pfizer monkey virus	MGQELSQHERYVEQLKQALKTRGVK VKYADLLKFFDFVKDTCPWFPQEGTID IKRWRRVGDCFQDYNTFGPEKVPVT AFSYWNLIELIDKKEVNPQVMAAVA QTEEILKSNSQTDLTKTSPNPDLDLISL DSDDEGAKSSSLQDKGLSSTKKPKRFP VLLTAQTSKDPEDPNPSEVDWDGLED EAAKYHNPDWPPFLTRPPPYNKATPSA PTVMAVVNPKEELKEKIAQLEEQIKLE ELHQALISKLQKLKTGNETVTHPDTAG GLSRTPHWPGQHHPKGKCCASREKEEQ IPKDIFPVTETVDGQQQAWRHHNGFDF AVIKELKTAASQYGATAPYTLAIVESV ADNWLTPTDWNTLVRAVLSSGGDHL WKSEFFENCRTAKRNQQAGNGWDF DMLTGSGNYSSTDAQM QYDPGLFAQI QAAATKAWRKLPVKGDPGASLTGVK QGPDEPFADFVHRLITTAGRIFGSAEAG VDYVKQLAYENANPACQAAIRPYRKK TDLTG YIRLCSDIGPSYQQGLAMAAAF SGQTVKDFLNNKNKEKGCCCFKCGKK GHFAKNCHEHAHNNAEPKVPGLCPRC KRGKHWANECKSKTDNQGNPIPPHQG NRVEGPAPGPETSLWGSQLCSSQQKQP ISKLTRATPGSAGLDLCSTSHTVLTPEM GPQALSTGIYGPLPNTFGLILGRSSITM KGLQVYPGVIDNDYTGEIKIMAKAVN NIVTVSQGNRIAQLILLPLIETDNKVQQ PYRGQGSFGSSDIYWVQPITCQKPSLTL WLDDKMFTGLIDTGADVTHIKLEDWPP NWPITDTLTNLRGIGQSNNPKQSSKYL TWRDKENNSGLIKPFVIPNLPVNLWGR DLLSQMKIMMCSPNDIVTAQMLAQGY SPGKGLGKKENGILHPIPNQGQSNKKG FGNFLTAAIDILAPQQCAEPITWKSDEP VWVDQWPLTNDKLAAAQQLVQEQL AGHITESSPWNTPIFVIKKKSGKWRL QDLRAVNATMVLMGALQPGLPSPVAI PQGYLKIIIDLKDCFFSIPLHPSDQKRFA FSLPSTNFKEPMQRFQWKVLPQGMAN SPTLCQKYVATAIHKVRHAWKQMYII HYMDDILIAGKDGQQVLQCFDQLKQE LTAAGLHIAPEKVQLQDPYTYLGFELN GPKITNQKAVIRKDKLQTLNDFQKLLG DINWLRPYLKLTTGDLKPLFDTLKGDS DPNSHRSLSKEALASLEKVETAIAEQF	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661
--	--------	-------------------------------------	--	--

			VTHINYSLPLIFLIFNTALTPTGLFWQD NPIMWIHLPASPKKVLLPYYDAIADLII LGRDHSKKYFGIEPSTIIQPYSKSQIDW LMQNTEMWPIACASFVGILDNHYPPN KLIQFCKLHTFVFPQIISKTPLNNALLVF TDGSSTGMAAYTLTDTTIKFQTNLNSA QLVELQALIAVLSAFPNQPLNIYTDSAY LAHSIPLLETVAQIKHISETAKLFLQCQ QLIYNRSIPFYIGHVRAHSGLPPIAQQ NQRADLATKIVASNINTNLESAQNAHT LHHLNAQTLRLMFNIPREQARQIVKQC PICVTYLPVPHLGVNPRGLFPNMIWQM DVTHYSEFGNLKYIHVSIDTFSGFLLAT LQTGETTKHVITHLLHCFSIIGLPKQIKT DNGPGYTSKNFQEFCTLQIKHITGIPY NPQQQGIVERAHLCLKTTIEKIKKGEW YPRKGTPRNILNHALFILNFLNLDDQN KSAADRFWHNNPKKQFAMVKWKDPL DNTWHGPDVPLIWGRGSVCVYSQTYD AARWLPERLVRQVSNNNQSRE (SEQ ID NO: 1561)	
--	--	--	---	--

POL_ MMT VB - residu es only	P03365	Mouse mammary tumor virus	MGVSGSKGQKLFVSVLQRLLSERGLH VKESSAIEFYQFLIKVSPWFPEEGGLNL QDWKRVGREMKRYAAEHGTDSIPKQ AYPWLQLREILTEQSDLVLLSAEAKSV TEEELEEGLTGLLSTSSQEKTYGTRGT AYAEIDTEVDKLSEHIYDEPYEEKEKA DKNEEKDHVRKIKKVVQRKENSEGKR KEKDSKAFLATDWNDDDLSPEDWDD LEEQAAHYHDDDELILPVKRKVVKKK PQALRRKPLPPVGFAGAMAEAREKGD LTFTFPVVMGESDEDDTPVWEPLPLK TLKELQSAVRTMGPSAPYTLQVVDV ASQWLTPSDWHQTARATLSPGDYVL WRTEYEEKSKEMVQKAAGKRKGKVS LDMLLGTGQFLSPSSQIKLSKDVLDV TTNVAVLAWRAIPPPGVKKTVLAGLKQ GNEESYETFISRLEEAVYRMMPRGEGS DILIKQLAWENANSLCQDLIRPIRKTGT IQDYIRACLDASPAVVQGMAYAAAMR GQKYSTFVKQTYGGGKGGQGAEGPV CFSCGKTGHIRKDCKDEKGSKRAPPGL CPRCKKGYHWKSECKSKFDKGNPLP PLETNAENSKNLVKGQSPSPAQKGDG VKGSGLNPEAPPFTIHDLPRTGPSAGL DLSSQKDLILSLEDGVSLVPTLVKGTLP EGTTGLIIGRSSNYKKGLEVLPGVIDSD FQGEIKVMVKAANKNAVIIHKGERIAQL LLLPYLKLNPVKEERGSEFGSTSHV HWVQEISDSRPMLHIYLNRRFLGLLD TGADKTCIAGRDWPANWPIHQTESSLQ GLGMACGVARSSQPLRWQHEDKSGII HPFVIPTLPFTLWGRDIMKDIKVRMT DSPDDSQDLMIGAIESNLFADQISWKS DQPVWLNQWPLKQEKQLQALQQLVTE QLQLGHLEESNSPWNTPVFVIKKKSGK WRLLQDLRAVNATMHDMGALQPGLP SPVAVPKGWEIIIDLQDCFFNIKLHPED CKRFAFSVSPNFKRPYQRFQWKVLPQ GMKNSPTLCQKFVDKAILTVRDKYQD SYIVHYMDDILLAHPSRSIVDEILTSMI QALNKHGLVVSTEKIQKYDNLKYLGT HIQGDSVSYQKLQIRTDKLRTLNDQK LLGNINWIRPFLKLTTGELKPLFEILNG DSNPISTRKLTPEACKALQLMNERLST ARVKRLDLSQPWSLCILKTEYTPTACL WQDGVVEWIHLPHISPKVITPYDIFCTQ	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661
---	--------	------------------------------------	---	--

			<p>LIKGRHRSKELFSKDPDYIVVPYTKVQ FDLLQEKEDWPISLLGFLGEVHFHLP KDPLLTFTLQTAIIFPHMTSTTPLEKGIV IFTDGSANGRSVTYIQGREPIIKENTQN TAQQAEIVAVITAFEEVSQPFNLYTDSK YVTGLFPEIETATLSPRTKIYTELKHLQ RLIHKRQEKFYIGHIRGHTGLPGPLAQQ NAYADSLTRILTALESAQESHALHHQN AAALRFQFHITREQAREIVKLCPCPD WGHAPQLGVNPRGLKPRVLWQMDVT HVSEFGKLKYVHVTVDTYSHFTFATA RTGEATKDVLQHLAQSFAYMGIPQKIK TDNAPAYVSRSIQEFARWKISHVTGIP YNPQGQAIVERTHQNIKAQLNKLQKA GKYYTPHHLLAHALFVLNHNMDNQ GHTAAERHWGPISADPKPMVMWKDL LTGSWKGPDLITAGRGYACVFPQDA ETPIWVPDRFIRPFTERKEATPTPGTAE KTPPRDEKDQQESPKNESPHQREDGL ATSAGVDLRSGGGP (SEQ ID NO: 1562)</p>	
--	--	--	--	--

POL_ MLV MS - residu es only	P03355	Moloney murine leukemia virus	MGQTVTTPLSLTLGHWKDVERIAHNQ SVDVKKRRWVTFCSAEWPTFNVGWP RDGTFNRDLITQVKIKVFSPGPHGHPD QVPYIVTWEALAFDPPPWVKPFVHPKP PPPLPPSAPSLPLEPPRSTPPRSSLYPALT PSLGAKPKPQVLSDSGGPLIDLLTEDPP PYRDPRPPPSDRDGNGGEATPAGEAPD PSPMASRLRGRREPPVADSTTSQAFPL RAGGNGQLQYWPFSSSDLYNWKNNN PSFSEDPGKLTALIESVLITHQPTWDDC QQLLGTLLTGEEKQRVLLLEARKAVRG DDGRPTQLPNEVDAAFLERPDWDYT TQAGRNHLVHYRQLLLAGLQNAGRSP TNLAKVKGITQGPNESPSAFLERLKEA YRRYTPYDPEDPGQETNVSMSFIWQSA PDIGRKLERLEDLKNKTLGDLVREAEK IFNKRETPEEREERIRRETEEKEERRRTE DEQKEKERDRRRHREMSKLLATVVSG QKQDRQGGERRRSQLDRDQCAYCKE KGHWAKDCPKKPRGPRGPRPQTSLLT LDDQGGQGQEPPEPRITLKVGGQPV FLVDTGAQHSVLTQNPGLSDKSAWV QGATGGKRYRWTTDRKVHLATGKVT HSFLHVPDCPYPLLGRDLLTKLKAQIH FEGSGAQVMGPMGQPLQVLTNIEDE HRLHETSKEPDVSLGSTWLSDFPQAW AETGGMGLAVRQAPLIPLKATSTPVS KQYPMSQEARLGKPHIQRLLDQGILV PCQSPWNTPLLPVKKPGTNDYRPVQD LREVNKRVEDIHPTVNPYNLLSGLPPS HQUYTVLDLKDFAFFCLRLHPTSQPLFA FEWRDPGEMGISGQLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYV DDLILLAATSELDCCQGTRALLQTLGN LGYRASAKKAQICQKQVKYLGILLKE GQRWLTEARKETVMGQPTPKTPRQLR EFLGTAGFCRLWIPGFAEMAAPLYPLT KTGTLFNWGPDQQKAYQEIKQALLTA PALGLPDLTKPFELFVDEKQGYAKGVL TQKLGPWRRPVAYLSKKLDPVAAGWP PCLRMVAAIAVLTKDAGKLTMGQPLV ILAPHAVEALVKQPPDRWLSNARMTH YQALLLDTDREVQFGPVVALNPATLLPL PEEGLQHNCLDILAEAHGTRPDLTDQP LPDADHTWYTDGSSLLQEGQRKAGAA VTTETEVIWAKALPAGTSAQRAELIAL	IPR043502, SSF56672, IPR000477, PF00078, cd03715
---	--------	--	---	--

			<p>TQALKMAEGKKLVYTDSTRYAFATA HHGEIYRRRGLLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEAR GNRMADQAARKAAITETPDTSTLLIEN SSPYTSEHFHYTVTDIKDLTKLGAIYDK TKKYWVYQGKPVMPDQFTFELLDFLH QLTHLSFSKMKALLERSHSPYYMLNR DRTLKNITETCKACAQVNASKSAVKQ GTRVRGHRPGTHWEIDFTEIKPGLYGY KYL LVFIDTFSGWIEAFPTKKETAKVV TKKLLEEIFPRFGMPQVLGTDNGPAFV SKVSQTVADLLGIDWKLHCAYRPQSS GQVERMNRTIKETLTKLTLATGSRDW VLLLPLALYRARNTPGPHGLTPYEILY GAPPPLVNFPDPMTRVTNSPSLQAH LQALYLVQHEVWRPLAAAYQEQLDRP VVPHPYRVGDTVWVRRHQTKNLEPR WKGPYTVLLTTPTALKVDGIAAWIHA AHVKAADPGGGPSSRLTWRVQRSQNP LKIRLTREAP (SEQ ID NO: 1563)</p>	
--	--	--	--	--

POL_ HTL1 A - residu es only	P03362	Human T- cell leukemia virus 1	MGQIFSRASAPIRPPRGLAAHHWLNFL QAAYRLEPGPSSYDFHQLKKFLKIALE TPARICPINYSLLASLLPKGYPRVNEIL HILIQTQAQIPSRPAPPPSSPTHDPDS DPQIPPPYVEPTAPQVLPVMHPHGAPP NHRPWQMKDLQAIKQEVSAAPGSPQ FMQTIRLAVQQFDPTAKDLQDLLQYL CSSLVASLHHQQLDLISEAETRGITGY NPLAGPLRVQANNPQQQGLRREYQQ WLAFAALPGSAKDPSWASILQGLEEP YHAFVERLNIALDNGLPEGTPKDILRS LAYSANKECQKLLQARGHTNSPLGD MLRACQWTWPKDKTKVLVVQPKKPPP NQPCFRCGKAGHWSRDCTQPRPPPGP CPLCQDPTHWKRDCPRLKPTIPEPEPEE DALLLDLPADIPHPKNLHRGGGLTSPP TLQQVLPNQDPASILPVIPLDPARRPVI KAQVDTQTSHPKTIEALLDTGADMTV LPIALFSSNTPLKNTSVLGAGGQTQDH FKLTSPLVLRPLPRTTPIVLTSCLVDTK NNWAIIGRDALQQCQGVLYLPEAKRPP VILPIQAPAVLGLEHLPRPPQISQFPLNP ERLQALQHLVRKALEAGHIEPYTGPGN NPVFPVKKANGTWRFIHDRATNSLTI DLSSSSPGPPDLSSLPTTLAHLQTIDLRD AFFQIPLPKQFQPYFAFTVPQQCNYP GTRYAWKVLPPQGFKNSTPLFEMQLAH ILQPIRQAFPQCTILQYMDDILLASPSHE DLLLLSEATMASLISHGLPVSENKTQQ TPGTIKFLGQIISP NHLTYDAVPTVPIRS RWALPELQALLGEIQWVSKGTPTLRQP LHSLYCALQRHTDPRDQIYNPSQVQS LVQLRQALSQNCRSRLVQTLPLLGAIM LTLTGTTTVVFQSKEQWPLVWLHAPL PHTSQCPWGQLLASAVLLLDKYTLQS YGLLCQTIHHNISTQTFNQFIQTSDHPS VPILLHSHRFKNLGAQTGELWNTFLK TAAPLAPVKALMPVFTLSPVIINTAPCL FSDGSTSRAAYILWDKQILSQRSFPLPP PHKSAQRAELLGLLHGLSSARSWRCL NIFLDSKYLYHYLRTLALGTFQGRSSQ APFQALLPRLLSRKVVYLHHVRSHTNL PDPISRLNALTDALLITPVLQLSPAELHS FTHCGQTALTQGATTTEASNILRSCH ACRGGNPQHQMPPRGHIRRGLLPNHIW QGDITHFKYKNTLYRLHVWVDTFSGAI	IPR043502, SSF56672, IPR000477, PF00078
---	--------	---	---	--

			SATQKRKETSSEAISSLLQAIAHLGKPS YINTDNGPAYISQDFLNMCTSLAIRHTT HVPYNPTSSGLVERSNGLKTLLYKYF TDKPDLPMDNALSIALWTINHLNVLTN CHKTRWQLHHSRLQPIPETRSLSNKQ THWYYFKLPGLNSRQWKGPEALQEA AGAALIPVSASSAQWIPWRLLKRAACP RPVGGPADPKEKDLQHHG (SEQ ID NO: 1564)	
--	--	--	--	--

POL_ FOA MV - residu es only	P14350	Human spumaretr ovirus	MNPLQLLQPLPAEIKGTKLLAHWDSG ATITCIPESFLEDEQPIKKTIIKTIHGEK QQNVYYVTFKVKGRKVEAEVIASPYE YILLSPTDVPWLTQQPLQLTILVPLQEY QEKILSKTALPEDQKQQLKTLFVKYDN LWQHWENQVGRKIRPHNIATGDYPP RPQKQYPINPKAKPSIQIVIDDLLKQGV LTPQNSTMNTPVYPVPKPDGRWRMVL DYREVNKTIPLTAAQNQHSAGILATIV RQKYKTTLDLANGFWAHPITPESYWL TAFTWQGKQYCWTRLPPQGFLNSPALF TADVVDLLKEIPNVQVYVDDIYLSHDD PKEHVQQLEKVFQILLQAGYVVSLLKKS EIGQKTVEFLGFNITKEGRGLTDTFKTK LLNITPPKDLKQLQSILGLLNFARNFIPN FAELVQPLYNLIASAKGKYIEWSEENT KQLNMVIEALNTASNLEERLPEQRLVI KVNTSPSAGYVRYYNETGKKPIMYLN YVFSKAELKFSMLEKLLTTMHKALIKA MDLAMGQEILVYSPIVSMTKIQKTPLP ERKALPIRWITWMTYLEDPRIQFHYDK TLPELKHIPDVYTSSQSPVKHPSQYEGV FYTDGSAIKSPDPTKSNNAGMGIVHAT YKPEYQVLNQWSIPLGNHTAQMAEIA AVEFACKKALKIPGPVLVITDSFYVAES ANKELPYWKSNGFVNNKKKPLKHISK WKSIAECLSMKPDITIQHEKGISLQIPVF ILKGNALADKLATQGSYVVNCNTKKP NLDAELDQLLQGHYIKGYPKQYTYFL EDGKVKVSRPEGVKIIPPQSDRQKIVLQ AHNLAHTGREATLLKIANLYWWPNM RKDVVKQLGRCQQCLITNASNKASGPI LRPDRPQKPFDKFFIDYIGPLPPSQGYL YVLVVVDGMTGFTWLYPTKAPSTSAT VKSLNVLTSAIPKVIHSDQGAAFTSST FAEWAKERGIHLEFSTPYHPQSGSKVE RKNSDIKRLLTKLLVGRPTKWYDLLPV VQLALNNTYSPVLKYTPHQLLFGIDSN TPFANQDTLTLTREEELSLLQEIRTSY HPSTPPASSRSWSPVVGQLVQERVARP ASLRPRWHKPSTVLKVLNPRTVVILDH LGNNRTVSDNLKPTSHQNGTTNDTAT MDHLEKNE (SEQ ID NO: 1565)	IPR043502, SSF56672, IPR000477, PF00078
---	--------	------------------------------	---	--

POL_ BLVJ - residu es only	P03361	Bovine leukemia virus	MGNPSYNPPAGISPSDWLNLLQSAQR LNPRPSPDFTDLKNYIHWFKTKKKP WTFTSGGPTSCPPGRFGRVPLVLATLN EVLSNEGGAPGASAPEEQPPPYDPPAIL PIISEGNRNRHRAWALRELQDIKKEIEN KAPGSQVWIQTLRLAILQADPTPADLE QLCQYIASPVDQTAHMTSLTAAIAAAE AANTLQGFNPKTGTLTQQSAQPNAGD LRSQYQNLWLQAGKNLPTRPSAPWSTI VQGPAESSVEFVNRLQISLADNLPDGV PKEPIIDSLSYANANRECQQILQGRGPV AAVGQKLQACAQWAPKNKQPALLVH TPGPKMPGPRQPAPKRPPPGPCYRCLK EGHWARDCPTKATGPPPGPCPICKDPS HWKRDCPTLKSKNKLIIEGGLSAPQTIT PITDSLSEAELECLLSIPLARSRPSVAVY LSGPWLQPSQNQALMLVDTGAENTVL PQNWLVRDYPRIAAVLGAGGVSRNR YNWLQGPLTLALKPEGPFITIPKILVDT SDKWQILGRDVPSRLQASISIPVEVRPP VVGVLDTPPSHIGLEHLPPPPEVPQFPL NLERLQALQDLVHRSLEAGYISPWDGP GNNPVFPVRKPNGAWRFVHDLRATNA LTKPIPALSPGPPDLTAIPTHPPHIICLDL KDAFFQIPVEDRFRFYLSFTLPSPGGLQ PHRRFAWRVLPQGFINSALFERALQE PLRQVSAAFSQSLLVSYMDDILYASPT EEQRSQCYQALAAARLRDLGFQVASEK TSQTPSPVPFLGQMVHEQIVTYQSLPTL QISSPISLHQLQAVLGDQLQWVSRGTPTT RRPLQLLYSSLKRHHDPRAIIQLSPEQL QGIAELRQALSHNARSRYNEQEPLLAY VHLTRAGSTLVLFQKGAQFPLAYYFQTP LTDNQASPWGLLLLLGCQYLQTQALS SYAKPILKYYHNLPKTSLDNWQSSD PRVQELLQLWPQISSQGIQPPGPWKTLI TRAEVFLTPQFSPDPIPAALCLFSDGAT GRGAYCLWKDHLLDFQAVPAPESAQK GELAGLLAGLAAAPPEPVNIWVDSKY LYSLLRTLVLGAWLQPDVPVSYALLYK SLLRHPAIVVGHVRSHSSASHPIASLNN YVDQLLPLETPEQWHKLTHCNSRALS RWPNPRIAWDPRSPATLCETCQKLN TGGGKMRTIQRGWAPNHIWQADITHY KYKQFTYALHFVVDITYSGATHASAKR GLTTQTTEIGLLEAIVHLGRPCKLNTD	IPR043502, SSF56672, IPR000477, PF00078
---	--------	-----------------------------	---	--

			QGANYTSKTFVRFCQQFGVSLSHHVP YNPTSSGLDERTNGLLKLLLSKYHLDE PHLPMTQALSRAWTHNQINLLPILKT RWELHHSPPLAVISEGGETPKGSDKLF LYLLPGQNNRRWLGPLPALVEASGGA LLATDPPVWVPWRLLKAFKCLKNDGP EDAHNRSSDG (SEQ ID NO: 1566)	
--	--	--	--	--

O4189 4_9RE TR - residu es only	O41894	Bovine foamy virus	MPALRPLQVEIKGNHLKGYWDSGAEI TCVPAIYIIIEEQPVGKKLITTIHNEKEHD VYYVEMKIEKRKVQCEVIATALDYVL VAPVDIPWYKPGPLELTIKIDVESQKHT LITESTLSPQGQMRLKKLLDQYQALW QCWENQVGHRRIEPHKIATGALKPRPQ KQYHINPRAKADIQIVIDDLLRQGVLR QQNSEMNTPVYPVPKADGRWRMVLD YREVNKVTPLVATQNCHSASILNTLYR GPYKSTLDLANGFWAHPKPEDYWITA FTWGGKTYCWTVLPQGFLNSPALFTA DVVDILKDIPNVQVYVDDVYVSSATE QEHLDILETIFNRLSTAGYIVSLKKSKL AKETVEFLGFSISQNGRGLTDSYKQKL MDLQPPTTLRQLQSILGLINFARNFLPN FAELVAPLYQLIPKAKGQCIPWTMDHT TQLKTIIQALNSTENLEERRPDVDLIMK VHISNTAGYIRFYNHGGQKPIAYNNAL FTSTELKFTPTEKIMATIHKGLLKALDL SLGKEIHVYSAIASMTKLQKTPLSERK ALSIRWLKWQTYFEDPRIKFHHDATLP DLQNLVPVQQDTGKEMTILPLLHYEAI FYTDGSAIRSPKPNKTHSAGMGIIQAKF EPDFRIVHLWSFPLGDHTAQYAEIAAF EFAIRRATGIRGPVLIVTDSNYVAKSYN EELPYWESNGFVNNKKKTLKHISKWK AIAECKNLKADIIHVIHEPGHQPAEASP HAQGNALADKQAVSGSYKVFSNELKP SLDAELEQVLSTGRPNPQGYPNKYEYK LVNGLCYVDRRGEEGLKIIPPKADRVK LCQLAHDGPGSAHLGRSALLKLQKQK YWWPRMHIDASRIVLNCTVCAQTNST NQKPRPPLVIPHDTKPFQVWYMDYIGP LPPSNGYQHALLVIVDAGTGFTWIYPTK AQTANATVKALTHLTGTAVPKVLHSD QGPFTSSILADWAKDRGIQLEHSAPY HPQSSGKVERKNSEIKRLLTKLLAGRP TKWYPLIPIVQLALNNTPNTRQKYTPH QLMYGADCNLPFENLDTLDTREEQL AVLKEVRDGLLDLYSPSPQTTARSWTP SPGLLVQERVVARPAQLRPKWRKPTPIK KVLNERTVIIDHLGQDKVVSIDNLKPA AHQKLAQTPDSEICPSATPCPPNTSL WYDLDTGTWTCQRCGYQCPDKYHQP QCTWSCEDRCGHRWKECGNCIPQDGS SDDASAVAAVEI (SEQ ID NO: 1567)	IPR043502, SSF56672, IPR000477, PF00078
--	--------	--------------------------	--	--

POL_ MLV BM - residu es only	Q7SVK7	Murine leukemia virus	MGQTVTTPLSLTLEHWGDVQRIASNQ SVG VKRRWVTFCSAEWPTFGVGWP QDGT FNLDIILQVKS KVFSPGPHGHPD QVPYIVTWEAIA YEPPPWVKPFVSPKL SLSPTAPILPSGPSTQPPRSALYPAFTP SIKPRPSKPQVLSDDGGPLIDLLEDPPP YGEQGPSSPDGDGDREEATSTSEIPAPS PMVSRLRGKRDP PAADSTTSRAFPLRL GGNGQLQYWPFSDDL YNWKNNNPSF SEDPGKLTALIESVLTTHQPTWDDCQQ LLGTLTGEEKQRVLL EARKAVRGND GRPTQLPNEVNSAFPLERPDWDYTTPE GRNHLVLYRQLLLAGLQNA GRSPTNL AKVKGITQGP NESPSAFLERLKEAYRR YTPYDPEDPGQETNV SMSFIWQSAPAI GRKLERLEDLKS KTLGDLVREAEKIFN KRETPEEREERIRRETEEKEERRRAGDE QREKERDRRRQREMSKLLATVVTGQR QDRQGGERRRPQLDKDQCA YCKEKG HWAKDCPKKPRGPRGPRQTSLLTLD DQGGQGGEPPEPRITLTVGGQPVTFL VDTGAQHSVLTQNPGLSDRS AWVQG ATGGKRYRWTTDRKVHLATGKVTHSF LHVPDCPYPLLGRDLLTKLKAQIHFE SGAQVVGPKGQPLQVLT LGIEDEYRLH ETSTEPDVSLGSTWLSDFPQAWAETGG MGLAVRQAPLIPLKATSTPVS IQQYPM SHEARLGIKPHIQRLLDQGILVPCQSPW NTPLL PVKKPGTNDYRPVQDLREVNK RVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRD PGMGISGQLTWTRLPQGFKNSPTLFDE ALHRDLADFRIQHPDLILLQYVDDILLA ATSELDCQQGTRALLQTLGDLGYRAS AKKAQICQKQVKYLG YLLREGQRWLT EARKETVMGQPVPKTPRQLREFLGTA GFCRLWIPGFAEMAAPLYPLTKTGTLF SWGPDQQKAYQEIKQALLTAPALGLP DLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRM VAAIAVLTKDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNARMTHYQAM LLDTRVQFGPVVALNPATLLPLPEEG APHDCLEILAETHGTRPDLTDQIPDAD HTWYTDGSSFLQEGQRKAGAAVTET EVIWAGALPAGTSAQRAELIALTQALK	IPR043502, SSF56672, IPR000477, PF00078, cd03715
---	--------	-----------------------------	---	--

			MAEGKRLNVYTDSRYAFATAHIIHGEI YRRRGLLTSEGREIKNKSEILALLKALF LPKRLSIIHCLGHQKGDSAEARGNRLA DQAAREAAIKTPPDTSTLLIEDSTPYTP AYFHYTETDLKKLRDLGATYNQSKGY WVFQGKPVMPDQFVFELLDLHRLTH LGYQMKALLDRGESPYMLNRDKT LQYVADSCTVCAQVNASKAKIGAGVR VRGHRPGTHWEIDFTEVKPGLYGYKY LLVFVDTFSGWVEAFPTKRETARVVS KLEEIFPRFGMPQVLGSDNGPAFTSQ VSQSVADLLGIDWKLHCAYRPQSSGQ VERINRTIKETLTKLTLAAGTRDWVLL LPLALYRARNTPGPHGLTPYEILYGAPP PLVNFHDPDMSELTNPSLQAHQALQ TVQREIWKPLAEAYRDRLDQPVIPHPF RIGDSVWVRRHQTKNLEPRWKGPYTV LLTPTALKVDGISAWIHAHVKAATT 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 myeloblas- tosis- associated virus type 1	RATVLTVALHLAIPKWKPNHTPVWID QWPLPEGKLVALTQLVEKELQLGHIEP SLSCWNTPVFVIRKASGSYRLLHDLRA VNAKLVPFGAVQQGAPVLSALPRGWP LMVLDLKDCFFSIPLAEQDREAFATLP SVNNQAPARRFQWKVLPQGMTCSPTI CQLIVGQILEPLRLKHPSLRMLHYMDD LLAASSHDGLEAAGEEVISTLERAGF TISPDKVQREPGVQYLGYKLGSTYVAP VGLVAEPRIATLWDVQKLVGSLQSVR PALGIPRLMGPFYEQLRGSDPNEARE WNLDMKMAWREIVQLSTTAALERWD PALLEGAVARCEQGAIGVLGQGLSTH PRPCLWLFSTQPTKAFTAWLEVLTLLIT KLRSASVRTFGKEVDILLLPACFREDL PLPEGILLALRGFAGKIRSSDTPSIFDIA RPLHVSLKVRVTDHPVPGPTVFTDASS STHKGVVVWREGPRWEIKEIADLGAS VQQLARAVAMALLWPTPTNVVTD SAFVAKMLLKMGQEGVPSTAAAFIE DALSQRSAMAAVLHVRSHSEVPGFFTE GNDVADSQATFQAYPLREAKDLHTAL HIGPRALSKACNISMQQAREVVQTCPH CNSAPALEAGVNPRGLGPLQIWQTDFT LEPRMAPRSWLAVTVDTASSAIVVTQ HGRVTSVAAQHHWATAIAVLGRP KAI KTDNGSCFTSKSTREWLARWGIAHTT GIPGNSQGMAMVERANRLLKDKIRVL AEGDGFMKRIPTSKQGELLAKAMYAL NHFERGENTKTPIQKHWRPTVLTEGPP VKIRIETGEWEKGWNVLVWGRGYAA VKNRDTDKVIWVPSRKVKPDITQKDE VTKKDEASPLFAGISDWAPWEGEQEG LQEETASNKQERPGE DTPAANES (SEQ ID NO: 1568)	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661
----------------------	--------	--	--	--

POL_ SIVM 1	P05896	Simian immunode ficiency virus	MGARNSVLSGKKADELEKIRLRPGGK KKYMLKHVVWAANELDRFGLAESLL ENKEGCQKILSVLAPLVPTGSENKSL YNTVCVIWCIAHEEKVKHTEEAKQIVQ RHLVMETGTAETMPKTSRPTAPFSGRG GNYPVQQIGGNYTHLPLSPRTLNAWV KLIEEKKFGAEVVSGFQALSEGCLPYDI NQMLNCVGDHQAAMQIIRDIINEEAAD WDLQHPQQAPQQGQLREPSGSDIAGT TSTVEEQIQWMYRQQNPIPVGNIYRRW IQLGLQKCVRMYNPTNILDVKQGPKEP FQSYVDRFYKSLRAEQTDPVKNWMT QTLLIQNANPDCKLVKGLGTNPTLEE MLTACQGVGGPGQKARLMAEALKEA LAPAPIPFAAAQQKGPRKPIKCWNCGK EGHSARQCRAPRRQGCWKCGKMDHV MAKCPNRQAGFFRPWPLGKEAPQFPH GSSASGADANCSPRRTSCGSAKELHAL GQAAERKQREALQGGDRGFAAPQFSL WRRPVVTAHIEGQPVEVLLDTGADDSI VTGIELGPHYTPKIVGGIGGFINTKEYK NVEIEVLGKRIKGTIMTGDTPINIFGRN LLTALGMSLNLPIAKVEPVKSPLKPGK DGPKLKQWPLSKEKIVALREICEKMEK DGQLEEAPPTNPYNTPTFAIKKKDKNK WRMLIDFRELN RV TQDFTEVQLGIPHP AGLAKRK RITVLDIGDAYFSIPLDEEFR QYTAFTLPSVNNAEPGKRYIYKVL PQG WKGSPAIFQYTM R HVLEPFRKANPDV TLVQYMDDILIASDR TDLEHDRVVLQL KELLNSIGFSSPEEK FQKDPFFQWMGY ELWPTKWKLQKIELPQRETWTVNDIQ KLVGVLNWAAQIYPGIKTKHL CRLIRG KMTLTEE VQWTEMAEAEYEENKIILSQ EQEGCYYQESKPLEATVIKSQDNQWS YKIHQEDKILKVGKFAKIKNTH TN GVR LLAHV IQKIGKEAIVIWGQVPKFHLPVE KDVWEQWWTDYWQVTWIPEWDFIST PPLVRLVFNLVKDPIEGEETYYVDGSC SKQSKEGKAGYITDRGKDKVKVLEQT TNQQAELEAFLMALTD SGPKANIIVDS QYVMGIITGCPTESERLVNQIIEEMIK KTEIYVAWVPAHKGIGGNQEIDHLVSQ GIRQVLFLEKIEPAQEEHSKYHSNIKEL VFKFG LPRLVAKQIVDTCDKCHQKGE AIHGQVNSDLGTWQMDCTHLEGKIVI	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659
-------------------	--------	---	--	--

			VAVHVASGFIEAEVIPQETGRQTALFLL KLASRWPITHLHTDNGANFASQEVKM VAWWAGIEHTFGVPYNPQSQGVVEA MNHHLKNQIDRIREQANSVETIVLMAV HCMNFKRRGGIGDMTPAERLINMITTE QEIQFQQSKNSKFKNFRVYYREGRDQL WKGPGELLWKGEHAVILKVGTDIKVV PRRKAKIIKDYGKGKEMDSSSHMEDT GEAREVA (SEQ ID NO: 1569)	
--	--	--	---	--

POL_ RSVP	P03354	Rous sarcoma virus	<p> MEAVIKVISSACKTYCGKTSPSKKEIGA MSLLLQKEGLLMSPSDLYSPGSWDPI AALSQRAMILGKSGELKTWGLVLGAL KAAREEQVTSEQAKFWLGLGGGRVSP PGPECIEKPATERRIDKGEEVGETTVQR DAKMAPEETATPKTVGTSCYHCGTAI GCNCATASAPPPPYVGSGLYPSLAGVG EQQGQGGDTPPGAEQSRAEPGHAGQA PGPALTDWARVREELASTGPPVVAMP VVIKTEGPAWTPLEPKLITRLADTVRT KGLRSPITMAEVEALMSSPLLPHDVTN LMRVILGPAPYALWMDAWGVQLQTVI AAATRDPRHPANGQGRGERTNLNRLK GLADGMVGNPQGQAALLRPGELVAIT ASALQAFREVARLAEPAGPWADIMQG PSESFVDFANRLIKAVEGSDLPPSARAP VIIDCFRQKSQPDIIQLIRTAPSTLTTPG EIIKYVLDRQKTAPLTDQGIAAAMSSAI QPLIMAVVNRERDGTGSGGRARGLC YTCGSPGHYQAQCPKKRKSGNSRERC QLCNGMGHNAKQCRKRDGNQGQRPG KGLSSGPWPGPEPPAVSLAMTMEHKD RPLVRVILTNTGSHPVKQRSVYITALLD SGADITIISEEDWPTDWPVMEAANPQI HGIGGGIPMRKSRDMIELGVINRDGSL ERPLLLFPAVAMVRGSILGRDCLQGLG LRLTNLIGRATVLTVALHLAIPKWKWP DHTPVWIDQWPLPEGKLVALTQLVEK ELQLGHIEPSLSCWNTPVFVIRKASGSY RLLHDLRAVNAKLVPFGAVQQGAPVL SALPRGWPLMVLDLKDCFFSIPLAEQD REAFAFTLPSVNNQAPARRFQWKVLP QGMTCSPTICQLVVGQVLEPLRLKHPS LCMLHYMDDLLEASSHDGLEAAGEE VISTLERAGFTISPDKVQREPGVQYLGY KLGSTYVAPVGLVAEPRIATLWDVQK LVGSLQWLRPALGIPPRLMGPFYEQLR GSDPNEAREWNLDMKMAWREIVRLST TAALERWDPALPLEGAVARCEQGAIG VLGQGLSTHPRPCLWLFSTQPTKAFTA WLEVLTLLITKLRSASAVRTFGKEVDIL LLPACFREDLPLPEGILLALKGFAGKIR SSDTPSIFDIARPLHVSLKVRVTDHPVP GPTVFTDASSSTHKGVVVWREGPRWE IKEIADLGASVQQLERAVAMALLLW PTTPTNVVTDSAFVAKMLLKMGQEGV </p>	<p> IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661 </p>
--------------	--------	--------------------------	---	---

			<p>PSTAAAFILEDALSQRSAMAAVLHVRS HSEVPGFFTEGNDVADSQATFQAYPLR EAKDLHTALHIGPRALSKACNISMQQA REVVQTCPHCNSAPALEAGVNPRLG PLQIWQTDFTLEPRMAPRSWLAVTVD TASSAIVVTQHGRVTSVAVQHHWATA IAVLGRPKAIKTDNGSCFTSKSTREWL ARWGIAHTTGIPGNSQGQAMVERANR LLKDRIRVLAEGDGFMKRIPTSKQGEL LAKAMYALNHFERGENTKTPIQKHWR PTVLTEGPPVKIRIETGEWEKGWNVLV WGRGYAAVKNRDTDKVIWVPSRKVK PDITQKDEVTKKDEASPLFAGISDWIP WEDEQEGLOGETASNKQERPGEDTLA ANES (SEQ ID NO: 1570)</p>	
--	--	--	--	--

POL_ HV2D 2	P15833	Human immunode ficiency virus type 2	MGARGSVLSGKKTDELEKVRLRPGGK KKYMLKHVVWAVNELDRFGLAESLL ESKEGCQKILKVLAPLVPTGSENKSLF NIVCVIFCLHAEKVKDTEEAQKIAQR HLAADTEKMPATNKPTAPPSGGNYPV QQLAGNYVHLPLSPRTLNAWVKLVEE KKFGAEVVPGFQALSEGCTPYDINQML NCVGEHQAAHQIIEIINEEAADWDQQ HPSPGMPAGQLRDPGSDIAGTTSTV EEQIQWMYRAQNPVPVGNIRRWIQL GLQKCVRMYNPTNILDIKQGPKEPFQS YVDRFYKSLRAEQTDAVKNWMTQT LLIQNANPDCKLVKGLGMNPTLEEM LTACQGIGGPGQKARLMAEALKEALT PAPIFAAVQQKAGKRGTVTCWNCCK QGHARQCRAPRRQGCWKCGKTGHI MSKCPERQAGFLRVRTLKGEASQLPH DPSASGSDTICTPDEPSRGHDTSGGDTI CAPCRSSSGDAEKLHADGETTEREPRE TLQGGDRGFAAPQFSLWRRPVVKACIE GQSVVLLDTGVDDSVAGIELGSNYT PKIVGGIGGFINTKEYKDVEIEVVGKRV RATIMTGDTPINIFGRNLTNLGMTLNF PVAKVEPVKVELKPGKDGPKIRQWPLS REKILALKEICEKMEKEGQLEEAPPTNP YNTPTFAIKKKDKNKWRMLIDFRELN KVTQDFTEVNWVFPTRQVAEKRRITVI DVGDAYFSIPLDPNFRQYTAFTLPSVN NAEPGKRYIYKVLPPQGWKGSQSICQYS MRKVLDPFRKANSDVIIIQYMDDILIAS DRSDLEHDRVVSQKELLNDMGFSTPE EKFKDPPFKWMGYELWPKKWKLQK IQLPEKEVWTVNAIQKLVGVLNWAAQ LFPGIKTRHICKLIRGKMTLTEEVDQTE LAEAEQENKIIEQEQEESYYKERVPL EATVQKNLANQWYKIHQGNKVLKV GKYAKVKNTHNTNGVRLLAHVVQKIG KEALVIWGEIPVFHLPVERETWDQWW TDYWQVTWIPEWDFVSTPPLIRLAYNL VKDPLEGRETYTDTGSCNRTSKEGKA GYVTDRGKDKVKVLEQTTNQQAELEA FALALTDSEPQVNIIVDSQYVMGIIAAQ PTETESPIVAKIIEEMIKKEAVYVGWVP AHKGLGGNQEVDFHLSQGIRQVLFLE KIEPAQEEHEKYHGNVKELVHKFGIPQ LVAKQIVNSCDKCCQKGEAIIHGQVNA	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659
-------------------	--------	--	--	--

			<div>DLGTWQMDCTHLEGKIIIVAVHVASGF IEAEVIPQETGRQTALFLLKLASRWPIT HLHTDNGANFTSPSVKMVAWWVGIE QTFGVPYNPQSQGVVEAMNHHLKNQI DRLRDQAVSIETVVLMATHCMNFKRR GGIGDMTPAERLVNMITTEQEIQFFQA KNLKFQNFQVYYREGRDQLWKGPGEL LWKGEGAVIIKVGTEIKVVPRRKAKIIR HYGGGKGLDCSADMEDTRQAREMAQ SD (SEQ ID NO: 1571)</div>	
--	--	--	--	--

POL_ HV1A 2	P03369	Human immunode ficiency virus type 1	MGARASVLSGGELDKWEKIRLRPGGK KKYKLKHIVWASRELERFAVNPGLLET SEGCRQILGQLQPSLQTGSEELRSLYNT VATLYCVHQRIDVKDTKEALEKIEEEQ NKSKKKAQAAAAAGTGNSSQVSQN YPIVQNLQGQMVHQAISPRTLNAWVK VVEEKAFSPEVIPMFSALEGATPQDL NTMLNTVGGHQAAMQMLKETINEEA AEWDRVHPVHAGPIAPGQMREPRGSD IAGTTSTLQEQIGWMTNNPIPVEIYK RWIILGLNKIVRMYSPTSILDIRQGPKEP FRDYVDRFYKTLRAEQASQDVKNWM TETLLVQANPDCKTILKALGPAATLE EMMTACQGVGGPGHKARVLAEAMSQ VTNPANIMMQRGNFRNQRKTVKCFNC GKEGHIAKNCRAPRKKGCWRCGREGH QMKDCTERQANFLREDLAFLQGKARE FSSEQTRANSPTRRELQVWGGENNSLS EAGADRQGTVSFNFQITLWQRPLVTI RIGGQLKEALLDTGADDTVLEEMNLP GKWKPKMIGGIGGFIKVRQYDQIPVEI CGHKAIGTVLVGPTPVNIIGNLLTQIG CTNFPISPIETVPVKLPGMDGPKVKQ WPLTEEKIKALVEICTEMEKEGKISKIG PENPYNTPVFAIKKKDSTKWRKLVDFR ELNKRTQDFWEVQLGIPHPAGLKKKK SVTVLDVGDAYFSVPLDKDFRKYTAF TIPSINNETPGIRYQYNVLPQGWKGSPA IFQSSMTKILEPFRKQNPDIVIYQYMDD LYVGSdleIGQHRTKIEELRQHLLRWG FTTPDKKHQKEPPFLWMGYELHPDKW TVQPIMLPEKDSWTVNDIQKLVGKLN WASQIYAGIKVKQLCKLLRGTKALTE VIPLTEEALELAENREILKEPVHEVYY DPSKDLVAEIQKQGQGWTYQIYQEPF KNLKTGKYARMRGAHTNDVKQLTEA VQKVSTESIVIWGKIPKFKLPIKETWE AWWMEYWQATWIPEWEFVNTPLVK LWYQLEKEPIVGAETFYVDGAANRET KLGKAGYVTDGRQKVVSADTTNQK TELQAIHLALQDSGLEVNIVTDSQYAL GIIQAQPDKSESELVSQIEQLIKKEKVY LAWVPAHKGIGGNEQVDKLVSAGIRK VLFLNGIDKAQEEHEKYHSNWRAMAS DFNLPVVAKEIVASCDKQCQLKGEAM HGQVDCSPGIWQLDCTHLEGKIILVAV	IPR043502, SSF56672, IPR000477, PF00078, cd01645, PF06817, IPR010661, PF06815, IPR010659
-------------------	--------	--	--	--

			<p>HVASGYIEAEVIPAETGQETAYFLLKL AGRWPVKTIHTDNGSNFTSTTVKAAC WWAGIKQEFGIPYNPQSQGVVESMNN ELKKIIGQVRDQAEHLKTAVQMAVFIH NFKRKGGIGGYSAGERIVDIIATDIQTK ELQKQITKIQNFRVYYRDNKDPLWKG PAKLLWKGEHAVVIQDNSDIKVPPRR KAKIIRDYGKQMAGDDCVASRQDED (SEQ ID NO: 1572)</p>	
--	--	--	---	--

POL_ FIVPE	P16088	Feline immunode ficiency virus	KEFGKLEGGASCSPSESNAASSNAICTS NGGETIGFVNYNKVGTTTTLEKRPEILI FVNGYPIKFLLDTGADITILNRRDFQVK NSIENGRQNMIGVGGGKRGTNYINVH LEIRDENYKTQCIFGNVCVLEDNSLIQP LLGRDNMIKFNIRLVMAQISDKIPVVK VKMKDPNKGPKQIKQWPLTNEKIEALTE IVERLEKEGKVKRADSNPNWNTPTVFAI KKKSGKWRMLIDFRELNKLTEKGAEV QLGLPHPAGLQIKKQVTVLDIGDAYFT IPLDPDYAPYTAFTLPRKNNAGPGRRF VWCSLPQGWILSPLIYQSTLDNIIQPFIR QNPQLDIYQYMDDIYIGSNLSKKEHKE KVEELRKLLLWWGFETPEDKLQEPP YTWMGYELHPLTWTIQQKQLDIPEQPT LNELQKLAGKINWASQAIPDLSIKALT NMMRGNQNLNSTRQWTKEARLEVQK AKKAIEEQVQLGYYDPSKELYAKLSLV GPHQISYQVYQKDPEKILWYGKMSRQ KKKAENTCDIALRACYKIREESIIRIGK EPRYEIPTSREAWESNLINSPYLKAPPP EVEYIHAALNIKRALSMIKDAPIPGAET WYIDGGRKLGKAACAAYWTDGKW RVMDLEGSNQKAEIQALLALKAGSE EMNIITDSQYVINIILQQPDMMEGIWQE VLEELEKKTAFIDWVPGHKGIPGNEE VDKLCQTMMIIEGDGILDKRSEDAGYD LLAAKEIHLPLGEVKVIPTGVKMLPK GYWGLIIGKSSIGSKGLDVLGGVIDEG YRGEIGVIMINVSRSITLMERQKIAQL IILPCKHEVLEQGKVVMDSERGDNGY GSTGVFSSWVDRIEEAEINHEKFHSDP QYLRTEFNLPKMVAEEIRRKCPVCRIG EQVGGQLKIGPGIWQMDCTHFDGKIIL VGIHVESGYIWAQIISQETADCTVKAV LQLLSAHNVTELQTDNGPNFKNQKME GVLNYMGVKHKFGIPGNPQSQUALVEN VNHTLKVWIQKFLPETTSLDNALSLAV HSLNFKRRGRIGGMAPYELLAQQESLR IQDYFSAIPQKLQAQWIYYKDQKDKK WKGPMRVEYWGQGSVLLKDEEKGYF LIPRRHIRRVPEPCALPEGDE (SEQ ID NO: 1573)	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659
---------------	--------	---	---	--

POL_ EIAV Y	P03371	Equine infectious anemia virus	<p> TAWTFLKAMQKCSKKREARGSREAPE TNFPDTTEESAQQICCTRDSSDSKSVPR SERNNKGIQCQEGSSRGSPQGQFVGV TYNLEKRPTTIVLINDTPLNVLLDTGAD TSVLTTAHYNRLKYRGRKYQGTGIIGV GGNVETFSTPVTIKKKGRHIKTRMLVA DIPVTILGRDILQDLGAKLVLAQLSKEI KFRKIELKEGTMGPKIPQWPLTKEKLE GAKETVQRLLSEGKISEASDNNPYNPI FVIKKRSGKWRLQLDLRELNKTQVVG TEISRGLPHPGGLICKHMTVLDIGDA YFTIPLDPEFRPYTAFTIPSINHQEPDKR YVWKCLPQGFVLSPIYQKTLQEILQP FRERYPEVQLYQYMDDLFGSNGSKK QHKELIHELRAILQKGFETPDDKLQEV PYSWLG YQLCPENWKVQKMQLDMVK NPTLNDVQKLMGNITWMSSGVPGLTV KHIAATTKGCLELNQKVIWTEEAQKEL EENNEKIKNAQGLQYYNPEEEMLCV EITKNYEATYVIKQSQGILWAGKKIMK ANKGWSTVKNLMLLLQHVATESITRV GKCPTFKVPFTKEQVMWEMQKGWYY SWLPEIVYTHQVHDDWRMKLVEEPT SGITIYTDGGKQNGEGIAAYVTSNGRT KQKRLGPVTHQVAERMAIQMALEDTR DKQVNIVTDSYYCWKNITEGLGLEP QNPWWPIIQNIREKEIVYFAWVPGHKG IYGNQLADEAAKIKEEIMLAYQGTQIK EKRDEDAGFDLCVPYDIMIPVSDTKIIP TDVKIQVPPNSFGWVTGKSSMAKQGL LINGGIIDEGYTGEIQVICTNIGKSNIKLI EGQKFAQLIILQHHSNSRQPWDENKIS QRGDKGFGSTGVFWVENIQEAQDEHE NWH TSPKILARNYKIPLTVAKQITQEC HCTKQSGPAGCVMRSPNHWQADCT HLDNKIILHFVESNSGYIHATLLSKENA LCTSLAILEWARLFSPKSLHTDNGTNF VAEPVVNLLKFLKIAHTTGIPYHPESQG IVERANRTLKEKIQSHRDNTQTLEAAL QLALITCNKGRESMGGQTPWEVFI TNQ AQVIHEKLLLQQAQSSKKFCFYKIPGE HDWKGPTRVLWKG DGAVVVNDEGK GHAVPLTRTKLLIKPN (SEQ ID NO: 1574) </p>	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661, PF06815, IPR010659
-------------------	--------	---	---	--

POL_ BIV29	P19560	Bovine immunode ficiency virus	MKRRELEKKLRKVRVTPQQDKYYTIG NLQWAIRMINLMGIKCVCDEECsAAE VALIITQFSALDLENSPIRGKEEVAIKNT LKVFWSLLAGYKPESTETALGYWEAF TYREREARADKEGEIKSIYPSLTQNTQ NKKQTSNQTNTQSLPAITTQDGTPRFD PDLMKQLKIWSDATERNGVDLHAVNI LGVITANLVQEEIKLLLNSTPKWRLDV QLIESKVREKENAHRTWKQHHPEAPK TDEIIGKGLSSAEQATLISVECRETFRQ WVLQAAMEVAQAKHATPGPINIHQGP KEPYTDFINRLVAALEGMAAPETTKEY LLQHLSIDHANEDCQSILRPLGPNTPME KKLEACRVVGSQKSKMQFLVAAMKE MGIQSPIAVLPHTPEAYASQTSGPEDG RRCYGCCKTGHLKRNCKQQKCYHCG KPGHQARNCRSKNREVLLCPLWAEPP TTEQFSPEQHEFCDPICTPSYIRLDKQPF IKVFIGGRWVKGLVDTGADEVVLKNI HWDRIKGYPGTPIKQIGVNGVNVAKR KTHVEWRFKDKTGIIIDVLFSDTPVNLF GRSLLRSIVTCFTLLVHTEKIEPLPVKV RGPCKVPQWPLTKEKYQALKEIVKD LLAEGKISEAAWDNPYNTPVFVIKKKG TGRWRMLMDFRELNKITVKGQEFSTG LPYPPGIKECEHLTAIDIKDAYFTIPLHE DFRPFTAFSVVPVNREGPIERFQWNVL PQGWVCSPAIIYQTTTQKIIENIKKSHPD VMLYQYMDDLIGSNRDDHKQIVQEI RDKLGSYGFKTPDEKVQEERVKWIGF ELTPKKWRFQPRQLKIKNPLTVNELQQ LVGNCVWVQPEVKIPLYPLTDLLRDKT NLQEKIQLTPEAIKCVVEFNKLDPE WKDRIREGAELVIKIQMVPRGIVFDLL QDGNPIWGGVKGLNYDHSNKIKKILRT MNELNRTVVIMTGREASFLPGSSEDW EAALQKEESLTQIFPVKFYRHSCRWTSI CGPVRENLTYYTDGGKKKGKTAABVY WCEGRTKSKVFPGTNQQAELKAICMA LLDGPPKMNIITDSRYAYEGMREEPET WAREGIWLEIAKILPFKQYVGVGVWVP AHKGIGGNTAEDEGVKKALEQMAPCS PPEAILLKPGKQNLGTGIYMQGLRPQS FLPRADLPVAITGTMVDSELQLQLLNI GTEHIRIQKDEVFMTCFLENIPSATEDH ERWHTSPDILVRQFHLPKRIAKEIVARC	IPR043502, SSF56672, IPR000477, PF00078, PF06817, IPR010661
---------------	--------	---	--	--

			QECKRTTSPVRGTNPRGRFLWQMDN THWNKTIIWVAVETNSGLVEAQVPIEE TALQVALCILQLIQRYTVLHLHSDNGP CFTAHRIENLCKYLGITKTTGIPYNPQS QGVVERAHRDLKDRLAAYQGDCETV EAALSLALVSLNKKRGGIGGHTPYEY LESEHTKYQDQLEQQFSKQKIEKWCY VRNRRKEWKGPYKVLWDGDGAIVIE EEGKTALYPHRHMRFIPPPDSIDIQDGSS (SEQ ID NO: 1575)	
--	--	--	---	--

A0A1 42BK H1_A LV	A0A142B KH1	Avian leukosis and sarcoma virus	TVALHLAIPKWKPDHTPVWIDQWPL PEGKLVALTQLVEKELQLGHIEPSLSC WNTVPFVIRKASGSYRLHDLRAVNA KLVPFGAVQQGAPVLSALPRGWPLMV LDLKDCFFSIPLAEQDREAFATLPSVN NQAPARRFQWKVLPQGMTCSPICQL VVGQVLEPLRLKHPSLRMLHYMDDL LAASSHDGLEAAGEEVISTLERAGFTIS PDKIQREPGVQYLGYKLGSTYVAPVGL VAEPRIATLWDVQKLVGSLQWLRPAL GIPPRLMGPFYEQLRGSDPNEAREWNL DMKMAWREIVQLSTTAALERWDPALP LEGAVARCEQGAIGVLGQGLSTHPRPC LWLFSTQPTKAFTAWLEVLTLLITKLR ASAVRTFGKEVDVLLLPACFREDLPLP EGILLALRGFAGKIRSSDTPSIFDIARPL HVSLKVRVTDHPVPGPTVFTDASSSTH KGVVVWREGPRWEIKEIADLGASVQQ LEARAVAMALLLWPTTPTNVVTDASF VAKMLLKMGQEGVPSTAAAFILEDAL SQRSAMAAVLHVRSHSEVPGFFTEGN DVADSQATFQAYPLREAKDLHTALHI GPRALSKACNISMQQAREVVQTCPHC NSAPALEAGVNPRGLGPLQIWQTDFTL EPRMAPRSWLAVTVATASSAIVVTQH GRVTSVAARHHWATAIAVLGRPKAIK TDNGSCFTSKSTREWLARWGIAHTTGI PGNSQGMAMVERANRLKDKIRVLAE GDGFMKRIPTGKQGELLAKAMYALNH FERGENTKTPIQKHWRTVLTEGPPVKI RIETGEWEKGWNVLVWGRGYAAVKN RDTDKIIWVPSRKVKPDITQKDELTKK DEASPLFAGISDWAPWKGEQEGL (SEQ ID NO: 1576)	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	<i>RT_Rtv</i>	<p>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]</p>
---------	-----	---------------	---

cd03715	CDD	<i>RT_ZFREV_like</i>	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	<i>RVT_1</i>	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]

			<p>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 retroviral elements, such as the retroviruses and bacterial retrons. The enzyme catalysing polymerisation is an RNA-directed DNA-polymerase, or reverse transcriptase (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 polypeptide that contains; an aspartyl protease, a reverse transcriptase, RNase H and integrase. POL polypeptide 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].</p>
IPR000477	InterPro	<i>RT_dom</i>	
IPR043502	InterPro	<i>DNA/RNA polymerase superfamily</i>	<p>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]</p>
SSF56672	Superfamily	<i>DNA/RNA polymerases</i>	<p>This superfamily comprises DNA polymerases and RNA polymerases</p>
PF06817	Pfam	<i>RVT_thumb</i>	<p>This domain is known as the thumb domain. It is composed of a four helix bundle [PMID:1377403].</p>

IPR010661	InterPro	<i>RVT_thumb</i>	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	<i>RVT_connect</i>	This domain is known as the connection domain. This domain lies between the thumb and palm domains [PMID:1377403].
IPR010659	InterPro	<i>RVT_connect</i>	This domain is known as the connection domain. This domain lies between the thumb and palm domains [PMID:1377403].
cd03715	CDD	<i>RT_ZFREV_like</i>	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]

Table 41 provides a listing 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 41, 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 41, 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 41, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or functional fragment thereof.

Table 41. Non-LTR retrotransposase systems with annotation of reverse transcriptase domain signatures.

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
Penelope	Penelope-2_AD		Acropora digitifera	TTTCACTCT	TAATTTATT	MPNLKLTKELRQEIVSYLYPLPSISSNLAFLRAIHIFTKT			
				CTCCAC	CACAACTTC	RFSSHINFISRCLOHKVVPKGFNSNFHASIFSNVSNFLKY			
				CGACGCA	ATGTGACC	LHEIQRQTNTFSHNIMRSTIRAMCSKRNELNKQIECRSQ			
				GCACAC	ATATCTTCA	LSNVCPSVLLRSHAKIQALNSKLFHLEQTKKHLENL			
				AGTTTCTT	CCAATGGC	TNSPKNRTEPLENQNHHTVV7TIPENLPLSNAEKSVLSKG			
				TAGAAAC	AAAGCTCCT	LNFPISKKSDEFTTRQDVEKFLRRVQLKAFFHNKEDKS	IPR0		
				TAGAAAT	CTACACTCT	DNTEKDAFETLTAKKSKWTPPEGQFASIDYFIKKCRHDV	00477	371	610

				TCATTTA CTACTTA ACCGCTG CGACCTC ATTCAAT CCTCGCT ACCGCGA CGCTGAA ACTCAGG	CATATAAA CCACAACA ACCCACAA TTCCTCTAT TCGCTCTGA CGAAGGGC TAAACGCTG AAACGTCA GCTTCTAA ATCTTTCAC GGTGGTAA TTCAACCTT TATCAACTC GTTTGATAA AACCAAAT TTTTGT	HKLKSNCTKLSNLSKEEWTAALNKNRDKLVIAADK GGATVVRWTDLYQQAIRQLSDPTFYTKVNDLTPAN QKIVKDTIQELIRKQELPVT AQNLITTPRTSCIYFKPKIH KPNNGRPVIVSACPTTELSSYLDKVMTPIVKSLSFYIKD SNHALETGNFNGENKIIFTMDITSLYTVIPNNEGLQA LKYFENQRPKPKSSETLLRLAELVLTLNCFSEFGDNYK QINGVAMGTGMGPSYANLFGVFIENKFFSNYHGPDPDL YKRYIDDCVGAT SSTKEELNLFNSVNSFHPALKYTWEL SENSLAFLDIKLSINGNGLSTSVHYKPTDSHNYLLHSSSH POHVKNAPFSQFLRLRLRCSDDTDFNNKCEMCQFFKK RGYPDSA VTTGKHRAQEIDRETALQTSQNEETDRIPFTL TYHPQNLAIKNVILKNFKILSNDPETKHIFSLPPLISFKRD KNLGNFLVRSFAKFNNKPGTFTCKRTRCKTCPFISNTVN ISGPNRSVKVTDHFTCISSNVICYCTCLCKKIYIGETGRR LADRFEHLRDAEQNNNTDASKPVARHFNLPNHSHHNM TICGLYLHHGNTESRKNLEQKLIQGLTSLPHGINERLSF H			
				GCCTCTT CGCTTGA ATCTCTG CTGGCCC ACAGACG TGCAAAA AAAGCCG CTCGTTC CCGCGGT ATTATTA ACCACAA TCTCACC GNCGATA ACCTCAA TCAACAC ATCAATA ACACTAA TGCCAGT ACTAATG ACATACT ATCTAAT GTTACAC ATAACGC	TAATATTGT TACCTCTTA TCATGTTAT TTCATTATT TTTTCACCT TATACTTAT CTTTAGTTT ACACTGTCA ATTATTTT ACTTATTTC CGGTTATTA CAAAATAT GCAACATA AAATCAC GTACCAAT GTAAAATTTC TCATTGTAC CTGAAGAA GGCTGGTTT GGCCAGCC GAAATATA GTACACCTA CAAAAAAG	MRLAEYFFDENSNRTANEHDTPFHNKSTWNPPNDRER ALNTFLDAVKLDITTKPKPTQDNLTVTERQAIGQLKQR QDIIKPADKSGSTVVMDKTWYIDECNRQLTDTKTYQR LDEDITADIQKRVTFYVNRMHKDKLNDKTKQYLIQSD VKPGRFYILPKVHKPGNPGRPVIVSSNSHPTERLSHFVDY HLQPLVHKLPSFVKDNTDFLNKLLTIGNLPANSLLTLD VSSLYTNIPHNEGINACERFLRTSSHKTPTSTLCDLIRMI LTMNNSFENDNHYLQIHGTAMGTKMAPSYANLFLGFPE ANALKNAPFQHTWLRYYDDIFMIWTEGLDNLKIFIDYL NNIHSITKFTSSHSTNIPFLDVNVSLENDGNISTDLYTKP TDKHQHLLYSSCHPLHTKKAIPESLALRLRICSTDA TFH TRTAQLATYLLKRGYNRNFVNKQIRRAADIPRQLTLQT KDINKPKRIPFTTFNPSLPHISHIHKHFNLLSSNRCKSV FQHPVVAFRRSPNLRDLLVTAKLSPNSTNPQLPCGSFR CGKNCATCPYISHGLTJTYTFSTGETRPIKFNTCETKNL IYMIQNRCLQYIGETKRRLKDRFNEHRRITIDNPNNKS KPTTAAEHFLSSPNHTANDMILPIEKIFSNRDSIRKAREA FLIQKGRITIDPDGLNIREETY	IPRO 00477	143	383
Pene lope	Penel ope- l_AD i	Acropora digitifera	.						

[illegible]

					CTGTTTAGT CTTTCCAGC CAATTACAT CTAGGCTCA ACTGACAG TCGACCAAT AACATCAC GAATAAGT TGACCAATT ACATCTACG ACCGGAGT TTTTATAGT ATCTACTGA CGTTACACA AATCACTTG ACTCTGAA GATGACTTC CGCTCAGGT TGTCGAAA CGTCAGTCA ATGTCATCT CAACACAGT CCTTCTCAG GACTACACT CACCCGGA CGATCGTAC TTTAC	DDHKLVSFDVKSLFTSIPLQLALDCTETAINNSTLQLPLL TNDLMDLLNLCLTSTYQYNGQHYKQLHGTAMGSPVS VVVAEIVMQNIEEQALASYKRTIPLWLYVDDTFTTLH KDEIDDFHEHLNRQNAHIQFTKEIEDNGKIPFLDCLVIRD NNRLQTTVYRKPTHDRLLDESSYNPSSHKATTIRLTR RALLVCNSHDSLADERKYLDNVFSKNYNCDVTRNT YRTEPNETNLTPTTTVTIPYIKGTSEIARIILQPYNIRVA HRPITTLRKLLTNVKDKDQPRDRQGA VYKIKCCDCQAT YIGETGRNLNTRLTEHRRATRNGDINNIAEHHLQTNHR IDWDSATCVTYNNYQRIVLESWFTNLEQTPINRCLQL PAPYKRLIDDINRQTTH			
I	I_Ele 14	.	Aedes aegypti	CAGTTGA CAGCCGA CCGCCGT CGCGTAC AGTCGCC TTTTCGA TTGTTC GTACTGT TAATATC AATTAAC AGCGCAT CAGCCAA CTCGATC	TGATCCTGC CCAATATG ACGACGAC ATTGATGCC AGCCACG GATACGAA ACTGTTTA CTAAATGT ACTATGTAT AGATGTAA GCTTGAAT GTTAGATTG ACAGTTATC	MPGALGGAPLESATRTLCPPNGFSQTYFSTAAPSLPV VGSTPGPTLGGALSGSVSTFELGSTNSKETLSFFVPAL PTSHNFILQWNLNGYRSRLCDLELIQNNQPWILALQET NNISADEMGRITLGGQYAWTVKRLANQRHSAAGILKGI PHKILDLPDLPIVGIQLLGSVSVACAYLPGGNIPNLH GQIQNCLQALPEPRIFLGDLSHHPVWGSPKADKRGSTL INAFEEEDMLJLNDGSPTEFFNGHFSSEIDVTAVLSFARNI QWSVSSDLHGSDDHPRIALTGTASIPKNTRRPRWKYDK ADWAEYDLLIRSLRNPPDNMADFISTIVESAKATIPRT SSTPGRKALRWSDETKTA VKARRKALRQLRRLPLDH PDRQSTLENYRKLHLECRNIIGDAKKA TWESFLES MNAS QTTSELWNRVNALNGRKTTPLTQVQGN AISDPEA VA SELGKYFASLSAIESYETHFXNSVKPSTSAINNFIIPNTA	cd016 50	575	833

GTGTCT	TCCTTTTCA	DPMEQPFSA RELSFALSRNNGRSAGPDEVGYPLLNLP
GACGGG	TCGGTTTCA	SVGKLALLDLYNKIWNNDPPAWKESLVVPIPKANIPT
TGCTCCA	TCCTTCTGCT	RDPTKFRPISLTCCAGKVMERMVNRRRLRYHLEANGLLD
CAGCGG	GACCGTATT	FRQHAFRQGYGTSTYFAGLSDLVLEAYDKGNHAEVISL
GCGAGTG	TCTCCAGGG	DISKAFNRWTPLVLEQLAAWGIGGRTLHFVRNFLADR
TGCATTG	TCTCAGCCT	TFQVVGNTKSTTPEETGVPOGSVLAVTLFLAMNGVF
TTTTATTG	TCTGCTGAG	SRLPKNIYVFVYADDIVVCGSTPTMTIRIAQTAVKSV
CGGATCA	CCTTCTCCC	AKWASDNGFQLSASKSIRCHICPSGHRITGPDITIDGQPIP
GTGCAGC	GTGTTGAAC	LRKTVRLGVIVDRALSRQHFDTVKSACRSRLNLIKSI
ACCCAGT	TAGCATAAT	RPHRSNNREIRFRVRAIVDSRLVYGLELTSIAMDRIVE
GGTGTGG	GTAAAAA	VLSPYNSYIRIISGLLPSTPSDSACVEAGLLPRIFIRSSIC
TGAGAAAG	ACACGTTA	CKTAAFLSKTAGEDRVFLDDEGNRALSTAANLSLPPVTR
TGAGGTT	ATAAAGAT	VHWLGDRSWRSVPPKIDNKIKNSFSAGSNSAALRRSV A
AAATAGA	GAAAAAAA	ELLQTSYSDFAALRYSDGSLTSAGVGIGVAGDIPDVMSL
CTCGGTA	AAA	PSQCSVFSAEAAAFAIAATTPSDRSILVLTDSASVISALQS
GCGTGGT		DSPHPWQIAYALPDTVFTWPGHCGVPGNEMADH
TCGAGTA		LARSLSGQRYTSEVPFMDLKRWKSTFRQHWEDSWY
CTGTATA		RTRTLFLRKIKNSTTTWSDLPILKQDRILSRLRTGHTLISH
KTATTAT		NMGGGPFHKECESCHIPASVDHVLCACPMYEHLRQIHG
TTCGGGC		LSDNIGEVLRDDATTIAALLSFLRDANLYSSI
GCGAGTG		
TATGTGT		
TAGTAAC		
TTGTGAG		
TGGTATA		
AATAAAA		
TTATCTTC		
AAATTAA		
AAAGCGG		
TTCAGTG		
GTTTICA		
ACTGTAG		
TGTGACT		
GACCAAC		
CAAAACA		
GGTGAGT		
CTCCAC		
GAGTGCC		
TCCTTCC		
GGCCCGC		

CGGCGCC
ATGCTGG
CGAACGG
CCATCCC
CTCCACC
CAGGGGA
CCCAGGG
GGCCCCG
GAGGAGG
ATTGGCC
AAAGTAG
GCAACAG
AGTCTC
GGCGACT
ACACCGG
TGCATTG
CTACCGG
AGTTTAT
GGACCGT
GAAAGTA
CTGCAGG
ACAACTG
CAGTATC
TGAAGAT
TGAAGAT
ACTGCTG
GAACCAT
GCCCCAG
GATCCGT
TTCTTCTT
CCGTCGA
AAAGGCC
ATTGGAG
GTCCCAT
TTTCGGG
GCCTTCA
AGGAAA
TAAAGGC
ATGTCCT

ACGCTTT
GAAGGTG
CGCAGTC
AAGACCA
GGTMAAT
AAGCTCC
TCCGCAT
GAGCCA
CTGCAAG
ATAAMAC
CCCAATC
CGGATCG
CTGAGCA
CCCGCGG
CTMAATC
AAAGGGA
ATGTGTG
GTGACCA
ACCATGA
CACAAATC
GGACTAT
CGGATGA
CTACCTG
CTTCAAC
AGCTCTC
CTCCACG
GGAATCA
AGGGMST
GCGAAGG
ATCACGC
GACGCGG
ACCGCA
TGCGCAA
AGCGGTG
GTAAACA
CGGCAC
CATCGTA
CTACCA
TCAAGGG
CACCGTT

ATTCCGG
AGCATGT
CGTTTC
GGTTGGT
CAGGTCG
CAAGACG
ACGTCTT
TTCTACC
CATCTCC
CATGCTA
TGCTTCC
GCTGTGT
GGAGTAT
GGGCATA
CCGGTAA
GCGTTGC
ACAAGCC
CCCAGCA
CAGACTG
CGGACGC
TGCTGCG
AGGTACA
CCCGGAA
CAACAAA
CCCCCGA
TCTTACA
AACACCC
AGGAAGG
AGCCGCC
AGTGCCC
ACGTGTG
GTGGA
GTGCGAA
ACGCGGT
ACCAAGT
CACCGCA
GAGCCGT
TGCTGCA
AGACCCAG
TGCAAAA
CAGGAGA

CAGCACT
CAGTCTC
AAGCCGC
CAGTGTG
CGGTGTA
CWVGAA
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
TCCGGG
ACGGGA
AGAAAAG
AGGATGG
CTGAAAT
GGAACGG
GCTCTGC
CAGCGTC

AGCGACC
GCCTCGA
GGCGGTT
AAAGAAC
ACGGCAC
CATCGGA
GAGCTGG
TCAAGCA
AGTGGCA
ACCTTAA
GGTTACG
AACCTCC
AAAAATG
TCTAGCA
GTAAGGA
CAAAACAA
ATCCCGG
ATCAAAG
TCGCACC
ATCGCTG
AGCTTCT
AGGCAGG
CAATCAG
TCTCCGT
CCAACGA
CAAAACC
GCGACAA
GCTTTGG
CTTACCG
CGCAAGA
SGATCAA
ACGCAA
CCGCCCC
AAGTCGC
CAGAGGT
TCCCGTC
ATGAAAC
CGCCAC
TCTGAAC

CCCAAAG
TGGACGT
ACAGGTG
TCCAAAT
GGATTTC
GAATATA
ACACCG
ACCTCAA
GAAACTC
TCACAGG
CGATGAA
ACCAACG
ATGGAAG
TGCGTCT
GATCATA
GCATGAT
GTACAGC
GATACCG
AAGCCAC
TGGAAAT
TCCAACA
CGTCTCA
CGTTTCG
ACACCTA
CGAAAGA
AATGCAC
TCAACCT
CGGATCA
CAGTTGC
CCATCCG
GCGATTG
AAATGCC
TCATCCA
CCAACTC
GAAGCGG
AAAAACA
AGAACCG
CAAAAGA
GGTGGTA
GAACCAA

ACAGTAA
AACAGTA
ATTACG
CTCAGTC
CTTGTCA
AGCACAT
GGTCTAA
ACAATTC
GACAATA
TTTCTTC
CGTCACT
CTTCAC
CCGCAAA
TTAGCGT
TGAGACA
GTCAAAT
CACAAGA
GTGACT
GAGCAAA
TACCGAT
GAACACA
CTACAGC
TCCGAA
CGAAGTA
CTCAGG
ACAACTC
AGCGACA
ACAGGGG
TCACGGA
GAGCCGG
GGCCCCG
TCAGTGC
GGACGCT
ACACCCC
AACCGGA
ACTGGCG
GACAACC
CCCGGA
TCCGGGC
GTAAC TG

TTGTTATC
GGGACGC
AAAAGGG
ACACAGT
GCCTATC
CCCCGAA
AGTCCCA
ACTAGCT
GTATGGG
TACGCCA
GGGCCCC
GTCGGTG
CGGAAGT
TTTGGAA
GAACCGG
AACCGGC
GGACAAC
CTCTGGC
GCCCAT
TACGGTC
TGGAAGA
GGGACGC
AAAAGGG
ATCAATA
TCCTATC
CCCTKAG
GACAGCG
TGGGAAG
CAAGCCA
CAGGAAA
CTCTTTT
CTGCCC
TGCCAAC
ACACAGA
GGGTGAA
GGTCGGG
ACGCCA
ATACCC
GAAGGAC
AACGAGG

GAAGCAA
GCCTCCG
TCGAGTC
GAAACCC
CCCCGGA
ASAGAAA
ACCGCG
GACCACT
ACACCTC
CAAACTGG
CACAAACC
ACCCCGG
TCAGCGA
CAAGCTG
AATCCCC
CGACTTC
CCGGGCA
GCTTGG
GACCCGG
AAGCAAG
CGACGCC
AGGTGTG
ATGACCT
GTTACAC
AGCGTC
GTTACAA
TCGCTCA
GCCCGCC
GACTTCA
GCCCGGG
TATTCGG
GAACCTG
TTCCAGA
CTTCAAC
TCCAAAC
GCGGAAG
CATACCA
AAACCGA
TCGCCAT
CCCGGAT

AATKATA
GCTCAAC
CCGCTGC
TGGCCCC
TCAAGCC
AAACAGC
CTCCCCA
GATCCCT
ATCAGGA
TCGCCCT
GGGACCC
TCGGACT
CAACAAAT
CCGAGCG
GTAAGGC
CGACCCCT
TTCCCGC
TGGGCTT
CCACCCG
GTGAATC
TTCGTCC
GCCACAA
CAACCTA
TGTC AAC
CGATTTT
CCAGCGG
GCTCCCA
GTCTTCTT
CGCTCTC
AAACTCA
CAGCAGC
ACTCCTT
AATCAAG
GATAAGG
TTTCTTA
GCAGCCC
TGTGGGA
AACGACT
GTCTTGC
CGCAGCT
CCCCAAG

RTE	RTE _E	Aedes aegypti	TTCGGC AGTAGTA GGTCCC	GGGCTGC AAAAATGG TGAGTCG ACGGAGT CCTCAC AGTTCCT ACCTCAT GCTTCCA CGGGTCG ACGGATG ACAAATG ACGCCA GCTAAGA GTTGCGT GCTTAGC TGGTAGT GCAGCCT GGGTCTT GTAGTCC TTCGAC TTCAGCT AGATTGA GGAGGTA CGTCCAC CGAGCGT CTGTTC CCAAAGG GGTGGG CTCAAC ACGCTCT GTCTGGC ATCCAGC GGCTGAA TAAGAAA TGCTCTTC CCCGGAA GCTACAC	cd016 50	503	762
				<p> MKLEILGLSEVRWPNFGEHRIPSGQILLYSGLRGEHAPR HRGVGFLLSAQAYAAALMKWEPINERIIVARFRTRVRNL TTVQCYAPTDAAELQDKENFYSQLNAVVDKIPKGDIIKY MGDENAKIGSDNSSYEPVMGRHGLGEMSENCELLAEFC GNNDMVIIGGSLFPHRPVHKVTWVSRDGFTEQIDHICIS RKWKRSLLDVRNKRSDIASDHHLLIGEVRLRIARRQ EERVGRFRNTRLEDA TVKRSFVEELETXAANISEGGSV EEQWTAIKNAFTTSENNLGELRTRRKQWITDETWRKIE ERREAKAAIERAKTRGAKALSRQYSALEREVKRSCRR DKRAWADSLADEGEKAAA TGDIRLLYDISRRLSGAKM NTTMPVKDTTGQLLTDPTDQLKRWFEHFEQLFQVSTRP PTPQHDPPRVRRISRNVNTEAPSLHEIETAIQSMKSNKAPG VDRISAEMLKADPAISAOQLLHQLFRNIWDTATFPADWM QGVLVKVPKKGDLTVCDNWRGIMLLCIVLKVLCKVILN RIQEKIDATLRRQAGFRAGRSCVDHIVTLRIILEQVNEF QESLYMVFIDYKAFDRNLNHNMMWGALRRKKGVPKEIIG LIEAQYEAFTRCVLHNGVLLSEPIRVVAGVRQGCILSPLF LIVIDEILIGAIQREQNRGLLWQPTIMEHLNLDLADDAIAI LAQRSDMQSKLDDLAEERSSAAGLNINVNKTKSLDVNT AGPSSFTVAGQTVENVENFQYLGSQLASDGGTKIDIGAR IKKARAAAFASLRNIWKNSQISQRTKIRFNSNVKSVLLYA SETWCVSAENTQRLQVFINRCLRIIRAAWWPHNWISNA DLHRRCHQOPIQTEIRQKRWGWVGHTRLRGGNEICKQA LEWNPAGHRSRGRPRGWSRRSLNKEIKEYDGNLTWPQ VKAIAGNRPGWKSFISALCTRMGAODP </p>			

485

AGGTATA	ACATGACA	KIHKPGNEMREIISAEGSPTHKLAKWL V KEFQSMKPFE
CTCCTGT	GGTATACTC	SRSVNNNTQEFASQLQASGYIQDDEIMV SFDV KALFPSVP
CGAAATT	CTGTCGAA	VKEAINLLEEWSLLNQHNESNWK NKVRYIKLTRL CME
TTTTCTTA	AAATTTTTC	ENYFSFRGNFYKQTKGAPMGNPLSPFLSELFMANFESIL
GGTAGTT	CTAGGTAGT	KKKDLLPQRWWRYVDDVFSIIKKDYL PKILEAINSIHKD
TGATTGT	TTGATTGTG	IKFTHEEEKDNKLPFLDLITRESSHFEFEIYRKPTNTMR
GTAGGTA	TAGGTACTT	VIPSTSNHTYQHKMAAFHHMIHRMQTLPLSETGKSKEL
CTTATCT	ATCTGGTGT	DYIFETARLNGYRNSTIQA VVDKRAKDYRKSM TTTLTA
GGTGTTA	TAGTATTTA	EKEDLKRVA VNFDTKITQPLAKKFRKYGV DLVFSRSS
GTATTTA	AGGTAGAG	QLKSRLGSTKDKINKLNRAGVYKISCPCNCNKIYIGQTKR
AGGTAGA	AGATGCGA	TLEIRFKEHIAEVKKADKELEKGLAYDFKSKV AEHV FQ
GAGATGC	CGATTTCTT	DGHQMTSDNIDIVRSVSSSWKLDVAESLEIYKQKPSLLL
GACGATT	CAGTGGAT	NRDQGNQGSRLFKFV PKKYKRA
TCTTCAG	ACGAGTAA	
TGGATAC	CTGACACTG	
GAGTAAC	AAGAAAGC	
TGACACT	TGCAAGTG	
GAAGAAAG	GTAGTCGA	
ACTGCAA	AATACGCG	
GTGGTAG	TATCTGTCA	
TCGAAAT	AAGATAAG	
ACGCGTA	CATTTAGGA	
TCTGTCA	GCGGAATT	
AAGATAA	AAAAGGTA	
GCATTTA	CACGGTACT	
GGAGCGG	CCATCTACT	
AATTAAA	TTAAAAGTT	
AGGTACA	CTCTATTTCG	
CGGTACT	AGATTCTGC	
CCATCTA	TCAGAGGA	
CTTTAAA	TTCGAACAT	
GTTTCTCT	TCATTAAG	
ATTTGAG	AG	
ATTCTGC		
TCAGAGG		
ATTCGAA		
CATTCAT		
TAAAGAG		
TTGTA		

CR1	CR1-61_Ae	.	Aedes aegypti	TT	TAATGTATC CTGTTAAGT GTTGTCCAA GAGTTGAT GCTTAGTTT AGTTTAGTT TTAATTTTA TTCATTAAG ACCATGAT GTCGGATG AATCTTCAA TACAAATA CAAATACA AATACAAA TAC	LDNFAFRGLKQRNLLRNHLDRITLDLIFTNDAITDCSSVD EACQPIVDIDPHHPAFVLTVRSPTVPFVEVSDEEVRDFG KVDYETLHVLLSRVDWSTVYQCNDVDNNAVSIENGKLIH ECFTHCVPLAMPRRKPAWNSARLLYLRRNQSKMQRRF SKRNDCKLAFNRASRSYRYNRYLYGKYVRRVQDN LRRHPKQFWSFVNSKRNETGLPSFLSLNGHSASSAEGKC ELFASWFAVFDTSQSQADPDAAASQSLPTDIDVDIFEV SHMMVEEALKKPSISPGDIPACILKHCKAVLIAPLK AIFNLSLRCCQFPKAWKMSYMRPIEKGSRKTEVSNYRGI TSLPAGSKCFEIMNDVLLQACKTYISQNGHGFEPQRSVI SNLCEYTSFCIKEMDKGGQVDSIYTDIKAAFDITVDHSIL LAKLNLGASSRICSWLYSYLTDRLQVLYKIGKAISAPFTT HCGVPOGSNLGPLLVLFENDVSHVLSNLCYLLFYADDL KIYLRINDCRQLQEQLDFVSWCTLNHLSISVSKCCII SFRMKSPVLYNYAINGTILNRVDNVKDLGVLLDHQLT FKMHYSETVDKANRQLGFIFKIASEFDDPLCLKSLYCSL VRSILEFASVIWSPYDAVWIAIESIQRRFRCALRSLPW SDPLRLPPYEHRCALLGIETLEERKMNAQAVGAKVLRS EIDSPALLRQLNYAPQVRLSRQLLYQPRRTMYGSNN PIDTISRFRQDNYNLFDENISTNTEKNRLNRQV	cd016 50	332	579
CR1	CR1-81_Ae	.	Aedes aegypti	T	TAGTTTGGG CCCCGTATC ACTCTGTAC ACATCAATC GACTTGAA CGAGTTCA AAAAAGTT TTATTAGGT ATGCACTCC AGAGACTA CCTTGCGA AATAGGAT GCTGCTGCC GCCATACG AGCATCGCT GTGCACTCG TCGACTTAC CTACACTAC GCAACAGG AGAACTCTT	SOHQRGVGLIAVKKALQSNVILINDANQLEISVRITY GTMSLVVCTVYLPNTEPALYQHAASVNLCKLLGTQ DEIVVLGDYNLPYLRWRFDGDIAGFLPNASSEQEISLVE DIVATGLMQMNHLLNSNERLLDLAFVSNASVCELLEPP LPLMKVDRHHTFVLTLDDVIQTPLENRKNETFFDFRQCN YDELNAEIAQINWRDILAFGLSDDAVTRFYSELNVIFER HVPRRSRACRPDNKQPWWNAQLNRLNRLRKARTYF RVRNLRNDAVRDLESQFNELNVSCFQSYIDRLSNMIM DDPKQFWSYLRSTTSRGGPQLINYDGKTATTPDDAVK LFSFFQSVLSITPPSSASYLSALKTFDLNLQPTSTFQHE VYKKLCSIDGSKGPGSDGLQPSFIKACASALAEPAALLF NRSIAESVFPKWKALITPIHKAGNVHDVKNYRGISIL SCLPKVFESMLLDLFLYPAVRHIITEDQHGFVRKRSTTN LMSYVSTLIDKIEKROQIDAVYVDFTKAFDRVPHQLAVE KFQRMGFPDWLTKWVLYLINRSASVRLGTTLSDPFEIT SGVPGQSHLGPLLFLFVNDLCSELKSKVMYADDLKIIF RVINSVDCALQMDIDKILDWCNRNGMDVNIACNV TFTTRKRSPTVFEYSMHGATIERVRLVKDLGISLDCKLRF VEQISSVIAKAYAVLGIIRNTKDFRDVYCLKALYYHLF AAFWNMEL	cd016 50	441	690

Crack k	Crack - 13_A Ae	.	Aedes aegypti	CA	CTAAAGCA GCTTTTCAT TTTTGATCT GCTCGAAA ATAACATC GACTGCTCT GGTCTGTTA GGCAAAAT AAGTTCAA CACCCCTGT GAGGTCCT ACGGAGGA CTGCACTAT TCCGACAC GCTGCTCAT CGAACGTT ATATGGGC AAAAATAAT CCGCTTGAT GTATGTGT GAAATGTT AATAATGT GTATCACTT GTACGATT TAACATAG CTAAGGAA ACATTAGA TTAAGGATT CTACGGAA TTGAAGTCT GTGCGATGT ATTTTATT TAAATCGA AGACCAAA TAAATCAA ATCA	HGGLAVFVRNELQMDVCRNIVIEGFHHICRFKFEANAI NLHAVYRPPSPDVRRLHEIEGMISLSKNNEKCVLVGDL NIPLNITNNIVGEYIRLLETYNMLPTNTVVSQSSNNIL DHVVCSTSMIEDVFNETIFTDVSDHCLICSSINMKCHTSE	cd016 50	420	676
------------	--------------------------	---	------------------	----	---	--	-------------	-----	-----

TTCTTTGTC	STLHKRIIDHSRLNELFERSILNIPQNLSVNERLNYVITHY
ATTCCACGC	NEILNDCSRLVTQRVKLKGSCPWMTLDLWKLKIKDKL
CACCACCA	LKKRRNNPNDNHTTELFHVSILQDKKAKTKRDYVE
CCCTTCGCC	GLLSGNNQKAAWKIKNVMGNQKSNKPTAIRNGAQLI
ACCACCCA	TDPQQVCSSFNDFHFCVGPKLSSNTNINRFGTLGSL
CCGCCAGC	QSTIFLKPSTRNEIILLINRLDSKKSAGPDGIPVSFIKHHHS
CACCAATC	FFASLLMETFNEIIVSGEPDCLKIARVIPVFKSGDPKDM
ACCACCAC	NNYRPISLTIIDKLEKLLVSRILEFSTKHHLYGQQYGF
CCACTGCTA	RQGSSTLTACHELVDEIYDALDNKQIVGALFIDLKKAFL
GCCACCCG	TVDHDLRLKLESYGIRGIAKDLLKSYLTDRHQYVSIGE
CCGCCAGC	HKSQRLVTGTVPQGSNLGPILFLFVNDVARLNLHGKL
CACCCGCC	RLFADDTAVFYRGTCNTILNQIKMDLEVLLLEYFGENV
GCCATCGCC	LSLNLKTKYMLHTPRRKIPAHGPISIGGHNLDKVEYEH
ACCATTCAT	EFLGLTIIDSVMSWAAHINKLKSVSLLCGILRKISSFMPQ
CACCTCAC	TCLKQIYFALVHSRLQYVTAVWGSASKSHLRELQVLQN
CATCGCTTG	RCLKIVLRKPLLPTVNLVSNRNDLSLLPIKALYEQVLV
CCACCAACT	QMQKIVKGTSTHNTVLHMSQQSSSRQANHIFLFRPN
TCAGTAAA	SEFGKKKKITYFGSKLYNALPENCKSFQNMHFKRFVRN
GATCAATG	ALKQKVEQFVL
CTCATAAAT	
ATTATAAT	
GTGAATTG	
AAAGTTTGT	
GAAAAATCT	
AAACTATA	
GCGCTTCCT	
TCAAAGAG	
CTACGCTCA	
TTGGAAGT	
GCAGTTATC	
GCTCAATCG	
TTGAATTAT	
ACTTATTTT	
TTTGAAAA	
GATGAGAA	
GGTTTTATG	
CCTATGGG	
AGAAAGTG	
CTTAAAGTA	
GCTACACTC	

				TCACGGGCT TTTCCCTTC TCCAAGAG AAAAAAA GCAGTAA TAAATAA TAAA			VNDVNNENGRCLDLCFVSTQDVAPSAIAPSAIVKNVS HHPPLVLLADSEPCPLTSPVSYDFNRAQQGIVEFLA SVNWLEVLDTNDVENAALTFTHILGHVIERHVPKKVHG TAGPPWQTQELRRLKTSKRAALRLYSKRGTSQSHRNHY VSVNNKYKTSRRCFRRYQQHIQQNLKTKPKSFWRVYN QQRKESGLPSTVKLDDEVASDRNQACRLFAKKFSSVFC DEMLSEYVALAASNVPVLMETLNAFSVDDGMIARAA TQLKSSINPGDPGIPSSFVKSIEQLLPIRHVFALSLSGT FPSVWKNAIHPVHKGDVKDVSNYRGISLCAISKLFEL VVMIEPLIAHCKQFISNEQHGFIPIGRSTTNLLSLTSHVSD SFTERVQTDVITYDLTAAFDKINHLITIAKLERLGISGLM LRWFKSYLTERQLTVGLEGHFSETFTATSGIPQGSHLGP VIFLLYFNDVNYVIKGPRLSYADDLKMXYLRVRSICDAIT LQNDLDAFAGWCSLNRMVNPEKCNVIFSRRKEPVYF NYQLFGKILQRVDHVKDLGVILDAQLTFKHMSYVIAK ASRTLGFIFRIAKEFTDVYCLKSLYCSLVRATLEYCSTV WHPHYQNGVQRIESVQRRFIRFALRRLPWRDPHRLPSY HSRCQLIHLETLQSRRDINRAMIVADLLQGRIDCPAILGQ VDISARHRLNRANSMLRPLRLRTNFGQNVAINGLQRVVF NRVAPAFDFHLPRHTVRRNFAEILTRRV	cd016 50	316	566
CR1- 121_ AAe	Aedes aegypti	AA	ATGAACT GTTCCAC TCCTAAA AGTTATA TTGGGAA GCTACTG TTTCCCCT TCTGCTG CTGTTGTT GCTGTGT GCTGTTG ATGCTAC TGCTGTT	TAGAACATT CAGGAAA TCGAATTCA GGAACATA ATTCGAGTT AGTTAAAT AGTTTAAAT GGAATCAC CACAGTGT AACATTCA AGAAGGAG TTTCCCTAC GTTACATAA			NICHINVQSLIARNFTKFHELKLTFDISKVDIICFTETWLN HTISNTMIGIEGYKLIRNDRNRHGGGICYLRNSLLYKV SMSQSSNDFYGT EYL NIEIKVGFDKVLLGVYNNPNIDC IDVLHDILENCTINYDCSFFVGDFENTDILKFSRRRFG MLDAMSYECMSEPTYFHQSGCSLLDLFTDSDPFAK LDQISMPGVSNHDLVFCSLRLSLNRVEDIVQYRDYVNF DSSALLDGFYNIEWNNLFCMEDPNEILNFNYHILKLHD DFIPLRRKKKNRPWFNSDISHAFINRLAYDQWKRST EIDRNTYKRLRNFAANDLVRQAKINHDKQNFNVDLP LWKNIKKLGISNDSFLDQDCLNANQINDYFISNYINDSS SVPIPFSSDGFKNVFHEEDIVNSIFSISNAVGLDNIPIRFT KILLPIALPVYKFLFDSIIKTSVFPRAWKNSKVIPIKKGN SASLSNLRPISILSSLSKVFEKLLKIQITKFTDNDLLHPLQ	cd016 50	467	726
Kiri- 9_AA e	Aedes aegypti	.								

GCTGCTG	ATTCAAAATA	SGFRQNHTNSALIKVHDDIARVIDRRGIAVLLLLIDFAKA	
GGAGGCG	AAGAAAATA	FDRVSHSKLVKKLSSKFSFSHNAVKLIESYLNERKQAVF	
TTGCTGT	AATAAAAT	YNGVLSDFSFIESGVPQGSVLGPLFLFLNDLPASLEFC	
CGAATAA	AAATAAAAT	SIHMFADDVQVYLCASTDVDMDNFARKINHDLNNILR	
GATTTTTC	A	WSQNNLLAVNPSKTKAMLLICKLKNRPLPPSIYFDGKRV	
TGCTGCT		QFYDKLENLGVFTSNLCWDAFNGQCCKIYGTLKKLNL	
CTCGATT		TTKHLDTAMKIKLFKSLILPHFTYCDFLFSNATASALNRL	
GTGGTGA		RVALNACVRYVFNLTTRYSHVSHLQDVLLGCQFSKFYSF	
ATTAGAT		RACLQVYIIIRKKTPLYDKLQFMRNTRTMSLVIPNHS	
CAAAATTA		SAYYGQSLFVRGINNWNRLSPTLKSTANIYGFKGGLLRE	
GACTTTT		LERLQ	
CGTTTTC			
GGCATAT			
TAGCAGA			
CATTATT			
AGCTGTT			
TTTTCTT			
GTAATGT			
GCTTTTCT			
CTAATGT			
CAAAATGA			
ACTTCAG			
ACTAAAA			
GTTAATC			
ACAAATGA			
ACATAAT			
AGTAGTT			
AGTAGTT			
TTGCTTTT			
CCTGATT			
TCTACAC			
TGCATGT			
TGCTTGG			
GTTTCGAT			
GGGTAGA			
CTGCTTTT			
GAAGCCT			
ATACTGT			
TCAACAT			
TTAGATA			

					TGGATGA CTTTCCC TTGATAA TGCTTCTC CGAACAA TGTAATT CCACGTG TTGTTCTT AATGCTG CATTAAA ACATGAC TGTTTGA									
													</	

			CAAGTAT GCGTGT ATACGAG ACGTCT TTSCGTG ATGAATT GCCAAAT TTCAGCT CTCTATC CGCCGCT GATGTAA CATCA	TATCGTTGA TTGTAATTT TGTTAGTAA CTGCTTCAT ATTTGTTAA ACTAGAAT AAGATATA CATAAGTTA TACATCATT GGGACCTG TTTTGCTG TTGATGTCA TCCAATAA ATAAATAA A	KYLLATSDDCYDIIVLTETWLDTRTISTQVFTPEYEVFRC DRNPRNSRKSTGGGVLAJVKKKLANIENDLWCSTEQ WVHVQFFGHSFLFCGYIPDRARDETLETHLQSVTSVI EIAKPMDEIVVLGDFNLPGISWISSGNGFLHPDMNHSV HSAASRLLDGYSTATLRQINHVITNEILSLDLCLFVSAQD VAPPSNAPSSLVKQVNHHPPLVLKLVQHYSTSESIAQV TYDFIKADHEGLEFLAAVEWTEVLDTTRDVNDNAVSTLA HILGHVIERHVPKKTLASKGLPWQTRRLRMTAKRTA LRRYTKNPTPRMRDQYVRLNAAVKRTSRHCENRYQQQ VQRDLTAKPKSFWRYVNEQRKVSGLPPLSLFLNGDVASE PQSVQQLFAQKFSSTFCDEVIPEVLVTQAASNVPVINDS WATLTVDDDMISRAASQLKSTMKPGPDGIPSAFLKRHID SLLVPIRHVFNLSLISGAFPSLWKNASVFPVHKKGERKD VNNYRGISSLCAISKLFELVMDPLMTHCKNHLSDQDQ GFISGRSTATNLRLTSHIADSFAGAQTDAIYTDLTAAAF DKINHAIAATAKLERFGISGNMLRWLHSLYIGRKLTVTIE GFQNEFLATSGIPQGSGLPLIFLYFNDVHYTLKGPRL SFADDLKYRRIHSDYDAMCLQQDLHAFASWCTLNRMIT VNPCKSVISFVIRRPVVFYEL YNTVIQRVDSVKDLG VILDPQLTFKHHVAYVAKASRTLGFIRVAKDFTSVYC LKSLYCSLVRSTLEYCSVVWNPYYQNGSDRIESIQHRFI RFALRQLPWRDPLRLPSYRRCQLIDLLETQLRRDIARA MVVADTLQGRIDCPVLLQSLDLNVRPRVLRNNTMLRLP FRRINFGQQSAYIGIKRVFNVRVASFDFHLSRQTIIRKF KTVFLSLDE			
CR1- 109_	CR1	Aedes aegypti		TAATTTTGT TAATCTTGG TTTGAATAT TTTAGTGT TGTATTTTG TTTAAGAA ACAAAGGT CATTGGGGT AATGTTTTA CCTGTTGAC TAATAAAT AAATAAAT AAA	AAPASSNILMYQNVGGINSSLAENRAVSDSCYDVYA FSETWLNANTLSNQLFDSSYIVYRQDRSASNSKRTGG GVLFVRSNFKSRVLSNPGSSIEQLWVAITTVDATLYL CVIYIPDRVNDLESIEKHTDSDLDWVVSQKPRDSIMIIG DFNLGSIWQYDXSGSLFPVASRSTISLTSRKLDDAYSTA CLRQITVGINNENSRVLDLCFVSEELRDASTVMQAPSPL VKTCRHHPPILLKMEICPQKRHDVSESVFYDFRNADFN GMNDFFENVDWDDVLRDSANLAASTISAILLYGIDQY VPVKVREPSKPAWSNSELKNLKRVKKAALRRHSKFR DSTRARYMQANTEYKQNDCLYNAHQDHLQSILKANP RSFWKHVNEQRKESGLPPTMSDGLHEADSKEDIADMF TQESNVFTNEHVDPDIVNDIAIRDVPRLSFVQQIITDDM VASASKELKSTGSGPDGIPSLVLKRCATETIAAPLARVFN LSLASGVFPNCWKQSYVFPVFKKGSQKQVSNYRGIAAL SATSKLFEVILQKLVQSYAHYISPDQHGFMKPRSTTNL	cd016 50	518	768

						ACFTSFILRLMEAGQQVDAIYTDLSAAFDKMNHNQIAVA KFDKLGMSMLLWLQSYLTGRNMSVKIGDHVSSPFR VWSGVPOGSHLGPFLFLYMNDVNFKLKCLKLSYADD LKLWYTIKQPEDAVFLQHELEAFQAEWCQINRMSLNSK CSVISFRKHAPHFGYFLAGTQLERVSTVKDLGILLDSK MTFKDHVAYAVSKASSQLGFLRYAKFRDVCYCLKAL YCSIVRPTLEYSSIIWSPYQNGIKRVEAIQRFVRFALR HLRWRDPLNPSYGSRCQLIKLESLSRRNVAKACFVG DLLQGNIDCSTLLSCLDINVRRNLRTHSFINPSARTNY GLQEPMRMSRVEFNMCYHVFDFNVSRRETNKSSYRRVL C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
--	--	--	--	--	--	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

GAATGGA
AGCATCC
GCCATCG
CCCGATA
ACATATT
TCCCACC
TCGCCGC
TCTCCAC
CTTCCAA
CCGATCG
AAAGCTG
CAACAAG
ATCAAAA
ATCAACC
CGCCCCC
ATTGCG
CCANCGC
CCCACCA
ACGCCCT
ACTAYTG
TTTGCTG
CACCGAA
CGAAATC
AAAAACA
GAATTCG
TCCTGTA
ATATCTC
GCTTCAT
ATTGGAA
ATAAACT
ACTTACG
CTGCCTC
GTCTTTC
ATTCCAA
CGTCCG
AAAGGGA
CCAACGA
ACGATCA
TCATACA
GCCTCAA

TCGCCGT
CCAAAGC
CAACCCG
TTGCTAC
CAAGGAT
TAAATCG
CAAACTC
TGCATCC
GTAGTAT
CATCCAT
ACCGATA
AGGAAAT
TCTGTTT
GGTCTG
CACTTGG
TTTATGT
ATTGAAC
TTTGAAG
CTCACC
TTTGTA
CAGAGAA
CACAGCG
GCAGCTC
GTGGTCA
TTCCGAG
AATCAAT
AGTTAC
AGTATAT
CACCCCT
TCGTAAC
TGGGAGC
TTGTTCT
CGCAGCT
AAAATGC
AGGATGT
TTTGAA
AAATGGC
CTTCCGA

ACTATCC
GCTGATT
TTTGGAG
CTGTGGT
GGATTTT
GTCTTGG
TCGCTTTT
GCTTCAG
GTGTACC
GGAATGA
GTCTGAA
TATCCAA
GCTGCCG
TTGCGAA
TAAACGG
TGCTTCT
ACTGGAT
GTGTAAC
GCGTGA
GCGGAAT
AATGATT
AAAGCTC
GCTTTAC
TAAAGCA
ATCATAT
CAGTGAA
TGTGTC
TATGAAG
GCATAAT
CGAATCA
ATGAAAA
TCCGAAA
TTTCGAGA
CAACATT
TTGTCCG
AAATCCG
CTCGGAA
TTGCAGA
ATAACCT
TAAACGA

ATTCCGG
AGATTGT
TGCAAAC
ACACCGA
TTCGAAA
TATTCAA
CGTCGA
CATTGA
TTCGGCT
GTCCGTA
ATCTGCT
AAAACGT
CTTCGT
CAAAATGA
TGATGTG
CCTCTCG
ACCCGCC
TAAAAAA
CTGGTCT
GTGGCAC
TGATGGC
CACAGTA
CTAGTTC
TTCAGAC
GTCATCG
CAGTCC
TGAACGA
ACGGATG
AAAAACAC
GTTTGG
CTTTACTT
ATCAGGA
GTGTCCC
CGAAAGC
GCCTGAT
GAGAAAG
TTATCGA
AATGGTC
AAGAATA
GGCTCGM

AACAGAC
GACTTAA
CTGTCGT
TAAACTT
GTTCCTC
TTGGAAA
GGATACA
AACAAATC
TCACGTT
TGTATCC
TTCAAAA
TAGGAAT
GAAAGCT
GATCTTA
GGGCAAA
AGCAAATG
ACTTCAT
CAACCTG
GCCAGTT
GGAATTC
GTTTTCG
CGAATTC
GAGAACC
ACAGTTC
CTCAAGG
GCGGGGT
TCTGGAA
TCCCTCG
GAGAACC
AGCCCGA
ACCAAG
ATCGTAA
CAGTATT
GAAC TAG
TATCCCC
ATCGTGT
CCCATG
ATCTTCA
TTCCTC
GTGATGA

			GACCGTT TTAAGCC ATTGAC AATCTAC TACCAA ACGTGCG GGGT			EQSIVCVKLQKSSIIYICGIYLRPN SQPVLYSSHSA AIQQLS ERISRFD SIVVGDYNLPQLIWQTD DDDINGLLPSNASSEQ EVTLLETMVASGFQINNIVNSNGRLLDLAFVNDNDNV ELLEPTPLLRMDNHHKPFVLRIDVND SCTQVGDSNHA DDDFRYCNFDELKSAIATVEWATLFGKDTDTVCIFY EVLHRLDEHVPRKRRRRTHPFKHPWWTSELQHLRNIV RKSRRRYFQSRTVENRDKLRLEASVDCQTASFRRYVA RMETTAKEDPSAFWTHIRNRKRNVKFFPAEMAFRDTVAN SPEGIANLFADFLENVHCTNPPTSPD SLRNCPTFDL NLE PFDFAQDDVLSALQKLDTSKPGTDNLP LLFKECDSF QIPLTHFNKSLQSRTPF MLWKTASVIPFKAGSTRGVENY RGISILCCLAKIFEELVHNVLYTASQPLISQFQHGFVKKR STTNLMFTTFSLSTEIENKHQVD AIFYDFSKAFDKVPH DLAISKLRLHGFPPNWIADWLRSYL TNRKAFVNINGTYSR VISVTSGVPQGSVLGPLFVLFINDL CFR LKSGKLFYADD LKIYRTIASHLDCCALQADVNELELW CQENGME LNIKK CKSI AFRRRQSRTEFDYKIGSEPLERVESIRD LGVIIDTKV RENDHISVINAKAFAALGFVRRSTNDFNDIYAL KSLYCS LVRSILEY AACVWAPHHTTQIVRLEK VQRSFIYALRQL PWSDPVNLDPYPARCM L INEMLDARRNNLQRLFVFDL IMGNIDCPALLEDVQFYAPLRQLRERDLLQIRRHRTSYG FNNPLSRCFLFNCASALFDNFVSKYVFEKNRLKDF GIYRPPHQEEELDLFESRLEME AACSRDMVMVMGD LNY PDICWEEHLARSDRS CRFLT C VEELYLVQEVNGPTRGK ALLDLVLAKGDDLVGDLRIEGLNGSDHEMIEFTIRRA GKSASRAEVLDFRKADFNELRKIVGETLWEEGLEGKV QEGWSFLKGTILRAQKRIPVCRKGGKGAGKPSWLTRG VKECLKKKKEAYGWQGVVTK EYTSIARICRIRK ARAVTEIRLVTEIKDNKKSFFRYTGSRRTKGNVGPLQN GTGQLITDSQEKAELEFNEFYFASVFLHANGRRPPGRILDG CGRGDCQPMVSVELVREHLERLNVFKSAGPDGLHPRVL RELAGVIAGPLAKVFE DSWCSGEVPE DWKRANVVPFK KGRREDPGN YRPVSLTSPGKILEKIIKKSICERLAEGMM	cd016 50	413	664
CR1- 64B- AAe	Aedes aegypti		TGATCTCAG CAGTATTCC TGCCCGA GCAGGGG GTTGGACTC GATGATCTT TCGAGGTCC CTTCCAGCC CTTAATTTC TATGATTCT ATGATT						
CR1- 8-A Mi	Alligator mississippi piensis	TA					cd016 50	375	626

							LKDSQHGFVAGRSCLTNLVSFYDQVTNRLDVGDAVDVI YDFKKAADMVSHDALIKLGDCLGAESTVRWVGNWL RGRTQRLVVDGSSVSWREVASGVPQGSVLGPVLFNFIN DLDEDVESTLAKFADDTKLWGKAVTLGDRMRIQADLD KLERWAERNRMQFNKDKCRVLHLGRSNQRHEYRLGG DVLASTMAERDLGVMDVHRMNMMSRQCDAVVNRANSI LGCISRCVASRAKEVILPLYSALVRPQLEYCVQFWAPHF KKDVENLERVQRRATRMKGLGLEDKPYEERLRLDLGLFSL GKRRLRGDLVAVYKYVRGEYRELGEKLFTRAPQGRTR SNGHKLEDDGFRLLDVRKNFFTVRVTRIWNRLPKEVVQS PTLEVFKRRLDRHLAGII						
							SPPAGFKLGSSGEDEYGGSHGTPNQETPTRSAQVCSK DPIRQQFKGGPGTATRIRCITYTNARSMGNKQEDLELLL ASNFDLVGLTETWWDSSHDWVVGIEGYRLYRRDRVG KKGGAVALYVKEQYTSLLIKLGSEGGQAEILWVRIQGA RGERDLVVGYYRPPHQDDELQDEFKQLEEA VRSRD MVMGDLNYPDICWEEQSARSNSRRFLTCIQDLHLTQ EVRCPTRGNALLDLVLATGDDLVDGLRIQGNLGDSDHQ MIKFTIQRRAGIIPSRARVLNFKKADFNRLRGLVSRVLG AKNLDALRVQDGTWFLKETILWAQSESIWQKGGKGG ARKPLWLTIRGHQDTLRAKKEAYRKWKQGVASKEEYTS LARTCRESVRKAKATLELRLASGIKDNKKSFFRYVVGSK KRVGSGIQDQGGGLVTDSDRDKAELNEFFASVFLNG NLDEPSTYITDTNKGNSRPTVSADTVKRHLEGLDVEKS PGPEGLHPWVMKELAGVIADPLARIYERSWCSGQVPED WKRANVVPFKKGRKEDPGNYRPVSLTSPGKILEKIKE HICGSPLANVMRRGNQHGFLAGRSCLTNLVSFYDKATN CLDAGVEVDVIYLDPFKAADKVSHTILINKLRFALDDY TVDWVANWLKGRQTRVVVDGVSSTWVKDVGSGVPQGS VLGPVLFNIFISDLDEGVGSLSKFADDTKLWGAANTLE RRNQIADLDRLGKWAESNRMQFNKDKCKVLHLGRK NCHHTYRLGGDLLSSTEAEKDLGVTVDSRMNMMSRQCD EAVGRANRTLSCIRRCFTNRSREVILPLYAALVRPQLEY CVQFWMPQFKRADNLERVQRRATRLIRGLQGRPYEE RLRDLNLSLSKRRLRGDLVAAKYFIRGEQHEVGDDDLF TRAPGMTRNNGHKLLERRFLDIRRKFTTVRVARIWN GLPREVVLAAPSLGVFKRRLDRHLAGIT	cd016 50	536	787			
CR1- 9D_A Mi						TCAGAAC AAGACAC ATGCTCG GGTGGGA TGGGCTT CACCTCT CCCCCAA AGGTAAG CATGTGT TCTCTTTC AGGTTGG CTGA	TGACCCCA GGGTGGT GACTCTTT CTTGCCCTGC CAGGGGCA GGGGGTCA GACTAGAT GGCCCACT GAGGTCCCT TCCAACTCT ATGAATCTA TGA	Alligator mississip piensis			cd016 50	353	603
RTE						TAATTAG TGCTTAT GCACCAC	TAAATGGG AAAAGGCT	Amborella			cd016 50	353	603

			trichopoda	AAGTAGG ATTAGAT GATAGAA TTAAGAG ACAAATT TGGGAGG ACATAGA TGGAATA GTCCAAG ACATACC AAATGGA GAGAGGA CCTTTTAA GGAGGGG ACTTAAA TGGTCAT GTAGGGA AGGATAG TAGAGGA TATGAAG GGGTACA TGGGGGA TATGGGT TTGGGGA AAGAAAT GAG	ATAATGAT GAT	WNSEHDVNVIVNEMADCIKKVAKEVLGESRGRNFDK ETWWWSEVQKAISEKRSYKWTWCNRNDQNLEEK NAKKNAKQVSKAKLKAYDNLYNKLGTRGEKDVFKL AKIREKKCRDLNHIKIKSEDRVLVKDDEVNERWRVY FQKLLNDEHNGNIVVEGTGTSEGGGHHYIRRMSEV KEALKRMKIGKALGPDGPIEVWKCMEIGIDWLTCLF NKVLRTKKMPDEWRKSIVPIYKKNKGDIQNCANYRGIK LMSHTMKLWERVIEQRLRHETSVSENQFGFMPGRSTTE AIFLLRQLIEKFRAGKDLHMVFIDLEKAYDRVPRDLW RVLEKKGVSRHYIDLVRDMDYDGLTSTVTSGETSEFS VTGLHQGSALSPYLFVLVMDLTRLQDEVPCWMLFA DDIVLIDETRNGVNTKLELWRDALESKGFKSRKTTEYM ECKFSNNRSRDEEVVKIDGQEVQKRPHFRYLSIIQENG EIEEDVAHRIRAGWAKWRCATGILCDRRIPMKLKGKPY RTAVRPALLYGAECWADKKQYTHRMMSVAEMRMLRW MSGKTRRRDRIRNESIRENLGVAPIGDKMRESRLRWAGH VWRRPSTTPVRRCELVQVEGLKRARGRPKRTWLEVVR KDMGTYGLTEGMAFNAEWRTIRIHVADPN			
L2	Unal 2	.	Anguilla japonica	CAGTGTG CATCTGA TTGTGTC GTGCTT CTGCCGT CCCCCGC GATTCAG ATAAGCG TATCTTA ACTTGAT TTGTCTCT GCTGTTG CTAGTTA GAGAACA	TAAACCACC ACCATTTCTG TATATTCTCA CTCTTTAAT TTAAAAAA AAAAACAA AAAAAAAC AAAAAAC CCCCCCCC TTTTTTCATG TCACCTTGTA TTTGCTTT GTCCTAATA CTGTAGCTT	MFNPIPLCALPRPKWSIHRRRLNLSNNIYPPLSTPQEFVT GGLWNCQSAVKKADFTTAFATKSLDFLALTETWITPD NTATPAALSAAGYSFSTTPRASGRGGTGLLSLAWKFCV LPSNPSTSSSFESHATVTHPSKFHIVLYRPPGPLGSFLDE LDTLSSLPEDGTPAILLGDENLSPETSQLSRVASLLQSF GLSSSPPTHKAGNQLDLVTRSCSIPQLTVTPLHVSDH FELSFPLYLHPLNSSNSSTVTFRRNLSTLSQSSLISTALS TLPPPASFNLPVDLATSTFNSCLSQSLDCLPLVSKPAR SSPPCPWLTESLRSRTSLRAERKWRKSRSSDDLTAHY TLAAFTSALTSKASFFQSKISACVSNPRKLSTFESSLLT PPPPPPSSLSADDFLA AFEKGKVA TIRNSFAHPVSPPPRD RPTATLTSFEMLESDVLQLITSHRATTCPLDPPPSKVLQ SISSEFLPYLSSIVNSSLSSGHVPSDFKMARVTPLLKKPTS DPSDVNTYRPVSLPFLSKTLERAVLKQLTSFLHQNNLL	cd016 50	513	773

[illegible]

				GAACATA CACCGTC AGGTTTC CAGATGT TTCAAAG ATCTCGA AAGGGGG GTCGAGG GGGTTGG GAATAGT GGCTCGC GATCAAT TCTCCATT AGAACTA TTCCTAC CCCTGAT ATT	TTGAATGTT TTTATAATTG TGGAATGA TATTTTAAA TTGTAAGTC GCTCGGAG CACTCTGGT GGAGAGCG ACTAATTA GAAATAAA GTGAAGTG AAGTGAAG TGAAGTGA A	DMASQDITAEFTQSYFINKINLLREGLPPTEDTLREAF WPSSDKLLELFVPLEMEDLEKTLAAARPTTCSLDPCPSR LVKECWEGLRDLPLDIINSSLEQGIPECLKEAVVHPLLK KPSLDAAPSNRYPSNLPFLSKVIERTVAGLQQFLDET AGLDPFQSGFRSGHGTETVLVAITDEIRQLDRGGSALL VLLDLTAAPDVTVDHLLIHLRMSGVRLALXWFSFL RDRRQRVWVGSGSESSCLCGVPQGAISPLLFNIYVR PLAGLVRSFGIECYQYADDTQLLLRLEPGTTTIPENFKLC LEAVMSWLRASRLKVNPAKTEILWPGQPPELIQSLPDFD GEVLVPSATVKSGLVVDSSLTMEAQITAISKQAFHLR QARKLASYLSTEALATVIHATVTTRLDYCNALYAGLPK TKTQKIRMVQNAARLLAGSSIRSHITPILKQLHWLPIEH RISYKILITFRARNGQGPLYLRDLRIPFHQWSPRSSQE NLLYVPGPREVHLEATRRRAFSTYAPILWNSLPPYVRA MSELRFVVKALKTWLFCAFK			
Pene lope	Posei don_ Ac	.	Anolis carolin ensis	AGG	GCTTTTCAA TGCTAATTA AAGTGATT AACTACAC AACATTAC ACTGACCTC TCTCACCTC AGACTTCC ACAGATAT ATATTAACC TCTTTGCTT AGTTTCTC CATACCTCA CAACCTCTG	MPAIDVGETSGENASGTWPHSPKDIQQPCDPGHESLRQ HIEHLYGEKLFQDTRRLEKLRKAHLCSLTFLLRCD TDTPQFLAAKRTFKTPQAHRIYRNLRLRERHTIRK ELANTDKELLHLHINISQKMNQDWDKIDSLTYKKMEK DMVLHTKRQKQKFKCRKQKPKPELDKSRITINLTDRQ LTEDQVSILEKGGNFAVTTRIPVENIANVESAIYQLPEE EAEVVRMETARILRNAKLPSPNITRKERQAIKDLNSDPEI IILPADKGNATVIMETKQYKEKIRQLDPTTYKKLKQDPT NKITRKTNTLIKNSINFDIRQQLCKSEALPPRLYGLPKIH KDSIPLRPIVSAIGSPYTNLAKFLATQLQTHIGLTAHYKD STHIEKISNLNSTKDILISFDVVSFLTQVPAADTLTIK QNFPEDITAFHHCLTTSYFQWDTGFYEQKDGVAAGSP LSPVANFYMEYFEKQALETAPKKPTVWFRYVDDTFTI WSHGEELSKFLDLNLSIHPNIQFTMEKEKEGKLPFLDV LVIRKPNQQLGHTVYRKPTHTDRLYHKNSNHPSQKRS TIKALTDRAQRICEPHLLQGELNHLNWLQANGYSTTDI RRAARPRTSHEQDKDPPRGKVFPLPIKGTITDRIGKLMK KHNLTQTYRPTKKIQQMLRSADKDRDPLSSAGVYRIPCS CGQVYIGTTKRSQTRVKEHERHCLRIQPEKSAIEHLM NQPGHRLIFENTKMLDHSNNYHVRLHREAIEIHKHVDN FNKKEETMKMNKIWLPLVQLNSKIKTVNKKQYSENRGFP DMNQPRAVNDSKQRMPIRQEEEDSR	cd003 04	470	554

Ving i	Ving 2_Ac ar	Anolis carolinensis	GGGGGAC ATGAGAG AAGCCTC CACGCAG GATCGTG ACACATC CGGGCGT CCCTGG GCAACGT CTTTGTA GACGGCA GATTCTC TCACAAC CAGAAGC GACGTGC AACTTGC AACATGC AAGCAAA CAAAACA ACCAACG GACCTTT TCAGGAC CAACT	TAGTATTAA TGCCAAAGTT TTTGTGTTT TTTTCTTAA AAAAAATC TCTATGTTT GCAAATCC ATTACAACT TGTACCCCT GGTTTCGCT TCTGACACG AGAAATAA ATAA	MTIISINIEGMSAAKEQLLYELCRDKKCDVLCVQETHRD KQQRPKVPGMILVAERPHAQYGSVFKPGLQVSSVH NTEENNIEVITVELNNTITSVYKPPNEEFCSFSPENFDSH QRTVIGDFNSHSHVWGYAQEDKNGEAVLSWAELHKL SLIHDSKLPPSYNSGRWHRGYNPDLLFVSDSIVQQCSKS VGPPINQHRPIVQLFPTIRPLEVFRFRFNRKADWS KFSDMLDDKIIISVAPIPEEYERFIKIVKTCRASIPRGCT HYLQGTPTDPAALDEYYRLHNEDEPSEQTLTAQSILSS ISEAKKEQWIKLITEVDMGRSSQKAWRLRLRNSNDPSQT NTHANIKADQIAHQLLKNGKPKLSTKVGRKPIVRQPDN ENNLHEPFTSTELDMALNCKNGKAAAGLDDLMEQI KNFGPKARCWLLKMMNNCIASCQIPKIWRKARVAILK PGKDRNDPKSYRPSILLCHLYKVLRLJLNRIEMKIDPCL IPEQAGFRKGKSCTSQVLNLTQHIEDGFERQQITGAVFID LSAA YDTVNHRLLRKTYNITKDYHLTRFIGNLLQNRSF FVEFQQRSRWRKQKNGLPQGSVLA PSMENIYTNQPL PEGTESFIYADDRAITAQAGSFEMVEQKLEALGALTAY YRENQLIPNPSKTQTCFHLKNRQASRALRITWEGIPLE HCSTPKYLGVTLDRLALTYKKHCFTIKQKVGARNNIVRK LTGTTWGSQPDVTKTSALALCYSAAEYACPVWNTSHH VKTVDVALNETCRITHTGCLRPTEKLYCLAGIAPPDICR EVAANNERTKALTSPAHLFGYQPARQRLKSRNSFLRST EILAGTPQARVQKQWAKTRNLNQWLT PDERLPPGHTE DWATWKALNRLRSGTTRCRANLKKWGYKVESATCDC GEEQTDDHLLQCILSPATCTMEDLLTATPEALQVASYW AKDI	cd016 50	459	714
CR1	CR1- 9_AG	Anopheles gambiae	TTGACAG TTGACAG CTGACGA CWGAGTG TGTTTCT CACTGTG CTCCGCG TAAATCC GTGAATT GTTTTCG GTTTCGG CCTWATT TTCGTCA AACTGCC	TAAATGTTT TGTTTCGTTA TCTAATGTA ATGTATTGT AAACAAAT TTTGACTCG AGAGGGGT TCATAGTCC ATCGATTAA TAAACTAA ACTAAACT AAACTAAA	MEGVCCDCSVPLESLECGMCSFCDAVYHSACTKSPFTV IEEVKRTASLHWSVCVGCNSALGNPHSKAVKTTGMQVVG FQAALTA AVEVMKAALVPPVMHEIREGFASFAAAHQ PSEHFNVPDPALPSGKRRRLFRD VVASTEAI VVDNAPL YPVNTDIDNTHQNRPSLPPIITGTDITTSIPTVSQLPRT NYMWLHLSRLAKTVTDQV VSMVKSQD TTDVIAFSL LKAGTTISSVSSLTFKVRIIAALRDKALTAGSWPVGLGV REFISLPQRPSHHRPDLTNTNTPIPVRHSPMHPVQLSPDM NCTLISAPQKQLQPHQTPCLKLFFNTEFLSNNHTTVS SKGTIVPLNLTSHLFRNNNTSITSLTDDLA PTYTRIPLN VRSTISITVYQNVRLGRSKADEFRLT VLEADYDVVVL ETWLDPSLPTALLFGDAHC VYRCDRNAANSSLSRGGGV LIAVSSALRSHALPVLAQTLEFACICIQLPYRLYIAAAY LPPNHSTDEGKINALLDITVNSICDITLGPNDRFLVGDEN	cd016 50	909	115 8

				GCTTAGC TGTGACT TAAGACTG TCCCTGG TAGTCCT GCCTTCA ATCGTTC GTCACAA CCATTCC TGACTTT GCTGTAT TTCCCTGC GTTTGTTC CTGAGTC CTGTGGC ATCTATC TGCAACG GCAACTA AA		QPALSWSPAQSNHETPFVFEFPHSTSSVRSQAQFVDGLHQN ALYQLNNTNSMGRILDLAYGNWSSAATSSLRVSEHP LLPIDCYHPLEFDELEITASSMNHATTVTEIKLNYARADL TKLRLICISFNQSECSNYTSDHATNDFSEFMRAALREC APIKPHRGPWWGDRTHALKKAKRAAYDYLRANRSTP CRLYNGAHALYRRYNRVCYRRYRQTERSLRKYPRRF WSFADNMIRKSTGLPGTIRYKDNCAHSPDLICELFATRFE DSFISNITPPEAVASALVNTPADAIHVTLPIVTTETLVTKI ERLKSSYSEGPDGIPATILKRCAVSVAPVLTKEFNESLRT GTFPSLWKVSWLTPHKKGSKNEAVNYRGITSLAACAK VFELIYEPLMASAGKYISTNQHGFMKPRSTTNLMQFV SGCYKSIDDDGMQVDAYTDIKAAAFDSVSHNILLAKLDRL GLSPPLVKWMKSYLCGRSYAVRMGSHQSRSGASSGVP QGSNLGPLLLYISDLCTALPDNSCLLYADDAKIYREIR EPEDHLQLQATLTFVSWCKRNALSVCIDKCAIISFSRSR APALFNYTLDGQNLERNVCVLDGLDADAKLSFEEQID QVVASGNRLGLVINMTRELDPDMCFKLYCALIRPLLE YASIVWWPASTRALARLESIQRKATRFALRDWPRRLDY RTRCLLGGIPPLAERVEHTRLATFITGILNGTIDCPELLSRI HLVYVPARILRRRPMPLAVAEITRTTFGSRNPLLCMCRLLNS ANDIYEPGMTTITELTSLSVRNAFNNNNI			
CR1- 10_A G	CR1	NT_07 8268.2	Anopheles gambiae		TCCTTTATA ATCCTTTTC CCTGATAC CTTCCCTAT CCTCTTTT TCTCATCCT CTTTTTC GACTATCAT ATCTAATAT TATTCACCT TTCTTTTCT TACGTTTGT ATTTAATC TAATATCTG TCTTTCATG TGCTCAAA AGTTCGCTT TAGTATTAG GAATTAG	SPRIPRIQLQCSPGFSSTTAAIPSQAINRCDPRTRITSVY PYRFSAAFPVDEPAATTAQSNSTSTPTPLHHAERFE TLPSVASPITVSAPLPLHASFPHPIDDLRFYQNRGLRT KLDDLRLASADTEYHVLIJLTETWLSNPSSLAFDNNFVI YRCDNRNRSRSGGGTLIAVSRKLAACEISPADSLE QCWIRLKLPCSLFIGSIYLPPTLSADASTLNTLDASIDEII SRMKKNDRLLLFGDNIPSITWASPDPAPEPLLV PAS SAHSSSLFIDIVRYHLSLSQLSSVSNFRQVQLDLFTFPKNQ SNATSTCTYVSQLAPVPIQIDPHHPPELLESLKITGSLTPSN RSQHSSRLTTAPSLDFRKTDENLNTLSNIDWSQLYSC KDVNLAVSTFTSLFLSTLPQCCPKRNPSPPPWSDATLK SLKRDKSRYQRAYWSTRSVNALRTYKAAAAYRLYNR TRYNGYILILQTRLRWYPRAFWFMNARRKGSLPSTM HLESVYASSPEDICSLFRNHFNSIFSPSPQIPNTASQCLSY TPQDVIDVSCFGFDEHSVSEALSKLKPSFAGPDGIPPSL LKKTAEVIAPLLSFIFNLRLCCFPSSWKGWLCPIFKK GNSADVSNYRGIVLLPACSKVFEIAVHSVISFQVKSWSIV NQHGFMPSRSTVNTLMQFASHVIPSLSGLQVDITYTDF KAAFDITVPHHLLLLKLKKGFSNMLIDWLYSFLNSRSIR	cd016 50	634	882

				ATAGGTTA AGTTAGGA ATTAGTTAG ATTAGGA ATTGTGAA ACCATCAG GCTGACAA TTTGAATAA ACTTGAAT GAAATGAA A	VKTDITFSSFCPSGVPQGSVLSPLLFLVFLVSDVHAILPS EGHLLYADDLKIYLTISTHSDCLRLQEFNLNFSLSWCSYN QLSLCPEKCSVISFCHNRNPIHYEYLSNSVLSRTSLRDL GVLWDDKLSLNEHLESVVASASRTLGLJCLRLTRDFRDPL CLKSLYCSLVRSTLEYASVVGPVRSGWILRFEKVQKRF TRFAFRMAGPNAAMPSEYRCNLLGLESLEVRSSQAR VLFGRMLNGLIDAPTLLEKIGLYAPSRNLRARSALNMV SRQTTFGSNEPLLCCVRFTEFTSTDPHPRL				
			TATATAT CTTCAAT CGAAAGG CATCCCC GAAAATT TTGGCGC TTCATGG ACAGCAG ACGACGC TCCAAGC ACAAATCA TGGGCAC GTCACCC GAAGATA TTTGCAA CATTTTC GCTGAAC GATTTTC CG	TAATTTTAT TTTCAAAAT ACATTTTCT ATTCTTGTG CACCTGTTC TCCCCTATC GATTTATGT ACCTTACAT AGTCTATAA GCACTAATT TAAGTACAT TCATGTAGG ATCTTCAAG ATCCCGATG GATTAATCA ATAAACAT ATTAACAT ATTAAACA	MPSPVDQTVISRATSYTPLNMIDYTIPTDIRTVCKLKS LKPSTAPGPDGIPSYISKCYQSLAPVLNINYSLSGIFP RLWKTSMVPIFKKGDRCASNYRGITSLCAVSKAFEL LLYGPMLSATSNYIMDNQHGFPVKRSTLNLTEFVSFK RNLDSSGGQVDIYTDLKAAPDLISHDILIAKLQKLGFSE QIVKWLSYLTERTSYKIKVNDHTSREVLTSGVPQGSN LGPLLFIYINDVQGLLAHRFLFADDDVKLFAPINDTND CINLQHAINLFSIWCNDNAMVLCIEKCRVLSFYRSRST RFNYQIDNISVERIDTFRDLGILDTRLTENDHLENIVSRG NQLLGWIIRSTQGFNRPMTIKTIYCAVYRSVLEYGSIVSD PCTEQWSNRIEAIQRKATRYAVRLLPWQGDVLPYSYHS RCLLGIQSLKKRREIAKCLFISGLLNHHIDAPTLATIEF NAPSRNLRTRQFISVPRYRTFRGQSDPMSAMCFTFINNF DLFDENIPQNVFRDLRLRTQLT	cd016 50	92	340	
		Anophel es gambiae	TTAGCCG TGCAGTC TCGGCAT TGAGAAG AATAGGA CTGACGT GGTGCTA GCGTGTC GTACGAG GCGACTA	TGATGATG ATGAT	MMQGESRKEAAGQKPPSSKSKSLRIGSWNVRLTYEA GKTAQVLQEMHRYQLHILGISETHLNQSGQKRMKEGEL FVYSGNDSGPHREGVGMISKTAQKTRGWAEHGPRII MASFGTKKELKVRDGRKKVINNMIIQVYAPTNDAEVEN KDDFYNRLQNVIDKLPKGDINIVMGDLNAKVGSDNSHN EEIMGRHGLGEANDNGERFQNFCTNNLVIGGTFFPKR IHKATWVSPDGTENQIDHFCISRKFRRSLEDVRLRGA DVASDHHILGKMKLKLRRYAQVAAGNRLKYEVSLK DPVKNEFKLELKNRFQALQDLDECEVNEHWDKLREA VNATCQSLTGPKDNTIRNEWISQDSLDKVKRRRELKQV				
RTE	CR1- 8_AG						cd016 50	562	819
RTE	RTE- 1_AC	Aplysia californica					cd016 50	562	819

[illegible]

CCGGGGA	HRRLIQKLSACGVKGRVWKVVRDFLTSRKQRVVINGSR			
ACGAGA	SEEAETSGIPQGSVLGPLLFVIYINDLPRGLKTTAKMFA			
AAAGGCC	DDTKLYVRSDTVNGPQDLQDDLDLTLQEWQRWLLRFH			
AAGACCT	PQKCCVLKVGKDNNDNEYFMDQAVDGS TSRHKLNVTKR			
ACACTAC	EKDLGVVVDNKLTFKEHVAQMTSKANRIVGLIRRSFDY			
TGAAATT	LSEDMFVLLYKSLVRPLLEYAHCVVWQPFKEGLCSEIEQ			
TGAAACT	VQRRA TKMLSHLKEKPYSERLKKLKLPSLEHRLRGDM			
GAAGAAG	IETYKYIKGYITITERPAFTKATTTQLRGNSEKLQKQRSR			
CAAGTGC	LCVRANFLENRVNTIWNNLPESVVSAPNVDTFKRRLDE			
TAATACG	YWKDLPTIYNPTCQN			
ATGAGGA				
AATTTAT				
TGAGGCA				
AGAAAAG				
ACGCAGC				
AAAGGAC				
AAAAATG				
AAGGAGA				
AGAGGAC				
AAGTTAT				
ACTTGAC				
TATTGGG				
ATGCGAA				
GAGATAT				
KACAAAG				
CAGGAGA				
GAGAGGA				
GGA CTAT				
GCCCTCT				
ACCAGGA				
GTTGAAG				
GCTAAGA				
AAGAGAG				
GTAAAAA				
AACTTGG				
GGGACAA				
TTTCGCTC				
ACTGGGT				
GAGGAGA				
GACGGCA				

AAATTCG
GAACATT
GGGAAAT
ACCCAAC
GAGAACGT
CGAAAGT
AGAGGGA
GACAATA
CAAGAAAT
AAAGCCA
AAGACAG
CAAGAA
TGTAAT
AGTTTG
GCCATTG
TTGCGCTT
TTTCATAT
TTTAATT
ATTCGT
ATCTCT
TCCTGAC
GTWCAGA
ATAGAAG
TATCTTCT
CATTTTCT
TCTTTTTC
ACCTACA
TTTCATT
CAATCGG
ATGCAC
ATACTT
AGAGAA
TGTAATT
GCTTGA
TACAAAT
GCAGATC
AATGTT
GAACAA
AGAA

TCAAGCG	MSIRIASLNCRLNKKLRRAIFRKCTQYDIICLQETVVT	cd016	507	772
ACAATCC	NTTAREWSLEWSDFFYSEGTNKSQGILVNKHFKYK			
GTGAGGC	TVELCIKTERILGLSFSXDDVSYVAINVYGPVVKNDKLD			
TAAAGCT	FIDKLYSVLKDEKADHTIICGFENVVHNTLIDIVTGEKH			
AGGCGGA	DDMSTQKFRDWIANCQIVDWRGVHGEKDFTWRSRT			
GACACAG	PFTARRLDYIFCNTELFPFVIDSKHAFVYGTDHKAVITV			
GCAGAGA	RTEKFKRGPSYWKFNNSLKDKKEYVNSMEKCIDEYLER			
CACACAC	TLDETDKTIRWELLKVMIKSHHTIDYCIRKCRALKRENLI			
CCTGCGT	TEELLKLNQVVKNPQDLTLLKQFESKKKELEYQLHK			
GCCGTAA	ARGAQIRSRQKWIEEGEKNTRYFLRMEQMRGGSNVITS			
GCGAAAC	LQDDKNNLYGHVDIVDSIKNHFERLYKKDHNIEINRV			
AAATATG	DDFMQGCSPVLDDVDKEICDSLSIVELDAIVNKLNIDS			
GACTCTT	APGIDGTTTFYKFFWSKLKPLTFNQQCLETGELSPYQ			
TTTTCCCG	RRGVISLLHKGKLNRRDNINWNPITLTNTDYKIFSKILA			
CAGCCTG	IRLQRMVPVLINENQIGFLKGRSISEHIRTLDIIYVASEK			
ACCGCTC	NLPGMVVSLDFQRAFDVKEGTLATLKKFNFGEGFIN			
CCGCCCC	MVGVLMANTCECCVQNGGWLSEWLNTERGIKQGCCTSP			
CAGTAC	LLFVLVAELMSIKLRNNSISGMTILSGEIHNFCKVLQYA			
GCTCACC	DDTLILKDQEDLASALLDIDSGSISGLKNKDKSQGM			
CCGAAGC	WIGSDRTRTEKLGGISWVPNGNTIKILGIHFSATEASRIE			
CAGGACA	LNWIDKIENMKNMKMMKRNPSLYGKVIICKTFLLSQ			
CAAGGTC	ISYIIQSLSLPERVLNTIDTLFFKFLWQKRHSNKKAREKV			
TCCAAGT	KRTVMCRSIEEGLGVISVKDQKVFELNLWLKRMATV			
GTCAAGA	TEGRNPNIRISDLAFTEVGGCDYFFSATANCKKTGKL			
CGAGGAG	CHASKFWSVMDTWIEMASIKKEPSSTIEILMEPLFNRR			
ACATAAA	NIKYKGKSLYLKRWMNSGIKYVYHLYDRRILKSKEELS			
AGACAGG	HITGTYAAFDFDYALMNALPTQWKQDVSDNNTEELL			
TGAGCGA	VNTRVINTNGKKTVEILSKSNREIRQLLKYRRHAEICG			
CGAGCCA	RNFWMRKMNIDITCKYNIAHRSTKESRLRLHLFKILHNI			
GCTCGGT	FPSNILLNRLGIKNTLCEHCGVKVDVIEHMFHCTLLKNF			
CACCGAG	WKYVVFHTICLKTSSRFPMTPDCNILFGFSERGOG			
CGTGAAA				
NCTCGAA				
AACAGGA				
GGTGAGT				
GGCACGA				
GGAAGAAG				
GCAAGAC				
ACGGTAG				

ATTATTTT GATTATG AGGAGCC AGTGACT ACCGAAA CGTGTGA GCACGAA GAGGTGG GTGCGTG AGTGGTG TGTACAC GTTTCGG TAGCCAT TGGTAAC GACTTGG ACTGGTA AAAGGAC TTTGAGT TTGGTTT GCGTCTG ACTGTTT GG				TAAAGTTG GCAGCAGA GGAAACA GTCGAAC AACATGGA GACCTCAC ATCGGTGA GGGGGCTC GGTGCGTC AGAAAGC TGTAGAGA GGAAGAGC CCGAAACA GCGACGCA TTCCTCCT GATGATCTC CATGCTAG AGTCCGGG	MDNDLSVKRTMQKENQENSMCEQC YRKFDTWRLR VHQGKVC LKKRRQCRSSDCKTRSKSSQDENHSGAISAS VDSPTLSQEQMEQQT EPGLRKPRLWPAANEKAKYKT LEEEVQKKLKEVKKSSNRGDAKEILSAFATIIYDSA KEFGICEGKKKTEKKSGLSRRQRVLAQLRKQKRD LQKQKKYAPPEEIEGLRILFEDLKKKSRDLQRH ERRCTRRKERREREQFLKNPYEAAKKLFT EARSGLKCTKEELD DAYVKQTYSDSRREEPLPYMRGLKRPAA PVVKYYLGP IREKEVD EFVKKARAKSAPGGDGSYKVPKYCDTLRHTL FEMLRGLWKDKEIVKDWCR AEGVYLPKEQNAKEIGQFRPISINVTCKIFM GILAKRTVAYLQSNQYVDET VQKAGVPGIPG CIEHAFTTWDITQEA KKNKESLNVVWLDLANAYGS VPHKLLMKAMDDFFYVPQ EVQNMREYYSNFQMR FSTANFTTEWHRLEIGIAAGCSIS VIWFLVMEMLLRSAECCEEE EAKVRSPPKKA FMDDVTLLTRDGD TVMQSVLNRLDKLITWAR MKFKAKKSRSLT NKGKQQLKFTIAGEQ MPTVKEQPVKSLGRWYSGTL SDRSQGV EIMKQAE NGLKAIDQTKLPG KHKIWCLOFALY PRLAWPLTMYE VALSRVEMIERK CNS	cd016 50	355	614
--	--	--	--	--	--	-------------	-----	-----

YIRKWLGLPRTINTSSLYRHRGALQLPLTTSIVEIYKAGKI RTVLMRESRDKEISGKPPDVKTARKWKA EVAANEIISA LEHGDIVGPAQPDRLGLNGDFRPRKMSQRDRRKAAT RHMQKMEAEERREIHLIQCSQQGQMTRWEEHVVERKVG WSDIWKWNTSRLSFLLRSTYDVLSPVNLVRWKVRED NKRCRGKLGTMKHILSNCNLALNRYTWRHNEVLKVL KMAATQQVEEGKYAPKIQKHGPTKIAFVPQGSKVAKRK KANEAMKDSGGTWEVAADLKGFERFFPIPTTKKPDIV LWEEKDKEVHLVELTVPHEDNIYAAHERKENRYEALVE QCEEAGWKATHPFVEVGCRGFVATSARKWMRVAGLT AKKSNELTNSLQETA EKASHWIWLKREDSWNES	ATACGGAG CGTTCGACG GTGAGCAG TGCATACG GCTGTTTAG ACTACGTAC	CAGGTGT ATGCTAC GGACATC CTGGAAA GACAGCA CGACAAC CCGGAGT GACGGGT GGACATC GGGAAAG ACCGAGG ACGTCGG GAAAGAC CGACGGT GGCAGGG GCTAACT ACCCAGG TCACTGC AAATNTA AGTTACG TCAACCA NTAGAAA AGACAGC CCCGGGA TCCCCCG AGAGGGG GGAGTAG AGGATGG ATCCACA AGGACGG ATTCA	MLFTNDRIPIYELRRRLGNQGLPQVLVGNQAGLVDGYW K KKKIRKQRKWKNNKKTPTTGTNGKEKLIDFNILQANV CGLDKKKTQLGKILYERKVHIALQLQETLHSSCDINITGY TPYPCKCEGCQGITIYIRNDIQGEVELLNW/HPTDAHKVT LWYGNSKLJILYNIYCPSSTLNLEQSTIFKKTVIAGDLN GHSPLWGYADQNNITGKRIEELCXSSNLILLQDEDSKTL LHRRHGTLHRPDLTLVSADLHNYCSEEVLDICSSDHRPV LTKISILKHRQRKRRTRWPFKANWDLFRKTAEDTVKL DLDNKSIDALNSELTSILDAAKKSPOGCRNTYKPVW	cd016 50	550	806
---	---	---	---	-------------	-----	-----

AAGAAAA	AGGCACCA	NERLEKATKTKNDARHEYEKDLKNMDKKIRYKKA VAE
AGAGGAA	CGTCAGAG	AKLVTKESKKEAWTQKCEGLNLHQGGREAWSLNLT
GCTAACG	TAGTCTAC	GEKRKENPKPLHTETEVLTSELKKAHFNKHFA SINKSS
GGTCAAG	AATGGATTC	KRNDLDRGLLSVLKGERDLMTPSIFQDFLT KRELDCAI
GGGTGAC	CTGCCACA	KKLKKRKSPPDKVHNEMQLNLEKAKLLLLKLYNRT
CCCGAAA	CTGGGATA	WTEGNLPKAWKLATINPILKKGKKANDPKNYRPI SLTSC
AGAGACA	CCGGGAAA	IGKLCERILNSRLYWLESSGLISQCQAGFRSKGRTE DQ
TATTACTT	TGAAAGAG	LFRLTQKIDGFORGEQTTAVFV DLQQA YDRVWRKG LGL
TACTATG	CTGATGCTC	LKLQNMGIKGNMYSWIKSFLTDRLIQTQMDETLSSKAV
TTTCTTAC	TGGCAAG	LEEGLPQGGSSISCTLFLVFLNDVSDVLKTEKALFADDLV
TCTTTCTA	AAAGGAGC	LWHTGNTTISRRLQDDLSLGTGYCRLWKMKVNCEK
TTTATTGT	TAGCCTCCC	TVYSVFTRSYNLANAEIPLKIDGTNLKKNPNVYLG VQL
TTTTATTG	TCAGCCTGA	DSKLNKQHTNNLRKKAMKRLNLIKRLASTNWGSDKN
TTTTACTT	AAATCCTAC	TLRGLYLGYTRAVFDYNLVLQNICSKTKTKNSMDITQ NQ
TATTACA	GAATTACA	ALRLISGGMRSPTAAACEHTNIEPLETRRKRAALEVYER
TTTTATTA	CCACCTGCT	SKRLNNHPNRVLVDKWKPNQRLKNTKSVLDTVLNLQ
ACTTTTCT	GCCAAAATG	EKHHLPERREPLERVPSTLPPHLTIKPEIRKTLDDKSNK
TTTCTGCT	ATCAAAGTC	NTYPVALKHSALETIDSYPTWIHSYTDGS AFKATINAG
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	

516

[illegible]

			ACAGGGT GGTGGCA ACCCTGT CTCTATA AAATACC TCATGTG CTCAGTA AAGAGAC TGTCAGG CGAGAGG GAGGCGC CGAAACC TGCCACC TGCCCCC ACTGCCT TG'TGGGT TAGTCCT TTTGGAC AAAGGCT AGCGGAG TAAACCG TACAAAA ATCCAGC AGAAGGC AAGGGGA AACCAC'T GCTGTAT TCT'TTCCC TAGAACA TT	VISRXTPIPTCNIKIND'TTLKQVQHFKYLGTTISSDGRCHT EIKSRICQAKATFHKMKQILTNINLSLETRKRVLKSYIEPI LLYGCEAWTINKQLERTIVATELWFIRMLRIPWTARKT NEEVLVEANYRKQLITNIRKROANFFGHTMRKEALENI VTTGKFDDGKRGRGRQREKILDKMTTWLGAARTTDTIA AXKDRXKWRSMIAHASRQGTG			
Ingi	Ingi- 7_AC	Aplysia californica	TAAAGGGA TTTGAAAAAC AACAAACA	MHTSLHNPXTNKMPLRQMKLYELGCRGLGYQGPRSLPSP GGGSKSATGTHTAGIGATSN'TQLNILQLNISSIQNKK MELAKTSLDLRIHVALLQETQLPKRIELHL'TGYTMYHC KCTKCGIATLIRNDITAEVSHPETDDETDMEVXVQQS DGT'TYTYNVYCPPRXTLQLNFPNTDHHKHTILAGDFNGH SPLWGYSDTDXTGTVLEELLRSSNLISLQNDKTPPTLLH RASGTLRSPDLTLVSADLHPHCSMTVLDDIGSDHRPILL QISSNRKREPTNPIRRSWNYKKANWELYYQKSESRLLD VCTTNPIDECYKDIT'TALKTSAKEAIPRGNPKKFRPYWD KELEQLTKARKKARRRAEKNPTAENRKHNYNLTAKVK	cd016 50	545	802

GTCTTG	LKTKSAKRQJTWREHCGELDLRKDGRKAWNLLTNLSGT
TCGAGCA	KRRKNPAPMKPTEKETCRAEVMNKYFASINKKGNQN
GCCTGGG	KLSHAIKQLTKRAERTPKANLEVEFEFTLTELKAAALKK
GACCCCT	CKKRKAPGPDHVTNEMILNLTSSSILLGFNRTWNEG
GCTCTGT	QLPTEWKKANITPILKKDPASQPSYRPSLTSCLGKV
GCATAGG	ERMVNDRLYWWMESTGTISTQAGFRKGKRTMDHLIR
AGCAGGG	FAQKTSDAFQDKDNVAVFIDLRQAYDRVVRQGLFYK
GTTGCAG	MQKLGIAGRMYSWIKNFLSCRTIQTSCDSKISSQRILEE
ACGGAGC	GLPQGSALSCITLFLFMNDLPDVISCQKALYADDLLXW
TCTGGCC	SKGSDMRALNQAINADLTITSSYCOLWKMEINTDKTVF
AGGAGGT	GLFSLKNKVTASQVNLKINGTIRHEDYPKYLGVLQDR
ACCATAG	KLSLKEHLGELTKKATSRLLNLLKHLSSLNWGADKATLR
CCCTGTG	QLYIGYVRSVIDYSLPLQSIASQSSLAQLDRVQNQALRF
AGGTGCG	VSGAMKTTPTSAACEIHTNIEPLDLRRKSATLETYERFVRL
ANGTGAG	EDTNENRKMADEWEPRTRIQKTSFMREATKLLEFDPP
CAACGAC	ENREPTNISSEPPYWQTPETITRLLDPSCDKTTPPEILR
CCTTCGG	TSALETVDSY AETPIHAYTDGSATNATRNAGYGSTIKYR
CTCGTAG	NNLGEKKQTLFGPCGTYSSNYDAERIAIVKLEQIAQD
AAGGGTC	FFEGAVTPGDVVVFSDMSVLA VEGGGDYRGDGNLS
CGTACAC	LVKQMGTSYGVVRVTLQWVPGHSGVPGNETADHLSKTG
ATCCACC	ASLPQTDTTASLQTCRGHIAKSRDEWLNWRWANAETGR
ATTATGG	SLFGYMDKPNPKDPSNLLCRADQTLIFRARTGHLHTNK
TAGGTCT	HINRINPMWEPHCRHCDYHEETVEHLLHCPNLA VLRR
GGCGTCA	RLHHTPTDLQSLLYSGVDQMTRTCGYLREALRCQERNA
CAGCAGG	RS
CCGACCT	
GTTCATT	
CAGGAAC	
TCACTGT	
CGACATC	
TTCITCCC	
TCCCTCTC	
GTGGGAG	
CTCTACA	
GACGGGG	
CATGGTA	
CCCACAG	
ACTGCCA	
GATGACT	
CATCGTG	

AGATGTC
TCCGCGG
ACCCAGA
AGGCCCT
GTAGCTG
TAAGGAA
TAGGCTG
CCTGGAA
ACAGATG
ACGTCC
ACAGAT
ATCTGG
CAAAATG
TCGCTGA
CCCATGA
CGTGACG
ATCTCT
GTCTGAG
TAGATT
GGCCACC
CCTCGGA
TTATTCTG
CCCCATG
ACCGGAA
ATGTTAA
TAGCAGA
TAGCCAG
GAGACAG
CAGGTGG
TGGGCAA
ACGCAAG
CTACGCT
TGGTGAT
GTAGACG
ACTCACT
GTGGAAG
GAACGCG
TTACCCC
TTACTCA

	RTE-4 AC		Aplysia californica	ATGGGAC CCCTTAC TGGA CTG CTCAGAC GATCACT TGTCATT GGTTGCA CTTTGTG GCTCCCA TCCTCA GAACTAG AGTGTGC CTACACC CCCTATG GTTGTCT ACTGTCT AAGAAAAT CCCCGTG AAGGTTG ATGGAAA ACAGGAC AGCGAAC ATTGCAC CTTTGTTC TCTTGTG GCAAAGAT GTTTGGG CACAA GC CAGTGT C AGGACAA GACTACC ACCTACC TGAGATC CTACAGG G	TGAGAAAAG GCAATGAT GAT	MKDSGQSRKDSSDPKPSLSNKSRLRIASWNVRTL YEAGK SSQVVQEMKRYRLHILGVSETHWIQSGKKQLSSGEQVL YSGREDKYHAEGVGHLLSKYA QKSLRGWEPHGERIIMAS FTTRSKSINNMTIVQIYAPTNDASDEDKDAFYSQLQGVID KLPRKDVNIIMGDANAKIGQDNVGYERIMGQHGLGQM	cd016 50	551	808
--	----------	--	---------------------	---	------------------------------	---	-------------	-----	-----

			GGGGAG CAAAGAA TACACCT GCCGTGA TACTGCG TGTCGTA AGAGGCG ACTAAAA GTGCATA GGTCTGG TCGCCCG AAAAGTG CCAGTTT GCGGAGG GCCAACT ACCAAC CCAGTAA AAAACAG TAAGTTA CAGAAAC ACGGACA AACACAG CTTTTAT TTTTCGAT CTTGGA GACGGAG GAGCTTC TGCATCA GCAGGAA CA	NENGERFANICAFNSFVIGGSVFPKHRIHKATWVSPDGV TENQIDHFCISRKRFRSLEDVKVQRGADVGSDDHLLLA KLKLRLKKGHTSTGKRKTQVSLLDGHKKEEFQLQLR NRFQPLADLGGEDTEVDTHWTKVKEVFSTCQEVLLGG KKWEHKEWISQHSLLDIRRQLKAEANASRTRGSRAD ALEQYRTTAKDVKKSRKDKKEEFMNLAEKAEAAVG GHMRLYQTRTLAGKYGKPEVPVKDQEGKSIFGKEAQ SKRWVEHFNLSLLNRPPSPNPILPARNELPICCDAPSKQ EILDAIKLKNKGKAAAGPDQIPPEALKADANVTADILLPL FTRIWNESKFKDWKEGHLVKLPKKGDLSCSNYRGIT LLSIPGKVFNRIVLERIKTGTDPRLRDEQAQGRKNRSTTD QIATLRIVEQSLEWNSQLIVNFDYKAFDSIDRATLWK IMRHHGIPQKLVNLIKSMYEGTSCRVVHDGQLTDSFDIK TGVRQGCLSPFLAIDWLMKETTRGHRNGIQWTPW TQLDDLDFADDLALLSHTQEQQMAKTSVLSLSESVGL RIHPGKSKVLRTDAQRKEQVTIEGKALENVESFCYLSII DREGGTGAIEKSRIGKAQAAFIISLDRIWKTREISLRKLR IFNSNVKSVLLYGCEAWNASQTCIKRIQVFLNRCLRKL RIRWTDRTNVEVLKTRQIPVDEVGRRRWRWIGHTL RKSGSCTTKKVLWDWNPQGRSRGRPRGTWRRITEIDVQ KSGKTWNNIKRLAQDREKVGWVFRGLYPGAG			
			TAAGCTAA ATTGACCGT CTGCATAG ACGAAAAG AGCCATAT ATATA	MSPRVTFDDGSDYSVALISYRPSQTGQPPVSKVLIRHVC LLLWVYGARTKTERQTLNPAPGKKRVKLTWPWKLGTWN VRTMTTGLDTNLEKTEDIRKTAIIDRELTRLNVDTALQ ETRLANSGTIRERNYTFWKGGLGEDERRRHGVSFAVHN RLNLRISTPQGISERTVMMETKTGKAHFISAYAPTLLT ENKDKQFYELQDAINNVPKADQLFLMGDFNARVGT ENKVVPTCLGPGLGKINENGQRLLFECTNNLCITNTF YSGKDRHKVSWCHPRS GHWHQIDFIIVRKKDANMVKQ TRFSYADCNTDHLHVIKAKIVTKKTPRAEQKKPKIN IRNTKDPALKAKEQHEVKEELQHLDSSSPDNMMWNSMRS	cd016 50	605	868
RTE	RTE- 3_AC	Aplysia californica					

TGTAGGT	TTYSASITFGIERPCKEDWVEASASTLAPLEKKRILIE			
GCATACG	HNTNPTESTKQQLRAAKSNLQRETRKCANQYWDKLC			
GGAGTAT	EQQASDTGNIRKMYENIRVAIGPTANKVAPLKSAATGEQ			
GAAAGGT	LTDKKKQMERWVEHYSDLYSEERHVDRLHEAIPQLSE			
TAAGGGA	MSELDTPTEEEELSSAIDELACRKAAGNDNIPAEVLKEN			
GTAACC	KETLLPYLYKLLTTCWLKAEIPHDMDAKITTLYKNKG			
CTAAACA	DKGDCNNYRGISLISITGKAFARIILKRLQKLAERVLPES			
AATCCGG	QCGFRSNRSTMDMIFSRQLQEKCREQLPLHTAFVDLT			
AGGAGG	KAFDTVSRSGLYMVLGKIGCPPILLQLIISFHEDMKATIQ			
GCCCCTA	FDEGCVLAPTLFSTYFAVLLQHAFKESPGDVFLHWRTD			
AGGCAGT	GSLNLAARKAKTKINKTTVRDLLFADDAALVAHSEST			
CGACAGG	LQTMMDRLSTACKVFSLVSVKKTIVILTQEGAEQSHIWL			
GAGGAAC	NDKELETVDKFCYLGSTISSTSLDYEINIRIGKAATFG			
GATCCGA	RLTKRAWENKRLTIKIKIYEACVLSLTYGSETWTLY			
ACTATCT	SRQEKRLDIFHLRCCCKILGISWRDKVTNVEVLARANLQ			
TCTGGCA	SMTTILCLKRLRWLGHVCRMDDNRIKPKQLLYGELAAG			
GCCCCTG	KRQGRPKLRVKDTCCKNSLSKLGVDVSIWEEIAEDXKT			
CAACTTT	WRKXIKERTKMLENSLRNNQAEKRQRKRENSGNTLRS			
ACTGGTT	TTLTCKHCDRICASNIGRISHERSCMKTTTLRPS			
CCAAATG				
TATTGAC				
CTCGTCT				
CTCCTTTG				
GACTAAA				
TCAGCTA				
AGTCGAG				
AGGGGGT				
TCGTGCA				
AATTGGG				
CAAGGCA				
CGTTCIC				
CATATAG				
TCAGGCC				
CAGGCAT				
TTAGCTT				
AGTCAAG				
ATTCAA				
GATTTC				
GGGAGGA				
CGCTGCA				

[illegible]

[illegible]

CATTGAG
GTTATCG
AAGGCCT
TCAGAAA
GCGAAAG
TCAACCC
CCCCATC
CCCGCCC
AGACAGG
CCATAGT
GATGATA
ATTCTGC
AATGGCT
AAGATCC
TACAGGA
GATAAAG
GACATCA
AGCCTC
CATGCC
CAAAATA
GCCACC
AATGCAA
GCCCCAG
CCGCTC
CTGGCA
AAAGTCG
CAAGCAA
GGGAGAG
TCGTACG
GAACTAC
TATCCGA
ATCAGG
ATGAACA
GGAAGAAG
AGAGAAT
AGCCAAAG
CTCTCGA
GCGAAGA
GTTAGTC
AAAAAAA

TCGGAGT
ACAGCAA
ATAGTAG
GTGCTAG
ACAAATG
ACAAATG
GTCAAAT
GAAGGTC
TACTTCG
CCGGTGA
AGAAACG
AGGAAGT
TATGGAG
CGTCACA
AGGACTG
GACAAAG
AAGCTGG
TCGCTTCT
GCGAAAG
TTGCCGG
CCCTGTCT
ACCAAGT
TCTAGTG
CATGACA
TGCCGCT
CTCCCTTG
AGCCACA
AAATGCA
GAACATC
TGAAGCT
GCTACAA
AAGCCAA
ACCATGT
TATATAC
AAGGCAT
GCTAATM
CAAAAGG
CGCTAAG
AAATCTC

AGCAGCC
AGGAAA
AAGCTG
GATCTTA
GCTTGT
GGATTGA
TCAGGCC
GTACACG
CAGATAT
GGCGATA
AAAAAGG
CTTGAT
GTGGCAA
TATGAGC
TCAAGTC
AACAGAA
ATATTCA
GGTGGG
CTCCGC
ATGATGC
AATGCTT
CAATTGC
CAAAAGT
ACGGACA
TATTGCG
AAAAATCT
GCTCCGC
AGACTCG
AAGTGTG
GCCACTG
TGCAAGG
GAGCACA
ACACTCG
AAACTGT
ACAGGAA
CAACAAGA
CGTCAGA
TGCAGCA
ACTGTGG
AAAAGGC

CACAAAA	CATGGGA	CCAGAGC	TGTCAG	TTAGGTC	ATGGCG	AAAAATA	GGGTGT	CCAAAAAC	AGAACCC	AGAACCC	AGAAATG	TTCCTGA	CCGACGA	AAAAATG	ACAAGAA	ATCAAGA	TGGCGAA	TGGCAGA	TTCAGG	AAGCAGG	AAAAAGAA	GAGCAGT	ATAATAT	GCACCA	TGTGTGG	GGCCGAT	GGTAGTA	TAGATGG	TCGTAGG	CGTCAGG	GCAGACC	ACCAAAG	ATCCCAT	CAGCTGT	CCCCACA	GTCGCAG	GGCTCTT	CGCAAAI	CTCAGCG
---------	---------	---------	--------	---------	--------	---------	--------	----------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	--------	---------	----------	---------	---------	--------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

AGTGCC
ATCAAAA
ATGACAC
CCGGCC
TGTGACA
CGGGGAA
ATCAAGC
AGATGAG
ACTGAGA
GCGAAAT
GGATTCA
TGAAGGA
CAAAATCA
ATCTTTC
AATCAT
CAGTACA
ACGTCAA
CAGGAGT
AAGAATA
AAGTGCA
GCACGAT
TTCTCC
AAGCTTT
GAATCCA
CTAAGC
ACCAATG
AGTGGCC
CTGCAAG
AACCATG
GCCTCAC
CCCACMG
AACACAA
CGACCGT
CACCCAC
CCTGCTT
ACCACTT
GGTCTTC
CCAAGTG
GCCACAA
AGCAGG

ACATGCA
TCTATAT
CAGCAAA
AGTCTGG
CTATTAG
CAATATG
AGAAGAG
AAAGAGT
GCCAGAG
GAATCGC
AAGGAGA
CATCACA
TCKGTCA
GCCTCGA
TACCGAC
CTTGTA
AAATTGT
GATACAT
AGCGTGT
ACAATCC
ACCCCCA
TTACACC
ACAGCAG
CCTAGAA
GAAGGAA
CACTCAA
GTGGCTC
CCGCAAA
TCCTATCT
CAAGAGG
GGCAACA
CATCTA
CTCGGAG
ACTTAA
TGTGCAC
CATCCGC
ATTGGGG
GGGACAG
TCAGTAC
TTAGGCA

				CCATAAG ATCGCAG ACAGCTT CATCGAA		Caenorha b ditis brenneri						
				TGATTTCCTC TTGTCTTTTC CCTTCTTTC ACTTGTCGC TTTTGTITG ACCTCCCAT TCGCTCCAT TTTAGGCAT ATCCCCCGT TTCTCTTTC TCCCCCGGT GAGGCCTC GTGAGGGG CACTTGCCA GGTCTTCT GATGTATCA TATTGATTT TTCCCTCTT TTGGAAT AAATAAAT AAATAAA	LLQKLINFGIPSSVSWFAVFLADRTFSVKINDSIGDTSPT IPSGVPQGSVSGPLFLIFINDLLSLATIPLLHIAAFADDI KJFSNDPFVVGJGIDTVVWAKSNSPLAHSKTAHLLHG SLNPFPHYTIAGNPIPTSSVRDLGLLTJEPNLKFSAHINRT VALALLRSKQLLKSFKSSSPQFYVYLKCYVLPJIEYCS VVYCPPTSLSLKLLEYPLRFTTKVFHRCNISYTSYPDR LAHLDLFSLRHRLKAQLQLLYNFLSGSSYFPNLDShVR FSSPPRRPMNLVNVNPRVKDFEFTTVPWNAITCHCPTFL RPFSENSLLQCSIQRF	SSF5 6672	45	118				
				C'TCTTTCT GTACTAA GCCTTTTC GTGGCGA AAGGGAT CAACAGA TAACACAC TTCAACA AACTCGA TGATTTC GTCACAT CGAGAGT GGCTTGG TTTGGAA CCAAGTG	IGLAGHYRVSRDTTFALNSNPCLSPHPASAKGKTVPG ARRRRSNLGVQITCQSDLEIKPRDKTRRTDYRVCTFN CRSLSSDDRLAELLETRRIQFDVIGLCETKRAAEAHLIH HDGTGVFLGKRNESSVGGVGFIVRSTLLPKIVEVRFLS HRIGLLTFKVSRKFNCTVIQVYAPTADSDLEEICDFYDSV EDVFECSRKYKLIGDFNARMGCRQNNERYIGPHAME PRNDTGELLATFCETNRLWHTNSMFKKPMHKRWTFVS PDGNHRHEIDHILANGKFVTDITVLPSTNGSDHRLRC NLHFNNCLAKLEQVRRRKPPKRVLDPAIAYAISATVTV QSDPDLDTDYDNLIQSLKELQDQAIVRPANHSSNRLSEE TRKLLNKRFRMDRNDPQFKSIDKCREAVQKDHEAFAS TRLLSAANQKSLKRVARDINEYKSVIPCLKSTGERIT SRVKMEQEIEKFYTELFKSAVNSQTSIPKATATPPFLPE EIRHVLSRFPNGKAAAGQDKISADFLKSDHNDVIDLITDRF NRYLHSRNVPKPWKTSKTTLIFKKGDRLENLYRPICLL	cd016 50	562	831				
RTE 1	RTE- 7_CB re	Caenorha b ditis brenneri	AF054 983									

				CAAAATTA TCTGGCC AAGAAAT ACACCGT CTCACGC AGGCCAG AGGTAGT GTCGAAT ACTCTCT GTCCGGG CGGCTCT AG			PVLYKVFITKLLNRMRRSLDEAQPVEQAQFRRSFSTIDH IHSLQRLLEVGREYQIPLTLVFIDFKKAFDSVEHQAIWKS LDEQAGADGAYIDLLKEYCKNCTNFTPHRPVAVPVTK GVRQGDPISPNLSACLEHVFRKLSWIELKGEAEDYDIP GMRVNGRNLNLRADDIVLIANHPNIASKMLQELVQK CSEVGLEINTGKTKVLNRNRFADPSKVYFGSPPTQLDD VDEYIYLGRIQNAQNLMPEIHRRRRAAWAAFGIKNT TDSITDKKIRANLFDIVLPAITYGSEAWTFTKALSERVR ITHASLERRLVGITLTQQRERDLHREDIRTMISLVRDPLNF VKRRKLGWAGHVARRKDGRWTTLMTEWRPYGWKRP VGRPPMRWTDLSRKEITTRDADGEVITPWSIIAKDRKQ WLAIVIRRENTTNS		
Ving i	Vingi - 2_CE	Caenorha bditis elegans		TGAAAACG GGCTGCG CCGCTGTT GCACAACA ACAACAA C			NMKQPAKRTVMTCLDQRAAFDKVPIKKLCVLDHESLN EAKISKWLTNLSKRIKANFGNHSSKWCTLLGGVPQGS VLSNLFsfymRNMPHPKTEITVYADDTTILAQDTPNF KAGELIQEHLDLLYDFFNERKMSISAEKSTTTTFTCDTK EFKTKIPVNWNGAEIPTTNQVRLLGVLHNTMCGSKDHV AETRKKMMVSTNILKALSGTKWGCSELLISTAKALVK PIAIYGAPAWSQLSETNLNLEAAYRNLRTCCGLTSD TPIEHYAEAQMPLKEEFELAKQQTYYATTKSGTHPNR NLESRRNTERRSNRPREPLNLEATKTHLSRLKGTGKAE AQKANHTRFVREFLNSAPPNRLGATPPPLNSSESSLPK ATRTALARLCGHSLIPRYKARIEGTPIGTCGCGDNEG GIDHYLSCGRSHPISEKELWTSPEAARALGLAYKPYD PGGGATTSTV	SSF5 6672	8 150
				AGTGACG GGTCTCA GTAATCA ATACTCT GTCTTCC GTACGGA TCGTGCT AAAGGTA AACGAGG TGGAGGT GTTTGTA ATGTGGT AAAAGTA GAAATCC CATGTGA			MPTRKSKILDVALSNSVISNLNIMAPIGSSDHTIKFTVQ LLQNISVPKMPIDFSKLDWPCCATYFENIEWGKVFDTH SHVNDIYECFLGHIRTMSQAVPTMPMPETSKVPVQYLR SFEQMVKNQFSKTMSTKTMNDIEYIYLARKYRKKLRK YLSNLEVKLFKKGNKSKFSKHAKTLLKPRNCSVPALIDN DENEIVSELKANLLASYFEAQYTNQNNTNFALDDDFV SDQPIPVWTDAEVLKVMASKKSCSVTSDQVPFTFIKLS PLISSVLAQIFNLMMRGQVPHAWTESIVPLDKITNANN PTDPRPISLTSHLCRVYERCILAKIELFLIDNKFWSQNOH GFRPKKSTVTNMLESLNDWSKSIDLGEQDIYLDIFS KAF DRLPHDLLDKLVSLRMKNLLIWLNEFLSNRSFRVRIG QTLSSRKLATCGVPQAGVLSPLFGIYVNDISSILPEKVA CKQFADDTKLYVSTPKTESENLLQSALDAVVDWTRGS KLTLNQSETVHVTVGKRRTDFKYHLDDGYPIERKAITRDL GFLISEKLDSEHWRKCNLAKFQLANMFNQYSTSNKKL	cd016 50	299 552

AACAGTA	TTATTTATG	MILLYKTFIRPRLEYGTVVSSPTKKSDKAIESVQNAFTTR	
GTTTCAG	TAATTTAAT	RLYSRIKKRYINPTDKDYKTAMERNELFDLQTLSTRRKII	
TTGAAAA	TATGTAACC	DRIQVIRMNNGKVLDLKTSEFFEQQETHTRSKKKYIWST	
CGCTTAA	ACTTGTGAT	GKSKLRRHFFVNRILATMIKQH	
CGCTTAT	GGTTTGTAT		
GAAATTC	CTATTTATG		
TCGTAAA	TGTGTTTTT		
TGACCTG	TTTCTACCT		
AACATTG	CACGATATC		
ACGTAGT	TTACTATCG		
TCCCTTTA	TAATAAAAT		
GAATCAT	AAATTAAA		
AAACATA	TTAAATTAA		
TATCGTG	AACTT		
CTCCTTCT			
TGCAC'TT			
ATCAAAA			
TACTGTT			
AGATTAA			
TTAGTAA			
AAITGAA			
GACCTTG			
TGGAACA			
TCAAGGC			
AAAGTTG			
TAGTTGT			
GGGAGAT			
CTTAATA			
TTGGAGG			
TATCGAC			
TGGGGCC			
TCAACGA			
CTCTAAA			
ACTAAAT			
TAGGTTT			
TTTATTA			
GTAGATC			
TGTGCAA			
AAACTAT			
AACTCTGG			

				AAACAATA TGTCAAA					

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

[illegible]

540

ACAACTG
CCTCAGT
CCTCCGC
ATCCGGC
CTTACT
CTTGTA
CGGCAGA
TTCAAGA
GTATGTC
GGAAGAG
GCGATCT
AATCAG
GAGTATC
GAACGTC
TGCCACA
CTGGACG
GAACAAAG
AACACCC
GGCCAAA
AACAAAGA
AGAAATG
ACTACAG
AGTGTT
ACGTTC A
ACGTGAG
ATCGGTT
GCCACCG
ATGACCG
ACTTGCA
GAGCTGC
TAGAGGA
GACAAA
ACGATCA
AGTACGA
CATTATC
GGACTGT
CGGAAAC
GAAGAGA
GCAGCTG

				AAACACA C					
				TAATACTG CTGAACGG ACGTACCC AAGATCCTT TATCACTTC TACCCACCC TTTTTTTTT CTGCTCTCA TTCTTGATC TGTTTCTAC CCAACTCCT TATTGCTCT CATTATTTC CCTGCACAT ACCTTTAAT GTATCTGTG TACACAAAT ATCAACAC GTCACATTG GCGTAAAT AAATAAAT AAATAAAT AAA		VLVNRHEIQFLIDRAKPDILCLQETWLSEKTPSPTLVNPP SRYQIFRTDRAHEKRGGVCTIVKSEIPAKVITSETKNM AYEILALDINISTTPIRLINVYRPPTCSYANTLLTKISD MTNXNGQIILGDFNIKNIKWIPAGPMTKDRLGTLSDFC KNNLKKQFVNXPTRKNNILGIVFSNTKLENVEISXPIGSS DHRKITFTLDILEKPTKHLIEWDFKGINWKVCAEELNLI DWPTIIDPVKTISEIYNFELTIMRRICELAVPKKRINKELY NVPEYLVTFHKTIENKFKDLLKRSWADYALYIELSRKF RKLLKFRRNQELLTLKKSGNSHFTHIAKNILKPKFSMV PTFVDSQGHHAATSDHEKAEVLA VNFESQYVNKKSNTSFT LPPDYKSDQEIPVWVTD AEMLKFKIQCKNCSCTDPDKVPF KFLKLIAPFIASPLSQICNL SMMIRGEVPEKWQQSIITPINK IAKPQVPTDFRPISTSQICRIYERCILSKLLPFLDKIRFWN NNQHGRPKRSTVSCMLESNDWVTSIERGSGQVDVIYF DYAKAFDRVPHDLLDLKLVSIKLNGLLKWLASYSNR TFTVVKTNQSSTKMAQCQGVPGGAVLSPLFGIYVNEIP SKLPMGVHCKQFADDIKLYSEIPKSDPQNKKLQQAID AICQWSETKKLALNNSKTILISLGTTKNCDYVINGSPIG RKQLTRDLGEMISPKLDFSEHWHKATNQAKYALSLFK KYCSDNSRLLTLLYKIFIRPILEYGTVISNPHKSDIKSVE SVQNAFTRRLYSRQKGRYIRADDPEYLSAPERNLSYGLS SLENRRKYFDRKKFLT KMIAGNVVDIDTRPFTFFHQNTRTR AKTKFIWARSKTIRQEFFTNRTLTAITL	cd016 50	470	724
CR1- 13_C Jap	Caenorha bditis japonica	.		TAATTTTTT GTCTACATA GATTTTCCT AACATCTCT TTTTAACTT CTTTCGTTT GTTTCTAAA TCTCCGTC ATATCGTTT CTTATACGT GTGTAATA GTTTGTTAT TTATATATG TGCTGATGA		KILVSGSSKNFFKFINQRMKESDVGMMIDGKTXVYDN YRKAELFAESFGKVYSTDDGQYTTCDLRTSAHLSIEIFEP YAVELELSKLASKVNTSPDEIPAFILKKICTSIALPLSIIFN TSMRTGFLPDLWKTAVIKPLFKKGRSDANNYPISLTS TICKTMRMIRKEMMKHIYGNKLLSNAQYGRFPHRGTE AQLLHYISCIQSNMEKKKIDMSSVYVDFRKAFTVSLPK LQIKMESYGICGRLLNWLSSFLTDRHQSVVLNEVFSPRM VVKSGVPQGSVLGPVLFLLYVNDIGDSLECDTLLYADD LKIFSTDSAAALQDRDLDRISNWCRTWQMDVAPNKCEFTF NYSKKNLSALSPLMIGNENLQRTEIRDLGVFNSNLSFK SHISVIRKKAHQIRINILNIRLSTIDVFIKCYVIYVRPILE GSVVFSPISKELVKTIESLQKSFYRCYKKGIVYNSYFEA LEQNLETLERRRLMADLVFIYKWLVSREINIGHDLFISS	cd016 50	134	384
CR1		.							

					ATAAGTGG CAGCTGCC TTCAATCA TAAACAAA TACAACAA ATA	ANLLPLRRHNYIRQLTGATRIGSQFLSNRIICWNSLP VKTFPVKFSSTAFKNRLQNLSNLFLLHFNHF			
CR1	CR1- 8_CR em	.	Caenorha bditis remanei		TAGAGCTTA TTCTTATTC TGTAATCCC AAACCAAC TTTCCCGC ATTACCACA TTGCTCAAT TCCTAATGC CTTCTTTTA TTCCCTCA TACCGGTCA TTCTTTTTT TTTGTGGC CCACATCTC TCACCTTAA CCCACCTT TGAATGGTC TCGTGAGG GACTCATTC TCAGCTCC CGCCCCATA TATTGATAA TTTTTTATA AGAATAAA ATACAAAT ACAAAATA CA	VIPIPKGSLSPENXRPISLTDPFARLFERFVICEYIKLHFA HKFSQNHGFLAYRSCSTSSLVHSISCYKSSLSNNSLDVI FFDFKAPDKVNHKLLQKLALFGPHLFIEWFSNLSG RTFSIKIEDFTDTSITQIPSGVPQSGVSGPLLFIFINDLLD LALPSLQVSAFADDIKIYSSNPVAVQKGIDLIETWASSN SLPLAHTKTSLLRLGSKNISFPYFIAGQPIETSKSVRDGL ITDSTLKFKSHINKTASALLRTKQLLKSKFSTSPQFYIFL FNCYVLPHEYCSVVYSPPPASKLSLSLETPLRFFTRKIFQ RCNITYSSYSDRLAQNLFLSLRHRRLKAQLLLLYKFLSG TSYFPHLDSYIRFSSSTRRPMNLICIKPKCSDFFSHTPIWN AITSQSSYFLSPSEFNTLISSITRY	cd016 50	1	246
CR1	CR1- 5_CR em	.	Caenorha bditis remanei		TAAATTATA ACTGTTCT TTTTTTCT ATTGTATA AGTGGTGA AAGTCCTC AAATAAAT	ELVTTNNFVCLTETKLDNTTDPDSSLSSNNFVVRKD RNKHGGVAILSKSIRFLPIEPSSLSSEIAGVDIMAGG VSIRVVGYPHSHSKLDGMISCLFLLSTRKNCVILGDF NMPHISWPSLTASDSHCKKFLSFVTKNGLTQHIASPTRL KPDNILDLCFTNTDILRDVRVGDLSFDHKLHIVTLVKVN RTKKLTKEVKLFRKADYASINYLNTDWDVLRFSNLNA	cd016 50	484	735

				AAATAAAT C	DGMYENLLSILHELJASYVPVKAINTLSKRHS AEILKLQK AKLNVWRKEGNSTNYKNISADLKV ALIKEERRVNDKEL TNGSVKDFEFKINSRYKDNQEI GTLKNDSGAPINCDSEK VELFSDSFVKFTEDNNVQPHFDKRTEELVSSPDEFYPII EHTLSKLTPKLNTPDGPALFLKNVCTAIALPLSIIPRESF RTSIVPTAWKTAIVKPLHKGSRA NPNNYRPSLTSVVG KVMKLVRLQLTNYLNSNRLSNCOYGFRSSMSTEAOQL LSYQADIL TNYICKTTHSVYIDFKKAFDKVSLSKLKIKL RSYGHDDFFNWLC AFLTNRTQRCINNVSFSDRSVLSG VPQGSVLGPLLLFLFLINDJGDAFESNYLLYADDLKLFSTN ADCIKDLVRLSVWCDNWQMGVAPEKCEVISFNHSHKK HSNSASLNFINDAIIPQTKIIRDLGIHFNLDLNFSLDVT LRKAHQRVNIFFNVLRHADFEIFIKCFKIYVRPLLEYGST VFSPTSKEQVRLIESVQKTFIFRVFRKFNIPYISYFESIKHC DILSLEHRRIIDLVFMYKILVSKEVRIYNDVHISLPRFSN NLRHPYVIKSKLSNSTKNTCQFFTNRIISWCWNKLPPHIF PSFPNSDVFRNCISFNVEPFLILKHSNF			
					MLDTADSGRPERRSALIAHEL SRLNIDIAALSEVHLHEE GSLKEHGASYTLYWSGKPKTERHLSGVGFMKNSIASK LENLXTGHSDRIISRLPLHNKQHVVLFSIYAPTLQADH AEVKFYTDLRRLTQKVPAGDKIILGDFNARVGKNEY AWKVLGKHGVGNCNDNGRLLLEFCAEQQLTTINTIFQ QKDSLKTTWMHPRSKHWHLXDYVLRQRNVGDVRHT RAMPSAECHTDHRLVRCKLTLYFKPKRGGIPRRRLK VNDLQKDTVNSFQVNLQARLLDHPIDPSPEALWQHIIK NSILQSSEESLGFSSKKNKDWFDENNQEIQELLTKKRTA HQAHLAQPSCHVRKA AFHLACSKLQKLRDIQNKWWI NLAEKTELCAVTGDHRGLYEV LKAVYGPTHQVQSPLLS ADGQMLLTDKTSILNRWSEHFQTLFSASCVVQGS AIQHI TQQPVKTELDIAPTLGETLKA IQQLKTGKAAGVDGIPSE VWKHGGQALHAKFHELLVSCWEQGGKLPDLRDAVIITL YKKKGEKSNCSNYQGITLLSIAGKILARILLNRLVPAIAE EVLPEQCQGFANRSTTDMFVLRQLQEKCREQNKGLY VTFIDLTKAFDTSVRKGLWQILERLGCPPKFLKMILLHE DQQGVRYGDALSESPFIINGVKQGCVLTPTLFTIFFSTM LQRATADLDEENGIIYIRXYTDGSLFNRLRLKAHTKTLN HHVPELLFADDAALVAHTEAALQRITSCFAEAADLFGL EVXLLKTEVLYQPAPQEDFHHPHITIGESLKS VQQFSY LGCIISSDAKINKEIDNRLAKAYRAFGRHLHKRVWSNKH KKSTKISVYRAIVLSTLLYGSESVVIYRHHLLRLERFHQ	cd016 50	531	809
RTE	AviR TE_C An		Calypte anna	G	CCCTCCA CGGGCAC CCCCTTG TCGTGGG GGAGGAG CTTGCGT GCCCTTG TGAGGTT GGGAGCT CTGCTGG CGGTGGC ATATGCC TCCAGTA GGGTCTC CCATGCC AGACAGG TCACAAC TGAAGGG TCAGACA AAGTGTG TCCATGG GCAGGAT GGGCTCA			

545

ACAGGCG
AGGWC
GAAACGG
CAGGGGA
AAGGTGC
ACTGGAA
GCCGCAG
ACCCAAC
CCTGCAT
GCAGGCG
GCTCAGG
ACATTGG
TCGCTCG
AGACTGA
CCCAGGA
GATGACA
GCTTGIT
GTCGAGC
ATCTGA
GTGACCA
AGCAGCC
TTTTTCAA
GGAGAGC
ACTGCTT
GCTCCAC
ATGGAGA
GGGGGCT
AGAAAAAG
GTGGCTT
AACCAAA
GCTCATC
TCCACCC
CCCCCCC
GGTTGGC
TASCCAC
GGTCAAC
GGGGCATC
TCTTGAT
GCATGCG
ACACAAAG

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

ATCATTCAA TTCTCACTC ATTGGTGG CATACATAT GTTACGTTC AGTTATGA ACTTCGAT GTCTCCTGA CCCTGTGCG AATAAATA ATCTATCTA TCTATCT	GAACAGG AGAAAGA TGCCATG AGAAAGA GAAATGT CATCATA CACGGAC TGCCAGA AAAGAAAG ACCAACA ACAGAGA GGAGAGA AAGAAAG AAGATGA GCAACAC AATCGAG AGGATTA TGGACGC ATTGGAA CTGAACC TAGAAAT TGAAAC TTCTACC GACTGGG AAAANTA CCCAATG AAGACGC CGACAGG	CR1- 6_CT e	Capitella teleta			MTMYIKDHIDYTMLDIGEDTCEVQAARLLMEQQQATIF ASVYRSPNSSKENSNDINAFIRNFSNERAYTLIAGDFNYP TIGWKMTGRTAEADIEFIEALRDGFLIQHVDRPTRGRGS NNPSLLDLVLTKEHTATPTVEYHQPLGKSDHSILDIRIEM KQSETFIKTIPNYNRGNFDRMRRLSDVDWEEGLGREE TTQKKWELFHETTNAVNRCPKIKRKLKNIPVDNK TLTKIRRKRLWKQYLETGNQTYHDYCRTRNQVRQLT RKNMRELEKNIAKEAKTNPKFWAYANQKTKYRETIP QLLKTGNPNDGMLTENNQEKAQTLATFNSVFSEKPN HWNLSPPSYPTIEQDLITEASVLHHLKSKKTKSPGPDG IHPRILIEIKAIKPLSIYQHSLDQGGQLPDTWKTANTAI HKKGDKRIAGNYRPVSLTSVPCKIMEKIIRKSIMDHLIRN NILSNKQYGFITGRSTLLQLLTVLDQWTEILDRGSEIDVI FLDFKKAQFDRVPHQRLIDKLEYINGKILTWISAFILNR KQVTHEATSAWAEVKSQIPQGSVLGPTLFIHINSLPDV TSSPTFLFADDAKIFREITKPTDSLQEDINNMHWLLE SLLTFNPDKSSMTISRSGTNHRQYQMNNSPLNKTIERD LGVLNDNRLTFDHHIQAKIKKNSIMALIRRTYTYLDNHT FLLYKSFVRPHLEYANQVWKPHLQKHITAIENVQRRRA TRILPGMKDLNVEQRLRLKLPTLQYRRIRGDLIETYKI ATEYDYDTQITQNFQFNRSRISRGHPLKLYHQPARTELRK NSPKIRIAQLWNQLPSEIVASPSIATFERLDHHLKNHPK HNPRRP	cd016 50	423	675
--	--	-------------------	---------------------	--	--	---	-------------	-----	-----

GAACATA
CTACTAG
CAGACCC
CTGAAGT
TGTGCTT
CAATAGT
GAAGACA
AGGTGAA
GATGTTT
ATGGGGA
GGCTGAAA
CAAAATTA
CAGAGATG
CAGAGCA
AGATATC
AAACGAA
TTAGAGT
TCAACAC
CAACATGA
ATTACAA
AGAAAGA
GCAGAGA
TTAGAGA
ACTGGTA
CAGGAAG
CAAAAAA
CTTAACA
TTAACAG
AGAGTGG
AGACTTT
GTACACG
TCGTGAG
AGGAAAA
AAGATAA
TCAGAGT
GAAAGCC
AGAAAGA
ACGGTCA
GAGCAGG
AGAGAGA

550

[illegible]

				ATAGACA AG			<p>QITCPDIKAITPNQLENTPYWLRRPELDISLIHF GKKTAN PNYTLRNEFYNLMDKYPDHKVIFTDGSKSDSAVACSAT ADNLRQIRLPDSASIFSAELLAIYQVLTLECSANDQQQ FLIATDSLSSLQAIGNFNKHPYVFKILEKCTLLHKKGIDL VMAWCPSHVGVMGNERADLLAKEALSFTTCTIRIPSSD FKPTHEFYKEK WQEQWSEQENKLYCIQPTLGKWA KS SREIRREEIVLARARIGHSHLTHGYLLRREMPVVCIPQCSI LSVKHIFIECVDFDILRRRFFNVSSMK</p>			
				<p>TGATTTTAC CCTAATTTC CTTGACACA TATTCTTTG TTGCCATTG TATTTTTT AAGCTGGC AGTCTTAT AGGTTTCC TGTTCTGCC ATGCTGTTG CATGCATTG ATTTTATTG TACATATTA TTTTTG CAT GCATGAAG CTTAATAAA TAAAAAAA TA</p>			<p>LD CPOIPDPHFTVENVLSVLASLDTNTAFSPQDIPPIILKK CRSLAPSLCHIEFVNSIRCKTVPLEWKKARVVPVFKKGD KSAVSNYRPSILCIVSKVMERCVENHMYPLSSFIHPLQ HGFCGRSCTSQLLDVYNRIGSILDKGGQIDVVFLDFSK AFDCVSHKLLLYKLKHFYGVDCNLLLEWIRSYLSERSQS VIVEGKSTWLGVP SGVPQGSILGPLLELFLINDMPYKTL HCTTALFADDSKCFKSINTHTDCELLQRDLNAMHEWSL IWQMSFNPSCKVLSTIRSHAPIHFENYAMNGVILDHVN FKDLGVFIDQTL SFSTHIESTILCKNKMCGIVKRTIGENAP VNVKSSLFMSLCRSHLDFASPVVSPHLKSKVKQIESVQR SMTRFILNADLSYIERCQFLEILPLSFRREITDILLFKY FNGAMNVDFTSIVAKRCFNTRLRS GPFIELLTPTRTDSF INFYFN RV AHMWNILPLEIRSCLSVSTFKSKLLSFYSNK MCFYDVNNSCWTWTSTCRNCGFYHT</p>	cd016 50	69	322
				<p>AGATCAG GTAGCAC CCTCCGG TAGCAC TAAAAAT TGCTCCG ACCAATT TGCTGTTT TCCGATA TCCTCCA CTCCGCG GATTGCA GCGAATC</p>			<p>MWKVIFVICVLSQIKAKTSGGTRPGFTTICLPQRHASPT SMPSLCLEFIRALAVKAPARLPTDLYKRLNGLRISRRRLS KRGTKGYGRGVPVQVSALRSADRYVNTARRDASLPTS LVTIEMTIASSSCVLLNCRSICNKA SEIRDFISDRRMDF LCLTETWLTGTDEDKVVS AALLPDDYAMVHRPRDARG GGVAFHRRSITLSQIRLDYSSFEVLDALVCTARRLM CVVYRPPSRNNLTVAQFMSEFAFDLSIVSLPGEPLIV GDFNPHLDDCGDADASTFSDLLTSFGLKQHVRGP THEK GHTLDLVITKTDALIWSSVSRDLCISDHYVFTTLAIPK VLNPTREISYRIKAIDKEQLDRLEASSLSPPENASLDE NVTNYQRTILGILDTIAPLRSTRVIRV NADWYNDEIRA EKQRRRQAERWRKSKLPADRELFKVAKNRTQSMIDE AKSLHFREKVSACSDDPKQLFRVVDTLTLLGNSTESPLPVH</p>	cd016 50	620	878
CR1	CR1- 13_C Te	Capitella teleta	.							
L2	L2- 9_CT e	Capitella teleta	.							

				TTGGTCA AGAC	DSRVELSNRFSDFVEKIAITQSAIPSVNDNPPRQAPLPQS QLATFQPLDVADTCALISGAAPKSCELDPLPSPLVKDLPS LRPIITRIINQSLATGVFPMEYKSAVVRPLLKKPSLDRDS LKNYRPVSNLSFVSKLIERAVASQLASHLRNNDLMEPR QSA YRPGHSTETALLCHINEIVCAIGEGKAVLLVLLDLS AAFDTVHHPTMLSSLERLGVGTGPLKWFESYLSERSQA VQIGDAKSDPAAINCGVPGQSVLGPILFTIYTSSLGALLR EQGVSYHL YADDSQ LLLTCDPRDPAPA VARMEACIRSV QAWMSCHSLKMNGDKTEILLITSKPMRKVLDPVQLCID GHQLQPSITTARNLGVTFNRHASMESHITSVCRLSYMQL RRLRIKKLLPRDVLESVHAFVTSRLDYCNLSYLGIPQ TQLDRLQRIQNSAARLVTD SGAREHITPVLQDLHWLPV KERVIFKVLIVFKSVHNLTPQYLSEMITVHIPTRLRSS NQNFLTVPFTRSSFINKSAFSAAGPTIFNSLPINIRQAESV PIFKRLKTHLPHNYFLT				
				GGTGAC AAACTTC GTCATCA AACACAC TTCGTGC ACAGCTG ATGTTGA GATCCTA TGTGTGC AGTGTCG ACCATTT TATCTAC CGCGTGA GATCAGC TGTGTGC TCTTCATC GTGATCT ACATCCC TCCCAGT GGCGATG CCAACAG AGCCACG GAAGCGA TAGCGTC TGTGTGCTT	TAATGTTAT TATCATCAT TAATGGCTT AATTGTCAT ATTTCATCAT TCATTTTAC TCATTATTA TCTGTTTGG CAAAATCA TTCCTTCGG GCGAATAA AATCATCTT ATCTTATCT CATCTCATC TCA	MIYCLPTYIRKLERQKPMIKTIRQWTEDAAEQLGGCFAC TDWNIFEDTTTNNINEYTNVITDYIKFCEETIPTKEIKIYPN NRPWVTNDLIH A VRRKRRLWMSGHHEECKEAQKEVNT IIAQCKDNYKKKIESFRSGNPKKTWSGLNTITGYKKKP MTRTSISNDEGWANELNVFYTRFEKEDEDLVFDQPAIAP IRFEEDVRRVFMGKQDKAQGPDGLSPRLLKACAQDL APVFTGLFNRSLSESCVPTLWKCSTIVPVKRPRTAKELN DFRPVALTSIVMKSMERLVLRRLLHQVQLSMDSNQFAY KARRCVEDASALMVHLVSEHLETPGSYARVLFVDFSSA FNTMRPSILISKLEAQGVAPPLCQWILSYLKNRPQQVRV NQSLSTIRHTNIGAPQGCVLSPVLFITYTDEHRATSQETR VIKYADDTALVALLSPGKEHL YTD AVSHFIDQCDNDDL KLNSTKTEKMCIDFRTHNTIITPIIKGQPIEQVTYKYLGI TIQSDLRWTSIHSTQCKKASARLYHLRKLRFHVDRHL QELFFATIDSLLFGITVWGGRCTARDNNISRVQRIAG KIMNTTVQPWGVSHRQRTIMRAKKIIREPLHPLHYTFQR LPSDRRYRALKAKTNRLRQTFVPSAIRLINSQPI	cd016 50	257	510
Rex 1	Rex1- 1_CT e	Capitella teleta	.						

L2-3_CTe				TGGACAT ACAACAA ACAAAGC CAGAGGC AGCTATC GTCATCA CAGGAGA CTTTAAT GGAGCTT CGTTACA TGACGCC CTCCCGA CCTACGT ACAATAC GTAAATT TCAACAC AAGAGGT CGCAGCT GTCTCGA TCTACTG TACAGCA ACATAAA AGATGCC TACAAGA CCACAAG CCTGCCA CCCATCG GCAGATC TGACCAC ACG	TAATGATTT TCCTTTT CCTCTGTT GTACAGCG CATTTGAAT GTCTTTGA TATTGCTGA AAATGCGTT TTAAATGTA ATAAATGA	MSLTHEHLSIKVIDVDRSVTSDFGVVFTINCPKPAREK QKLKVRKWKNVNISALDETIASHIPAIESAQDPFAAYNN CLTQVVDEFAPETITIEVTKRDTWPYNAELRHLKTECR RLERKWRRTQLKIDQEIYRKRCRLRLDRDKMTYISS KLQNATSTKDYNILNQLLHKQSSPLPRHDDDEIELAETF SSFFSEKIERVRSIGAHPDPPDV/TIEHHLSITEITPLTSFRPV SRDWISKTLSSITTPKSCNLDPIPTALLRKPCPSAVPLLCFTI NWSLANSTMPPDSHKTAHVVRPRLKKTSLDHDQLKNYRP VSNLSFISKMTEKAIV AQLKDHC FVNNIDVQM QSAYRA GHSTETAMVRVONDL LD DAVDRDGGAILVL L DL SAAFD	cd016 50	306	560
----------	--	--	--	--	--	--	-------------	-----	-----

					AATGAAAT GAAAT	TLDPHLILHALKYNIGVTDLSALKWFESYLTNRKQLVTIG SQTSDTKNLVYGVPGQSVLGPILFTMYTRPLSNISNCG MNHHL YADDTQLYIAVNVRSPTSVSSVKEISQECNQV HRWMTANFLKVNEEKTEVLVITKPSL TEIAPSSFNIGETI TSSASVRNLGVMLDNKLSMADQVRNICQKSFHHLHMIR KIRNFITTEDAARQLVHANVMSYIDYCNALLAGSPSYLID RLQHIONCAARLVKCADRHSPSLPLLKELHWPPIKFRIM FKINVITFKALRGLAPPYICSLINVYTPARSLSRNSNFLLS VPKFHLQSYGYRTFAHQAPILWNRLPSNLQNQQSLDAF KAQLKTYFYFNIAFSSI			
					TGACATGG CTTTGTGAA GCGCATAG AAATGCAT CATGCATCA TGTTGCGCT ATATAAAA TAATAAAT CAATTCAAT TCAATTCAA TTC	DDLNDLVSSFNRRLCDVTDNAHAPLRQITLKGDMKPW YSDAIHSARVVRRKYERQKRKSPLEVHRQIYVDQCKEV VQLIQNTKKEYFHQKFASASAKEVRLVDNLLHKEPNH TLPTNVPLGDLPQJTNKFFYDKVHOIRAEALDASPTLPSLT TPQPLVAERGGAPALETFEICSADLRKIIQTSQSKTCMQ DSIPTPLLKEGAVLEAALPAIVNIINRSITEGCVPAELKKA VIAPILKKTGLDPNNLAHYRPVSNIGFIKVMKEVIAHQ LTDHLTRHNLHDQLQSAYRKGCSETSLLYIKSDIELILD AGDAVLMILLDLAAFDTIDHALLTRLSEIGLEGSALA WVKSYLMERTQVVVRIGDKIESSDLDIGVPQGSVLGPLL FLYMIPLGSIIDNRGVLRHSYADDTQLYCRVNIKDPTKL QKQIRDMEECITSVRSWMKSNKLKLNDAKTEVLLISRH EPRQVPEIRVGNVAVIKPAPLVKNLGAYLDCRLSMKDQV NRTTSAVYHQLQRIAKVRPFLSDSACANAINATVTSRLD FQNALLMGLPDCQLRKLQVAQNNAARLLTRTKKREHIS PVLTAHLWLPKHRIEFKVLSMIHKSFHLSAPDYLKKL CRPYVPSRNLRSDDPFKVVQLRTRNLYGARTLGGGLGA RLWNALPMSLRKPMFTFTFKKKNLKTTHFFRISQ	cd016 50	233	497
L2- 13_C Te			Capitella teleta	C	TAGTATCAT CCTCGATCT ACAACTAA CATATGAA ACGAACTA TTTCACTCT GAATATGC GCTGCACC ACCCACTAT TTTTTCAAT TTTTTTTCT CTCCGCACC	KMFLDLMHSHAFFXLIDKPTRITATTTSLDNIFFSNVINX THLSGVLYTDISDHLPLFTVNQFICRNNISIVTKTIGYRSF NETNKLAFQNAPLAIDXDXLFXCHNCGAAATLFTSNFV DCYNESFPIKSHKINPKRNQPWLSPELKRINRNKRLYR RYHARPSVFNENYKSYKSMITKEIRVAKRTYYHDLLEQ NRNKIXKTWDIIEKEVIGIDNTKKPIASITINDSTVSEPIEA QHFNKYFSSIGNNLASSLPINHLDPMRYMNPNTPRSLFL SPVTSNELYTIIFNLKDSAGIDDIKPIIIEFSQLIIPKLLHI FNLSFTQGSVPCAMKKANVTPIFGGDPSTSLGNYPISV LPVFAKILEKLMYNRVYNLTENRILFNKQFGFRRGYST EMALLTALDELSALDMKRNVGIVYDLRKAFTDVNID ILLAKLFKYGIRGIAHQWFTSYLNDRSQRVKIGCSFSEM	cd016 50	337	594
L2									
Daphne									
Daphne									

					AAAGCTGA ATAAAATTT TAAATTGCA CAATTGAA CAGGCGAC TCACAAGT AGCCATACT TAGTCGCCT ATTGACATG ACCTATTTC TTTTTCGCT TCTGTATTT CCATTTCCTC TAGTATGTA TTTCCATAT TTCAATTGTC GCTTCCTGC TTTCTTTAC CAACCTCCC TTTTTCTAT TTTGGCATC TTATATTTG CTAATATTT CATTCCTTC TGTCGTGTC AATAAATG CAATTCAAT TCAATTCAA TTCAATTCA ATTCAA	KSISCGVPQGSTLGPLLLFLLYINDLPNVLTNIPPIIFADDTSLFMTGVINDMTAEAFNHQLSKINNEWLIANKLSLNQKTY SMLFTLSPHLYSSPITNTNTEIQRVHSTRFLGVIIDDRLT WANHITHVSNKIAKSIGIINKVKHSLKQDTLMILYYSLIY PYLTYCFLLWGKASHSHINKLTILQKRVRITITNSGFRDH TRPLENTLRHMDLFTYLCVYIFKCIKIFPRSFLLDILN YFTDYNSENAPRSSTRASLTRSIIVPLCRTTLRQKQLRFQSI KLYNNFLMPLEILDNSNSISHFRRLKNVLS				
						KRNQLTRATRRAKIKYYSDKINESASPKEIFQVTSFLLQSKREVLPKATTMSSGSLADEFAEFFENKITAIRNGISAAP VQSSPLSDVV ALHTFALTNAAEVSTLIRKSTSKSSAVDPI PTWLPKKCILTFSPILAAIVSASLVSGCVDPQLKAAVVTP LLKRPTLDPEVYKNYRPVSNLSYVSKLIERVVAERLKEH CALNNIQCKFQSA YKQNHSTETALIRVHNDILRAVDRD GGAILVLLDLSSAFDTIDHTLLHLEHQMAIGDTALKW FSSYLSERTQRVKVKESLSSPRRMLYGV PQGSVLGPVLF TLYTSPISAIRAQQLQYHLYADDTQLYMSFTPTSAASTT SSGASMESAVSAIKRWMDHDLKLNEDKTEMLVIQRP	cd016 50	155	425	
L2	L2- 11_C Te	Capitella teleta	.			TAAATAAA TAAATATA AATTAATTA A				

Prot 02	Proto 2- 1_CT e				<p>GCATCAA GGAAGGT GTGCAAG GGAGGAC GAATATT GTTGATG GTCGAGT AGGGGAA TCTGAGA TCTGTGA TGTGTTT CATGATG GCTATGA GCATCTG TATCAGT CTGTGCC TTCGTAT CAGAGTG AACTCGA GGAACATC CAGCGCG AACTTTC TGCGGGA GTAGATC GCACATG TAAAGTA GGCTTAT GTTACTG TGATCAC GTGATCA GCCCCAG TGATATT GAGTTTG</p>	<p>Capitella teleta</p>		<p>NPSSQPTHRPTLTLDGNINNPSTSRNLGVLFNSAFNLE DHVNALCRTMFFHVNRIWRIREYLTTSATRALMQAHVI SRLDYCNALFYGLPVVVRNLQKAQNAARVVVRKR SDHITPILLQLHWLPVEFRVQKICLMTYKALHGLAPAY IDELISVHRPTRLRSSDIRISPVQYKLKSVGGRCTSSAA PSLWNELPQSIRLSSHLLLEYKSKLKTHTFRKAFNDIYLNK STSYHFYIYILLYIAFEQFYFFLNF AIHKFEVINK</p>		<p>MSDHFHIGCPELYVHLSLFLNAMLHLHLDVSSVMSRSVLV PIPKNRKKSMNSASNYRSIAISSVKGKCLDNVILVKHADV LSSCDLQFGFKRTHSTTQCTFVLNEVDFYMRGSSVY VTLLDATKAFDRVHYTKLFRLLVDRSMCNATAKLLLC MYVNQSLVVRWQSVTSDSPVTNGVKQGGVLSPLMFC VYIETLLERLRSGIGCYIGSTFAGAIGYADDVALLAPSL ITAQKMLNVCEDSFREFVQFNSSKSKAIVENAPPNAV SMRKLTLNGSPIDYVADALHLGHVVGDKSASLNMENA CRDLVVRVNMLCSNFVHLDITLLFNSYCTSYGCSF WDFSPRTINGFVVCWRKCVRRLLDLPYTRSKXHLHV LKPEPRHQLFREFQSFYVGCCTSTNHIIVSLCTRLLPFSSS VACNLRQLMSFVNINYDQNNLIDNHSIRKRLSEFCEAT TDQESLCCASSVIELMKIREGSLISPINPTEATDLLKHM CSND</p>	cd016 50	35	295
------------	--------------------------	--	--	--	---	-----------------------------	--	---	--	---	-------------	----	-----

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

				CTCGTTTTT AAATTTTTT TGTCGCTCT CTCTTTGTG CAAGACTC GTTAAAACT ATTATTTTC TCTCTTCCT TTTCTATGT ACTTTTCAT GTGGCGCA AAAGCGAT CACGCTTCT GCCACTAC GTTATTTTC GCACCTTTT ATGTTGTGA TTGAATAA AATTATTAT TATTATAAA AAA	KHVMGLFLDLKKAFTVDNFILMKKLVHYGIRGNALD LLKNYLSNRKQSVKIENSVSSELPTVTCGV PQGSILGPLLF LIYNDLPNAXTSCNPIMY ADDTNLFLSGKNIDEMENKF NTDLVALTKYLCNRLSLNVSKTHSMLFTLNLSLRDSS LSLSIDGSIIDTVKCNFLGVKIDNKNLNFSEHITHCNKVS KSGIIRKISKTVNRPTLLMLYNSLILPYLYTCNMIWGS ANIHINRLFILQKKALRLINKAPLLTHTAPLFHECSXLNIH DLTYRSTIFIEKLSHHMLPTPVSKFFNLTPLVHGLNTRR QLNYKVIVPFCRTSLGQKSLNFENIAKL YNNILQPLSLLNL DSLEKFKKSIHCILVAKYL			
				TAAATTTCA CATAATTTT ATATTTCAA TATGTTTTT CGGACTATT CTATGTGAA GCGCATAG CAGTGTTTT ACCTTTTGC GCTATATAA ATAAATAA ATTGATTGA TTGATTGAT TATAA	TLKIAQVKPRPKPGVNTMDPKLFRPISNLPTEKLEKC VIRDLHLYIQQHNIIDPLQSA YRPGHSTETAMLRIQSDIK TSLDEGNGVLLVLLDLAAFDLDHNILLSRLQNSAGIT GV AIDWMKSYLSRQQFVQIEEKRSVTHDLDIGVPQGS CLGPLLFV MYTTLGAI FRKHGLSYHQYADDTQVYCSL PSDDQNSLKNVLTRVEKCLDEV KCWMTQNKLLNESK TECIIFSSKRNKQLFHQTTRIFGGSVITPVTSVRNLGAHL DCEMSMDKQINNITKSAY YHLRRIAKIRRHLDSSCTQA VFAVSSRLDYHNGLLAGITQKNINRLQKVQNHAARLV SQTAAHHCHITPVLADLHWLPIRARIDFKMLNAHKA VHV CNDAPKYMKELCMKYTPSRITLRSSDPHLLAVKKSQRN IGASTFDVYSCKTWNELPRDVRDIEETICFKKTLKTHLY KRCF	cd016 50	12	272
L2- 17_C Te		Capitella teleta	T	GCAACCT TTTTGTG GAAGCTA AGACCAC GGAAGGC	MACEDWGKNSLLKLGSENQPSVSDVSISAMQAPV AQS PVVSTARLSTLALPYAQVILNVPTAKASTRHPASVHPT SKPKKENMDSHNIWITNIRGLVSNFDSFRLEIEHAHPYL VIVTETFLSPDV PDQALLIPGYSLRRDRRTHGGV ALY YRSSRLRRLTSL ETQDHEVIVWFQASTREHDIIGCAAYW	cd016 50	587	851
CR1		CR1- 10_C Te	Capitella teleta					

CTCGACG	GACATGTA	PPGSDLDITQHLQHALLDEVSEIFNTVLLAGDLNAHHPD
AGCTTCG	CTGAATAA	LSEISINTAGRLLQRFAATVNALTQVIQAPTRVTAHSSC
AAACAAG	AGTTTCTA	LDILLTNIPEINFLGVYPGIGTSDHGVACVDLKTTSVITH
ATAGCCC	TTGAAAAA	TEDSATIRPRLHENTHSTNWKTVNDHFSNINWHHRFAA
CTCTCAC	AAAAA	CNVEETWTLFKDAYEEVLGRLSKKSRPTSRRTNHERHQ
AAGCGAT		HNTLNQELRLRATQQAARLHSTARTRETHQSLVTAR
CTCTTG		QEFILACANHIRRTDEATVCRMTESSNAKQWHKYCKQ
GGGAAGC		LYRGSAYKEAIPIMTVGTESYTTAKDKATILNQTFIAKSR
GGCTTCA		ASSRPRFSLKRRDSTLADILFNEYRVKKILQDLNINKA
TTGCGCC		SGPDGIQPSLTKNCSDSLHVPLAILFTKSFTEGVVPTWEK
TGATGGA		LAAVVPVYKNKGQRIDPNQYRPISLTSCVGVKVMKIIN
CGACAGC		DALLQHLMGNELIVDNQFGFLPHHSTTDDQIALLLHELLA
AAGACCA		ATTTKKSVAACFLDLAAAFDSVPHPAILQKLPAYGVRG
GGCCAAA		RLFRWIESYLSFRITQYVVVDQSSPIAQVSSGVPPQGSVIA
ACACGTC		PTLFLFINDLCHLSANTSLSHPDCHPEALVYADDTML
ATCATCA		YRTGKDPALICPAIEEDLRIAARKWADIWGMRFNASKTV
GAAATGT		AMYISRTTATPPPIITFAGATLEYSHHRHLGFILDSALS
TCCTGCC		HAHVNALTRKGATEVFLRRRLSYKVKDRDILLKIYKM
CACATATG		YVRPHLEYASPAWAALTVTQTGLLESQRRAMRILALP
ACGCGGA		RNHITDADYAAALNITSLRHRRNFALACFMFKLNSNRLP
CACCATC		RKLSKYRPIHTNPYNTRNCVLRAPYVPYPCRLRLDRSP
CTGGAGG		FLFGLKILNFISHTHSLPDISSFKSSSLRESSYIMPLAQIY
AGGCCCG		Y
AGTCGAA		
TACAGCA		
GCGCCTC		
ATCCGTC		
ATTAACT		
TGACCAA		
GCCAGGG		
AGCACCG		
CCAGAAG		
GAGGCC		
ATAAAGC		
TGAGCCT		
TGACGAC		
AGCAATC		
AATTGGA		
GCGACTT		
CTCAACC		

Dap hne	Daph- ne- 4_CT e	.	Capitella teleta	CGATGTA CTCTTTGC GAATTCT TTTCTGCC AACAAATC AACAAAGC CAACTAG AATCACA GGGTCGT CTGCGAC TATTATT GATAACA TTTTTCATC AATAATC TAACAAG AGCATCA ACTACAT CCGGAAT CCTTGCC AACTCAA TCTCTGA TCACTGT CCTGTAT TCTGCTTC ACACCAT TCCACAA TAATAAG AGCAACC AAAATAA CTTCATC ACAAAAG GCTATTTC ACTGACC GAAATAA AAATAAA TTCCACG ACCTCAT AAGCTCC ACGCAAT GGGATCA	TAAATCCT TCACAATCC CCCCTGCTA AAATTATCT ATCCCTAAC AGCTCATCT CCTAACGCT TTTGTGAAT ATCTGTATT GTAGTTTG TTTATTCGT TGTGTGTTT GTAAAATT GTAAATATC CATTGCGG GCACATAC CCATGTTCT GGGTCAGC CCCTCTCAC TCCACTTGT GTTTTTTTC TTTTCTTTT TGAATAAA CACATTTT AAGCTCC ACGCAAT ATTGA	MKNSIKMKNKLYCKYLKTKNPNDQKQYKSYKITLRSL YKAERTHYDQLFCRNKNLRSKWEVIKTVIHKQKNNN KQQYNMEVNGKKTSDDLITKHFNEYFINVGPNLKSIP YSNIDPISYITQTNRYSMFTKPTNETEIGNIIRALKNSAPG PDGIPSSILKENTAHFIQPLTHILNLSLQGHVPMEMKIAEI KPLKSGNKHLLVNNYRPSILLPAFSKIFEKCMASRLIEFT KHNLQYQYQGFREKHSNTMALHLLIDKITNAFDQNKV LVGVALDFRKAFTIDPSILISKLEKYGIRGCTLDWFRDY LHNRKQYVHLEGIKSSPLSIKCGVPQGSILGPLLLLYIN DLPQSSQLLPIHFADDTNVFYSGDDSEECIGHNQEMINLI KWIRSNKLSLNVKETQYIVFSKRKRIAPDNDVNINGERL NRVDSFKFLGVIDDRLSWKQHIAYIKTKVSKSIGILRQA RPVLNKTLLTVYYSFYIPYLYNCIDVWGHCSKQDLQSL FIVQKRAIRIUNYAERNAHTDQLFKSLLEILPLHHLIYFVSII FMHKFHHKRLPPVIDELFQTNASHISINTIRQRHQLHTPLL HTNISMKSIRYRGAFIWNQTFNEIDTNTLNRFKTCCLKNK LFTDFPIRTSPIQ	cd016 50	195	449
------------	---------------------------	---	---------------------	--	---	--	-------------	-----	-----

562

AAACCCC ATCCCTA CCTTTACT TGCACAA AGCACCT ATCTGCA CGTCCTC CACTCTG GTGAGTT CCTATCC AACTTTC TTCAATTCC TACCCAT CATGCGA GTATCTC ATACCTT CGCCATA CCTTACC TACCTGT ATCTGAC CCTAGCT GCGCGAA CAGAAAC CTCACTA CGAA	AAATAGTTT ACCACCTCG ACTCTCTGC TATCCTAGA CTCACATGC TCAACCTCA TCAGTACGC CACACGCTC CATCCTACC ATACCGTCC TCCCAGAAT AACCTCTCT AGCCACTG CAAAAAGT TTTTTATCT GAAGGTAT CAGAAATCT GGTCTCTCC TCAACCCCTG AAACTCGC GCACTTCCC CCTCGCCGC TTCTCCAAG CAACTAAA ACGAAAAT ATCTGAATG AACTCACC ACCTCTCTC TAACTTACA CATGTCCAT CACCATTTT TTTTCTCTC TCTCTCTCT CTCTCTCTC TCTCTCTCT CTCTCCTCC TCCTCCTCC TCCTCCTCC TCCATCTAC	KIERVTTTFKFLGITIDERL TWKAHIESLSKKLSSGGLYALR NTKRLTSRHGLLSLYHSLIHSHLTYGCLLWGSASDNAL KPLKIAQKKAIRLVYDAIV
---	---	--

ATGTGTTT GTTTTTTT TCTTTTGC TTGACAAAT TCCCCAAA AGTTTCAAT CTGCTTAT TTGGCCCC TGCCCCGT ACAGACGT TTGTGTC GCCGGGG GCTCAGGT CACGTTAA TGTCCAAAT TTATAATAC CCTGTAATA TTCTGCCCT GAAATAAA CATTCATC ATTCAATCA TTCA					KLGP TPAALAQDPPTNPTGGPPAGPNTTNSPSPTSANP GPAARSNTATRIVPKPAGPTKRPARRTPNPTPKLRKLAT GTPSAPPPAKRLHQTTIATSHKRRHTATLQNTPTPTNG QAQASPPPPDGRISILTLN'TRELLSHHESILHTV'THTPD ILFLTETKIPTRLHYSMGSAKKGLSSHYQTY'TSNAPYNP REGTRPHAGVMGV'THRFSRPSGTTDMNAHNTDSRLA GHVCHIRIAPHSTPVHAIGIYAPSDDATRAAIKYCTDQ AAKTAAKEHLIGGDWNAVLFPSPDRSTGQLNATDRA HIAFTTSAGLHPLTPPTHEHPRHPTYLQDQATAPHMSRI DDLVSHHHTHATITTTGGGGQEA VHECGDALDHRALLSS VPLAALRLYPFPTAPAAPSPDPAWVNYHPHTKQDQREKA RIKMEETIGQEIADLLNTLHPAHLAITNKLYPAGLTPAQ AERNTPSPAEFRALPEVSAADTT'SISNTLESILQKCLDVL KSYCDKKPPRASTTSPRHHTSKTVTRLLTATSKLKERLT AALTVTNAAAGATPPTTSPHPAPTTPPNPTTSPPTPPPAE ADTPTAIRSQLLNARTLLRAAKKEARKLNTKRAARAAR RLVATQPKRAHQKIFDETDQPRGTPALTDPTTKEVHTAP EKILSSIHOFHTAOLLPIGPRTGHDYDIPHSEKOYPPWSRP	cd016 50	107 786 4
---	--	--	--	--	---	-------------	-----------------

TATTTCAT TTCTATTC	DAPDRMPALSKPIDAPATLLPGIADSTTFFGCVNHLSSR KAPGPDGIPNELLQTIPTSLQTAIHQLMTLmwvKAHIPT HWTQSDTVLLPKKGDPLLIQNKRIALANTLYKLYTSLI TVNAALFSEQAHIPTESQEGFLRGRNTERQVQNNLLHAIE DAAATTGRDLFLYIDFTSAFNTHDKLLIVMHDLYPT DLIEVVRLYGKAATSVRTEHGTTPPIQIRGTVQGDVL SPLLFIIFIEPLLRWLHVGGRGYQYGCLTKAENDKHNLS SGAFADDLVTATTQISNLRIOVEKIEKYAAWGGKLKVV SKCQLQGILHGRARADPTLGGPRSKPHQAALARMLEGQ IRIDGTAIYKPPHESYKYLGIHITPLDWSAQVQYIREKI KKRGRIATSLATPDQKRLMIQLTHPAVAYSPLPMPYS THDV'TSLDSLJASTAKMCYGLKRGFPTRAILQTNALGL GVGSIRELYVRRAGKALVQALNDGGRLGAVTRAMLHL QHSVAQKIPADQLRTQCRFYTSRLQSLSLITDAGLELTQQ GSAYHTTLHGLSALLARTVPPGADAEQALPRLIQLLE LTSDFRELLTPDRTCVIDANTLKNTPWPAQRHRLALN RLTLALHAEITRAKTHNSTAPLTAQRITLDPDGHTFTGL HTAHTAPAAVDWGTNLNRFQRAPATGQTANAGETAA ATTAATAATAAEPEEDHHAAGENHAPAQTRRPGTQQRK RACIKMVRPEGRGPPADAPAFDWDPSLDPLTEAHC ADLIHTAPSLLGKSKPNPAAHLRGLLLGPTPPATFTHR HKLSEEQDAYLAATSDTWKHFESETYRTKALTTNTLY GQQERVQQVIAHTTSKGAPFYRVQWEPTVITPGSLTLY ARNRYRKPGGGLTWLLTPGNEPRVLVWEDLEEPAAIL DAQNKQELLDALNARLAVPLGPRPPPATDTHLSHDQ RQGRWTDASNDTARGQRTLRYVDINVEPCNPSRDHP TQAYTVQLGTVLDPGSPACGTHLHVYDPRGRCIGTISA EQLSLLWQRYQHAQTTPWKVHAKHDQGRGFAAIALR LLRYKPGGSKSHIPVTEKGECMLCEGLTQTLHTAIGSTT ELFASPLNVSDAATYYTLYEEDMLFGANYDAYSQPW DGVAVACPEPNTDGAHVRWALAWAEHTETLTNTRN PPTTATLLIPTQRNAPYTRYLAHPRVTLTLLGSISATTHA LLGQIHNDTGITLKHDLVLSNTPGAQLAQPIRDALGTS LYLSGMPVDPPQPLTWATHRWVTKQVRRMTRTRPAAP RVTPPPVPATPATLAPFAATQPLRLNPYEGFYTDGSVQK DALGQPSAWGAAIDARTLCYQIDVNPQHGCNTIIR CELSAIHQVLKLDVEVQALPEVTYTDSSAASLFLIRKTIHQ POLLTFSKHKALLDNILALLKRRSHAGRRTNLLKVKAH SGIAHNELADQAAVKTAGEQIPGAEWLVEESDNDPRS KYYWIRKIIKPTVEGPREKNEGHYAHDINRGIAJTLAAP
----------------------	--

[illegible]

[illegible]

GTGCAGC	FLVESFVHIGRKPMTPDLWQAVKGDIFCLHPYHLDMP
GCGAGGG	DPRTYPTIPNLPPPPQPPPPNQTKPKRGPTQALSTMIF
GATGGGC	ARQTTQLPTNHSAGDNAAPTHPGPAETTVTRPLTQATL
TCCTCGCC	NITREATEPARTQNERPATPNSAARTRRRHDNQLDP
TAGCCACT	EPTSPTRANTTTRPRANANMHAPNTNPQTANNTTQP
TACCCGAT	TWAQVVAHAEAAGNALRTVGRITTAAPHQRPETPPSA
AATCCACG	ARPTQRPRADPPDSPTHQTIVNPNTPRTPPGTPDGA
CCAGGCC	GAQDNTETRPDNTTWLHIAHNVRGLGNTLLSTICHLR
CATGGCC	EWNADIVILTETKLGPRTWLKDLRQEKLSYRTFTSTK
GGCGTTGA	PGTEQYKRRSAGVLMASKRYHSGSNVHTATIPPHLHG
ATTATTAT	YIVHCTIRTPHSTPLHILGVYAPEDMPTRRRIYKYCEETI
ATTATT	NTANKTDAHVIIGGFNAVLNEGDRTPPPDEADKHHAR
	FVSKNKLRLPEFGGTISMFSYQTRDGTQPTSRIDDLIC
	QNIHSALPTASRTPATLQVQPCGGLFDHYSLHARVPTFP
	RLWPSHTATQADTQPPETPGWRTVVLPIPKSTLEAART
	KIEHTLANPIATLASAIQPAATTAIQQLAQHTAGATNASE
	LQHALQTDPAIQTPDLDALANQLLEIQDGLSILVDMCP
	KKPPTRGVFAPRKINKRIRESHNEIKQLRTQCDALLQQQ
	HPYANITTTTAPNTGIPLPDPTTEQDAENPPQPPPAHD
	ATTAEPPQRRPPPATPAIANLIDKIKRAGDILRQILKK
	QTQHDIQKARDRFQHKLATRAKQAHKDFSTQEQRGIP
	ALHPQNAHCTDKASILDAIHAYFQGGASITGPRSGK
	YKPEDRTQVYPEDPSLKNAPDTFTLHSAATTPPTHALLPS
	IADSTNFDSCIDHLSRNKSPGDPGIPNELLRALPQTLRTT
	LHNLTLMWVRAHTPASWTASETVLLPKPGDPLLLKNY
	RPIALANTTYKLWTSITVAISDISHDLDFSETQEGFLR
	YRNTERTNLNVHALEDAGLTKQDIYVMYADFSAFNT
	ISHDRLEIMWDLGMPHDLIRVVQNLVSNARTYRSEHG
	LTAPIYIERGTVQGDTLSPVLVFEIPLRWLHVGGRGY
	SYGCLPPHLNARYNCSSAAYADDLAVLTNLSDLRIQC
	DKIHRYSEWAGLQVNHAKCRVTGILHRRAMQDKGLNG
	PTCNRTLRSMLENKIHGTTISIPYLPATEPKYLVGVQVTM
	TLNWGPQFAYLCKNITDKLSLQTCQASPEQCLRIKSCI
	QSVAAYSFVMPYETENDIRTLNMIARLARTCYRLSRG
	FPTRAVLTPAEMSGLVGSLPIYVQKATRALTLALNDP
	GRLGMVTHALLELQTRIAGTTDMHKLHSHYTYTLK
	QLSIMRAYNLTLWDHGAPYKGPNTLLTAISRSPHISH
	HLVLTQAGLDRLNLVGKERTPEGRACLIPSDILFQHL
	PRNQCHKHVRVALNRLSLCLSAECDGAREAPFNQAESY
	KCHLPLPQRYRALPNTPAFAALLPRNATPGELTTQSTMD

Daphne	Daphne	Ciona savignyi	TG	TAAATATTA TACCGTAAC TGTCCAATG	<p>AHLQHM TALHPNTEAQTPQHNPGLGPRQRQTTLAHML RTTGAAPTAPTQSPATIL TQTPTGTLPPLAQMPPLPT QTPPLTTPAGDPAPNTQTPPTV ASLTTAHQPKSKKQR PNAIKRLQKIDADPRHDPPLPPTDMRSHTEKSGFILNND PDENDSGPPNFV MWAPHYFANVQKDMQGTNRMPRLGI PRHRSRPQNRTRPNPTRTIPAHLADPYFLGSQVDDTNP LNPVTAYISANPPTTPPTCA YDSRPRWADTGDGFIAPEP PNWPNITGATPGSDHDPQADHPDQGMIRAEYMQERE RLRAHLDDHPQPQPRFKRWGIRLQLLTGKLPWPQTQAY HVAQA AKQL VYSLYHDIEEISHIQAGPITKAVTAPEKGA SLHTERY YEVAWKDITIVKGAAEAYTRMKYTGTA AAH MGYQACHKYFLRMKWDPTSEPVDKCPTLQGWTPAIS EHLRLAEPTPTPQPKPKDARLTPAQOQGRWQTQADVA ASHVRRTTKQNTRIDTNPCNPYLDVHPTGTRYTIQPDY TDERTLAHVYNPTGRHIASVTLGTLTLKARYEHALSTH HDATSPHRSQGGFPEDLAKLLLRTRVDSKGKCKSAPA RERSLPESKLALYNCLASTHSDPDRPVGLVERFASPLN CSGRATKYYSQHAADSLFGASHDAYTAPFDGLSLAHPD PDPDA TKALAWATASIEHLEHQATPAITFMIIPQKGGA ASRCLASPHVTILALIPPGVSASSMLDGLGADEAASSK QGLQLLIANTTGVQALNSNNLTRLQTS LAELGAVPPP TSQPPGRRQPLEWPARVTLPRDR TAPPGRPYSHRAKAR AKQLTRPQAPAAARGLDTCPTMPITLSALYTTRAAFPTRS SYLYDAATIVWTDGSCIKVAAGEDGNEMNKLGCACAWS QQTEHIYVEPGGLNSTNTIQRAELSAILVALEYIRPRLCL LGMTQVTIASDSLVS LYLRRLTAPQRLLSKHKNLLE NILHELQMYADAGVPVQLLKVRSHSTGLHGNEEADKGA AKVAANDMEAMGQAIRLAPDEPWSNIWWPKRNSDNF YLSDLNRGVLSL PPEAQOQGTNTTPVLEQWAEASAN MCPTLSNRITLASSTRHPWL VKQILYARYGYLFNAQLKR RYRLGGNGNCPLCHTPDSGGHILGGCTHPKMKGAYINR HNWAVQRVAKCISKGTGNGFMVMDAGPLADLPHYA SAQRPPAWLAAERTSSPLITRLRPDILFIPNLPAHETHRE GYRPPNLNKA RLPYILEIGYSSDVRLHDKREEKSSQQHQE LKRLEAAAGWTVHYSTNECLALGVTGAI PCNLTKLLTS QLGVTHTAARVCGERLHTHAVNSAGGIIMRRRVTEQEL KPRGG</p>	cd016 50	266	523
--------	--------	----------------	----	-------------------------------------	--	-------------	-----	-----

1_CS av	CCTTAGCAA TATTGTAAC TTCATTAAT TTAGATTTT CCATACCTC GATCTCTTT AACACATT GTAATCA CTGTTATAT ATTCAATTAT GCATTTTGT TTGTTTCT TTTTTTAGG GTTTGTGA AAGTTTTT TGTCATTGT AACGTTTT GCACAAAT TTTTACAC TGCTATTCT GTTATTTC TATCTTAT ATTAAAT TACATTTAT TATAAGGT GGCACCTG CTTCGATAA CCGCATGGT TTCTCGGTG TCTCCTTGT TAAACTCA GTTAAATTT ATTTCGTA TTGTTTAT ATTTGT TCTGTGCGT TTAACAAAT AAATATG TTTGATTGA	HEFSRSKNDIGRSWKJRELIIQIKSPTMTISKSVGITASDV TSDPINVANKFNDFVDIGKNLASKIKPTNHVPLKPRK SSSFISPTPDEVFSLNCLDVKKGTREDIPTKFIKLANS IISPISEIFNTCVKQGIYPDFLKNAEVIPIFKNGSDNLCTN YRPISLLSQFSKIFEKLLYNRLIKYLDKHKLLSNHQYGER ENCSTTLAVADITNMLLSNTDKGLYSCGIFIDLKAFDT VDHKKLLAKLEHCYGIRGVALDLFKNYLTJRSQYTFVH NVRSSIKPVLGVPQGSTLGPLLLMYINDPLVSNFVT KLFADDTFLCMSSKCPYELQAKVNLELIAIDDDWLRNKK LSLNYSKTSYVLTGNLKPIGINDFHLNINNKKIIRSQSA KYLGVLLDEKLNWNCHILNLSQKLARYCGIMFKLRKY VDIKTLLNIYSLVYSQIQYAIVCWGASNKKTLYPLNVI HNKIIRAMSYPARCHVTPLFSDLKILKLCDIYKLEMGG LMYKFNKTLTLDSENYIFQPLYCAHNHITRQVAHHDYF VPYFHLLGQNSIGYRGPVAVWTAIDVSLKKLGWFSFKK QLKNQYLLTYIT			
------------	---	--	--	--	--

			TTGATTGAT	
			TG	
CCAAAAT	TACTTCC	TAAGATCC	MAGHKITMSEGLLEVA	VRYGGVRNVSECPVPDCTK
AGCACT	AGCAGC	GCGGCTGT	TFSQANNLRHLNFGNTK	HRAHNFYFTCECKKIQIH
CCACAG	AGAGAC	GGCGCCGA	SNTKHNISNHYKQCCAT	GGGSPCETGQYFCPAEQAGL
AGACGAA	CGAAGAA	CTGCCCAT	GNLESALRHQSSHPEFL	PPRSQFSKSHPSNYTSLSLPK
CGAAGAA	AGAAGAC	CTTCTTGTA	DHLMKILYSGPLTPGQL	VCPIKICLRSSAARLFHDSVKLR
TTACGAA	ATAAGAA	GGGACTTTT	KHMLVDHNRTLVEYET	TCGCKLRPVDTSKNMRKTTSHF
ATAAGAA	TAAGCAG	TCACCTCA	EKCSGESFISSPPIQKTY	KLDLSTSTPPRKS
TTTAAAG	ACGAAGC	CTCCCCCA	KPIRTFKNPLTKSSQSK	SDNPPKPTPFSPRTLERSASWP
CGAAGC	TTTAAAG	ATAGTTT	ALSEVVDPLPKLKEKHP	SLPCALDKCPPSPRIKPSLTVPP
AGACGAA	CACCACT	TTTCGTTTT	CHTANNPSKPTSPESP	TLKPLPRPIRSPKPLEDWLTVRS
CACCACT	CCACCAA	TTTCGTTTT	VGPDREIVLNIGPRPR	PGPAAGSRITTSPPSTAPAKRVAAN
CGACGCT	CGACGTA	TTTCGTTTT	PIAAPLSGEPGATLDCG	TQGRKVQPPKKRPTESAGSLPPP
TACACCA	CCACCAT	TCACCCCA	AEPAIDLTLTGREGLAR	LVVEYHLSGDFGAFCDRLERWT
CCACCAT	CTCGCCG	CCCCACTCG	ALSTNNRPKPRRGYNR	GAAARATRNRRGRDDRDQDPQ
CAACAAG	AAGAATT	CTCTGGGC	DRDDQGGPGPVTGCR	PQRYKRAAALRSFGRDMKATV
CTCGCCG	CTCGCTC	TGCACATCC	RRIIDGERGDARCEID	PKTIEGRFDELSPVREGPECSLP
CAACAAG	CTCGCTC	CACACGTA	PWMAEAQAGEHAPSND	SQPGDAYDGPITALEVEMVLS
AAGAATT	CTCGCTC	GGGACCTG	TLNVGSAPGSDGLSYG	FWRALDPKGLVLSLSEFEVCRIER
CTCGCTC	CTCGCTC	TTTATAITA	RVPGPWKSSRVTLICK	DAEGDLDDLGNWRPISICQTVY
CTCGCTC	CTCGCTC	TTTGCCCTT	KIYAAVLARRLQSWAL	DGGVISRSQKGFMPFEGVYEHV
CTCGCTC	CTCGCTC	TATATGTAC	FLDLSVVADARATRRSL	AVCVWDLRNLAFGSDVHTTIVE
CTCGCTC	CTCGCTC	CACCTTTTA	ALSRFGAPAGLVEMIS	DISDIYTGSGSCRITRAGFTPDIPVGR
CTCGCTC	CTCGCTC	AAATATATT	GVRQGGPLSGHIFNLV	MEVLLRGVEANNACGYRLSCAG
CTCGCTC	CTCGCTC	TTGTACCCC	GASVRVLAYADDDVAL	VGSSRAEXKIQLGVCERFAAWA
CTCGCTC	CTCGCTC	ACAAGATG	GFSEFNKKCAAMVLKH	QRRGRRLLDSAPLRLCGEEVAI
CTCGCTC	CTCGCTC	CTTTTCGCC	LGPDSFYKYLGAHTGY	GRQTGGQLVDRVERQVVRLFT
CTCGCTC	CTCGCTC	AAAAAAA	SFLTPTQKLSALKRIVL	PAMSFHLRVRPCAEGLHRLDN
CTCGCTC	CTCGCTC	AAAAATTTT	TVRRRCVKTALRLPKG	SCRAFFHTSPDAGGLGITSVVAEC
CTCGCTC	CTCGCTC	TTATACAT	DILTVTQAFKMLSSPDH	LVSLVAKGRLGMHAARMGRS
CTCGCTC	CTCGCTC	TTTTATATT	ETASACAMADYLSGDS	VMGHXSWKTYGYRMPADLWTA
CTCGCTC	CTCGCTC	TTGTAAAA	TRAAASRRRLSLRFS	PQPQGEFGLESFTFKIAPRERSL
CTCGCTC	CTCGCTC	ACAGATTTT	LHHRQNLWWRNQWAA	LPNQKTVAAHSAYAAASNW
CTCGCTC	CTCGCTC	TATAAACTT	VKGPSLAPQALFFGLK	ARLNQMPTRSVKACYSRAPNY
CTCGCTC	CTCGCTC	TGCACATTT	DKSCRRCGAEVETLPH	VLNHCPSKMSILERHDSVLA
CTCGCTC	CTCGCTC	TTTATATAA	VLAAPIRGTIFASVD	DRTSREHFRVGEALRPDIVARRH
CTCGCTC	CTCGCTC	ACTTCGCAC	DGSVVVADVTCPECSA	SALDDTAARAKIEKYDQLCANL
CTCGCTC	CTCGCTC	TTATTTAAA	RQLYRKVPVESHALV	VGSLGSWGRNTNTALAAALGIRGA
CTCGCTC	CTCGCTC	ATGAATCG	VRSRLAKOYLNL	SVEGSHNIWLRWSGGIPKDLVR

572

573

				AGATCTT ATAATTA TAAAAAT AGATTCA AAGTTAA CACCACC AGGCCGC TACAGAG CATTTTAT TTAATCA ATTGTT ACCGACC TCCTGCT GCTTCTTT TTACTTTTC TCCAGAC TACTACC CGGATAC AACCCTT GGAAACG AGAGGA	ACCGTCGA CAACCAAC ATCCTCAAG TCGGCGGTT GCTGGAAC AACTCATA ACATCTTCA ANATAAAT TATCACCTT GTGCAGCA GGAGGCCG TGCTTTTAA AACTACTCT GTAGTGGCT CATGATAAT ATTTCGCTC CTTTTTCG CCCGTGTA ACTTAGTNG ATGCGAAT AAAATCAG TTGAATCA		MAQRRLKCRYARVFPNHKLHLYNKCSSVNFRCPIH ADFNKSTLHFCQKLLNFEISDINTHIRFLQNSISRLKDHA SLKLDYNIRERHYKFVNIKQNTYRRSNHTLHKFKKIL PKTNTTSSNDSIHNLNIAIPKDVAAEIASLGPENFLPSTF RERDFLCIKNLEVGLSHLDTNDRVKTEIRDKVSQYLVN ACKQKPAISRKDRQYAKNLKSTKNFLNKNPDIFTADK GNVTVCLKKSDYTERMHNLLSDKSTYKIISKNPLNSLQK RTSNILKNLNSNGFLSRTYHPNALSLTDTTLAKCYGLPKI HKVDAPLRPIVSLINSPTTYFLAKVIYDELKKSNNVPISHIN NSFELKNLKNIKPTTNYRLMSLDVSSLFTNIPLELVLRS IDRRIVTIMNKCRIPVDEIRQCTEFLFNFTFTDNIFYQQ IFGTPMGSPISPLFADLVMDDLENDCLSLKNNYNIPSFY FRYVDDVLLFINVDHIDQILNIFNSYHPKLQFTYEMEIDK KLNFLDLTIIRNSNYFNFDWFEKPTSSGRILNYFSGHPVS YKRNIVFNLVDRAILLASTRFHGVNIARVKNLLIKNNYP FNFINKNIKVRLKRHKYSNHNRSNTNTYPLCKIGIPYIDN NFEKISTMLKPYNIHTYPLINNLQHIVRLGKDTTNKHK QTNVVYKFCYCKNCPSYVGETKRAAEVRISEHMKSKNS	SSF5 6672	363	534	
Pene lope	Penel ope- l_Co Fl	.	Copidoso ma floridanu m	GT	TCCAATG CTTTAAA ACGCTGG AAGGACC TGCACAT ACAACTA TCAACCA						

[illegible]

[illegible]

5_Co Fl		floridanu m	<p>TAACAAATTT ATATCCATC AATCTCTTT TTATCGGTT AGTGGCCA CTTTCTTGT ATATTGAGT TTGATTCTT TAAAGTAA AAATAAAC TTAGTTTCA TAAATTACC GCTGATTTT GTACTACAC CAGGTTTAC CACTGTTGT AGTACTAC ACACACAT GTTCTCAG TAGCTTTAA AAAGAAAT CCCATATAT TGTAATAA TCGCAGTTG AGTAGAAT AAACATCTC ACATTGAC GACACTCTG CGATTATTG TTTAATAAT GTAAGTAA AGATAAAA CTTTTATAC ATTGTTAGT GCGTTTCT GTTGTGTAC ATCTTTATA TACATATTT ATATTACCA TTGCTGTAT</p>	<p>NQNETKRVYLELFKSLNKPLHKNQISNNIKFILDV/K THIKNNKSSLDILISNMLNCCNKFLKNNKNIMTKSDK GGAFVIIDKQDYINKVEESLNDKKLYEVVEHNPLPSLIN KTSNLLNELNGSDLFGDSGNKKVFNTRDKCILAKCYGL VKVHKPNQPMRLIISNIGSPTYFLSKWVSEFLKNNLPPSP HNVKNSFELKNDIKNFIIPDGFIVSFDVVAFENNIHFES VKVALESVWHILNLKINLEFESFLNIIRFIMENNYEFSNN KNYRQKFGSAMGSPSPSLADIVMRHIKMLLFLGPFEIK LIKVVYDDLLACIKISESDKVLNLYLNSIDKHIQFTMETPN NNSINYLDMTLTWDNNGSIKTNWFKPHTANLHLYW SHSPLSQKIGVIYSLVDRAIKLSDTCYHNNNNLKKVYELK ENNYPPDFIKKHAERRKKIYYNNKNNDNNNNNSKKLND DNKNYIVLPYIKNCDSNFKSLLNKVSYKTIFKSVNKQKF LNQLGKDSIKNKDKVCSVYCIKCKNCDKCYVGQSSRVV SKRINEHKRDVSKKNESAFNHVQDNGHSHFDNFKVI DNEPNFYKRLISEMVFIIKYNTV/NKVTDTQSLSSVFEFYI NKYL</p>	
------------	--	----------------	--	---	--

[illegible]

TTTGACTGAT
AAAGTTTAT
AATATTAA
GATTATTTA
ATTATTAGT
AATTCATTG
CCGTGGTTA
TGACTGACT
AGTCTTTAT
CAGTCATGC
TGGCTTTGG
CATTTTCAT
GTTTCGCAT
TGCATAGGT
GACTATAT
TTTCAATTA
ATGTAATTT
GTCTAATG
TTTTTTTAA
TTTAAATA
TTGTAAATT
GTACAGTTC
CCCTGAAGA
GGGAGTCA
GGGAGTAC
CCGAAATA
TTTGGATCAC
ATTGTCCAA
CTAATATAT
GGAAATAAA
TAGCATGTC
TCCAAGATT
CCAAGTTT
TTTCATTTAA
ATTTCCTTA
ATTATGGCA
AAGAGCA
AAGAAGA
GTAGTTTAA
TTATCAAA

					TAAATTTATT CATCSAGG ATATACACT TTGATGTAA CATTGTGAT TTAATATGT ACCATAAC CCAAATAT AAGTTAAT AATTGATAC ACCTTGTAT AAGCGGTG ATGAACATAT AATTGTAA TATTACCAA GGTGACTTA TAGGCCCA ATGGGCCG GGTGCCCA ATATTGATG TACAAAGTA GTTGTTATA AAATTGTGA TCATACGGC ACAATTGTA CATAATATG CACATGTCA CTATAATAA ATAATTTAC TTACTTACT TA																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
--	--	--	--	--	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

CCACCTG	GTCTATGAA	VRECVPEKTTTIRPRDKPWYDSMLRKTMRIRDLRNKA
TTTTATAT	TTTAAATTAT	LRTNNDADWTKFRHVRNKNVNNMKKHATINYDNIESA
AACTCAT	TATAAAGTC	LHDSRNGKIYWKLMDLKFVKATTEIPLQFFENGQPT
ATTTTTT	AATAACTAT	LAFSDIEKTETLNKYFSSISNIESANRSLPNLYYMCNNFF
CATTTTCT	AAGTAACA	SDIHIEVQEVFDIISILPVNKAIGPDNISHRMLKYTKFSISK
TTTCATATT	TGTAATATT	PLCMLFNKSLVENSFPNYPWKIAHVLPLFKDDPSVASN
TACTACA	ACATGTACT	YRPVSLSCVSKIMERIIFKHVYNFFHENNLIFYKYQAGF
TTCTAAC	ATGACAAAT	LPGHSTVFQLLETYSIVKSHEDGKMCCMIFCDLSKAFD
GCAATAC	GTATATATT	RVWHEALIFKLQTYGISGHLQWFKSYLSDRTOQVMYK
GTTTATG	GTATGTATT	NKLSSTLCINAGVPQGSVLGPLLIYINDVATNMLSFCR
CGTTATTT	ATCATTAGG	LYADDNSLQHCDLYTKNIEIVLNHDLKILSDWSKDWLL
AACTACC	AGAGGAAC	KENPSKTKAVFTTKKHVEFPKLFQGCQLDYVSTHRH
TTTGTAT	TTGTAAGTT	LGLLFSNDLSWSQYINNIVNSAYKKLGLLKKLKFILGRT
ATTGTTA	GTAAGAAC	NLSKMYTITFIRPVLEYASVVWDGCMFDMERLEKVQLS
TTTACCCTT	TTGTTTCCG	AARIVTGLPILASTDSLYTETGWESLVSRRNKAKLITMF
TTGTTTTTG	ATCCTTTTG	KIHTNQVPEYLSNIIPPVTKNVSIYHTRNSENYSVPISRL
TCAITGT	TGTTACTTA	LYKKSFPDITIRKWNLSKLEVREATSLNIFKKNITESIPQ
GCACATA	CACAATAA	PPKYFSFGKRTINILHTKLHNCILNYDLYRRNIIESPNCI
CATTTTTT	AAATGTTC	CGQKEDVYHFFVCKNYVNAARNLFEELFNRLDETILIN
ATGCTAT	AAATCAAA	SHLLWGSNDLSYNTNCFIFSAVQKFINESGRFNN
TACTTTC	A	
ACCCCTA		
CGTTTTTGT		
TTATCTCT		
TTTTGTG		
AACAATA		
TTTTGTCA		
ATTGTCA		
TATTGAG		
GTATCGT		
CAITTTIC		
ATTTGAA		
TACCTTC		
ATTCAGC		
TGTAAGT		
TTACTAC		
TAGTGGT		
GTATCTT		
CCGGTTT		
TTGGTCTT		

[illegible]

CTGGGAA
TTGACTG
GAAAGTG
GAGTCTT
TGAGGTG
GCATCGG
TTCGGTC
GATTTCA
GCASAAC
AGGAGCA
SACTTAC
TGTAGCG
AAATTC
TTTAIGA
AAAAGAC
AGCTGCA
GGTCTGT
GACTCC
GGTCTGA
AAAAAAGT
TGAAATGG
ATCCAG
GTCGTCC
ATCCGAA
TTTKTCA
GACCGGK
AGCTACA
GACCTKC
AGCTATG
AAAAAAGA
CAGATTA
ASTSTCCT
TGAAASA
GCCTATC
AATTA
WKGAMG
TGGATTC
GGTATCA
ATGAACA
ATTCCG

[illegible]

[illegible]

[illegible]

				CAAGCAA GTCAAAA TGGCGAC GAACAGC TCACC GA ACACACC CGGCAAG TCCGGTA GGTCAAA CATTTC TC ATGCCGA TTATGAA ACATTIG TGGA AAA AATGACT ATTATTG CTGATGC TGTTGAT AAGTTAC AGAAAGG GCAATCG ACACTAC AATCAAT TGTTGAA AGCAAAT TAGACAA GTTCAAA AATGAGT TCGTTCA GGGCATC AATGAAA AATTCAA AGCCATG A		NADTLKAGGRALGMISKIRSIFFKFGPYKEIGISTFEKL FLNCVAPVLDYCSEVWGLQNYHSIEMVQNRLRFYLG VHRFTPLPALHGEFGWKLCYHRHCINAVRLWNRLLTMI EDDLTKRVLFWDFNQTLGSRWTSRIKSILFESMNLHHY DNLSCDVNEFENKLFITYENEWLKGINKLPKLRLYKSI KTNFKTENYLEMNISKSQRSILAQFRCCGILPIRIETGRYK GEPLDERICNFCVLNEIEDESHFLNCTIYSDFRRELFNNI GYNPTLVMSNLD C IFFLLSNFPRQT SKYLJESYL YRQSLL YKKRNA			
RTE X	RTE X- 21_C Gi	Crassostrea gigas	ATCACGT GATATGT ATGGAGG TGGACGT GTTTTTAC CGGAGTA	TAACTTTAG ATGTTGTTC TTTTTTGAT TCTTTTCGC CTTTACTTA TGTC AATT	MNLSFPSKPPKKKKQRTKSSTQGKMATTQQGACTNS VYPNDNLSQSVQPIQQVG N LTYHQLPQM QYIQSAHPM QTHMQHPQNNVQSNTPQVTPOWVS DIV/TOLKNMNQKL DTIDTTLSSMKREIKGLSNRVVEVEKSEQEFSKSLDNEK NTSKALT ELSDLK M ALDQSLQENV RITESV LDMQC RSM RDNLLFFGIEEAOSSGEOENCEDIUTSLCASKLEIRSELIER	cd016 50 920	118 1		

			CGACCTA CGGGAAT TTCATAC CATATAT AGCGAAG TTTAGTT AAAGGTT GCGGTAA ATGTGAT ACACCAT TGATACG TGAGTAT TATACCT TGTTCTT	TTCATATTT ATGTATGTA CAATGTAAT AGGGTCTCC TGTAATAAT ACCCACTTC TGTAATTTA CTATGTTTA ATTTGTAAA ATCTGATGA GCTGCAAA GCTTAAAG AAATAAAG AATGAATG	AHRLGSKNNKPRPIVVKFNRFPHREVMVRKCSYKDKDT HISIGEQQPKEIQERRKKLLPIYKRAKENNQKAVLARDK LFIDGKLYSDDGPVQMKATPTQPKASRRGQKHGRGK RGQENQJGLRIIAWNIHGGLENKWSERNFKSFLCTYDII CLSECWIGKDNISDQNFQSFVPRKSGKSGQGTVVLI RNVYTPFISFVESLHDITVWLKIDKSFCHTKDIYIAFTY FPPSNSTFLKERDIDLFMELENQLSYMSIGDIFLFGDEN SRITLDPYIVNDQVHESVLDNNLQYSSDNQLTERVNP DLGHNEYGQRLLSLCRSTGVRIVNGRHSGGHSNDYTFN GARGLSTIDYLLTTYDMDFISKFIVCNFTTFSDHAPLHI EFKCTLLDPCDTSSHYRTRNVPRWDSYVDSNDFFNS IEGTCNFVDFEKVTSQSDLDCEMLKITDMLSKDIGNFFY VSSKNSNQSKPANRSRYSYSTYLSLTNDKPFENENLSLY RDYITALRTFNVCNAQNHLNLVSKKRTYKKLEAKLKR SYKQIEGNRLNILKKNPRQFFAKFRKKPKSVGNVPIEIF FDHFCKITNV DGSFDTNCEQANMEVLPIVKNKPTDECVF YELDKVIDVDEIRKAISNLKRGKSHGEDGILNEYFINFQQ HLLPLLYSLFNCILDGTGFFPKSWSSSIIVPVYKKGDCNTP DNYRGISLVSNMAKLFSTILNNRLLDWSKTNDVITDAQF GFKPNCGTRDAIFALHSIVTNTLCRKRLYCAFIDFKKA FDSIDRSKLWLKLSKVGHQGLNLNVIKALYNDVKSCVV VNGHLSDYFCNNVGLMQGEVLSPILFNLVYVDFEMNFL KSGCIPHELLSLNFLMYADDMVLFSSEVEGLQLLNE LSHYCKTWNLCINVEKSKIVFRNKGKIKCNEKWCIGN DVLEIVEQFTYLGHIIFYNTKFTKAEKQLSDQGRKAMFA FRKNIRGMILNHETLLSLFDCYVGGIINYASEIWGSHKG NNVEKLHLDCKQTLGVKRSSFNAAVYAEIGRNTSYY KKVLFNFKILEGCMVGIPLITKRFCSILKFWKDVLCNS NCIQKCYEEMYINCELHGHNWASDVKKILNEIGFNEV WNKQTIDSLLTKVINQRIADLAKQDIFGKIENSQKCHFY KYLVDGFYLYYLRKSPVPVKFQQCISKRLSSHRLAIET GRYNNTHRAQRNCFCKNYVEDEFHILVCPVYIEYRQ KYIKKYFYEKPSMFKLLQLFNNEFKDMCNLKGYYYS MKLRNDICN			
			TATAATT GTATTTC ATGAACA TTAATAT ATCTCAC ATAATT	TAAATGT GTTTATAAT AGCTGCAA GGAAATTG TAAGCAA TAATGACA	MDDKVQQSYLSILTLNCQGLGDINKRRDIFKVLKSKNY NIYFLQDTHFEVEKEESLIKGLWGYKAYFNSYKNSRGV AILNNNCDITILNEHKDSDGNFLLLDVFTIESLPFLINIYG PNSDSPAFYINVLNKIQDIYSNQHVIGGDFNLVFDKDL SMNYKNLNNPKARLEVLKIMETLTLKDVYREHNPTLKK FTWRKKNPIKQARLDFFLISELNLNLVYKYENSYSRD	cd016 50	530	791
Tx1	Tx1- 5_CG i	Crassostrea gigas						

TGGTATG	AAAAAATA	HCPVVLVYRTNEFLKGGFWKFNNSLLKDKEYVYLIKE
ATTTTAT	TAAGTCTAC	KIKNVKQYACPVYNLENETIDDDNIQFIISDQSFDLTLL
AATATGC	TTTATATGA	MEIRGKSISYSTYKKKKTTEREKLEQEILLLEQNFTVET
TGTCATA	TCCTCTTGC	KDELSNKQTELELRKRNLRQGCIRSKVKWIEEGEKPSK
TTAACAA	ATATGTATG	YFISLESNFSSKIIPKVEKEDGSTVFNQFEILNELRMFYD
GCTCAAA	AGTTTITGA	NLYQKRETIIDSENIYEKLYDIPKLNKQESDLLEGPLN
TCATCAG	TAGATAGT	KLEVYNFLKNMKNDSKSPDGFSEFFKFWNDLGSFIT
AAAAATCA	GTTCCCTC	RAINNTYELGHFTESNKLGVITCLPKAGKPKQFLKNWRP
AGGTACT	TCCTTCTCC	ISLLNVLYKIASGCCIANRIKTVLYKLINEDQTGFLKGRFIG
TTTTTTT	ACCTCTTT	ENIRMIYDIMQYTEQNQIPGLMLIDFEKAFDTISWKFIY
AAACACA	CACCCCTCC	QTLDFNFGISIKNWIIQIFYSDIKSCVIQNGIASDYFFPQR
TAGTATG	CCACCTCAT	GCRQGDPISPYLLCAEVLGILIRNNKEIRGIEIDGVEYK
CGCAGCA	AACTAAAA	LSQYADDTSIIFDQSPQMDGILRELDYFACISGLKINFSK
TAGGGTT	ATACTCNA	TKMVWIGSKISKEVFFHHSRWKLDWNNPTFDLLGVKFS
TAAGGTT	ACTTTCATG	VNLNEMVELNYPSEFLVTRLIKQWKLRLNTTVGRLTV
TTTAAAA	GGTGATAA	IKTLIIPKLNHLILTIKPSVEYLOQFEKEIYIFLWGGKVH
AAACAGA	AAGCAAGT	KVKKNIIQDYYRYGGLKMDYMEFVCALKSTWIRRLIHS
CAATAAG	TTTCGAGAT	NTKWVKLFESELGLKIETLWEKGSDFITNISKNISNKFV
TTATACA	TGTTTGTGC	KEVLDDWVKIVMSYSQNDIKVLNEHIWFNPNIKINSSIF
TATAAAA	ATATATATG	IKNYFNCGYKFKIDLMTEEGNFVSLETLENSNVKTNFLE
CTCCTCA	TCTGTCTGT	YRSLKGAVERLKAALNNNKVEYGPFPNLSIFPKQSKG
CGACCGA	CGGTGTATA	CKDMYYVLIDNKKESIRSVKKWTESGFAYTEIEWNNIFE
TTCTAATT	CACCTGTCT	LPFKTSKESKFQWLQFQILHRLPTNEYLFLKLNIDSPACS
ATTATA	GTACGTGTG	FCNQDIE/TIEHLFVDCFHVKEIWL YFEEWLRSKFNFLMT
CTATTGT	CTTTTITAT	FDKKTLFGKFKNRNMYRLQNILITVKQFIFASKYKNV
ACGATTT	GTGACTGTG	PNLSTLMLKRTVTVDRIHIEKLLLLKNCKFAEYEWQTI
TTTTTTT	TTTATAITA	CDLL
AAGTTGA	TACTTATAT	
CAAAAAA	TCTAAATCT	
AAGGTG	CAAAATGAA	
TTTTTGGT	GIGAATAT	
CAACTTC	GTGATGTGT	
ATGCAGT	ATGCAC'TGT	
CAACTAT	AAACTAAA	
TTTTTAGT	CATATAAG	
GTTTGA	AAGATGTA	
AAATTAG	AAATAACTTG	
CAAGCGT	TACATGAA	
GATGTGC	AGTTTGAAT	
TTTCGCA		

[illegible]

TACAGGG
GCATAAC
CCTACAC
TCTAAGC
TTGCTGG
CGAAGCT
TTTCTCCC
GTATCCT
AAACAAT
AGATTAC
AAAGATG
GGCAGAA
GAAAAATA
GT'GT'GAT
ATCTGAT
GCGCAAT
ATGGGTT
AAAGAAA
GGT'GTA
GTACTAT
TGATGCA
TTATATG
TTTAAA
AAGTATC
GTGACT
TTAAAAT
TGCAAAT
GGCAGTA
AGCTCTT
TTGTACA
TTTGATG
ATTTTCG
GAGAGCG
TTTGACA
CGGT'GTA
TAGAAAT
CGATTGT
ATTATAA
AATGTTT
AAGTTAG

[illegible]

NIP LNER CLCNKGLIGDEFHYILECSALEEIRKKYINTK
YWKRPNFFKFSELMTNCNCKSLRKL'CVFISKIFDAVCSS

TATATCC
CTCTAT
AGGATCC
GCTCCGA
GAGCAT
TCCAACG
CCACGC
CAGTCA
GACGGA
TCCGCCA
GCAATCA
AACAGCC
GAGCGAC
AATACAC
CCCGCA
AACCGA
CCAAAGA
CAGCGA
CAGGCT
TTAAACG
ACAAAGT
CAAGGAT
CAACCC
ACCCGA
ATTGGAT
AGGGAGA
CGCAGGA
GAACAAC
TCTCGG
AACAAAC
AGCTACA
CGTCTCT
GAAATT
CTCTCGC
TTAACGT
ATCGGT
GTCAAAAC
AAAAACT
AAATTAC
CCCGAAT

[illegible]

DEWNA LSTDV RGS DSV RIFKE KIHVS VKNEN XLTC RTRP	AGAGTAA
NFFTYGDRFLNIHTKL RHNCALNNDLFR CNINSP LSC	ACTAAA
GKJEDTYHYFFTCKYKDARNVLFNDVFQIVNLNIVNT	AATTCCG
HVLLWGDSSISDNENKHLFTLVYNFIKNTERFN	ATATCTTT
	GTGACCG
	AGGACCT
	CACCAGA
	TACCGAC
	AAGGAAT
	CATACAT
	GAACTGT
	CAGCTGC
	TAAACGC
	TGCGTA
	AAGTTCA
	CTCTTTCT
	GGACCAA
	CGACGGG
	AGAAATT
	TCACTAA
	ACTGTCC
	GAAAATG
	GTCAAAA
	ATAATTG
	ATCAAAT
	CCATCAG
	TGAACTG
	CACGAAA
	TCGCGCC
	GCTACCG
	AGTGACG
	AACAACA
	AGGCTAA
	AACGCAA
	ATCCCGT
	ACTCTCTT
	CTGAACC
	ACAAACC
	CATGTGA
	GCAATGT

ATAGTAT
AAATTAG
ATACAAT
GTTCAT
CAGTAT
GTGTAT
ATTCTAA
GTATCGT
GTCTCA
AAAGTAG
ATACATG
TATTGC
AGATGTA
TGAAAAAT
AATTTT
AGTTATA
TGTACAC
CAATCTA
ATAGTGT
TGGTAA
ATTACAT
AGCTACA
TTTTTGT
TGTATA
TAGTAGT
ATTATAA
CTGTGTA
AACCTAT
GAATTGA
TATTCCC
ATTTTTTC
TCAAATG
CAATCAA
TTTGGTTT
TTACCAA
TATTGTA
CTGTATG
CCATAAT
GTTTTTT
TTTCTAC

L2B	L2B-1B_C	.	Culex quinquefasciatus	G	ATTTCCT GTTTTGT TCATTTGT TCCTTTCT TGTTTTTC ACTTTGC ATTCCTTA TTTGTTTG CATACAC TTTTTCAA GTGTTTG CTTACTG TGTTTTTA CCTTCAA CTTTTGTC TTGATGT TACGCAA CGAAGGT CATGGTC GTTCAGT ACATGTA TGAATTC ACACCTG TCTTTTCA ATCGCAG TGCTCTCT CATAGTT TCAATAA CGTTAAT CAACATG TCTTTCA	TAGAAAGTG ACCCACGA TGGTACTGT GACGAAAG GCACGTTAT GACGGTCG GCCATCTTC ATTATCGGT ACACATCG	CESALNLLLLKWKQCKIEKGNIVLSVFVDLKRAFETIDRT KLIRVLKKSIGRGTVLKWFSYLENRKQVTRYNDVSA ETAVDLGVPPQGSVLGPLLLFILYINDIKQALKRQVNLFA DDTVLFVVGDSFDECFDMMNEELGGFSEWLKWKKLQL NITKTCMLVTTTRQTKDCCRSVQMDGGIVERVETMKY LGVMLDEKLFNEHIDYTIRKAAARKFGVLCRINRYLTAE TKVQLYNSLIAPHFDYCSSILFLATQRQMKRMQVLSQKI MRILJRCDRLLTPRRSMLECLQWMSVKQRIEYNLTIVFVF RVKERMAPQYLTTEAMVRGRDIHEHDTRGADDLRLQIW	cd016 50	1	195
-----	----------	---	------------------------	---	---	---	--	-------------	---	-----

RRKACTQNSLFYKGFKLYNQLPEAAKMTSININEFKRSCK
EFVQRPLE

CR1	CR1-10_C Q	Culex quinquefasciatus	TC	<p>GAATAAAT AAA</p> <p>TAAATTAG TTTTAAGTK CATCATTGG GGCAAAAT AAGTGCCT GTTGAGAA GAAATAAA CAAC</p>	<p>RRADFESLRALRGIEWSSVLDPVDIDLAVDTYTXILGR LIEQFVPIKIRKVVSHHLPWQTQELRXLKTKQRAAFKVS NGSVHLRDYYLRINHSYQRLSRRCFANYQRKMQKKLK TDPKFWKFNESRRESGLPTSMRLADXEGDDADKICQ LFAKKFAGVFCAEPTIDEEVRAAENVPRRDELLASIEID EGTIQAAAALKHSCSPGDPSTLLKRCSSSLSAPLLH LFRSLSSGKFPVAVKQAYMFPVHKKGDRTNVENYRGI SALCATSKLLELVVIDPVFCHCRQDLADEQHGFPRKST ATNLCFTESIINSEDAHSQTDIYTDLSAAFDKINHRIA VAKLERLGRGSLLAWFKSYLTGRSLRVRGDALSIDI ATSGIAQGSHLGPVVFIFYENDVNYVTGPRLVYADDL KLFARIDNSNDAEALQRDLDRFAGWCXVNRMLNPGK CQXITFSRRHFPVTHDYRINDTSVERVTHVKDLGVILDT QLTFKQHLHYIAKASRQLGLVIRMTKFTDIYCLKTLYC SLVRSTLEYCSTVWSPHYNNVAVCRLESIQRRFVRYALR QLPWTNPLXLPPEYEQRCRLIDLTLQLRRDLARAMTSA DVLLGRIDCPDLRNQILRVPTQLRHTPALEVPRTN YSANGAINGLKRAFKNKVSFVSLDVSRTTLRFKFSLLR SLLRM</p>	cd016 50	250	500	
CR1	CR1-25_C Q	Culex quinquefasciatus	A	<p>TAATGTTAG TTTTTTAAG ATTATACAG CCTGAGCAT TTTAGATGG AGGCGGTG AAATAAAT AAATA</p>	<p>RALRIFLGADKRXQPDASXSASSCPTCTNKTVNAESH AKVNAETNTKVNVEVNPEIYAGHGADPKGYGTNFACE SQWFTYGSRDRLTNLTSLPSLPQQSSVQHLPASSPGRA VESIMRAPRLPDATAPSPSYRTRSPASRSHQRRPGAGVG VGGVSGTAPLGKYPQLQVSLPDACPTYSSSLITSSSTR TANASPPAQAIQSSPGRTVESIMRAPRLPDATSSFSYRT RSPAGCSHQSRPGAGVGVGGSQPVLFKGYPFQSQRV LPDASPTHRSCTSTSTRTANASPLPAQAIQSSPGRTVE SIMRAPRLPDATSSFSYRTRSPAGCSHQSRPGAGVGVG GGVSGPVLFKGYPFSTPHTSPDARSDFSATVPPSTFVTL AAPGCVYHGTMEDLGSANSGMPAALATASDPASASSER GFLNVYQNVNRLRTKNTSVLLALSSCDYDVIIFTETWL VPGIKNSEFAAGYDIFRCDRSETSSLRGGGVLVAVKT GRNCKDELVGCGLSEQIAVRVLPHQALFICCIYIRPAS NIDVYQLHTGAVQQILGMAGVRDLVLVVDYNLPAIT WFFDDDTSHFLPINASAEQEVLLLESMLSCGLLQINNLA SAHGNNLLDLAFVSDATAAELIDPPCPLLKPDAAHKPILL KLDYDHGGNDDLLDDADEIFDNQCDYALLNERISAVD WDTLLDTSSVDESTANFYERLNSILHEVVVPKKIRRMMSHR</p>	cd016 50	933	118	4

					YKLPWVNDLRTLNRRLRKATSYMESKTAWDKVTV QNIEDYNALNQESFRSYNNNTQSRSLKSDPSKFWSYVNS RKNAARPTDVSYRDLNSTSPAIAANFFADFFKSVHSSD AIPRDHEFEYVPTHYLPPLPVNLNDEETLKALSGVDGA KGPDPGIPPSLIKACADSLTPVKRIFNDLSLRKKGVYPA MWKVASITPIHKSGSTKKVENYRPIISILNCLAKVLEKIVY DRLYAVVRPIISDDQHGFQMQRSTTTNLLTFVHSTVNAL ESGEQVDIYIDFEKAFDKVPHALTVRKLRKLGPAWIT NWLSTYLTNRKAFVNLRIARSTFDIPSGVPQSHLGPLI FILFVNEICALTNSNTLMYADDLKLFRSVKNNVDDCCSLQ MDVNALLHWCESNGMKVHVKKCKVISFSRSRNPLRFD YTMGDGCPPLDRVNSIRDLGTVDRKMKFNEHIALTTGK AFAMLGFLKRNTHAFDDPFALKVLYASLVRSVLEYAAV VWAPYHVTLAARIERVQRAFRFALRKLWNNPLILPS YEGRCLLFRLQSLAERRITLMQRLFVFDLLRGNIDCRSIR NNVRFNEPARQGLRGERPLLWIPRHIFDYGYNPLDVCF RQFNEICSCFDYVVSKTVFKTRIS				
					GESESRGREGHGRVTRPPLAASPGRAVESHVRVPRLPDIT TSSPSHRLIRPACCSLQSRPGAGVGAGGVSQTATLGKY XSXSQXPLPDARPTSSCAISSTSPVTPNAVRAFWSSP GRAVESHVRVPHLPDITTTSSPSHRLISPAACCSLQSRPGAG VGAGGVSQTATLGKYHSSLQNPLPDARPTSSCATSPP VRTPNAVRAFRSSPGRAVESHVRVPRLPDITTTSSPSHRLI SSACCSLQSRPGAGVGAGGVSQTATLGKYHSSLQNPL PDARPTSSCATSPPVRTPNAVRAFWSSPGRAVESHVRV PRLPDITTTSSPSHRPISPCCSLQSRPGAGVGAGGVSQT ATLGKYPGLQNPLSDARPTSSAAFPSTINAKGTSTPCK LRIYYQNVRLRTKIDAFFLAVTECDYDVIVLTETWLDI EILSPQLFGSTYRVRNDRNERNSDKKSGGVLIAISSRL DCVLDPTPVCDTLEQLWVRLTVGASQRCVIGAFYLP LNRSIDLIKDHLEISGNVLSHLNLDNSFIQFDYNNQPGLK WIPSEEGGLEIDPVQSRVPLVSSYLIDGNLHGLSQLNPC VNVYNNCLDLVLVNEVALPNCAVLAADVPLLPQDIAHP ALEMFVDVPAPATAEPPPLNSQQSFNFRKANYAAIQEEL SRIDW/SFLETAPSIDAADVAFOQSIILNNLFTEHVPLRRPQ KPPWSNPTLRKLNRLKTAALQEFNSHRSPFFKSKFNLAS KRYRGYNRTLRYRYVIRTQRDLRRNPNKFWFVNSKRK ADGLPPTLYLDDVAAAATSEKCALFAAFHPKDAFNRSFA SEAEIIEATRNTPINAIECSLTEVANQVSAAINKLKLSYS PGPDGIPSSVIKRCSGALLSPLTCLFNLSLRQSTFPDKWK	cd016 50	910	115 7	
CR1	CR1- 9 CO	.	Culex quinquef asciatus	C		TAATCTGT TTTGTGT CATGTGT AGCATTTGT CTTTAGTTA AGTTAAGTT TTATCATTA AGACCACA AGTTGTCGG ATGATTCAA ATAAACAA ATAAACAA ATAAACAA AAAAAAA			

						ASFMFPVHKKGDKSSVKNYRGITSLACSKVFEILVNNV LFDCKNYISTDQHGFFPKRSVVTNLAQFVSHCLQEMD AGAQDAVYTDLKAADFDRVDHGVLLKRLERLGVASL VRWLKSYLTNRILRVKIGNSSEDMFISLGVPGSNLGP LLFILFANELSMMLSEGCRIFYADDVKIFTVVRSLVNC ALQRLNHFNEWCVRNFLTISIEKCTSTFHRKLKPVVFD YAIAGNTLARVTQVRDLGVTLDELSCFTIHLHDIVSRAN RQLGFIFRIADDFRDVACLRSLYCALVRSILEFCSVVWCP YQSTWTAKIESVQRKFTRLARRARNPAGWNTYEERC VLQLETLERRRHVSQATFVAKVLAGEVDTPWIRAQIQY TPARPLRQRQYLQLAHRDTNYGQHEPVRMCSRNFVFS HLYDPNISTPQFQQRARLWYSHQL				
						MNTRALEYRLACSAISYDVIAVTEWLKNNLTSSQVFG NGYEVFRCDREQFINSQKDDGGVAIAVRQGLNARIVS SDEWKCEQVWVAIDLADHTVYICVVYIPPDQTNNASV FEINSSSVTVSSMARPYDELVILGDFNLPRLIWRPSRDG FLYADPELSTQTLHVTEFLDNYSSALLQQINQFPNENNV LLDLCFVSCPDIAPIAIAAPLVKDVRRHHPPLHLSLPVH LASEFTPIVCEVRYDYRKSDVPSMLELLRSIDWENTLDK DNLDDAVQFTFTNIIGYVIDRHVPKPKVTNARVPWQTSE LRRVKAARKSALKKYSKHRITLPLHNYVYRLNSCYKRM SKQCHADHEVRTQKLLKSNPKGFWKYINEQRKEIGLPS SMFLGDKTASNTEDICQLFAAKFSSMTQEDLSQQQITA AANNVLPNGLALNRITIDRSQILEAATKLKSSNTAGPDG VPSVLKKCIDGLDPLAHLFSLSLSTGSFPKLWKHAYM FPVHKKGDKRNADNYRGISALCATSKLFELAVMDPVFF FCKNLISDDQHGFMPSRSTTNLLTFTTFVTDSTAKSQT DAVYTDLSAAFDKLNHSIAIAKLERMGIGGVLLRWFFQS YLTGRKISVKIGDWLSAVFEAFSGIAQGSGLGPLVELLY YNDCHISISVPRLSYADDMKIYCRIDNASDAQVLQQQLI NFAEWCKINRMVNVNPTKCSVISFRRKRNPKEFYQICGTS IPRENCVKDLGVLLDTELTYKQHSIYIVSKASRQLGFVF RTAKNFTDIYCLKALYCALVRSTLEYCCSVWSPYENS VARIESVQRRFIRYALRSLPWRDPFRLPSESRCELINLD TLAVRRNVCRSLVADVLTSRVQCPAILRGLNIFAPIRTF RNSPFLRVPIRQNYAMYALSALTGLHRAFNFAVSLDFENL SRDVIKRRKFLFFRR	cd016 50	499	750	
CR1- 5B_C Q	CR1					TAATTTATG TATTAGTTT GTAAGGTA AAACACCA TTGGGGCCC GAGCCTGTT GGTGTCCTA AATAAATA AATA	Culex quinquef asciatus			
RTE	RTE- 1_CQ					TAAAGTAA TAAAGTAA TAGGGCT CCTTAGG	Culex quinquef asciatus			

				CTAGCCA CTT		ERIIVARFRTRVRNLTFVQVYAPTDAADLQEKEEFYSQL SXVVNEIPKGDIRIYAGDFNAKIGSDNTDLERIMGPHGX GEMSENGELTFEFCGNQDMVIGGSLFPHRSVHKVTWVS RDGKTENQIDHICISRKWRRSLLDVNMRSADIASDHL VVGEIRLRVARVVRQEEKVGCRFNVQRLSNPEVXRAFV GELQARALEIPDSTVEEEWQFIKDAFAVTGENTLGMRLT QRKEWISDATWDKIEERKQAKAAIVSARTRATKARTRQ TYAELEKAVKRCCRDRKRAWTDSLADGEKAAANGD MRLLYDISRRLCGARTNTRIPVKDRSGQLITDPADQLKR WFEHFEQLLOISHPRSPQYHPQNVRRINRVNSSPPTLRE IEEAVKGMKSGKAPGIDRISAEMLKADAHLSARMLHRL FSKIWDATFPVDWMQGLVKVPKKGDLTNCNWNWRI TLLCALKVLCVKILCRIKSKIDESLRRQAGFRSGRSCV DHIVTLRVILEQVNEFQDSLHLVFVDYEKAFDRLDHENL WEALGRKGVPERIINLIKAQYEAFCRVVHNGVLSPIR VVAGVRQGCILSPLFLVIDEVLGALDRVQNRGLLW QPIRMEHLNDLDFADDIVLFSQRGNMQNKLLDILAACS SAGLVNVN'SKTSMANVTERAQGFVAGENVEKVDIT FQYLSQIAPDGGTKLDISTRIKKARSAFAGLRTLWRSN QISLRTKLRFNSNVKSVLLYGSETWCVSSQNTSKLVF VNRCILNLR'AWWPNWVWSNAELHRRCCQQRPHVEIRE RKWRWIGHTLRLKDAHEICREALDWNPPQGRGRGRPG SWRRSLHQEISAVDGTLSWRQVKAENRQQWKSFNA ALCSTGEQXA			
				GAATTGT GGTTTGG TTGAGCG TTTTAGC AGAAATGC ATTACTG TGAATAC AAAGAGA CGAGTTG TTCCTTTT AATTCG CTATATA TTTAAAT TAAAGGA ACAACTCGT CTCTTTGTA TTCACA	TGATGAAG TCTGCAAGT AGTAGACG AAATACGT ATCTGTCAA GATATTAA AAATATAT AGCGGAAT TAAAAGGA ACAACTCGT CTCTTTGTA TTCACA	MHSAKTLNQTTTMTATKPSEHNKFIGLKIIAGIHKDIAF LKKC'VEHRVTP'TSHKIRTKGHTDSKIIRRV'ELELVKESIK KLYGRLSIKITL'ECSLHLKLAKEPIEF'EAFLTKVKVAEK CESERKKILDN'KFRQLISHIAPT'MKPVVKPLEEFVNRS SETFSEEQTL'LNKGLKYAVSSKPNLEINIVIDIETVINANS ID EKQTKPPPLPAQINSVRMTVSKLIREATT'KKHQDNAE VRVVKELKEKPV'YLLKADKGNRVVIMDRDEYDKELLE KLG'NREKYSI'QRGYTLIDMVKKVENTIKECAGILLKSVG KTA'KSLRVPNPTL'PRIKALPKVHKPGNEMREIISAVDAP TSRIAKWLVEEFKSMKPPPSRSI'ISSQVFTIKELLESPIA EDEIMVS'FDV'TALFSP'IPVKN'AGITL'RDWLQSQYE'GEPW RQKTIQYITFVKLCMEQSYFQFRDCITYQQRMGASMGNP LSPFMSD'VYVMAALEHELQTKNLLPERWWRYVDDVFCII KRD'LSLT'VLD'TINSARRS'IKFTTYEMEVEGKLPFLDILIRE PGCPSSFSFEIYRKPTINTRRTIPATSNHSFQHKMAAYHYF IHRMTTLPLSEAGKQKELEYIFETAEINGYSRTTIQAIDK	cd003 04	469	548
Pene lope	Penel ope- 4_CQ	Culex quinquef asciatus							

604

TCGATGA
TGTTGTC
GGTTACT
TGAAGT
TGAAATG
GGGGACA
AATCTCT
TCCGTTTT
TCCTCCA
TGGAAT
TTTGATA
TATGTGT
GTATATG
CTCATCC
AACCTGG
TTTGCAG
CAATCTC
TTAGTTT
TGCCAAAT
GTAGCTC
ATATCAC
AGTGCC
GCAGGAT
ACCTCGT
AGATGCC
GGGTGTG
CCTAAAG
TCGGTAT
GGGGTCT
TTCTGTG
TCCAACT
GATGGAT
TCCAACT
GGCTGTT
TCTACTG
CTGAACA
CTAAATC
GATGCCA
AATTAG
CCAGTTT

TTGTCGT
AAATGGGC
GTGTTAT
CCTGTAAG
TCGAATT
CTCAATC
TGTCTCTC
GCCAGTGG
CTCTGCT
GATGGCG
TAAGGGT
TGTGAGT
GAAGTTC
TGTGCAG
TTTTCTTT
CCTTCTTA
TCGATGA
TTGCTTG
GATGCTG
GTCTGCG
TGTAACC
GTTGATT
TCCGCTG
TTTCAA
TATGATC
TCCAATT
CTTCTGT
TTTCTCTG
TTTCGCTG
AGGGGTA
AAGTGT
CATCTTA
TGTATAA
ATAAATG
ATATGCT
GCCATCT
TATGTTG
GAATGAA
TGTGTTG
ACCTGGC

TGGGATT
GTTCTTCT
GGTGTT
TGTGGCT
TCCTGTA
GATCTCG
AAAAGAAA
ACGAAGA
TGGGCAA
CCTGGTT
GTCCTAG
GATCAGG
ATGCTA
AGAAAGG
TAACTTT
CTTCGA
CTTCCATT
TCGTAG
TGAACCT
GATGCTC
CTGCGTG
CACTGTT
TATGGTG
TCCAGTA
CGTTGT
TAGTGAG
TCCCTTTT
GATGATG
CAGAAGA
CGTCAIC
TACGTAT
CTCCACC
AACGTTC
TGGTAGT
AGGTTT
TGGTTTG
TAGTTCG
TGTTCCA
AAGCTGC
CATATAA

ACATCGC
TCATGAA
CGGTGAC
AGTGGGT
TACCAT
CGATGCT
CCCATCC
TCTGTGT
GTAGATG
CAATCCC
TGAAGTAG
GAAGTAG
CTCTGTTC
CATGCAT
AATTGA
CGAATGT
TATAIAC
TGATATC
GTCGTCG
CTCGTAT
TGTGATT
GCAGCCA
ATCCCGT
AGAGTTC
CTATTGC
GTTTTTTA
CTGGTAT
GCTTGGG
AACAGTG
CTGTTAC
GTCGAAG
GAAACCA
TTATTCCA
TCTTCAG
CTATCGG
TCCCGAC
TCTAAGA
GCTCCTTT

GTGAACA
CCTGCGA
ACTGATG
ATCGAAC
GGCTCGG
GATCATC
AACCCTAT
CGAAGCT
GGCTGTC
GCCGAAA
GTTTGGG
AATCTTC
AAGGCAG
GCAACAA
ACGTTTA
CTCAACA
AGGATTC
AGGTAACT
GGCTCAA
CGTGGCT
GTTTAAG
CTGTTAC
CTATACA
GGCAACA
AAGAAGA
ACGAGGA
GGTAGTA
CCTACAG
TAACTAC
AAATACG
GATACAT
CAGCGAT
GAACCTT
CAGTCGA
GAACAGA
ACACTGA
TGAAGTC
AGTAGAC
GAAATAC

				GTATCTG TCAAGAT ATTAAAA TATATAG CGGAATT AAAAGGA ACAACTC GTCTCTTT GTATTCA CAGTA				ARNCRRVYACNXDTANLQPAELQPKTPGPAVTPVPES TESLTYDFTRCDFEAVNJALSQINWTGILAHCSLDEAVS RFYDALXDVIQLHTPLRRLHPPSNRKPPWNAATLNLRN RLRKARTLHKKRRQCPNIAEVRELEELYESQQNAAFSEY IHRIETDVKQNPFSFWTLKSRKKGSGIPTISNGCVEAN SVEEQASLFADFFSGVYSTSLPAPCNEVVDHIPTHDLHLP VVHFSIEDVSRALDDVDVTGPGPDSIHPSFLKNCAASL ALPTIIFNRSLAESTFPTAWKLASITPIHKAGNVHKVEN YRGISLSGFAKLFEHLVHGALYPVLKPIIANEQHGFVKR RSTTSNLLFTSTLLTNEKGQQIDAVYVDLSKAFDKVP HALLLEKLRRYGLPEWVIRWHSYLAADRKAFAVKVRCST SHLLAIPSGVPQGSVIGPLLFILFVNLDLCSVLESCLKMYA DDLKVFRSVSSPLDSCALQQDIDRLLRWCTANGMEVNV QKCKIISFTRKRTPLLLSSYRMGQNELERVNTIKDLGVTD SKACFNEHIALTTAKAFATLGLFRRNAADFVDLFALKTI YCSLVRSQLLEYAVQVWAPYAVQAERIERIQRAFTFA VRRLPWVRPQGLSSYDDNCERLKLPAASRRVLLQRIY VFDLLSCNIDCPTLRDKITLHIPARNLRNPAPFLEVPGHR TNYGYYSPLSACCRVFNDVADVVFVFGMSKCVFKTRIKN RT	cd016 50	297	547
CR1	CR1- 27_C Q	.	Culex quinquef asciatus	TGACAAAAG GTTTAGACC GGTTTATTT TTAGTCAGT TTTATTTAT TTATTTAT TATTTTTTA CATTGTGCT AGTTTAGTT TGWTAGCT				ALKKLSKRAALKKFSKCPISVNRNLNLNQISNDYRRLN KSLFLAHQSRVQNRLKQNPKGFWKFNQQRKESGLPST MIRDGVEVSTSEICDAFRVQSSVFADNNLSDPDVAAA ADSVPSGLVCSDLQVSREDFNAACSKLKSSTSPGPDGV PAIVIKKCANSLSEPLLRIFNLSSLGVFPNNWKESYVFP VFKKGCKQDVKNYRGIAALCAVSKLFEVIVLEFLRHNL SHVVSDDQHGFFPKRSTNTNLVTYISKIQRAISDGYQVD AIYTDLSAADFKNHRIAIAKLRRVGLHGSLLTWVNSYL TGRIMLVKISDNLSLPSVTSVGPQSGHLGPFLFYIND VNFLRCPKLCYADDFKLYAIKRPDTMEDLQAQVDLF	cd016 50	190	441

					ADWCRLNQMVNLNPSKCSITITTRKRRCPTTFESYKILGNPL TRDSFVKDLGVLLDSKLSFKNHISYLCASKQLGMIFR MTKYFRDVNCLKILYLSLVRSTLEYGSVVWAPYYHNDI GRIEAIQRKFTRFALRHLGECPNYETRCSMLHLDPLSVR REAAKAFVADLFRSNINCPFLTSELNVNIRRRVLRSHSF LVIPGARTNYAFNEPVSSMCRVFERCYPFFDFHLRPRRL KHDIXNFLREAYVF						
				TAAGTTTGA TTGTATATA TCCTCTATG TCATTGTGA TTGTAAAT CTGTTGATA ATAAAAAA GAGGTTTTG TGCTTATTT GAGAAGGA CCCATCGGT AGGCGTTCC ACTCAAAC GGGCTTTTC CCTCATATA AAACATAA AACA				MTNTVNNVQELREFEPINATLKNLSKENGFKVGHNLNTC SIICHDDDVKNIFTRTGDCIGFTETWLNDSIDKDALIKFD NFEILRHDRNSTTRVXGGVALFVRPEYKTKIHKTSRDS IEYMFCELKVVSKLWLVLSNPQPIKSSQLQVVVGVIY KPPTLXDFSEFFEDLGNISSTYNNIILMGDENINILVSNPM SNNFRLINSFSLSVNKSFPHTFHNTGSSLLDYFICSENL XSTLFSQVSVPAVSHHDLIYTVFNLPVPVASLKITSFRNF NRIDKEAVRNDPFPNFQNLGVDCETLASNFNHLKLIQ LVNKHAPLKTITLPTSCPWYSREAYNACVDRDLAYRF WKNRNTLENKLIYNRFRNRATNLIRIAKKQYSAKLFDPT LPSKTIWKNFKSLGFRDDPDNLDKFKPEDVNNCFIQKI ENYPTGNRQPTIIPDIVEEDQFKFRAVTEDEVHKAILRIK SDAIGYDEIPLKFVKLFLDLLVPYITLLYNKIISTNTEPTIY KLAIVKPIPIKNPSNPNDLRPISILPSLSKSGCEILLKDQIV QYVTFKKLITPFQSGFREHHSTSTIITKVCQDINASFDNS YVTVLJLLDFTKAFDLIDHRIFIDKLKEQFQSSDACDLM KSYLSDRSQVLQINGVRSQFRSTKRGVPQGSVFGPLAYT LSVNDLPSFTHCNKQSFADDTQIYLSCPVDQLPNKISLI NSDLSAISRYCKDGLVLNTSKTKAIXFSRQNTVLPPLP PIQIDNENVEIVDECLNLGIKMNKHLTWXDHVSLIKKKV YSTLRLTNCHRHTLSIDTKIKLVHTLLMPHFYIGNVFSFK MNVDTERHLQVCFNSCIRFIYNLRFPDHVSQFQRSIFDLS LSKPYQFQLLLFLFLKLNFKEPQYLFESLVFSRSRNRNNIII	cd016 50	519	778
				TTGTTTGG GATTAG ATGATAG TAGATT GCCAGC AGTACAG GATTTTT GTTGTTA TTGTTATT TTGTTTGT TTAAGSA GTTTATCT GAAAACC AAAACAG CACTATT ATCTTTTG TTTTGGCC TGATTGC TCCTTTCC TACGCGT GTTTTTCT CTACATA				Culex quinquef asciatus	Kiri- 5_CQ	Kiri	

				GATTTAT ATATTTT	ATAAGGTC CTTATCAAT TAAGTCAAT GTATACAA AAAAGGT GTACCTTCG GTATATTGT TGAAAAAT AAATACA AATACAAA TACAAATA CAAT	PVHHTNFMGNSFAVRGARLWNDLPHNLKSITLSEFKR RCKEHLKNSI			
Tx1	Tx1- 11_C Q	.	Culex quinquef asciatus	TG	TAGAAACA TCCAAAAGT AAATTATCT CTACCGAG CTTATTGT CGTCTTATT AGCAAAGT GTGAAATG TAAATTCAA GAAAAATG TTGTAAAAT AAAGTGAA AAAAAA	LYENIFGLIANNFVKVAYADDLNLVRNNHEFDVL QLVSYFSIKIKLNFQSKYLRNGCLSGPHQKEVESL NVLGVLCQNFNKLVEINYDYLNLKLYTLHLQKRILN LFQKAWLLNTYVLKWIYICQIPPNKHLAAIKSICGQF IWQGFVKVPRNELYLPVKGGLALTDVESKAKSLFIKNI LYSNADGNAPLDNFMGLQINNNTLRNMREWIVKADEL KLESHLSSKKIYDFHLSLNICKCTMEDMPLNLMWSNLL NNLDSNLFNSKCKTVLFCIFKDIVLNKKKHLHLYGIRGTD SKFCNCGLVESVSHLVKYCQTSSEIWRWLKNILNKRL KILVKDPDELLSLSIGDRDFKYKAALWLTQTVIVNLEK RFGDLNEFKQIRIVRENNKAIFERHFKHLNIF	IPR0 00477	11	83
CR1	CR1- 19_C Q	.	Culex quinquef asciatus	A	TAGGAAAC AGTCTGTT AAGACGAA GATAAATA AATAAATA AATA	IXRTQRRXFLGXFKCCREHSGKSATPPVPTLNDADDAEF AARVCTAIDSKDGTGACFGVDWRSGRFTHGSSDRITCGTI VPTIPASFRSSQHLVPRVLQSSPGRAVESLMRAPRLPDA TTSSPRQRLRTPASRSHQSRPGAGVGGVGGVSGPAPFGK YPSRSQHFLPDVVRTSSSCASTSATQTAVSPSFRAIQFS PGRVESPMPGAPRLPDATTSSPRQRLRTPASRSHQSRPG AGVGGVGGVSGPAPFGKYPSRSQHFLPDVVRTSSSCAST SATQTAVSPSFRAIQFSPPGRAVESPMGAPRLPDATTSS PRQRLRTPASRSHQSRPGAGVGGVGGVSGPAPFGKYPS RSQHFLPDVVRTSSSCASTSATQTAVSPSFRAIQFSPPGR AVESPMGAPRLPDATTSSPRQRLRTPASRSHQSRPGAGV GVGGVSGPAPFGKYPSRSQHFLPDVVRTSSSCAGPSAT QIANASTLSATVPSQLANDVQRPQSGVDRDDSSLLITY YQNVRLRRTKTNQLHLSSCCDYDVIAFSETWLNKIV DSELSSDYTFYRADRSRETSLVLERGGVGLGVKKRHFPE	cd016 50	101 1	126 1

[illegible]

615

TCATTCACC	KPVQKQIQVWPFGSSDALQDCFDTTDWNMFKAATYN
ATGGAAC	NXTDLQYSETVTAITKCIEDVTVTKTTTVRANQKPW
GACGTCATT	MTGEVYRLLETRNAAFRAGDEGLRTARANLSRGIREA
CTACCACCT	KRQYSRIAQFSDSRDTRSLWQGIQTITDYKPPQRTCD
CATCAGTCA	SNISLLNELNTHFAFEAQNSTTAQKTPPPSDQVMMLT
GTTTAAATA	PDSVRRSFSRINARKAPGPDNIPGRVLRDCAELTDVFT
AAATAACT	DIFNISLSQAVVPTCFKATTIIPVKKPSPSCFNDRPVAL
GCTCTTGAG	TPILMKCFERLVMHHIKSALPPSLDPFQFAYRSNRSTDD
CCCTTTGTC	AISTTLHSALTHLEKKDSYVRMLFIDFSAFNTIIPQQLIH
ACTTTGTCA	KLVELGLNLSLCNWLLDFLTGRPQAVRVGSGNTSSTITLN
CTTTTAATC	TGAPQGCVLSPLLFTLLTHDCTPSHNSNLFKFADDTTV
AGACTGAA	VGLISNRDETNYRSEVSRLAGWCSDNNLSLNVEKTEIV
TAAGCTATT	VDFFRAHTQHVPLTINGATVERVSSTKFLGVHITDLSW
TTTTTATGC	TDNTAALAKKSQQRLYFLRKLRRARAPPIMYTFYRGTI
ACTAAAAA	ESILTSCITVWYGACNASCRTLQRIVRAAEKIIGVSLPS
TCCTTTTAT	LQDIYGTRLTRKALCIAGDPTHPSHSFFSLLPSGRRRLRSL
CTGCACTGT	QARTSRLKDSFIHQAVRKLNSLPNLPSPSSASGTTEL
TTGTTCACT	
GGTTTGCAC	
TCTATCTGC	
CATGTGCCT	
TGCGCTGCT	
TTTATTTAA	
CCTTAITTT	
TATTTTATT	
TTTTCTATT	
ATGCTTTT	
TTACATTCC	
CTTATTGTA	
TAGTTTAT	
CTTATAITTT	
TATATTGA	
TCTAGATTT	
TAGGCTCTA	
CTGTTAGTG	
TTATCTGTA	
TGCACCGG	
GGGTCTGA	
GAGTAACG	
CAATTTCGA	

[illegible]

[illegible]

[illegible]

CR1	CR1-10_D R	.				TTCATATAT TGTAACATA TACACTTGT AATCATGTT TATCTATCT GCACACTA CTGATTATT AATAGCAA CCTGCACAT ATATTCATA TATTGTAAA TCTGTTCAT AGCTTATCC AACCCTGTAT ATAATGTTT ATAGTACAT CCATCTGTA AATATCACC ATAGTTTTC TATAAATGC ACTTATAAA CTTATTCCCT GTATCCTGC ACTTGCCTGC TATTGCACT GCTGGTTAG ACCTAAACT GCATTTCGT TGCATTGTA CTGTACAT GTGTAATG ACAATAAA GTGGAATCT AATCTAATC TAATCT	MATGGTGVVITLLVYCYFTSLTAIRVDPVRMFCSIYPL SYHNDNPHIYKDLGYKRRKGNELAKTSTGYTALAVT MVYPSLYNKITKKKKAMTTTQSSRTGLIACLMLSGDI HPCPGPHHFKTPREANPQTEDKRALTLQVCSSTLLCS SAQQANPLLSLCPTNPNTLKL SVQGVCEVKGAAAVPGWS	cd016 50	800	105	6
-----	------------	---	--	--	--	--	--	-------------	-----	-----	---

AAATAGA	TCAGACGT	DGERQGEKTHPRLCVGMSDRSFGREFAGRVCAFRGNK
GTAAAC	GTCATCATT	RPDPETQQQQQTEFLNDSASTMATNSVNRKRLNPGVVK
AACCTCT	CCGATTAA	NRKHAFENTVNNSKIVWDPKMKPKGLLGRHLNIRSILS
TTGCTTG	AATGAACC	KSEQITHLLTNSLDYLCCLSETWLYKYAPSAAFNISGYT
CCATTTT	TAGCATGTT	MFRKDRGAKGGGVMIYVKDNLQCEEINWLCDEFEFC
AAGACTT	TCCAAATTGT	LGLNVRLSPQMSFKIIVIRPPSSDGTFFDRFEKLLKESN
GCTGTCT	GCCYACCA	LNKETVILGDFNINWEDTSCRNKLKRITDKFDLVQMING
TTTTTTAG	TCCTAAATGA	PTKTTSSSTQIDLFTNRPERIHKTYNMLAGISDHNLTLV
TCTCTGT	TCTGTGTGT	ARKLSKKRFKFSAREHEFIFIGIPNRDRDGFKNAVEQIDWD
GTGCAAA	TAAATGTGT	RLLLGTDLEKDRGSFVGTEKTVKNSALNLNANTIKIYY
AACGTCT	TTCAACTAT	PGLTLTYHTKTNERERFCVKVKSASDRNMFTTLRN
TGCCTCC	TTTTATCTC	RVVKNIRKAKADYELTLIEKAKGDSKTIW'NQLKKVVRH
TATTCAG	TATTACAAA	DTKNKKMLEVMIDGHITNNAACKIAEAFNNFLLTQFQIW
CGTTGGC	ATGTTGGTG	SHCVHQKFLTCALFFINSVSNMVPCLCPSKISNMCSLNS
GCTGTGG	TTTTAGYTT	QPIFSISNIESDALQAIKSLKSSRTKETFGIDSLMLKELSS
CTTAGTT	AATTGTAA	TLTTPITKLLNRSIEAIFPCVWKSSAVIPFK'NGDHYLSL
GGTCAAA	ATATCTCT	NYRPIHIPTVSKVAEKLIANQLIKHLN'TTFSLHPMQFGF
GTGCCTG	TTTAAACATC	RSQHSJETANCFLIENIKHLLDKGGVVGAVILDLLKKAFF
TCTGGTA	TAYATRACC	TVNHKILLSKLTKFNMSPELKWFESEYLSNRSQSVRINN
AGAAACT	ATGTCTATT	YHSSPLSLPTGVPQGSILGPLLFSLYINDLPVVCQNLIQM
TGGCTGA	TTAGTACTT	YADETVIXTHGKNTTQVAEKL'TQAMAHVSEWLNQSCL
GGCCAAG	TTTAACTAT	KLNVSKTVAMYFSKTNVNNLTDIYVDGDKLQIVSDFK
	GTTTGATW	YLGVDQDSKLTFKSQIKKVCNRRKFNIRSHLSLQAAXIF
	GATTGTATA	MHSMILSHITYCLTTWSQANKITTLKPESLYKQSIKILXK
	ATCTTGTA	KPLRHHHCILQKHKLLSWENVIKY'NICLLYKIIHGLSS
	TTGTTTTTT	PPLSQFVNIRNSTHRA'IRGAARGELYRPTPEKYLW'SGCL
	TTAGGTGT	FSYSC'TGVEFCPWQHQRYNLL
	TCAGGGAC	
	AACAGATA	
	AAAAATAG	
	CCTTCKGGC	
	TAATCTGG	
	TGCATTTC	
	AGAAATGC	
	TAATTAATG	
	TACACTGTC	
	CCTGCCAA	
	ATAAACAA	
	ATAAATAA	

[illegible]

TGCCAAA	TCTGCAGCG	GSDPAIYLHLSAISAVSLAVGDADQLVLGDFNLPHLS		
ATTCGAC	AAGTGTTC	WLPNSDSNILFPENQHEFDGLTNLSLQGVNSIKNSRERI		
GGTAMCT	CCTAATGCT	LDLIFVSDVSVATVSRCDPVSTPDSYHPALYLSVHINFE		
ACGTGCA	GTATCAGC	TTNQSASQRGNTIRDNYESICLTTRHYTGLNSHFDRLN		
GTTCAAA	AATTCAGC	WSTLYSFTDLDSA VLVWENESVCSALDTFVPSYRWKLIK		
ATTATTA	ATACGTAA	ASLVLAQAFVLHIPLSQVTVGNSLKPWFSTRHLSYIKNTK		
CTACTAT	CTGAGAGA	SLLKDIKNPRFFYAEYIAECRQNIKSDPKKFYSFVNSKRR		
TCAATGT	CATAATTAG	SNVPPCLHLGNLSESSDINISLFAITFQSTYSLPHINNI		
CAAAAAA	CGTCAAAT	HIPPFKISDV L NGMLTIDESFAGDPKIPSCIUKQCAASLC		
ATCAATA	GTATTGACT	GPITFLNLSLQLSFFPSVWKTSLPLHKKGSKTDVGNV		
ATTTAT	TATGTATTC	RGIACL SAIPKLFERLVT CQIQHMCSSIIISTSQHGFVKKRS		
ATCCGCT	TTATGTTAC	PSTNLLEFSRLHCSFLQKKQADVITDFSKAFDSVNHS		
TTGAAAA	CTTGCAAA	LIYKLSLIGFPYLLDWISSYLSNRSQVLFKNVSSLVN		
GGAASAG	CATTTACTT	VTSGVPQATSHLGPLLFIPLINDLPQVMMSYCTTFMYSD		
TAAAGGC	AATCTCGTT	VKICMSFSNLSHLLIQSDLNKQAWASTKLLSLNYSKC		
TTACCAA	TAAAAATT	KTMSFSRTTFTHTTSYTLGNSLLERVVKINDLGVLFDQKL		
ATCAMGT	ATCTGATTT	CFNSHISVMVNKATGVLAFAIKRWAKEFNDPYTTKSLYV		
CGACCAT	TGACAAAA	SLVRPIMEYCEVWNQYVYQQNLIESVQKQKQLLFA		
GAATAGG	TTAATGATA	NLNWDPRHLPHLSRLKLIDLNTLKHRENFCKIF		
GGGTATG	ATTGTTATA			
TAATGCC	AGCCGAGG			
AGCGCCA	TTTGATAAT			
TCCGCCG	TAACCAAA			
CAGTGT	TGCAATCCA			
ACGATCT	GTGGGGCTT			
CTCTCG	TCCAGTTTC			
ACGCCCT	GAATCCGA			
TACTAGT	ATGAATTTA			
TTTTACG	GTTGTGTTT			
GATTTCTT	AGAAGAAT			
CTCTCGG	TATTGCACA			
CAACGAA	ACTAAAGA			
AAGTTGG	TGTTATGAA			
AAAAAGT	AAAATAGA			
CATGTCC	TTCAT			
AGGTAG				
ATCGAAA				
AACGCGG				
TCGCGAT				

GGGAAGA
ATAAACC
AGTTAA
ATTCTAG
TCAGATA
TGGAAAT
ATTCTG
GGAAGCG
ATTAATA
AACGTTA
TCAACAC
TTAGTGG
CTGAGGA
TTCCACG
ACAGGAA
GGAGAGC
TTCCAAG
CGGAGC
AGCCAAA
CAAGCCA
TGCTTTA
GAAAGAA
GGCATTC
GTAGTTT
CGTGACC
TGMCGAA
GACTCCC
AATTGAA
ATKTCAA
TTTTGTA
AGTCAAA
GTCTAT
CAATTGC
CAAACTG
TTTACATT
TCGCCCA
AAGTCTT
AGAAAG
GACGAAA
TTTCAAC

AGATGTA
ATATATG
CAAAACA
TAGACCG
CATTAAA
CATATAT
TGTTCG
GTTTCCG
TTGCCG
GTTTCG
CTTTGTT
TCTAAGA
ATAGATC
GGGAGTT
ACAAAT
TACGATA
GTATTGA
ATTATG
CTTACAT
AGTCATT
ATTGCTTT
TCGGCGA
GCCGAGC
AGCTCGA
TTGTCTA
AGAAATA
TATAAAC
TTAATTA
TTTATTTA
TAATGAC
AATTATT
ATTATA
TAAAAAT
TATACAG
TGTATTG
GCCAGCC
ATTTCAT
TGAACAA
ACGTATT
GCCGTGG

TGCTCTA
TCCTTAA
ATTACGA
ACTATAG
GCAAGTC
AAACAGT
CATTTCG
CAACGAT
ACTGACC
GGTTGTG
TTTTATT
GAGCTCT
TGAAATT
ACCTTTC
AATCGCG
GTTACAC
CGAACGT
ATATAAT
CTTAAAA
TTGGAGA
ATAACTT
TCTCAAC
CCCGCTT
TCAACAT
TCACGCA
CCGTATA
TAAACAT
TACATCG
CGCCAAT
TAACATA
TTTAAIT
GAATCGC
AATTCAT
CTGCTGT
TGCCACA
AGKGCTT
CTATTCC
CTATCGC
CGTCGCG
TCTGTTCT

TTACACC
CGCTCGC
CAATTGC
TCCTCA
AGCCAC
GGTCTC
TCCTCGT
GCACTTC
ACTTGA
GCGCTC
GCCTCIC
TCATCGC
CTGGATA
TTCGAAT
TGGCTCT
CCGAATC
GACGACT
CTCCCAT
CAAGTAT
TCTAACG
GTCTCT
CTCCGTTT
ACTCGA
GTGGGTC
GCCTCIC
TCATCGC
CTTGCTT
GCTGCGG
ATCTCG
AATTGGC
TCCTCGA
ATCGTCG
CCTCTCC
CATCGCC
GTATCTG
CCACTAC
CTGGTGT
TCTGCGT
GTCTCTTT
AATCTCC

CACGGCC
TGCTCTTC
TGCAAGA
TTCTCTTA
CTTGCTT
CTTAAC
GCATAGT
GTCCAAA
CAATCAC
TAAAATG
AGTGTG
TCATCTG
TTCCGCA
AAGAAAT
GCGAGTT
CAATGT
GTATCA
TTGGAGA
TAATTAT
TTGCTTG
TTGGCTTT
GCGACAA
TTGCGCT
CATGTAA
AATGCGC
AGGAATG
GGTCAAA
ACGGCCG
TGTAGTT
GATCTGA
TCTCGAA
GACGAGC
GGCTTAA
CTTGGTT
CTGCCCT
ACCTGCC
TCGTACT
TACTGG
AG

RTE	RTE-1_DF					ATGTAGG GCCAGAC CAATAAC AATCGTG CAGTGT ACGCCCC AACCGAA GACGCAA ACGACGA CATCAAA GACGACT TCTATAC CGCCCTT ACTACCA CTCTCAC TGACATA CCTCGAG GCAACAT CACCGTT CTC	TAAAGGAA AAAAAAA TAGGTTTAG TTATAACT CCACGTCAC TTGGCGCA GACGTGTC AAAAATGA TGTCATAT TTAAATTGT AACAAAA AGTGATTT TGATGTGA TTGTTTT GTTTCGAT TAGTTGCC TGAAGACG GTTGCCGA GGAGCAAC CGAAATAT	Drosophila ficusphila				MGDFNAKIGPNFTRMRSATGGHGIGTRNNNGDRLVEV CQMFQIIGGSMFPHRDIHKYTWTSFGNTRNQIDHICIS RKWRKSLDVRNKRASIDSDHELIIGELSIKLDNRNPR NADDSNRHPSPLNIHLLGETDISARLSTLRTHLLPQQ GQTWEQTCDTLRTTAEGHIGTQRRRNWISSATWDTIN KRNSKLLADRSNRRDEHRAKCAVKRAARNDKRAL LDRLADNAERAAENNMRSLYQTIEQISGGYRRKTQPV KDANGILLTSDDDDQIHRWRQHFMESNSDMPSRSDDYR SVAPANGNRRISSAPSVREIGNAIKKLNKKAAGMDGI PAEILQVDTQLMAEILHPHFSSAWESEHIPDSWKCGVIV KIPKKGDLSDCENNWRGITLNTSYKVLATILNERLTEKF EPTLRDEQAGFRPHRSCVDQANTLRITEQAVEWRAPLY LLFIDFKKAFDSINRAAIWALARKGVPHKVIALIKALY DDANLAVLHNGKTSAPFQNTGTGKQGCPLSPLFNIV DEVREVCCHKRGITWSFTRHLEDLDYADDICLISHKLS DIQSKVDDLVLASRVGLQINTAKTKAMRLNHASQGN IVDGNPIEYVESFPYLGTVISSNGGVTEDEVTSRLAKARSA FGRLYRIWRSQRIRKTKIRIFNSCVKSVLLYGSETWLT KTHIQQLQAFNTKCLRIICGIFWPNRISNIDLWNMTGETAI RLQIRKKKRWIGHALRKDQXSIVRTALEWNPQGSRR VGRPKATWRRTALQELSQYNITWDTAKQTAXNRVRWR NLVEALCSQEEL	cd016 50	381	637
Pene lope	Penel ope-1_DF							Drosophila ficusphila				IVNNKWTLKEHTXTPSKVFLDIVKFCIQDNRYLYETK LXIXLKGMPMGSPASPVVADIVXETLLQNTTEKLRRPPR LLTKYVDDLFGIHEDEIENLTTLNAFNXXNIKFXMELEK DNCIAYLDTVIRRGNELKLKWKPIASGRIINFNSRH PKMMIRNTASSCINRMLRISDAIFHEEINEIKTILTNDDFP KHKMISLLDKYKSDNNQRPKERTYISVTVVPLSER LAQSDCYDKENILIAHKPDNTLKQIFNKTCKIAPTDDK DVVYRIPCNGDGKNNCNSIYGTGTSKLKTRLSQHSDF KLHQNTQKTALMSHCAKQGHSPNEETTLAQEQHY RKRYMMEMMLHIINTPTNIRMNYKTDTENCAHIYRHLG K	cd003 04	47	129

[illegible]

[illegible]

AAATTTTAT
ATGCCCTCG
AGTCCAAA
TGCTCAACT
CCGAGAGG
CAGTCTTT
GTCAACGCT
ATCGTCTA
TGGAAACA
ATTACCCCA
TGAACATTAA
AACTATTAG
GAATGCAA
CTCAATTTA
ATTCCAAA
GTAAAACT
ATTTTAAAG
AAACAACG
CATAACAA
ACAAATAA
ACAAACAG
CAAACTTCA
TATTAATAC
GTAATTAA
AGATAATA
AAATAAAT
AAAAATAA
ATAAATA
AAATAAAA
ATAACTAA
ATAAATA
ATATAAAA
TAAATAAA
ATAAATA
TTAGTTCAT
AATAATAAT
AATGTACTA
GGTATTAA
GTAGAAAT
CCTGTCTAA

[illegible]

CR1-4_DR_h	Drosophila rhopalodia	<p>GTGAAGC TGAGCGA CGGAGGG GAACCTC GATATGT GCAAAGC TCACACG AGCTTTC TGAGCTG ATCAACG ATGTGG CGGCGCC GATAAGA GCACATT TCGAGAT TAAGAAG CTTCACT CAATTTT GTATCTC CTATGTTT ACACATA ACTATAA TATTTATT AGATTTA AAATTAA TGTAGGT CGCTATG TTTTCTTT GTGTCCT TTTCAGT ATGTTTC CAATGCC TTAATT</p>	<p>FKLLRKTVTKLIDSAKTNFYSSRFQTTASTSRTWKEIKSL GIGRKPAPEQXVDIEELNTSLSKCNVPEICENNMYLRS VPRNLCDSEFRFCVNQCEVLVNILKIKSDAVGLDDXN PKFIKVLPPKILKYITYIENTAFTTSTFDRWKIAKIPVPK ANKEYRPISILPYLSKAFENIMASQINSFIHDNALLNNQ SGFRKHRSCTSAILKIANDIKEYADNSNVTLVLDDYSK AFDTVNHNVLCACLVNQKFSTAVKLMTSYLQNRQQ SVVVSNNVSSFAFKRGVPGQSVLGPLLFTLYVNDLPSM LSSCDVHMYADDVQMYVSRPLNRINECVDICNNELVKI SNWAKENGLGNPSKSKCIHKKKLSLNIGPLLLNDSPI EYVERATNLGHFENNTLNWSDHINTAVGKTYGLLRSLY LVRNPLNVRMLLAKSCLPTLFYGHLEFANCDSNSQH KLKMAFNSITRFVFNLRFDHVTRLSYNIFELHLADWIK MRVLVYVYQKIIFSCLPIYLYETLRFSTSRRSNIIQIRH</p>	<p>SSGPRASDSFNQSRKKLNNFSLSHYQNVRSLLGKLSKLH VNSVSFDANVIAFTETWLNSSVSDCEIFAGSYVYRCDR SMRLGGGVLIADVKSDDLVSQRIPIGTDSILEFVAVRIKLH DYFVFISSYIPPRSDISLYLQHLAQIQSIHSTLGYNDHLI VMGDFNLDPISWMTSDISIALVPTTAHEFVHGLIDLSLG QLNSILNCSKKVLDLIFVSEFANAFVTRISPLSNPEDAYH PTLEVTLASAPLOSCSEIEPASFRCFRKAFAELNRHLS</p>	<p>cd016 50</p>	<p>514</p>	<p>765</p>
------------	-----------------------	---	---	--	---------------------	------------	------------

				AAAACTA CCTGAGC AGCAAAC TGAATGT TTCTCGA GCCGAAC TAACCGT CTTGAAG TTCAAGT TTAAGCA GAGACGT GAAATAT CTTCGTTT AAAAATG GACTTCC TGAGGAA CATCTTA ATAAGGC TCTCGAT GTTTCTAT ATGGCCT AAACATG TTATTGTT CACGAGT ATGTAAA CATGCCT AA	GTAAATAAA TAAATAAA	ELDWSEFMDSNDDPLHIVEQFYGILYSAPDFTVPSPSAN APQKPPWFSRRLSFLKNMSRLFKPKRTGLQVDHSMY LIARSNFLVHNSVCYNGYLNRCRIEFQDNPKQFYAFVNS KRKSQQFPSCLDLDTTSSADADIANLFAKFFQSTYSSFS YDNTSSYPYPFLTPTVIFTPLISEDVLQNLKTIKMSFSSG PDLVPSCVIKKCADALCKPLSKLFQJSLHSHFFPEVWKES FIPLHKKGSKSDIKNYRGIKLSAIPKFEQLJTAQLQHQ CSSVMSPQTQHGFMKRRSTTTNLEFTSVIRSAFKNNIQT DVIFTDFSKAFDSVNHQLLLKLLTLLGFPIMLTKWISTYL SNRTQRFVAFKSTISNSVHNSGVPPQGSHLGPLLFGIFIND LPQVVLSSYVMYADDVKLCLPLSNPNSHDLLQADLN NLHWCQNLFLNNSCKLMSFYRKRLLLTSYSIGSND LERINSVTDLGVLFNHNLFSDHIGTIVSKAGVFAFVK RWSKEFEDPYTTKHLYISLVRPILEYCSCVWSPQYTNHQ DLIESVQKQFLFALRGLNWEHRLHPSYVDRRLRLNLP SLKDRRTCHGVFMFLHKLITGLVNSSYLLGQIRFSIPVRST RYFRPIALDICKTNFEVHDPFRVLCSEQYNELYFLIYSECS VNLIVTNLTSLSPPT			
R1- i_DS i	Drosophi la simulans	CACCACG GAAGACG AGTTGAA GAGTGCC TTGAGG CCACGGC AGAGATC CCTGCAG CAGTCGT GGCTATC AAGAGCC TTTCGGCA ATCGCAG	TGAGCAAA TCCTGATCG TCCTCGCGA AGCAATGC TTCACGGCC ATACCGCG GGGAGTCC TGGATTTC CCTCNTAAA TGTTATTAT GTTGTATAG TATATAGTT TAGAATAA	MMRLIQLNLNHCCTAAQDLLVQTVRERRVELALLSEPYR TPASPDWAFDRTKKAIIWRCSETQITDVLSDIGFVRA RVGRWWVYSVYLAPSLTLIAFSRALDRLAADARGRTQ VLIAGDFNAWSESWSSTTNARGRMVLEAFATLDLALL NHGNRHTFRAGLSVVDLTFTSGSSYRLARWRLSEDY TGSDDLAIICDLGSPSSQAQTPAQARIKYKTDSLDTQVF REQFLPTASRQGAELTAVELMRRLKAACDASMQSSRPH SRQRAPVYWNQEIETARRECLARRRYQRRARGAESFA ERQSEYRARRKTLKLAIRESKRKCFDLDCDSADSDPWG SAYKVVVKKAYTRTPKLLDPAMLSVAEHLFPVMDGL RPADPATGDHVDSDATVSSDEILEAKLLKDKGKAPGPD GIPNRALRLALHLQPDFAKFTKCLEEGVFPSCWKIQK LLLLPKPGKPEEPSSFRPICLDGTGKLEKLCVIRLERAI	cd016 50	473	720		

TACGGAG	GGGTTCAAT	ADAGDLSRSQFGFRKARSTVDVNRVVEVAAQAIDGT	
GGAAGAC	AAACGAAT	RWKGGKEYCLMVTLDIRNAFNTARWDRILEALIGFSV	
TGCGGTG	TAAAAAAA	PAYLVRIIRSYFSDRVLLCESSEGVFRHQVTGGVPQGSV	
ATAGCAG	A	LGPLLWNTMYDGILRLPLVGRSEIVGFADDVALLVVDK	
TCCCAGC		YPGKAEECNQIRAIHFWLSSMGLLELAPEKTEAVLISS	
CAATCTG		RKAVETATVEVGCTKVSSSRAIKYLGVMIDTRLSFREHL	
ACGGACC		AYANSKAAAGVNRALSNIMLSNRGPKQASRRLLTSVTRA	
CGTGTAT		TMLYAAPVWAKALGTESYAQGLNATHRLSALRVCCAF	
CAAGCGT		RTVSNAAALVIAGIPPLDLLAKENEVVFNQTHGRRRLSPG	
GGCAAGC		VKKAIRAEARQHTLETWQHRWTNEPKGRWTQQLIPNL	
TGAGGGT		KPWLERKHGDTTFHLTQLISGHGCFRSLMRFGHEDAD	
AGGATGG		DCPWCGGGRSETAEHVLFSVCVKYARERSSLETVLGGRL	
TCCCAAT		NADNLVPFMLQGEAEWQAVNSFAAAITTELRAERAR	
GCTTGAT		RVHE	
CAAGGAG			
CTGGAGC			
CACGCCA			
AAGATGC			
TTCAGAT			
GCCTGGA			
AGAAGGC			
CACATAG			
CGGCCCA			
CTGCAGA			
AGCACCG			
TGGACAG			
AAGCCAG			
TGCTGCT			
TCAGATG			
CGGGACC			
GCGGGAC			
ACAAAGC			
CGCCGAG			
TGCCCTA			
ACGAAGC			
TAAGTGC			
TTCITGTG			
CGCAAGC			
AGAGGTA			
GCCAAGC			

637

638

TTTCAAG	GGCAAAAG
ACAAGGA	ATCGAAAT
CCTAAAA	CGTAACCCC
ATTAAAC	TGAAATTG
GCCGATG	ACCTTTTT
GGACCGA	TTCCCTTAG
GGACGTG	AATTAATAT
CATAAGC	CAACGAAA
ACTTGCA	TCCTTGTA
AAGAGGA	AAGATCGA
CCCCCTG	AATAATTTA
ACGACCA	ATCAAAAC
CAGCCTA	ATAAAACA
AACCTGG	GAGTATTAT
CGCAACG	ATAATAAT
AGTACTG	ATATAAWA
GAAATGG	AAAAATTC
AAAAGAA	ACAATGTTT
AATTGAC	TGAAGCCG
ACACTTA	TCTCAACGT
TGGAAC	CAAGAAGC
CTTCTTCA	CCACTCGAT
AATGCTT	TTTATTGAC
AGAAAAAT	CCTTGGCGC
AACAAAA	GGCMCTGC
ACCTGAC	CGACCCGA
CACAGCA	CAAAAGGT
CACCCCA	CAAAATCG
TGTCCTC	TGACCCCTG
CTAAACT	AATTTCGAC
TCTAAGT	TTTTCCCTG
TCCTCGC	GTAGCAAC
AAAAATA	GATTATACT
AGTAGCA	GACAAATAT
CCATGAC	CCCTGAAA
ATTAAAC	AAGAATAA
ATAAACA	CAGATCAA
ACACACA	TAAAAGCTT
TACTATC	AAAACCTAC
AATACCC	GCAACCTA

TTAAAAAT	CTACTCCC
AGGTAT	TACTAGGC
TGGAAT	GCAACATA
CAITGG	AAACGAAG
AGTTACC	ATTGGCAC
AATAAAA	CTAAACA
GAGACGA	GTAAAAAC
ACTAGAA	CTAAAAATGT
GCTTACA	TTTGATGTC
TAAAAAA	GTTCATG
AGAGGGT	WTGGATG
GTCCAGA	CCCCCTCGT
TCATGCT	CCTCTGAA
GGTAASA	CCCTAGGTG
GAAACCA	CGCCGCTGC
AGCTAGA	TGACCCGG
ACGAAAT	CAAAGGGT
TCTACTA	CAATCTCGC
TGCTTAA	GACCTTGA
	ATTGACTT
	TTTCCTGGC
	AGCAACGG
	TAAACAGA
	CAAAATCC
	CTGAAAAA
	TTATAAAC
	AAATTATA
	AAACAAAT
	TATAAAG
	CATAAAC
	CACGACTC
	GGCGAAT
	GAAWAACG
	ACGATTTG
	GCACCTAA
	AACATTAA
	AAACCTAA
	AATGTTTK
	ATGCCGTTT
	CCATGCTGG

[illegible]

[illegible]

				TAGAGAA TGAA	ATTATGTAT TATGATAA AGTA	YAVVPVPTKSDKVKNLTTWVKSRVLQWSITQVTPSRP RPQFAKPSRPQETIAKKVEAKLADGDVIRGAILRLTSDDT IAPNNQETLDALITKHPHPPEPTYFPNAPDKAEVLAPVDP KEVLIAISSFAPGSAGGLDSLRLPQILKDLVISENGEPATNL LNAMLKFLDLVLRGDPKSPICPIFFGASMTALRKKTTGGI RPVAVGNTWRLISKIAVARMSSTLVEIFSPHQLGVGIK GGAEAAHAARIYWNYSHDSPKAFKLDFRNAFNEJRR DTLLQIIREKPPNFYRFVSQAYSSPSEIFFAGTPISSLCGV QQGDPLGPALYALVIQPIVDLIDTELNMWYLDSDTIADS PEKVLAMAMQTTITRMGTVEVGLHLNTTKCEVGILGVVDD QTRDEIFELFRSTAPGIQKISPESAMLLGVPLTDEAIPIL GAKTEEMEKISTRLLKLSHSAFFLLRASISTPRLIYFLRC APTWRFRDALTYDAKLKNAMEGIINCSLSPQSWLQSS LPVNRGGLGVRHAIDVAIPCFLASTYSSLPLAERLLPPLH HGIDTAIQEGERWKSLSPLPLEIRGQAIWEIPQOEG VINHLTTNATTLEDKARYHAMSPQNVGAWLNAIPSPQL GTHLPNETFRIAAALRGDQCQPHKSCGEQVSTKGYH GLSCKNSAGRHSRAAANDVIARALRSAGVPIKEPAGC AREDGKRPDGLTLPWARGKSVVWDFTCSDTFAPSYLP QTSIRPGVAAEKAEDRKKRYEQLMDSFLFVPIAIEITGI WGKDLGSLFIKEIGSRITNITEEKRA TSFLLQRMSTTVQRG NVASILGSLPKGHKLEEFML					
				CTTCCGT GGAATTA CTGGACA TTCTAGT CAAAGGT GCGCTCA CCTTTCGT CACGCGG TGAGACG CCGTAGG AGAGGAA AACGGGC CGGCTCG CTGGTGA GACTCCG CAGGCGT GGTTC TC GCACTCC	TGACTATAA CACTCCACC GGTCACTCC CCCTCATAC TGACACATGC CACTTTAAC TGCAATTCA TCACTTTGT CGTCACTCG TCACTTTGT CTCTTGCT GTTACTTGT TCGTTAGTG CACTTTATG CTTAATATT TTTCTTTT AACTTTTA ATATTTAA	MRRDRDFTSSVLCFTETWLCGSPDSALQLAGFQLLRA DRDTLSGKTKGGGICFYNNWCNDVTVILQHCSPDL ETFEINCKPFYSPREFASHILVGVYMPAGNVHEAQRTL ADQIRRVERTNPDSL VIVLGFNKGNTHEL PKYRQFIK CPTREENTLDHCYTTVSRAHYAVPRAALGHS DHVMVH LIPAYRQKLCKPAVREFKWTSEALDLRACHDC TD WDVFRATDSLDEFTEAVTSYIGFCEDSCVPSCTRVSYN NDKPWFTAELRLRLQKQAFRSQDKDLYTEAKYKFS KAVRDAKRLYSEKLQQQFASDSASVWRGLKHLTNYK PKTPHSMNDLRLADELNEFYCRFERQCDDPHPSSTNLL QSPSPSPSGAHASSISPPPPATTL SILERDVNRLFR RLNPRKAAAGPDSVSPHSLKHCADQLSPVFTDIFNTSLET CHVPACFKASTIIPVKKPRITGLNDYRPVALTSVVMKSF ERLVLTHLKS LTDPLDPLQFAYRANRSVDDAVNMALH YILQHLDSPGTYARILFVDFSSAFNTIPSLQLDQLSQLHV PDSTCKWITDFLSDRKQHVKLGHKHSASRTISTGSPQGC VLSPLFLSLYTNSTCTSSHQSVKLLKFADDTTLGLISGGD ESAYRWESDHLVSWCSQNNLELNALKTIVEMVVDFFRN	cd016 50	472	725		
Rex 1	REX1 - 1_GA	Gasterost eius aculeatus									

				<p> GCTGCCA GGCATCT TTCTCTCC AACGTGC GCTCACT GTGCAAC AAACTGG ACGAAAT TCAGCTG CTG </p>	<p> ATTTTAA CTTTATTCC CTTGTTTTA TACTAACCC ATAGCCTTA GCCTTATTC TACTAACCC ATTGCATTA GCATTTCAT TCCACTTTA TTTTATTAC TTGTGCACT GTGTCTTG TCTGTCTAC TGTGCGCA CTAACCGCC AAGACAA TTCCTTGTA TGTGTGACA TATTTTGGC TAATAAAT GTTTCTGA TTCTTGATT CCTGA </p>	<p> RAPSPITLCDSPVITIVDSFRFLGSIITQDLKWELNISSITK KAQQLFELRQLKKFNLPKTMVMVHFYTAIESILCSSITV WYAAATAKDKGRLQRVIRSAERVIGCNLPSLQDLFASR SLKRAKKIAADPSHPGQKLFVPLPSGRRRLRSIRTKTSRHT NSFPSSAVGLINRARSPTA </p>		
RTE-1	RTE-1	Gasterost eus aculeatus	<p> CCTCTCT GGGAACC GTACACT TATCGTG GTGGAGA GGTTTGT GTGTCCC TATGAAC CTGAGGG CTGTGTT GTCGGGA GCCTTGT GCTCTGT GTAGGGT TACCCTT GGCAAAAG </p>	<p> TAAGCGGTT GAAGATGG ATGGATGG ATGGATGG ATGGAT </p>	<p> MKKRRKREGVTRPGGSPFPFGARPRWRARQRASGGRV YRGARPGTARRSDAAAQHSTPHGLTTCGGNRWGRVRC YKGGSDSGSRNRNPRQMLALGTWNNVTSLGKKEPEL VREVERYQLDLVGLTSTHSLGSGTKLLDRGWTLFYSGV AHGVRQAGVGILISPLRSAAATLEFTPVDKRVASLRLRV VGGKTLTVVCAYAPNSSEYSAFLETNLNGLHGAPVGD SIVLLGDFNAHVGGDDGNWRGVIGRNGLPDLNPSGCLL LDFCASHGLSITNTMFEHKDVHKCTWYQSTLQGQSMID FVIVSSDLRPHVLDTRV/KRGAELSTDHHLVVSWSVRGW GKTRDRPGKPKRVVRVNWERLEEAPVQKIFNSHLRRSF SHIPVEVGDIPEWSMFKTSIAEAAA VSCGLKVLGASRG GNPRTPWWTPVVREA VRLKKEAFRVMISGGTPEAVAV YRQARRAAAAAVMEAKQRVWEEFGVTMEKDFRSAPK CFWKTRHLRRGKRGITQAVYSKDGTLLTATEEVIGRW KEHFEELLNP7TTPSLVEAELEAEEGSSISLVEVTEVVK QLRSGKAPGIDEIRPEMLKAMGVGGLSWMTRLFNIJAWK </p>	<p> cd016 50 </p>	616	870

				TGGTCTC AGGCGAG GGGCCAG ACTAAGA ATGGTTC AAGACCC CC		SGTVPKEWQIGVVVPLFKKGDQRVCANYRGITLLSLPG KVYSKVLERRVPVEPRIEEEQCGFRPGRTDQLFTLS RIIEGAWAYAQPVYMCVFDLEKAYDRVPREILWEVLRE YGVRSLLGAIQSLYAOSESCVRVLGSKSKAFVGVGL RQGCALSPILFVFMDRISRRSRGEEGLQFGGLRISLLF ADDVLMASVCDLQLSLERFAVECEAVGMRISTSKSE AMVLSRKPMDCLLQVGNVSLPQVKEFKYLVGLFTSEG KMECFGRRIGAGAVLHSLYRTVVTKRELSRKAKLSI YRSIFVPTLYGHEGWMTERTRSRVQAAEMGFLRRVA GVSLDRVRSSAIREELGVEPLLLCVERSQRLRWFGLVR MPPGRLPREVFQARPAGKRPGRPRTRWRDYISALAW RLGIPQSELVDVAREREVWGPPLLELLPPRPDPG LDXRXHGIXXVQEGSYWKXGGVALYVKEHYTSLXVKI GCKEGXLETLWVKIRGERREEXLJGVYYPRLTKNSLE NWQRLHGMXFNYPDISWDEYSARSDRSQSFLTCEIEL YLTQEVYEPTRGKALLDLVLAKGDDLVSNLRXGKLG SDHEMITFSHCKPGKASRVEIPDFRKADFKLRRLVG XGLRNEEGNTRRVVFPQGSNPKSSKKVYSLSQERQOK SWETPLAQQSHGPPPEIKARSLQGMVERNYYQRGVLS CPHLKSLILGKLSTNSGDWLVRDYGPMPKGIQEEWSFL KEAILKAQRKSIPSHKKGSKRAGRPWLNREVMDLLKL KREAYKGWKS GTTTKEEYSAAARTCREQIRTA KALTEH KLATRIKDNKKSFFRYVVGCKRNKGSIGPLINEMGRLLT DTQDKTELLNEYFMSVFHQSKGISLDRIQDDHGGGDH PPVDVVLVGEHLQRLNYSKAGPDELHPRVLKELADVI AQMMAKLFEDSWHSGEVPEDWKKANVVPFKKGRRED PGNYRPISLTSPGKILEIKIETTHRLTEGSILKGSQHGF MGRSCQTNLISFYDQVTCHLDRGNEVDVIYLDPKKAFD LVSHDVLVMVKGKCGLDYFMVWVWLGWNLGRTQRV LIDGTESSWQRTVTSIPQGSVLPVLFNMFINDLDSGITS NLAKFADDTKLWGSVVTLDRLAQAADLDRMLMKWAD LNLMSFNTEKCKVLHLGRNPNQHTYRLGSALLASTTTE RDLGVMTDNKMNMSHQCDVAAGRANQTLACHRGILS KAKDVILPLYSALVRPQLEYCVQVWLFHFKKDVVKLR VQRRATCMIRGLETKTYEERLRLDGLFSLEKRRRLRGDL VAAKYIRGVHQGLGEHLFTKAPQGRTRNNNGHKLIEDR FRLDIRKNFFTSRVPVWVNRPLPEVVQSSTLAIFFKKRLD AHLAGHI	cd016 50	521	770
CR1	CR1- 1_Ga v	Gavialis gangetic us		AGCTTAG TTGCCCTG TACACTA ACKCCAK AAGCATG SGAAACA AACGTAC TGACCAA CAAGCAA CACGACC AGCTCCA CCTMTGA	TGATCCAG CAGTCTTC CTGCCACG TGCAGGGG GGCTGGAC CTGATGATC TGTAAGGTC CCTTCCAGC CCCTAACA ACTATGAA A				

CR1	CR1-7_Ga _v	.	Gavialis gangeticus	CC	TGACCCCA GTGCTCTTC CTGCCATGG CAGGGGGT CGGACTAG ATGATCTGC TCAGGTCCC TTCCGACCC TAGAAACT ATGAAACT ATGAAA	LVRGLRVLDPLGDSHCLLEFTTQIRKVAKASSQVVALD FRANFNELRRIVGEALKTQRTGGEPVQGEWSLLKETIL QAQRKVIPRRVKGRRAYKPPWLTKNIRECLLAKKEAY SQWKGGVITKEDYSAVARGCRGMVRKAKAEMEQGLA SRIKDNKKSFFKYIGSRKKAPGNVGPLQDKTGNLVVTP DDKANFNFFASVFLSKDQNCNLPAGPPLPGGGSPGP RVSEDLVRELLEGLDVFKSAGPDDLHPRVLKELAGVIAS PLARLYEHSWCSGEVPEKWKANVVPFIFKKGKREDPGN YRPVSLTSLKLFELFEKILAHIQEPAGNGMLRGNQHGFI SGRSCQTNLVAFYDQVTKSLDAGVAVDIVFLDFRKAFFD TVSHPIRLKLGDCGVDTHTVQWVANWLESRTQRVVV DGSFSTWVDVSGVPGQSVLGPALFNIFSDLDDEGVQST LFKFADDTKMWGDVGTIEGSNRLQLDLRLQGWAEEN RMGFNTDKCKVMHLGRNNRHHTYRLGNSPLVSTEVEK DLGVIIDSKMMSRQCGEAAKANCNTLSCHRCISRSK EVILPLYAALVRPQLEYCVQFWAPHFKRDVDSIERVQR RATRMKGQOQGRSYEERLRLNLSLQKRRLRGDLVAC YKLVRGDDQALGESLFPAPQGVTRNNGHKLAEGRFK LDIRRYFTVRAARVWNQLPSEVVQAPTLGVFKRRLLDD HLTGVI	cd016 50	288	539
CR1	CR1-17_G _{av}	.	Gavialis gangeticus	ATACAGG ACCTCCA CCTAACT CAGGAGG TGCACAG TCCCACC AGGGGAA ATGCCCT GCTGGAC CTGGTCC TTGCCAC AGCGGAT GACCTGG TGAGGGA ATTGCGG GTTCTCG GCCACCT GGGCGAT AGCGATC ATCGCCT	TGACCCCA GCATCCATT CCTGCCCAA GGCAGGGG GTCGGACTT GATGATCTG TTTAGGTCC CTTCCGACC CTAGCTACT ATGATACTA TGA	MGVQDQGSFLEKETILRAQRESVPMHTKGGKSAKPPW LSKGIQECLRAKKEAYIRWKQGAITKEDHTSVARGCRA AVRKAKAEMELGLVTRIKDNKKSFFKYIGSKKAAAPGNV GPLQDALGNLVIAANEKADLFNKFFASIFLSRDRDISPTG STEGLRGGTARPKVRDDVVRELLEGLDVFKSAGPDALH PRVLRAGVIAAGPLARLYEHSWSSGQVPDDWKKANV VPIYKKGKREDPGNYRPVSLTSVLKIFEKIHEHCEGP AGEIMLRGNQHGFIKGRSCQTNLSIFYDQVTKSLDEGVG VDVIFLDFKKAFDTVSHPIIKKLGCIDAYTVRWVAN WLDGRTQRVVDGFSFTRWDVSGVPGQSVLGPVLFNI FINDLDEGVESLTKFADDTKMWGEVGTPEGRDRIQLD LDRLQRWADENRMEFNADKCRVLHLGRRNQHTYRL GNTLLISTEAERDLGVIVDSRLNMSRQCEEA VSKANRTL SCHRCITSRREVILPLYAALVRPQLEYCVQFWAPHFKR DVASIERVQRRATRMIRGQOQGRPYEERLEGLNLSLHKKR RLRGDLVAVYKFTRGDQGGIGDALFPAPHGVTNRNG HKLIESRFLDIRRHYFMVRVARTWNQLPREVVLAAPT GVFKRRLLDRHLTGVI	cd016 50	225	476

[illegible]

GTTGAAG	LQSIPTYAMSIFLAPKGIIDDIQAMLSRTWW/SGKDKGGF
TCTGGCG	WTMPWKTLCCKPKGMGGLGFRDVRLENLALLGRQVW
TTAGGTG	RLNNTDLSLCFKVLSKYPFDG:NIFHAKKVDKASFTWSS
GGGATTT	IAKAVEAFQDGFQVGVNGEKINWADNWGMEGLNG
TCAGAGG	AAVISNTLNHSEKFKDLWHLDRRSWDENKVQAYGQ
GTTTCTTG	DWGDKICDIPIGDEVQEDKIWFHNTHGCTSKSAYSWL
GCTCTGT	LLKEMGCGPHRFWRALWKLDTLPKVRVFTWRVGHEI
GACGTTC	LPTNGKIASIRPGFNKGCPRCGAEVETLLHALRDCPTSR
GTTGCTTT	AVLSIGGWSRSFSISKSYDHCIDWLEDLMRVLDKRAMAD
GCA TGCG	LMITILWNCWNNRNNFIRGKEEEAQLIWERASNLSKEF
CAAGGAG	RICNMLHAPLISQNIAEKRWW/KPPKGFLKINFDATVGET
TTCTTGC	GIGYGTVIRDEEGFVLGGGGGFKEGRRSVEEAECVAFE
GACTTTA	QSIDVACKLNKIDHVLFETDHAGLVNRINNLANDEVTIIG
TGCGCTT	ARLKDCTVAFTFFKSASLI
TAGCGGT	
GGCGGGC	
CATTCGA	
GGTGGGA	
CGGGCGA	
TACATAC	
GTTGGCTC	
GTCACGA	
TCTTTTCT	
AGCATCG	
TCAGACC	
CTGTTTTG	
ATGGGAT	
TGTTTCAT	
TGGCAGC	
AATAAGG	
TTTTGICG	
TTTCTTTT	
GTTTGTGT	
CAGCTAT	
TTTATTG	
ATTTTCTT	
TACGTTC	
AGTAGGT	
TGCATGG	
GTTCTTGC	

ATGTATC
TTGCATG
GATTGT
TCGTCT
GCACGA
TTCCACT
GGTGGTA
TTTCTTAT
TTTCTTGG
ACTTTC
AACGTTA
TTCAATC
TGAGGGT
TTTTTTGG
TTTAACA
GGCATT
TATGAAT
GCGGAGG
ATAAGTT
TTAAATG
TTTTCCCT
ATTGTTC
GATTAG
TTGTTTT
GTTACAG
TAGCTAT
TAGGTTT
CTGGGT
TTCTTGG
ACTGCTG
GTTGTAG
GTTGCAG
TTTTTTGG
TACTTTCT
TGCATGT
TTCCCTT
CCTTGG
TTTCGGG
GATGGGG
GTTTGAT

650

TGAGTTC	ACCAAGAG	GIVASACDAAMPQSHPRRRRATYWWTEELNLRDSSV
ATGGTCG	GCCAAACCG	QARRRYRARRRRGGDVAGTKEGYLRAREALKAAlVEA
AGGCCAG	GCGGGGCA	KRAAWEQMVSSLDLDDPWGRPYKRVMGKIRPWAPPTE
GCGCCGA	ACCCGCCG	SLDRQALEGLLDALFPRAEGGPEIPTPLGEGDWDEGL
ATCAGAC	GGGACCGC	AVSPEELTAARRRLGAKGKAPGPDGIPGRAWALALGEG
CGGCCGA	GCCCTGGAT	NLSAATCRVLSGCLREGVFPPEWRRRAKL VLLPKGKAP
CCTSGG	TCCCTTCGT	GSASAYRPICLLDEAGKMLERIVADRLVQHLSSKGPNLH
ATAAACT	GGAAACAGG	DRQYGFRRSTLDAIQCFRDLTRA VVEEGGGVLLAVS
CTTCCCT	GCGCGGTA	LDITNAFNTLPWPEIGRALEHHGVVPLYLRRILSAYFGGR
GAGGGTC	CACAGGTG	DLAFPVEGGVQGRRAMERGVQGSVGLGPLLWDIAFDR
CGCCGGG	GCGAGGGG	VLRTPLPGCHVVCYADDTLV VAGGRAGGKPASERRR
CCCTGAA	CTCCCGGA	HWHRWWEFSETWVSRWRPKRPRPCSFAGVGGRRRRP
CGGGGCA	AGGAGACC	GSSPGGELGGSPPSGDGIGGGNHQRPQSGGAPKDP
ATGCTGA	CTGGGGCC	GRVLSRQSGGAAAGDPGPRRWRPRRRADDEISRPDAG
TAGAGGT	GGGGGGTT	QPMELRTPLQGTGPQGAQGRGTGCSQPDAPHRGPRLEGS
GCCTGGC	GGGAGGG	PPVRHGGALDRPVRGAHMGATAGGLPRRKARAAAAGPE
CTCAACG	AATACGGG	ARANQGDGVPYRVVHGGDHPGRRPPCGAPRRGEAPP
TCGGGGC	AGTAGTTCC	LLEGEAPRGEPVGGQGESEAPGRGPDGPGEVEGLAV
GTCCGCC	CCTGACTCC	RPPGARASDGRGCPPLPVVAGRRGRGLSFLAQVLTG
GACCGGC	CCTCCTCAC	HGCFGRYLRRIGKELTTRCHHCPALEDTAHHTLAACPA
TTCGGGA	GGTCCGAC	WAGERVLAGAVGCREGEJSLPRMVKAMCGSEEA WA
GGAGCTC	AGGTAGAC	AAVSFCEAVMSQKETAEWGRRRPHRARGAAPGGRGTRQ
CGCAAGC	ATCCCGCA	RARGGAGRGGLPDRPGAPSAARARHRSQGGSRGRR
TGGCGGC	GGGGTTTA	AAFPAPPPSDRTSRDRGERCPGWGWARSPTNPLPD
GGAAAAA	GTCGGTAA	
GGCCCCG	GAGTCCGA	
GTTATCG	CATAACCCC	
CATCCAG	CTCCCCCCC	
CGGCCGG	AAGGGGG	
CAAAGTT	GGGTATCC	
CGCAGCA	ATGATGGA	
CTGCGAA	TTTCCCCCTG	
TTGCTGG	CGTGACCA	
GCTGGAT	AAAAAAA	
GCGACGA		
CCGGGGC		
CGAAGAG		
GTGGCGG		
CCGCGGT		
GGCCCTC		

GCTGGGG
GATGCTC
GCCTGAG
GAAAGTT
CTCGGG
CGTGATG
AGCGTCT
CCCAGCG
GGGAGTC
GGCGCCA
TGGTGGT
CGGTGT
CCGCTGG
CGCGGCG
CGCTAAG
GTGCGCG
CCGCGGA
CGGGGT
CCAGTGG
GCTGGAC
CCGCGCC
GGGGTGA
CGCGCT
CGCGCT
CGCGCAA
CGCAGC
GGTGCAC
CAGCGCC
GTCGACC
GCAGCGG
TATGTGC
TACCGCT
CGCGTAA
CCCGGCG
CACCGCG

CCAAGGG
GTGCCAG
GAGGCG
TGCCTG
CCCGTG
TGCCTG
AGGCGG
CAGGCC
TCCGTC
ACAAGAT
CGGAGTA
CCTGCCT
GCAAAAC
CCCTCCC
GGAAGGA
GAAAGGG
GGGAAAG
AAGAAGG
AAGGAGG
AGAGAGG
CAGAAGG
GGAAGGA
GGACACC
GGCGGG
GAGGCC
CGAACCA
CCACCG
CGGGGA
GGACCG
ACCGAC
GAGGCT
GGGAGG
AGASCG
ACCACT
TCCTCC
CGCCGAG
GGACTGA
CGGGAG
CGGGGC
GGATCC

GCCCCCC
 CCTGGCA
 AGAACCC
 CAGAGGA
 AAAGGGG
 AGGAGGG
 GGACTCC
 TCCCCCG
 ACCTAAT
 AGACTGG
 GACTGGG
 AGCCGGA
 GAGCCCC
 GATCCAG
 GAGCCGG
 AGGAACC
 TCTGCCC
 CGCAGGA
 CAGAGGAG
 AGGGGAA
 GAAGAGG
 CCAACAG
 GGACGAG
 CGCGACC
 CGGAGGA
 GGASGCC
 CAGGCCG
 CAAAGGT
 GAGGAGG
 GTTGAGC
 CCTCGGG
 CCAACCC
 CCAACAG
 GTGTAGA
 GGAAGGG
 GAGTCGG
 TGGGAGC
 GGTGCCC
 CCGACTC
 CCTTGCA

Daphnia			Hydra vulgaris	GTGTATA TAAAAAC	TATTCAAG TTATTTGAT	MDSTYRNDFEKISNLFQNTFNLDNESDPDINYNFNEINGS KFECSYFYPPDKIKSIFKNNDRDYFNIIHINIRSLKKNFDN	cd016 50	570	829
---------	--	--	----------------	--------------------	-----------------------	---	-------------	-----	-----

14_H Ma	CTTCAAA TATTTTCA AAATAAC AAAAACA TACAAA	ATTTTGTG GACTGTGTG ATGAATATT ATTTAAACA ATATATGCT GAATACTTT ATATTATAA TATTGTTCA ATATTGAG AATGTTATA TATCCATAT TATAAAAT ATTATGTTA AAACATGCT TAACATCTG AAGAAAT TCTAACGCT TATGTAATT TTTAATATT TGTAATAATT TCTACCTTT TAATTCGTT ATATAGATT GTAAAGGG CTTCATGAT AAGATCAT GTGATCTTC TGGAAATC CTACCAGTA ATATATTCA ATGTAAATT TACCTTGTT CATATTTT CTGGAATA AAAAA	FLNMICQTENFFNVICLTETWCTEEDFKSNSNLHLTGFN TVFLERNVNRGGGVLFYKENLAYNIRSDMSVSDDDNIE ILTVEINKSKNILLSCCYPPPTGRTTEYLSNYLVHNIEK TSQEKKKNYILGDFNLDCFYNEKKYITTKFYNLHFETGT IPIINKPRTITKSSSLDNVLTDYFNITLKKGIKTDVSD HPIPFCLGADHKTNNGKITLKKRIYSNNNLNLFKEQL LHQDWSYFNEDINIIYKSFENIFYKIYEANFPIREIVLN AKDISCPWITKGIKKSSKIKQKLYIKLYKNKSDKHKNY TVYKNLFEKLKKAKQIYYSDLLIKHKHNSRRVWQIMK EITSKHKTNTNMPNAIKFENKLTDSNEIAAEFNFFESI GHNLSNKIPYVDNNSVDEFISSPNTINFSLTYNEFETAF KSLKRNKAIKPDNDINGNIVDSFNEIKDILFKVFKASITQG SFPNSLKIATVTPFEKTTGDHTNITNYPISVLPVFSKILERI MYNRIYTYLIDNKLLYKNQFGQRNSSTEHAILOVTRSI ADSPKNSQFTLGIHDLKAFDITVNHHLIKKLESYGIKD KTLNWFKSYLSNRKQFVYNKSSSQIKNITCGVPQGSIL GPLLLFYINDLYKASTLTITVNFADDTNLFQSHENIEILF NNMNELKRISYWFRKNKLSLNTDKTKFILPHPTSKKK RIQNPLDLFIDNTIHKREKFTKFLGVUIDENLSWSQQIDIS MKVAKNIGVLYKARYILNKKQLTQLYYFSFIHCHINYANI IWGSAHKTCLKSLYRHQKHAARIINFGNRIFSSSTPLLIEM KALNIYELNIFNILCFMFKCKENLSPSVFINLYNLKPKNK YELRDSKNIQEPFCKTKIDQFCISYRGPFLWNKIVLPNFD FSFKWSYSSFKKKTKKEIFFSTQNIIFY			
CRI- 53_H M	CGCGGC CGTTAAA GATGGCA GACGTGA CTTGTG	TCTCTATGA CTATTATGA AACTGTAA TAGCAATG GCTCATAAT	MVRKEDFEALVSRVLKLEAELQIKDKINDLKIVVNDM KNQSKLVNSNSDNTVSDVVKGRKKNTSVQDLNLNV VVKESHDRERQKNLVVFGFKTSTNKDESVAKDDKQE LQNLFPQINVKVITQRVFKLTNKTIDKPPPVIVVLTNKEE RNAVLKTAKNLRNSPETNNFINADMTEAERFKMKALR	cd016 50	757	100 7

				AAAAAT AAGAATA ACAAAAC ATAAACA TTTTGTA ATAATAA CTAAATA ATAAAAG ATAAACT TATCTTGT CAAAATT GACA	AGATTCAA CATGGCGTT GAATCTCA AGGAGCAT AAATAAAT AAATAAAT TATTT	EERAKLNEANQDLKYYGIRNERVVKLLKKLYESTQPY QQISPKKVLXNLLNNTKNPNDRHLLNSNKRLLVYYTNS TSLNKKNDFRALDIEKPDVFAVTETWYTPLSDVNIEG YVIYRNDRLNGKGGVCYVNSSIKSYEISNIELKSDLI EQIWTWIEYGQEKVLLGCIYRPDSSCVTNKNIINSIAA KRAQDKKQFTSIITGDFNYNTLKWKDSVEESHSSQAEAD FVACLNDCFEQCVTEATFQSNENELTNILDLVIVDDSN RIFGLSHLPPLGLSSKGHHILKWFYKMKDRIDDFSSSTK KFAYRRGNYEKINVIKSNWADFQNKCIDKCYDKFL YHYNQCKNYPIVKLNKASEVWMSNSLRKEIKYKNN YWTNKKMNWSIAGSKEKYNKMKRNILKQINNSVKNY ERNLVNQVKNPEKLLYSYIRKKQVKVKNHIALNDNIGS TMYDKHSIVDFNKYFHSVYIKEENINDCNFEVQTNRSF GKIFVDQALVKSYSKLDINKSSGYDEVNYPYVLKVCSES FSIPLSIYQQSLTNSDIPDLWKMANVSPIFKNGSKLQAIN YRPVSLTSPCKVLEKIISDNLKNYLESNNLFCKQHQHFL KGKNCITNLESDFLYQLMIGNSIDILYTDQKAFDRV PHKRLISKLCGYGIEKEMLNWIETYSNRKQQRVIGDTKS DWLEVLGVPQGSVLGPLLFLYINDLPSKIKNKCCELYA DDNKIIAVVNNQVDSKSLQDDINNLTWSDKWMINFI KKCKIMHFGHKHKNFEYFIKDSKLDSTNEKDGLLISN TMKWVNHQVNMVNVKANSRLLGSKSFTYKDKNTMKI LYCTFVRPILEVASPIWNPYNKNDIHKLEIVQQRATKLIP ELRHLPYEERLKKQLTSLKIRRLRYDLIQYYKIFHNLD KVCWYQQPKTAISISANGPASNIRGHKARVDIELIKRNR PRENFFTNRVAKWVWKNLPEEISAKSVNIFKDKLKHMM VIMKL	cd016 50	330	590
CR1	CR1- 7_H M	Hydra vulgaris		AAACGTT ATCATCT ATGGACT AGAAAGCA ACCCAAT GTTCTTCT CAGACGG AAAAACT AGAAAAT GATAAAT CGCAAAGT AAATAAT TTACTTA	AACATTAA ACCGATAC AACACGG ATCTACCGG AAGATTAA AAAAACCT TTCCGTTCA TCAAAACTT ACGAGCTTT GCCATTTC TATCGTGGC CCAATGCTA TGGAACCA	MMPIINKSTRITQHSCSAIDNIFTNLTFLHSTFKSGHIRSDIT DHFVFAIFKKPKDLQKKDWYFKKRIFTTRAKNMFKLE LSTVWNVNFLETCAANNSYDLFLDKFFNLVDKHFENKII RIKTGLLNPWFTKSLKSKRKQKLYIRFINSKKERDK HNYLEHKRLYERTKIARTTYFKNKLLELKGNCCKTW DAIKELGKQIHKNALPSVITVDGTQISDVKSVATKFNE FFVNGPNLASKIKHTPAHFLSYLSKSDKSLAFKKLTSKE LDDSIKKLNKNKATGYDAISASVIKDCYELIKDVLDDFN KSLKTGVFPDKMKVAIITPIFKSGDESLLVTNYPISVLV FSKILERIMYNRIFAFLYDNNLLYNNQYGFRRSNYSYTHA LMQLVEVISEASSHNKFTLGIFDLKSAFDTVNHQILLEK LKYYGISHEYHKWFTSYLTDRKQFISLNDGFTQNKLKIS CGVPGQSILGPLLLFLIYVNDLYKASNTMTSTIMFADDTTL			

GTGCTAT	TTTTACAAG	FYSHSNIISLFQV ANNELEQISNWLJANSLSLNADKTKFS		
CGAATGC	TCTGAGTGG	LFHSKRKFQSIKPSLPMLYINNINIERSFINKLLGILIDENL		
GATAGAG	TAAACTTAA	SWRDHINHISNKISRSIGILYKTRQILDKSTLTQLYFAFIN		
CAAAAGT	AGAAAGCA	SYLSYGNIVWGSTNKSCLDKLLKQQNHAARIHFQNRFT		
GAGATCT	ATACCTCTC	NAKPLLKSMRALDIYELNIFQNYTYVSTSVKKN AEGFR		
ATTCGCC	AAAAATT	KYFP		
GTCTTAA	AAAAAGAA			
ATCTAAA	CATAATCG			
GGAAC TA	AACTTATTG			
TCGATCC	CCAATCAC			
TTTGCTTA	GATTCCTTA			
TCGAGCT	CATGATTTT			
TGAAGAC	TTCTAACAA			
GTTAATA	TACACGATT			
ACAAAAA	ATTTCCAAT			
TATGGTT	TTTTATTAT			
CTTAAAG	CAAAGGTT			
CATCAAT	ACTGATGA			
TAAATTA	AAAGACTTT			
AGAAAAA	ATCTTCTTC			
TCAAAAA	AGGTTCTCC			
ATATGAA	GGTCTTGTA			
AAATGTT	ACTATTATA			
ATATAAA	AATAATGA			
TCGCGAT	AATGTACTT			
CTAACAG	GTATTTAAT			
CAAGTCA	ATTICTAAA			
AAGATTC	TATTTTACG			
ATTCTAA	GAAATTATC			
AAAAATA	TTTAATTGT			
ATTAATA	AAATGGTT			
GAACGAA	GCTTGACA			
ACGAAAA	AAAAATGA			
AAATGCG	AATAAAAA			
TTAAGAA	AAAAAA			
GTGAACA				
AGAAAAA				
ATTAAAT				
TTCAATTA				
CGGTATG				

ACAATCT
TACTCGA
ATTTTCA
CAATAG
CCTAG
CAACCAC
TAAITTA
GCAACGT
CATCAA
AGTATCT
CCCCCAA
AAAGAC
AAATATT
GACTATA
CTCGAAT
AAATCT
CAATTA
TTTCAAG
TCCACTT
CTCAACA
TAAACAA
ACTGAAC
ACTTCTT
ACTTGA
TTCCGGAT
AAAGCCG
CTAATTC
CTTAAAA
GATGTGA
ATTATAT
TTTCTAA
TCAATTC
TCAGTAT
TTCAITTA
AAATATC
GTAGTAT
TACTAAA
AACTTCG
ACGAACT
AAAGATC

CTACTAG
ATAATCT
TAACTTT
AACTTIA
GCTTAAT
TTGTTTA
ACGGAAG
GCTGGAT
TTTAGAT
ACAACAT
TAAATCA
TAAATCT
AAATTTA
TATTGCC
CAACTAC
AAAGCCA
TTAGTITT
GAACGTA
AGCTTIC
AAAAACA
GGCGGTG
GTATATG
TGTCTAT
GTGCTTG
ATGAATT
TACTTAT
CAAAATTA
GATATGA
CCTGTGC
TTCCTG
AGCCGGA
TTTTGAA
TCTTTCAC
TTTAGAA
ATCTACA
ATAAGTC
ATCTAAA
AATATTA
TTGTGAC
AACTACA

661

ATTTTATA	CTTTTTCGT	PIFKSGDETNISNYRPISVLPCFSKLIERIMYNRLYKYLSE
TTTATACA	TCATATTTT	NKILYNNQYGFKKNHSTHAXELVKHISNGFNSDCYTI
ATAGTTT	TGATGATAT	GVFIDLSKAFDVTNHEILLKKLENYGVKNQSLWFKSY
TTCAAAT	TAAATAATC	LTNRKQFILKESKSKNNLITCGVPQGSILGPLLFLLYIN
TTTAAATTT	AGCCCAAA	DLYLSSKVLNTILFADDTNLFYSHKDTVLPKIFNEELDK
AACTTTT	TACTTATGT	INWFISNRLSLNVEKTKFILPHKPSKAENIPKLPNLLIN
AGAA	TTACATATA	NKIIKRESTNFGVLLDEKLSWKFHIIKYEIGKISKNIAM
	TTGCTTTT	LYRTKPFLNNEKSLNLYFSFHSYLSYCNISWGSTNHAK
	TTTTTTTT	LKKIYSKQKHA CRIVFGANRKTQCEPLLRQLGALNIYKL
	AATTAAC	NIHQVMIFMFKAKLGFSPVTFQSYFNEISHKYPTKFSVN
	AACACATTT	NFVVPINFLKSSYQIQYRGPLVWKEFLPIFSKKQNFVLN
	TTTTTGTGT	CTLQSFKDDSKRHLLNMEFDILDFKYL
	GTTTTTATA	
	ACTTTCATG	
	ACTTTTGT	
	CGTTTTTAC	
	TTTTTAT	
	TTTAAAAAT	
	ATTCTGCTT	
	GAAGATTG	
	TTAGCAAC	
	AAATTATTT	
	TATTGCTGA	
	AGTTTTTGA	
	TTTGTTTTT	
	GTTTTTGT	
	TTTATACIG	
	TCTTTTTGT	
	CTTTTATCA	
	TTTATTGTG	
	CATGAAAT	
	TTTTTGTG	
	ACGGGGCT	
	CGGTGATA	
	AGACAACC	
	TGTCCTCTT	
	CTTGCTCCG	
	GCCAGGAT	
	TTTCTTTTA	
	TTGTAATTT	

CR1	CR1-42_H M	Hydra vulgaris				CTTTTGT AATATTTAT CAACGGCA AATTAAAT AAATAATT AAAAAAA A	MKNKTLCHFCKTVTDNHKAIYDCSSWVHAKCNQID LLTYNLLSEDSSEWYCFNCISKNNMFGLLTDNELLTSL CKNPQTASSNFLVSPHSRNILKELNTISGSAINCKFYDV VDENKAIKPNSDIYHLNLASLPYHIDDLCSLISTLNIPM AIGTESNLYLNDTNTDITINGFNIHCPTAKKGGALLY LQSNLYVVRHDLKIYLPKLESFVEIIPHKTNVLIGC IYRHPSMNQNEFTTFLPLLEKLSKEQKLFILMGDFNV DLLSYSESNTSYLDLCSYFNPSPILPTRITTKSQTLID NIFMNFYSSDHLSGNLTISISDHTAQFVCPGTSNQPKNV KISRLKRCCLKKFDPNKFIEISKINWELIJKDDNINESIN NFLKSNRVLDQCAPYKTLTKNQIKLSKPWISKGLLKS ISVKNKLYKKFVRCKNNTKNELFTKFKFYRNKISNLLK SSKSYTYTYFNNNNINNIKNTWKGIKDIINIKGSLHKTP QLNLNENIITDHA TVSNIFNYYFVSIREKLLNNIMPSEYT YSDFLNFPNANSFFIDPVSEDEVSLSISNTLKNDSYGP SIPTFLKL'TSHIISKPLSIMNNSFKNGKPPDVFKVAQVIP IYKAGSLLDIFS NYRPISSLNLSKLFKAMHNRLYKFLD KFKCLYQHQQYGRSKHSTLALJETTEKIRRALDNKHFA CGVFVDLQKAFDITVDHSILLTKLEYGIRGIPLKWFTSY LKNRTQFVSINSTKSDLQNCNGVPQGSVLPPLFLFIN DLNTSPKFATAYHFADDTNMLLVDKSLKKINKHINHDL GNLVQWLRNKLNSNKTLEIFKSNLAPIKQKNFRLS GQKIYPLDSIKYLGKIDSNLSFQGHNDLAMLKSRNG MLAKVRHYVNLTLNIYHAFQSHLRYACQVWQSH KQIILRLSHLQNKALKIYFYQYNNSSKNVLYLISKILKISD LVVFLNCLFVWNYHHKNLPLSFNNFFSITNCHYPLRSISS LNLVPKHQSAKYGNKSIKYQCISSWNKLPPDLKSLNSIS EFKIKLFNYLLKKYN	cd016 50	640	894
						TGCGCGA TGAAGGA AAGATGG CGGCTW AATTCTG AGTTGAA GTTTGGT ATACTTTT ATTTTAW CGAGTCG AAATTTA TATTTATC CATCTAT TATTA TCAATGA AAAAGAT AAAAAAT TAATAAA GTAATTA GTATCAA AATAATT TATTCGT GATCAAT ATTAATA CAAAGTA ACCCTAA AGTAGCA AACAAAA AAACGAG CCATATT TCATTAT CCATCAA ACAAGAA	ATATTCTA CCAACTAAT ATTGACATT CTTACTAT ATATGTCAA TTCGTAATC CTTCAGITT TTAATTGA TATTTTAA CATTTTGA ATTCTCTAC TTATCTATC TACATCTAA TTTCTCTGG TTTTCTATC TGCCATTTT TAAACTTTA TTATTAAC TATCTATAC TAATATTAT TAATTTTATG TACCAATG GGTTGGCAT TCITTTATT ATCTATCTA TTTATCTCT TTTCAITCA AAAATTCTC TATTTCTCA ACTATTCTA CTTCGTACT AAATCTTTG TTGTGTCCA			

664

CCTATAT
AACTATC
TATATAT
CTATATA
GCTTTAT
TAGTGAT
TATATAA
ATGTCAT
ATAACTT
AATCTTT
CAGTTTA
TGTTAAT
TATATAT
TTACATA
ACTATCA
TTATTTAC
ATAACTA
TCACTAT
CTACATA
ACTATCA
TTATTTAT
CATTGA
ACTTTAT
TATATTT
GTACATT
TCTCTCA
TTATTTCT
TTTCTTGG
TGCTTGTT
TTAACAT
TTTATTTG
TATATTT
ATTTTAT
TTATTTCT
TTCTTCAT
GATTATT
TTAACAT
TCTCCC
AAGTATA
TATTTTA

TTATTAT ATTGGCT TGTTAGT GTTTTGT TAGTTGT CAACATC TAJTAA TCICTTCA ACATCAT CTACTTG ATTTGT ATAACCT ACCTGTA CTGIGIT ATATATA CTGTCTT ATGCAAA TCCATTTT GCCTCAC AATTACT CTCTTAT ATATTAT TATATTA TTATAAG TTTATCTT ATTATTT ATACTCT TAGTCAC TCCTTTTT TTTTCATC ACCTTCT GTGTTTA TCAATAA TAAACA TAATAAC ATCTGC CTATCAA GT

CR1	CR1- 20_H M	.	Hydra vulgaris	AACATAC AAAATGG ATAACGA TAAAGTA ATTACAA TGGAAT GCTTCGA GAACTCA TGTCAT ACAAGGA GATACAA TTTTCAG ATGCTTT AAGGAAA GTTTCGGT CAGTGT CTGGAAA AAGTTGA GAACCTA GCCAAAG RCATTGA AGAAATA AAACGTG GGCTTAC TTTTGTTG GTGATCA AATTGAG ACAAAGT AAGCAAG GTTGATA GTAAAAAT GGCAGAA ATTAAAG AAACTAT TGTTGAT TAAATCA TTCCACG GAAAAAT AAGCATT GAAAAATA	CATATATTT TATAGTTTA CAGTATTGG TTTTCAGTA TTTTTTTAT TTTTTTTAT TCTTCTCT TCTTATTTT TCTTGTATT TAATATTTT TTGTTTTTG AATTGTGTA CACTAAAA AAATATGA GCTTTGTTT TTATTGAAT TTTTATTT ATTTTAAA AATTAAT GTTTTATA ATGAGTGA CTATAGGG GCTCTATGA AAAGGTTG TGATGTAA ACACATCAT CCTTACCTT CTTGAGTC CTTGACTGA TTAGAACA GAATATAC ATAATATA AAATAAAA AAATTATGA AAAAACGT TTTTTATCA TGTTTATCT TTACGACGT GTATTTTAT GTATTTTAT	MSNSLYNLVSYKMISQPRVNSKGGVAVYLDKFEFK KNNTFSKTNKHIEITLCEVLNTSGKSFLLSYRPPCGKQ VNFLNLKXFMSQMSKQNKHVFFAGDTNMDALSYEF ANIKNFNLLFEFNIPTIFKPTRITNQSSVIDNLTNCSLN TKFESGLFIADFSDFHPVFHIAREVLSNIDNKKCFVTYRN LSIKLKKLNNFNKQLQVEQWDKVYSSTDPNIAFDHF SSTFQNIQFNVCPKITTEIKNKEYKNPWMTTLLKASKRK QKLYVKFLSKNKEDEKIYKNYKKFFQHSMEKEKIRYF SNQLDKHKFDIKRTWAVINELTGREKKKHTCTCLPQKLI SDKKNYNKNMSRIFPIKTIITSTTKICQEFNKYFVNV GPSLASNVQPNKTFDNFLGEKPTTSINNEKITKLEFNKAI IELKRNKSCGYDGISSNVAYIMDSISKPLFYLISSSPENSI FPNKLKAKINPILKKGDCNNVSNYRPSLLPVLSKVFER IINRVNNYFTLNLLYKNQFGQSKSCSTEHAIKADKI FKSPDSDELVLGVFVDSLKAFDTVDHSILLSKLKYGIIG STFKWISQYLSNRKQFVLEKSGALDTCGVPQGSILGPL LFLIYINDMHKASQKISTIMYADDTNLFNHPBAKTLFET MNIELEKFNLWFIANKLSLNCOQKTSFTLFHKKNNQLIS LKKKQSNLPLKPLKLLINKNEIKREKCTKFLGVLFDENL SWRNHIGLLETKMSSVIGVLKSRPFLDSKSRNMLYFSL VHSQLYANIVWASTHKTKLLKLQSLQNACKVINYLN RLVNPAPVMKSMVLDIEKLNLTQILIFMYKYKNLLP SVFSDIFLISEKYNLRSNTRYNYHLGKIKSHQSNFLITN RGPSIWNHFNQKNKIDKLKSISSFKEKTRKMHLLNS	cd016 50	484	751
-----	-------------------	---	-------------------	---	--	---	-------------	-----	-----

GAAGACT
ATAGTGA
TGAAACA
ATGCGAA
TTGCTAA
AGCAATT
TTGTGAC
AAGCYAA
AATACAT
AGTGGTA
ACGAAA
ATACGCG
AAAGTCA
TTTACAG
TCGATTA
GTGGTAA
GAGAAAT
GATAAAC
GAGAACG
AAGGTGT
TGAAAGY
AAGTAAT
TTATTTAT
TTACGTTT
TAATAGT
TTTAATTA
GATATTA
AATTTTA
WGGTTTT
TAATTAA
GGATAAT
ATTAACA
ATGAATA
TRNTAA
AACCAAT
CATAAYA
ATGAAGC
AATTTTA
ACAAAAT
TAATATC

CR1	CR1- 3_H M	.	Hydra vulgaris	TCCAATT TTTGAAA TTTGTAA ATATAAC AAAAATG ATTAYAA TTTTTGA CTGAAAT AGATATT AAWTGCT TAAATAA TGAATAC TATAAAC CGGAAGA ATTTAAG AGAAAAA AAGCTTA ACAAAGA TTTAAAC ATTTAA GTATAAA TATACGG TCAATGA ACATTTT ATTATTCT TTAAATT ATCTTTT GACGTAA TATGTAT ATCAGAA ACTTGGG AAGATGC AAACTTG CCC	ATCTTATAC CTTTTACA ACTTATTAT TTATAACCC TAACGCAG TGCTTCTAT	MKICHCKFCNKKIKDNRHAIQCDCCSFWIHAKCIPVD KHAYNMLSDSSEWFCPCISNNMPFVLSDELKSTL SCKNPSNKIKFHEPPDYLNLFKTMNNVFNPTTKCYL DVNELKSNPGTEIYLHLNISSLPFHINDLSSFVGTINTPP MVIGITESNLYANDSNITDITIQGYNIEHCPTESKKGAL LYLNSNLNYIVRSDLQIYATKFLESIFVEVIYPLKSNITFG	cd016 50	634	890
-----	------------------	---	-------------------	--	--	---	-------------	-----	-----

TAAATAA	TCATCTTTT	CIYRHPSLNITEFLSIHFNPLEKLSHEAKNVLMDGFNID
AAGAAAA	AATTTTCT	LLNRYESQVISNYFESLCSHSLPTILPTRVTAKTTLID
AGGATAA	CATTTCTAT	NIFMNSFTDIVSGNLTSISDHMAQFVCIPIPNPIKKVK
GATAAAC	CGTGATATA	MFKRSFKFDSDFSQIEISDINWELLIKDDDDNINSEINFL
AGTTAAT	TACTAAT	KTFEKILDYAPYKELTKKQIKLQSKPWITSGILKSIN
TTTTCAAT	TAAATATCA	NILYKKFTRSKNVNTKNILFTKFKLYRNKISNLLRYSK
TTCAAT	ATTATATAA	LYYASFENNINNVKNTWKGIKEIINIKPSTSTKSFNLKI
CGCTAAA	TTTATTAT	NENVISDHTSVANIFNNYFTTIQDELLKNTPPSKCAFNF
AAGTTTT	ATAAACTG	LKIPSINSFFIEPVTENEVANLIKDTLKNKSLGPNLPTF
TTTAAAT	ATATGTTTC	LLKLVSIIISKPLCTMMNNSFKNGIFPAFKVAKVPIHK
TGTTAAA	TATAGTATT	MGSYLDYSNYRPISSLNLSKLFKAMFQRLEHFLEKHK
AATCTGA	CTCTTATCA	FYKHQYGFNRNKHSTTHSLIEITEKIRQAIDNKHFAAGVFI
TTTAAAT	CTGTTGTTG	DIRKAFDTVEHTILLEKLKHYGIRGIPFLWFSSYLCNRTQ
CAAAACGC	TTATTATTA	FVSINGINSGLAKEFNGVPQGSVLGPLLFIINDLNVSLK
TTAAAAA	CTATATTTT	FSTAYHFADDTNLLINKSLKKNMNMHDLANVVQW
TAAAAAT	ATATCGTAA	LRSNKLSLSKSKTEIIFKSAKTINKQLNFRLSGQKINPV
CAAAATAA	TGAATATTA	NSIKYLGKIDSNLSFASHLQDLALKLSRSNGILAKIRHFV
AAAAAAG	TTGTAATTG	NHETLLNLYHAIFHSHLRYACQVWGQSKQLAFLRLTYL
AACCTCG	TTATTATTG	QNKALKLIYFQHTNSNCILYFLSKVLKCLDLIQLSNCLF
AATTATT	TTGTTGTTA	VWNQNHNNLPLTFINFFSYRENCYILRSALNFKLSVPK
AAAAAAA	TTATTTTIG	YRTVHYGYESIQHKSQITWNNNLPSQLKSLKSFSKFKTAL
AAGAACAA	TTGTTATTA	FNHFLEKYSL
CTTTAAA	TTGCTATTA	
CAAAAGAA	TTATTGYTA	
AAGAACG	TTATTATTA	
AAGAAAA	TTATTATTA	
AAAAAAA	TTATTATTA	
AAAAAGA	TTATTATTA	
AAGAAAG	TTATTATTA	
AAAGAGA	TTATTATTA	
AAGTAAA	TAGTATATT	
CAAAGAA	CTTTTATA	
GAAGAAA	ATCGCTACT	
AAAAGAA	ATTATTTT	
AAAATAA	TTTATCACT	
GTAACAA	ATAATCACT	
AACAAGA	ACTATTACT	
AAGAGAA	TTTATTATT	
GAACAAA	CCTACTTGT	
ACCAATT	ATTATAATT	

TTAAATT
ATTTTACT
ATATATA
ATTTTAT
ATATTTT
TGATCAT
TATTATA
ACTTTAT
CATTAAT
ATATAAT
CTTAATC
ATTTATTT
ATTAATG
ATTTATA
TTATTTTA
ATTTTATC
TATGTTT
ATATTTT
ATCTTAG
TTATTA
TTAATAT
TATAAAT
CTATACT
CTTTATTT
ATAATTT
CGATAAT
AATTAAC
TACGTAC
TACTACT
GATTACT
GTGTCTA
TTTATCC
ATTTACT
GTTTATTT
GTTTGAT
TATTTTCT
CTTTTAAT
TTTCTATA
TTTCTGT
TCTATAA

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

				TGATAATT TGGGATA CTAATAT TATCGTT AAACCAT TCAGGCT TCATCGT GTTACCT CAAGAAA CGAAACA CCACAGC AGAGTGC AGTTAAT ACGTACT ACTCTGT ACAACCA ACTCAAC AAATATT TTCTACA TCTAAAG AACAAACC ATCCTTT GCACAGC AACATTG TACTCA ACTCGAA TACTAAA TTCTAATT CAAA	TGTTGTTTT AGCTGATTT TTAATTTAT TGTTIGACT ATTAAATA TTATTTTAC CTGTAATTT ATTACCGG GTGATAAA GAGAAAA ATAATATA ATAATAA	MIFLFYIIFLNFNIALSYFLPKMADETIVLTNLFNFSESK KSILLNNYSDDPINFYNNDDEVIKNINPLYYPNPNSKNIANG MDDNSFSLMHINIRSIQKNFESLKQFLYKIKINFQICLSE TWCHDKNIENNPFQLPNYKVVIHQVRASNKEGGGLCIFH KNSLLYKVKPNLSSTTDYESLCIEIHNETSKNIIHALYR PPSGSIKVFEENHIKKIKNKSSINKTVYVFGDNLNLIDYD KNKNIKFNFNVIQNSYPLINKPTRITKESATSIDIQTINE FINKKIKTGIFKTDISDHFPFIJSQCIYNAQNKRKKKIITTR LINDNSIKHFHNLLSCVNWDTLKLNLNABKAYDIFLTTF LKLYNKAFFPEITKFIKTKTLNPNWITKGILKSSKKQRLY NKFLKKKTCLKNETDYKNYKRLFESILKRSKKHYYSEQLI	cd016 50	592	694
--	--	--	--	--	--	---	-------------	-----	-----

AATGATC	GAAAAACT	KHKND	SKQTWQIIKEVIGKKNL	YRNLLPKNLKFN	NK	YI
GAATCGA	GAATATAT	VNKS	LVAEITLNQFFVNI	GPTLASKIAT	TQLNFKSYL	TSN
TGGATTG	GACAAATTA	NINVM	XNYKLTEKELLD	AVSLLKPNESL	GFDDISSN	VIK
CTAAAAG	CTTGCGGTG	SIKYIT	PLHLHFNL	SLKQGVFEK	LKIAKVIF	KSGDPSN
AAATTAC	TTCCCAAG	VANYR	PISILPCFSKIL	ERIMYNRL	YSFLNINN	ILYNKQFG
TACATTA	GATCAATTC	FKSGH	STDHAIHL	VHDIKAFDEK	KFTLG	VFIDL
AAAGAA	TAGGACCA	TV	DH	KILL		
WCTTGTG	CTTTTATTT					
AAWCTTT	CTAATTTAT					
AAAAAAA	ATAAACGA					
GAAAAACG	TTTAAATAA					
AAAACMG	ATTTTCTAA					
AAAWGCC	CATTTTAAA					
GATCAAA	CTCAATTTT					
ATAAAAA	ATTGCGWG					
CAAAACAA	ATGATACTA					
GCAGTTA	ACTTATTTT					
ATAAAGT	ATTCTAATA					
ATGAAAA	AAGATATT					
AATACAA	AATATTTTA					
AAAGACA	TTTAAACA					
TTGAAGA	GTAACAA					
AAGTCTA	AGAACTTAT					
ACGTTC	AAAAACTA					
ATCAAGA	ACCGAATG					
TAGCCAW	GTTTAAATC					
GATAAAA	AAATAAAT					
AAGTTAG	TGTCCTTAA					
CGAYTTA	ATATAAAT					
GAAAAAA	AAAACAAA					
AAAAGTC	ATATACTTT					
ATCGAAT	ATTTTCATCG					
GCAAATA	TCCTCATAA					
TAGGTGA	AAAAAGAA					
GAATGAA	ATATTCCAT					
AAAAATA	TAAAACTTC					
AAITTAG	CAAACTCTT					
ACARCTG	TTATTGATA					
GAAGATC	ACTCATTTAA					
GACAAAG	TAAAAAGA					

677

TATCATATT	TAAGAAA
CTTATTTT	AAAACCT
ATGTTAAA	CTTGAAG
CTTGATAAT	AAAGCAA
AAAATATC	AATGCAT
ACCAATTT	AGGAAAA
ATTIAATC	ATGGTAA
ATTTTAA	GTATTC
AAAAATAA	GTTGTTA
ATCATATAT	TATACGA
ACCCTACA	TAARCTT
AGATTTTCA	ATTGTTA
AAAAACAA	AAGAATT
TTATATTC	TATAAAA
ATCCAAAA	AAAAAAT
CACATTATT	AATGATA
CAGCAACT	A
AAATTCTCA	
ATTGCAAAT	
CGAGGTCCT	
AAATTATG	
GAATACAA	
TATTAAATA	
ATGAGCTTA	
AAACTAATT	
CTTCCTTAA	
ACCAATTTA	
AAACAAAA	
CTTAARCAA	
AAACTTTTG	
ATAAATGA	
AAATGAAT	
TAAATTCT	
TCTAAAAA	
GCTTTTAA	
ATTAGCAA	
ACAAAAAT	
AAAAATTA	
ATCTATTAA	
TTACTCTT	

AAATATTAAT
TTAATATTT
TAAACAAT
GICTTATIG
AAATCTATT
TATATCTTT
ATAAGTTTT
ATTACGAC
GATTTTTTA
AGTTTIAIT
ATTTTTTTT
KAATTTCTT
TTTATTAAT
TTCTTTTTT
ATTARTACT
TACTTAITTT
TACTAATCA
ATTAAATTT
ATAAATTTT
TCGAAATAC
WAAATAGT
TATATATTT
TATTTTTTA
TTTTGITAA
CGTTTTACT
TATGAATGC
ATTTTGGGG
CTTAGTGAT
AAGACGAA
TTTTGICTT
CTTCTAGCC
CCAGTCATG
TAAATAITTT
GTTTTTATA
AATTACATT
GTAAATATAT
TTTCAACGG
CAAATAAA
TAATAAAA
AAAAA

680

AACCAAG	ATTATTATA
AACGATT	TATGTATCA
GTCTTAA	GACTTATAT
AGTTATT	TATTATATA
AAATTTT	TCACGACTT
CAAGACA	ATATTATTT
AAACAAA	TGTATCACG
AATACTT	ACTAACATT
AACGCTA	GTAAACTA
CATTTAA	AAAAAAAAA
CTTGGCA	AAA
GGAACATG	
GGATATA	
TGTTTAT	
GAGGATT	
TCGCTAA	
AGAAACG	
ATGGAAA	
GCCGAAG	
GATGTIG	
TGGGAAG	
AGGTCAA	
AAAGTIG	
CGACTTG	
AAGGTAA	
GTAIGCA	
GTTATAA	
AGTATGA	
TAAAAAT	
GTTTCTA	
GAGAAAT	
TCGAAAG	
TAAATGC	
GCTCATT	
CTTAATA	
AGAACTC	
CTTTTAA	
TTTATTIG	
TGATAAT	
TAATC	

CR1	CR1-11_H M	.	Hydra vulgaris	GGCTAAA GGATTAT CAATTAC TCAAAAT AAAGAAAT TACTGGA AATTCAT GAAAACA CAATATT AAAAATM TTTARCG ATAAAAT TGATAAA ATGGAAA ACAAAAT GTAAGTA CTTAAAA GMGAAA ATAAAAT TTTAAAA AATGAAA TAGCAGA TATTAAA AGTGCAA TGACTTTT CATAAAT GAAAAGT ATGAAAA AGAGACT AAAGAAAT TATACGA TTTTAAA GAACATC TTAAACA CAACATG AAAGTAA ATGAAAA AGCTTTA AACAAATG ATGAATT	TTTTTTGTT TCTTTTTTT AACTTTATA TTTAACACT GCATTAATA AATTTTATA ATTIAACAA TTTATAAAC AGTGTGTTA ACACTTTTG TTAAAGGT ATCTTAAAA CTGTATAAC AAAAACCTT AACAAAAA GTTTAACAA TTGTGCTAA AGATATCTT AAAACGTG ATAACGCA TCTATGAT TTATAACGG ATTCTTTT TCTTTTTT CTTTAAAT TGTTTATGT TTTCAATTC TTAATTCTT AATWATT GTTATCTC ATGTATGTT AAGTATAT ATTCTTAT GAAAAGTT ATTTATTTT GAGAATAT TTTATTAAT TATAATAT ACGAAATA TTAACACG AAAAATTTT	MDSNEPVNFESTLNFQTNDFLLDSDSDPDLNYFSEAG ALQNKCSYFYNHEIKEFLDRDNFNVHINIRSLKKNFENF RNXIEETNFISHLCTETWCSDDDVNFNSNHLPGFNLIS LARKTNKRGGVJYVKEXLXFFTRLDMSISDGDKEVLT IELTKTKNIISCCYRPPSGVIESFNSFLRNDIIKKSNEHEK KENYLIGDLNLNCFEYHVNNIKRFYNDFENGAIPLINK PRTITQSTSLDNIITTDVFNESLKKGIKSDSDHPIFFSI NTNNEXVTKTHSLCLKTKTFNDKFLYDKKEIANELNKYF VSVGPNLAKKIPKAINPNNLFIPNTSQNSFELSFEFES AFKMLKPKDAVGPDDINGNIVINSYDIKDILFKIFSCSIK QGIHPPDQLKIAKVIPIKGGDLTNVSNYRPI SILSVFSKVL ERISYNKIYNHLSVNNLLYSNQYGFKKNNSTEHAILEQFT RSIIDSFENSQFTLGVFIDLAKAFDTIDHKILFKKLKY YGI TGNVLNWLKSYLNNRKQFVYVDV VSPVNLFDITCGVP QGSILGPLLLIYVNDLHKASNLMTMFADDTNLF LSHN NITTLFQNMNIELVKISDWFKLNKLSLNIDKTKWIFFHPN RKKHLLPSEMPPLLFIDNLIKRVTVTNFLGIYDENLSWK NHLENCSKISKSIGILYKARSVLNKH TLQLYYSFHCHI NYANIAWGSTNKSXLPL YRQOKHVARLINFKDRFTHA KPLLPEMKVFNVELNIFNVLCFMFKCKTNISPV SFHNL Y XKKEKNKYILRNDNFIRQPVQTFNFGKFCISFRGAF LWN KIVLKNFKDFSQECNYSRFSFKQLKKVIFSIDDISIYF	cd016 50	418	677
-----	------------	---	----------------	---	---	---	-------------	-----	-----

TAAAGAC	TAGCGGTTC
GAACITG	TCGACGAC
CKGAAC	AAGACCTT
TGAAGAC	ACGGTCTTC
AGAAGTC	TACGAGTW
GAAGAAA	CCCGCGTTC
TAATTTA	TTTTATTTC
AGATTTA	TTAATATCA
ATGGAAT	TTTATATTTA
TGAAGAA	TTATATCAT
AGGAAA	TGTTAACGA
ATGAAAC	ATATTTATT
GTGGGAA	AAATACAA
GAAAGCG	AGTTGTAAT
AAAAAAA	AAAGAAC
AATTAAA	AAAAAAA
GAANTCT	
TAAAACA	
AAAACTT	
GGTATTA	
CGAACGA	
AATATAA	
ATTGAAA	
GRGCTCA	
TAGAACA	
GGAAAAA	
AGAATGA	
AGTTGAA	
ATGAAAA	
AACGAAG	
AACTATT	
ATTGTAA	
AATTTT	
AAACTAT	
AAAGACA	
AAGAAAC	
CGTTTA	
RAAAAAT	
ACCGATT	
ACTCAAA	

684

				TATCCAG TAATCCA A	NVKGHHLLKCCYLMNDYKFSSTEHHKSIRLYQKGYE LFSCHFNNVDWENGFKSKSINEMYDLFLYHYNATNLF IPRINYQNRNRNRKWTTPELKHKIRTKNNIWLANKRNG WNEESTKIQYNKLIKRIKNDIKNNIKQFESTLASKFSNP KLLYKINEKKSVKTHIRALEIDNGEINDQFAIATELNK YFHSVYVNDNCTNIEFHNPISNLLNTVLITQNDVQTRL QALDKSVGFDFVHPYVLKECSSISYPLYLIFKKSMEI GTLDDMWKKNVMPFKKGSKLKASNYRPVSLTSPCK VLRIADCMQHLIKNNLLSKKHGFMKSKSCTTNLLE YLDILJDAFHNRTPVDVLYTDFKKAFFDSVSHKLLSKLL TFGISGKLLTWIVCHLANRKQRVVLGKSVTDWVDVLSG VPQSVLGPLLFLVFINDLPENFINECRLYADDNKIAPIS SQNDSQLFQNDINKLDEWSTKWKLGLNFECKIMHFGS NNEYSYFMNNNTLMEIEKSKLEKDIGVYISSDLKWA KQVYAAASKANSILSLNNTFKYKDKDLIKTLCTYTVR PNLEFSIQAWSPLYTKDINELEKVQRRATKMIPELRHLE YKKRLEILGLJTLETRRLRCDLIQQFKIFKGIEIDIKQNO TMNSICSSGPAANIRGAKHRMKPELVNCLKRQYFFSN RVVDGWNLSLPAEVVQSTTVKSFKLNLSQVDEAITNKI KTKKAYY			
CR1	CR1- 55_H M	Hydra vulgaris	TATAA	TACAAATTT GTTGTAAC TGTCAGCA CTTGATACA ATTTAACAT TTCAGTAAA AAAAAAA AATGTTT GTTAATGTA TACTCAGTG TTAGTGCC TTCGGGTCC TTCTGTTGT TTTCAAATA AATAAATA AATAAATA AATAAATA TAT	RWQPKFCIKYYNENFLPKLIKIFQLTIKIMVRKEDFEAL VSRVLKLEAELLTKDKEITQLKILVNDLKQSPNVLNSS ACAWSDVVKGKSKNAQQLDLNVVVKESQDRDLRQ KNLIVFGFNTSKNKNDVAREEDILGIGNFNQINAKATI QRFKLSIKSDKPPVIVLTNKEERNAVLTAKTLRNL PETNNIFINADLTVAERLKMALREERAKLNETNKDLQ YYYGIRNERVVKLIKROFISSQDQLLFSKNLLFKNSNK FIEKNNKRLNLKNQELFNHPKNNYCGKWALTNRSL GNKIADFQHFVHLHSLDICCVIETWLNKNLPDSLLCPRF YTVLRCDRSSKGGGIAIFRSFIQFTKVDIPPEFHELEIVC ADLNLNGQSRVIGYRSGGFGKV AIDYMQSSVKCFSL CSTKKNVILGDFNLDPIDWHYHGPDSVIHNSFLHFIN RYSLTQHVDQPTRENSILDILSSSSSISNLCIEPIGTSD HNVHIFPNLSCLYNKTEVPVQQFFCWKNADYQVINKYL KSVDSWIIHQTCFNIEACWHTFSTILNITQHXPVRC EEHKLSHKIHYPYVVKRLNNKAMAWKRSKVTKSND KAAYSIAAQCSDSIRKYHAAKELMKIRKNDNLGSGFNF VNGKLTCSKLNDIRSDNTLCHDKEEISCFNNFFASVF TVDNGLSPKFTDRVNVHSTCLSSINFSPSTVEKALRAK PTTSTGLDGIPNVFLKNCANSLPSLHSLFDTSFKDGAIPL	cd016 50	787	104 3

[illegible]

687

CGAAAAGA
AAACATTC
ACTAAGAA
ATATTGCT
AGAAAAGT
TTGATAGA
AAAGTCAA
AAGATTAT
TTTAAATGA
GTTACTTAT
GCTGAGGC
AACAAAT
AAACAAGA
AAAACACA
TAAACGAA
ATAAGAAA
CATTTTAA
GAATTGAA
AAATTAAAT
TGATAGCA
GAATGTCG
GACTATTTC
GTATAATT
TCGTTTTT
CTTTTCTT
TTTTCTTT
GTTTTTCG
AATTTTAT
TTTTAATT
TTATTATT
TTTATTAT
TTTATTAT
TTTAGTIT
CGGGTCCA
ATGGACAT
CTTTGTAAA
CGTTTTTT
TCGTTTTT
CGTTTTTT
TTTAAACT
TTGATTTAT

[illegible]

[illegible]

TTTTTTTT	LIGCIYRPNDASLIKYYAESFKLAHSLIKKPYDDCLIVGD
TTTTTTTT	ENFPNIKWSNDGIESFNPNPTFYDFWNIFQNSLFTQHIT
TTTCCGT	FPTQTDYGSNNMDDLFTYNSNLSNLSPSTILGNITK
ATCTTA	GHFSITWRFLVSELFTESESQSFNYKKGDYDSISNFIASID
TCITTCTT	WEFLHLKTAQDMENILISNYRTACTYFIPKINKTKKKY
TTTTACTT	DPWVTPHIKNLIREKKSAWHINCSTKWENSELKIKYNRL
ATTTTCT	NRSVKTEIFKAKKAHEKEIISKQNNPKIVYKYINKQLKI
AACTTTT	KDRIRSLKDTSGRIEALPDIVDLLNSQFKSMFTLESSENF
CTAACCA	PLLNKSGILSECIYNNELDFSISDIIQELADLKPYKSFQSD
CTACTTCT	GLHPLILKICAKSAFIPLTKIFHKSFNSGVVPQWKLANIS
TTTTCTAC	PIYKKKGDKLLAENYRPISLTSVACKIFEKLVKKKILKHI
TTATTGCT	NDNMMLSANQHGFLYHKSCVTNLIETFDIISSSIASHQFI
TTATCTA	DVIYLDFAKAFDSVPHNRLLLKLSFGINGKLLQWIKSF
ATTAGCT	LVGRKQRTINNTSSAWVDVLSGVPGSGVLGPLLILYT
TATTAAT	NDLAKLTETSEIFADDTKLFSKTLSSQSNFVDLQDRRLRV
TACTTTTC	YEWSTWLLKLTSKCCVIHYGRNNPHYNFYLQNNQG
TACTTAA	INQSLRSTKERDLGVFTSNLKWSAHIQTVVVSKSLQILG
AATTATA	MLKKTFTFIDEKMALKLYKVFIHPHLEYGVSIWAPFLKA
GTAATAT	DMALVESVQRKATKWAYSLRHLQYCERLKRNLPSLE
ATTAACCT	FRRLRCDLIQIFKLJHRFDIVNDKYLPKYNSVSRGHHSKI
ATTCGTG	RREYVLNCLPRYNHLYNRSKAYWNNLPEQITLYSTVLS
TCTATAA	FKNRDLSWLWKKFGTNIYKTGCL
TACAGTT	
TTTGTGTG	
TACTTCG	
CATACCT	
AAATTA	
ATATAAT	
TCTTATAT	
TTTATAC	
CTATACT	
TATATTTT	
ATTATAT	
ATAICTTT	
ATTTATA	
TATTACT	
ATTATAA	
AAATAGT	
TATAACT	
GAGATTT	

TTTTTTAT
GCTTCAG
TTTCGCTT
AAATTTTT
TTTTTTTT
TTTTTTTT
TTTTTTAC
AGAGTTA
TAAAAAG
TTTTTTTT
TTTTTTTT
TTTTTTTT
TTTTTTTT
TTTTTTTT
TTTTCCTT
TTCTTCGC
CAAAAAGT
TGGGTTT
GCACCTG
TTCTTTGT
TTTTAGTT
TTCTGAA
CAATAAG
CTAAATT
TTTGGCG
TCTGACA
CAAACTA
TATTAGT
AAATTAT
ATATCCA
TATATTA
TATATAA
AATACAT
ATAAAAA
AGTTATA
TATAAAC
ATATACA
AATAATA
TAAAGAA
ATTAAAT
ATATACA
AATAAAT
ATGATAA

[illegible]

				CTTTAGAAC ATGTATTTA TACTTTGTT CTAATAAAT ATCAATCA ATCAATCAT CAT	YCDMDLSQDLVVALNRLQLQWAETWQLTIACSKCFALRI APPSLCKGIKVVNYKLGESLAWSSNPKDLGVTMDCNLN YKKHFNVIHQANSRAYILKCFKTRDPCILVRAFTTYIR PILEYCSVPWSPHQIGLJKSVESVQQKFTKKIRSLRNL HNRLQLLGIESLEVRRLKNDLVTCFKIFRKLVCIDQSMF FAIDNNSHTRGHNYKISKQSCRDLIKKI			
				TGGTAAA GATGGCG GAGTTAT AGATAGA ATTGTAA ATTTATC AAATAAA GTAATAT ATCTTAA CATGACA GCAATTA AGAGGAA AAATATT TATAGAA AAAAAGA CTAAAAA GTAAAAA AAAAATAT ATCGGAA AATTAAAT TAGAAAA CCAAAGT ACTTAAA GTTTAGA GATGTGT AAGAACT TCATAGA CTCAAAAT AAATATA ATTAAAG AACAAAGA TACATTA CAAAATT	MLIRNLCKFCNKTVDVSHKGILCNYSWHLHPKCNYN KHTYSLIEKSDTDWYCSDCINKNLPFCSLDFELSLTSS QTIPTNNLKSLSPPHFKLFNNKINSKSNCKYYDILE LNKVLNNCNLYHLNIASLPYHIDGLRSLSSLTTLPIVI GISERLNINDSSKTDVNNINGFTIEHSPTEAKKGALLYL RSNLNYFIRSDLVYSPKYLESIFVEIVNPLKQNIIVGSIYR HPSMNSNEFTDHFENPLERLSLENKQVLMGDFNMMDL LNYNESKVISSYLDLTLCSYFFPSIITRITATSKTLIDNI FLNFQTPDLJSGNLTISISDHMAQFICPLSPKKTKVEIFR RTFQNFQDKFKDISADHWQIQJTNNDNVNESLHYFFK TFNKILDQHAPYKKTAKQIKLSKPWTNGILKSISIKN KLYQYIKAKNHNKIHFNKFWYRNNICNLLKLSKKS YYINFFNNLNKXNKNVWKGIKEINIIPKSSFSKSNLKFN GNTISDSKAVSNIFNNYFVNTQQTLLTKIVSPKCNFSDYL NNPINSFFIDPVTENEVSNLKLSQSKSLGINSIPTFLK LVSHIISKPLCTMINSFENGLFPDVFKIAKVPLYKKGSR TDHTNYRPISSLNSKIFEKAMHNRLYNFLDKYQCLYK HQYGFRRSKHSTTHALJEITEVIRKALDDKLFACGVFVDL QKAFDTVDHSILLSKLEYGIRGISLQWFTSYLYNRFQF VSINGNTSSLVKCTTGVPGQSVLGPLLFLVFENDLNL YTTSYHFADDTNLLINKSLKINKYINHDLNLIQWLR SNKLSLNTSKTIIIFKSKQSKSKHLNRLSGQKINPVMY LKYLGIKIDSNLTFSQHLKDLKSLCRSNGMLAKIRHYV NLETLINIYHAIFGSHLRYACQVWQJRTQEILRLSNLQN KALKIHYFQHYQNSDHLFYFLSKILKIDDLVQLLNCHVF NYKQKCLPSTFTFTFTVRESCRYHLRSANSYNLHVPNHS SVKYGLKSIKYQSVKSWNSLPSPLKSIDSTITNFKNCLFNY FLARYYT	cd016 50	635	891
Dap hne	Daph ne- 78_H Ma		Hydra vulgaris					

696

ATGTTAT
TTACCTA
ATATTAC
TACTTAT
CTCCTTCT
TTACTTAT
TATTACIT
ACCTTAT
TATCATC
AAATTAT
TAAATAA
ACATAA
TTATTTAT
ATTATAC
TATGTCT
AAAGCTA
CTAGAGG
TATTAA
TATTAA
ATATATA
AAATTAT
ATTCTAA
ATCAAGT
TAATAAT
ATTCTAA
TTTTTATCT
AGCTGTA
ATACTTA
TTCTCCTG
TGTAAAG
TATTTAA
ATCTCTT
ATATCTT
ACTATCA
TTTCTTTA
TAACTGA
TCATTCTT
CAATAT
TAAATA
GCAACCA
TTTCTAG

[illegible]

					AGACAGCG ACAGAAAC CCAGAGAT ACTGGGAA GCCAGAAAG TCAACTAA AGGCTAA TGATAACTA TTATAAAGC CTGCCATT GATATAAT AATAATAA	AGVYSELFTHFAAAGQNRSAFEGEIEAIAIALGQLCCRD TKFTNAVILSDSQAISQISGNREPPKTAIEIHECKKLYQLL KEKNKTIVLQWIPGHCHIGNERADTLAKKGTTLQANIF LHNEDLNQERIQDSKKQRT					
					CATTGG CCGGCAG GGTGCGT TTTTCGCC CTTTTTC GGCATT TTAAAGA ACTCTCG GACACGG ACACCTC GGTCGAT GTAAACG CATAATC AGGTACC	MECTLYYQNVRLRTRKSSEFLSGVIGNSYTIISLTETWLE DTIPSSQYFPCCYSVFRKDRDYAATRQRYGGGVLTAVD TSLCARRRYDLELYSECVWIEVTCRDGLNFIIGNYYFPP HLSSDVFAHFRLDLEGLDVTKFRVHIYGFENLPIDW LANYNLITSDVTSLKASSLFDIFYENGFEQRNTIQNSAGN VLDLVYSTPINYISPTIDPLTRVDHFHPPFAVSFCFPLRK QALCSYEYLLYSSGDYLNMYKYFCYDWWKPLNNKCA DSAAEALTEVVSSAINDFVPRVSRNLNKYPAPWFSKELRS CLRRKLHYHRLFRKRGSLDKWYAKFSECRALAKWLFKK DESAHLHGLESDMYAKQNFWKYVPTKKKRNSSVAPM RNSSGYTCDSEIANMFAEHFKSCFSSASVPDNNATCHG TSDFLSFIPIDEPEIREISIRNMKSLKSGSDGIPSFIVKGA ELLTPVLLHIFNLHSIGIFPSIWKRSVVVPPIHKAGDPSTV NNYRPVSLLCSFSKVFEMIMHKRLYRHFHMKMNPEQHG FLKGSVETNLCFLDYAVPLVCNREQVDVIYFDMTKA FDKVNHSSTLLTKLDSYGLCLTYCQWFKSYLTCRTNYVR VSGMSDIFYSSPSGVPQGSNLGPLLFLMFVNDISLCAKN SKVLLYADDIKLFRRVQSHSDCVLLQEDALQISQWCAK NYLPLNEQKTRTMSLTKRSDSLAFSYTIGGSVITRVSVIR DLGIFDSSLNFNHHVLTICNASFRNLGAISRFRTRKFSSPV CLVRLFCCLVRTKLEYASSVWNCINLTASCCIERVQKRF IGIYDRYFDRTCYYSYEQLLKKGLCTLLDRRRLDLM FLHKVVNGAVDSPLLSTVCFHAPLNRNRPQSPFYSAL NVASPLTRMQLEFNKLDQSAVDIFSDLVFRHRESIICT RRGIIFISYAAPLGGSSADTLRTTALGLVYSTAEYCAPV WMNSAHTRKIDVQLNGTMRIVSGTIRSTPTSWLPILSHIP PADLRNSALLREYRKITNNPRLPIHHDTPDLERNRLRS RHPTMKTAKELHNTDFNLIESWKERWRQACPEQCOAP	cd016 50	493	745		
CR1	CR1- 1_IS	.		Ixodes scapulari s					NA	NA	NA
Ving i	Vingi - 35_L Mi	.		Locusta migratori a							

GGAGATC	TCATTATATA	CITERPPGYELPRKAWSSLNRIRTNHGRCADSLYRWGKI	
GTTCTGA	CCTTAATTTT	PSPNCDCGADRQTIRHIEDCEMRAYDGNVSDFLTAKDA	
TTGTCTG	ATGTCTCTT	AIDYIRNLDVCI	
GCTGCCA	TTGTATCTT		
GTTAGCG	TGTATGTTT		
CTAGTTC	TACACATTT		
AJTTAGT	TTGTATTTT		
TCACGTT	TTGTATGTA		
TTACAAC	TAGTTGTAT		
GCTATCC	AAGACCTG		
GGGGTGC	CTAAATGAT		
ACCACGG	GTGCACTGC		
GGAGGCG	CATACGAT		
AACTAGC	AAATAAAAT		
ATTACAC	AAATAATA		
CCTTACA			
AACGCC			
GACTTAT			
AGATAGC			
ACGCTTG			
CCGCAGA			
ACATTCC			
AGGTGAC			
AGTGGGG			
GACCAAA			
TATGCCC			
AAAGAAC			
CCTAAAC			
AACGGAT			
TACCACA			
GGGCTCC			
GTCCTAG			
CCCCATT			
GCTATTC			
AGCATGT			
ACATCGC			
TGACATG			
CCCCAAA			
CAACTCC			
AGAAAAAT			

TTGAATA
TGCAGAC
GACTGGG
CTTAGCG
TCAACAT
AAGAGCT
TTGAGAC
AATGGAG
TGACCTC
GTCGTCC
CAACTGG
CGACTGG
CGCATAT
TTCGGGG
GCCGAGC
CTACAAA
ACGGGAC
CTGCTTC
CATATCT
GTAATAA
ATCAGCA
GGCAGAC
GACGACG
TCTACTT
GAGGGAA
CTGCTC
CGGCATA
ACGGTAT
CCAAAGT
ACCTCGG
AGTCACA
TTAGACA
GAACCTC
CACCTAT
AAGCGAC
ATCTGGA
GAACACT
GCAGCCA
CAACTAG

702

[illegible]

CTTAATC	CATCTACTT	RSGLPIAVRVRSLLSSLES LPIPPPTTHYPSRAPPWSVPVRS
GCCCGG	CGACGGGA	FGIPSYSPHFSLPFCSSLVTRPQLRLDLSRDSKASVAPL
TCCGAG	GGATCCTCC	NLRRQFQAVLGSFHGAIEIVTDGSLVDGRCGYAYAH
GACGGTA	ACAGTGTC	EHIEQHSLPCGSSIFTAELIAIRRALEHLRACPGESYVICS
CAAACCT	GCTGTGGC	DSLGLQALDQCQPKHPLVRSIQESVTFLSLAALWSSFG
CTATAAA	ACTTTATTG	PRDTLEFRGTNAWIGWLSRQLLQAP
AAAAAAA	ACAGTGGT	
CCTCAAT	CCACATTCT	
CTCAAGG	GTTGGACTG	
TGTGCTA	TCCCTTTT	
CTGCCGA	AGCTGCCTT	
GGAGATG	ACGACAGT	
CATGGCT	CTTTTGACC	
GTTGAGG	TGCCAGCTG	
TTGAACA	CGCTTCCTT	
GTCGCTA	TTATGCTTG	
GCGAGCA	GTGACGAC	
ACCTCTG	GCTGCTACA	
AGGAACC	GCTGATCGT	
TGCCGCA	GTTTACGT	
CCTCAGT	TTTATCCGT	
TGTATAA	GCAGCTGG	
GGCTTAC	GTTTATCA	
CCAGGCA	TTTGATCTA	
AGCGGGG	AAGGTGGG	
CTCTGCC	TACCTCAGC	
TGGGTGG	CTGATCTCT	
ACTTAGT	CAGGCACC	
TCCCTAG	CATCTCTCA	
CTGCTCG	CAGATTITT	
TGGGACG	ACTCTTCTT	
AAATGAA	ATTTTGTGT	
ACGAGCC	TGTTCTGACC	
AAATAGC	TGCTTTTAC	
GTGGTGG	TCGTATGTT	
ACGTAAA	TTAGAAATG	
CCACCGT	GCAGAGTG	
CTCCAAG	GCTGGGCTCA	
CGTGGAA	TCCTTTATT	
GAGTTGC	ATCCTCGTG	

TTCGGT CAGTATA ACACCCC ACAAAATC GCTAAAC TGATCCA GGGTCG ATCTATC ACAGGGA CCTTCCTT TACAGTC GATGAGA GTCTACG GCCAACC TAGAGCG ACGGGGT GTCCACT TCGTCCG CCGCGTG CATAGG GTCCGAA GGATAAT AGGTAG CCACCG TGCTTC ACTTGG CCTTGA GGCGATA CATGCC CGAAAAG GTCAAGG TCATGAT CTACCG TGTGACG TAAAGCC TTATATC CTCCCC CTATGC GTGTTG AAATGCT

GGAAATT
TGGACAT
ATGTCTT
CCAGATG
TAGTCTC
AGCCCCA
CATGTAG
GGAATGC
GGGCGTC
CATACAA
CCCGGAC
GCTCCGT
GTACTCC
TCCACCT
ATCTGTG
TAAAC TG
TGGGCAG
CAACATC
CAACCTG
TTCCTCA
GACTGTA
CCGTGCT
CCGCAAA
GAGCGAA
AGATTAT
GGAATCA
AGACCTT
GGATCGA
CTGACAT
ATACGGA
GGCAGCT
AAAAAAT
ATGAACG
GCTTCAC
CCAGTTT
CCATGAC
TTCGTCT
CAGTTAC
ATCTCTG

CCACCCT
TCTGATA
ACACCTA
CCGTGCA
ACACACT
TCTGTGG
GCCAAA
GGGCCGT
CAAAGCT
CATCTGC
TCTCTCA
AGGTCGG
GGGCACT
CCTCTCA
CTGCCCC
CCACTAC
CACCTGG
GGGGCAA
CTCCACC
CCCACCA
TCGGGGA
CACCAGT
CCCCACC
TCCCAGC
CGGAGAA
GGAGACAG
CCTCTCT
CGGCTCC
CCTCGCT
CGGAAGG
GGTCCCT
CGGGGCA
CTCCCTC
TGAGGCT
TCTCTA
GTGCTCC
AGCAGC
ACGCCAA
TGGCTGA
AGGATCC

[illegible]

CAACTGA	GTGTAAGA	VINTTRKAIMRISSEAIGNDGISITMIKNVADSLLPVLTID
ACAGCAC	GAGGCCT	IYLSLFSANQHIACVSKALEYIVHDQITEHLCKFSLFDK
ATGTCCT	GAAAGCCC	YQSGFRKNHSTNTALIKVTDDLKFAMNNREATILTLDDF
CITCTCA	TAATCTTGC	SKAFDVTNFDILLNKMRLKFSNSALKWFESYLKNRQQ
ATCGGAA	CAGGTTAA	SVVCGNEKSSWRNVVSGVPQGSVLGPLLFSLYVNDISS
TTACAGT	ATAAATGA	VLTFCKYHFYADDLQLYQSTRPEDINTAIAQMNEDLSSV
TACTCAC	GTAAATAA	VTWAKNLGLKLNAKKSQVILVAHQKLISSELRLDQLPPIQ
TATCAAT	ATAAA	LDGIPIYQKTVKNLGVIIDEHLNWADNTVTVCRKTSAC
ATGACAC		LYALKKFRNIFPDQVKCKLVQSLVLPNLHYCDMIQHGT
GTACATC		SSENTRRLELTMNACVRYICNIRRYDHISASYTQLGWLR
TGATGGG		PDKLRDYYHTLCLLHRLLEQHHSTSPQNLTIFRAIHLTQD
AGACCTT		PTYLVPLCPFIKRRKHLQTPSLSLLFASGTSCPALFARFSL
AACATGA		LRLKSN'TSLCHPLKTLAKTHVYSQVSLCVIVKPMFLSP
CITGTTA		LFLLFLSPNQCSSLRYSSTSSFFHHNLLFRYISLCEYY
AGAGACA		THFSPFPPLPPCPLLLAAIHKTPYCN'TCHVSYLSNTNIN
CTCCATC		CVYVLLLLVLIENFNCRFSTITSVILEM
CACAATA		
AACTAAG		
AGATTGC		
TCCAACA		
ACATGAA		
CATTCTG		
CCATTAG		
AACCTAC		
TCACCAT		
ACGGAGC		
ACACTCT		
TATAGAT		
GTGATCG		
CAACGAA		
AAAGATC		
GACAAAG		
TACAGAT		
CTAAGTC		
AAACGTC		
AGCCCGC		
CTCTCAG		
CACACGA		
TGCAGCG		
CCAAAAG		

711

TKQKVAARNNVIRKLTGTTWGAHPNTVRTSSLALSYSA GEYASPVWFKSSHAKIVDVALNDTCRIITGCLKPTPLNK LYSLAGIAPPDIRREVAAKSERGKSETSEAHPLYGHQPA NQRLKSRKSFRLRSTEALVGTSSQOCRTELWRARYTGTAG WFTPEKLPPTYTEKWSTWRSNRLRSGVTKCKSNLVK WGFLVESPLCECGATQTTSHLLQCTQCPATCTPEDLLHA TPNAIQVANFWASKV	AGTGCTTCT GACACGAA TAAATAAA TAAATAA	TCCAAGC AACAGCT GCTTCAA GAACTGT GCCATGA ACAGAAAG TTTGATG TACTGTG CATACAA GAAACTC ACAGAGA TATCCAT CAAATGC GACCGAC AATACCT GGTTTGC AGCTAGC TGCAGAA GTAACCC ATGCTAA ACGTGGA GGTGCTG TTTTTGTG AAACCAG GAATACG AGTCTTG AGTGCAT ATAACTA AATGACA ATAACAT CGAGATC ATCACCT TGGAACCT TAGTAAC AGTGTGA TCACCTC GGTGATA AGCGGTC CTCCTTCC GAAAATT					
--	--	--	--	--	--	--	--

CTCCTTG
CTCCTCT
AAAAAAT
TGATTTCA
GGAAAAA
TATCTTG
GTAATAT
GGGATAT
GAAAAATG
TGAACCC
GTTATGT
CTCGAGC
TGAAGAC
AACACCC
WCACCTT
CACCTAC
GACAGCA
AAGTACC
TCATCCTT
CAAAAGTG
AAGGTGG
CAACGTG
ACCTCTA
TTTGTA
TTGAAAC
TAGCAA
CAGTGT
ACAATCA
ATGTGTA
ATTTCAT
ACCCAC
ACCAAC
ACGGGT
CCAATAA
TGTTGCG
AATCTT
CCCATCA
TAAGACC
CAAGAA
ACTGGTT

TTAGACG
AAGGGAT
AGTTCA
AAAAAGC
AAGATTG
CCGAAAT
TCTCAA
AATGCTG
GATGAAA
AAGTTTA
TACACTT
AGCTGT
GCCTGAA
AGTATGA
ACACTTT
TTGAGCT
CTTAAT
CCCTGCG
ACGATTA
TCCCAAG
AGCTGCA
AACA AAC
TACCTGA
GGGCATT
ACTCAG
AAACAGC
TACCTTG
CTCGATA
TTATACC
CAATATG
ATAATCC
CATTACG
TGATGAA
ACCCAAA
ACAGCCA
GATCCGT
CCTCTCT
CGAGGCA
AAGAGAG

[illegible]

				CAGCAGT AGAGTAA GGTTATG TGCTTGT CGTGTG AAGTGT CATATTG TTCCACG TTTGCTA AGGTGT GCACCAT ATTAATT CTCTGC CGCCACG TATCAAG CAACGTG CTGCTTA GCTCTTG TTCTGTA AAGGATC CACTTA TCGTCTG CAATTAG ATTTGTT TCCAAGG AGTTATT TCGGCTT CATTA	TTGCACATT GCTTCCTTG AGTGCAGTT GTGCCTGA GAATGGCA GGGTGTCCT CTGCCGA AATATCGG CGGTGTCTG ACGACGTC ACCCGGCA GCATACCC GTAAGTTGT TTGAACATT CATTAC	CVFRPPSKIKALLGSVKDDDLGLRKAGVYRIPCSCGMSYI GQTRTVEDRCIEHKRYTRLQQPNKSAVAEHCLGTGHP MEYNNTEILACTSTYWDSVIKEAVEIKLASDLINRDGGF CLNSAWNPAISLVKKQRDKINAASPVD			
Vingi - 26_L Mi	Vingi - 26_L Mi	Locusta migratori a	ACGAGTA TACTAA AAGGATT GCAAAAGA ATACTCA CGGCCAT CCCAGTA GAAAAGC CGGCTTT CGTSCCA GCCGAAG CGTGCAG	TAGATGTTT GTTGTGAAT ACATGTAAT GTAATATAT GTAATGAA TGTAACATA ATGTAATCA GACACTGA AATGTACTT AAATGCCA TACGCTAA ATAAATAA	TYENNSLSKSDKLIGNMIGNRGFQVIMGNISSQKFLQNG LPQGSVLAPLLFSLYIADMPETACTKFGYADDWXIATR HTNFDTMZETLTSLSALAA YFRKWRLRPNAKTKEATC FHLNNRGADRTLEVYFDDTRLNYNKEPSIQHTSVNGGL GLMPRRQKRLASISITEERTELWKYTLTTTHASTTTRSQV SRVTLDRTLSEKTHVTKTAAKVRTRNNLIHKLCGTTWG STATTLRTSALSLSVSAEYCAPVWLNSSHVKKIDVEL NNTMRIISGTIRSTPNWL PVLSHILPPKMRRRETALLREY NKIEANPELPHISDIPDLRIKRLSRHPPLSLDSLTGADIN SWRREW/LQNAQPLATHQLLEKRMAPKRPTTCHLMPCIT EAVPGEDQPRKIWSTLNRRLRTGHGRCADSLFKWGLLSS	SSF5 6672	34	112	

[illegible]

TAATTTTT
GGTCAAGTT
TAATTTAAC
CAGTAAA
AATTTAACA
TATTTTAC
AACGCTCC
ATGTTTCTA
GACATTAG
ACCTGTATA
AGATAAGA
TACGTTTIG
ATCAATAATT
TTTCAGCCA
AATTTCTATT
GTTTCACTT
TATATTCT
TAGGATA
CGTCATAAT
CTATATAT
TGTCACGTA
TTTTATCAA
TGTGAGGA
ATTTACACC
ATAACCTG
AAATGTAA
AATTTCCGC
AGGTTAGTT
ATAAACCA
AAACACTA
TGTCCTTT
GAAAGTTA
TATATCTAG
CAAGTTAA
GATCTTGA
AACTTAAA
CTTTGAATG
TTTACCTG
TGATGCGGT
AACAGCCG

[illegible]

CCCTCGG	CCATAGTA	PRRLDLSREPKTSTPLVLRHQFYALLREFRNSEVIYTD	
GGTTTGA	GTCTAGTCC	GSSDGDHTGYAYMHGAYVEQRSLPGSSVFTAEIVAIY	
CCTCCAT	CCCTAAA	RALKHIRTCPGTSFLICSDLSGLQALDQCYPRHTLVRAI	
CAAAATT	CCCCCATA	QESILALDDSGRSVTFWTPGHVGIPEGNESVDQLAKQAT	
TCAGAGT	CCAAACCA	TCRPLAIGLPEADFRPLLRNKIRELWNAEWSALDVPNKL	
AAGTGTG	CCAAACCA	RTIKD'TTSVWKSSQLGRRRDSVALCRLRIGHTRLTHGHL	
AGCCATT	C	LRREDPPVCQCGAPLTVVHLLHCPILAPLRHSFNLPNT	
TGGGGAA		LSLVLGDDASVANMVLRFRLRAGGIYQLL	
GAAC'TCC			
CAACCTA			
GCCCCTA			
CGGCGCG			
GATTCTT			
ACCCITT			
CCTTCCA			
AATCCCT			
CTTCCCT			
GCCGTAG			
CTCCCTC			
CCCGGCT			
AGGTCAC			
CAGCACG			
GTAGCCA			
GTCCGTG			
TGGTGGG			
GCTGTTA			
TGTACCC			
TCATGGT			
TGAGCCC			
CCTGAAT			
TGCACAG			
GGATCAC			
ACTGCCG			
ATGCCCTG			
AGCTGTG			
ACCTCCT			
CATGAAT			
GCCAAAGG			
AGTCGAT			
GCTTGTC			

TCTGCGG
GGCATCA
GGACTCC
CAGCAAC
GGTCATC
TCTGCCAG
GTGGCGT
TGCTGTG
GCTGGGT
GGCGCCC
TTGAGGA
TGACCCC
TGTTTCG
GAGTGGG
TGGCATC
GGGCGG
ACATTC
GCAATGA
AGAGGAA
ACAAACC
TCTGAAA
ATCTGG
CTCTAAG
TCAGGTG
TCTCAAG
TGTTGTA
AGTTTAA
TTTCAAT
GCTGCAA
GATACGA
CCCCAAA
TCTGTTCC
TGTCCTA
ACTCCAT
CATGGGG
GGAACGT
GGGCCCA
CAAGCA
GACTGAA
CCATACT

[illegible]

726

				GTCTTCA TTGA	CCATACGAT AAATAAAT AAA	REALKSVKAGKAPGLDGVHPEFLLTVGNMMPKNGLLNFI QMSQSAEECQGRKGTWIGWSTPRIPSYCGEYAKKWL AEFYSDVMATGNIPQKMNKSKIIALLKPGKPSNQPKSYR PIALLSVLYKLRLRLYNRLSPKILESIPDEQSGFRPNRSC TDQVLSLTTHIEAGYQSKKKHQLHLSTFPDNYRGRLS KKQKTSIAFIDLAAAYDTVWIQGLLYKLIKIPCSSTIRLIN NMLTGRAFOVITGERTSSTKVLNNGLPQGSVLAPLFFNL YISDIPNTQSRKFGYADDWAIAAQHESFEATEDILCSDL KLLCGYFRKWRLQPNPTKTEVSSFHLCHNSARRELVIF EGKRLSHNNYPKYLGVTLDRTLSTYKEHLTNTANKMRT RNNHKLCCGTTWGSTANTLRTSAVGLVYSVAEYCAPV WINSPTNRIDVVNLNDRMTISGTIKSTPTVWLPTLSHIP PEMRRKNALLREYNKITNNPQLPHSDVADLERRRLKS RRPALLSAKEMMNTNFEVDLWKQHWQEHCPKQCSF MTGEKPPGFNQPRKVVWTLNRIRTGHGRSADAFYKWG KTPSPNCDGAEKQTVRHIVEECPFRAYKGTENDFLAAT EEAINYIKCLDVNL			
					TAGTTCCTG CCTCTACCA CCCTTCCTT TCCTCTCCC TGCCACTCT CTATCCTTA CATTCGTCT CCCATCCCG CCACACATT TCATTCCCT TCCCTTCCT TTTCCGCAC TGTCCTGCA GACAGCAC TGGCGTATT CCCCCACCC ATGCCACA CATCCCCCC CCCTTCCTC TCCTGCTTC AATCCCTT GCCACCTCC TCACCCCCA	PPVQISSIHSTHQPNPTKPSKPGPNLSLAAPHLHRKILLC NPKYLHPIQTDVLALQVLEQHSHTTSKKLSNLFYSYR LGLPLSTPSTSRHTPQPLIADKPCPADLLHLPHQKLPH TTTQNPQPKQTVMNLSSKSLSPKIEVSVLSKGLTFCPTPKF NHAGLVKDLLSFSQSLQWKHHFATNPTQIHTNTNIEPC LTQFTPPSSQDPPPTLPNHPLLTFQNFLTSLNASPSFKSL KMDTNLTPSERNALYHLKTDPLMILPADKGSTTVVLD RKDYVAEGLRQLSDSSTYTPCHNDPIPKTQDDLQSLKS LGPSQDLSPEISILLTPPTPTPTFYMLPKVHKPNHPGRPI VAGYGAPTERISALVDQHLQPIHTNLPSYIKDTNHLHR LSTVPVPLPGTMLVTVDATSLYTNIPNAHGLAAVEHYL SQCPDCKPTTSFLVTILTRNYFSEFITYKQICGT AMGTRMAPSYANLFMGHLEESFLTQNPKPLIWFREFID DIFIWTEGEDTLSTFLQNLTNFTSSIHFTWSYSTQRTAFDL VDLFLKDGLISTSVHIKPTNHQYLLHFDSCHPHTKKS PYSLATRGRRICSDKQSLSKYAKRLTEAFTDRNYPNLV QKQISRALSVQSPSSPTPTVRPQRSTPLVTQYHPGLEQL NQVLROGFDFYLLSCPEMRNLTLPNLTPTPSYGIPPTKPT EIPYQISSPPPTTVFRRPPNLTNLTNLTPTPSYGIPPTKPT YPCNRRCKTCPIHHPTTYSSTPTGISYPIKGRATCESSH VYKLSCHNCAAFYVGMTSTQLSIRMNGHRQTVARRQL DHPVAEHAALHSVTHFNDCYTACAIWILPTNTSFSLELR WELSLOQYILHSRNPPLGLNLR	IPRO 00477	326	566
Pene lope	Penel ope- 47_L Mi	.	Locusta migratori a						

728

				GTGATAG TACCTGT GAGGGCC CCTTGCC AGGGATA CGGTGAA GACCTCA ACGGCAG TGAAGGC GGAG			QNVKMILLNTLKSEVGRET RKCKNVNERTEDNRGTT ENKSVRTFGIEDSRGVV HRPHDIDIEPEEAIDE TGDDDDIPVDLLKDIGN LDVTMTVLEKKQAKKCS RLEHKIEEVMGEDQFGR AVNEEICACFIDWQAFD RRLRNLYLQORVKVRLK SILFNLYGEMLAAREAL VLAKNQRLQHMMNRLV SKREKHLKIIVDNRLEN RARIAMAKAAFNKKRIL ALYGAETWTMRKKEKK VTNDDVLRVVGEEKSIL DVIEGKINGTAGRRRRR REHWHHKFSVRRRHGP			
				CACCTCT AGGGAGC TACCAGT TTCCAGT TTCATCA GCTCTGT GGAGCAA GTAGTCA ACACTCT ACCATCA GTGTTGC AGAAAGAG ATCGGAG AGAGACG CAGGGCG CTACCAG AGCAGGC CCCAAAA CAAACAT CTCAAGA GACGAGA GGTGGCG CTAG			TRARKPWITSEMLEKMEER KVRPHVPENEGTGEG TEQVEVLKTWEEYIRELYDTE SEVEKAIDMKSRKA TQVINKIYETGEWPKDF AKIARLLNR DAIGMMRILSERVL NINILKEIGINWRD VGRVGRQGCCLSP VGGRLINTIKYADDLV GMEINIEKSKVMIRI HCTNEI KVLKVKCYIWSI WTCWRRIEKIKWTD KANWIGHILRHEGLIH DLKDGSRYLKEEAED PATRQNTT			
				GTATGCGA CGGTAGC ATACTCGC AGGGCAC ACCGAATTC GAAGCCAG GAGAGTAG CGTGCCA CGATAACA TTCTGTAGT GGTCACCA AMAGCAGA GAGATCG CCATGAAA GTGCTAGG AAGCGCA TGGCGGC GCGGACCA CCTATGGA ACCTGAG GACGTCG CCACTAGA GAGCCGCA			EFAGRRKLGVLPTGR XLENDPTDKVDRKTR VPPRLYGLPKVHKQDV MLAPYVGKCVHHIRN VSLFTRVPLQDTLDL GQFYEQTEGVAMGSP QPTCFRYYVDDTFV TMELENEGQLPFLD YLQADSCHHPAQKNG DHLQRVFTNNGYSS GSLPTICRAYICQDQ SECSQTTATRAETS LPYAGPISAKISRL GLRKPQIYQIPCECG RQATKSAIAEHCL ESIEIKMGEDLINR HPPVELQTTTRDSS			
Pene lope	Penel ope- 38_L Mi	.	Locusta migratori a				cd003 04	203	287	

[illegible]

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

AAGGGGG
GATGAGC
GGTAAAG
AAGGGGC
ACCTCCT
TGTCCAG
GGATGAG
AAATTGT
CCCTAAA
GGCGGAG
GAACCAT
TAACATG
GTCAACG
GCAGGAG
GATGCAG
AAGGCCAA
CGGGAAC
CACTGCA
TTAACGA
TCCCAT
CAGGATT
CCCAAA
AGAAATA
TGGCTAC
GTCCAA
TCAAAAT
TTTCCGG
AGTTCA
ATTCCC
CATCGG
ATCTCG
AGGAAA
TGCTTG
AGCAAGG
CTGTAG
GAGTAA
GGAATA
AAGAAC
G

733

734

736

GAACACA	TGCTGCT	MMACAG	AATGTGT	TTATTGTC	GTGAAGA	GAAAAAG	GGGCTCG	TTTGAGA	AGGTGTC	GGCCTTC	TATATCC	AGAAAGG	TCTCGAG	GGAAATG	CTGGTGM	TTTGAAA	TTCAGTAA	AACGGTT	GCCTAAT	GGACAT	TGTTGGT	GGAACA	TCAACT	CCCACT	GGCGGCC	AACTAC	AGACGGC	TAAAGTC	CTTGGGG	AATATCA	AATCGAG	ACGGAGC	TGCACAA	CACCCCT	AACTATA	GCAAGG	CGTTGTC	ACCTGTA	GGGATCT
---------	---------	--------	---------	----------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	----------	---------	---------	--------	---------	--------	--------	--------	---------	--------	---------	---------	---------	---------	---------	---------	---------	---------	---------	--------	---------	---------	---------

[illegible]

AGGATGA	ATTGTACA	VFLGDKGRWRLNDGLPQGSVLSPTLFNL YTS DMPNT
ACATAGC	TAGTTTTTG	SAKKIQFADDLALAYQARNLEECEGTLT TDLHIMGDYF
TGATTAT	TATTTGTAA	KSRLKPNPTKTEVNAYHLNNREANRKLNVVFENVPL
AGAGTGG	TTCTTCTAT	EHNPPKYLGVILDRALTEKNHLEYVAKKVNTRIDIVRK
TTTCTAA	ATATGAAA	LAGSGWGANADTLRTATLALVYSTAEYCAPVWKN SAH
GGACTCT	ATTTTCAAA	VRKVDVVKLHQAMRIISGTTKSTPLPWL PVL CNIPPSIRR
ACAGACA	TAAAGACTGT	EAAAIREFHKSSTHKNSLLYNQLRNIPTSRLKSRKPPWP
ACATTTA	AGCATGAC	MQDSPLVYDVCQEWKDVVKKFFPPANSRLITDPTKR VN
TATCTTA	CATGTAAG	GFGLPRKLWTS LNRFTGHGRCSACLYEWGYKDAPAC
GCAGTGA	CCATACGA	DCGAPVQTMQHLIEECPPRKFPGLSELHLVTTT RAVEW
CGATATG	ATAAATAA	LRDLDVQL
TGGAAC T	ATAAATAA	
ACTATTG		
TCAACAT		
CTATAAA		
CCCCCAT		
TAGTCAT		
CTGGCCT		
ACGCCCC		
CATTCAA		
ACTTTTA		
CTGACCC		
TGCCATC		
TATGTIG		
GAGACTT		
TAATAGC		
CATCACC		
ATGCTTG		
GGGATAC		
AAGAATA		
ATGATAT		
AAATGGT		
AACATTT		
TACATGA		
CTGGCTG		
TCCTTGA		
ATAAJTT		
GTATCTG		
GTCTTTA		
GTGCAAA		

GGACAAA
GGAACGT
TTCGATC
AGCAAGA
TGGCAGA
AAGACAC
ATCACCA
GACCTAT
GTTTAC
ACTACCA
GTCAAGC
TAATGGA
AATCAAC
TGGCAAA
CAGGAAA
GTTATAG
ATTCTTT
GCACATA
CTCAACA
TAGGCC
ATTATCA
TTGACTA
TGGCATC
CAGATAC
CATTAGT
AATAATC
ACTGCCA
AAGCTCG
TTGGAAT
TTTAAAA
AACTGAA
CTGGCAA
CAGTATG
CTGACAA
ACTGAT
GTGACTG
TTGACTC
ATCCCTG
TGC AAT
TAATAAC

TATGACA
GATTTC
TAAAGCT
GTCAGAG
AGTAGC
AAAAAGA
ACGATAC
CTAGAGG
GAGGAAA
CAAATAT
TCCAGGA
TGGTCCA
ATGATGT
GATGAGC
TATATGA
GGAATAC
AACITAC
ACATTCA
AATAAAA
CAGCAGA
CGAACTT
CTTCAAG
CTTTGAA
TGAGCAC
AGAAGAA
GGAAATG
GCAAGAA
GCTACAG
AGAGCCT
CGACTTC
ACCCATT
CGAGTCG
TAAAGCG
TGGCAGC
TATTGAA
GAAACTG
GGTGCAG
ACTAAGC
CTGAGAG
ATGCACC

[illegible]

TTCCTIG	TACTGTTTC
GTTATTG	TCTCTGTAC
CTAGCAA	AGTAGCCT
TTCTGAA	AGTTGTAAT
TATGTCT	TGCGGCTAC
CTAA AAC	TAGATGGC
CTTAGAA	ACTACTGA
TTTTTTAA	GGTAACGT
TGAAAAAT	CTTCCTGTT
AATATTT	CACAACTCC
CGGAACT	ACTTAGCAC
GCACGAG	ACGCCAGT
GACCCGA	CTGATGCC
CCGCAGA	CCGTGCAG
ATTACAA	GCGTGCAG
AAAGAAAG	CCCATAGTG
TCAGGTC	GCTAGCTTT
GGCCGTC	TTCTAAAAA
AGTGATG	TTTCITTTT
CAAAATA	AAATTTTGT
TTTGCTTA	GACGTAAG
AGCCATT	CTTTACAGC
TCATAAA	TGAGTTGTA
AAAACTC	TATTTTATT
TCGTTAA	ATGTGACTT
T	GTTTTATT
	TATAATTTC
	TGATGAAG
	GCAATCAT
	AGTCATTGC
	CGAAACCA
	GGTTAAAT
	AATTTTAA
	TTTGCGACC
	GAGGGCTG
	CWTTTTTAT
	CCTTIGACG
	TTGTAACGG
	TCGCTGACA
	ATAGCAGC

[illegible]

AGGTGAT	KSKIFIPNSVKNNKKPTRKNFETNNIDADTKNIFKSKLQ			
ACTCCGA	GETWEEMKSIQIMKAAEESLPTKIKRKNIWWDKNEA			
CGTGGCG	LDNRQTLWQKFKSNPTQINLERFREGRKIASYTIRNEKR			
CGTCCCA	LFEKLSLQQIQVHFTQHKSRDYYSFQKRHLNGYEPPSLN			
GGTGGCG	FRDKSGKLGTDQENCNILAHYFKDLLNCKAPPEKPEF			
GATAKGG	KTTEEIPAESHPPSLDEVRAIKSLKNNKSSGEDGITAEI			
GGGTCTT	LKLGEEKVTLHLGTIKEIWQTEKIPEDWRTALIHPLHKK			
AACTGGC	GDRSDVNNYRGHSLLSVYKVL SKILLNRIEPLVDAQIGE			
TTGCCAG	YQGGFRKGRSCSEQIINLKNIIRTRAGQPHVTTFIDFRKA			
CGGACTT	YDSIDRESLIRTLLEFGIDRKTREIVKQTLDTKSKVKFR			
GAGCGAA	GKISDPFKIQTLGRQGDGLSPILFNITLEKIIRETLAEYTQE			
ATAAAAT	KVKGISFGGSVKGLTIHALAFADDIALSTNXDAIKQIEI			
AGCTCTC	LAKHAAKFGLQISQEKTKFMSNKFENVNKLHTIIGCIER			
GCGGACC	VKHFYLGLETITDITGTGTEKEAVKL RVAKLRRRAQILTMKL			
AAACACA	YNKKCLSRNAKLL			
AACCTTC				
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
TTGTTTAA
CCAAGAC
AACCTGA
AAAACCT
AAATCTA
AGAMGA
AATACAG
TATCGGA
ACACTGA
ACATTAA
CTCGCTC
ATCAAAA
CAGGAAA
ACATAAA
GAACCTGA
CTCAAGC
CTTGAG
CATCTAG
ATATCCA
AATTCTG
GCACCTAC
AAGAAAC
CCGCTAC
ACTGACG
AAAACTC
TTTTGAA
ACAGGCA
ACTACAG
AATTAT
AAAGGAA
AACCAGC
AGAGTAT
ATCTTTG
GAAAATG
CCCCITA
TTAGGTA
CAGCTTT
TATGTT
AGAAAAA

[illegible]

[illegible]

Daphnia	Daphnia			ACGTTTG CATAGCA AGCGGCT AACGAAT GCAGGGA CCGAATA TCCCCCG GGATAAG GGGATCC CCTATAC CAAGTGC CAAAGGC CTCCATG AGGCGG ATGAGCG GGTATAG GATAGGG CACTCTC TTGTCTTT GGAGTGT GAAATCG CTCCTAA AGCGGGA AGAAACCA GCAAGGT CAACGGC ATAGAAT GGAGAAG GCAAGAG GGAAACC ACTCCAT TAAAGAT CCTTTTG GATATCC CATGTCA AGCAGCA	Locusta migratoria		TGEIAEDFKKTHIIPKPKGTIKCEEHRITLSLISHMSKILT KIYNNRIISKIEQNLEDEDQFGFRNRNRTREAILMLRQIMD KMLDLKRPLFIADFVLEKAFDNVNWNLMEIVTNVGLD FRDRRIVHEL YKNQRGVINIHGEJQEAHIKKGVRQGCNL SPVLFNL YIEEGIKECKESMEGIKINGVQIKNMIRFADDIAII EISEENLQKALSRMNEILPAKYNMKINTKKTEIMVCSHE QITTDIKINDVPLKQNVFTYLGSKITPDGKSSADIKSRIA QAKLAFNNKRMLLTSRNIKEETKKAFIKCYVWSVALYG SETWTLGVCDKRRLEAFELWCWRRMQKIPWTAKMTN EEVLRIGEKRALVKTIMKRDKLIGHLYRHNFEIINVM EGMVEGKRGRGRPRQLYKQIQRDIGVESYQEMKRLSE NREVWRRRIAANQSSD	cd016 50	111 851	111 2
---------	---------	--	--	---	--------------------	--	--	-------------	------------	----------

ATGCTGT	ACTTTGTC	DQVVVGGAGNSLDRDGEYDIGDGLERIAQTGTGNNVH
GCACATC	TAAGATTGT	FVQLFQRHDPCLNTAVSRVNMGLEKALMTEGMDHIA
GCTCTCG	CGTGTGCAT	VVPVETIQRSDFTRHGLHLNRYGKGRALKLIGDSVVG
ATTTTTA	AGGACTAC	GGXTHGKIPVVVGVRAAPFLDGGCTFFRLKTADRYSC
TAATTTT	TTAAAAAT	EVPLRXEPPLTKSKNISNKGITVFHQNIQIRDKVNEXLI
AGGATTA	TTCTATTAC	DTDSIIGISEHFLSGVTIQRPLPGYRLAAFYSRSLQCG
ATATAGT	CATCTTGT	GVAMYVKKRFFGVNRRFFKFNDFESSIDFSKHCTEQIFEC
TATTTTA	GTAAAAAT	CAGVVNVNGTKLVVIVYRSPDSDFKMFPLKLELVLSL
ATTGTTA	TAATGTATT	YRKYNLIVCGDENINCVGDSRRKSQLLDLFNSYNMMQ
TTTCGTG	GACTTGTC	TVFFPTRVQGNSSSTIDNIFVHSSLLEGHSVRIVNGLSD
ATAATAT	CATGACCAT	HDAQLLILKSICSTNVKHKYQLLRKKGKVFVLKQMLNI
TCAATAT	GGAGTCTC	NINYSEKVNPAIETFLSLIKEQEWQDVYSADNADDKY
AAAGGTG	ACTCCTCAA	NAFLNTFIMLFESCFPLERSKQGTSSNRQPGWLTGIRVS
TTCTAGA	TTGGGTCCT	CRTRRELYQNVRSSQNEADVAHYKQYCKVLKNVIKNA
AAAGGAT	ATGGAACA	KSMWYAKRIANSRDKIKTIWVVKEVSGQQQVEDITS
CGTCAGT	GAAAAAAC	TFLENAPRLLPEMVLHDTSGIEISIKSLKTKDSCGYDG
AAITTTA	ATAAAATA	VSRNVSVTDKLDICAFNNHFLNIVGELNKTILVSTGNHI
GCGTTTG	AATAAATA	VSSRLKYCSPYVSPVLSHCNFSFQNGQFPDRLYSVVK
TAGTGCC		PLYKGERDNVDNYRPIPMSPVFSKVEKAVYSRLIDHF
GTATTGC		SSHNLSSNVQFGRSGLTTEAIFSLCEALDGLNKLRLT
TTTAAGT		LGVFHDLTKAFDCVDHKILLQKLEHYGIRGVAHNWFTS
GGTGTTA		YFNNRQKVLIRSVENGYDVQSQWGTIKCGVPQGSVLG
TTTGTTT		PLFLIYINDMPSSITGCKIFLFADDTSLVVKDVARNID
CGCGTGT		TVSNNAVQEISSWLVENKLMNHSKTQFLQFLTHNLTK
GCGGCAC		ADILITHNGHMISETQFKFLGVHIDSKLSWKAHVQDLV
CGACGTC		PKLNSAIFTIRTISEISDFSTRKVVYAYFHSMLTYGIVF
TTACTGC		WGNSPDAKRIFLAQKRAVRAISGVSSRTSCRPLFSSLGIL
TTTGTTTA		TLASQYVFSLSMFVNNISLPFRVSSPHSVNTRQKSNLH
CATCTGG		LHHTTTLTLVQKGVQYSAASIFENKLPQELKNLSTNPHAFK
TGGCGAG		SKLKSYLEMAHSFYVEELLEKLNN
TGGTTCC		
AGCCGTA		
GCGGTCT		
CCGAAAG		
CTAAGTT		
CCAATAA		
GTGTTTG		
TGTTATT		
GTAGTTA		
AATTATT		

AAAGTGA
CTAGATT
CAGGAGG
TGTATC
ACGCTAG
TTGTTTAT
TAAATTT
AGTACTT
TTCAACA
GGTGACT
TGCTGTTT
TTAACGT
AGTTGAG
GTTGGGA
AACTTTA
AGTTACG
GTATTTG
ATAACTC
TTKTACT
CTCACTG
TGTTTAG
TTAGTTT
ATTAGTG
ACGCTTG
AATTGGA
ACTTGTTT
ATACAAG
CAATAGT
GTAAAGT
TTCAAGT
GGTAGTG
TTATTGT
TCTTTCT
GATACTA
TTAGTTG
TTGTTACT
TGATTTT
GTAGAGA
CTGAGGT
TTGTJGA

[illegible]

CCCCCTCCC	QFLPTLRGDLVPLPSPTRPESNTTPDAVLVSPNLWDACR
GTCCCCCGC	RKSLTSLAVTTTPYSSPSTPAAPVAPPVLCPLLGLGRL
TTGGCTCAC	QAEVLIGSDHAPVLLTFNPRGRPGRAIRPVPPRLVH
CCCCTCCTT	DYSRADWDAYRRSINTSVASHPLSLHTPDDIVHASSFLH
TCTCCACCT	QTVADAVATHIPTKAIHPHPPLPPQAVLLRESRRLYR
CCCCGGGA	SFLRTRDRDTLLRHRQLQRHNRLLNAKKRRDWRQTCT
GCTCCTTCC	RLNATLPINSSKYWKAHRLTGSHSVAPYPLLNHNLPLP
TTCCGGTTT	DNASKAQLFADHLSNIFDLPDNPRFDYSRFPDVHERTDT
CCCCCATTC	SVPLAPGFQYLDHIPPSDLNTPITAQDIVHTLTKRNTA
TCCCCTCCC	PGHDIRITYRHLRACPPSFHSALASLYNVILFTGFYDPLW
CCACCTCCC	KTSRLMFLKPKPPEASSYRPICLTSVFESKLVESILSRRI
TTCCCCTTC	HEHLRLHQLPLSQCGFRFSFVDDQLRLVHLLSVQVN
CCCCGGGCC	SRRAAILVSLDIEKAYDRVWHSGLLFKLQTYALPINYVR
TTCTCCATC	LIAFSLSKRPSYVTVNNQNSRLFYPSAGVPQGSVLSPLLY
CTCCCCCCT	LLYSADMRRPQPVPVHLLQYADDTAFALAPSLARSQHL
CCCCCTGCA	LQSHLSEFSSWCNQWLLTVNPTKSQAIIIGRTTPSFRHRD
GGGCCAGT	LYLTLHGRPIQTKTLRYLGLTLDRDLSWNPHLKTVRQ
ACAGTGCT	KALSRLNLLKCSGRTWGLNPSITILYTYRSLMRPVLCCF
AGTGTTTTT	HPAWISAPSYYESLKILERHALRLAYRIRLPSPTIRLYDL
AAGTGTTCA	IPFPHLVQFLQHRLKYIHRKIESHNPLVSPLLDPLPLPR
GTGCTTCGT	LYRHIPPSLHLHTLYVLSEGGFRHLPLPDDVLHNRDPDIYP
GTGTATAGT	SFQL
GTTCGTCG	
CCCAGTGT	
TTTTAACGT	
TGCGTACAC	
GCCGTGG	
GAAACCAAG	
TGTTTCAGTG	
TTTTTTACC	
TCAAATGTCG	
CCGTATGT	
TATGTCCTG	
TCGTCTGCA	
GTGAGTGT	
TTTTAGCCA	
CCGTGGCA	
ATGCTTGT	
TTTAACAAC	
CATGATGT	

[illegible]

[illegible]

				AACTTG AATACAA ATTACCT GGTCCCTC CAGGTTG GGGGTTG TGCAGTG GGCCAGC TCCTCAC TCACATA AAAAATAT AAAATGC TAAAAAA CCTAATA AT	TAAACAAT CCCAC TGCC AATCATAG ATAATAGC GCAACAAA ATTATCAAA GATAAAAT ATCCCAAC ATATTTCGT TCAATTTTA AATAAAAG CTTGCCCTC ATAGATAT ATATTCTCT GTGCAAAAT AATGTAAG CATATTACA TCCTGTAG AGCTGTCA AGTAAATG TCAAAATCC GTAACGTGA TAAAGTCT TGCATGAC AGCACAAAT	RRRLVLINKQDYIQKVFEEFDKNGIIPCDPINKYIAQT RKIHKSSSILSDKDKRKYPMPNPKTPQLRGHPKTHKELS PIRPVVD SINCPYTLISKKLNLLQKYTYKKYTIKNSR TLAANINKIDIPINGKFVSFDIINLFTNIPVKETINIELNLL SXNKLPSINTYIEIINLLHLVLEQNYFKFEDNFYTPDGVVA MGCNLA STIADIFLNTLEERFFLHQETNKIIYYRYVDDI IVLIDGNETDINNHLNKLHPKIKFTMELETNKTNHLN LTQNVQNKHRFEIYRKPTTTTNNIIHAHSCHPSSHKLAF- HSMINRLINVLDDAAKQKEMTILQYIAQQNEYDPNNIL KLYNNKTSRPPTTRTEITLQRETSKLNKPIYISTPYLGKI SNKVQQIFRKNTTTFAEQNRHRLKHKFVTHITTKPSKFLN SGYIKISCSNCKYYIGQTGRNFTRYKEHTRESEKNPST FYQHLQSENHTVDTIDNALEILHTVQKGYEMNILEEME VCLHAIKNTNNLLNDOTDFKHKNYFDNFIHILVHDNG	IPR0 00477	124	263
Pene lope	Penel ope- 55 L Mi	Locusta migratori a							

[illegible]

<p> GTACCC ATGCCAG ACAGGTC AGTTGGT GAGATTC CAGACAA AGAGTAT TCCCCCT AGGGGGC TCGTCC TTGGGAG TATGTGG GTGGCAA CCACATG GGCCCTT AGCTAAG TCTGGCA TTACTTCC ACTTACT TGTGTCA GGCTTCC TCTGCAT TCTTTCC ATCGGCC TCCCTTG GCCAACT CTTGATT TTCCGAC CCCGACG GTATTAG GTTTCGA GGCCCGA GGGAGTC TTTCATT TCCTTCTT CTACTCTT ATTGTCA ACTGCTG CCTTGTA GTTAGAA </p>	<p> RAIQQLKNGKAAAGMDNITPEALKVDVSTSVEIFPLFKK IWEEKIPTDWKKGLLIKLPKKGDLSECDNWRGITLLSV PSKVLRSVILERIKKRVNSRLRKVQAGFREGYSTQDQIN TLRHLEQSTELQSSLYLLFVDFEKAQFDSVMRKKMWKA MQEFGPKKIITQEMYENYTCQVLHQGMLSEPIPVDT GVKQGCLLSPILFLMVLDLVMQKAMNRRRGIQWGLTG RLEDLDFADDLCLLASSYKDMAMLNDLKTASEVGL KINGKTKEMRCNNRNEALFFFKPRDRASQAVLLPGKF SHPGRTIEDIGNRINKAKGAFAMLRPIWKSRELNSYTK LRIENTNVKSVLLYGCESWKTNKQLISRLQIFVNRSLRNI LRIWPPQVISNEELWRRTHQTPIDIEIRRRKWKWIGHIL RRGKDNVARQALEWNPQGSRRRGRPRNTWRRSVVAE GTEKGGKAWTEIKTLAQNRVRWRKFVDALCST </p>				
--	---	--	--	--	--

				GCATACT CCAACTK AACATAG AAGGAAT AAGCAGG TCGAA	ATATATTGT CCTTGTCAT ACATATCTT ACAAATGTAT ACAATTAG AACTATGTA CTTGAACCA AGCTCTGTA TCATCATTC GATAAATA ATAA	SWGCSAQTLRSTALALSVAEYCCPTWMNSVHSRKID VQLNQTMLITGCIHSTQTAGYQFHPTSSSTRSDALLKI WRNIQKNPQLPHSDTSTAHEQRLKSRKPSWKTAAHLD VTNFINEAWKSHWESSVTIWLKTLQCEFFQVSHSP DASGESSTASEQDKGTADIYSTNGAGQHLPSTVELPYK Q				
				CCCTCA GGGGCTC AGCATTA ATTGCT GAGTACG GGCTTGG CGAACCC GGGGTCC TGAGCCG TGGGCTG TGTCGCG CCAACAG TCCGCTA CACCATA AGCTCTG TGCGTGC TTCAACG ACCATCG TTCGGCG CGACGGT GGAACAT TGACGT TTCAGAG CCCGGG ATCTTGG CTTAATC GCCCGGG TCGCGAG GACGGTA	TGAAGATT GGTCTCCAA GACAGGGA TTTCAGACT GCAACTGA CCCGTAAG GTTCTGGGA TTTGGACCT CGGAATGG GCTGCCTTG AACGAACC GAATAAGC TGCGCGTA GTCAAGCC CCACAGC CGTGTGA GATCCTCCC TGCGAGCCT CTCGTAGG GAGTCTGTA GTCCTCTGC AGACTACG TATTGGMC ATACGTGGT TGACACAC GGCCACCTT CTACGCGA GAGGACCC ACCAAAAT	GQVWRSGLFFENASGAVLIQTASPAQSRALLACDKLGD VPVSTPHKSLNMVQGLIYHRDLLSQTDEELRANLERRG VHFVRRVYRGPKDQVATGAFILAFEGDITLPEKVKMI YRCDVKPYVPPMRGRCWKFGHMSRCTSSPNCRECG RPSHPDAPCTPPPVCVNCGQHPPPCPNCVLRERKIM EIKTLDRLTIQRHVKSMGSIRAAAPSGKDDNIFCRSHSGT SFTAINQCSSLCCGPGPTTSTAPSMGGGTSSVAPTTPP VGATTPSGTISVPASRRRNSLLRLLSLGRGPWDTPFRG FSQCTGDPQRLKGRPAGRRASRSSNPDAATTEKPSQA SKPKQREGQRKKKKETLVATKPAQAPSSASEDEVEI LASPEDLAIPGSPESRIADTGDSEQLAADVMDEPVLLYG FSARRPVLLQWNCGGGFNRLAELQQLCRFPVLCALQ ETWFPQRTPALRGYRGYKNAAYDRASGGVCYIVL ATVPSAAVPLHTTLEAVA VRIRAMQEFITCISYLLPPDEAV SIADIAALIAQLPPPELLLGDFNAHHPLWGGTQTTRGR DVEALLSQRDVCLLNTGAPTHFCVARGTYSADLSVCSP SLLPHFRWKVHDDLCGSDHPILVSLLFQAPAGHQPRW SLRRADWQRYASAAVVPVPLPADVDMAVHQVTSSIVS AAEATVPTTSGSPRRKSPVWWTPEIAEAIRDRRRALQRH RRHPTMENLIAFKRLRARARQLIRQSKRDCWEQYVANI GSRASSQVWVSVRRFLRGHRPPAAIPSNIIQGTVCTDPTA IAEHMADHYACNSATAA YSAAFRALKVVERRHLSFHT RRDETYNAPESQWFEQTALAACPDTPSPGDRHSQFLRH LSAESQHQLLALFNHLW AEGAPPSQWRESIVPRNVQCA IQSVGIPNCPCMPRYLPGRDPDPQVPSPFSRVPPTASR PVQPSMGRGGFPITMAGKHRYPVMPKPADPQVADSYR PISLTNVLKLLERMVSRRLMWLEHRGLLSRQGAFRQ GRSTVDNLVALESARITAFSRREHLVAVFFDLRKAYDTT WRHHILATLHGWGLRGLPLFIQNFLLSLRSFRVRVGTSL	cd016 50	993	124 6	
Nim b	Nimb - 22_L Mi	.	Locusta migratori a							

CAAACCTC	GCCGTTGTG	SSPRIQENGVQGSVLSVPLFLIAIDGLAAAVGSSVSSSL
TATAAAA	GTGCACTAT	YADDFCLYFSSTGMGV AERRLQGAITKAQSWAMTHGF
AAAACCT	TGACGTGGT	QFSASKTRVMHFCRRRAVHPDPELYLGXDLRVEQTHR
CAATCTC	CCACATCCT	FLGMQFDARLTWVPHLQQLRKCKWRHLNVLRLCSHTA
AAGGTGT	GGTGGACT	WGAARTSLQLYKALIQRLDYGSLVYDSAAPSALRVL
GCTACGT	GCCCACTTT	DPITIVASVSWGSPSPESAGSSYWLLVARIHASPDPNHR
GCCGAGG	TAGCCGACT	LLFPAAVLPSPRRRPRSGPIAVRVSLLSLSLPIPTF
AGATGCA	TGAGGCAG	HYLSRTPPWVFPRLQLRLDLSSDSKASVSPMVLCRQFQ
TGGCTGT	TCCTTTACC	ALIDSFHGAEVVYTDGSLVDGRAGYAYVHGDHVEQHS
TGAGGTT	CTGCCTTCT	LPCGTSVFTAELVAISRALTFRDCPGESIVICSDSLGLL
GAACAGC	GAGCTTCCA	ALDQFMPKHHLVRSIQDAVFTLGESGRSVFVWTPGH
CGTAGC	TTGATTTTA	MAFRETNESITWPSRPPHHGP
GAGCAAC	TCTGACGAT	
CCTGAG	GCCTCTATG	
GAACCTG	GTTGACCGT	
CCGCACC	GTTTTAGGT	
TCAGTTG	TTTATTCGC	
TATAAGG	GCAGCTGG	
CTTACCC	CTTTTATTTG	
AGGCAAG	GTCTATTTA	
CGGGCT	GTGGTGGG	
CTGCCTG	GCCTATTTG	
GGTGGAC	CCCGCCTTC	
CTTAGTT	TCGGGGCTCC	
CCCTAGC	CACCCCTCC	
TGCTCGT	AGAACAAAT	
GGGACCA	AGGCCGT	
ATGAAAC	GTCTTTTAA	
GAGCAAA	CGGTTTATG	
TAAAGCG	ACTIGTTT	
TGGCCGG	TCAATTAAG	
CGCAAAA	GATTTTAGT	
CGGCGTC	G'TGTGCAG	
TCTAAGC	AGTGGCTG	
GTGGCAA	GCTCATCCT	
AACGAGT	GTATTATCC	
TCAATGC	TTGTGATCA	
GTCACTA	GCCAGCCA	
TGACCCCT	CACCTCTCT	
AAGACGT	GCTGTCTTA	

762

CAAGTCT	ATATGATG	KVVQNLSKSSAGIDEIPVCLLKDCTQSIHPLAHINESF		
TGGTATC	ACAACATC	TKGCFPETLKHAKVLPPIHKKGDPESIESIENYRPISSLSSF		
ATGATAA	CATACAATT	SKVMEITIMKNRLMAYLTKYNLLCKEQSGFRSGKSTSSA		
ATGIGTA	TTTATTGTA	IAQFTNAVLEALDKGDHV TGIFLDLT KAFD TVDHTILLN		
AAAGTAA	TACGGATG	KLEALGVRGVANMWFHSYLNQNRVQQVEITKTNNSNHTI		
ATCTAAA	GATAAATA	KHLSPEYINIGVPQGSVLGPILFLIYINDFPQSVKHGKSV		
TACATAA	AAATAAATA	LFADDSNILITDNSQIKLNQKAEDTLVHVSRRWVENNKLT		
AAGAGAT	AA	LNVKKTASTSTQINPQCKKNNINFHHRKRPVTNPLKIK		
CATGACT		DENIECVSTTKFLGMHDSQLKWTEHITVLGKKIASACY		
GGCATGC		ALRIINSVCSSTRTAYFANIHSLSYGIIFWGTNEQNRQ		
CGGCAGT		AIFKLQKRAVRIMTNSSRAHCTELFKYMGILTVPCEYIL		
GCAATTGT		QQSSTLEGPLYRIQIYGN TNC SMRIHFTAVIYKENINQY		
GAATGCC		PPNSSIHEHETRSSRHLLNRKNTTKTQNSILYNGIKLYN		
ACGAAAA		RLPQDIKNKDKIQDFRNSLKKFLEKCYTYTVQDYLN		
GTGAACT				
AGTGAAT				
ACACTAA				
AGTCAAA				
AGAAGAA				
ATAATAC				
GTGTGCT				
GACAGAC				
GATTTAG				
TGGCTAT				
AAATGCA				
AAACACC				
AAGAACT				
TCTGAAT				
AAATATC				
AAGAACT				
AGTAACG				
AAACACA				
AAGAAGT				
GGAGCAA				
AATCATT				
GCCAAGA				
ACAAGAA				
AAATATTA				
TTGTTAA				
TGGACGT				

TCTGTGC
GACTAAG
GTGGCAT
GTACCAA
ATAAAAA
MTGCAAA
CAAAATGA
TAAAGTGC
AGCTCCG
GCGTTAA
TGCTTTC
AAATAAAA
TTGCCGG
TATAGAA
GATTTTG
GCCTGA
AACAGAA
GTCGCGA
ACGATCT
CCAGTGA
CAAAATTA
CCATCAG
AATGCAG
TACAAGA
AAATCGAA
TTCTATTG
CAAAAAG
TGAAAAA
ACGGGTT
CTGATAC
TGGTGAC
AGCCACA
CCAGGAA
AAATTCC
GAGAAAA
TAAATGA
ATATAAC
ACATTTT
CTATTGC
CACAGGA

ATAACAA
AGCCGAG
TGCCT
CTATGT
AGATACT
GAGAAAC
AGTAACT
CAGTCTG
TGACTCA
CTAATAA
TGATTTT
GTCATTA
TCATGGG
AGGCAGA
ATGATGT
CTATCGA
AATGAGA
GTAAAAA
TKCCACA
CGGTGT
TAAAAATC
TACACTG
CGTGATC
TAAAAGT
GCCAAAA
TCATAGT
TACTAGC
ATGCCTC
ACAGATA
TGACCTG
GACGAAA
ACTCAAT
TGTAAT
ATGGAAC
TTGCAAA
AATAAAC
AGACAAT
TCCCAA
AATGTGC
AAGTCT

Ving i	Vingi - 14_L Mi	Locusta migratori a	CCACCCG GGTCGCG AGGATGG ATAAAC TCTATAA AAAAAAC CTTCAAT CCTAAGG TGTGCTA CGTGTCTG ATGGGAT GCATGGC TGTGAG GTGGAAC AGTCGCT AGCGAGC AACCTCT GGGATA CTGCCGC ACCTCAG TGTGATA AGGCTTA CCTAGGC ACGTGGG GCTCTGT CTAGGTG GACATTT ACTTCCC TAGCTGC TCGTGGG ACCAGA TGAATAT TCTTTGT TTTTCCCC TCCAGT GGG	TTGTGAAGC GTCTGCACT CTGGCCTAA GCTTTTAGG TTGGAGATT TTAGTTTGA TGCAGATG GCTGGCTG AGTTCCTCC TTGTGTGTC AGCCAGCC GTGCTAATC TACCTAGTC CTCTCTTGG TAACTGTGA TATGAATGT TTTTATGGC CTCATTTAA CTAGCTTGT CTCTTATCT TGTTTGACA CTCGTCTCA CTCTGTTT CGTCCGGGT GCTAAAGA CCTCGACGT CTGCGCCT GATCCATA ATCACACC ACACCACC A	FRSTAAEYNCPFSMWELDSACLQLVIRALDMTGYITVC YSTLQNGTEKSSSHFLMSFGRQVIFPTHGVKQFEGQFGL QVHYCRRIQLPFLHVGARFCLSAARDTCPGHDIRHYSM LQHLTERNREILLNFVWVASGHFPDSWREAVLVPIK PGKDRTPSSYSRIALTSCVGKTLERMVNRRLVWMLES RQLLSRFQCGFRKYRSTFDNLAFLEAAIQQAFLRRQHLL GIFFDIERAYDTTWRHLILRQLHEWGFGRGAFPLLCPPFL HAILGIESVTCRIVLSKRITVFLRVAFMGFSWSLPLFVVF LSPRYPYRVGDVLSDRFEQENGVPQGSVLSVTLFAIAI NSLTSVVKSPVQCSLFDVDFSIICSSSLATTTTQQLQTL RRLDDWAKKSGKFSTEKSVCVLFNRSRSLNLPRLMR DTVLHFKEIVRFGLJFDSKLTWLPHLKDLKKRCLKALS ILKCLSHNSWGADRTCLLRFYRAFRVRSRLDYGCTVYGS ARPSYLKMLDVVHHEGIRLATGAFRTSPISLCAEAAEP PLSIRQLLGRQYKMLSTPHAPAYHSVARPPMAGCFIL GRGQHDPLGFARIVSHRWVYLLMFTSGLEQISSLPSE AQHYFRLGSFKKGTCTPDFTYRLNFSKFIGITILRWSTPMV QAGDSLGCIVFPTMLPGFASHSLFCLHSRASRDLISRNA ACSEHTVSYLLRFTRTATIDSLYPDEKIHSSYRTNCTPSS GGVSRCHSAGSWHVGIrgNEVADGLQRRPAAATTYTIC HSPAGCPLCGSAEVL SAVGRHVAAVTGNKLRVSKPTVR RGNLLPVTDA	MLTISVNI EGLSSKENLLHEMCQSNKCDVLCLQETHR DNNHRRPNIPGMKLV AERPHAQYGS AVFTKPDQLQVKST CHTEENNIETITVELKDYTVTSVYKPPSEIFAFTPPANFV KEKISFVIGDFNSHSHVWGYASEDKSGEAVLAWAESQ	cd016 50	461	717
-----------	--------------------------	---------------------------	---	--	--	--	-------------	-----	-----

				CAAAAGA GTGGACT CGTTTAA GAGCCCA CCTGACA ATTATTTC GCTAACA TCGAAGG AATTTC GTATCCA AACAAAGA TTTAAATT GCAGAAA TCTGCAG AACCAAC AGCTGCG ACGTCTCT GTGCTTA CAGGAGA CCCACAG GGGCCTC TCAAAAT AAGCCAA ACATTGT AGGT	AGTTCGTAC ACGAATAA	GANAKVLRSTALATCFV AEYGCVPWERSHHTKKLDV VLNNTCRKITGCLKPTRV DYL YLSGIAPPNRRSFLSSIE RNKQTNDRHLL YKNQPV PKRLQSRFSFMYSVPTVAS IKSLWQEYPEALITEEAL PQGSRLDWQHWKTLNRLRV GVGRSKQLMKRWGYTDS DECECGAIQIMHLLKCPV DASRECTQEDLLKYTGT AAAYTDFWTSKI			
CR1- 3B_L Mi	CR1	Locusta migratori a		TTCCTCGG TATTGTA ATCATGA ATAGGA CTGTGAC TGCTGTG TGCGGAT GCAGTCT GAGCTGG TGACACT TCGCTCC CAGCTTC AAGCTGG AGGCTGC GGTCGAG	TAAAAGTG GTTTCGATGA ACCCCTCTGC CATGCACCT CAATGTGA CTTGCAGA GTATCTATG TAGATGTA GATGTAG	MEPLREMAAKEGKKANV HSVCIPGGVIPDVERVLPDA MKGTCGSQLQVVAHVGT NDVCRFGSEILSGFERLSEV VKTASLACKMKAELTCS IVDRIDCGPLVQSRVEGLN QRLRRFCDRVGCRLDLR HRVVGQVPLNRSQVHYTQEA ATRVAGAVWRGLGGFFFR ELCGVDWAVFFFRLEGL GNTQKGFQSQRVQVKDR TNTDLKTIGLTVVNCRCV GKVPELQALJESTESQII IGTESWLKPEISSAEFTK GQMVF RKDRLNKIGGGVFIATG GLSCSEIEVDCSELVWVE VI PDNQNKLIIGSFYRPPD SDSIAEMFKESLDQISK TYPKH TLIGGDFNLPSICWQY MFKTGGKHKTSSEIVLN AFSEN YLEQLVHEPTRIVNGC ENILDLLATNNPELIKSI KTDGTGIS EHRVVVARLNIGLSK STKIKRKYIYKKQKIRL TJLSKKI YLFKKADKNINAFLEN LQSFETNNVSVDDQMW LTFKEI ISTTIDRFIPNKLTKR GLTRRGTONRLDHCYK SKEKSMIP NLNAQKIPKTNLLQK LEIQHTKQINKEGADPP WYTKQ	cd016 50	772	102 2

GGGCGTC	VRLLQQRKKHAKFRITKNPQDWQSFTEIRNLARTS			
GTTGTTG	MREAYNSYHNETLSRNLVENPKKFWYVYKTSKGKTQSI			
TTGGCCG	PSLHNRNGDATDDSSAKAELLNTAFRNSFTREDEANTPE			
GTGCTGG	FVSRTAASMRNLEVDTPGVLKQLKSLNKSXSPSPDCIPR			
GACAAACG	FLSEYADAIAPYLTIYNNRSLSERVSKDWKVAQVTPIFK			
GATCCGT	KGNRSDPQNYRPISLTSVCSRILEHILSSNIMNYLEENDL			
TCCACCG	LSHSQHGFYRSCETQLALYSHDVMSAIDKGCQIDSVF			
ATCGGCC	LDFRKAFDTVPHKRLLIKRLAYGISPQLCDWIRDFLSERT			
TGCGGCA	QYVVIDGKSSSRTEVISGVPQGSVIGPLLLFLIFINDLGDNL			
GCGACAT	SSSLRFADDAVIYRLVKSPEDQKQLQNDLDKISAWCK			
GCCCTGA	NWQLTLNNDKCEVIHMTTKRNPLNFSYTINQSNLQAVN			
TACTGCC	STKYLGVTTNNLNWKEHIENIVGKANQRLRFIGRTLRR			
AGCAATG	CNSSTKETAYTTLVRPILEYYCCALWDPHQVGLTDDLEKI			
AGGCAGA	QRRARFVLSRYRGESVTEMMHELGWKSLKERRYLLR			
CACCCCA	RDLEAKFLSPTFSSECEHILLAPTYIGRHDHKNKIREIRSK			
CCGTGGT	TERFRCSTFPRAIREWNYREIS			
TGCGTGG				
GAGGTGT				
CCAGTGC				
GTGGCAG				
GCGGCGA				
AGACATT				
CCGGGGG				
ACCGATC				
GTGAGGC				
CTCCCCG				
GTTCGTC				
TTGCGAA				
CAGGTTT				
GAGGCGC				
TAICTGTT				
CGGACAA				
TGTTCTT				
GAGCCGG				
ATGGTGT				
CGCTCAT				
CCTGAAG				
TCACCAC				
GCAGCCT				
GCAGATC				

[illegible]

[illegible]

				ATAATAAA TCCACAAG ACCACGGC ATAATAGC CCGGAAT TTTTAGTAA TTTTGACCA TTCCGGCCG TGAAAGTTT ACATTTTAC A		LRITVTQLIKKSSIEQCVCQKSLFNSVALPPRLYGLPKVHK EKENVPPLRPVSAIASPTYLIAKFLTTLRPHVGRSDSYI KNSRHFISKLDILVNPEDILVSFDVVSFLTMVVPNEVLQ QLNLFPPDIADLFKYCLTTTYFKWDDDFYEQTDGVAM GSPLSPVIANFYMEKFEEQALKTANKKPAIWFRYVDDTF VWVRHGREELDRFLQHLNGINSRIQFTMEKEADGKLN LDVSVEFKKGNRLGHTVYRKPTHDRYLHRDSNHHPO QKRGHIKTLADRARNICEPELDAELKHLNALLKNGYS STEIRRALRPTRRVENDERVPVSKVFLPEKDVTDRIK ILTKRNVAVIYKPTRKIQDHLKSAKDARKPLEKAGIYRIP CSCGDVYVGTTRTVKKRLDEHKGNCRRGEIDRSAVA EHVFQPGDHLIQDKTQVLAATNGYHERLYREAIEIVKH PNNFNRMEEGVKLNEIWPVLKKKCNLSLPPWDDINGRR QSATTNCARMSKTSDDVTPLRASPRKRSLSASVSEPGCAS DIQKCHDPP					
						MSATTTIPNGSCQGNQCNQNDLNRAGPVLNVCHLNI EGISRSKCDFLSKNLLSNNTDVLALQXTHAEDLKNCAR GNIHGRLVGEIGHRTYGVATYARQDLDDVSIHESCCDD DIFVLAVKVS GTTIVNVYKPPNIQWQD TVLPVFEHPSLY VGDENSHSHWGYSSDNENGIALAEWAENNNLSLVFD AKDSSFTFSARWKRGYSPDLFCSCDQTLRSLSVTRKVL GTFPRSQRHPVLISVGIKIPLIRSFPRPRWNFNKADWEAF TSKLEDNIRFIPPTTENYDRFAGLLKATAKACIPRGYRKT YIPGWSQECEDLYNEFISTEDSNTADLLNSLDNARREK WVKVQSLDFTHSSRKAWSLLRKLSTSVSEKKLPPTH DPSDIKRIVALS KSKRQSDRLISYKLN EVKKCQQA PL EVS KPFCLDELHVAISDVKTGKAAAGIDDIFEFIKHAGLK TRLWLLNLFNNILKDSLPKAFKRSKVIAVLKPGKPADR ADSYRPIALLSVCYKLLERLIYRIVPILTAIPVEQAGFK QGRNCCDQVLA L TSFIETGFEKKCTTSV AFLDLSAAYDT VWKKGLLV KLYNLIPCPILRLNLSMLSGRLKILYLGNT SGFRQLHDGLPQGSVLAPILFNVYADIPATQSRKFTYAD DIAIASQRDIHVTTTLTSDLTLRNYFAMWKLKPNPN KTEVACFHLNRMADYKLNISFDGVQLEHSEHPKYLGV ILDRSLTFKYHLHKTADKVKTRNNILQHLAGSSWGASA KVLRTTALS LVYSAAEYCSPVWLSAHVRDIDVQLNRA MRLISGTLMATPLQWLPILCNIVPPDLRRKEALMREYK ILSTPELPILQDLPPQVSR LKSRKPP LKTA VQLHESGFSPK SNWARTWDSFNGXNSIWIDDPSELQGM DLPREWSLT	cd016 50	490	744		
Vingi - 27_L Mi	Locusta migratori a	GGGGGCG AAGTTGT TCGCGTC CGGAAGC AATCCTC AAAGCTC CACGTGC CCCCTGG GTGGTTC TAAACGG ACAACAA TCTCTGG TAGCTGG TCCTGTC ACCAACG TGATCTC CCAGCT	TGACAGAG ATATGTTT ATCTTATTA ATTAGACA CTAGTATAT CTTACTAAA GTATTATAT GTAGCTGA CATTATACC ATA								

[illegible]

				GCTGTAC AAAGTAC TTTGAC AAATGGG TAGTAAG GAAGGAC CAGGTCT CCCAATT GTCAAAT ACGTGGC CCCAGTG ACTGAAG AAACTAA TGCTGCT GCAAGAC CTGAAGT CCTGAAG ACGCCAG CCCAACA CCAAAAT GGTTTTCT AGACAAA ATAA	TAGTCTTC CTTTGTG TATTTACC CTATGTTT TAGCTGTT TCATATAGG CTTATTATC CTTCGGTAT ATCTCAGT TTGATTCGG GCGCTAAT GACCTCAG CAGTTGTGC GCCCTAA GCCCTAAC ACCACAAA AAAAAA	MGDNFAMTSSSSSTLPTQKRKLDSESKKVLPNPKF LVVSRIDGHDFSSVPFIQRGVDAIAGPVKTCSRQNG TLLETEGAFQVKLLRTRLTHFPVKVEPHRTLNSSRG VVFSRLDGLSDEELQEYLSDDQGVTAHVVMKRVNN LTPTQTNVCVLDLHRPKQTVFLTFDKVDLPSKIKAGYEIS VRPYVPNPTRCYQCQRNHTRQSCPNAAKCVTCGRDA HEGDCPPSPRCINCMGDHAAANSRECPVYKDERAIQEUR VKEKYSTSAARKLLASRKPAVLPAVTYSNVLSRPRTKE TGKTQCDFTFSDTVVRSASPKIARSTPMPSAHKSQSITS VSQKSRA TKSEARTSKKDPVREDFLKTSRSSQPSAAHSSN RTASKKAHKRRSSSPPRRDSSVTPSSRRRPSPSVSPG RSTACRSAGISPAAEAA PSDQSRKEAEQMSEPMEHDCPS VDCSSAA SQPGQPQRSPLAFSLFFFS AFEMALLQWNI RGIRSNRDELRLRLHLCHPLIVGLQETKL RPTDHINFPQY TAVREDLPPEVIP AHGGAML LIREDEVYYHP IPIQLTPLQ AVA VRVTISTFTFSLCNITYTPSSS AVLRADLLHLVAQLPT PELLJGFENAHHP LWGSPASCORGSLADLENHNLVCL	cd016 50	987	124 4
--	--	--	--	--	---	--	-------------	-----	----------

ACGGGTG	NTGPTFLSDTTHYSHLDLAICTTPLAHRFEWVYVLSDR
TCCCGTG	YSDHFPCLSLHHTTSPCSTSWNISKADWGLYSSMAVF
TGAGTTT	HDQNFPCDSQVNYLTDVILTA AEHSIPRTSPHRVPVP
AGGTGGT	WWNAACSDAIRARRRLRFQRRPTLENCINYSKRLRAK
CTCATGT	CRRVIKESKACWTAFTSSENSTPSSVWGNLRLRLSGT
ACCAGCT	KVHSPGPGLNANQVLVEPVDICNALGRSFAEVSSSANY
TTTTCTT	DPAFLPRKQAEARTPSTSPNRESYNATFTMQELERAL
TTTCTGG	ALSRSSAPGDAIHQMLKNLSPAGKCFLLRTYNRVWTE
TTTTCTT	GEVTRWREAIUVPIPKPGDKHLPSSYRPISLTSCVCKV
CTCCTAA	MERMVNARLVWILESRHLLTNVQCGRHRSSIDHLVT
CCTTTGCT	LSTFVMNNFLKRQTVAVFFDLEKAYDTCWRAGILRT
TCCGCCT	MHTWGLRGLPLFICAFKDRMFRVRVGSVLSDTFRQE
TTCCCTCC	NGVPQGSVLSVALFAINPIMDCLPADVSGSLFVDDFTI
TTACTGT	YCSAQRTCLLERRLQHCLDSLVSWSVANGFRFSHEKTV
ACTCTTCT	CINFWRYPREFLPSLHLGPVALPFVETTTFLGLTLDRLS
AGCCTTC	WSPHVSSLAARCTRSLNVLRVLSGTSWGADRTVLLHLY
CTCTCCTC	WSTVRSKLDYGSFVYSSARPSILRLNSIHHRGLRLAIGA
CTTTAGT	FYTSPVDSLYAEAAELPLYRRDILLCRYACRLLSMPHH
CTCCGCC	PSYPSFENDSLSHYGLSAIALLPPGVRRRLRLQDFTL
TCCGCGT	PSTFQIGESPVPWLQAQVRVHLELTSLPKESTSDAVYR
TTGAGAC	SRFLELRAWLENSVFYITDGSKTNEGVGCAFVIGDDGFQ
TCAACCT	YRLNDHCSNFTAEFLALYQAVLYIRRHRSPSYVICSDSL
TTCCAC	SAIQSLRDPFSAHPVVHRINRSLQLLAADGPRVHLLWVP
TTCCCTCT	GHAGVPGNEAADAATAAATLLPRTATYCVPSNDIRGAL
TCTTTCTCT	SRRLSSWHNVWSSLTNNKLRTIKPLPVAWRTTSRRSR
TCTTTCTCT	EEVILARLRIGHCRFHRHLLTGDPAPLCSGHPPLTVRH
CITTCTTC	VLLSCPNTFPVRIALNLPDTEELADDARAAAARVLNFIN
CTCCCTG	LTNLSKDI
TGCGTTC	
CTGAAGG	
CCGCCCA	
CGCACTC	
GCTGCGT	
AACAGGT	
GACTGGG	
TAATGCG	
TAAGTCC	
CAGCCCC	
GGGTGG	
CATGTAG	

Daphne	Daphne				GGCCCGC ACGTACC CCCTGGT AAAGGCC AGGCCCA GGGAGGG GTGATTA CCCGAGC TGATACC TTCCGAC CGTGCCG ATTGGTC CCTCCGT CCGTTTCT CGGGAGG TGTGACC TGAGGTG TGGACAA TCACCTA AGGCGGG TGAGCCC TCAGTGA GGGTTC CCATTAG GAAGGAG CGCGCCA TCGGAGA CGCCGGT AATC	TGATGTTC TTATATAGC TAATCAAA ATTGAGAA TATAAGTTT TGTTGCAA CATTGTACT ATATACCTC ATAAGTATT CCCCCTATA AATATTGTG	KTTYIVTTAEHRRAKDQEDLLCVRIYGFSTRKAVTNQ PGDSSNTQVGAkakssvTVQESQCSFRIHQNFGLGNKH DLDLVAFVNKPDVLVTEHWYPKDKISVMHPLMMNL SSYSRMNKSgggVAIfAKSGSNFIVSFYCVIDVSSLSEV VIELSAINYRWERQHIIIVGVYRPPSANIDIFLTNLNQLLV EIDNKYSGKNGCINIITGDINIDTLSNNSTTKRLQLLQE FNLNLLIKEPTREVGGTRTCLDNIITNMDKFTYKTAVLK LAISDHHAVEFQCVPNINLHLVKHKPPPGKMKRYETKR TINDKNIRNLTHVLQSATWAETNCYSYENFASDFFYYFD MTCPTKSVIVNNNEKCRKVPGRKMQKSTWIDAEVTQA RNMLTLFOCKMLENRENTKLKEAFNKYHEYKYNLIGD	cd016 50	581	845
--------	--------	--	--	--	--	---	---	-------------	-----	-----

				CCAGCAT CAGAGGC AACTGCT ATGCCTC AGCACCA CTGACAG CAGCAAC AGGACCA ACTGCAA CACCAAC CCAACAT TATCACC AGTAACA GCAGCTA A	CTATGTATC CCTTGTTT GTTGTAAC ATTGTACTA TATACCAT CCTACCTCA TAGTTCATA AGTATTCCC CTTATAAAT ATTATGCTA TGTATCCCT TGTAACCA ACTGACGT ATTCAGAA AAATGATA CTTGATCA TTTGATGTT GAATAAAG TTTATTAT TTATTTA	KRSKHVAHILNNGNNTAACIWNVINSFRAKGDVNIDL VISYNGSVIECQQDIANIEFESSVGKSINDRLHNNKYA FKGKASSYSIYFTPTNSNEIRNIVKTLRNSKSAGYDGLNI NIVKSCIDSISGPLSTAINNLTSGSPDSLKIAQVPIFKK GDRLEKIENYRPISELPVLSKIEKVYINRIVTFLNKHNIIFE NQNGFIKNKSTVTAQAQYQLHSHFEDEVITAIENKEYVT GIFLDLEKAFDCVNIPIILLDKLWNIGVRGIAHDLJKSYLT NRKQFVVIRTSSGTYSRTVTDMSKYGVPPQSGVMGPLLFLI YVNDIKYCTQNSKVILYADDTISVCKHKELDQLEILSNC LTNEIVQYFNESHLSVSTAKTAVMDFNPSRQTDFFEQLCI GKDIASKENCVKFLGITQDTLKWVHVHVDTLASKLAKNI FAINMLSKHCNDTQVLRATYHALMSSNISYGIEVW/GAT TQANLNLKILKQKKAVRICGAKPKDCRSRSLFPKAGILT VVNMYIYKSIMLIRKLNLSNDTHNNTNRNKKDKFHLL SHRTALYERNPLYAGQRLANTIPSCITSLPTEKFKVTLKR ILITNPHYSLEIFYCFQAQKENFIPDSLTSV			
Jockey	Jockey-9_LM	.	Locusta migratoria	ACACATA GCGGACG GTACGCAC ACTAGGA GGCCATC ACAGCGA AGTTTGG GGCGCTT GATAAGC CCTTCTAT ACATTTC CTTCTGC CGCCGAC AAAACAT TGAAGAT TGTTTTC GTGGCCT GCCCAGT AATTACA GTCTGGA ACACATA	TAGGAGCC ACTAGGA TCACCAAG ACGACGCA AACCGCAA GGACAATC AACAAAGC AGGACACA GAATACCA GCACCCAG TTATTTTGG AGCGTAG CTCACCAGC TTGGACTCC ACCGACCA TCGGAAC TCATATGTT TCCGTCAAT GCACGAGC	MPPLRQMQLTRDQNOQMNRNHTARRQNGGENST TQPLTRSERTEQQRDRTHQAREEDVATAALSSSTVPAT PRENAPRFEDMVTLLQKRLDLIKEPRQITAAVTAQAI STGPVMARSOQGLTAAWNANGIRTOEGEFRLREEA VDVCLVNTELKPGVHVRAANYVCYRTDRTHGGGTA VYVRASLAHHRVQLPPLQAIQESHSGSGRHHHQDHRR SLQTTKDCGTTCTRRNFTVTPGKVFAGDFNAKHHIWN SRVINNHGWHILREAAQHNDAMAIGPNDPTFFPENGQGP DVIMLRSDPPREWGTAGCYVVALJGVYPYVISASTRNV LSSDHVPVVDVDDVADGPAKRRGLPTNSTNWDRYREH LGNLQDAPMVEDAGVEDNVRTFTKNAIQAVKRATPR LQRDPPDRSRQLPPDILAMITAKNRMFREWQLTRNPAT KRAVNRMQREIKAAVKAHRHQEWSNKLATLRVDDGT GWKATQMFRLRRLLKIPLQVGPITVSAPEDKANAIADVF EANFTPVQDPIDPEHVRLVQERLPTFLNAQKSVQKRCPG FYAASINEKLEAPTSSTKCFDRCRQKQRTLQQRHCIM RTGHXPQAWKHAEEVVPVKPGDKDRLPSNRYRPSISLLPAJ SKVFERVYLKRLMDHRRRENILPDEQCQGRPGHATTHQL LRLVEEALDXTERREYFGAVFLDVSRFAFSVWHDGLLF KLDDXRPAAPARPINSVVSPPQDVPGAHWKRPXSXRPNV	cd016 50	621	939

AAACAAC	GNTSGQYSLTFRERSXPSGMTVYSSSCWTSGLPLPHVRL		
AATTGTC	IRSYLQDRTFQVRTEELSTRRPIRAGVPQGSVLPILYILY		
GGACATG	TADMPRIDRVHLALYADDTALFSRSMNARTLGTRLQRA		
GGGTATG	SDELSRWATKWRLSFNGAKTQALIAARKRVVDLPRLS		
GTTTTCG	VRGTAVPWGRTAKYLGVTLDRLTWQQHVDVIGRA		
ACTCCGC	RGRRLDLYPFLNPTSTLPPQLGVAIYTTLIRPIFEYA		
TCCGTGA	WGNAAQTHIKRLQGLQNRILRLALHLPRDFPTEDLLRV		
CTCGGGT	TEIPPIQARFRQAAKLFYQKAARSPNALIRALGHRVQRL		
GAATTCT	NATRWPHLLRSD		
GGACGAA			
CGAGGAA			
AGAGACA			
CCACTGG			
TCTTGGTT			
GTCACCT			
GTGACAT			
ACCTCAT			
AACAGGG			
GTCTCTG			
GCAAGTT			
ACCAGGA			
TCGCTGC			
GACGCCT			
GTACCCA			
CGGAAAA			
ACTCAGA			
TCAAAGC			
GGGTGA			
ACCACAG			
TGCTACA			
ACTGCCA			
AGCTGTG			
GGACACG			
TCGCCCG			
ATTGTGCT			
AATTACC			
TCCGCGC			
TGTGTGA			
AGTGCGC			
GGGAGAA			
GAGCGAGC			
GA			

[illegible]

				CCAGTGA TCTCCCA GCTATGT CAGCTAC CCACCAA AATCCAA ATGGACC TTCTCAG GACCAAC ATAATT TAAGAAT ATAGTCC AGTTC AGTGTT GTCATTTT AACGCCG AGGGTT GAGCAAA TCAAAAT GTGAGTG TCTAAGT AA	TAATATGGT GGCTGACA AGAACCAT ACGATAAT AA	LLSVCYKLLERLIYNRISQILDASIPVEQAGFRPGRNCCD QVLSLTSFIEKGYEMKKKTSAAFDLSGAYDTVWRKGL LLKFYNIIPCPLLLRFLNSTLSNRLIKVYLGHKESAFRQL NDGLPQGSVLAPLLFNVTADIPATQSKKFIYADDIALAT QHTDLPTEATLTNDLATLQEFRTWKLKPNPNKTEVS CFHLNNKLANQKLNLYLQVVKLTNNHPKYLGVILDRS LTFKSHLQKTAAKVKTTRTNILQQIAGSSWGANAQVLR TALSLVYSSAEYCAPVWLSAHVREVDVQLNRAMRIIS GTLLATPLPWLPVLCCHIPPEIRREQALIREYKILSTPQL PITQYLPQNLRLKSRKPLRLASQLIAEDFSPNTTWAAS WDXFEGRNKHLPSPSTKVGGMINLPREWTLLNRFRITG VGRCNVWCKWGMITLDQSCDCGADTQTMEDIAIDCP CVLSLRGWLVTI			
Pene lope	Penel ope- 80_L Mi	Locusta migratori a		GGGCGAA ACGTACG GGGAAAG TTTCATGT ATGGACC ACGGCAT CTCAGCC CGGAAGC TTTAAAGT A	TGACTGAC AGCAGCGC CCTCTGTT TGCCTATAT ATACGACG CTTACTCGC GCTCTCGTT GCAGTACG CCGCTTTAC CTCCGAGG ATGCCTTC GCAGACGA GGGCGAA ATGGACC ACGGCAT CTCAGCC CGGAAGC TTTAAAGT A	MITPAVKAFIYDQTPLRGYYNIGSAVREHAKKSQQLC SLTFLRCRDTSTVPKFLKVKRTFRSAQAHRIYSRMESA MLRERHQTRRALAETNSKLFQLHLLISNTMQCDDWVK IDNLTFRTEASAEIAGNRQKRKFDLSQKNVLVGTMED NSRTVINLSDKVLTADVLAKGGNFVTPRKVPTEI VANIEAGIRLSQSSADVIRMETARILRTCKPPTSLSQA QRKALRDINADENIIVLSADKGNATVVMKSEDIHEKIM DILDPSNYKQLRKDPTSKILKMTNQLVKASSFSDDKKH LCKTEAYSPRLYGLPKVHKQLVPLRPVSAIGSPTHEVA GHLASLLQPYVGGTDSHIKNSCHFDIIEKELSVDPDDILV SFDVVSFTMIPVNEAITCIADIFPSDIVALFKHCLTTTF QYKNVFEYEQIDGVAMGSPSPAVANLFMENFEQALQT ARKRPAKWYRYVDDTFVWTHGEEELDAFLHLNLSNP RIQFTMEKETNGQLNFLDVTVIKREDGSLGHKVVYKDT HTDRLYLQKDSNHHPRQKRRDRLSNERTDHVVIRSIGK THLTVIFKKILTTTHDKKRGVITLVDRAHKICEPAHLD DEIEHLRVTFKKNNGYSREIDRALRPRTTTRTDEERQPP KGVFLFPINKVTDIGIKLSKYDVEIVFRTTKIKEFLR	cd003 04	435	526

					CGGAAGCT TTAAGTAAT GACGACAC CTGCCGTGA AAGCCTTCA TTCTATGA	SAKDKRHLATPGVYKIPCSGKVYGTTRKSVNTRLGE HKRNCRLGHTDKSAVAEHVYQEGDHEIKFNETSIAKT SHYYARLYREAIEIFKHRNNFNKKEEGLKLDKIWMSTL QQQHDNRLISIEEVGNSRDHSTADTT			
					ACCAAGG TGCTGCT GCACAGC AACAAGT ATGCAAG TAAACCT ACAATGA ATAGTCC TTTTTCAG GACCAAC CAAGCTC AGAACAC CCAGCTC ATGGTAA CAACTTA AGAAATAT GTCACCT GAACATG AGTCGAT TTTCATCT GCTAAGA CTCAGAT CTTTCAC GCCTGAT GCTGAAT CGGTGTG CAATTCA GGAACCA ACCAATC TACAGAA AAGAAAT CTTTTGG ATATGTT CTGGCTG	MLVILSHSHVWGYYQEEDRNGAILNEWMDMYNLQLVH NAKDPGTFERSGRWQKDYSPDLCHSKSGNNNTTQRKVIK SFPHSQHRPVIIQFGTEIPLVKSVPPLRWNFQLANWEQFR KSLDNSIQWIPPLPSNYERFLGAIKVNAKRYIPRGFRKEY IPGWSEECDELYAEYQDSHSDSTADDLLRALNKNRREK WHKTTAGLNFTHSSRKAWSLMRKLGTDPRLSNRVAVH NANSVAARLMNISKSKEKAYTRTVKKELHIKRSQLEP HSDYSREFSIEEVETALKMVKNKAPGFDGVYPEFLKA AGPKTKLWLTDFENAILRTGKLPKLFQAIIAILKPGK DGTIDASHYRPVSLLSMTYKLLERLILNRIQETIDQIIPAE QGGFRKHRSQCDQVLTTLTTLVETGFQRLGKSAAVFVDL TAAYGTVWREGLMLKLIQIPQKLAFLVNNMLSDRRF KVFLNGQSSRWRTLNNGLPQGSLAPILFNLVYHDMPTD NSTKIYADDLALVFQSNLDLKHCEQVLTSDLIKLSDFH KWRLQPNPKTEVTAHFLNNKEAQRKLHVTFDGEIVPH NYPVYLGITLDRSLTFKEQCLSKKLASRVNLLRKIA GSNWGADANTLRSALALVYSAGEYCAPVWENSAHIT KIDVQLNEAMRIITGTVRSTPIQWLPVLANIAPQDLRRK AAKVKLIKNVKAHKNSPLYDALQDPPACRLKSRKPPWC DTESTTSFDVTSKWQERWENPPHNGHLVQDPSKKVE GFHLPRQVTSRLNRIQTGGRCAYNNMKWGFPLPSSDC DCGAEETVGHIVKQCPLRAFNGDLESLSHQVTEALDW LSNLDISI	cd016 50	335	588
Ving i	Ving - 15_L Mi			Locusta migratori a					

783

784

785

[illegible]

				TCTGTCG GAAAAGT TGCGGA AGTAAAG TTAAAGG ATGCAA CAAAAAG TGAAAAA GCAAGAA AACAGTG AAACAAA GAAAAGT GTAGTGC ACGAAAA TTACTTCC AGGTICT TACAGAT ACGGTTT GTGATAT CGCTGTG ACAAATG CAAGTAG GCCAAAA CAACATA AAAAACA AAACGAA TCCGTAA AAAACAA TCAAAGT GCAAATA AACAAAA TSTAATA CAAGTGA CAGAAAA G			KLKKAIRCC ⁷ THTE ⁷ FTKTSKESHQMLAKLSPRESCKNKF KEFKILTVPSLYYMRQDTNCTKS ⁷ VYYETIVFLMVNCT APLNSDIHNHNTRTKEQYHFFSKRNLNMTRDPINAGCRF FNKLPLSIRCIKERSAFKRKLKEHLISGCWYKVKEYLTE NL			
RTE- 7_LM i			Locusta migratori a	CAGGGAC ATGCATC TTGCGTG ACGTGGT GGCGAGC	TGACAAAG AGAGGATC AACCCATG ACAAAGACC TGCCGCAA	MDRNKLDVLMCEVRWGGCGELES ⁵⁰ GSDCTTXVKIK MECMELEYLRLLYYSEXEDKNMGYGVLIGRKRLKDKVM KVEYINGRIMMIRLKGKKEDLVIIQQVYMPTSNHSEEEVE ECYEKIENLAEREKRNACIVVMGDWNAVVGEGSEGKA LGKFGI ⁵⁰ LGERNERGORLLEFCNENLMVVGNITLFDNHKR	cd016 50	523	779	

ATGGGGA	GGCAGAAAT	RRYTWSIRFDNKRHQIDYIMIQQRYRNCLKSAKSYPGA	
CCCAAGC	ACAATTGAT	DIYSDHILVVAEVLVKLRITWKQKREFINIEKLDKGI	
TGGCCTG	GATGATGA	CEKLENAFSEIIMPQEMECIERRWDHYKNGLKRQLRK	
GCAATGC	T	RKWVKKAAKETLGVKERKRVKKEWITPEMISKMDRR	
TTCCCCCTT		KWKTVNTEEGKRNRYRRLNNELRRETRKARVWLTEKC	
ACCTGTG		DKIENLZKEGYDLMYKEAMDLVFREKNSNGIKEIEA	
CCAGGTC		ADGKMVNETQEVMRRWEEYIEWLYDGDTRPSEDELGL	
CCTCCTG		EEENTVAEDDKGNAILTSEVEASVKEMKNGKAEGIDEIP	
TTATCTTT		VELLSLQREGKKELIGLYNQIYMTGNWPKDFCESILPI	
CCTATCG		EKKKNSKKCEEHRTVSLISHTAKVLLRTLNRRLYGKLN	
ACCACCC		AVMGEEQFGRRGVGTRDAIGLLRVIGERYMEKKRKV	
TCGGTCA		YAIFIDLEKAFDRVRWDKLMAILRKHGVDWRDRKLLTN	
ACTCTTG		LYLNQTARVRIGGVISEEIELGRGVRCQCCCLPTLFNLY	
TTCATTTC		EDIKESFNGSRGVCIGRRRVECIRFADDMVVMASAKE	
CGACCCC		MEQMLDDLNTKLEEYGMKINKKTKSMVIGGRNTKCR	
GACGGTA		LKIGGEEIEQVHSFKYLGSMIMDDMYCTAEIKKRIAMA	
TTAAGCT		KEGFRKQRLLCGPLKKDLRKLAKCYIWSVVLYGAE	
ATGGAGT		WTLRKEDERRLEALEMWIWRMEQVKWEDIRNEEVL	
CTTCATT		RRVGEERNMLEVIRKRNWIGHCLRRDCLLKEGIEGM	
TCCTCTCT		VKGKRGKGPYQMLDNIQGNKYEEMKRLALDREKW	
CAGTGCA		KRINP	
AGGTAAC			
AGGAAAC			
CACCACT			
GACATTA			
TTTTCTCT			
AGACGCT			
CATGGGG			
TGACCTC			
TTTGTA			
AAGATTT			
GGAGGTA			
AAGCTTC			
CCCTCAA			
TCGGATC			
TCCGGGA			
GGGGAAT			
ACAACGA			
GGTTTCC			
AGACATA			

Ving	Vingi	Locusta migratoria	AGGAGAG ATGGAAC AGGAAGG AAGGAGA AGGAAAA AGAAAGAA TTATTGA GGATAGG AACGTGG AATGTTT TGACCCCT TTTGCAG GAAGGGA AATTAGA AAACGCT AAACATG AA	<p>MRMVSNKRSFSEPALKMTSVNIEGLSANKELLSDICSK CKCDVLILQETHRGPTSHRPNIKGMILTLERPHEQYGS FTKPGLMVRSAAAMTHVDDIEILTVELKACTVTSLYKPPG ADFTREPSNFQIQPLKFITGDFNSHSTAWGYQETDSNG EKVERWAESSGLKLVHDPKLPASFNSGRWRRGYNPDNI FVTEQIAQQSFKSVEDALPRTQHRPISLTVKAVVCPVTV PEKRRFNFKKARWRLFTEELDSKIDEILPAACSYDKFQIL IRRVSRYIPRGCGRNEYIAGLDDNIEILLKYEELYNTDP FSEDLTETGEDLLRALTTTRREKWCDLVTSLDFKQNSR QAWSLLRNLSNDPTSSKTLHTSVTPNQIAHQLLNGKTT HRSPRPSIVRDMENEHHSLEAPFTLEELQA AIDKMKNK AAGEDDLRVEQIKHFQPKTLRWLTHLMKQLRQHHADS QSVESKSGSPIKTRKGANRCQKLSPSILAVPSVQGP NGANNCVNTMQIPRVWRKARVVALPKGKEPTDVKNF RPVSLCHLFLKALERMVLRNISHVDEVLSPEQAGFRPG RSCTGQILNLQHIEDGYERKQV/TGVAFVDLSAA YDTIN HSRLVHKVYQATKDFGLSRFVQCILQNRFRFVYTLQGKS SRWRIOQMVLRLRAVYGLWTFKICAVHTEQAFLCHTAR KKQOMENPKNGLAQGSVLA PVLFNITNDQPMHPDVR QFIYADDTAVAAQGSFTEEVEEKLTRALELLGNYHQ HLKPNPAKTQVCSFHLRNKEAKTKLHVSWKQGQLEHC YAPKYLGVKLDRTLSTFKQCHLDTKKKVSTRNSIIRKLTN TNWGDAPVOVLRTTAMALSLSAEYAAPVWKNASHVK cd016 50</p>	524	819
------	-------	--------------------	--	---	-----	-----

[illegible]

[illegible]

					GAAATTC TAACAGT GGAAACT GACAATG CACAATA ACAAGTG CGTACAA GCCCCTA GCCAAAA ATTCATA TTAATGA ACACATA ATTTTAG AACAAAA TAACAAC TGCTTCA TCCTGGG TGATTTT AACAGCC ATAGCAC TACATGG GGATATG AAGAACT AACACCA ATGGAAG TCTTGTA G	TAAATGAA CAAGTCA AGCAATAC ATACAGCTT TTTTAAAA CTTTGTGGA CCTATTATA ACAAAGT CGTTATT TAAATTTC ACCTCTGTC TTAATGTC AAATGATT	MIWDTVKSELDEQLQVEFLNLKSRQDQKLKLMRQIDC RVNNAQNVDNFQYPRVVNETDISFNDEELRLLNKGLK FNVQSLNDKTVKELIVATKMAVDVAKLNENEKGHV YKLANAIAETNKCSCNSMHNNSDHTVTSINDKLISNK AIVTKADKGN'TTVIVSEEQYINKTLQFFSENGVTEIAKDP TKKVQAELKQLLKRTSFLTEKEKFQCYVMNPTAPRLR SQLKVHKNGHPIRAIVNSMKGPLCKITTKLKDILKEFYV FGDNFSVSNNTTLINEIKDKPIADTSHLMSLDIKNLVYANV PVDEAVELIKANLLKHRLSQQEIVEFIDLLQFILKYNFY AFNGKFIQKDGGLAMGSSISGTHADIFINNLENRLFTNMQ HPVDKILYYKRYVDDTLLLDGGQTETEIVNKNFNSLHE KIEFTVEQEKDGTLLQFLDLNITKQDGRHLFKVYRKPTTT DILJHNSSCHPNPGOKLSALRCLVDRAFKVPMNEKDRITE	cd003 04	353 448	
--	--	--	--	--	---	---	---	-------------	------------	--

[illegible]

[illegible]

TTTTTGGC	GGTCTCCTG
GAAGTCA	CCGAAATA
GATAAGC	TCGGCCGTC
TTCGTAG	GTCGATTTT
CAGCAGG	GGTACCTG
GTCAAGT	GCAGCATA
TGCAATG	CCCGTAAAG
TGCATGT	TTTTCGAAG
CGTCCTT	TTGTT
GTGTGAA	
TGTCGTG	
ATAATAG	
TGTGTGTC	
CAACTTT	
TGCCAAA	
GTGTGTC	
ACATAATC	
AATTCTC	
CTGCCGC	
TGGTGC	
ATTAAC	
AACGTGC	
AGGCTTA	
GCTCTTG	
TCCGCGA	
GAGGATT	
CATTACA	
CCCGCCG	
AGAATTG	
GATTTG	
TGTCCAA	
GGAATTA	
TTTCGGC	
TTCAATTG	
TACATTG	
CCTCTGA	
AATTTCG	
GAATTTT	
CTTGGGA	
CTGGTIG	

GACGGTGT
CTTCTGT
GTCACAA
TCCGATT
GGGAATT
TCAAAAT
GTAAACGG
CCCGGCA
TTCGGCG
AAGTTTG
ACAGGTT
CCGTAGC
TCAGGGT
CTGCTTTT
ACAATC
GACGTTC
TGTGATT
AATCTCA
CAGACAA
AACCTTT
GACGACG
CAGCTCT
ATCCGTT
CTGGGGA
CAGGTTT
GAATTTT
GCACCGA
CGCTAA
ACGTTTA
CCTGTGT
TTGATTTT
ATTAGTT
CTATCGA
ACAGGCT
GTTTTTA
AACTACC
TTCTGAC
ATTAAAG
AGGAGGT
TAGGAGG

Daphnia	Locusta migratoria	GenBank	Accession	Gene	Protein	cd016	503	794
GAAGCTT								
GTCGTGT								
GTTGACT								
AGGGCAC								
GTCCGCC								
TGTGAGT								
AAATATA								
CAGCAGC								
TGAGGCA								
GC'TGCCT								
TACGTTC								
ACTTAGG								
TGCATGA				TAGAAATG	MLFHQN'VQSL'NNKLSSEIKIMLSNELIDVSVLCMSEHWL			
ACAGTGT				TAGGTCTAT	RDDALQTVSILNYELAA'YTCRQQILGGGICIFVKKGLKK			
AATACTT				GATAAAGT	SFFFLFFFFFEISAIELPSLRDNFNMMHLQIPNGDIKTFY			
TAAAGAG				GTGTAGAG	QNLENTLTQTVKNKQ'THICGDFNIDFKGSAKEKTDLMINI			
AATATTT				AAAATGTA	LVSHTLKITIHSP'TREIVT'LSLQVYNRSNNNKPTKEQYKA			
TGGACC				AATTTACTA	EAVKTGLGDHGGQTISFHIK'GIEVNKL'TSTYKTRTYNDQ			
TAAATTC				TTAGTTATT	EIEIFKNLLKSESWEQIYD'TS'ENISEKFDIFIEKIKIYYFDV			
ATTTACA				ACGATGTAT	VFPVKTRRLKDTLPK'KIPSQNKWITTGIRISCANKRFLYR			
GTAAGTG				TTATTTATA	LTKTCDTTPAFDRYF'KNYKTVLRKVIKSAKKLENDRIYS			
TCTAGCA				TAAAAAAT	NSSNKMKAVWNVK'KETGTGNTKKNKNIQLKENQQ'TTS			
CAAGATG				GTAATCATG	KPKEVANMFTYF'INIAEKLIEKNVIKNP'SQLATKIKHID			
GACTGCA				CTGTGTAAA	KTIFLTPNTKEVQ'QEIKLLSKSHSSGIDNIPDFIVKKCAD			
TCTAAAT				AGTGGCAA	FLSEPLTYLINLSF'STGSFPKCLKIAKV'KPLHKKGSEEV			
AGGCTAG				ATATCCATT	SN'YRPVSQLSAFSK'JLEKLFIKDCPISQNPNNMVLKNNHST			
GCACAAA				TTATTC AAT	ETA'VFSFLYETLQALDRREHT'ISIFLDLSKAFDTINHNKL			
TACTTGT				ATGTAGAG	LSKIQNW'GLGAAHXXIK'SNPHDRKQFQKAVLIEITYESN			
AGCAAAAG				TAACAAAA	GT'VSTYHSDSLVVKHGV'PQGSILGPILFLLYVNDLYNN			
TATTGGA				ATTCTAGCA	TNGKLYQFADDT'SILITGSTKEIFESNIAATKNIRDYFDA			
GTGTGTG				GTGAAAAA	NGLIINVEKPLQW'YLRHRIDHQLTQILJCKWPDNCKCGK			
GATGTAT				CATTATGTG	TVAMDIFTTQNRSPAN'PNIDINGKTIETKTECTKFLGIWIQ			
GTATATG				GTAATTCGT	QNIKWDSHVDYINR'KL'SKTCYIMRTLKNLSYNNLKSII			
ATAAAAA				ACCAAGTA	FAY'AHSQMKYGI'VFWNSGASMKTFRIQKKIIRMEGV			
CAAGGGA				AAATGTTA	DQRTSCKPLLKRNT'PTSKHVHNGKPNYYKRIKIKKEILPL			
AATAAT				ATCTACTTT	PSMYIMENLIIEK'LN'SKSPCTTRNETVHCHNTRQKNV			
GATAGG				TAATAATTAC	HVKQ'TNTALFQK'GV'LHPGIRLYNALPKAIKSKTDITTKF			
CAGGGGG				AAACAAAG	KTAVKDYLV'MHCFYSVDEYLEFONYYSIS			
TGAA AAT				AAAAGAAA				
CCAGACA				TATGTTAGG				
AGAAATAC								

[illegible]

[illegible]

[illegible]

TGCATGA	CKIINQSL E E G Y P P D S L K Y A E V K P L F K K G S K E D M G N Y R P			
AAGATGT	ISSFFCPKYLKYQKQFGQK GKSTIYAINEFVNKVSTSL			
GCTAAAG	DKSQKAAGIFCDLSKAFDSVNHSLLLYKMEKYGIMGTA			
TGAACCT	LQWCTSYLTERKQRVVITSSKGKYSKWKRLSHRGYHKA			
TAAGTTC	QFVETIPQGV PQGSILGPILFLFYINDLPININYHSVLFAD			
GTAATG	DTSVLIENKKADEIPDNVLKTLENLDVWFHLNGLKLNK			
ACGAGTT	KLRCLDPRKPRCLVPSKWSKIKYKKTQVMQFKTKQSK			
KCCATGG	LNDIQIKHNSQELHEADHVKFLGMKLDRIYPGPHTILGN			
TCGTGTG	ETGQNLSSSHIDYLNKLNLSLAFAMSLANVTSLDTRK			
GCGAATG	IVYRSYFESIVRYGIFWGNASTMIRILTLQKKIETCGAK			
TTTATGT	PRESCRPLLRDLNLSVPSLYLFEILSHIQNYFFIITSTVM			
GATGAAC	KQEQRFHAPTHRLNLYAQAPQYMGMKICNKLKNTHF			
GGAATGA	RKVETESMKRKLSTLVQKCYYSIDEFLNDDLSIEQDKT			
ACTGTTA	YNYVL			
AACACAA				
TAAACTC				
GAAAAAT				
GAAATCA				
TCAAGAT				
ACTAAAT				
AATGACG				
TTGAAGA				
ACTTAAA				
AAGGTAA				
TAGTTGC				
GTTGAAA				
GAAGAAA				
ATGCGAA				
ACTTGCA				
AGTGATG				
GAAATAC				
ACAAAAG				
TGCGAAA				
AGAAATG				
ATCACTG				
TGTGGAA				
AAAACCA				
AGAAAAT				
CCAAGTG				
TTTCTAAT				

803

[illegible]

ATCTAATCT
TA

RTE	RTE- 26_L Mi	.	Locusta migratori a	GTGGAGT TGCTAGT CTCCCAT CGGGCAC CTCCCCA GGTGCG GATAGGG GAATGCT CACCAGA TATGGTG GGTACCG GGGAAAT AAAATAC CCGGGGT GGACCAA AACTAGC ACCTACT GCCTTGT AGTTAGG TATTGGA TTACCAA AGGCTAA AGGAAGA AAACCTT GATCAAA AATTACC TGGCCCT CCAGGTT GGGGGTT AATGCA	MGQLPIHWKKTLPVKKPIRLQQDGYNRRRRRLRKRTVL IGVWVQGLQNKVGEVIKEMEARNQDIVL/PETKKKG HGIERIGPYTHFYSGVPKNQRAKRGISILMKNKFNRMVT NWEPIDENLKNINIQEHKTLIGVYGEISEDEPIRKKDEFY NRLNGVIEDIGKTREIAGDFNSRTGSKIGNKIIGPFGEN VVNDNGNRLDICEQNSLKILNGFYQHKDIIHKYTWHKD TLNQSIIDYIIKQQTRIKIHDIRAYRGVTLSDHYLVNSK FLPFRGKNSEENQNCITTEDLQTEIKITKYNLNSFENEST VFLYQKRLDEKLIETEFNTEDQYQYIISKLHESAKEAL GETDEKHTXHSLYYWDEEIKKEVKLKEKYLKWLNTR NVQDKIELNAQQAIRKKVAEAKNAWEQSCQKIESYI GGKRNTEAWKTIKLNRRNSKEXTHIKYTHDSWEKYFQ NLLSENRRILGKPEGILKXQHILNLLMTLGNIFKIYYQ KTEFLGNPEYNEITNELPDNYIHLEIETVKSIAKSLKNG RATGIGGIPAECLKFGTEKLIJLLRNLFERCLNGEDVPND LKVGYSIVHKKGRKDDCKNYRGITVTNTFSRLYGRIK YFLEQYKEKEAEQAGFRAGRSTIDHIFCLQQHIEKNKK LGRAQPIHLVFDIEKAYDSVPLPNLWKALLSIGINPTIHK AIQNLKNSIKIKIGKFLSSGFQVTKGLRQGCMSPTLY KIYTEVTLQNWKKKQCMGIPINDRTIYSLQFADDQLIA QDYEDIEYMTKRLKIQEYKKAGLNVNMNKTVMVIGGI NEDLILEEGMGTTTTTEEYKYLGVKITNDGKQKEIRSRI NAGKTTISLLNGILWDKHITTDNKIRIYKTIVRSIITYGSE VWTTKWQLKSKLLATEMDFWRRSARISRRERIRNEVIR EKMCKKNSHIEFIEHKHLKWYGHVRRMGQERLPKCILD WTPNGRRKKGRPPTTWMQEVQSAMRKNKNIPEDLWEDR EEWRTIIXNLM	cd016 50	592	851
RTE	RTE- 38_L Mi	.	Locusta migratori a	GCCATAG CCGGTCC CAAGCCC GGTTGAG AAAGGAG GAGGGGG AGGAGTA ACACCTC GTTAAAA AACCTTG	MSTQFLKPVNKLTTWKS PNSTLGEFQIDHVAISNKNKTE ILNVTRKGFESDHHLLQIKIKFPNRRKKPQMKNKTRID PEYLKLNKKEISQIRQSNATNWKELAEVLKSTMHLGK PPRRKRRWWNSSCDQAEIERTQAWKTFSSNKTTEKW QEFRLIQKNTSKAIRREKRRYDINRLKEIEQEFKNNTRN FYTFKENLTGYQPNLCFRRPDGSLTETNRNQCILANI FTHSSTASRRQSNISKTSPNPSKPPSLKEVKEIINSLNK SPGEDGHAEMWKLNDSTLTQATHRIISKVWETEQIPAE WKALHPLHKKVQKLIQITTFGEFHCSQLQGNFTAPSN QNFKGPVKQTRRTGGPVPVNRVSGRVSRSLRRTNMIEP	SSF5 6672	393	550

GTTCACC	KDILQMRRRCRNTVVTVDFFKKAYDSIDRQTLFHTLEEFR			
TGGCCCT	IDRKTRSHQQALTGTTTSKVKFMGETSEAFHHHTGVRQG			
GGTGATA	DGLSPLLFNLVLEKIVRTWEKEIKGIGLGIKKEQIRV/KC			
GCACCTC	LAFADDLAILTNNREEARL/ALEKLEHISQKTGLQISYEKT			
GTGAGGG	QYMENKLQIVSPSPSMEKWHKFLTSDIKDSVHGKQAA			
CCCCTGC	DSLPTTTQYGVKVAQVSHFRYLGEMIQPSGLNSIANQERV			
CAGGGTA	SRLQRAYKLTWNHYNKKSISVNAKLRHYNVT/VLPEAL			
CGGTGAA	YAAETTVIHGHTKIREIEKQERKILRKIYGSVFREGIWK			
GACCTCA	RPTREIYKDMETTTDITIRKRRMKFYGHISRMNETRLTRQI			
ACGGCAG	FEJANKSKNKT/KWTETEEDIKELGITQDNNNNREQFRNII			
TGAAGGC	KKHKLKQEHNTENRTGKKWSEKREHSEKRMKRIWEERK			
GGAGATG	KKKS			
AAGATCA				
TCGGTAC				
GGCGGAA				
CTGGCGG				
AAGAGGC				
AGCCCTA				
ATCCACT				
AGCGTCT				
GCCCTGG				
AATCCAG				
CGGTTAG				
TGGTCAG				
CGTCTGT				
TCACCTA				
GTTGGTG				
CGGCTGC				
TCCTAGT				
TCCTCTTG				
ATCTCAA				
CAATCAA				
GTCGTAA				
ACATGTG				
ACCTTCG				
TAGAAGA				
TCACCAA				
AACAAGT				
TTCAACT				
TGAATAA				

GCATGAT
GGAACAA
AAGGAAA
GAATTC
ACTACC
CAGGCAT
CAGTCT
TAGACTG
AATTC
CTGGTCA
CTGGATC
TGGGGA
CCAGGAA
CTTCATG
CGCAGAA
GAGTCGG
AGCTTCT
CAAAATC
TCTAAGA
CCAAAC
AGAAAA
CTTCATA
GGAATCT
TCAACAC
AAACACA
CTCTCAA
AAGTAGG
TAAACTT
AAACAAC
TCACAAA
CACTGTG
GAAAAAT
TCCACAT
AAAAATC
CTTGCAG
TTCAAGA
AACACGA
TTTACCG
ACGAAA
TCAATTG

TACAGGA
AACTTCA
GAATCTA
CAAAGTA
AAACCTG
MAATCCA
GTTGAAC
AAAGGAC
CCAACAA
ATTGGC
ACAGCTT
TTGTTGT
GCATAAA
TCAATTA
CAGACTC
AATAATA
AATTCCT
ATTCAAC
ATCAGAA
AGAAATAT
CCACAAT
CTCATTA
AAATCTG
AAACAA
GGCATAC
ACACTGA
TAAATGC
ACACGCT
CCCACAA
ATGATTA
CAATAGA
AAAGACC
CACAGGT
AATAGAG
GACTTCT
GGGAACT
ACTGGAG
GAAACTT
CAAAATA
AATTCCT

[illegible]

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

TAGCCTT	CTTGCGGA	NNRIRKLLRKQRELHCKHKHSQSLVDKQKLNEAKCSVR
GTCTGTG	GTATGGTGT	RAMREAFNEFESKVLCTDLAENPKKFWSYVKA VGGSK
GCTGGTT	AGATGTAG	QNXQTLCDENGHETEDDITLKAELLN VFFQSCFTEEDCTV
TTCAAAA	AGATG	VPSLDCRTDYKMAIDIEIDRGIEKQLSKLRGKAAGPDG
CAAGTCA		IPVRFYTEYAKELAPLLAVVYRRSLEERSVPKDWKMAQ
GAGACCT		VIPVFKKGRRTDVQNYRPISLTSVSCRILEHVLCSIMTF
TACTACA		LETRNLLCRNQHGFRKRRSCETQLALFVHETQRAIDTGS
GTCTGTA		QVDAVFLDFRKAFDTVPHRRLMNKVRA YGLSDQLCDW
TCAGTCA		IEEFLDNRMQHVILNGEKSEVRVMMSGVPQGSGVGPLLF
GTCCAGT		TIYINDLADNIGSSLRLFADDAVVYREVVTMENCIAEQE
GCCAGCT		DLHRIDAWCREWQLNLNVDKCNVLRHHRKKDPLSFSYT
GTTGCCA		IAGQQLAVNSINYLGVGIRSDLKWN DHIKSIVGKADAR
GTGTCAC		LRFIGRILRKCSPKTKEVGYSTLVRPQLEYCSPVWDpYQ
ATCGTAA		VGLIEEVEKIQRAARFVTGSFGNRESVTEMIDKLQWKT
ACAGCGC		LQERRSVARYGLLSKFREHTFTTESSILLPPTYISRDRHE
TAAGTAT		DKIREIRAHTEA YRQSFPPRTIRDWNRRENIR
CTAAATT		
AATTGT		
GTCAAAT		
TTAAATT		
GCCAGTC		
TAGTTCIT		
CAITTC		
GITAGTA		
CTTCIGTT		
TTAAGAG		
GTTTATTC		
TCGCGTA		
CTTGCA		
GATTTCG		
TTTIGGCT		
ACTGTTTT		
AATAGTT		
TAITTC		
GCGTATT		
TGATAAG		
TACIGIT		
GGCGCG		
TCAGCAG		
CACACAG		

TCAGCAG	CACCAGC	AGAGTTA	CAGCTAA	GCAGGAG	CGAACTA	TGTGAC	TTATTGT	GGGCTGA	GCAGTGT	CATATTC	TATAAAC	TCGCGCT	AGTAAAC	AGATACT	TAACTCT	CTACAGG	CAATTGT	TGATTGT	TTAGGAT	GGATAGG	ATGTGTG	ACTGCTG	CGTACGG	ACGACGG	AGGAGCT	GGCGGCT	GTTGCGG	AGCAGCT	GAGCTTG	TGTTTGG	TCGTGCG	GCAGTGG	GGGCTCT	GGTGGCT
---------	---------	---------	---------	---------	---------	--------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

CGCATGG	GACACCA	CAGGTGT	TACATGC	TTCACCC	ACGGTCC	CTGCTGT	CGGGACA	TCTTCGC	GCGTGCC	GGACGCC	GTGAGCC	CACCCCTC	TCCTCAA	GGGGTGT	GGCGGGT	TCAACGG	CTGCTGC	GGTCCCTC	GGGGCGA	AAGGTCA	ATGTGGA	GGCTGGC	CGTCGGG	CGTCGCC	CGCTCTG	CCTGTGA	GTGGACA	TGTGGCC	GCCTCTT	CAGCATT	CAGCAGG	GTCGCG	CAGGCAC	ACGGGGG	GAGGGGT	WTATTAG	TGATTGG	GAGCTCC	AATGTTA	GGCGGGT
---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	----------	---------	---------	---------	---------	---------	----------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	--------	---------	---------	---------	---------	---------	---------	---------	---------

GATGGAG CCCCTTA GGGAAAT AGCGGAA AGTTCGG GGAAGAA GGCCAGT GTGCACT CTGTTTG CCGGGGG GTCTCAT CCGAGAT GTGGAGG AGGCCCT TCCGGCG GCGACAG AGGACAC TGGGTGC ACCGCAC TGCAAAT TGTGCT CATGTCG GCACCAA TGACTCC TGCCGCG TGGGTTC AGAGGTG TCCTCAG TTCATAC AGCGGTT TGGCGGA GTTGGTG AAGGCTG AAAGCGT CGCTCGC GGGGTGG AATCTGA GCTAACT ATTGTA GTGTCGT

TCCACGA
ACCGATC
GCGGTCC
TCTGGTT
AGGACCG
AGTGGAA
GGCTTAA
ACCAGAG
GCTCAGA
CGATTCT
TCTAGGC
TCTAGGA
TGCAAAAT
TTCCTGA
CCTCCGC
TATCGGG
TGGAGAA
ATGTAGG
GTTCCCC
TGAATAG
GTCAGGC
GTGCACT
ACACGCA
GGAAGCG
GCTACAA
GGGTAGC
GGAGTAC
GTGTGGA
GTGCACA
TGGGGGT
TTTTAG
GTATAG
AATTCCT
TCCCTAG
GCCCGAC
AAGACGC
CTCTGA
GCCCGA
CACGGTA
GCAGTAG

GCAGAAAC
ACAAACAG
GGAATGA
AAAGATC
AATGTAG
TAGTAGT
TAACTGC
AGGAGCG
TCTATAG
AAAGGTC
CAGGAAC
TGCTCTC
ATTAATA
AACGGTC
ACAATGC
CCACATA
GTACTAG
GGACGGA
AAGTTGG
CTGAAAC
CAGATGT
AAACAGT
AATGAGA
TTCTAAA
CTCAGC
TGGAAATG
TWTACCG
CAGAGAT
AGGCTGS
ACAGTGA
AGGGGGA
GGCGTGT
TTATAGC
GATAAAA
AGTGCAC
TAGTATC
GAAAGGA
ATAGACG
GAGATCC
GAAGTGT

[illegible]

TCATACA
GSTCATK
GCTTCCA
TCTAATC
ACTGGAA
AGAAGAT
GTGAGTA
ATGAATC
ATAAATC
TAGCCAT
GAGTTAT
TATACAT
TATAACG
TAGGTCA
GATCAAT
ACCACTT
AGATTTA
AGGAAA
TTCTTTTT
TAGAGGT
AGAAAAA
GGGCTGA
GCCAACT
GGTTGTC
CCAGATT
CACACTG
CAAGACC
AGATTAA
AGATAAG
AATAATC
TAATAAA
TGTTTGT
AATATTA
GAGAGAA
ATATAGC
AGTAATG
ATGATTG
CCCTCAT
ACAACA
ATTATTC

ATCAACT
TTGTCAIT
TAAGATA
GGCTATC
TCAATAT
TGAAGGT
ATGATGT
GAAAAACA
CGCAATT
TTAAACG
ATTTGTC
AAATGAG
AACTTAT
TTGAITC
ATTATGC
TTAAGTG
AGTATGG
GTGTAAA
GGAATATG
AAATATG
TTTTCAITG
TATTAGA
TGATTTC
CATTTAG
TTTAAACAG
TTTTAGC
AGAGAGC
AGCATAT
TCCTGGG
GGGTGTG
TCAATT
ACATTAA
GAAAGCA
TTAATAA
ATTGTAG
TAGTGAA
CTAGATC
TAGGTT
CTGTGC
AGTGAGA

TACATTTT
GAAGCCA
CAGGTGT
ATGTAIG
GATMAAT
TAAAAAT
GTATTGT
AGCTCTT
TATAGAT
CAGCTAG
TGGTGAT
CCACTAA
TATTTTAA
AGAAAAA
CTTGAGA
ATTATTC
AATTCCT
TCTCAGA
CCTCAGT
GGAAGAA
ATATAGT
TAGTGAT
TGGTGGT
GACATCA
AGCTGTA
TTTGAIG
TAAATAA
AGATGCA
CACACTG
TAAATAT
TAAAAAC
CTGTACG
ACAAATC
AACTTCA
ATTTTAC
AAACTGC
AACAGTC
TGCTTGT
CTTGACA
ACTACTT

[illegible]

CITCAAC	TGCCTTCCT	RALQRHRRHPSLEHLIAFKRLRAQARRLIQQKRDWSER
GACCATC	GTA CTTCTC	YAATIGSRASPTZVWKQVRRLLFGHRPLAGIPSLHIRGTL
GTTCGGC	GTGTACTGG	TDPAAIQYMAIDIMLVARHLHLTVRTSAPSKSVLNSGS
GCGACGG	CTGTGGAGT	FFFTCAARKNTMRHSVSGSLRVPAPFRASSTGRNPQSSH
TGGAATG	GCTTCGTCT	SGYSVHRPSCYRAVHGRHYARSSASASYSADFALKER
TTGTACG	TCAACTGTA	VEQRQLSFHMRREQYNAPSEWEFESALAACPDTSFG
TTTCAGA	AATCACAT	PDSVHSQFVRHLPAXFQQLLLALYNHLWVKGAFFSQW
GCCCGGG	GTGTGATCT	RDSIIPVMKPGADPQVAESYRPSLTNVLCKIFEHMSR
GATCTTG	CTTTGCTGG	RLMWILEDGRLLAPQQGGFRHGRSAIDNLVSLESAIRTA
GCTTAAT	GAACAGGG	FTRREHLVAVFFDLRKA YDTTWRHHILATLYDWGLRGP
CGCCAG	GACTGATA	LPTFIKNFLSLRTRVRVGTSLSTHHIQENGVPQGSVLSV
GTCGCGA	ACCCAGCC	PLFLVAINGLAAA VGSVCSLFADDFCLYYSSASIGVTE
GGACAGT	GTTTAGTCC	RRLQGAIRKAQSWASAHGFTFSASKTRVMHFCRRVVHP
ASAAACC	CCTTTAAAA	DPELYLADEQLHVAETHRFLGMLFDNRLTWGPHLRQL
TCTATTA	ACAACAAC	KQKCWRHLNLRCLGHTAWGAGRQTLMHFLKALILSR
AAAAAAA	A	LDYGSQVYDSASPSALRILDPILHGGVRLAMGAFRTSPV
CCTCAAT		PSLLAEAGIPPLRFRRQQLASVVAHVHATRLTIQIAGFC
CTCAAGG		FQTPSPFLHNIDPDQVFQSRSGYVLSFPHSSPYLYLPHSP
TGTGCTA		ANLISLLGPSLAYSSTCNCPAIQRFRHHHFINFWTSLA
CGTGCCG		RSVMQKSYTLMA PRSWHPSPVSPATVVGFCRPRSCYT
AGGAGAT		PHHPNRRLLFPNTVLPSPQHPRSGLPVAVRVRTLSSLE
GCATGGC		SLPLPPAFSCQSHIPWSPRLQLHLQLSCDPKASVSPHTL
TGTTGAG		HQQLDILGTFRDAEVIYTDGSQVEGRVGYAYVHGDIH
GTTGAAC		EQCSLLPGCTVFTAELIAIRALDHLRTCPKSYVICSDSL
AGTCGCT		SGLQALDQCCKHHLVRSIQDSVFSGLXSGRSVAFIWT
AGGAGAA		GHIGIPGNERADQLAKQAIASQRLLDGTPERDIRLGLHHK
CTCTGAG		ISMAWNSEWSTLSTPNKLCALKPSTA AAWKTSLRPSRRDS
GAACCTG		VVLCCLRIGHTWLTGHLLHREDPPNCRGALLTVVHI
CCGCACC		LLDCPLLAALRQSFNLPALPSLLADDSMVDRLVGFIR
TCAGTTG		AAGFYHMFNGGYLTFRPPPLNSFNLLFLSCQSYQLAA
TATAAGG		LVCTG
CTTACCC		
AGGCAAG		
CGGGGCT		
CTGCCTG		
GGTGGAC		
CITTATAT		
CCCTAGC		
TGCTCGT		
GGGACCA		

AGATGAA
ACGAGCT
ATAAACT
ATGCCAA
TGGCTGC
TTCGAAT
GAGGAAG
ATCAAAT
TTCAATG
CCACGAA
ATATGAC
CCCCGAT
CGTTCCC
TTCCCTTG
GTACATC
ACGGGAG
GAAACATC
GGGCTAA
AGATAGG
GATGAGT
TATACTC
TCCAAAA
TATCTCA
TATGCAG
CAGAAAC
CAGCGGG
AMAGTTT
CATGACG
ACAAAAC
CTCTTTTC
TTCGTAT
AAAAATC
GGAGGAT
AAGTATG
GTGAAGT
CGCGGGA
ATGTCAA
AGATGAG
GAGTGAT
TCAGTGC

TGATAGA
AGTGCA
TCATCCA
CTCAGTC
GAAAGCA
TTGCTTG
CTTGCGA
CAAGCTG
GGTGATG
TCCCTGT
AAGCATA
ACACCCC
ACAAATC
ACTGAAC
ATGGTCC
AGGTCT
GATTAC
CACAGAG
ACCTTCTT
CTACAGT
CTGATGA
CGAGTTA
CGCACCA
ACTTAGA
GCGACGG
GGCATCC
ACTTCGT
CCGGCGC
GTGTATA
GGGTCT
GAAGGAT
TGTAGGG
TTGCCAC
CGGTGCT
TTCACTCT
GGCCTT
GAGGGCG
ATACATT
GCCCCAA
AAGGTCA

AGTTCAT
GATTAC
CGGTGTG
ACGTAAA
CGTTAC
ATCCCTC
CCCTAT
GTGATGC
TTCAAAT
GTTGGAA
ATTGGGG
CATATGT
CATCCCG
CTGCAMC
TCTAGCC
CCACATG
TAGGGAA
TGCGGGT
TGCCATC
CCATCCG
GACGCTC
CGTGTAC
TCCCCCG
CCAATAT
CTGTCAA
CTGTGGG
CAGACC
ATCCACC
CTGTCA
CCGACT
TGACCAT
TCTCCGC
AAAGAGT
GTAAAT
CATGGAA
ATTAAGA
CCCTGGA
CAGACTG
ACATACA
CTGAGGC

[illegible]

AAACTCC
AAAAACA
AAAAAAT
CTTGGTC
CTTCAGG
TTGGGGG
TTGGCAA
CAGGCTA
CCAACCT
GTTCCAG
AAAAAGT
AAAAAAG
ATATAAC
AAGGCTT
CGGATAG
GACGGAT
GAATGGA
AATCGAC
CCAAGCA
AGTGAAC
GGACAGA
ATACTGG
ACTACGG
ATTGGAA
CCTGGAA
TGTTAGA
AGCTTGC
ATCAAGC
ATGACGA
ATACGGA
GTATGAC
AAGGCAA
GTAGAAA
GGTACAA
AATGGAT
GTAGTAG
CTCTACA
AGAAATT
AGATGGC

CGAACTG
CTCAACC
GATGTGA
AAACTCA
CAATTCT
GTATGGA
GTTTCAG
AAGAAAC
CATAAC
TGGGGAC
AGGTTT
AAGTAAG
GAAAGCC
TTATGTA
ACCAGTA
CTAGACT
TCAGTCA
GTAAAT
CAAGAAAT
CTCTTTT
TGACTAT
AGTACTG
GGGGACA
CAAGCTT
GTCTCTA
TCAACTG
AATTCGA
AGGACGC
CTTTTATG
ATGACTG
AGACTGC
AAGTGAC
AGTGTC
AAAGGAT
CACTTAA
GGTGCTG
TTAGGTG
ACTTAA
TGCCAAA
ATTGGAA

[illegible]

GTACGTG	TGGGTTTGT	RRGVHFVRRVQRGPKDHRVDTGAFILAFEGDITLPEKVK
TGTGGCG	GGCCCTCCC	VMYHCCDVKPYVPPPMRCMRFCHMSSRCTAVPV
CGCACGG	TGCGAGCA	CGECGRPLHEGNPCVPPPCVNCHEHSSRSPECVYK
GTCCCCG	GCAGGGTTT	KEKIQEYKSLDRLTYQEARRKFERLHPVQLSTSYASVT
AGCCATT	GGAGCTCTC	SFTPPSSSHSHPLPPPPAPPLPPSQRPSSSQPEKCSPL
GCAGCCT	CCCTCCCCT	ALAGEGAPSQSSRRPPDRSSASSRQPDQADNTKVA
TTCTTCTT	CCCCCTGAC	RSSSVPLDJDSLSLGTCPSPAREKEKEKRKKKQSQD
ACATTCC	GACTTCCAT	KAPVVLEAPPLSSQNNPTLVSDPLFMDVVPSSSVTD
TGGGCTG	TTCTCCTTT	VDPALPSFVTCISPAFHMDPRSVLQWNCNGYYRRLLE
CAITACC	TAACTTGGT	LKHLIASYSAACVVLQESHFTDGHSPILRGFAFCNRV
TCCCCCA	TTTAAATTCA	RPLQASGGVCTLVRTDVVSSWIPRLTILEVAVRVHLH
GTTCCTCT	CGTTGTTTC	TPLTICNLVLPDRPPTSAALTALLQQLPPFELIGDFNAH
CCTCCTC	CCCTTTT	HLSWGSAYTTSRGRVLDQFIADHDLCLLNDGSPHFA
CCCATCC	ACATTTT	AHGTSSIDLSTSPPPFSLFAGHRMTTFAVTITFCRS
TAAATCC	GGCTTCTCA	RHLFLHRPFTFSPPLFFFRWSPHDDLCDSDHFLILSL
TCCCCGT	TTTAGGAGC	PSHWTRDRIPRWVFRHADWPLYTSVVFTSPSDSIDEV
CCCTCTC	ATACACTGT	ISDVSNATRAASIAVPQSGPPRRRPVWVWNTIASAIR
ATCTCCC	TCTCCTGTA	ERRRALQHLKRHPRAALVNFKRLRAKARYIHKQSKRV
CCCCCTCT	TCTGCTTGT	CWERFLSSLGSSVPLSRVWAMLSIQGCQRPSSIPGLALP
CTCGGTG	CATTAAATTA	DGVCTDPLVIAEHLAAHFAKASATASYPVAFRLKKQA
GTTCCTT	ACGGGGAT	EASNLCTTHREEPYNKPTTEWELRLALSSSHDSAPGPD
TAAACGG	TCTAGATT	SIHNQLLKHLSPPQRQYLLQVFNHWWLQGVFSPQWRDSI
GCACGGC	GTCGCGAC	VVPVKGKDPSSLDYRPSISLTNVPCKLLERMVARRLC
TATTCCT	GTTCCTTGC	WVLESRLSPYQCGRGREGSPIDHLLRLETAIRKAFSK
ATGGTTC	AGAAATCCC	RQHLVAVFFDLRKA YDTA WRRHILCTLHDWAEAPYQ
CGATGAT	GCTTTTCT	FLSVFYPAACSVGLRGLPIFISQFLSRMRMRVVGSSF
GGTGTGA	ATTGGTTT	SSPQIQENGVPQGSVLSVTLFLAINGLVSSVGSVTPAL
GCAATTT	AGTCCTGTT	YVDDFCIWYSSHSVASAERLQQAIRRASTWALSHGFQ
TGATCTT	CTACTTGAT	FSPSKTRVMHFCRRTTVHPDPELFLGTQRLEVVAQHRFL
CGTCGTT	TTTAAACCC	GLLFDKLLTWLPHIRQLKSCMRKLNALFLAHSSWGA
TTCCCTCTG	CTTTTAGTT	DRITLHIYRALIQSLDYGCQVYGSAAPSALKLLDPVH
TTTTTCGG	CCTCTAATT	HCGIRLATGAFRTSPVDSLLAEAGIPPLQTRRHQLLVSY
CATAGTT	TACTCTCGT	AIAIRKFPGHPTYHVLFECDGLRTPDTRPRVGLPVGLRL
CGGTCCC	CACGGTCTT	ASLCRDFNLPSLDCMRRAIPRTPPWLVPRPQIRLDLYHA
CCTCGGG	AGCCGACC	PKASVSPHFRLLHTVLEEFQAGATIFYTDGSKFAGHTGY
GTTTGAC	ACACCTG	AFTSPIGVECHSLPESCVFTAELT AISRALHFVKQVALDR
CTCCACA	GCTTGCACT	VLICSDSMSSLOAIDRCYSSHPVVIDIHDLLADLRTACS
AAAAATC	CCATACAG	VAFWLVP SHVGHILGNEQADQLARVAVTPCPFPWVIPGA
ACAGTTA	AGGTCATG	DLRVQMRSLIQDWNNTNWQATALS NKLSIKSTTAAW
CTAGTGT	GGACTGAT	SSSLRPSRRESTVLCRLRIGHSLTHGFLLRNEPPPHCGC

GAGCCAT	GACCTAGCT	GALQSV
TTGGGGA	GTTTGGTCC	AHIL
AGAACTC	CATCCACCC	VECP
CCGACCT	ATCAACCA	LLAL
AGCCCCT	ACCAACCA	RAKYG
ATGGCGT	AC	IPHS
GGATCCC		LP
AAATCCT		SV
TCCTTTCC		LAND
CACTTCC		SRIA
TTCTTIGC		
CATAGCC		
CCCTCCC		
CCGCTAG		
GTCCCCA		
GCACGTG		
TAGCCAG		
TCCGTGT		
GGGGGGG		
GGGGGGG		
GCTGTAA		
TGTACCC		
TTATGGC		
TGAGCCC		
CCTGACA		
ACGCAGG		
GATCACA		
CTGCTGG		
TACCTGA		
GCTGCAA		
CCTCCCC		
ATGCAGG		
CCAAGGA		
GTCGATA		
CCCGTCC		
TCGTGGG		
GTATCAG		
GGTCCC		
GGCAATG		
GCCATCG		

[illegible]

[illegible]

[illegible]

TGGACGA	CATACTCAC
AATTCTG	AATACCAG
AGTGAAT	GAGGAAGC
TAGCAAC	TAGATCTCC
GAAAGAA	ATTATGAAT
AAAATTA	TAAAAAAC
TCAACAT	CTAATACTG
CCTGTGT	TACAAAAA
GACGACA	GGAGTCCA
TAGCAAA	CTATATGAG
GTTGCGA	CTGCAAAAG
CAGGAAT	TGTTCAATG
ATCATAT	CTCTTCCAC
TTICAAT	TATCCTGAA
CAACAAA	ATCACTTGT
CAGTGTG	CCATGAGC
AATACAT	CTCCCAAAT
CGCCAAA	TTAAACAA
TATAAAT	CAGCTCAA
TGTGCTG	ACAATATCT
AGTTTAT	ATAGATAAA
TAGGCCT	GTCCTTTTA
AAAAAGT	TTCAGTTGA
TCACGAG	AGAAATATT
TGGTAAA	CAATTGTG
CTACACG	TTAACAAAT
AAACGGT	TATTTTAA
CGCGCAA	ATTTACAAG
GTGCCTT	ATGTTATT
CAAAATA	TTGGTAGCA
CCGTTTC	TGTATTGTA
AGTGTC	ATTTTGCA
TTGAAGA	ATATACTAT
KAACACT	TCITGCTGT
GATACTG	GAGATCTA
TTGATTT	CCTTTTTTA
GIATIGA	TGTAATTTT
AAAACAA	ATTCITTTA
AAGTTTG	ATTTTGTG
	GTAATGTA

[illegible]

[illegible]

Dap	Daphne	Locusta migratoria	ATTCGCC GTTAATA AACAGTT TAGTGTT AGTTTCA TATTTTTC GAAGCTC ACTGTTT ATCAAAC TTCTGTGT CCTCAGA CGTACTT ACTTGGA ACATACA TAGGTTT TTAGA	TAATGTGAC AATCAAAA TTTAAATTTA TAACTTACA CTCTTTTAT TGTACATAT TTATGTTAA AAAAGCTAA TCATGTAAC AATATATAT TTTAATATG TAACATTTA TATATGTTA ACTTTTGTA CCATAATGT ATGCAATG ACGATGCTT TATACAAC ATCACAGTT GTCCACAG GCAAAAATA AATCAATTT CAATTTCAA T	TAYITLVRPLLEYCCPLWDPHQVGLTEDIEKVQRRRAARF VLSRYGGEASRKCCCTNWGGNHEGSEKGSFCTITLRRG SVTEMLHELGWKSLKERRYSLRRDLFAKFSQFTTSECE NILLAPTLYGRHDHPNKIREIRSKTERFRCSFFPRAIREW NGREKL	cd016 50	498	740
Dap	Daphne	Locusta migratoria		TAATGTGAC AATCAAAA TTTAAATTTA TAACTTACA CTCTTTTAT TGTACATAT TTATGTTAA AAAAGCTAA TCATGTAAC AATATATAT TTTAATATG TAACATTTA TATATGTTA ACTTTTGTA CCATAATGT ATGCAATG ACGATGCTT TATACAAC ATCACAGTT GTCCACAG GCAAAAATA AATCAATTT CAATTTCAA T	EQQCMPPQGTCSDDQACPHKSWSVKSPSFLTTGMNLC NYKPVQRLEVKDCGSMFFHLNIQSVRNKLPELEFLFLE FQCDILCLMSTGHTLSNEHWLKEHEHNMYTPLGKFLET GFFRSNIEKGGVSILMKNDLNCIRIVDLKHLASEGQIEI AAIHLPPSADTLKSHNFSCLDNKPTRQNACLDNISNLS RDDVNLNVKPMMLSDHDGILLKLVKPKDHRPTQPIVN RPVRQLKPINITNLTORLLTFDWDATVMENTTLNNAFD YPPQIIMLLIIRPNLFNEYCPPPSTKVYKKEETKEKLT VPHLNSLRILLKIYYEKSCTDSEYRHSYNKLNKNIYKKDI ISAKKEANGTYISQATNKCQAAWKIIEAEAGLQVVNPPI PRADELNKNFINSQSADHSNLDNDNADSLSSLETETX VDKLKFRWHAIYPKDILDVARSLNSRSTEDIYGLSNFTIK RIIEAICIPQLINRMFLEGTYPDCLKQTIIVLPYKKGDR TSPQNYRPIAIVPITSKIEITMKNQLNDFCNSLLNKAQY GFRSGSLTVKAVQEVVHSVLHGFFENRQHSYATLVDLSK AFDTVSHGILTKKLETYGITEKELALLKSYLANRNQMV VNSTNKSDDLKVQYGVPPQGSVLGPFLFIVYVNDLPNLLP FKSVMYADDTITFTISHNDVNAVQMNNSMMMEKATQW FQENRLTLNNNKTTEHILFSLGNINNEQNHPVKLLGYLD SKLCWNTHTEKILCSKLARVLYLLRKLKDCVSKELMLM AYYAFHAPYIMEYFFGETPLERKKSSGGRKRPDAFLA RTHAYGHLCFFPCTLHYGILLWGNASAGAKEVELWQKKA IRCIFGLSKLTSCCKQYFVQHKILTVPSLYVKALMHARQ NVDEYESRSDVHLYGTRNRNLLDIPWARIKVKNSFLH	cd016 50	498	740

842

				AAATTTC GGAGAGA ATTTGGA CGTTGCA ACTGTCT GTTTGG GCAAAGG ACTGAAG TTTTGTAC CTACACC CAGTGAC TGGGTG AAGAAGC TGACGA CATTCCT ATGGCAC TGCTGAA GAAATGG GAAACTT GCCGTGT TCTCTTG A	CAGGTGTCC AGTTGAAA TATCGTGA ATGAAAGTTT ACGACGAC CGGCTGCA AACC CGAA ATCTTTTIG AACA		VLCHIFKLFKVLHDLRSLKAIDNYLIPQAGFRPGKSTTS QVINLTQLLEDGYENRKITGVAFIDLTAA YDTVNRLLM KKLYATIRDFKFTSIIRSLQNRFSVEFQKGQSRWRAQ KNGLPQGSVLAPLLFNITNDQPLPAGTENYIYADRA MAVQGSDFEEVEKRLTDALDALCQYYADSQKPNPSK TQVCAFLRNKQANRKLRSWKDTPLEHCANPKYLV TLDRALTYKQHCINTKMKVSGRNNILRKITNSSWGAKP EVRITALTVCFSAGEYASSVWYNSAHAKQVDIALNET CRLVTGCLKPTPIDKLYHLA GIAPPVAVRREAAAYSERAR AELSESHLLHNIQPATRRLKSRRSFLRATDNLSDQPGSR VDRWKRSRERHWMPEKEELAPGHNEWDVTVWKS LNR RTNTARSRDNL MKWGYSTVSNLDCDGVQTTQHMYN CPECPSHCTLEDLMLATPNAVDVARFWAKCI MDRRKVILGLSETKWGSGKKTLRNGYILHWCGSQT ETRNGVALLHLHKDVAIHTDIRYVSDRIQMKLHVDNTYI TIFQVYAPQTGCTADEKEQFLENLEDLIVEDNVIIMGDL NAQVGKERQGCENIIPHGYGKRNPGEQQLDLCVRNK LIUKNTWYOKRESHKITRYSWDGQHKTVIDYIITDOKAS	cd016 50	1	232
Vingi - 9_LM i	Locusta migratori a		GTCTCCC ACAAGGC CAACCCC TCGTCCA CGTCGTG	TAATATTTC TGCTGTGT ATAGTTCGT AGTTTACAT GTATGTAA GTAATCTCT GTATGTTAT ACCTTTATA AACTTCTGT ACACATGA TGCTTCTGA CACGAATA AAAAAA						
RTE	RTE- 24_L Mi			TGATGATG ATGATGA		cd016 50	497	746		

844

845

[illegible]

847

				CAGAGGA AATTAGA TGCATAA TAAAAGA AAACAGA AATGACA TA	CAGTTAATG TACAAAAA GTAACAAT CATTCAT GTTAATTG ACAGACAA TAAACATCC CACTGATG ATGCTGCA ACGCCAGC GAAACATG TATGGGAG AAACTAAA AAACATTG TGTTTGCTC AAGCGGA CCATACCTA TAAAGCAA ATTTATTG TAAGCAAA CAGC				
RTE	RTE- 37_L Mi	.	Locusta migratori a	TATTGGT GGATTGA GTAACCC TGAAAGT CTGCTGC CGGTCCC AACCCGG ATAAAGG AGGAGGG TTGGCAG CCAAAGA AGCCGTA AAAGAAA AGCCGGG TCACATC TCAAGGC ACCTCTG AGGCTAA CTACCCC	TAAAAAAT TCAGGCATT TGCAGGATT TGAGGACA GAACTAGA CCGAGAAC GAAATCAT CGTGGACA GAAGAGCG AAGAAAAGA TGCAAAAGT ATGAGAAAT GAAAGAAAT ACTGGAAG AGGAGGAA GAAACGAA GAAAGGAA GCTGATGTG AATTAACGT	MMASSLKGSTTPGGNQSTIVKDNVSEPTDSGELTPTCKR LKSCKVSYFATFNVNSLLKIGLKLHLDTLTNQNIATA LOETRYTDEAAFESEGFRLKGGKGRVMKNVPQLGT GFIINRQILGATDFSSKSSRLSTLTFRGANKMYTMNVH APTNETNKTDPEKVEEFWDXDEETLSKIPTKNVILLGDF NAQLGKERKYRNIIGNYPAHNRNTXNGERLVEICKTFD LIMKSTAFKRLPRKQMTWVCQILCKTDDMGMPNPLLG EFQLDHVAIPKRAGKEIQNVKVLRGANLSDHFLXKIKI KIIPKVHQPRNTIRTVKYDTGKLKTSDFKTSLSQKDYQ GWEQLKDAAEKRRKAWADWCSHKLTLENLNFQEV KNTAKTIRNLKRKYEQARXTQIEEDFQKNNTRDFYRTF KTNLNRYDPPTLSFKNKEGKIVYGESENSQILAEYFEEL LNCEAPADTFSYNDTEPNPDSPTTEEIKNIINTLKNK APGEDGIIAELFKYANENVIELKQIISKIWEEEKLPEGWT SALVHPHKKRRYIPFTKKGDKTDPGNYRGISLIPVAYK VFSKALLNRVEKQLNPQIDEYQGGFRKGRSCAEQILNLK MIMAYQKQRCCKFVLTFFVDFKAYDSVDRQTLFQILKE MKLDRKSSSELVRQTLNNTTSKVFRGKLSSEPKINTGVR QGDGLSPMLFNCVLDKVIREWRCQMKGVGGIRLGCKK	cd016 50	556	814

					AGTTGTA ATCAGT TGACGCC TAGGCGT CAACSC CAATATA AATCAAT CGTTTCA GC	GGTCAAA GATGCCG AAACGAA GAAGAAG	KGIRVDCLAFADDIVLLSESLEEA VNTKELQRQAAT GLQISIEKTKYMTNTEAPGWMAIDRGKIQKVEFKYLG EWIDTGLTEKDAMEARINKINLAYKLTMINYNKKNISIN AKLKHYYQTVIRPEALYAAETLNLRTGVIRKLELVERKIL RRILGPRRTEDGQFRNRSNQELYSKIEKISNIKKRRCEFF GHICRMDQGRLTQQISQFLLYKSKIRWQEVQRDLEE MGIRDQEIHN RVIFK			
					GAAACCG TTGCAAC AAATCGA AAATTIA GGACGAT TTGTAGC AGTTTIG TAAACAC ACATTCA TTGAAGC AACCAGT TCTGCCA GGAGCAC TTTACCG GACATGG ACTGCAC AGAAACC ACCTGGG CAAGGAT GTGTGG CCAAAGA AATCTC CATGTGA TAAAGC TGCAGGA ATGAACA GTCACGT AAAAATT GCACATC CTGCAAT ATGTTCA	TAGTTGCC ACCATCT ACAAATTAG AAGAAATTC TCAACTGTA ATATAATAT TGTAATATT ACTGACAA GCTGTATC TCTTTTAAA TTAACAAAT TATGTTATC TATGTGACG TTGTCATT GCTGTAATG GCCTAATG ACAAATAA ATTAATTAT TATTATT	MSQSPVASSVAVVGNKAAVTPLREDSPVASSVAAVG NKAAPVTVPPQZASTATASTAPPNEEAEEKKRSSREE EVVSSVAGRRKMVPLRLKDFLVEGSKMKNLEISDLKH PIAPHQNVHSGKGNSSSELKFATNLLPEHQGIKDKKEFFV DFGLDNFDFCFSEHWATENELTLVLDTYNIANSYCRK LHLRGVAIYTNQSLSYSTFKLDDLYLCEEQNFEAAGIFT VSLKLVISLYRSPKGHSVFLEKLEMLLNVFRESKWLHY DIVIGDLNADFVTTNKGVS LNCNDKQRSVTELKNLL RQHNLSINSRPTDKACLDNIFVNFKAEDSCMCRRL MYVTVPFSDHDAVSVNYISRFHPDKCDIPVPKVITRPI TDDKISFRMALADREWFDCYGDTHCSLSDLSAKVV FDRFFKTLTLFNDSIPLKNAKQHSPPKCKVIDKGFKTK VTNRQNPWYTEQLTNLKHQLMVLYNISNALKTEHARS AYVKCRNEYKKAIVEAKKAHNFRSIENSTNKCQAAWK LINSVAKDPRNKSNNVSPQKFNDFIKSVEEVGNAIKPD IRSELLSKNVSRSTTNINVTFFSEVSPSLILKIVKQLKSSD SADIYDISCNMLKRVIDCIVYPLTFCINKCISEGYFPDELK VSRVVPVYKKGDTESPPSYRPIVPAFSKVLCEVYQQL SVYFENLGLISEQYGFPRKNLSTVDAIDS VVKYLHQVFE DKGFAQVTFCDLSKAFDCVEHTPLLDKLEFYGIRDLSLK LLKSYLQNRQVVCIGKEMSSTEQVKVGPQGSVLGPF LFLIMINDLPSFKSSTVLYADDTTLHSSNDXSLKTCV AKTLDAQSYWFKANGELNETKQQMVFTLKDCLPSD DPSYVFKVLGVYDDKLSWGQHVKYISGKLSRVYLLRR LMDYVPEKYVRITYFSFQSIITYGIILWGNSSCVQDILIL QKKAIRVTGSPHKEHCKPLFCCLKVMTVNLNYIYVYLI YTKRKLPETKLRENVHYNYNTRRNKCFITSYHRLSKSLNS YELVGINYLI SCHNNMYKIYPNKHNLKRCMTG	cd016 50	665	908
Daph- ne- 29_L Mi	Locusta migratori a									
Daph- ne- 29_L Mi										

850

ATAATCT
CTATAAA
TATTGAG
GGATTIT
CTAGTCA
TAAGAGT
ATTTATT
AGCAGAC
ATATGCA
ATAAGCA
TAATGTT
GATGTAT
TATTGGT
GCAGGAA
ACACATC
GGGAAAC
CACAGAC
CACAGAC
CTAAAAAT
TGCAGGA
ATGGAGT
TAGTCAT
CGAAAGA
CCTCATG
ACAAATG
CGGGAGT
GCGATIT
TCGTTAG
ACCTGGG
CTCAATG
TAAATAT
AGCTGCA
TTGACAG
AGATAAA
TAAACATT
GAAGTAC
TTACCAT
CCGGGCC
CAGAAAT
TCTCTGT

GACGACT
GTGTACA
AACCAACC
AGATAATG
GACTATTG
TTTTCAAT
GAAACCTA
ACAATTTT
TACATAT
GACCATTA
TTAAATTT
GTAATGG
GCGACTT
CAACTGTC
CATAGTG
AAGCGTGT
GGGATAC
AATGAGA
CTGATGA
AAACGGG
ATACAGC
TTGAATC
CTGGGCT
GACAGTA
AGACCT
GCAACTT
ATTATG
ACCCAAA
ACTCCCA
GCAATCGT
TCATATAG
CGGAAGA
TGGAGGA
GGGGATA
TAATCTT
GACAACA
TCTTTGTG
AGTAGGA
AGTTGGC
AATGCAG

TGTTGTGA
AGCTTGT
TGATCAT
CCCATCC
CAAGTAC
CAACACAC
AGACCCA
TTAATTGT
GTTGTGA
ATGCCAT
CGTGAAG
CCAGATG
AGGTGCC
CTTTCGA
AGCGGGT
TTAATTTC
AAAAAAG
CAAAATG
GCAACAA
TTTCAAG
AGCAATT
GGATGAA
GAGATGG
AGAAAT
TCCTCCTC
GACCAGA
TGAGTAT
GGTAGAT
TTATCGA
TCTTTGA
AGCAAGT
TTCAAGG
AAAAATA
TCCCAAG
AGTTGCC
ACACCCG
TACATCC
AAGGACT
GACTGGC
AGCAGTA

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

GATAAAT CCAAATG GTCTATT CCCACAC TCACCAA ACCAACA GTTATCA TTGGATC TTCCAAC ATCAGCA ACATCCC GTTCCCTA CATCCAG ATACCCA AGCCGAA TCTTACC CTGGTGC CCGAATT CAGCATA TCAATAC ATTACTC TCCAAGA TCTCTCC AACAAAC ATTCCGA AAAAGGT GTTATTTC ATGTTGG CCTAAAT AATAAAC AGGAATC AGGAGAA TCCATAC ACAGAAG AATGCTG GAACCTC TCAAACT TGCCAAG TCGAAAT TTCCGTCT	TATWAATA CTCTTTTCT TTTAGAACT GGTAACAA GAAAACTTT CATATTATA GACTTTTGT TTATTGTAG ATTACTCGT GGTTGTTT TTTAACTTA ATACTAACT ATGCTAACT TCACCTTAT CTTGTAATG AACATTAT CTCTTTAT TGTGACCTT GGCCAGTG CTTTACATA GAAGTTTTT AGGAACCT CTTCACTCT TAACACCA ATAACCA	LQTSGTAMGKRFAPSYANIFMAEWELEISSISPFEPFFYG RFLDDIFIWTESLTNFWQYFEILNNHHPSVKLTANISEQS MDFLDTTVFKGFSFSQTNKLDIKVFFKPTDTHQLLYKNS FHPKHTFTGIIKSQILRFFRICQKVDFDAACRLVESKLRE RGYSKRFLRKIKNDTVLDLLKINEPPPPNNFGAIPCRGS RCRTCEHIHMSHYFQNKNGKIFEIKERLCTCDKNIYLI SCKKCDQYVGETQNALRTRFSGHRFDIHSKNRPISHH FNLDPHSIEDLLMPIESIPDSDLKPKTKRLQRESYWMD QLQTVAPNGLNIQHDNQMPFVATFNHTAAKTSIVKQ YYSELQNRFPNVYTKRLVIAKYRNKNLSNFLVRGLDPI ESNNT				
---	---	--	--	--	--	--

				GCTACCC TTTACGC CACACAG ATTCCCT ACTCAAA GGAGCTA CAATTAG CTGAACC TCATTTA GTCAATA ACCTCCA GAAGTTC AACAGTC TATTATA TCATAAA CCGTCTA CCAAATG CCATACT ATTCTC GCTTCAA TGGAGTG GTCAGTG TCATTTC GGATAAC ATCCACT GGATGTC CCCCACT GCAAAAA GTATTTA TTTTTC		TGAGTTGG ATCCAAAATT TGGTGTATG ATAGAACA CTATGGCGT CATTTGATC CATGTAGCC GACCCCACT TAGTGGGA TAAGGCTTG	MFSCRFLIRKSDRKICLDCKVIPGESLTTQHRVMVMDV RVKRRRAKRSSQNGAPIKWHLKGKELRIFQHILEGG FRQPHGSANDMWERMAHEIKKVAKETLGESRGFGPRG KESWWWNDVSQSVRIKRDCFKDWSRCKNVETWDKY KIARKEAKKA VSEARTQA FEGLYQSLGTKEGEKSIYKL AKGRERKTRDLDDQVKICKDEEGRVLVQERDIKGRWKK YFHNLFNEGYEILPDSNRNLDIGEEDRNYNYYRRIQEHEV REALKRMSSGKAVGPDNIPIEVWKS LGDRGIVWLTNLF NEIMRTJKKMSDEWRRLSTLIPIYKNKGDIQN CAN YRGIKL MSHTMKLWERVIERRLRKETRVTDNQFGFMGRSTME	cd016 50	317	567
--	--	--	--	--	--	--	--	-------------	-----	-----

AACTTTG	GT	AIYLLRRVMERYRTDKKDLHLVFDLEKAYDRVPREIL	
TCGAATC	GT	WKALEKKGVRJAYIRAIKDMYEGASTSVRTQDGTTFD	
TCTATGA		PITIGLHQGSTLSPYLFVLVDVLTEHIQELAPRCMLFAD	
CATGGAC		DVVLVGESREEVNGRLETWRQALEAYGFRLSRSKTEY	
CAAATGC		MECNFSGRRSRSTLEVKVGDHHPQVTRFKYLGsfvQND	
AAAATAG		GEIEADVSHRIQAGWLKWRRASGVLCDKKVPLKLGK	
CGTGAAT		FYRTAVRPALLYGTECWA VKSQHENQVSVAEMLR	
GCTAGGT		WMSGKTRHDIRNDTIRVGVAPIVEKLVENRLRWFG	
CGTTGCC		HVERRPVDAVVRVRVDQMEESQVKRGRGRPRKTIRETIR	
CGGAAGC		KGFRGQ	
AACGCGT			
TGTATGG			
CTCGAGT			
ACAGTGT			
CAAATGA			
GCAAGAG			
CCGCTGC			
ATCACCA			
CCTGGAT			
GTAGTGA			
AAATTAA			
GCAAGGG			
TTCCCGC			
ATCTTCG			
TGAACGG			
GTGCGGG			
TAAAGAA			
GCTAGTT			
CATGAGA			
ATAGGAT			
TCGTTTIG			
GAACTTG			
AAATATA			
GGTACAC			
TTACAGG			
AAAATCT			
ATGGAAG			
TAGTAGA			
CACAATG			
ACTAGGA			

GGAGAGAT
CAATTTC
ATGTGCC
TACAAGA
CAACTAAG
CGGTGG
GGAAAAA
GGCGAAA
GAATTAG
ATAGTTC
AGAGATT
AAGCTTT
GGTATAC
TGGTGAA
GTTAGAT
CGAGAAA
TGGGGTA
GGCATCT
TTGTGGA
TAAGGAG
TGGAAGA
AAGATAT
TGTAGAT
TTTAAAA
GGTAGG
AGATCGG
ATTATAG
CCCTAAA
GATTGTA
TGGGAAC
AAGACAC
CTTTAAT
GTAATTA
GCGCTTA
TGCACCT
CAGGTAG
GGTTAGA
AGCACAC
CTTAAAG
TGAAGTT

[illegible]

YSDLRRLTQKVPIDDKIIILGDFNAQVGKNFEAWKGVLG KHGVGSCNDNGRLLLEFCAEQQLTTITNFTQQKDSXKTT WMHPRSKHWHLIDYVLVRQNRVDVCHTRVMPSAEC QTDHRLVRCKLNHFHFKPKRGGIPRRRLNVNQLJAA VRDSQANLQTRLEDHLIDSPEALWQHINKNSILQSSVES FGFSLKKNKDWFDNNQVEQLLTKKRSAHQAHLAQ LCHVRKAARLACSKLQKLRDIQNKWWLNLAEKTL CADLGDHGRGFYEALKA VYGYTYQVQSPLLSDGQTLLT DKISILNHWSHFQTLFSANRVVQXGIGIHILQHLLVKNL LDIAPTMGETLKAQQVKTGKAAGVDGIPPEWKHGGQ ALHAKFHELVRWCWEQKLPSPDRDAIIITLYKKKGVS DCSSYRGITLLSIAGKILARILLNRLVPAIAEELLPEQCG FRXNRSTDMVFLRQLQEKCREQNKGLYVTFVDLTK AFDTVSRKGLWLILERLGPCKFLKMIILLHEDQQGV YGDALSEPPTXGVKQGCVLAPTLFTIFFSMMLQRAMV DLDEQNGIYRYRTDGSLENLRLK AHTKTLNHLVRELL FADDAALVAHTEAALQRLTSCFAEAAELFGLVSLKKT VVLYQAPQEDYHHPHITIGESLKVQQFSYLGSISSD ATIDKEIDNRIAKAYRAFGLHKRVWCNKHLKKSTKIS VYRAIVLSTLLYGSWSVYRHLRLERFHQRCLRTIL NIHWTDYVSNVTVEQAQVTSIEAILLKLQLRWAGHVS RMKDHRLPKIVFYGELATGCRKRSAPKKRYKDSLQYL SLGHIDYHQWSTLASNQETWRHTIHNAVVSFENTRISL EEKRQRRKNRAPPIPKETFRCAFCNRLCSHIGLFSHQ ACSKRG	GAAGTCAA GCCATG	GGAAAGA GCTTGCG TGCCCTT GTGAGGT TGGGAGC TATGCTG GTGGTGG CATATGC CTCTAGT AGGGTCT CCCATGC CAGACAG GTCTCAA CTGAAGG GTCAGAC AAAGTGT GTCCATG GGCAGGA TGGGCTC ACTAGCC TTTGGGC AACCATC CTAGGAG AAGGACA ACTCTAA CTTCAA CCAGGC AGATGGA GCTCGTT TAGCCCT GTAAGGC CATCCAT CTAAGAG AAGGATA CTCTAAT CAAACT ATGACCT GAGGACT TTGCTGT CACCGTC
---	--------------------	---

CAAGTCT
GCTAGGC
CCTGGCA
GATAAAC
CTCAGGT
GTAAAGG
GTGGGGC
CAGTTCT
GCGCATG
CTGTGCC
TCACCTA
AAAATATC
CAATGCA
CAGGCCC
GAAGGGT
TTACCCA
CACTTGC
AAAGCCC
TGTAGCG
ACAGGCG
AGGGACG
AAACGGC
AGGTGAT
AAGTGTG
ACTGGAA
GYCGCAG
ACCCAAC
CCTGCAT
GCAGGGC
GTTCAGG
ACATTTG
TCGTCTG
AGACTGA
CCCAGGA
GATGACA
GTCTTGT
TGGCAAC
ATCTGA
ATGACCA
AGCAGCC

[illegible]

[illegible]

						PPAD EYPCPHILHSEVEAAIRTLKQRKSPGFDEITAE MIK AGGPKLTASLHEL CQKIWL EEPITNWGKTILIPKKG D LARCENYRTISLISHLEKVFLLJLLRRLQHQL EPYLAEEQA GFRKDRSTVQQILLVRLLA EKA KRKGK KIYNCHIDFQKA FDSINQQRIWAVLASFGVDRKVVKLLQNIYENTRLAVK TGNDMGTFWKMLIGSRQGDPLSPLLFITYLERTMDSVT KLN LGVTIQGSIISNLR FADDINLMNEDLQALQDOQATAL QASAEKSLRINTSKTMTVFGARNIEQQIEINNQPVEN TEQFEYLGSIITWDNNCSVEIKNRIAKATGALASLKH I W KSSKITVPTKLRILSVCFVSVLLYAAETWTLKENDKKKL LAFEMRCYRLLKIKWQDKIRNEEIRNTLNIHTTIIDLIR QRKLR LFGHICRM DNNRJIKHVVT SXMNGKNKQGRPPR AWMDITDWC KSAQEA FHRAQNRNQWKTATRTND MRLTVXLANSNCTIKDVSVLNQFNTGSDHRLVHAKVK XDIQERXRLXQHGGQKV KKEEYSLKLTDSKCSQKTK VXYGIHTMRQCRNRNGNSTKKSQRQENIHQNTGADRKE KEDEDQIHYQPDRIEDEVRKSPRRRQFNILKKEGGVPL YKQLAGIFSDCMRTGQIPDSWKNAVIVPIHKKGDKTDL RNYRPISLLSTIYKVFTIILTRRLEQVLD SNQPREQA GFRK GYSTM DHIHTNQLKECKEYHISFCMAFIDYEKAFDSV STEAILTALSNQGIETTSTQNXCKDSGIA PRPPSVRRNL LRIYEEATAIRLHKDSRKIKIQRGV RQGDSISPKLFTAC LEEIFRQINGWEHKGININGEYLNHLRFADDDIVLFAQTPN ELQEMLSDLKTESEKIGLQINYEKTKVMFNAEVIPQVVTI EGRTIERVEEYKYLQGVITTTQGDQMREIQHRTQLGWHA CGRYSNFLRDNRPICLKRKLFNMCILPVMTYGCETWTL TKKMEQKLATTQRSMEYMLGFTRRDRKNTNIWIREQT KVEDILVTVRSSKWHWAGHIARYTDNRRTRKRITEWQPR QGKRYPGROYKRWRDDIRKVIQSQWMQIAQXRYIWKQ KEAYISSSG cd016 50	175	448
RTE	RTE- 1_M Ma	Mesobut hus martensii	TTCTTCG AGAAGAA AGAACAG TGAAAAT GGACATG GATCAGC CCCAACG GACGGAC AAANA TGGGGAA AGCAATG GTGAACT ACCCGTG AAAAACC TCACCTC GAAATTC TATGGAT CAAGACC AAAGTGT	TGAAATAA GAAATCTG ATGATGAT GATGA	MIRSSQAENDARGANPPGGPIADVATGKKDHL SCKDLF TIGTVNVRGLNIGKLEIKREMTRTGLDLLGISELHW TG SGDLKSDDYTVHFSGNDHNRRRGVAFVANKRLAKYIE NARCISDRIMTIRLKGKPLGVITIQVYAPTADADKQDIVD FYTAIONALDQTPKKDIYIMGDFNAKVGDGEDGQVVVG RHGLGDRNDAGDRLVQFCQENKLRITNTFFKQHKRRLY TWTSPNGQHRNQIDYILCQQRWGRGSITA AKTLPGADCG SDHQLLIARIKIRLCKVKKTVA AKKFDVSRIPQTYTIEVR NRFEALSDDGGQTEELWAMIKGTIETALEYVVPYKMPTK GAVWLTEKTLQIAEKRRREAKAKGEIDKARRLHAD FQRE cd016 50	537	789	
RTE	RTE- 4_M Ma	Mesobut hus martensii		TAGGATTG GATTGGAT				

				TATCGAT ATGGTGT TGGACGG TGGGCC CCGGGCT GGACAAC ATCGACC ATGCTAC TGGGGAA GAGCAAA GG			ARIDKNNALNARCISLEEAFRKGRTRELFATVKQVKTPF SARQSTIKDRDGNELTDQDQIKERWREYTEDLYTSIDDDQ QNEQTDQGLQEPDLLESEVEWAMRQLPSNKA PGIDGI PAELLRPIPRAALTALCQQIWKSCTWADWRRSVFVPIP KKGDTDRDCANYRTIALPHASKILLKVIQKRLGNVIDREL PDVQAGFRGRGTRDHIANLRWIMEKAREYQKKLYMC FIDYTKAFDCVEHSKLWVALRELGVPEHLIRLSLYQN QEATVTRYGDTDFWKIKKTRQGGCTLSPFLNLYAEVI MRKLDLESEVGVKIGGRTISNRFADDTTLLAESEKDL KRMIMKLKNESEKMGLQNLNKKTKIMTTAAKEVHIKIN NEEIEV VDSFIFLGSKVDCNGDWRPEITRIALGRTAMV NMDPIWRCCKDVSTTKRRLVTAVFPPIATYGCETWILTK ADRRKIDAFELWCWRRLLRIPWTARVTNKSVALERINPD FSLEGKITRQKLTFFGHVVMRANSLEKDVMLGRMSGTRP RGRPKTRWLDTIKKDDTGKNIEELRELVRDRKKTWRDMT HRIAKNRTRLNG			
				TAGCCATCA TTGTGTGG GAAAATTTC CCAGGACT GTATCTGAG GTCCTTCAG ACAGCTCTC AATACAGT CTCTAAGTG GTGTGATA GGGTGGA TTGTCTGTT AATCCAGCT AAGACTGT CATTGTACC TTTTACACG AAGAAGAC AATTGGCG GGGCTAATT GAACCATC ATTCTCGG CAAAAAGT TGCAGTTTT CTACCCAG			YDLALIQEPWHYKGQIRGLNCKEGKLTFTSDSGPRTCV FVRNNINVPVWDFCSRDTTVRVKLDDEGGVSRELIVSS AYFPYDKDELPTPEFXLLVEFCQKNHFDLLVGVDANS HHVVWGSKDINRRGEYLLLEFLMGTNLFLVNLKGNPTFV VVNRREVIDLTLGTQNISDLIENWHVSGEPSSDHRVYICF NLQGMVPSQVTYRDPRTTNWASFREDLKNLGGIQSRF CSVVDLEMTVDQIQAILLSYHQNCVPRTARSSRQVPW WNEQLGGLRXNVRRLFNKAKRTGDWDYRSALTCTYN KAIRKAKRSSWKRYCQGIEDIPVSRVLVKIMSSGNFCV GSIRLPDGSMTSTCKDSLKELFRVHFPGSTLIEGSSCLQV QSDLDGHVYHAKREDWGLVRGVINQSNIRWAVSVFKP FKSAGPDGILPVFLQQGIDILASPLCRVFRACLAFGYIPW SWRQVRVSPFKPGHSSYTEAKSYRPSLTSFLKKTMERL VDRFVRDGAALLKTPLCRSQFAYQSGKSTENALHEVVSLI ERAIGYKEIALGAFLDIQGAFDSTSEQISLAASRHGVCS TIIRWIGSMLKHRVVYSTLYDEEVKILTDRCGCPQGGVLS PLLWCLVVNELLTLHSMGKVSSLLGMLMT	cd016 50	468	649
Loa	Loa- 2_M Ma	.	Mesobut hus martensii	T						

[illegible]

869

870

871

ATAGAAG
GGATGAC
GGAAAAAC
GCCAGAG
GTGAGAG
ATTGATA
TGGTGC
GAGACAT
GAATGCA
AGAACGG
GATAGGA
AGGAGGA
GGAGTCG
ACGAGGA
AGAAAAAC
GAAGAGA
ACAGGGA
ATCGAAA
GATGAAA
AGATAAA
CAGAGAA
GGGGAAG
AACTTCT
GGAAAA
ACAAAG
AAATGG
ACACAGT
ATATTGA
ATGGAAA
TACCGAA
GGAGACA
AAAACGG
AGAATAC
ACGTACA
TAGGAGG
AGCAGGA
TGCTCAG
TGATAGA
TTAIGT
ATAACAA

ATGAGCA
AGGAAA
AGAGAG
TACAACG
CAATAGA
GATAGAG
CTGGGAA
AAACACT
CGGGAAA
ACAGATG
AGAGAAA
AGACAAA
ATCACAG
AAAGAGC
TATATGG
ACTGAAA
AGGCAAT
AGGTCAA
TTCCGAG
AAGAACT
AAATAAA
GGGGAAG
AGGAAAC
ACAATGG
AGATAGG
AAGAAA
TAAAAAG
AAAACTA
GAAAAAG
CAATCAG
AAGGAAA
AAGGGA
GCGGCAG
AAAAAGA
AAGAAAA
ACACGTG
GTGGGAT
GAAAAAT
CGAGACT
GAGAAAG
CAGAGGT

AAGGAAG
TTAGGCA
AGGCGAT
TAAGAAG
GAGGCGG
GAAGGAA
AGAATAC
GTCAAAG
CAAAAGAG
AGACTGG
GGAATAT
TAGTAGA
AAGAAAA
AAGCAGG
AAGAAAT
GGAAAAA
AGAATAA
AAGAGGC
TGGAGAT
AGTAAAA
CCGGGAG
AAAAATC
TGGGAGG
TGTAAA
AGTAGG
AGAAGAA
AAAAAGAG
GACAAGA
TGCAGCA
GTAAAAAT
AAGAAAA
GAGGAAT
GGTTAAC
TYACTTC
AAAAGGC
AACTAGG
AAGAGAG
ACAGAGG
AAGATT
GGGAACA

AGCAYAC
AAGGAGA
GGATGAA
GAAGGAG
AKAGAGA
GGATAAA
GAAGRAA
GAATAAC
CGAAAAC
GAGGTGA
GGAAGGC
AATTAGA
AAAAATGA
AGAAGGG
AAAAGCG
CCAGGAG
CAGACGG
ACTACAA
AATGAGG
TATGGAT
ACACGGG
GAGGAAC
AGGTAAT
AGAGAA
ATACTA
AAATACT
GAATAAC
ATATGGA
AAGGGGG
AAAAAGTA
CCAGAGG
CAATGGAA
AACCGGA
CTAATAA
CCCCGCT
TTTCAAG
AAGAGGA
GGAAGCA
AAGAACT

ACAGAGG	CTCGTACAC	MEALKTQNIYSERISKIEQRLGLEKRLKALDELPAKTR			
GATAACT	ACTCCACGC	IHNAESTITELQAQIQDLSRSPTMQQDNGSTVPNTAEIC			
CTCATGG	ACACGCAC	SLRSEAEVKRRQEQTSCNVVVVTGLHY'TRETS'LHLA			
ACACAGG	ACACCTACT	FSVVNALDPTVLRDDV'ASVRTMGRLDATNSSARGDGR			
GTATAAA	TTCTCGCTC	LPPLAV'TLSSSALARSIVIAKARKRKLHTSELDA'TLLEE			
TTATACG	ATTCCACCT	KALSPDHQGLININELLPSDVHKLRT'RARLEAKKRQGCR			
CAGAAAT	TCACTATCG	TFVREGRLYTCAATMTASVLRSS'TPTPSWRLFPGRPLP			
AATAAGG	CCAYACTYT	TSSN'TEATXQY'THQ'TLSSAQERSKV'KSEYLFCEVSAKG			
GAAAAAG	CACACTATA	VSPFVGVVYRPPHAPFFOGSNFIDOL'TTHMHNYSK'I	SSF5	6672	971
TGGAAG					
GGAGTTA					
GAGGAA					
GAGAGAT					
GCTTGAC					
GATACAC					
AAATGGG					
TTTCAGA					
AAAGGA					
GAAGGTA					
CGGCAGA					
CGCAATT					
TATATTTT					
GAGTAAA					
GCAATAG					
AAACTGA					
ACTGGAT					
AAGAAAG					
GAGGTAA					
AGTCTAC					
GCCTTTT					
CGCGGAT					
TGGTGAG					
TACAGTT					
GACAGCT					
CTGCTGA					
TCCATCA					
TTTTTCTG					
TACTGCT					
TTTCTGCT					
GCTTCTT					

GCCTTCG	CATACTCTA	MGDFNSDQLSSSEDAKFIAFIDENSLSSVPYGATHHKQ
GGCCCTC	AACATACTT	GSDTWLCLCLIDEQDRLLSYWKNTSPFINGHDLITATLD
TTCTGTG	CAATCTATT	VQIPRYVPNTYSYRNFKGSAEKLRFELSACDWSSLTSS
TAGATAG	GTTATTCT	LDECISTLNALTNAINHLAPLRTVTPRQRHPWFVTVS
TTGCACG	TTTCTTACT	HVXXXRILVALRDLVSRDLRRFRDSRLSDLRIR
CTCCTTTT	TTGTAATGC	LARDNAHKQVEEARLNYYSRLSTLTDVAEIWRELEKL
TGCAACC	TGTAAACTT	GISATKAPSPSRFTIDELNKHFSISNDPLAPAVEDYLLT
TTACCGT	ACAACTCGT	LESIDLPEHFEFTITESDVLAASVSHFDTQARGSDGIPQV
CGATACA	ACAAATATA	VISKALPVLAPLLSHIFNLSEPCFSAWKLSLVRALNK
TAAATCG	ATGTACGC	VSSPTALTDYRPSLLCFLSKALEWLVRQVSEYLESRL
CCGAAGG	AAAATAAA	LDNFQTGRTGHSTHVRVGINKKKVTLLLLDFSKAFDT
CTGCAAT	ACTAATCT	VCHVRLRLKLSLTFGFSKQVIRWFASVLTGREQAVVGDN
TTTCTCT		SERFSPRRPAGSVWSLLHCTSTTSVALIPMFILSMRM
GTCACT		TCKFTVNATLRSSILYQTRRAFLSSASRRVRLVLAVALYI
TTTCTACT		NDIGFCLDSVSHLIYADDLQIYSQCHLEELDSLNNKMS
GCAAACT		ANAERIMGWAAQNRLKLNVNKTAKVILGSPYYINALPK
TTCTACA		VTLINMGVPGQLHLLIWCQVSESSVRNLGLVLDLTKLT
CGGTCC		WKEHVTQVCKRAHSLMYRLYFFRKSTNLRKHLVQA
TAACTA		LLFPHIDYCSLVYCDLTQELDLQLRLVNTGIRYIYGVRR
CAAATCT		DEXISPYRRELQWLJTAGRKYFTACFLRKMFNSVVP
GCTTGGC		YVLAFFDFRVTLRPVRGVTPLEPAFATETLRNSFHISA
TTGGCTT		SYLWNNLPISHIRNTSSTVTRKLAKDHHFFQLENT
GGCTTGT		
TTCTCA		
CTTGAC		
GGCGTCT		
CITTCCT		
CCGCAAT		
CAACATC		
GTGGCGA		
CATCTGC		
TGCATCG		
TGGGAGA		
ACTTCA		
AGCGGAC		
TTACACAC		
TGATCGA		
GGTCGGT		
ACTTCA		
CGCTCTTT		

CCACTAA
CGTCCA
TCTAGCC
GTAAGAG
TCAACAT
ACCTACT
CTTCACT
CCGTCGA
TAACTTA
TCATCCT
CTATCAA
ATATCTG
CTTTACC
GACTGCT
AGTAAGT
TATCGAC
GAAAAAT
GAGTGCA
ACGCGCG
AGAAAGG
TGTGGAA
GTGCTT
TTCAGCC
ATGCAGC
AGCTGTA
AACAACT
TATACAG
CACAAAG
ACCCCAA
GAAGTGT
GAATTTT
GTGGTCT
GCACTAC
CAGGTCA
AGTGCCA
AAACACC
GTAAACA
TCAAAAC
TATCGTG
GTGAACG

ACAGGCA
GATTATA
GCCTGTC
CCTCATG
TGCAGAC
TACAGAG
ACGCCAA
GGGTGCC
AAACTGA
GGACTAA
ATCCACC
TCTGCCA
CAACATC
AGCAACG
GGAGTAA
CTGCAAC
AACAGCA
GCAGGCA
AGAATAC
TCCCAAA
CAACAGC
AGCCGGT
GAATAAG
AGACGGA
GGTCAGC
ACCAACA
TCAGCCA
ACACGTC
GAGAAAT
CCCTCAC
CAACCG
CTCGGCC
GGAGTAA
CAACACC
TACGCCG
TCCGTGG
ATGCCGC
TTTGAGG
GACATTC
GCGGAGC

[illegible]

AATAGCC	AAAACTATT	SVDVTWVCGACQRRCTTLRSWSCHVLHCKGRQEPKDL
GATCTG	TATTCGAAA	PFKEHCSLSDSQIGLSQHERHVHPEVRNDKRAAEAN
GCGCGTC	TAAGGGGA	KPKGSGRRPSIWSDEDLLESEYHGARNINEKIAE
CGGCAGT	GGCTAACT	HFDRGTGRQVSDARRRKDYAALRCRGPGQPAEGVEAI
AATGCCA	GCCTGCAA	EEVDEGEIPEGEELVATDGAALSGPPENGGSAPAEQVN
CGTATGA	GTTGAACG	APALESSQDRECSPAVGSDEQIEDSSDDDEFSDALGEI
GTCGGTT	CGAAAAGTT	SLPEPLSVERTTISPPRRDWKGPWRWEICNASEEAGSY
ACCCATC	AGACCTTCC	ANWVTGLQELVRNNALSEIGLDSLYDQLIMRHPSDD
TCTAAAC	CACCTAAA	NEQDRLQLNARGPPRRGHRKNRRRRRLTAADRKRFAF
GCGTAGA	GCCCAAAA	ARCQDLWNNPKKLAEIVANDLSILQRRQAPGRTEQ
GGTGGGG	GTGATCGG	TLYNELWGRVGNIEAPRRTEPIPVSRIFTPTTQEIIMGR
AGCTAAA	GGAAATGAA	IRRIKNDSSAAGPDGVTKDDDLRGRGVSIASLKFNSILLAG
GGCCAGG	TCCGCGGGT	YYPKAWRENRTTLLPKPEKDPADVKNWRPITISSMVSR
CGGTITA	GACCCAG	VYSGLLDQVRRAVIKQCDRQKGFTEENGCFNSIQLLDD
CCCGACG	AGTTGGGT	AVSNAKKAGGVITILDVSKAFDTPHAIQGCLEKKGIP
TCGAATT	AAACCTTG	ETVAAAYISSMYRDCSTAIRTRSGDVKIGMKRGV/KQGP
TCTCCAG	AAACGTG	LSPLIFNLVLEPLLERLQETSGVEIEGMNLSCAAFADDIV
GTCTGTG	GAGAAAGG	CFANTAPEAGRQLRMVADYLGRLDMSLSVSKCIAVEY
TCAGTCG	GAAAGAG	VPHKWTWYTKNPGLEVNGNAVPSISPSETFKYLGAKVS
ACGGAAT	TCCCGCCAC	PWKGLLEGESDAFREVISRVQRPLPKPMQKVDLLQMY
AAAGGTA	CGAGCATC	IFPRYTYGLITSPPAKAVLKTIDRIIRTRIKEILHLPESVSS
CTACAAC	GAGTGCTG	FLYTPRKQGGGLGLEVEKMLVLAALRNGLRARQSHDPV
ATCTACT	CGGCGCCC	TRAAAMNSNAADDRKSYADALRLHWPLTTKELDTYKY
ATCTATC	GAATGAAA	QLRLSYAQKWAQKQWQGGQGVVEFAQDPVGNISWLQRY
GGGATCG	CCGATCGC	DLLPASRYDAIKLRTNTYPTRALMKIIDGRVDSSCRKC
GAAAGACG	GGATGGTG	QGSSETLGHILGRCRYTKDKRISRHNEIKDLLKARLAKN
CCTTACA	CAAGTCGT	HQVMDEPQITVRGQRFKPDLVVKTNEGRVHVIVTVRY
GCGTTTT	AGGACGGG	EHRTYLDDEGRTEKIGKYRQILSTLRRDLHSNAEEVPIVI
CCGATTT	GCACGACC	GSRGAIPRETRKALSKLGIGKSDWLITSLIALRSSLEIVNA
TTGCTCTT	TAAGCCCTCT	FMDD
TGAGCAT	GTCACGGC	
TTTTCTTC	GGCGAAGC	
AAATTGC	CAGGAATC	
GATAACC	ACCATGCA	
GACCCGA	AAGGTGTG	
TCACGCG	AACTGGGG	
GGGCTTT	CGGATACCT	
GACAAAG	CCACGGGG	
CAATGCG	TTTCCCTGG	
TGGTCGG	GCATCGCG	

TAAGATG	CGAGCGAT
GTTGCAA	GGCCAAAG
TCCTTTCC	TCCGCTTC
ACCTCGT	TCAGCTACA
TTCTTTTA	AAACAAA
CGGAACG	ATGGTATG
AAAGCAA	AGACTTCGT
TGCGTGT	TAACACTA
GGGGAAC	ATTTTCCG
GTTAAAA	AGCCTAGC
ACTCCCT	AGGCTCCCT
TCATGCA	TGACAAAG
TCCCAGG	CTTATGAAT
ATTATC	CTGGAAAA
CTGCTTA	GGACACAA
CTGCAAA	AGTGGA
GCAATGC	AAGCGCTG
GTGTGGA	ATGGTGGA
GCGACTT	CAAAAGTC
TACCACG	AGTTGAGA
AGTCGCT	CTTGATAATC
CCACCGC	AGTTGTTT
AAAGCAA	GAATAAGA
TGCGTAT	ATTTTATTA
CGCGCAA	TCGTTGACT
AAGCAAT	TTTAAATAT
GCGTGTG	TTTATTAAT
GGGGACT	GACTGTAA
TGTCAAA	TATACTGAC
GATCCCC	TTGGGACC
CGCCGCA	AAATCATCT
AAGCAAT	CTGTTACCC
GCGTGTG	GGTACCGG
GGCACCA	TTCTGTCA
CGTAGAG	TCAAACCG
CAAAGCG	GAAAGTCC
TGTAGGC	GTCCACGT
AGACTTT	AATGTGGT
GTCAAAA	AGACGCAG
GTAGTTC	GAG

[illegible]

					ELDTKLQRLVNTGIRYIYGVRRDEHISPYRRELQWLTTA GRRKYFMACFLRKIENTSVPTYILAFFDFHVALRPVRGE VTPLDIPSFKTETL							
					CTACTAT CTTACAA CACTCTC TACGTTA TCGACTG CCAGTAA GTTGTCTG ACGGAGA CTTAACA ATAACAT CGAAAAAT GAGCGGA ATTGGTG TGAAGAA TGACGAG TCGCTCC TCATGTG TGCTCAC TGCAAAA AAACAAA AAGAGGG AAAAATGA AGACCTG CGAGTTG TGTGACC GGCAGTT CCACACC ATATGTG TCGCCTT CAGGTCT ATAATGG TGA				MKIISWNVRGLGSSRKRSVIREVLRKEKPTIFLIQESKLT EVDRRWVRSVWSARGLSWAFTQSWGASGGIITWKNE EMDCTEELIGNFTVTIKLRKEDGFEWALTNVYGPNSY RERDDFWXELAAIRGLWNGPWCLGGDFNVIRFLNETTN PNHRVTSSMRKENLFDVSYELRELPMQLQANFTWSNFQE EPLCCKLDRFLFSSDFESA WRGLAVRALPRPTSDHCPVR LSQENLGGGCPFRFENMWLLHPDFIKKVQLWWEEDQ VEFRWKGYKFYKKLQALKARIKAWNKEEFGRVEERKE	cd016 50	496	761
					TAATGAAA GTTACCGTT TCCTATAAA AAAAA							
					GGTGGGA GTAACAG AAGAGAT AGGGGWT GGTGAGG MTGTAGA AGGGTCT AAGGAGG							
					Nelumbo nucifera							
					L1- 11_N N							
					L1							

AAGGGA GMGAATC TGGAAT TAATAGT GATCCTT TTCGAGC AGTTGGT AGAGGTG GAAATTC ACTCCAG GTTGGGA TCAGAAG AGGGGC TGAAC TK ATTGGGA CTACTTCT AAGGAAA AGGAGAG GGAGATT AATAAGA TG GTGC TGGA ACT AAAAAGG CAATTA AAAACTT AGCAAGA ATAAGAA AAAAGGA CAAGAAG TAACTCA GTTAAAA GCTTCAA ACAAAGG GGAGGGA GAGAATT CTAAAAG AAGAAGA AAAAAGA TGAGTGA AGTCGAA	WIANRIQIDGGEAERLTVGEREERWKLKQELQEIRK EEISWRQKSRCKWLKEGDGNTKFFHTLANGRRKNFV GRLEWNGEVITSEEEIARAFVEHFKSIYDSPRREREIPD LDWLPGNRVLMDQLEKNFEVQVDAIFSMERSKAPGP DGFMAFYQDCWDIIGDLMVEVLEFFERGGKINSANA TFIALIPKKVNAHPKDFRPISLIGSMYKVISKVLAEIRKK VPLSLSPNQGAFAVSGRIIDQILIANECVDYRKRQGEKG LICKLDEKAYDLVDWGFLEVLRLRGFGGERWCMWVK GCISSCHYSILVNGVAKDFPAKRLRQGDPLSPFLFTLI VDVLSVQITRADARNLLRGWRVGREIVVTHLQFADDL ILFCEGDPPELENLKVLENFEKVSNLRVNWNKSCVAGI NRQEVVASLARS LGCEVSDWPLKYLGLPLGGNPRSVS FWYPVLERFSKRLAGWKRNYSISLGGRTILKSTLANLPV YFLSTFKLPVKIANQVEKIMRDFLWEPGDTKKDHRVKW EVVSRKKDKGGLGIGNIRIRNAALLGKWLWRFPEQES LWHKVIRSIHGMAANGWDAGLDPRASRFPWKDIVGIL PVFSPFIRYDACVGNKILFWEDLWIGESPLKDRFPNLYSL SLKKSQFISFRASGSDWDLGLRRGLRDSEVDSANDIFSL LSSVFLDHSSPDSRIWAETSTGFSFCSFFENALVLPSPV SFPVKEVWSAPVPPKVR AFCWEVVLQKINTILEVLQKR PNWCLSPHYCSCLRNGETVNHLLISCPFSALIWDWFS LFD CRWTPPTSVADLFDPSFRRQPKKGRVMWRTTAA ILWAVWEERNRIFQGEKSAPEVIEGITIKAVHWLKTIL DLFKDVSIDDLIRGWSILARSSTSSPTTQAVWSPLQCNV KLNFDGSSIGNPGSGIGGVFRDAEGKILALYSGPIGLGD SLRAEVLALLQGLRWAKQLGFQQLVEGDSKLVVGWL QNSLGTWSLQRHVRELLSLTRELGAQVYWSRRSANAL ADNLAKQGVNREVFVLCVVSVSYLLFRGHLLLWKVTVV SSGFI				
---	---	--	--	--	--

886

				<p> ACTGTTG CAGGAGT CCAAGAG GATGGAG TTTAATA ACAGGTG GGCAAGG AGTGTAT GGAAAGC CAGGGGG ATTAGTT GGGTGGC TGCTCCC TCTTGGG GTGCTTC AGGGGGG ATAGTGA CAATGTG GAAGAGG GAATGTG GTTGAAG GCTTGGA GGAAC'TG ATTGGAA GATAC'TC AGTTTCC ATGAAGT TCAAAACA GCAGGAG G </p>	<p> RPIPRVGSFCWEAYLERFNTGDDLQKKRPNTCLSPHWC VLCKADGETGFHLLVSCQFAKSIWEWAFSLFGRRWNP SDLSLSPSPFRWHMKRGRMTMWRGLRAAVCWSLWEE RNSRIFSEDERGAELRAVHIRAVQWLKSLELFTDYTV EDMLRWGEIANCTQTKPVIRVSWTPPPAGFLKLNFDG SCRGNPGSGIGGVFRDENGVISALFSGPIGVGDSLRAEI LAVMEGVQKAKEMNVVRKLIEGDSEVVIGWLKDNCG WNYXNERRKFRWLVRMEIRMSWVKRSANSLADGLA KQIEREAMFWAKL </p>			
L1	L1- 10_N N	Nelumbo nucifera	<p> GGAGATT TGTAGGG AAGATTG AGTGGA GAATTGT TGGGAAA GTTGGGG CGTAAAC TGAAGAA CCCAAAT </p>	<p> TAGTTTCAT ATCTTTAAT AAAAGTTCT TGTTATTGA TCAAAAAA A </p>	<p> MRXSWNVRLGSGARKRATIKVXKKFNPDVVLQESK RMDFCRRARSVWRAGIKWAAAPSWGSSGGIVTLWK EDVLIGLEELIGDHSVSLKFRQVEDGFVWLTNVYGP DYRQRKKLWEELSDIRGLWEGPWCLGGDFNVRLTEE TNNPRQVTTSMRSFNQFVDXFELKELPMLNNTFTWSN LQSSPICKLDRFLFSTDWEAHWGNLTSRVLPRLTSDHW PIVMCKEXLRGDGPAPRFENMWLTHPGFIEKIEWWQ HYNPPVQWEGMRHFHKKLQFLKSQIKEWNRLEFGRVEE RKQVLVEQIEELDMKEGRGLFGEEDRERRWSLRKQLED LLIKEEISWRQKSRCKWLKEGDRNTGTFHVSANGRRRK </p>	cd016 50	498	763

			GATTTAA AAGCATT TCTACGA AAAAAAG AGGATAG AAGNTC GAGAAA GAAAAAT GAGCGAG GTTGAGA GATTAAG AAGAGAT ATTGAGG GGGATAA GAA GTTT GGGGCCA GAGGGGA CAGAAAG AAAGGC CCAAAGTA GATTTTG GTGGTGA TAGGGCA GATTCIT GCAA AAT TG TG	NFISKLEVEGVISHPVEIAEAFVGHFKSIFRRPLNERIWL EGLDWPTLSPENREWLERRFXLEEKEAIFSMDRTKAPG PDGFTMAFYQDCWGIVEEDIGSVFEFHRRGIINPGLNST FIALVPKKKEASQPRDRPISLVNSLYKIAKVLSLRKK VLPSIAPNQGAFAVMGRQITDQILJANESLESFRKSKKKG VVCKLDEKAYDXVDWSFLEVLKSFSGFDSCWXWIW GCIRSCHYSILVNGTVVDFFPASRGIRQGDP LSPFLTLJA DVLRLIQRADSRNLJKWRVGRREEIP LTHLFADDTIL FCEGGDIEWIRNLKLAQLFERVS NLKNLDKSSLAGVN LEERELREYADV LGCATSDWPMA YLGLPLGDNPKKVSF WNPVVERFGKRLASWRKNYLSLGGRITLJRSTLANLPV YFLSVFKIPTKVSNTEKLMRDFL WEPGEKEKRDRHRSW EVVTRGKKLGLGIGNLRKNRMALL GKWL WRFP LERD KLWT KIVASHIGHLEDNGWEPGILARTTAASPWSAIHKLL PVFSPCI RYGLNSGNKIKFWELWVGDSPLMDQFP SLY ALS DLKNEPISSFLSPDSPSGFGWNLGLRRAIRDIDL DH ILSL LAILENVFLDPXKEDTRFWVPSSSGVFSCASFETLT QPSPDIGLSFPVKEIWSRPPIPRVGSFCWE TFLERINTDL LQKKRRPN SCLLPNRCTLCKADGETGYHLMVTVC SVAKS V WEWVFSLFGCVVXPPG SVQSLSFSPFRWPKRK GKT M WRSIRA AVCXWL WEERN SRIFS DKERTIGNLLDLILLRV AHWK TMELFSDYSVADXVRNWKEIASSALCKRTCSPS WSPPPSGFIKLNPDGSCXRNPGPSGIGGVFRNEAGAILAM FSGPIGEDSLRAEILA ALEGQRARELNIRCLVLEG DSE VVIRWLKGGSGLWSYNNEKNKFKWLSRDIVVHYNNW VKRTANSFADGLAKQGVNRDSTFWANFLXFLVFFGGEL SLFISPLYSFT	TGAATGTGC AGAGTTTAT TCAGGTTC A TGCCGAGT GGGCCAAC AAAAAGGG CCAAAGGCT ATGTGGAG ATGTATGA GGGTGCA GTATGTGG ACTGTTTG GAGGAGAG	AGTGTTT GGAGAGG TAGGGGG ATCAATT GGGCGIT TCTGCC TCATGGG GGTCTGC AGGGGGT GTGTGA CA	L1- 17_N N	Nelumbo nucifera	cd016 50	426	691
--	--	--	--	---	--	--	------------------	---------------------	-------------	-----	-----

GAATCAC	RSKKKGVVCKLDLEKAYDMVDWSFLLLEV MNVNFQGD
GGATATTCT	RWCKWVMGCIRKCHYSILINGTVVDFFPALRGIRQGD
CAGAAAAG	LSPFLFTMV AEVLGR LISRADSRLIRGWRVGRSVPLT
GAGAGAAC	HLQFADDTILFCEGGDLNRIRNLKLTQLFQEQMSNLKVN
CTTGGAAGT	FEKSCLVGVNMGEGETSECADILECAVSNWPLNYLGLP
AATATTGG	LGDNPQKDEFWIPVVERFEKRLATWKKNYLSLGGRTLI
AGGTAATC	RSTLANLPVYFLSVFKIPAKVASRLEKQMRDFLWEPGEE
CATATTAGG	KKDHRVRWEMVCRSKREGGLGIGNIRSRNLALLGKWL
GTGATACAT	WRFSLKDKLWVKIVGSIHGVADNGWDSGIVARTTRA
TGGCTGAA	APWKAIAILPLFSPWIRYDLHSGNLINFWHEHIWIGDSPL
AACTTTGGC	KDQFPSLFTLSTLKNPVSFLVPSNSTSGFSWNLGRRE
CTTGTTTGC	IRDTEFDCILSLLSILEGVFLEPTKTDRIWVSLTPSGSFTCA
AGACTTCTC	SFFKALTQPPSLGLSFPVKEVWCLPIPPKVGSFCWEAFLE
AGTAGATG	RVNTCDLLQRRKRPNSCLCPHWCALCKADGESEYHLLIS
AGATGGTA	CRFTKSIWEWCTSLFGCRWPPL
AAAAACTG	
GAAAGAAA	
TTGCAGAAAT	
TAGCCTCAG	
CTAAGAAAG	
GCATCACTC	
TCTTGCTGG	
AGTCCCCCA	
CCTAATGGC	
TTTMTTAAA	
CTGAATTTT	
GATGGCAG	
CAGTATGG	
GTAATCCA	
GGTCCATCT	
GGAATAGG	
GGGAGCAA	
TAAGAGAT	
GAACATGG	
AAATCTGTT	
AGCTATGTT	
TTCAGGACC	
AATTGGGG	
AAGGAGAC	
TCCTTGAGG	

GCTGAGATT
 CTTGCAGCC
 CTTGAAG
 GGTTCGAA
 GAGCAAG
 GAGGTCA
 AATTAGAA
 AGCTGATC
 ATTGAAGG
 AGATTCTGA
 TGTGCTTGG
 TGAAGCTGGT
 TAAAGAG
 GATAACATC
 AAGGATAT
 GGAAGTAC
 AGCAATGA
 AACAAGAA
 AAATTAAGT
 GGGCTGGCT
 AAGGATGT
 TGAATATCA
 GTACAAAT
 GGGTTAAG
 AAGAAGTC
 AAACCTCAT
 GGCAGATG
 GTTTAGCTA
 GGCAAGGG
 GTTAATAG
 AAACCTCTT
 GTTTTGGGC
 TTTTATGTGA
 ATTGTATTT
 GCTGTCTCAT
 AGACTGTITT
 AGGGAGCT
 AAGATTGTT
 TAGCCCOCT
 GCTTGTGTA

[illegible]

[illegible]

					CAAATATCT AATTCAAT AATTTCAA GTAATATAT ATTTATGCT TCATTTCA ATTTAACCT AAATATTT TTCCGATGG CCGATAATT ATTATATT ATATTTTC TTTGTATT TACATATTG TATTTCTT TTTCAACTC CTAGTATAG TCTTAACTG TTTTTTAGG AGTATAACT ATTAAAGA TTAAAGATT AAAGATTA AAG	MINNPSNVVPVYFCLPILAFWSELPYFHLKAVDRGSHV ATLTSEPVASTGWSYSNQELKTIRKSQPKQRLEPKLWQ KLGDLGIRKPFRRRKRQSRHVCTGLPVVTTVDSTGHP QLSSGCTNSSSSQTARFPVIPDTIIANMQSLLPKLDLY AAIMQPPAGIICLTETWLNESHDPDLIHLPGYLCHFRNDRT TSRGGGVCYINSQYPCRRLPEFEDLSIESLWISVRPYRL PRSVSTLLIGVYHPPSPNGEMNRQLVDHIDKNCESEFLSKH PDGLVVVTGDEFNPSTGIKAQYVTHKGTGLTQIINVLTRD TGILDWCLTNPKYFQPLQLSKMGSCDHYIPILIKPNLS VTNPKAHKSSVWRRNLRNSNLDAFGRWITQFCWNTILS LQTAREKFERFYTVITDAVNVTYLPQRVQVCATNKPWI GTSLSLIAARRQKALTTLGKDSRLYKELRNQVQRECKV CRERYYPNKVAALKESNISRWQEVKCLSGSVVSNW WHQLMDDSTPSVQILAEERNEFLGLTSNFAPIQQVTC SGPVPELLITPGKAFKALSRIKVKKSSGPDVPVPSVIWKE FAPELSDVVCNLYNSSLMGEFVPOELKESIVHPVPKCSPP	cd016 50	612	864
					TAATAGTAT TAACGCTAT ATGTGCTGG CCCTTTGCC TTAATTCAG TCTGTAATA CTGCCATGG CAATTGAAT AAAAACCC TATCTATCT ATCTATCTA TCTATCTAT CTATCTATC TATCTATCT ATCTATCTA TCTATCT	CCAAAAT GGCTGCC ACGGCTT TTGTAGC TCTGTGA ATCTTGT CTCTGTTT TCATGTG TCATTGC TAGTTGA AGCGGTT TGACTGT CAAAAG	Nematost ella vectensis	CRI- 18_N V	

[illegible]

CAAACTC	ATTXEHDLXNYFXDLFQTDQLISENNMTGNTYNCEISSD
TACTAAC	DILXLSPLHQDEILTALNSFKAYKAPGPDGLHPFLPKIL
AGATAAA	DGCRIVLRHTKPDQMGYQFFYQKYWTDVGDVSSFF
TACCTTCT	CQDVFCSKIIPPELNKTLICLIPKVKHPTRIQQFRPISLCNT
GCCAWCT	VYKLTIIVNRLKPIISKIIGPTQTSFQQGKRACDNAIIVQ
CTAACCA	ESLNALAKMKGYPSMLLKLDEKAFDRLEWSFIRRAL
ACCAAAT	FYFNIPPSMISLMSCVTTTISILINGSKTQYFTPTRGIRQ
ACCTTCC	GDPLSPYLFIMCMELFSRSISLSDVYLKWQPIRLSNKGPL
TAAACA	LSHLFFADDIILFSKITTCSCHAIIDVNNFSKYSQGKINF
CTCTAGA	EKSKLFFSKNCSLILKILLPLTLKKEPLLNTWDFLCS
TCTAGCA	KQDKMQISSSLIISKINLLDKDFVITSFNIKEGTSFGKY
GATGAAG	LGPPMFQARPNADQFLLDNFKNKLAWKSRFLTLAG
TACTCCA	RATLIKSTLNLQPNHIMQYIHIPAHILNKMDQYQRNFW
TAACMCT	GTTLEKKKLHLVRWNEITRPKEAGGLGIGLQKLRQKNHAL
ACTGACG	LASLAWRLFKTPQAMWANILLTKYLNTPNNIHSQIWS
ATGACAC	NVINGWHQCQLGIKWNIGKGTHIKFWEDPWIGPGTTIRS
ATTTATA	KIHGPLPLDQQNITISSLINNKGWDWNKLPFELPTLLKV
CCAAATTA	AHNTIITQIATYISKYSNKVDNIFWGLIPTGHFTTTRSMYS
CATCAAC	ILEAQHDVSTNSSTKYNWIKLPVQPKIKGFVWLLHKK
TGATAAA	RLPTKDYLRHINLIQDATCQFCHQEDETIRHLFLDCTNA
TCACGIT	KTFWQDLGILDFFINQKLHYTNSFNWVHDLITCKDFHFPY
TATATGA	QMNPATYPLSLWNITWLNARNNCYKHTMTASAKLTT
GCCMTGG	QQAAKFTYLVAKKHTXVKKRQQLKWKPPKEHFYKL
AAATACT	NTDGAHTTSAKIGIGLIRNSKAEWILGFAAAVPXDSTS
CTGTGAT	AELRALLQGLQLAYSHNLVPLEVEVDSQDIITLLYSNNV
AGTTAAG	SFDNLTDCRYFLDRLGKPVVLHAYREQNRVADKLAK
CTWTTTG	AGCNFDIHSVVRILEDNPDFAWSVFEQERSAVQGHVLE
GCAAAMG	HFRPHSDNYDVTNRYCNSDAVGSSINIGLYNTLAPTGA
GATTACA	SACNRMQYRHPPPSVRNT
CACCAAA	
TCATGCA	
TGATAAA	
CTGATGA	
CTCTGIG	
GAAACCA	
ACAGAGG	
TCCTCCTC	
TCATTGA	
TTTGGGT	
TCTGACT	

[illegible]

TCAGATCTA
TGGAGAAG
GGAAGAAT
TTATGACCA
AACAAGAG
ATAGAGAA
CATTATGAA
ATGCAAAA
TGGATAATT
TTGATTACA
TTAAATTGA
AAAGGTTTT
GCACAAAC
AAAAACCA
TGCAACCA
AGATTAGA
AGCAAAAGC
AGAAAACT
GGGAAAGA
ATTTTACA
GCAAGTGT
CTCTGATAA
AGGCTCAT
TTCTAAAT
ATATAGAG
AACTGAGT
CAAAATTAC
AAGAATAT
AAGTCATTC
CCCAATTGA
TAAATTGGTC
AAAGGATA
TGAACAGG
CAGTTTCA
GAGGAAGA
AATTAAAG
CTATCTATA
GTCATAAA
AAAAATGC
TCTAAATCA

CTATTGATT
AGAGAGAT
GCAAAATTA
AACCAACT
CTGAGGTA
CCACATCAC
ACCTATCAG
ATTGGCTAA
CATGCACA
AACAGGAA
AATGATAA
ATGTTGGA
GAAGATGT
GGGAAAGT
TGGGAAC
CTAATTCAT
TGTTGGTGG
ACTGTGTA
ACGTGATCA
ACCATCTCG
GAGAGCAA
TTTTTGGAACT
ATGCCCAA
AGGGCTAT
AAAAATGT
GCATACCTT
TTTGATCCAG
CAATACTAC
TTTCTAGGTT
GTATCCCAA
AGAGATCA
TAAATAATG
GGAAAAAGG
ACCCACAT
GTACAAAA
ATATTATA
GCAGCTCT
TTTGTGGTG
GCAAGGAA
CTGCAAAAT

GAGGGGAT
GTCCATCAA
TTTGGGAA
TTGGCTGAA
CAATATTGT
GTATATGA
ATGTAATG
GAATATTAT
TTGTCTATA
AGAAATGA
TAGACAGG
TAGACTTCA
GAAAAAACC
TTGGAAGA
CTTACATGA
ACTGATGCT
GAGTGTAAG
TTGACGAGA
ACCAAGGAG
AACATTTGA
CACAGTAA
CAGCMACA
GTGTGTGA
KGACTRTTT
TTAGCTGAC
TTAGCTCTT
CTCAGCAAT
GCAAKGAT
CCAAAAACA
TTTCCAAAG
GACTCATG
ATGCAAAA
TTGCTATCCA
CATCCAGA
GAAAAAACC
GTGGAATC
KGAATGCA
GATGAAGC
AGACTTTT
TTTTTATTC

[illegible]

901

				CTGTTTGTAG TTTGTGTGT TTGTTTATA TTCTGCGTT GTTAAGCA CTTTGGTCA GCTTCTGCT GTTTAAAA GCGCTTTAT AAATAAAG TTGACTTGA CTTGACTTG AC	KPSGSPSDA VPPRLFKEVITTLGPAVLAVVNGSLSSGVFP RNCKHATVQPLIKKPNLDPVLSNRPISKLPFLSKILEKI VYVQLKDFLDENGLEVFQSGFKSLHSTESALLRVFNDI LLATDAGDHVVLVLLDLTAADFTLDHNLISRLQLVGI RGTAQWFRSYLEDRTFCVNLNNVQSSAPLLYGVPOG SVLGPLLFSLYLLPLGSLRKHNCAPHCYADDTQIYVPT KKDAHSISPLIKLEELKIWMALNHLNNEEKTEVIVFGP SGNSESHPSTLVTPESHPIALGSLTQSVKSTVTLNLGVK MDNGFNLDKQISSVVRSSFFQLRLAKVKSFLSRKHFFET VIHAFVTRLDYCNLSYFGAPHSSLSRLQLVQNAARLL TGTWKREHTPVLASLHWLPVHFRVRFKILLFVFKSLNG LAPPYLSSELLHLSPSRSLRSADQLLLDVPRTRRKLRGD RAFAVAGPSLWNAALPMQVRQARSLENDFKSKLKTHLFSL AFNSV			
				AGATTG TTTTTAT GACAAAT TTAATCTT TTCTGGA ACAGAAC CTCATCT ATAAAC TGATGGG ATTCACC CCAGTCC ACTGGGC AGCAATG TTTTAGCT GCMAACA TCAGACA CTCTGIC CAGACAT CACCGGC CCCCATC ATGCCCM CACCAGT CCAGACC TCCTCCTC ATCCGTG CAGACAT	MRGSAGRKANTNWLRLKEAPKEEAASTTPPSPFNM ALFNVRSLSNKTFIINDFHIDYNIDGLFLETWLSLTDAST VLTEASPPNFSFVSTREGKGGGTASIMNSPFAVKQITL DEYPSFEHHAFFVNSPPVLCITVYRPPKPSVFFIEFSEFL ASVHTSFNMILITGDENLHIDGKSDYCTTQFLDLLSLDF KQHVMPQTHNRGHTLDLVISHGLPTSVSAVVDLAVSDH YCVLENTGFIKQETSVRTVKKRYLTPEVVEKFTSVRSY SPPVLPAPCDLLVESFNSKLKSSLDVAPLXSKKRIKAT PPWRTEEIKQLKRRCSAERLWRKHLEINYQLYQNL KLYNQTVKQTRNSYFANIISVNKNPRVLFTTIESLNSN VNKNTMPASASLCEDFADHFRSKIDSVRSNIIKGQKSLFI GNDGLILPEEALDSFDLVEAATLGRVFSQVNPPTCFLDPI PISLKFSGFSESEILNMINSSLQTGIFPAAFKTA VVKPL LKKTNLDPSVFNNYRPVSNLPFLSKVLEKLVFIQLNEFLL QNNIFEINQSGFRTNHSHTETALLKIVNDXRCA	IPRO 00477	521	582
L2	L2- 4_OL	.	Oryzias latipes	T				

				CGGTCTG TGGAAGT GACTGAC TGGATGT GGTATCA CCCCCCC ACATTCC CCAAAAA ACTGCTT CAATCGA CAGTGCT TCCTTGC CACCACC TACACCC CCCCCAG TCTATCC CATTAA ACCATTA TTACAGA CAGGTCT TCC					
RTE-1_PP0		Papilio polytes		TAAGGGAT AAGGCCAA GTGGATGA TGATGATTG TAGA		TYVRSAGSDRFGVSVGLHQGSALSPYLFLVMDALT LNIQEEAPWCMLYADDIVLVGEDECEVQSRLEEWRVKL EGAGLKISRTKTEHLFCDFGGTFSFPIALNGVPLPTCSD FKYLGSLVQSDGGVDRDVKSRINAGWMKWRQVTVGT CDPRMPLRLKGQYKSHRPVVLVYGSECWATKVADERRL HVAEMRMLRW MCGVTRL DKVKNVYIRGSLKVAPVVE KLRSKRLAWYGHIMRRNDGDIKQVLRNLNVGYKGRG RPPKVVWMDCVKEDMLKKGVDSDMTADRDVWKSSTY CADPP	SSF5 6672	18	134
RTE-4_PP0		Papilio polytes		ACGGGAG GACGGGC AGCAGCG TCCCTTTC AAAAACG GTGGAGA AGTCTGC GATCGCC AACCCGC CTGCCAA	TGACAAATG ATGATGAT GA	MMATRKKVRPPVPGGPAVPGYGSQGGGAGGAKNPR LGYYGHTRRPLALATLNTRLTDTERIAEEELSLRLRW DIIGLSEVRREGEETILKSGNLFFREGDQSQGGVGFIV HKSLVNNVVKIESVSTRVAYLILISKRYSLKVIQVYAPT STHSDDEVTMYEDISRAMHSSKTFYTVVMGDFENAKLG IRDGDELRMGKFGYGRNRHRGHLA GFMEKEGLYMM NSFFKKQGHRRKWTVMSPDGATKNEIDFIMSTKKQIFND VSVINAVKTGSDHRIVRGTLNIDVKLERCRLMGSTLRPA PVHIHNAETFQLELRNQFECLGSSVDDLNDGFVHALQS VGSKFYKTHRRKSSPKKLSDHITLELMKERRDMKLHSSDD	cd016 50	549	822

				CGCTGGC GATTATG GCAAATC CCCCCAA	AERYRLNRQISTSIKDIRRINTNRVKLAIERNQGSKVF SRDLSIGSQQLKLKTGEGRVVSSRPELLSEVERFYGQL YATQEPVTDLVEDPRARLTRHYTEDIPDVSLDEISMALK QLKNNKAPGNDGITAEALLRAAGEPALKVLOKLFNSVILE GHTPKAWSRSVVVLFKKGDKALLKNYRPSISLLSHVYK LFSRVITNRLARRFDDFQPEQAGFRKGFSTVDHIHTLR QVIQKTEEYNLPLCLAFVDYEKAFDSVETWAVLQSLQR CRIDRYIEALRCLYENATMSVRVQSEDTKPIHLQRGVR QGDVISPKLFTALEDVFKVKYCHLPETLHRGVGRRF QGPEILVVITYTFLHVTFRVQVIGSLLLRLIKIGTMLEDL NRASERVGLRMNMDKTKVMSNVHVSPTITIGGHTLEV VDEYVYLGQTIQLGRSNFEREVNRRIRLGWAIAXNNR RSGCGISSRPKTYRSVSRQKSLTSVCCQHTGQRRGHSL RATSGGSKSLKGQWSGCSESCEIASGTMRSAAESVT DIARRIANLKWRAWAGHIAARTDGRWGRRVLEWRPRIG RRSVGRPPARWTDDLKVKVAGVHWMRAAQDRALWQS MGEAYVQQWTTYG			
RTE- 6_PP o	Papilio polytes	GTACGTG CGACTCG TGAKTGC TATGTAT GA	TAAGGAGA TGATGATG ATGATGAT GTATG		KSEHSRQVRSRRNRQVQC GCGPAGIGXSPYLLFLVIDA LTVEIQEEAPWCMLFADDIVLVGEDALEVQSRLEWRWD KLETAGLKISRAKTEQLCFDGGSSFIPIALDGAHLPVC SDFRYLGS LIQSDGYIDRDVTNRKAGWMMKWRQVTGVI CDPQIPLKLKGKIYKAVVRPVMYMGSECAVKGTDER RLHVTEMRLRWMCGVTRMDKIRNEYIRGSLRVAPIV EKVRSKRLAWYGHVMRRDESHVTRRTLVMSEVEGYIGR GRPRKRWMDCVKGDMIKKGVMTMTYDRVMWRTKT CCADPTGVG	SSF5 6672	25	135
RTE- 19_P Po	Papilio polytes	GGAAAAAG AAGGGAG ATATTAA AAAAATAA AAGGAGT TTGAAGC AGGTATA TGTTACT GAAGATT TTCCGAA AGAAATA TTAGAAA AGCGAAA AGAAATTA	TGAAAAATG TACATACAT ACTTTATTA AAAAAAA ACTTTTAT ATTAACTT TAACTATTT ATATTGTTA TTAATAAGT TGTAATAAT TGTAATAATG TCAGCGAA TAAAGGCTT		cd016 50	581	833	

				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	STLSSITKALTILFNEINSEQIPQWTKSTITLLHKKGDK QDINNYRPISLMSNIYKVFSKVLGRVTQTLDANQPREQ AGFRADYSTMDHINVVVKQIEKCNQYQNYLAFVDYS KAFDSLHNYIWKTLASQGIERYKINIIEIKNSTAKIQL EQVGEEFKIERGVQRDPLSPKIFTAVLENIFRNIDWDKF GININGTKLNHLRFADGIVLFAETPETLRTMLQQLDQES RRAGLSMNPMTKMTNAEKDIELRHEIEYVDEYIYL GQIISVNQHSKEIDRRISNAWKRYWSLKEVVKKSIPMS IKTKLYNTCILPCLTYGCQTWPSTTKHNKLKANCQRHM ERSMLDIKLRNRWRNVITIREKNKSKEKTKVNDVFKEI KKLKNWWTGHIIRTNKGKWTKDMDWYPREGKRQRG RQIKRWEDDLPGWRRRLARDRETWKKLGEAYVEGQPD TQLDAGK			
CR1	CR1- 2_PH	Parthale hawaie is		CCAGTTA CCCTCCG TGGCGGT CGGTTGT GACGTGT ACGGTGC TCCGCTA GAGTAGA TAAAAATC AGAAAGC TATAGAT	TAACGCCG CCTGGACCT CGTCCGAA GTATCATCA AAGGTATC AAAGGTA	MGGNKKGRNKTQFKPVILNSTDDLDQELEPQETRGVNE CTLCLTSKQSKMIQCVLCEENNHFASCLALNQSGRYHA PVCPSGHGCSVNERSTEQHWVCTGCRGPTTNATPKVA CTKCEFLFCAPCAKLRPGMIDMFNRPKSFWMNCERCL QVSTQNRKCVASDTKESVGESPKTAGEEPLOKYAETGD HDVKEISYEIAQSTVQKRSSSGDWISKNEKKPRICKYY AKNQCKYGISGRGCKFFHPKRCPREMKYGRQCNSKRE CNFFHPWLQCESFRNGTCFDTTCPEIHLKGTIRAAPTAH HRHTHAERGENSGRPWQNSNRITNKYGENNHQYNNVE RDTANDRDMSVYNQDVRRLDQNMQNFQTLYEKISS METYMEKVQHGGPRAPWNPPRVENPTQLWGDPGTTH	cd016 50	959	121 5

				TTTTCTCG AAAGCAA TAAACTA ATTAAGT GTTAAGC AGTGTGC CAGCCTG GCTATAA ACTCGTA CGGACCG TGTGACC AGCGGGA CCAGTGA CCACACC GTAGCTC GCTTAAT AGGGGAA GCGTAGA CGTCAGT T	QLKQTATHHAKQNEHQELKSNKDEILSKVHINIRGLL TKNCRQMVLQDIINDAGSKLLAVTETHLDDAVNDAE VKIENYVLHRSRDKDRKCGGVALYVRDHLGSHTKQKF SNGVDCIIVNIEKRLMCVVYRPPKCTTEKFAEMLQV INTHKEVQSKYQLYTTGDFNPFLLKWLTTDLDPDTGSF VLSRTSSATRDEKDQALSLVRIMEEFSLTQMISSPTRNEN VLDLFFVDNAOLIKNIEIQDTTILSDHRLVEITTLQGERDS HRPQKNQINSISSRLDLNVHDKGINWDGLNAELQSVN WDVLLNNSIESESINETFENILLEKCMKYLSPAKRKFNS RHHKFRKGLLKRAKLIKQSQSTKLLQRKSECKKKLDE IERQLARSHDEEKQEKEKKAIEKIKSNSKYFFAYASHHS SVKSKIGPLVNDTGSIVXDDQEMSEILSRQYQKVFSDKIP NPEIPRIDLTNYLTDFEFTKLSIELAIDELSGNSAAGPDD VPAKILKMCKLSISYPLYLMWRKSMDDGVVPLTYKHAI ISPIPKAGNSMDPANYPVSLTSHVVKTFERVFRKHLVH HLESNNLLTSNQHGRHKRSTTSQLEHYDSILDLLCCQG KSDVYLDFAKAFDKVDHGILLKLLFNMGIRGKCLRW IEAFLTNRTQSVIINKVSKRPVLSGVPPQGTVLGPILFLI MINDMGHQCFAVPSSFADDTILAAVQSTADVRNVQL DLTNIFHWTSVNNMTFNESKELLRYGKNFSLTESTSYT TRDGEIERKQILRDLGVRIQDNATFEDHVARVTVGK KWSGWVLRFTSSREADTMLTLWKQLILPRIEYCSPLWM PSKRKDLERLEAIQRSFTAKIRGMKDLNYHERLAQLQM YSIQRRFERFTIYVWKILEGITVNVNGRIQARHSNRIGRT CEVRIPIKTGGDAIATKYSNSFVVRGARLFNKMPCRHIRN VTGTETAKFKHLLDACLLKVPDKPAVSGYVNCGDNSLL ATPLVTDCSITSAQDQWRA				
				GCGTGGC CTGCGCG TTATCAG GCCGCC CCATTGT CCGGCCA CCGGACG CCGCTTT GTTTGTGA TGAGCTG GCTTTGG GTTTGGG GCTCTGG	TGATGGCTC TGCTCTTGC TGGTGGCG GTATTGGCA GCTTTAAAC GATCTGTAC ACAGAGCT ATTCTGTGA TCTCGGATC TTCTTTT TTCATTCTT TTGTTTCT TCACCCGCT	MVYVYLLFFACLITRYPDVEENPGRFGSRSCRILCTN VRGLYGNLAELAVASSNYDLVLCSETLVSSHRRVSELM IPNFEKPVLRRLDSCPGFRGLAAVVRKGVVVSQSEYEC PCCEMMVLRVCGSFRNFYLSVYRSPSADDDRVFCLLD AMARIQSSDRRSFAFMGDFNCHHVEWLGSGRTDCHG HAAFDFAASGCEQLVEGPTHLGGVLDLMTDVSDL VGSVVSPIGSSDHGALSATLQVRQRVPRVTVRREIFLK SRVNWDAVSGAILALPWGVVRESTDPGSTDLRLSDMV RAYVPSRFTVTRTGDEPWFDDACRRAHQLKREAYARW RTSRSAVDQFLFREQRRRAGATYEEAEREHNLRSRELL ASSVSAHKWWSTLKSALVGVSSVPLIGTGGDLVADP AAKADLLRYFDGKQSRDRVRIPLACDPEPRFSGFAFRS SEVHRLLLDLDPYGGTDPGSMFPFLRRVAGVLAAPLA	cd016 50	514	772
CR1	CR1- 4_PH	.	Parhyale hawaiensis						

GTAGCCC	GGGCTGGTT	SVFRRMLRLGSPFECWRV AHVTPVPKGSVSSRTSDYRPI	
CTGAGTT	GCGTGGGG	SITPVL SKVYERLVASRLGCFLEKQGVLPPTQYGRKGL	
GGAGGAA	GTTTGTAA	GTCDALLHISHVLQALDGGHEARLVQLDFSAAFDRVN	
ATGCCAC	CCTCATGGA	HLGLHLKRLMLGIGGSVLAVISQFLSDRRQCVVVDGKR	
TCTAAAG	GTTTCTCCG	SEFLDLVSGVPQGSVLGPLLFIVYTADLFHLVNNLLVGY	
GCGGATG	TTCCACAGT	ADDVTLA VCDKPSDRLEVSESIQHDLEMIDSWCSRWG	
AGACCTT	CGCCAGTCT	MRLNAGKTKTMLVSRRTVGPFPALMLDGVSLPVVD	
TTTGGTC	GGTGTGATT	DLVILGVNFDPKLTPEPHIRSVASRSQKIGLMRRAWRV	
CAACGGC	TAATAATA	FGSVELTRRCNCFVLPLLEYCSVVWGSAADCHLCLLD	
ATCCCTG	ATAA	KVVRRC AFMLGGVPCDLAHRRRVGGVCFLYKIYNCP	
AGAGGAG		SHPLSEFLPDPFVRVRYTRRTGALHEFALEPLRCRTGQY	
ATGAACA		SRSFFSIVCGCLEWS	
GGCTGGT			
GCCATAG			
CTGGGTA			
CTCAGTC			
TTCCGCA			
GTCTGCT			
CCTCTCTT			
CCTTCCCT			
CGGGTAG			
TACCACC			
TGATGCT			
CITTTCT			
TTCCCTTC			
TCTTCTTC			
CCTTTTC			
GGCCAAA			
GCCCGTA			
CTGGTGA			
TGGCAGG			
GTCTGGG			
GCGCGTG			
GATAGCT			
GGCCGCT			
TTTCGTTG			
CTATCAC			
GACTGTC			
GATTTCG			
TTTTTTGT			

CR1-3 PH	CR1-3 PH	Parhyale hawaiiensis	<p>AATTTT TGATGAT TTCCTTTG CTT</p> <p>CAGAAAA TAATCAA CATTGA CTGTGGT GTATATA CAAATAC AAGTGTA GTGTAA AATGTGT TCGATTA WACTCTT ATTCAAT GCTCTCG AGTGCTA CCGTAG ATATTTT TCCATTG CTTTTGGT TCTTTTGG TTGGATT CTGTGGT CGAACTT AGAAAGC GCTGATC CTGAGTT ACTAAGT AATATTA TATTCTAT CAAAATT TCCTCTAT CAATATT TTTTCCAT CTTCTCC CTATGTA CTTGCCT GATCCTC T</p>	<p>MKQDECHSVFPLEFFYYEEVELESCKYCCKIQALQAEI EKLESDREAMQGRGDGEKEKAPREAHRLSSPSPWKTV MRNKRQRSTGAILEQTNHIKVNNRFAAFNDGDDNVA TSVRPKIKLKTNGEQHRPPQSRDQKKSETSVQPSHE RERNPAPIGQRAIKTELGDIQIKQYYAFKRGCDNRKV ICRPQASTDATAILKDLRCKKMDHVIHVGTNDLRTYS ETQLQLVKDLIAAAKKSDHVLTLSILPRLRAGNRLYSR IRYNGQIKREAXNGLEFIDLPSFDDDWRLFLPDGVNL NGAGHHVLGEMLLAAAASETAALEEGLRTRPKTHIEQPR HPAPHALPADLPAMEQATATNPAAVPCNPQISSPSGM KVGDPKGNEGTHITNSLDTDEEITPTPDASVQMSRLGPP PLLPISVNLVSTNSMSEGTDPQPTNARQTDLDSSNT ERGSVSVSKEKSQSTPVAKKRGQFPRTSIPAHENR STRKSGARNPPKNSCQPPQITRIRSKANSAGNDLLSAR LGGKRPIVQIAADVKSQDNCQAEIQNISLXLQTGALA ATDNAFADLTPPIFSQQSDAPSVSVTRSNANTSVKFLF GNVQSITNMNMLQAIATLAYDVIALNETWINLSKN LEAEVALKGYKLFNDRPTETGRGGGSIYKEHLNPLCK GFAATKEYEFTTHARLGIQSLNIMLAYRSPRSTAEIDA ALYNKIENALNTPNETILMGDFNLQIDWSTLCSPPAGS KLLELSHHHLTQHVNEPTRYGNILDLVFPDCEIVDL AILDRLGDHNMIEFTLKTGNEFTNASRLKLPDYRRANF QGMATEIETDFTPMRDAVEQCSHILKTSILRAAERYI PNKRVISNDPLWMNRIVKRCIAGDKCVTKPGSVIQQR RTGPTVSPLESQSITLHCRQVRYKAWKRNPTENRS SHIRACRDVKKIVRTTKREKEVSVAAMAKRNPKAFYKY VNERLKKETIGPLKRPDXITNDLRDKATILNDYFTSV FTVENVTNIPEAPMQVAPINTCVFTEIEILQELTHLNVY KSTGPDNLTPRVLEMAPVIVKPLLEFNRSRLTGEVPQ DWKLANVPIHKKGDKKSPSNYRPSLTSVVSKTIERLLK KRITDHLRMHQLISDTQHGFRNGRSLTNLLDFFSKVTD IKDEEKGVDLVYLDFOKAFDKVPHQRLMKKLEAHGIQ GDYLRWIRAWLTGRQQRVKLEGETSDWTQVSSGVPQG SVLGPLLFSIYINDLNLNRSKISKFPADDTKLCHPATKRA ELQQDLHLMEWSTTWQMSFNVDKCSVMHVGLVNPS TPLKMGNGTELAAYTOORDLGITITNKLOHSSOOVDKSVK</p>	cd016 50	113 6	138 3
----------	----------	----------------------	---	--	-------------	----------	----------

GCCAAAG	GGACAAAG	TATGACT	CCTGAAC	TTCCCGA	GTGGCTA	TAAGTIT	GCCTTAC	TATAGGT	GGGATTA	AAAATTG	TGTGCTT	CGGTC TC	AACTTTT	TTTAGITT	TCAAGTA	GATAACA	CCTTTTCA	AGAACGC	TAATAAT	TTTTTTT	CAAACCT	TTCTAGA	GACTAGG	AGGTACA	CTAGCGA	ATAATTA	TTCTCAA	ACTCCTT	ACGTTT	GGATCTT	CATAAAA	TATTATT	TTTATTTC	TTTTTTTT	CTCTCATT	CAITTTCT	TTTCTAA	CAGACTG	TAGGTTT
MASRVLGFTTRNFYKSKDIIVPLYKALVRPHLEFAVQF	WSPHLCKDIDNNMERYQRRATKLPQLRNKPYECLKEL	GLISLEQRRRLRGQLIETFKYLNNGENDVPTTGLFERDLND	RLRNNGQKLKHKARRITLSQQFFPYFIVSIWNNRLPSDVV	SAESVNAFKNRDLIHWERHPPQLRTPPPGVASSELLQE	SQALNGQSFGRGHLNPSSLSRSNYTSLSSAYQRLIVLAC	IHDRLFGWDSSPPSFSEFFLFLPFFSSLSPLVLYILLT																																	

AGTTCAG
AAATGTT
CTACATG
AAATGAA
AAAAAAG
AGCTTGA
AACATAT
GATAATT
GAGTTCA
TTCTCTAA
CATTGCC
AAGATAA
ACAGTTC
GGAGCTA
TAGGCC
GATGTTT
TGATAT
TGATAAT
TACCAAT
CTTAGCG
GCGTGAG
GCAAAAC
AAAGTAA
CACTCCA
GAGTTCC
TGGCAAT
TTTTGTTT
AATCTT
GTTCTCT
ACTCTC
CCAAAC
TCTCAA
CGACTTC
CAATAGT
CTTCATT
TATTCA
AAGATT
TAATACA
TCTTATG
TAGTTTC

TTTTATT
TGACATG
TTAAAT
GCCTAAW
TGTGTC
TGCAATG
TACGTAA
AAATTAG
AACTGA
TCTTTTT
TTTCAI
CTTCTA
GGAACAT
TTGIGIA
CGGTGCA
ACGGGT
ATAATG
ATAAT
ATATA
GTTCTTT
GTTACC
TCGTCAA
GGAGGT
ATGTTTT
CTCGTG
TTGTCTGT
CCCTGT
CTGCCTC
TTGTAAG
TCACCAA
GATTACT
CTAAAC
TAATGAA
CGGACTC
TTATGAA
ATTTACA
GGGATG
TTGAC
AGTAACA

GATGATT
AACCTTT
GGTAAAA
ATCCGT
CACTGT
CTGTCTG
CCTCTTT
TTCGTGC
GCCAACT
CAAAAGG
TTCGGAA
CGGATTT
TTATGAA
GTTATA
TGATTTG
TTGGGTA
TAAACG
TAGAGTA
ATTGACC
GAAATTT
TGGTGAT
ATGTTTG
TTCGACT
CTGTGAT
CTGTG
TTCGTCCTC
TGIGATG
ATATGTT
TGTTCGT
CTCTTTGT
TTCGCT
TGTTTG
ATCTTGA
TATTATG
CTTGTA
GCACAAT
AACTGA
AAGTCG
TGGATGG
ATCTCA

TGAAATT
TACAGG
CATGTAG
TAAATGT
AAAAAG
ATAAAGG
ATAAAC
TTTGGTG
AGGATCG
GTATGTT
GTTTGG
GTGTTGA
ACGGATC
TTCA1GA
AAATTAC
AGGAATT
TTTTTAAT
TCTACAG
AAAGGCA
CAGCTGA
TTAGATT
TTGGTGT
TGATACA
AAGAAAT
TCATGTT
TTCAGAG
CCAATT
GAATTT
TCGTAC
CGTTGCC
TTGCTTG
GCGGAGG
TATGCGC
TCTAGTTT
GTAATGA
ACTTGCA
TGTAGGG
CATCTTA
TTTACTC
CTTATCA

AATTATG
TCCTTTCT
TGCATGT
AGGTTTC
TTTTCCTT
TAATAACA
TTGTAACT
TTACTAT
AGCTATA
TCTAGGT
AGGCTTA
TTTGCATT
TTTTCATT
GTTTCAT
AGGAATT
GCTTGGG
GGCAGGA
TAACACG
TTAAGAT
ATTTCIT
TGTTTTTC
ATCCTTG
TGTGTGT
AATTTAA
CTGCCAA
GTTCTTAT
GTACACT
GTTTAAT
CAGTTAA
GCATATA
CAGTATT
TGCAGTG
TACAGTA
TACGGAG
GTGCTAC
TTTGCAT
GAGGAAA
CTTAATA
ACAAAGT
TCATTAA

[illegible]

GGGCCAA ACGCAGG CACTCAT CGTTGGA GTGGCTA CGCGGGT CGCTGTG GCGCTCC CTTCTTC CTCTCTA CTCTCTC TCCCGCT ATTCTTC TCGCTGT CTCGTGT TTTTTTC TTCACGG ACAGGTC CTTCTGC CGACGAC CTGCCCG CTCTTGT TAGGCAG CCTTTTC TTACTTGT TTCCCCCT CCGCTG TCCCTTT CTCCCC TATCCAT TCTACTA AGCGGTG ATCATCC ACCCCAA CCCCGGC CCGCTC GTCCCTG CTCAGTC TGCTCGC GGAACGT	GDIARGFCQFTATSPALATPPRLARRLLRKRHPHAGF- SPFTPAQVREAIVASGSSTAVGPDQLSILHLRHLGPQGII YLTRLNLSLAHADIPALWKLAVIAILKPSKPADKGSY RPISLLCPAAKVLERLLLPIFEAYQPAPHQHGFRPMHST TSALLPLAFQATRGMINPKPHDRTVALLIDFSKAFDTVP TEGLSSVSGTTLHHNVVRWLAAYLRGAAACCLYNGS RSQFRAVKAGVPGSGVSPSLFNLYVSDFPIPRDQLCYP SVTSYADDFTVVASHRDPLTAAAAALSDHSASVSNWAA NKGLTISTKSNSLWTTDTHQSTLDPGITMNDLSALPLE RRPKILGVTFDTHFTFPHITSTAARASSRLRLKALAGT RWQCKETLLTYRAFVRPLITYAAPVWYPGARQS AVR KLQTVQNAALRTATGCLRMASVDHLHAESKXLRIDQSL DLQCRQYLVSALRPEHPAHSVVTQAAGPRLIRHNLATK FQPDINHLLTDNIMPSPPEYRSALSSLRSAVQASLDRLPV NRVLGFAPPEISPDEAVLSRAARSTLSQLRSGFCSALMS YRHRIGLADSDSCPECHRAPHTVNHLFECPAHPTALLPT SLWDSPLSVTDFCLSLPSFSYLPHPARPPPEPPT				
---	--	--	--	--	--

[illegible]

918

[illegible]

920

CTGTCCC	GTCTTATGT	GELMAAKRSVRRLERWRRTRAESDRRLRVRAELRAYS		
CAGCTC	CCCTCTGGG	VAVLTAKKSYFSTRIASAQCRPAELFRVVRGLLQSGTRD		
AAWAACT	TCCTCTTTTG	EMEEPLAPRCDLFAKHFADKIRIRLDLDSGCCVSGAQN		
CTGGGAA	GGGTC'CTGT	VLSAPPSLSLDQFQLQTEDVDKVLGGVRPTTCSLDPC		
CCTTAAC	ATTTTAAAT	PSWLKSSREGMTGWVQGIINASFRSGMVPAAALKEAVIK		
GCCTGKC	TTTTATATA	PLLKKPSLDPRDLNNYRPVANIPFLGKVLERAVALQLQA		
CATCAGC	TATTGTGAT	FLEETDYLDPFQSGFRPGFGTETALVALVDDLRRREIDRG		
CTGAGC	TGTGTTTTT	SASLLILLDL'SAAFDTIDHGHILLERLAELGIGGTVLQWFR		
GTGCTGT	AAAAATTTT	SYLSGRFQKVELGGCYSTPWRLCYGV'PQGSILSPMLFNI		
TGCTGAA	GTAAGCCG	YMKPLGEVIRSF'GGRCHQYADDTQLYFPVTSEPGVAAT		
TGCCAGG	CCTCGAATA	DLNR'CLEA'VMDWMKSNKLNPDKTEVLWIRRP'SVRE		
TCAGTAT	TTTTTAAATA	CGVVPVLDGVTLP'LPKEQVRS'LVLLDPFLT'FENQISSVA		
CCCAGAA	GAGAGGCG	RS'AF'GQLRLIRQLRPFLNKHDLATVIHALVTSRLDYCNA		
GACAGCT	GGGTAAAA	LYVGLPLKSLRKLQLVQNA'AAARVLTNTRYTEHIMPVLR		
GTCATCC	ATATTTTAA	QLHWLPVCFRAQF'KVLVMTYKALNGLPGGYLKDRLLP		
ACGATTT	ATAAA	YEPARG'LSAGEALLRV'PPLMEIRLTSTRSRAFS'AFAPRL		
AATTCGT		WNSLPRDIRMAPTLALFRRQAKATLFT'RAFIGV		
AATGAAG				
GGGCTGA				
CCTGGCT				
TGTATCA				
CTGAGAC				
CTGGCTG				
AGAGTGG				
AGGAGG				
CGTTGAC				
CITTC'CC				
AATTGTG				
CCCGGCT				
GGGTCT				
CAGTACA				
GCATCAG				
CTGTGGC				
TTAATGG				
TCGGGGA				
GGTGGGG				
TGGCAAT				
TGTC'AT				
CGAGATA				
CCATCTC				

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

923

[illegible]

DFSKTKVLCKPTYQQRFVHEAIQYKRRHNFNREDGY
KLSNHWKDVINQVKS

[illegible]

					AGCACTTTA AATTTCGTT GTACTGGTG ACAAATGAC AATAAAGG ATCTATCTA TCTATCT	AQQLYFLRLRKNNTQKLLVSFYRASIESILTYCICVW YTSCIVAQRKALQRVINTAQKIAGCPLPTLEELHSSRCIK KAQNIKDSHPGHSLFELLPSGRFRFSVKTRNRFKHSF YPTAITLTNTAN			
					CCCAGGT CGACGTA GCCCTSA GAGKCTT GGGACAC ATCCCGG CAGACGT TSCTAGA TCTGAAA TCGCTGC TACCCTC AAAAACG CTAAGAA ACCCATT AGCAACA TCAGCAA GGACGAG CGTKTTG AACTCTC TCAATTA AAAAACA GAGATGA CATAKTC ATCCTCT CAGCCGA AAGGGCA ATGCCAC TGTCATC ATGGATA A	TTNNXIQXLJNDGXYTKINRDPINRLATKXNNIITNST LIDSNKQSLKPSAARPPRFYGXPKIHKEGTPLRPIVSTIN SPTYKLXXFLPETLLPFTGKTTSAXLNSAHLVSLIKDISLS PDDXLVSFDVVSFTKVPVPDIAIEDLVQQGLNPDVPT LVRXCLNTNTYFTWNGTFYKQNEGTPMGSPSPVAVNVF MEKFETDAIQSTLRPSLWKRYVDDTFVWPHGLPALN QFLNHINSIHPSIKFTIMEVEKDGKLPFLDVLEKRPDGS GHSVYRKATHTDAYLKPDSHHHPAQKNSLIKTLHHRAR CISDQHHLASELHVHQALRMINGYSDRTIRKAVSPAMG RQPSDPTNTITLPIYIAGVTKKISKILNKHKIQVRENTVH KIRDLIPSVKDVIPDMQYPGVYKLSGCGSSYIGETKRH VSDRVREHKADLKHGRTNTSAVADHCFNAVGSHEIDFS KTKVLCCKPTGYQRFVHEAIQYKHRHNENREDGYKLS NHWKDVINQVKS	cd003 04	186	270
Pene lope	Posei don- ID_P M	.	Petromyz on marinus						

[illegible]

[illegible]

[illegible]

ATAAACAG
TTTGCCTTT
AGAGGCTA
ACCCCTT
TTCTGATG
ACGACCG
TTGTGTGC
GATCGAAA
CGTCGTGGT
TTTGTITT
TATTATATT
TAATTACTG
AAAGAACCC
AAGAAATA
GTCTACATT
ATAGACTG
CCAGTACG
CTCAGTAC
TTTAATAAC
CATGTCTAC
AGTCTACAT
CATAGACT
GCCAGTAC
ACTACCA
GCTTAAA
ACCAGGT
ACAGTCTAC
ATTATAGAC
TGCCAGTAC
ACTACCA
AGGTTAAT
AACCAGGC
TACAGTCTA
CATTATAGA
CTGCCAGTA
CGCTCACCG
AGGTTAAG
AACCAGGC
TACAGTCTA
CATTATAGA

[illegible]

TTTCTCCA	TAAATATTC	HVTVPEHLQMHVSSSLRYLMYSPSSNVVREAWEDFK
CACCAAT	ATTAGCGTC	DRFRWRVFFTEKLIDSPNDDPDPDYVVPKARKRFD
CGCATTA	TCCCTCACA	DGVRTTPAYIELGLRKGDKFDDYISNVMPVKATRSLSL
CAGCCGA	GGCCTAAA	NLVWLDKDYLSANKYIVSPDKNLGVCIQADWAT
CGAGCGA	CCCCCTGGA	DQTRTLWNDPSNYRKLGSLLLEADEIMEEKHERIFRAAR
CATTTCG	CGCGAGTT	LARRLGKRQLAEFLVSKIHGKDAKTAGYTLGQFYGVPK
CCTCTCG	GTACCCGTT	VHKRPVKMRPIVCHSVSQAPSATYVSKMLKPLVSRQP
TCGGCCC	GAACGGTA	HVIHGSKHLVRQLESQLDKHKRAWVSGDIVAFYPNIP
GCCAGTC	CAGCCAGC	LENCIRVVQMWNRNSPEAEKLSPEERQLFLMSFMTANK
CCTTCGG	AGTTATAG	RLVIQFGDEYAEQIRGLAMGIACSPDLANLYGSHYENPI
GCCTGGC	AGTATTTTG	LARPDMSKFAFFGRYLDLDDVLGIVYASSAEAMSYANE
GGTTGCA	TTGTACAGC	IRYDGVIEWVSVSEYHTPFLDLFLYLDPVKEKRLHWKPYS
AGGGGTA	GTCATTATT	KPLNHRERIPWASHHPKDVKKGTFIGEMSKRLATLCSKLE
TATATCC	TAATGGAT	HYRDAISDLNLYLARGYPDDAVRSWVKAHFSQRWER
CAGTAGG	GCAGTGAG	RLGEPVIRDEVFLKSHFNPAWSAFNVHGLSEVVESW
AAACCGC	CGCCCGCCT	LSSLKLDGANQSQDRTWHAESTDGSSELKIFRQTTLG
TGATGAG	TCGTGGCGT	QFQGEQSSLYPGESRPEGVAGRVGVANVAKPRVTYDPR
TCTTCAC	GGCGCGTTC	DVMDVDELVALSEPSFAHDLVSVVWTDNALRVPKSLRI
GACGAAA	GGTGATCG	RYAMAEELRLDADPQVNPALHVPPLGVEDQVGDLPASD
GCTCAGC	CCGTGATG	AESVAGDDGTNSKPYRICTAKKDSGIVQEDLDIRS VG
GCGACAC	GCCGCGCA	FHDRAWLVSRKKIRNLGDIVNSWKKDIIQAALNSDSVP
AACAGTG	CGTGTCCCA	HMHVDEWQ
TCGGCTG	CGTGTTCGT	
CCTTCGG	CTCGTGGCT	
TTTTCTCTG	GCTAATAG	
TGAAGCA	CGTCTGTGA	
TTCCGCG	TCATGTGAT	
CTCGCCG	CTTCGACCC	
CTCCTA	AGGGCTGC	
CGCTCGC	GGTAACTC	
GTCTCGC	TGGATTCTGA	
GCCCCC	TGCGCATGT	
AAACCT	TAATAGCC	
AATCAAA	GGCCCCGAA	
TCGAGTA	TCCGGGTCG	
AGCACTC	GCTCCGAA	
TTTACC	CTGACTCCG	
	CTAGGTTCG	
	GGAGACTTT	
	CGCCGGGG	

TGGCTGCGT
GCCGGACTT
CGTCCGG
CACTTTGTG
CGTTTGATC
CGATCAGA
TCGAGATCT
CGAGACTTC
GAAAGGTG
ACTTTCGG
TAGTCGCT
GCCCGTG
GGGTGTGT
TGGCCCCCT
CCGCATGA
CCTGTACAT
TCAAAACT
TCATACTT
CGCCGAGA
AAGAGTCC
CTTAAGG
GATTGTGT
CTCCGACC
TCGGGAG
ACGTTCTC
CGCGCAA
TCGCATTAC
AGCCGATG
AGAGACAT
TTCGCCCTCT
CGTCGGCCC
GCCAGTCC
TTCGGGCTT
GGCGTTG
CAAGGGGT
ATATACTCC
AGTAGGAA
ACCGCTGAT
GAGTCTCA
CGACGAA

NeS	Utopi	Phytopht	cd016		
L	a-	hora	50		
	1	infestans			
GCTCAGCG	AGCTCGG		MLADPAALAAAGLARAPPPSPAPQDPSPAFPAGPAGQNP		
CGACACAA	CCTCGCG		RAAAPARVEVHTVAPPGRAGGMLPDPGLVEEPIQATY		
CAGTGTCTG	GCTGCTT		AHDAAQFECALCPV AESMAVLVQHRRSAHRGTRFKD		
GCTGCCTTC	TCCCAGG		IFTSGCQCSLVFYARIV AASHAVACARRNQRAVVPAPTIP		
GGTTTTCCT	CGCCGCC		VAPTRPEATPQPTGYLAAAMTAAAAAASDSTVVAAT		
GTGAAGCA	GACTTCG		NMQSAVPAAAKTTGLQVPPELEPALPQRASCHAGKRR		
TTCCGCGCT	CGCTCTG		RLNADAEVTPCTPTARVSPQTEVAMAPHDAPQDDTVLQ		
CGCCGCTCG	GCGCGGC		REAAEPQDPAA TPGAQVQRVEDTTAAQDDTVQQDHD		
CTAGCTCTG	CCACACG		ADTAQVSPRRRTPTRWGPRPSTQEPSMTGEPAA TLAA		
CGTCTCGCG	CCGCGCG		RRPLTPAA TGTRATRWGPCHRAIGAAAIARLV TGLSTEP		
CCCCCAA	CGAGCCT		AQPQRRQPPPPQEPSQPEPLAAAATAAADIAATVAADI		
ACCTAAA	CCAAGCG		AAAAAAMDDVGGFAADETWLLRFDGACRRNPGPG		
CCCTAAACC	CGCCCGT		GAGAA LFAPSGAVVWTCSHYMPRSRSETNNTAEY TALLL		
CTAAACCTT	TGGCTTT		GVQSAVHHGASHLEVEGDSSLVIAQVKGTFACRNARLR		
AATACCC	CGCAGAC		QLRNRVRHALRSVDTHKLRLHIDRQANAHADRLANRAL		
	GCGAGGC		DQRTSSECGTHGSCMDSCLA VPTALAAQETPPAAPST		
	TCGAGGC		SATPAEGNAMDDIAEIAARDEGETFPVLPIGPGSAPER		
	GACGGCC		QPLRLRLQLSDEERDAADALQELADITMASKIEDADSW		
	CTCGCCA		TSGEGYISSIPERIEVLQPYATAPPQPGRSRRQRRRPPR		
	GCCCCCA		VTRNQREHRLDEALDDMAATQQA TP RDQRAVRRARRR		
	AGACCCC		VGRVRSMAQQELRHEFAKDESKCVAKILKTASTETAA		
	CCCTACG		EDEHPTCPIDAA TLHAHFTGVNAPRTDFDY DATSGREF		
	ATGTGGC		RAAMSDLPATVEIDAADFADLTIDEVEDQLTRAAKTS SP		
	ACCACCC		GHDGIDYGYISRFQAQLVPLLHAVFQFCWRHRRVPLRW		
	GGCAGGG		KVGIVRLJHKKGDPDPROPTNWRPICLOPTYKLYSGLLAH		
		Phytopht hora infestans	cd016 50	919	118 7

CGGCCG	TTTTGGCA	RLSRWLEGNDRPLPMAQKGFRAFGNGCHEHNFMATLLD
CAGGTG	ATCGTCTGC	QTRRQHRKLYQVWYDLRNFAGSLPQQLMWRVLRHLG
CCCGACT	GGCTAGCTT	VDSGFIDRCRIYRDSAFVANAADGATDPVRQEVGVY
CGGTATC	GCTAGCGG	QGCPLSPLLFVAALVPLVRRLEKLDGCVGPLADGVRPC
GCCGGGT	CGGACGAG	TTAYADDLKVFSDSAAIRKCHDTVAGFLAWTGLRAN
GCTACAC	CGGTCTCCG	PGKCSLAVTTNARGNPTRDSSMRLEVHDAAITLSLH
TCTCAGC	GGGGCGTT	ESYRYLGVGDYDHRHRLQLEPKLKQLKREAVALLT
CGCCACA	CACCTTTCC	SGLAPWQVVRALKVYVYPKVEYALRHLRPLQSLQAF
GCTCGGG	CCCGCGAG	DRVVSKGLRHLLSLPRSATSEVLYAPTSSGGLQLPLVE
CCTTGGC	GCCAACTA	LHRALQLAHAWQMLHSHKSDPAIQAVARAQACQVVRKR
GGTCCGC	CACCGATCT	YRLQEDHWGRDDELVRSLNSELAAAPHAELRRNG
CATTGGC	TCTCTACAC	DIASLWSDVQRWLRVYHLRFEHCDETEAHGPLSFRVPH
CCTTGA	TTTTCTAAT	HNKWLTHKTVLRHVKLHLKIRHQTRWKGMDVQDKTV
GCTCGAC	TCGCCTCCG	RVHGGVGAKFMTTGAGLSDDDYRFGVKGRLNQVDTN
AGCGACA	TCTTCGGTC	SVLKRKRRLRAHTTCRDPDTCSSAETLAHVLNHCESNMDA
GCAGCGA	TTCGGTTGT	IRQRHDDALEQIGSKIRGALDRAKSPTLRLNQTVPEYT
CGACGAG	CGGGCTTTT	GAALRPDIVLRNVAAKTMVIADLAVTFEDQAARHSS
GACGCTC	TTCTTTTTG	LQLSHDHKTLKYQPIVAELQHKGWVQTAANVYGTLS
AAGACCC	ACCAATCA	VQPSNFKAYTEKFKLHKREARQLDLQLSSHCIASHRIW
CCACGCC	GAGCGCGC	GWHCRRQHRDRQRSGTASRASRGSGGAPRRTSQAPARR
GCCGCCC	CATGCGCCT	
CAGAAC	CTTCTGGCC	
CCACAGAA	AATCAGAG	
GACGTG	ACCGGGCC	
CGAGTGT	CTGTCTCTG	
GCTTGCC	GACAGCGA	
CCACCCG	GGCCTCCAC	
GCAGGGC	GGCCAGCC	
AGGCAGC	AATCGAGT	
	CTCGGCAG	
	CGACGCGT	
	CTTCTATA	
	GCGCAGCT	
	GACGAGGC	
	CGATCTGGC	
	GGCCCCG	
	ATTGGTCCG	
	ACTTTCGGC	
	CAATCAGC	
	GACGACGA	

GGGGGCGAG
GGGTTTACA
CTTTTGCCC
CGGTTTCGA
CTTCAACTT
CAGGCGAA
AATGGGA
TTTGGACCC
TCCACGCGC
CGTGCCACT
CGTGGGA
CCGGCGGC
GATTCAG
GGGTGCAA
CTTCGGGCA
CGTGTGCA
ACACATGC
AGGCCCA
TTGCGCGCC
AAGCGCA
TCGGGGGA
CGACGCCTC
GGCGGCC
AAGCGCAG
CCCCGCCCT
TCCAGCAC
GACCTCGC
GCCGTTTGG
CGGATCGC
CATCAAGA
CGTGCAG
AGCCAGGC
GGGGTCGG
GCAAAATA
TACTACTC
TAAGTATGC
CCGAATCCC
TGGCTCTC
AGGCTGAA
TGCGGCCC

[illegible]

GCTCGGG	CCCGCGAG	LKSQLQGFDIAIVRGLRHLLRLPENSHDGLFFSPTSAGG
CCTTGGC	GCCAACTA	LGLLSLVELHEALQVAHAWQMLHSDPAIRAIARTQVG
GGTCCGC	CACCGATCT	QVARKRFKLVEEHWRGREDDLQRFNLNTELAASPHAT
TAITGGC	TCCTACAC	ETRRNGDIGSLWINDVRDTLQTLGLKFAAGDEEEAPGLL
CCTTGGA	TTTTCTAAT	QLRVPHHTKWLSHSTVLRHVKLHMKLRMDTWKSKV
GCTCGAC	TCGCCTTCG	SQGTIVREHGGVGSRFITAGAGLSDAEYFAIAIPRAHLI
GGCGACA	TCITCGGTC	DTNSTLKRRRLRANDTCRAPGCSYTEPPAHILNKCSNPM
GCAGCGA	TTCGGCTGT	DAIRKRHDDALERIADALRRKVEKSGGRLEVAINKTVPE
CGACGAG	CGGATTTT	YDGAALRPDIVLRNTETKRAIADLAITHENQPTDATSS
GACGCTC	TTCITTTTG	ALQQSRDNKITKYQTVAAAMMRAGWRVRVTGIVYGS
AAGACCC	ACCAATCA	GSVLPSNFKVYTELLALLKRDARRLNRQLSSHCIASAR
CCACGCC	GAGTGCGC	IWSAHCRRHRERQSRGNASRSGGGAPRRTSQASAR
GCCGCC	CATGCGACT	R
CAGAAC	CTTCTGGCC	
CCCAGAA	AATCAGAG	
GACGTCG	ACCGGCC	
CGAGTGT	CTGTCTCTG	
GCTTGCC	GACAGCGA	
CCACCCG	GGCCTCCAC	
GCAGGGC	GGCCAGCC	
AGGCAGC	AATCAAGT	
	CTCGGCAG	
	CGACGCGT	
	CTTCTATA	
	GCGCAGCT	
	GACGAGGC	
	CGATCTGGC	
	GGCCCCCG	
	ATTGGTCCG	
	ACTTTCGGC	
	CAATCAGC	
	GACGACGA	
	GGGGGCAG	
	GAGTTTACA	
	CTTTTGCCC	
	CCGTTTCGA	
	CATCAACTT	
	CAGGCCAA	
	AATGGCGA	
	TTTCGACCC	

TCCACGCGC
CGTGCACCT
GCTCGGCA
CCGGCGGC
GATTACG
GGGTGCA
CTTCGGGCA
CGTGTGCA
ACACATGC
AGCGCCA
TTCACGCGC
AAGCGGCA
TCGGGGGA
CGACGCCTC
GGCCGCTC
AAGCGCAG
CCCCGCCCT
TCCAGCAC
GCCCTCGC
GCCGTTTG
CGGATCG
CATCAAGA
CGTGCAG
AGCCAGGC
GGGGTCGG
GCAAAATA
TACTTACTC
TAAGTATGC
CCGAATCCC
TGCCCCCTC
CGCGTGA
CCGCGCCC
CATACTTGA
TCTAAGTAT
GGGAGGAT
CCCTGGCCT
CTCAGGCTG
TACCGGAG
ACCCGTAC
GGCCGAAT

941

CTCAGCC	TAGCTTGCT	TAVISPLLHALKRLPGTGVQLSAVDRPGASAYADDLKV
GCCACAG	AGCGGCG	FSDTKDGTTRQHQLVTDFLRW'TGMV'ANPSKCS'TMSVQR
CTCGGC	ACGAGCG	DNRGVLTANLTLQLDGAQIPALGMTEAYAYLGIGDGF
CITGGCG	TCTCCGGGG	DHVRRLVELAPKLRRELKADTTALMQSGLAPWQVVKAL
GTCCGCC	GCGTTCACC	KVYIYPRVEYALRHLRPFQQLQGFDRHLARGLRHLLR
ATTGGCC	TTTCCCCCG	LPTSATTEFLYAPT SRGGGLLPLTEVHGALQIAHAWQT
CITGGAG	CGAGGCCA	LHSPDPAIRRIARVQLRQVADARHRLDAEHWKERGEEL
CTCGACA	ACTACACC	CERLLNSQLGTSAHAPKRRNCIDIGSLWVDVQRHLRSL
GCGACAG	GATCTTCTC	GLQLQTAPADTHTGAPAQPLQLRVPHHDKWLTHKDV L
CAGCGAC	TACACTTTT	RHVKLHIKNNHWRWTSMDQGGKTARAHGGEGSGFL
GACGAGG	CTAATTCGC	TQPRGMWEADYRFVAVAGRLNQVDITYSVLKRRLRSHD
ACGCTCA	CTCCGTCTT	RCRQPGCHRAETLAHVLNHCPTGMDAVRGRHDGALKR
AGACCCC	CGGTCTTCG	IERELHASATDRDRRVELRVNQTVPSLAGPALRDLQLY
CACGCCG	GCTGTCCG	NHTKKTVAVVDLAVAFEEQASDDASSALSJASHKRA
CCGCCCC	ATTTTTC	KYDRIKRHLERQGWKVHLSALVYGS LGAVASGNYQVY
AGGACCC	TTTTTGACC	TTHLGLLKRDAKRLDRQLSVECIQSSRRRIWNLHCSQHRT
CCCGCAG	AATCAGAG	RQHQA RPSQGP RGSRA TETGGTPSQTSRR
ACGTCCG	CGGCCAT	
GAGTGTG	GCGACTCTT	
CITGCC	CTGGCCAAT	
CACCCGG	CAGAGACC	
CAAGGCA	GGGCCCTGT	
GGCAGC	CTTCGGAC	
	AGCGAGGC	
	CTCCACGGC	
	CAGCCAAT	
	CAAGTCTCG	
	GCAGCGAC	
	GCGTCTTTC	
	TATAGCGC	
	AGCTGACG	
	AGCCGAT	
	CTGGCGGC	
	CCCCGATTG	
	GTCCGACTT	
	TCGGCCAAT	
	CAGCGACG	
	ACGAGGGG	
	GCAGGGGT	
	TTACACTTT	

TGCCCCCGT
TTCGACTTC
AACTTCAG
GCCAAAT
GACCAATTC
GGCCCTTCA
CGGCGGT
GCCACTGTCT
CGGCACCG
GCGGCGAT
TCAGCGGT
TGCAACTTC
GGCAACGT
GTGCAACA
CATGCAGC
GCCCATTGC
ACGCCAAG
CGGCATCG
CGGCATCG
CGCTCGGC
CGCTCAAG
CGCAGCC
CGCCCTTCC
AGCACGAC
CTCGCGCG
TTTGGCGGA
TCGCCATCA
AGACGTGC
GAGAGCCA
GGCGGGGT
CGGGCAAA
ATATACTTA
CTCTAAGTA
TGCCCGAAT
CCCTGCCCT
CTCAGGCTG
AACCGGC
CCCACTT
GATCTAAGT
ATGGCAGG

[illegible]

[illegible]

[illegible]

					TTCCCAATT CATTAATTT GAGTCAAA TAAATAT ACTATACTA TACTATACT ATACTA	LDSEVTNFVDMNDHNATSNPDLDKESIAEIIKAIKTKV NKSSGPDLSILNEMFKCYSPLLACISKLFNHLSSGCFPV HWSTSYLVPHKSGHIWDVNNYRGIALSSATSKIFNTIHI RIQHFLDKNNFFQDTQFGRKNHRTTDDNVYILKSLINKY TNKVGKHYLYACFVLDLKAADSVMREGLLFLKLANAGV TGKIIHLIKSMYSNTKYAIKTNQGYCKPFSSSGVKQGC NLSPLFNINYLNDLSESLVQTEPVLHDMFNCLLFADDI VILSSGSLQDSNLNFDYCKKWLTVNIKKTKVIIFN KRGRKINDIFFELGDEQIEVVSSTYTLGITFQNSGNFTVAV NNLVKKAQALGVLKQTLNIIASPINAIKLFRSSVM PILLYGSEIWAANFKPNITPELFIQKLFSDSSKIESIHKFC KQLLGVHKRASNIARAEGLSYPLTEICTQMYKFWLR VKNKECSNIVHCAYQSECTINSNSWANFVSKLLEITRLN ETDHNANIFRNRLKRIYIEKIWLPTLSKKEHNKRIYNK YKSDYEFEPYLNKINLNHRVSVTKPRISAHKLPIEKLK YQKIPSEQRFTLCSQDSVGDHMHALFDCKLDEIVKARS YFWNTLSYNDNIIKLKPKEDMFVHMGIKDSFIANSLGK FNTILQCNT			
					AACAGTG GCTATGC AATAAAC AGTGGCT ATGCAAT GAATAGT GGCTGTG CAATGAA CAGTGGC TATGCAA TAAACAG TGGCTAT GCAATGA ATAGTGG CTGTGCA ATGAACA GTGGCTA TGCAATA AACAGTG GCTATGC AATGAAC AGTGGCT	RGEGNSKKERCIOENAKSGTEENKVYKRLKKQAKK VVAKAMREENEKVLSSELKENPKVYKIVKAMKGDGRD VTGGRCMRGSDGKLNFSERDRSKVWKEHMEKIMNEEN EWDQDVEADLVEGPVERVREEVVKAIRDMKTGKAAG PSEVSAEMIFASGDIGIVIMELCQRLDGRGMPPEWEV SVVPIFKGKDAKSCGAYRGVKKLEHAMKIVERVIVN RIRKLVEVDEMVFQFMPGKGKTIDAVFILRRMQEEYRAK GKKLYMCFVDLEKAFDRVPRKVMWAMRKRGIPEAM VRVMSLYEGAKTRVRVGSSELSEFEVKVGVHQGSVLS PLJFAVVIVDVVTESAREGLMREMLYADDLVLMDDTME GLREFVKWKAFAESKGLKVNLGKTKVMVSGTEEKAL VSKVDPCGMCGRVKVNSVRCVKCGKWHGRCAKIR VTQRLARDFACGRCKNGVGRVVEQEEKLCDESVKE FTYLGDKVSVGGGCEAAVTARARFAWVKFRECCGELLY GKRFFLEMKGIVYKSCVRSALLYGSETWCLSENELGILR RTERAMVRVMCGVKLMDRKITEDLMQMLGLKEAVDR LAKANGVRWYGHVLRREDGHVLRALDLEVSQPRKRG RPKKTWKMQVEESGKVGMRMRDALNRAKWREGVR AIATSLR	cd016 50	188	431
RTE	RTE- 1_SK	.	Saccoglo ssus kowalevs kii		TGAATCCG GCCACCCCTC GTCTTCGGG GACCAAAC CGGATTAA AACTCTGG ACTACTACA T				

				ATGCAAT AAACAGK GGCAGAG GAGACAC AGGCTGG TGGAATG A	TAATTGTAA TACGTATTT TATATATAC CCTTTGATT TGTTTGTAA TTTTGAAAC TTTTTGTGT TTTTGTATA TTTGTGTTT GTTTTTTCT GATCTGTAT CCTTACCAT ACAGCAC CACTGATG GTGCAGTTG TGTAATAA AGTTATTAT TATTATT	INKNKFFKNEMINKLETCDKTDQNEYSLLKQLRCQ QNSLPDSTINSDEWLDHFSKLYNNNNNADFKQSLKL LEDTKLSNHELNNVISYNEIKKAIKLNKKNKSTAEDSISN EMLKTKNSLLPCIQKLFNFFILESGTYPYPSWKSITLVPIHK SGDRSPSNYRAIAISSCLSKLFTGIMNCRLTNHLEKNKL MNEVQFGFRKNRRSTDAILFLKTAVEKTLYGKKGTAK CKRLYSCFVDLKKAFDSVSRDALFLKLQKTGINGNFYNI IKSMYTDQYKVKLNSGITKGFHPTSGVKQGCNISPTLF NIFMNDLPGIFGENVDKPIDMTFPLNCIQYADDLAISES VDGLQCNLKNLDYCTKYLVN NVNRKTKVMIFNNSN RLIKNMFKINNEITSVKFCYLGTIFTPNGKFKINQSN LKNKAMKALFSIRHTILN HKMISTKLSLDFDQLISPILL YGEIWA NEITNENNYLEOACNHFYRHFILGVS KVTPLIGI RGEIGRIPIRIKATLSLIK YWHRLSLLPENHILKSALKETV SSNSSWIKHVKTILGHTD LSEVNKIYKVGNIAELTKLEE KLKNKYIQTWYQQLSHVPKNGSGKNKRLRTYNLFKMEY EPEGYLTEIQDFNLRMMVTKPRLSHDNLEIELGRRKFPFR MEANYRLCKKCNCTGTVEDEKHFILDCCFFTQERNLTLE DCKLEMCSITNRNLFTCIINSSTALAKYLKNTIYKT	cd016 50	150	418
				AAGTTAA ATCTAGT GAGTACT TAGTGGT TAAATAAT GATAGTA ATTAG	GTAGTTTAA TTGTTATAC TAATTTTCA TGTATTTT TCCTCCATT CTTAGTTAC ATGTATTCA	IFHYVCLYPENSVIHNRQDIDIFEKLSQDVIKYSNIGEVL LMGDFNARTGITLADHIINDNSTYMPIDPNYVDDEYNNR NNKDQKVNSFGKKLELKNKNNLLJLNGRTPGDLLGKF TCYNFAGGASTVDYALVSKSFLSNIQYFRVNEINNSDHL CAISLATITRDLSSDNNEIKCTPLPDRYLWDTKSIRYQ DALHSTEIINKLVSNYENNNXNSVTEL TNICNAADNS LKKIHKISSNNKMNKTRQKWFDDDCWNLRKKKFREA AKLVEKYPKDCHIRGQFMVLKKYNKILKIKRKKYKQN LWDSVTSQESNPQYWKNFKIFSKNDTSSANNDIIQAD EWNHFKNLNLLKNSKTKYDSDFGNKIINETREMETN KEINKELNEPITIGEVKAAVKKLKLNKSSGEDKIINELIK YQNYLMEPITAVFNVLKSGTFPKGWKGSIIPIHKKG SRSDPDNFRGVALSCLGKLFTATLNNRLIKYLEKNLY SPYQFGFRNRRTDCLFILKSIVDKYHSEYSKNSKSKK	cd016 50	458	726

949

[illegible]

[illegible]

AACTTTGT	AQNKEEKQEMEELIQERSSVIYKKNKANWMEYGEK				
GAAATAGC	CTKFFENLQYRNATKKMYKCYQKNV/LKLVNDGVTHD				
ACAGTTAC	SPNXILKEEVKYFNMFQAPPSPLETNCMDFFNN				
AAATAGAA	VKLTVQKDCSEGQITEELLDAIGAFKDKTGPLDGP				
ATCAAAC	VEVYKTFDILKGPLACFNHSYINGRLSDTQQEGLISL				
GGATGGAC	LKQDPSGIYKDPVHLKNWRPLTLQCCDAKILAKLAHR				
ATCAGAAA	IKKVLSDIHPNQTGFLHGRYIGDNIRQVLETEIYEISGT				
TAGAGGAA	PGLVFIADFEKAFDKVRLEFYKCLEYFNFGESLIKWVK				
GGACTAAA	VMYSNPRCKIVNNGYISEFKLSRGVKGQCPLSAYLFIA				
AACAAACA	IEMLAVKIRSNNNIKGLEIRGLKTKVSLYADDSFLLKP				
AAAAATAA	QLESGLIEDLDTFailsGLKPNYDKCTILRIGSQKNAN				
CTATTGTAA	FTLPCSLPIKWSGDVDILGIQIPKERNDLTPHFYRKLA				
AATAGACT	KIDKILLPWKGKYL SICGKTLINSLVISQFTYLLMVLPTP				
GTGCTGTGA	SDLLFKLYEQKIFNFIWNGKPKIKRAYLYNEYFEGGQK				
AAATGTGT	LNIALDLKSLKASVIPKLYLNPWVSSKLVRMSPHMFKK				
ATAAGATG	GLFPFIQITPAHFLFEKEIISKISLFFKQALESWLQFQNP				
AATAAACT	PEKTEQIIQQILWLSNIVIDKKKIFEMFKKGIIILVNDIIN				
GAAAGTAA	RTGGVMSTMQLTQTYGNVCSQNYNQNLIAALPQKWKR				
AAGCCGAA	QVEGEKSKELVCRPCIKHEKWLKKSVINKNYQFHLRTK				
GTGTTTATT	KLTAVPYKLQNSWEEIDVPIPWVHMVYELIRKTTTPDSKL				
AGTTTACTC	QIFQKLLYKILATNRMLYIWGIQSSQLCRFCCEEAESLD				
CAATTGGG	HLFWYCPYVARFWSQVQEWLKNCNICLTLQJAILGD				
GGATCGGT	LKSHSQSNNIIILAKMFIFNLQSVEAMRIERFNCVKKHH				
GGTAGGGT	STVEKYMANNRPKWMMLRDRWEGLNAGEGWD				
TTGCGGGG					
AATAATAA					
TAAAGGTA					
TATTCTTTA					
AAAAAGTA					
TGTATGTCT					
ATATAGGT					
ATGTGTATG					
TATATATGT					
GTATATGTA					
TGCATGCGT					
GTATGGAT					
ATATATATT					
TACCCCAA					
AAATATAT					
GGGGGATT					

[illegible]

ACATACA
CACACAC
ACCTGAT
CTCGCTC
TCTTCTCT
CACTCTC
TCTTCTC
TCTTCTCT
CTCTCTCT
TCTCCCTC
TCTCTCTA
TTGTGGA
GAACGTG
TTATAAA
TTTAATG
TGTTTAAT
TTGTGCT
ATCTTTT
TATGTAT
TTGTCTA
CAAGTTT
ATATATG
TTTACAA
CAAGCAA
GGGGAAG
ATATCAG
AATAGGG
GCATTGG
GGAATAA
TAACATA
TTGTACG
GAATAA
TTTGTACT
GTAGTGA
AGCGTCT
TCTAGAG
TAATGCA
AGGTCTC
TTTCATG
ATCATGT

GAAACGT
CAATTGC
TATGGAA
ATGTGG
GATGTGT
TTTTCAT
TATGTA
AAGGTAA
TTATGAT
GCAACTA
TGATGGT
GATACGA
TATGGAT
TACAAAT
ATCATAT
GAATATT
AAGGCTA
TACGCAC
TACATGT
TACACAC
ATAGACA
CTCAAAC
ATGCAAA
TTGGCAG
AGTTATT
ATTATTT
GGACATC
ACACATT
ATAGTCT
TACTCTT
ATACAGA
CATCATA
CACACTC
TCATTAT
ATCATAT
CCAATAT
CATACAC
ATCAACC
TACTCAG
ATTGCT

	L1- 5_SS a				ACACACA TAATATC TTGTCTC AATTTCCT AACATCT GTGGCAT TGGCTCA CAGGCAA ACAGGCA AGGAAT AAAACAA ATATTGC TAGAACC TATTTCCT GAATTCCT AAATATC AGACACT TGAAAGC TCTAGTTT TGACCTT CATAATA CCTCTGA TGGCAGT AACTTTA CAAGCAC TAATTAT GGTCAAT TTAGACT GAGAATA GTATCTA GTT	TGAAATTGC ACTTTTTAA CGCACTCTC AATTGTMA TATGGTAAT CTACCTTTG GGCAGCAT GTGCTGGCC CCGCCACCC	WFGALSYLFSDLHAMQATIGTTNERRAGLTFTSWNVKG LNEPIKRGKVLAHLKALSSDIIFLQEETHLKNNSHNRKC KWVGQVYHSNFSAKTRGTAILVRKGIPFLHKTIVDKE GRYVIVIGEIHSTSVLLNIYGPNIDNPFFKRVLALIPDIS HTNLVIGGDLCVLDQYLDRSSTRRTPTSYSSEFLNTYI KNSNFLDWIRISNPTGREYSFYSHVNVTYRIDYFLVDA KLLPYTCNVRYHDIIISHDSPLTFSLRLGDIVPNERVVRL NPQLLTEPTFCEHLKDQITFFFDTNDNTETSPALLWETLK AYLRGCIISFOTARSRONRGELVELEGQHLLDRENASHP	cd016 50	535	793
--	------------------	--	--	--	--	---	---	-------------	-----	-----

TCC TTGTT	GAGCAGTT	SMEKHKKITSLKFEYNQILSAKIAKSFLYAKQKYFEFGD
CCSATCTC	GGGAGGGG	KPHKLLARQLRKNVSDRMHKKVKSASGELLSSPKDIND
CTAACCC	AATAGGGT	RFRQFYETLYTSKADPNPLMQNFLEDCNLPALNQEDSN
ACAATAT	GGATAAAG	FLNKEISLEEIRETIKSLKSGTKPGDGYPGEFYKTFNSM
GTAATTG	GGGGGGAT	LSPYHKMLVQANEDGALPSTLDEAFITVHKKGKDPEE
TACTGCT	AAAGGGGG	VGSYRPSILLNDQKILAKTLANRLSTLIGKLVHSDQTGF
GTCTATG	GTATAAGT	IPNRNSFFNLRLFNIMYSQRLPNVDLAVISLDAEKAFD
TCGGTTG	GGGGGGGT	QVEWPYLFKVLQKCNIGDGFHKWQIQLLYRNPCARILTNO
CTGATGG	TACATCGAC	SLSPRFNL YRGTRQGCALSPMLFALIEPLAQAIRSHAAI
TTTATGTT	CCTTTTTC	HGYN TKDITLNKISLYADDILLYVTEPQDSIPAILXVINLF
TAGACCG	TTTCTACCT	GTFSGYRINWNKSELMPRLQNTSWLEHLPVKLSSEKFT
GCAATTGC	GTCC TTGTT	YLGIVVTQYSLLFKENFPSLMQKLKANIQFWRTLPISL
TTAGCAA	CTGTATGTC	LGRINA K MVFLPQLLYLYQNIPVFIPKSFHKQLDSIINPFI
TGTAATC	TTGTTTGT	WDYKTHRIGKKHLCKSKMEGGLSLPNFIFYWAAANLR
TTGTGAT	AATATTGT	AVTFLLDDAPPSSWLSMERECHPFSIGAVILSPVNLEM
GCTATAT	TTTGTGCCC	SLYRNNPIHSTVRIWKQIKVHFELRPISEMLPVARNPSFA
CTTGCTT	AGAACTCT	PSNLDNTFERWGEIGISTIGDLYIGGTFFASFESLRETYNL
ATCTCAG	GGGTCTTTT	PRSNFFRYLQIRDYVRKHLPTFGNAKPSMDFGCIKICPTS
GGATTGA	GGTTAATGT	DKLISRIYDAFQSVSAPSTDAIKAKWEEELGTDISVADW
CCCAAGA	CTGTCTCT	EDSLEYHTCSINSRHLIQFKVLHRLHYSKTKLHRIFPG
TGAMGCT	TATGTTTAG	TSPMCDKCAEQQTLLHCFALCSNLYGYWCGIFGVLS
TAAGTGC	ATAAGAAC	VLETSIDPDLIILGVSDSLNGLTNPQKQLISYSLISAKK
GCAATAT	TGTAACCT	LILFWKKREAPSTKLWLSELANTVHLERIRYILNNKLS
GIITTAAG	GTCTTTTAT	TFDQIWWQPFSLYLDQAAL
TTGCAAC	ACCTATACA	
TTATTCTG	CCATTCTTG	
TTTAGGT	TGTAIGTGT	
TTAATAG	TCAATAAA	
ATGCTAA	AAATATTG	
TTGAACA	AACAAAAA	
CTTTAAT	A	
CGCGGGA		
GCCTCCT		
CAGTTCA		
GATGTTA		
GTAGAAG		
TTCTAGG		
ATAGTGA		
GAGGTGA		
ACGTACA		

[illegible]

				TTAACGG TTTACAM CGACAGG AGCTTTG TATTTCAT TCATTTT TACSAGG AACG	CTGTAGCTT GTGAGGTG TGGAAACA CTTGTGTG TTTTATGGA TTTTGTCTT GTGTCTTT TGTTCATG TTGCTCTGT CTGTATGCT ACATCTTGC TTGTCTAT GTTGTCTG TCTGTATGC TATGTCTTG CTGTCTCA TGTTACTCT GCGTGTGCT CACTGTTCA ATGATTGTC TATATTGTA ATTGTTTT AATAACCT GCCAGGG ACTGCGGT GAAAATTA GCCGGCTG GCTAAAC CGGCACCTT TACTGAAA CGTTGATTA ATGTGACT GTCCCTGTA AAAATAAA ATAAACTC AAACTCAA	SNIPKPLYAIPDFASGTQAVLISSPHPVQAGGIVNSATELP LIKTKQGNLNPVRKHKKCNFFQCVNHSRVWIDPRAKP KGLLGGHLNIRSVIPKSDQIQHLLTDSNLDLCLSETWL HKNSPCAALIVPGYNVRRDRIEGRGGGLMIYKEHIRC KQIEWSCDNELEICGLNVTLSPQMSFTLIGMYRPPSTKS VFEDQNNMLRECDGKEVILMGDFNINYEDKSCRKTL KRITNTFDLTQLVKGPTRVTCCKTQIDLVESNKPervT KSFNNMVTGLSDHNLTLJARKLSKRPKLSVTRKPDQLRI PKSELNYFENAINGNWDLLSYTDVEADSEVFLSTIQA TINGFLKKIKSPGQKSTLPWLNGEIKWMKERDYALK TALKSKLEHRRRFTMLRNKVMKEIRQAKANFFINIIGE AKGNSKLIWENLKKLTGKDHSTAKRLEIMVNNLTQ DAVEIAIAFNSYFIDSVRVLTQNPSTGLLSVLVNDAAQP VFIREVSESKVNKVISLKNKSKSDVFGLDSTFLKNYKE SLIGPTKVTNTSIGLGVFPRVWKSAITAIFKSGDPADVS NYRPISILPVVSKVVEKCVAEQLSHLNNSPFTLHSMQFG FRAKHSTETANCFLLENVSKMDKGGVVGAVFLDLRK AFDTVNHEILTKLSKFNFPDALRWKMSYLEGRTQCVR VSNELSPTLSYDVGVPGQSILGPLFSLYINDLPSVCTGS EVQMYADDTVIYVHAKSKQAAQELJTMVMVQVTKWL SDSCLHLNVKKTVCMFFTKRATDATEPDVYVSGEKLQV VSDFKYLGHILDSNLSFKKHVKVQITKFNLANFRFIRN CLTTEVAKLYFKSMILPHLTCLTSWAQAQCYTLKPIQS VYKQALKVLDKRPNSHHHCYILRKHELLSWENLVQYT DACLVFKILNGLAPPLSIFVKQKTQTYGSRSTRSAMRG DCIVPLRKSTFSKSAFVRASHVWNTLPSDTHNCTTYHT FTKCMKTWLKVNQTCENHP			
Rex	Rex1-4_SS a	.	Salmo salar	AGATGGC GCCGACA GACATGG CAGCTCT	TAGCCGGCT ACCACCCG GTTACTCAA CCCTGCACC	LLDIRAAVTHQHYDQEYDFPELPLFVPPRAIELIPEAAP RRRRRRGIRSGLLVRLRRRAHHPPLPSILLANVQSLDN KVDELRARISFQDIRDCNLCFTESWLSPDILSPSIQPA FSVHRADRNKELSGKKKGGGVCFMINYSWDCDCDNVQK	cd016 50	539	848

GCTTCTA	TTAGAGGCT	LKSFCSPDLEYLTIKCRPYLPRFSSVIVTAVYIPQADT
GCTCCTA	GCTGCCCTA	TTALKELHWTLCKLETTYPEAAFIAGDENKANLRKTL
AGCAACT	TATACATAG	PKFYQHIDCSTRAAKTLCHCYSFPERDAYKALPRPFKGS
TTGCAGT	ACTGGAAT	DHDSILLPSYRQKLKQEVPLRSIQRWSQSESMQLDC
ATTTGTTT	CACCTGGCC	FDHADWDMFRVASENNIDVYTDVTEFIRKCGDVVPT
TTTTGTGT	ACTTTAATA	VTIKTYPNQKPWIDGSIRAKLARTTAFNHGKVIGNMV
GTTATTTC	ATGGAACA	EYKQCSYSLRKAQAKRQYRDKVESQFNGSDTRRMW
TTACATT	CTAGTCACT	QGLQITTDYKGTSHVADTDVLLPDKLNTHFARFEDNT
ATTAGCC	TTAATAATG	VPPTRAPKDCGLSFSVADVSKTFKRVNPRKAAAGPDGIP
CAGAACG	TTTACATAC	SRLRACADQLAGVTFDIFNLSLSQSAVPTCFKMSTIVP
TTTTTTGT	TGCTTTACT	VPKKAQVTELNDYRPVALTSVMKCFERLVKDHITSTLP
GTTATTA	CATCTCATA	DTLDPQFAYRPNRSTDDAIAIAHTALSHLDKRNITYVR
CATACAG	TGTATATAC	MLFIDYSSAFNTIVPSKLIJLKGALGNPALCNWVLDFLT
CCGGAAA	TGTATTCTA	GRPQVVKVGNNTSLILNTGAPQGCVLSPLLYSLFTHD
TAA	TTCTACTGT	CVATHASNSIIKFADDTTVVGLITNNDETAYREEVRLA
	ATTTTAGTC	EWQENNLSQLRQNEGADRGLQETAEAGARPYPHRRD
	AATGCCACT	RSGEGLQVPRRTHHGPWRSRGARKITSPSNVNKTKELI
	CCGACATTG	VDFRKQQRHAPIHIDGTAVEKVESFKFLGVHITDDLK
	CTCGTCCTA	WSTHTDSVVKKAQQRLFNRLRLLKFLGAPKLTNRYRC
	ATATTTATA	TIESILSGCITAWYGNCTARNRRALQRVVRSAQRITGGT
	TATTTCTTA	LPALQDITYSTRCHRKAKKIIKDINHPSHGLFTPLSSRRRG
	ATTCCATTC	QYRCIKAGTERLKNFSYFKAIRLLNSHH
	TTTTACTTT	
	TAGATTGT	
	GTGTATTGT	
	TGTGAATTG	
	TTAGATACT	
	ACTGCACTG	
	TTGGAGCTA	
	GGAACACA	
	AGCATTTCG	
	CTACACCCG	
	CAATAACA	
	TCTGCTAAA	
	TATGTGTAT	
	GTGACCAA	
	TAAAATTIG	
	ATTGATTT	
	GAG	

Rex 1	Rex1- 3B_S Sa	.	Salmo salar	TAGTCACTT TAACCATAC CTACATGTA CATACTACC TCAATCAGC CTGACTAAC CGGTGCTG TATATAGCC TCGCTACTC TTATTTTCA ATGCTTTT TACTGTTGT TTTATTTCT TTACTTACC TATTATTAC ACACATACT TACCTTTT TTCCGCACT ATTGGTTAG AGCCTGTA AGTAAGCA TTTCACTGT AAGGTCTA CACCTGTG TATTCGGCG CACGTGAC AAATAAAC TTTGATTTG ATT	CFCHRLRQKRASGYQSDSHSPRIQRFFLQQQAGLTG YSPNTXQGRHPRYSQEEATRRRRRAGCLVTRRRRVGK LPLPSVLLANVQSLDNKLDEVRISYQRDIKNCNILCFT ESWLNDMDIQLAGYTLHRQDRTAHSKTRGGGLCIFV NNSWCTKSKEVSRCFSCPEVEYIVINCRPHYLPRFSAILF VAVYLPQTDAGTKTALSQLYKDISKQENAHPEAALLV AGDFNAGKLSVLPNFYQHVKCATRGKKILDHLYSTHR DAYKALPRPFGKSDHNSILLIPAYKQKQKQAPVTRSR KWSDEADAKLQDCFAXTDWNMRDSSDGIEEYTTTSVT GFINKCIEDVVPVTVTRTPNQKPWITGNIRTELKGRAA AFKERDSNPEAYKKSRYALRRITIKQAKRQYRAKIESYY TGSDARRMWQGLQITTDYKKGKHSRELPSDTSLPDELNH FYARFEASNTACMRAPA VPDDCVITLSVADYSKTFKQ VNIHKAAGPDGLPGRVLRACADQLAGVTFDIFNMSLTE SVIPTCFKQTTVPVPKNTKATCLNDYRPVALTSVAMKC FERLVMAHINTIIPETLDPQLQFAYRPNRSTDDAISIALHTA LSHLDKRNTRYVRLFDYSSAFNTIVPSKLITKLRLGLN TSLCNWILDFTGRPQVVRVGSNTSATLILNTGAPQGCV LSPLYSLFTHDCMARHDSNTIHKFADDTTVVGLITDND ETAYREEVRDLAVWCQDNNSLNVTKEMIVDYRKR RAEHAPILIDGAVVEQVESFKFLGVHITNKLWSKHKT VVKRARQSLPRLRLKRFMGPEILKRFYSCTIESILTGCI TAWYGNCSASDRKALQVRVTAQYITGAKLPAIQDLYT RRCQRKALKIVKDPSPSHRFLSLLPHGKRYRSKSRTK RLLNSFYQAIRLLNRDOKAPOQLPPSHKTPQVQIKWL PGLFALCAPPSSTLCLSYMHSFNNTFMYILPQLARPTS APAHWLTGLSALCPTTRQPLFLRCCYSLFIYA	cd016 50	549	801
Crack - 2B_S Sa	Crack - 2B_S Sa	.	Salmo salar	TGACCTAG ATAGTTTGT GTGTATGCA TTGATATGT AGGCTACG TGTGCCTTT TTAATAAT GTATGTAGT TCGTCGAGC TGTCTTGT CTAWTGAT	MFFSLALFALFHIMSISGKLPKGLKIAHINICSLRNKVH EINNLLTSDNIHILAISETHLDNSFDDTSVAIQGYNIYRRD RNAYGGGVAVYIQSHIPVMLREDLMSVIEVLWLQVHL AHLKPELLGCCYRPPSANSQYLLNMCEMLDSVCDVNR EYVFLGDLNIDWFSSCPLKRLKLLTVTSACNLVQVINQP TRVFINTTGRSSTCIDHIFTNAVELCSKAVSVPIGSDH NIVAISRKAKVPTAGPKIVYKRSYKRFCCDSYVDDVKNI CWSDVINEEHPDAALDEFMKLLLPIDKHAPVKLLTVRT VKAPWIDELKNCMVVERDGAAGVANKSGCTSDWLTTC KLRNYVTKLKKKKLLYYEAKINDIKNDGKKKTLEYFK XNYGQKDKFNSIFHRIRWLIHHKTIRMMMEKLLWSTLNX	cd016 50	557	816

Rex I	Rex1-3D_Sa	Salmo salar	AAC TTGW GAGGAAC GTATATC TTCTCA ACGAGGA CTGTCTT GAAGCTG TGCGCCA GAAGAGG AAAGAGC TGATCCC AGCCATG AAAGCTG CCAGAGT GCGTGGG GACATTG CKTACAT CCGCTAC GACAGGC TCATTGT CCACCTT CCCTCCC AGAAAGC TGGAAAG GATGAGA GAGCCAA GCCTATG GGTTCGT AGCCTCA ACCCAC ACACACA CACACAC ACACACC AATTGAT TAATGGA CTGCTGA ATGT	GTTCTGTAT TATGTCATT CTGTATTAT GTTTCATGT TTTTGTGTG ACCCAGG AAGAGTAG CTGCTGCTT TTGCAACA GCTAATGG GGATCCTA ATAAATA CCAAATAC CAAA	IMGRKTNSTPSFIESDGLFITKPFDVANYFNDYFIGKV GK LRQEMPTTNSEQLYSCIKKQIMKEKQCKFEFCKVSVGE VEKLLSINNDKPPGIDNLDGKLLRMVADSIATPICHIFN LSLEESLCPQAWREAKVIPLPKSGKAFTGSNSRPISLLP ALSKLEKIVFDQICYFSVNKLTTDFQHA YREGHSTCT ALTQMTDDWLKEIDNKKIVGAVLLDFSAAFDIIDHNLLL KKRMCYGFSTSAISWISQVLSNRTQRVFFNGSFSNVKHV KCGVPQGSLSGLPLFSIFTNDLPLALNKA CVSMYADDST IYASATTANEVTETLNKELQSVLEWVASNKLVLNISKTK SIVFGTNHSLSSRPQLNLVMNGVAVEQVEETKLLGVTL DCKLSWSKSHDSMVVKMGRGLAVIKRCSAFLTPHSHKKQ VLQALVLSNLDYCPVVWSSAARKDLVKLQLAQNR AAR FALNCNQRADINTMHASLSWLRVEERLTASLLFIRNIN VLKIPNCLHSQLTHSSDTHTYPTRHATRGLFTVPKSRTN SRKRTVL YRALJAWNFLPSHIAQINSKPGFKKKQIKQH LA AQLSPI	cd01650	372	624
-------	------------	-------------	---	--	--	---------	-----	-----

ACGTTT	GCATTGCC	AYKQKLKQAPVTRSIKKWSDEADAKLQDCFASTDWN
ACGTGTC	CCCTCCTCC	MFRDSSDGIEEYTTTSVIGFINKCIDDVVPVTVTRTYPNQK
CCCAACC	ACACCACT	PWITGNIRTELKARAAAFKERDSNPEAYKKSRYALRRTI
GATTGTG	GCCACTCTC	KQAKRQYRTKIESYYTGS DARRMWQGLQTTIDYKGGH
TTTTTTTG	TGTTGTGAT	SRELPSDTSLPDELNYFYARFEANN TETCMRAPAVPEDC
TTTCGTTTA	CTATGCATA	VITLSAADVSKTFKQVNIHKAAGPDGLPGRVLRACADQ
TTTGCGTT	GTCACCTTA	LASFVDFIDFNLSESVIPTCFKQTIVVPKNTKVITCLND
TTTTGTA	ATTAACCTCT	YRPVALTSVAMKCFERLVMAHINTIIPGTLDPQFA YRP
ACTTATTT	ACCTACATG	NRSTDDAISIALHTALSHLDKRN TYVRMLFIDYSSAFNTI
TTTTACTT	TACATACTA	VPSKLITKLRITLGLNTSLCNWILDFLTGRPQVVRVGNT
ATTTGT	CCTCAACTA	SATLTLNTGAPQGCVLSPLLYSLFTHDCMARHDSNTIHK
ACATAAT	ACCGGTGC	FADDTTVVGLITDND E TAYREEVRDLAVWCQDNNLSL
GTGCGCG	CCCCGCAC	NVIKTKEMIVDYRKRTTEHAPILIDGAAVEQVESFKFLG
CTACCGT	ATTGACTCT	VHITNKL TW SKHTKT VVKRARQNLPRLRRKRFMGMPQ
CTCTTAT	GTACCGGC	ILKRFYSCITESILTGCITAWYGNCSASDRKALQRVVRTA
GACCGAA	ACCCCCCTG	QYITGAKLPAIQDLYTRRCQRKALKIVKDSHPHSHRLES
AATAACT	TATATAGTC	LLPHGKRYRS AKSRSKRLNLSFYPQAIRLNI
TCGGAC	TCGCTATTG	
ATCAGGA	TTATTTTAC	
CTGCGAT	TGCTGCTCT	
TACTCAC	TTAATTACT	
CACGGAC	TGTTACTTT	
TGGCAGA	TATCTCTTA	
ATCCTTT	TTCTTATCC	
TTTCCTTT	GTATTTTIT	
AACGAGT	TAAACTGC	
CTGACGA	ACTGTGGT	
GCCCGAC	TAGGGGCT	
GCGAACG	CGTAAGTA	
ATATACT	AGCATTICA	
GCTTTCTC	CTGTAAAGGT	
GGGAACA	CTACACCTG	
GGCCACG	TTGTATTTCG	
ATCCCCG	GCGCATGT	
TGATTTG	GACTAATA	
CGTGAAG	CAATTIGAT	
AGGAGGC	TTTTGATT	
GGAGAAA	GATTTG	
AAGGGGC		
CAGAGGG		

CGGGCTG
CCTTCTG
AGAAATC
GTAGGCG
ATGCAAT
AAACCCC
CACTTCC
TTCCATT
TGCTAGC
AAACGTG
CAATCTT
TGGASAA
TAAATA
GATGACC
TACGGCG
AAGATTA
AACTACC
ACGGGA
CATCAA
AACTGTA
ATATCTT
ATGCTTC
ACGGAGT
CTGGGCT
GAACGAC
GACACTA
TCAACAT
ACAGCTG
GCTGGTT
ATACGCT
GTACCGG
CAGGATA
GAACAGC
GGCGTCT
GGTAAGA
CAAGGGG
CTGGGA
CTATGTA
TTTTTGTA
AATAACA

[illegible]

966

CGACCCG	TCATGCTCT	AEQRGKLLRASALLQSSCTIRSSLLAFLQNESTLMDL
AAGGCTC	TGTGTTAAT	EHVSTFLNSHAEVFRPTWTPSRPKHPSHAPANVSRKKR
AGAAACT	TGACTGCTC	RKIEYAHIQRLFHHRPKDASNTVLDGRWRNPYVANH
ACCCACT	TCCTCTGGG	MIPIDFCFWTTVTFTKTNPSDSREITPIPMTPSLDIPILPSD
TCTTGAA	TTGACGTCT	VTWALKEMHGTAGGIDRLTSYDLMRFGKNGLAGYLN
CTGCTAC	GATTGCTC	MLLALAYLPTNLSTARVTFVPKSSSPVSPEDFRPISVAPV
TTTTTGCT	TCCTCCTT	ATRCLHKILAKRWMPFLPQERLQAFNLNRDGCFAVNL
GTTTATC	CCATATTGC	LHSVIRHVHTRHTGASFALLDISRAFDTVSHDSIIRA AAKR
CACAACA	TTGCTCTGC	YGAPELLCRYLNNYYRRSTSCVNRTELHPTCGVKQGD
ACAGTTG	CCGCTTACT	LSPLLFIMVLDEVLGLDPMTHLTVDGESLNYIAYADDL
TGATTCT	TCCAATAGT	VVFAPNAELLQRKLDRIISILLHEAGWSVNPEKSRTL DLS
ATTCCTC	TGTCATAAT	GGHSKITALSQTEFTIAGMRIPPLSAADTFDYLGIKENFK
ANATAAT	ATGCTCTTG	GRCPVAHIDLNNYLTEISCAPLKPQQRMKILKDNLLPR
CCTTGIG	TTTACTTGC	LLYPLTLGIVHLKTLKSMDRNIHTAIRKWLRLPSPDPLA
CTTTTGTC	CATGCTCTAA	YFHSPPV AAGGLGILHLSSVPFHRRKRLETLLSSPNRLLH
AACATTA	CGACAAT	KLPTSPTLASYSHLSQLPVRIGHETVTSREESNSWVRR
TTCTATA	ACTTTATCT	LHSSCDGKGLLLAPLSTESHAWLRYPOSIFPSVYINAVK
CCAACTG	ACCTTAGTT	LRGGLSTKVRRSRGRVTNGLNCRGGCAHHETIHHIL
TACCACC	TGCTCTCTT	QHCALTHDIRCKRHNELCNLVAKKLRRQKIHFLQEPICP
TACTTCTT	GGTTTCGAT	LEKTYCKPDHIIIRDSIAYVLDVTVSDDGNTHASRLKIS
CATCTCA	TGCCTTCAT	KYGNERTVASIKRFLTSSGYITTSVRQTPVLTIFRGILDR
CGTTTTA	ATGTTTCATG	ASSQSLRRLCFSSRDLGDLCLSAIQGSIKINYNTYMRGT
ATTCJGG	GCGGAATC	
TCCTATT	TGATGTTTA	
TCCTATC	TAATGACTA	
ATTAGTC	TTCTTATTA	
ACGGAGA	CCACCACTA	
GGGCCTA	CAACTACTA	
TGAACGG	TTATTATT	
TCCGTGA	TCATTACTA	
CGCGAAA	TTAACACTTA	
TTCTATCC	TTATAAACA	
GCGATT	TTATTACTA	
CGACCTC	TTATTATTA	
TCCTGCT	TTACTATTA	
AGTGGTC	TTACTTCTA	
CCCGAAG	CAATTAATA	
TACGGTT	TTATGGCTA	
CCTCTGG	CTCCTCTCA	
CCTGTCA	GCACACCA	

GTGTGT TAAACT ATATAAT AACG	ATAAAATA TCAATCAA ACATCTCAA TTATATCCA CCTATTAAA CTCTCTCTA TTTCCCTCG AGTTATAA ACTTACAAT TCAGTCTAA CCGAATATC TCTCTTTTA CAAACTCTA AGTATGTA ATTTGTGC CAAAACCA TTTGGGTCT GTACAAATTT GATACCTAA AAATAAAT GTTATTAGC C	MPVSTGAETDITSSLPASSIVSPNYTL PDSSTCLICFAI FPTHNILLSHATAIHHSICPPTPVQDGSQMSCVLC AAAF SSNRGLTQHRRHISEYNELIRQRIAVQPTSR IWSPFDD ASLLSIANHEAHRFTKNDLYQHISTVLT RRTAEAVKRR LLHLQWSRSPTAITSSNNHTTIDIPNTE ARYIFPVDLDE HPPLSDATTPDASTHPLPELLVIL TPLPSPTRLQNI SESQT SHESNRNSMHTPTTYACDSD ESLGVTPSS TIPSCFHSY RD PLAEQRSKLLRASASLLQSSCT RIRSSSLLAFLQ NASTLM DEEHVSTFLNSHGEFVFPRTWTP SRPKHPSHAPAN VSRK KRRKIEYAHIQTLFHHRPKDAANT VLDGRWRNPY VAN HSMIPDFDCFWTTVTFTKTN SPDSREITPI PMTPLSDPILP SDVTWALKEMHG TAGGIDRLTSY DLMRFGKNGL AGYL NMLLALAYLPT NLSTARVTFV PKSSSPVSP EDFRPISV AP VATRCLHKILAK RWMPLPQERLQ FAFLNRDGC FEAVN LLHSVIRHVHTR HAGASFALLD ISRAFDTVSH DSIIRAAK RYGAPELLCRY LNYYRRSTCV NRTELHP TCGVKQGD PLSPLLFIMV LDELLEGLD PMTHLTV DGESLN YIAYADD LVVFEAPNA ELLORKDL RISLLLH EAGWSIN PEKSR TDLJ	cd016 50	492	749
-------------------------------------	--	--	-------------	-----	-----

CTTTTGTC AACATTA TTCTATA CCAACTG TACCACC TACTTCTT CAJCTCA CGTTTCA ATTCTGG TCTAATT TCTCATC ATTAGTC ACGGAGA GGGCCTA TGAACGG TCCGTGA CGCGAAA TTCAATC CACGAAT TCGTCTT CTTCTGCT AGTGGTC CCCGAAA TACGGIT CCTCTGG CCTGTCA GTTGIGT TAAAACT ATATAAT AACG	GTTTACTTG CCATGTCTA ACGACAAAT TACTTTATC TACCTTAGT TGGTCTCT TGGTTTGGT TGCTTTCAT GTGTTTCATG GCGGAATC TGATGTTTA TAATGACTA TTCTTACTA CCACCATT 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 ATTGATAC	SGGHSKITALSQTEFTIAGMRIPPLSAADTFDYLGKSNF KGRCPVAHIDLLNNYLTEISCAPLKPQORMKILKDNLLP RLLYPLTLGIVHLKTLKSMRNIHTAIRKWLRPSDTPL AYFSPVAAGGILHLSSVPFRRKRLETLLSSPNRLL HKLPTSPTLASYSHLSQLPVRIGHETVTSREEASNSWVR RLHSSCDGKGLLLAPLSTESHAWLRYPSIFPSVYINAV KLRRGGLLSTKVRRSRGRV/TNGLNCRGGCAHHE/THHI LQHCALTHDIRCKRHNELCNLVAKKLRRQKIHFLQEP PLEKTYCKPDIIRDSIAYVLDVTVSDDGNTHASRLKIS KYGNERIVASIKRFLTSGGYIITSVRQTPVLTFRGILERAS SQSLRRLCFSSRDLGDLCLSAIQGSIKIYNTYMRGTQRLN E				
--	---	---	--	--	--	--

[illegible]

TTTAAAGTTT
TATATAAA
AAAAATTAA
AGTTTATAA
ANATTAAA
TATATAAAT
GTTAAATTA
AACAAAAAA
AGAGAGGGA
AAAAAATTT
TTTATTAAA
AATTTTAAA
TTTGATATT
TTTATAAAA
GATATTTAA
ATTTTAAAA
ATTTTAGGT
TTTAAAAATG
TTTAAAAAT
AAAAATTAA
ACCAATCA
AAAAAGCAA
AAAAAGTNA
ATTTTAATG
AAAAATCAG
AAGGTATTT
CGGCATTT
TTCTCAATGT
ATATAGTA
GTTCCGAA
AGAAAACTG
AAGAGAT
TATTTAATA
ATCGAAAC
TTTTTTTCAA
AAAAATAA
TTATTACTT
CGCAGCAA
GGGCTTAT
TTTGTAAT

[illegible]

						IIPIDENNIYKYLGVGFGGKSNHNIDNILEKALTDAEATF LSDLHPSQKLFAYKTSIHSLRIPFPRTCIVKHMILDCDRN RIVDNRTKQLGFDQKIKRLKNSIGDIFQINSNNFIYAHCS MGGLGIVPSIDEYLIQSTIQLTRLLMSSDVMTMRDFIIGELI ECVRARFNYLENEIEMAMKWLNGDIDDEKVGKPCFLFL KFKKSYYRLKIKFSIICKLYIDNIFSLKNDKFKVINNE NIDNASKIMHILTGEWYALEWYKMTHQGHIAETVGKSKQ FSKHLJKSTTLGDNQYFFLIKARNNSLNSNPSNHHKDK KDVNCRLCRIQWETQAHILNHCTLSQARRNKHNVHLE IVHNFASKNYDIDIEKSPNVDTTLRPDLIKTTDKNKIK LSLVDIKVPYDNETNFINARKQNLDKYRNLAETGKDN NCKAAVYTIIVGSLGSWDKENNKDLQKLGGLGVNEIGEM ARMCIKAAIEASYMIFMHNHVTNNK MLSQHSRSQALVKEEGCDRLGEPTLNLAILMEMKPKRN PLGRNPLSDAPYRNPGMV L NQGPGRLPLGDAPCRDLG MVSSSEQSGRHFSLGALHRRPGVV KNEQGSSHLRRVR RARKLVEPTRIRVGSWNVGS LTGKRELVDVATRRRVN ILCVQETKWKGQKAKEVDNTGFKLWYTGTAANRNGV GVLIDKSLKDGVDVRRQGDRIILVRLVV GDMVLNVIS AYAPQVGHDES AKRLFWE DL SLV RTVPSSEKLFIGGD LNGHVGTTSAGFEAVHGGFGYGSRNQEGEEVLDFAV A FDLMIANTFFRRKRSHLVTFSSGQHY SQIDFLTRREDKR ACLDCKVIPGECVV SQHKL VV ADRFQMRARRDKQAKI ARTKW WKLKEETSEVFKEKVIKEGPWKEGEDANNMW EKMANCIRKVA SEVLGVTKGGRGVDKDTWW WNEEVQ RTIKEKCEYRRLFHDRSVDNIEKY KEAKKIAKRAVSV AKGRAYEDLYQRLGTKEGEKDIYRMAKVREKTRDFN QVKCIKDEMENILV KEDEIKHRWQYFDKLFNGENENT TLQLDDSFDDTNRHFRVKIQESEVREALKRMKSGKAMG PDGIPIK VWRCLGDI A V W L T K L F N H I F R S N K M P E W R R SILVPIYKNKGDIQSC T N Y R G I K L M S H T M K L W E R V I E H R LRRITRVSMNQFGFMPGRSTME AIFLRQVMERYREKK KDLHMVFIDLEKAYDKIPRNVMMW ALDKHKVPTKYV GLIKDMYNNVTRVTS D G D T D D F P I R I G L H Q G S A L S P Y L F A L V M D E V T R D I Q G D I P W C M L F A D D V V L V D E S R A G V NRKLELWRETLESKGFRLSRTKTEYMRCDFTGTTHIEEG DVSLESQVVPRKDTFRYLGSLQMLQRDGDIDEDVSHRIKA GWMKWRQASGILYDKRVPLKLGKGFYRTAIRPAML YG AECWPTKRRHVQQISVAEMRM L R W I C G H T R R D R V R N E DIRDLRGVAPIEEKLIQHRLRWFHGVQRRPPEAPVRSCT	cd016 50	641	890
RTE	RTE- 1_Slt	.	Setaria italica	AACTCCG CCT	TGATTGTG GCTCTTGT GGGTTTCAA CTCTAGCCT ACCCCAACT TGCTTGGGA CTGAAAGG CTATGTTGT TGTGTGT				

[illegible]

AGAAAGT
TTTTTAT
GGAACCA
ATCAATG
ATGATTG
TATTGTC
GGTTATT
TTTTAATG
ACAGGTC
CAAGAAA
CTTAGGC
TACGGTC
CTGTCC
AATTCAA
TGGTGAA
TTGTAAT
CTATTAT
GAAAACT
GTTAAAT
GTCCTGA
GGATAAA
ATTGAGT
GAGTTCG
GAGGTGC
TGCCATA
ACAAATG
CGTCGAC
ATAACGA
TAATAGA
ACGGTAT
TTCTAGA
TTTATTG
ACTTAA
CGCTTTT
CTCTAA
ACCGTGC
AATACGA
TATTGCG
TATAATG
GGTGATA

[illegible]

[illegible]

[illegible]

[illegible]

CTGTTTGA	TTTTACCA	PEVTEDVDVNVRCGSFYSTLNNAINIYFPVRKVLHFRD
TTGTGTA	GTTAGTCAA	KPWTGEIKSLJHKRQIAFHQKGILWKRLRNKVNSSIQ
TCAACTG	TGTATTTT	KAKVKYRSRIQKLKKNPSRWYQEIRILTKGSRQSLNI
TCATCTT	TATGACATG	PVDGVDPSERDKIANAINEHLSVSSDLQPLNLAALPAYL
ATGTAAT	AAGAAATGC	PSRSLPEVYPWQVNVLSKIKCSTANGPDNISARLIRLFS
TGTGACA	GATGCATAT	VELSGPLTYIINASFQSKVPASWKAIVVPIPKVNPARI
CTGCAATG	CTATAATTA	DKLRPISLTDHFAKVEGFIAKWMYQDIAQHIDPCQFQGN
ACATTTT	TATTTTAAT	QKRLSTSHCLINILHYVFLNAEVPKTSSTILLTDFQKAFD
GAGGTGA	ATCTCTAAT	RIDHTVAIKKLEJGVRPAVVYVWISDFLSERKQVRVYHS
TGTATTT	GTTGCGCTT	DLSEWCQINAGVPQGTGKLGPLFLVMINDFKSSSESIHMF
ATATAAT	ATAATATTC	KYVDDLTLVECRKTNQISNMQEAVECLSLWADINKMQ
AGATACT	ATATTTGTC	LNPTKCIKMDICFSRVPVNQSSILENCVLSVSVQVLLG
GGAAACT	TTGCAATGT	VTIQNDLKWEAHVRDIEKRANAKLYMLRSLGKFGLSHE
GTTTTCT	ATCTTTGTA	DLTIFIGYIRPILEAYATQVWVGGLTSAQKQRIERLQKRA
AAATTTT	TTTTATCAG	LRIIFGKDYHSYANALTKSKLQTLSERRNDLCVTFKKT
GATTCAA	CATGTACAT	LSCTDQFKQFIPQRAHNRTLRRTNKIPEPRCKTKRMQNS
CGGATGT	AATGTGTTT	PIPDILRLYNKSFL
GTACACA	CCCTCTCTG	
GTACATG	TCCATCTGT	
CAAGTAA	CCCCCTTCT	
AGCCATT	CCCCCTCT	
TTACACT	CTCTCTCTC	
GGGCTGT	TCCTCTCTCT	
TACACAT	CTCTCTCTCT	
GTAATAT	CTCTCTCTC	
ACTGCGT	TCCTTTCTC	
GGATGTG	TCCTTTCTT	
TGAACGT	CTCTCTCTC	
CGTTGCT	CTGCTTTCC	
GTGTGAC	CATCTCTCT	
TGATTTG	TTCTTTTTT	
CCTATCC	AACTCAAA	
GTGGTGT	TTTTTAAAT	
GACATTC	TGAAATATAT	
ATCTACA	GTATAAGA	
TACAATG	TTAATGGTA	
AATAGTC	TATAGTATA	
TTGATAC	TACACGCTC	
GCACTG	TTCATAATT	
	ATGTAGATT	

981

				CTTTGAA TTTATTCT TATTAGC AAGCTAA GAAACAC TGAA				XPGRSTVXHILTLRRLVEGIKAKHLTAVLTFVDFXKAFD SINRKKMLKILXAYGIPHTTIVTAVGVMYKDTTAQVRSP DGETDSFTILAGVLQGDTLAPYLFVIALDYALRTAXEGF EDLGFTLQERKSSRYPAVMITDTEFADDDIALISDDLEKA QELXQVESAAQVGLQINSTKTEFMLRNLREGEICTAD XTKSKQVQNFQYLGWSVISKQDLNIRKAKAWTALNKL TAIWKSNLGRKLKVRFFTEVTVLLYGSEAWTITSLEK QLDGCYTRLLRAALNVNWRDHMTNIELYRDITXISARL RDRRLRFSGHCWRSKSE			
				GGAATAC CATCAGA GTTGGAA CATGGAA TGTGAGA ACACTGT ATCAGAG CGGCAAG CTGGAAA ATGTTAA ACAAGAG ATGATGA GGCTAAA TGTC AAC ATATTGG GACTAAA TGAAACA AGATGGC TGAATAA CGGCGAG TTTGTGA TAGATGA GTACAGA TTGATTT ATGCAGG AGGTGAA AAGCATG AAAGAGG TGTTGGG TTACTGC TGGATCC				MVITNTWFEKHPRRVYTWKSPGDIVRNQIDYIAINSRFK RAVKQSKTYPGADCGSDHVPVICTLQCKLKKLRPKVV KILDFKQLSKPDIRQEFTVQVKNRFELLIDEGEENTWET MRNILVETAEYVPTKERKRRNKWMTEELLMMKDRQ KIGNRQCKKYRDLDDKQIRQKCNKESWLNQCVEIEN QFDKNSVYQRINEISGKSGCSNPGCIKGKDGSLFEM KDKMSRWYQYIGELFEDSREAMPSFADEVKGEKILKSE VRWAMSKMKNKAAAGPDGVVTEMFEALEEYGVDRLT EIINKIYDDGTFPDELGRSIFITLPPKPGAVDCEQYRTISLI CHITKIILRIILLRARSRTIPMVGREQFGVKDAGTRNAIF VVRNITERAVEMQKDVYMCFIDYSKAFDKVKHSELFEE LKLDIYRKDLRLQLSYWNQFACIRVEGECSEYNIRR GVRQGCVMSPDLFNMYSEILRELEDENGIQVGGHSITNI RYADDTVLFSSEKDLQRLNVVVKESERMGLEINYTK TKCLVVSCKSSPKCNLKLKEQIKQVSSFNVLGSMIAED ARCEITEIKRRIALAKSAFSKLGKILTNRTMSMRTRIRVLN CYIYPVLMYGETWTLASDTKKRLESCEMWFLRRMMK IPWTDKVSNEEVLTRAGVQRKMIREMLIRQLKFMGHIT RKEGLENLALTGKIVGRRSRGRRRLWMDSLQKLEER GVKELGVVLQKARSRLGWSNMIKVKRYGT	cd016 50	321	578
RTE	RTE- 3 SP	048B0 5	Strongyl ocentrotu s purpurat us								

AGATATC
TCAAAGA
GTGTGTT
GGGACAC
TGGTGTG
TAAGTGA
TAGGGTG
CTCTAG
TGAAATT
ACAGGGA
CACCTT
TCCAAATC
ATAGTGG
TTTATGC
ACCAACG
GCAGATA
GTGTGA
CGAAGTT
ATTGACA
CCTTTTAT
GATTCCT
TAGATAA
AGCTAAA
TCTCAGT
GCAAAATC
GAATGAC
ATTGTCA
TTGTCTT
GGGAGAC
CTGAACG
CTAAAGT
CGGTCAAT
GGTACAG
AGGGCAA
GCAGTAG
TGCCGCA
TGGAAATT
GGTGAAA
GGAATGA

[illegible]

				CGTGACA GCATTGA GGGCGAT TGACGAT ACGGAAC TCCCGAC ATCCTTC GACCAGT CACTGCA GGGTCTA GTGGTTC ACACCCG TCCTGGC TTGTCTC GCCGGTG TGAAACT GCCACCA GTTCCCC GAGAGAG TGCCGTT AGTAAct TGCAAGA GCTACCG GTAActA TCCACCT TTGCAAG CAGCCTT CAT	TAGTATAAC AAAAACCGT GACTTATTT ATTCTTTAT CATGCAACT GCTTACTGC ACCTGCACT TGCACCGC ACCTCTTG ATCACTCAG TTTTTCGAA GAGAGTTA	MNSHAHTPDILVDDINESKYYTPQEFKNFSYKPDNISLL HINSRLNKNFEYVENLLHSLNNFAFVIGISETWLHQNS PDMFNLPNYKLIRADRKGRRGGVAFYTQNLFKIRS IKLTQAESLFIENN NVKNVIIGLYRPPDSNCSELFCEDEL DFYLHKIGDENKHNFILGDFNINFSDDNSNSTNCMQL MHSYGFFSIINKPTRINPHSSTQIDNIFSNVHNNTITGGIL CSEVSDHLPIFLTCECKLSCSKSFDELYRKESKRNIELLK QDLFFEGWAEVYNVDVNTAYKHFSNKLQYYYEKNIP LGKMSNRNKP RNPWITKGLLSIQTRNRLYKHLHRNP TDTNLT YKIYRNKLT KLIRTSRRLHF SNKIKKASSNTTA TWKIIEVMGKKTNPPPTDNITLNGTKIHDP SLHANSFN SFTNIGPELAS KINRTHAHFKDYL SNONOESLFLNP NP	cd016 50	531	787
--	--	--	--	--	---	--	-------------	-----	-----

ACTAATC	TGTCGCCAA	SEINTQSLNNSKSHGFDRISMSLLKEIMYPLAKPLTHIF
CTAAATT	TTCTTTT	NHLSQGVFPDLLKIAKVNPLKKDDPHEISNYRPISSLPS
GTCAATG	CGAGGCCG	ISKVLEKIVHNRLHKFTTKHKILNSNQYGRKNYSTDLA
TTGTATTC	CCATCCATC	LIQYDKITSATAANKEHVIGIFCDLSKAFDTLNHAILSKL
CTTCCAG	TCCAGCCTA	SHYGIRGQPLLWFKDYL TNRRQYVTFNSYSDSDPLLIQCG
GGGACAT	ACCGCTCAC	VPQGSILGPLLFLLYVNDIINTSPLLSFILFADDTNIFYSHK
GCAGTAG	GATGGTGG	DLNSLNNLTNLYEINKVSDWFKSNKLSLNTKKTTHFMYFK
AACTTTG	TCCTCTGCG	HHSNTEPIYIKIDGMLLEKKTNSKFLGVLIDETLTWN
TATCATG	TTCTAAATT	DHLHQVTMCISKISGHSKFLKFLPHATLFLLYNSLVLPIYI
CAGGAAT	ACCTCTCTA	TYCNSVWANGSGSTKINSIFKLQKKAIRICTGSHFLAHTD
GTATTCIT	ATTTATTGA	PLFYKLKTLKIFDLNTTQVAVIMFKYINNLLPSPFDNMF
TACCCTC	AAATATTTC	RENNSVHSYSTRISGNFHL TNPKLSITHKSVRHSGPDIWN
AAATGGG	ATATTATGT	DFPDNIKSCSTIYSLKATLKRQIIQSYAPHSLN
CITGAGA	TATTGTATG	
CAAGTAT	CTTCGTACA	
TCGTTAT	TGTTTATCG	
CCTTTGA	TATGATATA	
TCCTGAC	ATTGTACAT	
GACTGGG	GATTTT	
ATTATCA	TTTAAATTAC	
AGACCGA	TTTTGTAAA	
AGTCCTC	CATTTTATT	
ACTCTTT	CATGTATT	
ATCAATC	TGTTATTGT	
ATTATTA	TCGAAATTT	
TCTGTAT	TGTACTTTG	
AGCAATC	AATTIGAAT	
CAAAAAT	TTGTATATT	
TTGAGTC	TTTGAACGC	
TATCTGTT	AGAAATAA	
ATATTGT	AATCAATC	
AATTGAC	AAACAAAC	
TGAACCT	AAACAAA	
TTTTTGT		
TTTTTGA		
GCGGGAA		
TGTCAAA		
ACCTCT		
TTGCCAT		
TTAATAT		

TATTGTA
CTCCGAA
ATGACGT
TATTGCT
ATTAAGC
ATCATCT
CAGAGTC
TGCTTAC
ATGCATA
CATTTTTC
TCTCTCG
CAACAACA
AATCTCT
TTGTTTGG
CAAAACT
CCTGTGT
CAATCAC
AGAAATCC
ATTGTTA
TAGACTG
AACTAGT
TTTGATG
AGCGGGA
ATGTCAA
ATCCCTC
TTTTTCCAT
TTAATAA
TCGTACT
CCGATTT
GTCATTA
TTGTTATC
AAGCATT
TTGTTTCT
TCTTTCGT
TCCAGC
CCCTCTA
GTCAAAA
ATACACC
CTTTACTC
GGGCGGT

				CGGGTC CGTCTG TTGCTCCT AAGCAAT TATTGTA ACTGTA CCAAGAT GGGTCTT ATTGTAT TTTCAATT GTATTAT ATTTGA AGTTACC CGCACAC TTGCACT TCGTACA CACACAA ACGACC AGCATCA CACATGC AGTTGTT CATCCTC CAAAATT TTTACAT AATACGT ATCGTIA A		TAAATCCG GCAACGTT TTTTTTAC TGGCACTAT GTATATATA CACCTCTCA CCTGCACAC CTGGACATC CCCATGTCG CCACACAC ATGCACCTG CACACATCT	MGRKPKYPCGVCSRAVTTTRTTGVQCEGTCLKW/FHAECI GLTSTEYVAFDSSSIWICNHCGAQNLHSCLSLQLXP DTNRFGLDQDQGNVISEENNSCRPTXDSSNPSSPGSPA YSSSPQNIKQRKSKHKPRPRVLRFLNVNCQSIRAKMQPF SILVHQHNPDIVIGTESWLSDGINSNEVFLDEQYSILRRD RSNDPHGGVFIATKTDLIV/TREELETDCEILWCRIEMA GSKTLRVAAYRPKEDDEVSLAELQRLSLRDLRRHLVV LAGDENLPGWKWDGPYVGPCAYPALHHQFGEIINDHG LTQLVKEPTRQANTLDLILISNPITLVNSVKVVPGISDHNC PVADIDMSPPRRYQPKRKVPMYKKADWEGFAELMQHV ADYICKSSASVHDLWQLFKSGIEGGIQTfVPHKMLKQ KODLPWITSGIRKLIYKRDRLHKOV/TRLRKONMLVPFPL	cd016 50	632 884
--	--	--	--	--	--	--	---	-------------	------------

[illegible]

							VGRTQYFLHKIGAXSSPNCACGEIQITMDHVDDCPIYKS PHGASGLRELDDETCWLNKDLPI				
							TDTRGKQSKFRKLRDRKLIQIFYTNCDSLNNKRNELL LLLSQSHDVLLTEVFPKNHQNPLTHRDFILEGYDLHC NLEDGNRGVAIYIARDLYHLVGECDLQHSFHSVWILIT LRGSDKLLIGMYRSPSLDNHEELKRAFTQVTSINSSHV LIAGDFNYPNIDWKLGRARSHTQEFTVADCYLHQHI GEHTRYREGQRSSMLDLVFTNEDGMINHIDYVPGIGRS DHIGLFGLTVYTPDTHPTPHLQFYRGQYDDMCKELD QIEWELENKSVEDAWEFFLVLTTEKMEKFIKSKPKH RRKKQRWMTREALRKQSQKHRAWKEAQRTHKHPVDIM RARRLAKQLTLTKLCRNFEKDLAMNVKSNPKAFWR YTNKSLKTSRCLGELRTPDGERISTDKGADALNTFFAT TFTDEDKGHPKLPKRHKGETLTSISISEDVYKLVHLS KYKCAGPDGHPCVLKETAAVSSEPLSIYNASLKEGKLP SAWKVSHVVPVHKKASKMDVKNYRPSLTSVPCKILES LIRDEIVAHLMENDLLCDEQHGFPGRSCVTQLLTJME AWTEMLDKGTPIDVVYLDFFKAFDSVPHRLRLLSKLETY GICGQIKEWIADFLQGRQQRVRVNDKLSDWTKVRSQVP QGSVLGPTMFVLCINDLPDGLQGSVKIYADDTKLYGKA STKDDLCSDLDLALMEWSTWQMHNFTQKCKVLHI GPGNVKHIYALGDDVLEATSEKDLGILIDDELKFHNHV AYAVKKANGVLSIRRSECDRSTIPKLYKGLVRPLID YGDSIWYPRFKTDSKTIERVQKRAKMIYDIRNRPYIER LRALNLPSEYRRRRGDMIQVFKIHGLERIPVDTFEHS TSNTRGHSKIVKQRCSTKIRQDAFSQRVINDWNTLPEE VVTATTVNAFKARLDRVWVKDKRFCSPYE	cd016 50	509	757	
CR1	CR1- 24_S P				Strongyl ocentrotu s purpurat us		TAGAAATAC CCGACGCA GAACTGAG CCAGACAC TTACCAAG AAGACACC CAACTACG AGAAAGAA GAAAATGA AATTGAACT GGAAGCCC GACTGCTG AGAACACA GCATTCTTC CAGAACTTC GGGTAAAG GTAAAGGT AA				
							TACAGTCTT TCCTTAGGT CAAGTATA GTTTGTATG GTTTGTGTTG TTCCTTTGT TCATTTT TTTTTTTGT TTAGTATAC TTTTGTGTT TGTTGTCT TATTACATT ATATTCCAT				
RTE	RTE- 16_S P				Strongyl ocentrotu s purpurat us		GACCATA TTTTTTGTT TTGAAAT CCCTGCT TGACATT TACTTGT CTAAACG GAAGAGA ATTACT GCGCTTT TATAGAC TATAGAA				
							MVDRVSLWHKLLRSGVDGKVFTVIKNMYSQAKSCVIV DQMSEAFQCDVGVROGENLSPLLFAIYLNDFEFLSK HFGLKFLVENVCADILDKEVNKDLFLQYTLTYADDT LILADTAELQVALNALSQYKIWKLSVNAKTKVLIFS RGKVRKFPVRFGEDEIDVDDFVYLGQVNYNGSFKK AISKQVSQARKALFMSLKVQRLKPLDQCELFQDLVL PLIYGCEVWGYEDLSQIEIHRKFLRSALCVNKFTPDCM IYGETGRAWLANSIKHRMVFWLRVKGSPSKYSVKFY FLQRLISQNVDMSTSEKWLSCIEKILNHAGLGVWLKQG EGFSNSWILSTLRLNLDISKQDMATITWTSVCLNYRI FKEDCLFERYLVLLSKKDRITLSRFRCSHRLPINNRFK EAVNEISIEPCCDMHESGDEFHYVFCFFRKERELYL	NA	NA	NA	NA

GRKVVIVKRPCTSHMLNLFSCSGIGQLKKLVFLRIMSYP
ARLYTCKSEQSNHMSPIKRNVTTRSGRLJKQPTILD

[illegible]

TGTTTCATC	TENNPTIRNLRKKNKRYREPRCRTNRYQNSPIPFMTKLLNF				
CCCTCACTT	N				
CAATAGAT					
AACCTACC					
GTGCCTGCG					
CCTGCCAAC					
ATGGTTAIG					
TTTTAGGAT					
CTTGTTTAT					
TACATCCCA					
GCCTTTTAA					
AATAGATT					
AATTCACCT					
TAGCTAAA					
CTGGATCAT					
GTGTTTTTT					
TACTCITTA					
CTAATTAAAC					
TGGATAAA					
TCACTCGAG					
CATTATAAT					
ACATGTACC					
ATATTTAAA					
GCATGACCT					
ATAGTATTT					
ATTCATTCA					
CAAAAAAT					
GTGATTTC					
TAGAAAAAT					
TGAACGATT					
AATGCTAA					
ACATAGATT					
TATATTTT					
GTAATTTTC					
CACACGCA					
ACACTACTA					
TTTTCAGAG					
TTGAAATGT					
TCTAAACCA					
GTAAATTGC					

CATGTACA
 GTATTACGC
 ATGTAATTTC
 TTGTTAAGT
 ATGTAATGT
 TTTAATCAT
 GTAAAAGT
 TATCATTTT
 CTTTATATC
 CCTCCCCCT
 TTTTCTCTC
 TCCTCGGCC
 ATACCTGTC
 CTTTGTAT
 TCAATATTT
 CACTGTGA
 AATGTGCTG
 TATAATTGA
 TGCATGGGT
 GTTGATCGG
 TTGCGTCTGT
 GCTATTTCAT
 CAGGGGGG
 TATATGATC
 TTAATGTTT
 CTTGCCCCA
 CTGGTGACT
 GGGGTGGC
 GCACGCGT
 TCGACGCC
 GGTGCCAA
 GAGTTGCAT
 TATTACTA
 TAGTAGTTT
 AATTATTAGT
 AAGTGC AA
 TTTTACCTT
 GTAAATAATT
 AATCTATGA
 ATTCATGCA
 CACTATGCG

						GCAATTTAG TTTTTAGCT ACGAAGCA TGGAATTA ATGAACATT TATTGTTA CATGTTGA TTTGATTG AATAAACC ATTATCTA TCTATCTAT CTATCTA				LTKTEKDLLSRGLNSAVTPEELPVVDLITATETAIRNNKL PETEAEQIRLKVSAALANAKAPASNITAQEKRALASLAK DKDITILPADKGRCTVVLNTTDDYDSKILLSGDSNTYEK LKRDPSTYKKKVIDLLQKLEKEQAIDKPYRYLYPGE ATPCFYGLPKIHKEGTPLRPVSSINSVTYNISKFLASILAP MVGNTPHHIKNSQDFAEKVSNLQESGETMVSFDTLSL FTCIPTSEATEAIHKRLLLDKNLQERTTILSPAQICTMLDL CLNTTYFQFRESFYRQKHGCMGSPVPIVANLYMEEV EKRALLSFTGAAPSRWFRYVDDTWVKIQTKEVEKFTTH LNQD'TFVKFTREDV'KGDSLAFDCEIRIEEDRNLSIEIY RKPTHTDQYLLFDSHHPLEHKLGVIKTLQHRATKVPTTS HGRKKEQDHLTTALKTCGYPDWAFKTRKKRDPSENEE EKDKRHSVSIPIYMAGVSEKFRRLQKHNPVQFKPSNTL ROKLVHPKDKTPRHKQSNVVYAVCQCQCECRDLYIGET KQPLHKRMAQHRRATTSGQDSAVHLHLKESGHSFEDN QVRVLAREDRWFERGVKEAIVKMEKPSLNRGGGLRH HLSPTYNAVLSHFHKQPKHHTHSFTRAGVTSFGDPAVK RKTSQ'TETRSTLPTTLPTTPTPETS'IQTTGILMQRLSVDISM PMVQKGYIFKLCTEPDQN	cd003 04	296	379
Non - LTR Retr otra nspo son	BRID GE1_ TN	.	Tetraodon nigroviri dis	CTGGGGA CCGTGGT TACAACC CGGGCTT AGCTGCA GAGACAG TACCTCC CCGTGGT TCCCCGC	TAGGGGC TTGGCATTT CTCATTTGC TGCTCCTGA AAGGATAT GGGTCTTCG GTCGCGTG GTAGGCAG ACCCATTGC	MGSWVNFVSVAQTQGEFPVD'TARRAPVPVTSYPESEC HXPLPLTFCNSDVTIWWGGVRPEPVDCLGDLPEXYDALP GVAGPREXVGGSPGEGVRSPIASPSGTAVQHDGFSPI LVPGAEAAEVSTPVVKVPQDHPACPCCGTRVVKV'TALS EHLRRAHGRKRVLFCQSRCGRMNEKHHSIACHFPKCRG PPVEEGPLGAPEW'CCCEECGQKFN'TKSGLSQHKRSVHPL TRNVERIEAARPKGKGRGAHKGCWTEAEVAQLIELEG RFKNQRFINKLIAEHLPSKSAKQISDKRRQLAAATKTSSP EKRV'TSSTSGESSPEVEK'VEGIKREYRRRVGEWLCAGSL	cd016 50	542	806				
R2	R2- 1_TG ut	.	Tinamus guttatus										

GGACCCC	TCCGAGTA	XDQTSFQKILEDVESGEIVTGPLEELASFARGKLAAR
GTAACAT	GGGGGCTT	VRHRKHPAEAVPAREEQRWMMKRRVGRRLGLYLRFRQL
CGGGTGA	GGCAGTNT	FALDRKLAGHLLDDVESIKCPLPMEEVADVFRRRWEEV
CTGAATC	CCATTGCCT	APTGSFSRSLGKADNGAFKPMISAKEVMKNVXEMSR
TGTCTCT	GTGCCGA	RSAXGPDGLSLRDLMDKDPQGRMAELFNLWLLAGRVP
GCCCCGG	AAGACGT	DQVKAGRTVLIPKSDPGKIGNIDNWRPTIGSVXLRMF
GAGTAGT	GGGTCACT	SRILSARLRACPINRRQRGFIAAPGCSENLKLLQALIKS
TCCTCCTT	GGTCTGTCT	AKRDHRTLGVVFVDLAKAFDSVNHQHIFQVLVQKGVVD
GCCCTAT	GCCTACACC	GHIIDILRDLYTNAGTYLESQSQRSGFIKILRGVKKQGDPL
TGACCAG	TCCTAGAC	SPILFNALDPLLCRLLEDRLGYKYGDQQIXSLAFADDL
CGGTCGC	TTGTAACAT	ALLSDSWEGMQQSIRVVEEFCQRTGLRVQAPKCHGFLI
CGGCTGC	CTAGTCTGT	RPTKESYTINDCDPWTIADMQLDMDPGSSEKYLGLGID
TCAATAG	CAACAAGA	PWIGLSRPELSEVLTRWVKNIGGAPLKPQLQKVDILRSYA
TATTCTA	TCAAAAATC	LPRLFIADHAGLSATCLHSLDLSIRSASVKGWLHLPSTC
GGCGTGA	TTCACACAG	DAILYVSYKDGGLGLPRLASLIPNVQARRLVRIAQSEDD
AATATAG	ACGACCGA	VIRSVVLQEGIQEEIRKVVWISAGGRPEKVPSTGEFPVM
CGATAGT	GCTTGCTCA	EAQAADEALSEWERRAPRTYPIPCWKWKREMNWTNL
CCTAGTG	GTCTTCCTG	KSQGHGIRNFENDRISNDWLLHYGRIPHRKLITAIQLRA
GTTGTCTT	TACCCGCA	NVYPTREFLAGLGEAGPRGCRHCPAEWESCIIIGYCP
ACTGGGC	GAATTTTGC	AVQEARIKRHNDICGVLAEARKLGWVIFIEPHLRDNTN
CATAGCC	TCCTGTCTT	ELFKPDLVLVKGSCAKVVDVTIRYESGLTTLSDAAAEK
CCTTGCTT	CCTTTGGCT	ARKYQHLAGEVRLTSAITVDLGFPIGARWKWYVGN
CAGGGGT	GTGTCCTGG	NGLLSDLGFSTSRVRIARALSKKALLSSVDIIHIFASRA
CAITCGC	ACGTGGGA	RQAQTSE
GAAGTCT	CTATTCCAT	
CTCAGGA	CTCGTCCCA	
GAAC TGG	AATGCCGC	
GGGTGGT	GTCCAATTA	
GTTCTTCT	TACCGGATT	
GGGTATA	TGACAAAG	
GCTAAAC	CGGACGGC	
CCCCTAG	CCGCTTTAT	
ACTGTGT	AAGCCGGA	
CCGATCC	AAAGGTGC	
	CTTGTA AAA	
	TTGCAAGGT	
	TCATTAAAT	
	AG	

CR1	CR1-3_TC _a	.	Tribolium castaneum	ATAATCT TTAACAA AAATTG CTTAAA ATAAAAT GTTTGA AAACCGA AATTAAA CTAAATT AATTGTG TCTAACA CTTTAGT AGAGGAA AAAGGAG AAGACAT AAAAGTC ACCAAAG TCTTCAG AGTCGTT TTTAGTA CTAAATT TGCTTAT ATGCAAG CGCTTAA TTAAGGT AACAGTT TTTTGGTT TCTTGCTT ATATCTC GAAACAA GTTACTC CTATCAA TTTTTATC TTTTTACT AAAATTA AAGCTGA TAAAAAT TCCTACA AAATAGT TATTGTC	MSVCYKCKSSVNVNIACLNCKKAYHPSCCLRAFNYSAL GQDLTISACPCCDKLSLQNPTSPNLNISTNTIFEDLGIL KQETDLIKKTLDDLKTFTITESPTTCTFDKIDEYVRKSPD DLESKLISLLKDTLNRREIEKLNKSIKNEVIAEISTNLNRTN DNIKTANSSDLTLKSTNSKNVSTQSVSKTNRYKVRIFAD QKGRHCIDLLETYPVSKYVIQSVVVKPYSSLQELLKTEN NNEQLGNNDFFVIYAGNHEAQKGKIPDENTLNELINHK FLNTNIIIGCPYFVNRPNLNFVYELNCNFYNVAKNNK NIYYLDSNDKIFDQFKTDNFVINKAVKRCIKNLGDLITE CQDHTNRSAVEQNPVATFLDTSLPSVDKSDSQSFLSVFL NRFVCLQPSHKSISINKKTVNSFHGIYQNVNRLKSKSV TFENSFLSTDEFDLIVITESWLDLTVLDGELFSDNYVVF RRDRDLRHLGIHRGGVLIANNKYNCRQLDISNISRRS SPAIDLVAIKINVTSYQHLFFVVIYPPSISFDFDNFFDL LNNEDEINNDNVFLGDFNTPLFSGFSKNNFSLKHFAS PHNFNQENNRNRRNNLDLIFSMDCAVSKSSFSFVPED PHPSLEFTLDIVAKTPDKFYMINRDHVAFNFKKANFDC LYTTLQVVDWSELEKFSVDNEACRVFYETLYNCFELSV PKLFCRKNYRTFTWNTDINCLRRKSNILKKYRRTKN MHFYEEFKLRSLSKSLITAYSQYIGHMENTISKDPKK FWSFVQAKKRNRRIPGVLYHNGNAISNPNDIVNSFAKFF KSVFSDMTLTSDSPSMMTNLNSFTVNEINEHEIRVLKR SKDITLTAGCDGIPSLRLDCAHIFSKPLVYFNLILQJSTF PNIWKIAHVCPVFKTGITQIQNYRPISLCNFSKVFESIY DRIYSSVKRFISPSQHGFEKRVSTVNLACFTQFISDALD RKGVVDVIYMDFKKAFDQIDHLLLLKLDQYGFSGSLL SLIKSYLANREQRVKYRNYSDSYVATSGVPQGSNLGPL LFLLFINDICGSLSTCAKLLFADDLKIYTEIKSIEDCLTLQ NNINAVVWCNENRLYNPSKCNVMSYTKKREFLEFV YDISSVTILHRTIFILDLGVIFDTELAFSEHIRDVTARAIS YGFYRNCRDFKNLSVMKTLFFSLVRSKLEYGALIWHPI YKIHIDQLENIQRRLKFLVFIDGNYPYRGYDQNLNLLNRF GLQSLQFRRICIIKFLYNLIN	cd016 50	915	116 6
-----	-----------------------	---	---------------------	---	--	-------------	-----	----------

ATTTTTCA
TGTAGGA
CTAACCA
TAAGCGA
GATATAA
GTTTGAA
AATAAAT
AATAATT
AAAAAAA
GTGCTTT
AACAGAA
AATGCCT
TGTGATT
TAAATAA
CACTTAA
TTTATTG
GGTCCAA
TACTTTAT
TACTTTAT
TTAATTT
TTGGTGC
TATGCTA
GGCTTGT
GCTAATG
TGCTATA
AATAAGG
TGTACA
ATTTTGT
GCTGTAA
ATAAGGT
TTTGTGC
ATTTTGT
GCTATAA
ACGTCAG
TTTTTTAA
TTTTTCGT
TAATTAT
TTTGTGT
AATAACT
TAATAAG

TGACTAA
GGTTTAA
GGTTATG
TTAAACT
GTAGATG
ATAGCGT
TTTGCAIT
TCTTTTGG
ACTTTCA
AACATT
ATACTTT
ACACAAA
GAGTTAT
ACTGATA
TAAATGA
CGTGATC
AATGCTC
GTACTAT
AACCAAT
AATCCAA
TTTTTGTC
TACTTTT
ATGAACA
TTAGTTT
TTAGATT
TCAATGT
AAATTC
ATATCAG
AGCGTA
CGCTTCT
AGAAATG
CTGATAA
TAATTAT
TATCTAT
CCTCTGTT
CCTTTTGT
GTCTTTC
CTGTGCC
TCTGTT
GCTTGTT
TTTTATT

1000

TAATACT
AATAATA
CTGATGC
AGTAAAA
AACGTTA
GAATTTT
AAAATGA
TATGCAT
GTAATCG
CAATGAT
ATCAATG
CTCAGTG
GAGCAAT
TTTATAA
ATATTTTC
AATGATA
AATTAA
CGAATGT
TTTCCTCT
AAAACAA
ATAAAAAG
AAAAAAA
AACAAAG
AATAATT
TATAAAA
ATGATCC
TAAAGTA
AAATATT
TTAATAA
TTATCTA
GATTGTG
TATCTGT
GATAAGT
AAGTGTG
ACGATAA
ATATGTT
ATTCCTTA
TAAAGTA
TTGAAAA

[illegible]

1003

1005

TTTTATTA	AAATACCCCT	PSASSAAFLSDFDTWLSFFLCTDSPAILGDFNFHIDDAS
AGACCTC	CTGAATTGT	QPWPSRFLRLTSSFELQQWNTNTPTHKDGHFLDLVFTKK
TTTAGGG	GTCTGTATG	LSLSGFNSEPHPLSDHNLVTFISHSPPASPTVLVRDT
GAGAGCA	TTACCTCC	RIDLPLAQAQSRSSLSLKEMSDPDTLVSEYNHILSSTIG
TTTGCTG	CATTTAGAC	SFAPLQAKRTRALNRPWLNSQTRFLRSCTRSAERMWK
GGGTGAG	TGTAAGCCC	KSRTQADFIHYKFLSCLNSALSANKSTTVPLIHKPDSY
CATTGCG	TACGGGGT	APVRDLLNVCGRNPVRKQTLFTNFYCPASILPCQKPTR
AAGGGTG	AGGTCCTC	VLQYPYKQSQIQPTTPVFSIQYFLHPSSTVSVNQTHSPQD
CTATTGCG	TAGCCTTTT	FADFFKDKVESIRNQIPTSTNVNQLLPKPPSACLNSPPP
AAGTGAT	GTTTCCTTG	VTMSEVSKLLLSPLTTCPLDPMPSLLKHCV AELTPLL
TCATTTA	ACACTGAG	HIFNSSLASGTFPSPFKQACVKPILKKATLDPACLSNYRP
CAAGTGA	CACTTAATC	VSLPLASKLLERIVFSRITNFLHTHDLDDLPLQSGFRPGH
TCTGGAA	TGTAATTGA	STETALCRVTNDLQVAKAKGHFSILLLLDLSSAFDTVDH
GAGTTTA	ATTATAATT	SLLMQILYSLGIRNQAAASWLSSYLSNRSFTVSYDNKTSS
ACAAGGA	TATATTAT	PVPLNVGVPQGSVLGPLFLSLYTLSLGDLIHSFGFNYHL
GTTATGA	GTGAATTGT	YADDTQIYLSLTPSLTAETEAQISNCLLAINSNWMNQRLK
ATTGACA	ATTICTAAT	LNLTKTELIIFPPKPGTPLFTISIDGMLINPVNSARCLGVI
TCTCTCTT	AATGCACTT	FDSSLSFSDHINTTVKTHFFLRNIAKIRPFLSPDTSKTLM
CCTCTTCT	ATTGTTACT	HALILSRLDYCNLLLTGLPNSHLSPLQSVLNTAARIILLSS
GTTTATT	TTTTATTCC	KRVQAPPLLLKSLSWLPIKQIRITYKLLITFKALHSSAPNY
ACAGGTA	AATGACCC	ISSLVSLYVPARNLRSSQSNRLVVPPTTTTAVSRIPKPSLAT
AATCTGC	CTTGTTTGT	PYIWNAPESLRRESSFNVEFKTKLDYLWEHLDNT
TAAAAC	TACTACTAA	
GCTATT	TTTATTGT	
CCATCTC	TTGCTGTAC	
TGCTGTG	AGTGCTTTG	
GCCCTGT	CCCTCAAG	
TGCTGCG	GAGCGCTA	
ATTTTTC	TACAAATA	
ACCTTTA	AAAATATA	
TTGGTAA	CATACATAC	
ATCTGCT	ATACATAC	
AAAAC		
CTATGCA		
TTCCCTTC		
CATCTCT		
GCTGTGG		
CCCTGTT		
GTCTGCA		
TTTTTGCA		

CCATTTA
TTGGTAA
ATCTGCT
AAAACATG
CTATTGC
ATTCTTTC
CATCTCT
GCTGTGG
CCCTGTT
GTTTGCA
TTTGTGA
CCATTTA
TTGGTAA
ATCTGCT
AAAACATG
CTATTTC
ATCTGCA
TTACACC
TACACTG
TTTAGTG
AGTTATA
TTTATGC
ATTGTC
ATTGCC
ACTCTA
AACTAGT
CACATAA
TATTGAT
TATAAGT
TCTCCCT
GTAGTCT
TGTAAGT
TGTACAT
TTTACGA
CCCCCTA
TAAACTC
TCTCCA
AACAAAC
TTTGTA
GTCAGGT

TATCTCT
GTATTCA
TCTTGAG
TCTTGCTC
TGCTCTC
ACTCTCT
GTCACAG
CTGGGTG
ATCAGTT
TGGTGT
TGATCAA
TTAGAAT
CAGTGT
GATTCCC
CAGACCT
GTGTGCA
GCAGCTC
CCAGCTT
CTACTCT
CCCTAATT
ATTAAGA
CCTCTTTA
GGGGAGA
GCATTTG
CTGGGT
GAGCAAT
TGCAAGG
GTGATTC
GTTTCA
AGTGAT
CAATTAC
AAGTGAT
CTGGAAG
AGTTTAA
CAAGGAA
AGGAGAG
ACAACAC
AGAGGAC
TCCATT

[illegible]

					ATAGACAT A				
Pene lope	Penel ope- 6_XT	Xenopus tropicalis	A	AGTCTG TGTTGAG GAGTGGC CCAGCTT GCTCTCC CTATAATC TTAAGAA	TAAACCTG CTTTGAAGA GTTGCATGA TACTGGGT AATCCTTTC AAAGTAAC TCCACAGC ATTCACACC TCTGTGAGT TTCAGATGT CACCTTCT GAAGAGTT ACATAGTG CCATTGTGA TTGGATTAG TGGAAGAG TATATATG TGAGGACTT CCCATACCA GTCAGTTGA ACTGAAGA AGCTGCTCG GATGAGTA GTGAAACG TCTTCAATG ATTACTAAA CAAGTCCA GTTGCTTTA AATTTATTT ATACTAGAT AT	KGHRASNIQKAQKHLNERNVRQINFTEALKQKIEQLK QSLTKQLPDVTLERVVQFIEHAQMAQHIKSKERQISKFT SLLSRSSGLKRTEELTWQKDETSKNTKDTWVKNLSDR VLTQPEKDVLAAGLNYAVTPQHIPVVVELITATESAINNN HIKVEAEQLRLKVSAALSSARAPPSNLTQERRALTSL SKDPNITILPADKGRCTVVNTTDDYHCKMTLSLSDSNY EPLRRDPSSSYKKKVIDCLQOLEKEEAIDRALYHRLYPR EATPCLYGLPKIHKDGAPLRPIVSSINSVTYSIAKYLANIL APLVGKTVHHIQNAKEFVTKIQGVKLEAETMVSYDVT SLFTCIPTTEATEIVRNRLQKDNLTSSRTKLSPNQVCLLL DLCLNTTYPKYKQFYRQKHGCAAGSPVSPIVANLYM EEVERKALLTFNGTTPSHWFYVDDTWWKIRSNVAAF SEHINSVDNNIKFTREDVHENKLAFLDCLISIQEGGILKT EYVRKPTHDTQYLLFDSHHPLEHKLGVIRTLHHRAESVT TDTESKDKECKHLRGALKACGYPDWAFVKTRATKPNR NTRKRNRRPEAERRRNIVIPYVAGVSEKLRRIFNKHHPVF FKPSNTLRQKLVHPKDPPTPEKQSNVVYAVQCSEECTD LYIGETKQLLSKRMAQHRRANTTQDQSAVFLHLKDKG HSFENSNVQILDKEDRWFERGVKEAIVKVEKPSLNRG GGLRHHLSATYNAVLTSVPRQFQNSHHSCNSNK	cd003 04	413	496
					MSLLLVLSEMFVLPKVTSFSPPPADCPYSITSSLLHST HCCSYNLYTLKLNLAAPTNTSQKHNNFKSFHLLSLSL LLLAAGDISPNPGCRIPISYPRNRSLLV/KPSQTANLSN LIQIPTLSSSYLPSCALNLTNRSVTNKAAIHDLFCSSKSFH LFAITETWLSQNDNAIEAALSYGGLSFSHTPRITGRGGV VGVLISPHCRFKLIPIPTFRFPSPFVHAVQLFHPLSVRV AVVYRPPSSKSLATFLSDFESWLSSFCLSDTPAILGDEN	cd016 50	616	874	

CCTCATTT	ACCTTCCAA	CHDDPSQPWPSRFLHLTSSFDLRQWNTATATKDGHFL
GGGAAAG	TTTGTGCT	DLVFSKNLDLSAFNSEPPPLSDHHLISFSISHVSPSPSTNA
CAGCGCA	GTATGTTAC	QPKTLIRNTQRVDITALSNSLSSLSFSCASSDPELLVNTY
AGGGTAG	CCAACCACT	NSILSSAINTHAPLQPHQSHRAHNPWLNQTRFLRSCA
CAGCGCA	TAGATTGTA	RSAERIWRKSRVKADFLHYKFLMVCLNTALSQAKQVY
CAAGTGA	AGCTCTACG	YNTLVNNNKSNPRRLFSIFNALHPSPNESIIEHTPQDFA
ATCCTTC	GGGCAGGG	DFEKSVDISIRKQFSQPSDTGNLNLKPSPLSVFSSFNLT
CAAAGTG	ACCTCCTTC	ESEVSQLLRSSPPTTCSLDPMPSSLKQCAPALTPALTHIF
AATCAAT	CTACTGTGT	NSSLASGTFPSLKFQACV/KPILKKATLDPCLSNYRPVSL
CACAAGT	CTCATACCA	LPLASKLLERLVFSRVTNFLCTHNLDPMQSGFRPAHST
GAACCTGG	CATGGCACT	ETALCRVANDLQTAkakGRYSVLILLDLSSAFDTVDHS
AAAAGGA	TATTCCTCG	VLMQILYSLGIRDQAAASWFSSYLSNRFSVALANKSTP
GTTTAAC	TGCATTTAT	VPLSVGPQGSVLGPLLFSLYTSLGDLISSFGLKYHL YA
AAGGAGT	ATATATTTA	DDIQLYLDTPALTDDVQTQIGNCLLAISSWMNQRLKLN
TAGGAAT	TTGTATTTA	LAKTELMVFPPKPGSPPTTTIDGMTINPVNSARCLGVIF
TGACACC	TTTATTATA	DQSLFSNHNNTAKTCRFFLRNIAKIRPFLSQVTAkTLIH
CTCTCTCT	ACACTTGTC	ALILRLDYCNLLLTGLPDSHLSPLOSILNSAARILLSPK
TCTGCTT	CTCCCTGTG	REPVPQPHLSSLAWLVPVKQRIAYKILLTFKALHSSAPHYI
ATTCTAC	TGTAATTTT	SSLVSLHVPGRLLRSSQSLRLFTPTPTALSRLKPFYLA
AGGTAAT	GTATTTTGT	PYLWNSIPESLREHSPTLFKKKLSCYLLEH
TGCTGCA	AAGATTGT	
CTATTTC	ACAGGGCT	
ATCTCCA	GCGTACCCT	
CTGTGGC	TGTGGCGCT	
CCTGAAT	TTATAAATA	
ACTGCAT	AAGTTATAC	
TTTGCTA	ATACATAC	
GTTTATT	ATACATAC	
GGGTAAT	ATACATAC	
TGCTGCA	A	
CTATTTC		
ATCTTCA		
CTGTGGC		
CCTGAAT		
ACTGCAT		
TTTGCTA		
GTTTATT		
GGGTAAT		
TTAGCTA		
GTTCTATT		

TCCATTCT
CAATTCC
CAATATT
GCATGCC
CCCTATT
CCCTGTC
TCCACCA
ATTGTT
GATTTTA
TGACCCC
CATCGTA
ACCTCT
CTAAACA
AACTTGT
TAGTTCA
GTGTATC
TTTGGGC
TGATCTT
GCATCTT
GCACCTC
TCTCATC
CTGGGAC
TCAGCTG
ATTGATT
TAATTGT
TTAATTA
ACTAGCT
GCAATCA
GCTCTGA
TTCCCCA
GTCTCTG
GTTCTAG
AGTGGCC
CAGCTTG
CTCTCCCT
ATACTT
AAGAACC
TCATTGT
GGAAGC
AGCGGG

[illegible]

TTTTAGAC	AAAAAAA	STRVLVNGHPGSIHRRGLRQGDPLSPMLFVLVMDVLG	
CCGTCCT	A	QLVSKADEGLLQPLSTRQLQHRVSLYADDVVLFINPV	
TGGTGCC		AEDMTMISDILQLFGEASGLYNNNQKSSVYPIRCHEDNL	
TGCTTCA		EIIQQFWPCEISSPCKYLGPLSLHKLRRREQVQPIVDKIA	
AGGAGC		NQLPGWKADLLTKAGRSVLVQSVLTGVLIYSVMAVDL	
TTTGGAT		PSTTLKDIDKIRRGFLWRGRKNAMGGHCVAWGSVCR	
CGTCGAA		PRELGGLGFSITELSWALKMRWAWLQKTEPGRPWASL	
CCTCCCTT		PLHTPKKIREFLQVALLYSEIGNGASTLFWSDRWLDGQRI	
TGCCGTT		GDIAPRLIETIPMKLINKRTVQEAIEDNGWINDIPGSLSVG	
GATGGGA		ALADFLRIWDLVAHVLDLDPHKEDKHIFRLAANGKYSTK	
TGICTTCT		AAYEGLFIGSVEFEFERIWKTWAPPKCRFFLWLVAHK	
GGCAGGA		KCWTADRLEKRGDLHPKCPCEQERETIDHLLVNCVF	
GAACGCT		SRECFWLLLRQFGLQGLAPQPTDINFMEWWHQANEAIQ	
GATGTAT		GSFRDGLNSLMLGAWILWNHRNRCIFDGLSPNVANFLI	
CCAATCC		HVGDERRLWETAGAKGLTSLVASLSHVA	
GCATCAC			
ATTCTCG			
TCTGCC			
CACGGAG			
GGATGGA			
TCCGATC			
CTAAGCT			
TCATGCT			
ATTACTT			
CGGGCGG			
GGATGGT			
TCTGGTC			
ACGTGCA			
GCAGGGA			
TCCCTC			
CTTTTCA			
AGACCTC			
CCCATTC			
GAGTGGG			
AGAAGGG			
GAGAAC			
TGGTCAT			
TCGGGAC			
GTCTCTC			
AGTTCAT			

GAGCTGTT
TGGGCC
AAGAGAT
GAGAACA
TGCAGCA
GCGGAGC
ATTAGCT
CCGGGGT
GGCAGTT
TCAGATT
TCAGGAG
AACTTTG
GAATCTG
ATGGGGG
GACCTTG
TCTCTCA
GGCTGGT
ACATGGT
TCTTGTTT
CACCGGT
GAAGGCG
AAGCAGT
TGAAAGT
TTATTTTC
GTCAAAG
ATTCAAG
AATGTCC
TGAAACCA
GGAGGAA
GCTGGAA
CTCCCAA
TTCTGTC
ACTGTGC
TTCTCTC
ATCCCCC
ACAGACG
TGGCAGC
TGCTGAC

ACAGATC
CATCTGA
GTATCCT
GTTTCAG
ATGTTGA
CGCTGTG
TTGGCAA
CTCCGTC
TAGAACA
TTACCTC
CCTCCCT
GTCTGAT
GTGGCAG
CAGACAA
CGGCGCT
CCGTCGG
AGGCTAA
TGCGTTA
GACCTTG
ATGCTGA
GTGGGCT
GCTCCAT
CTGAGGA
GGACAT
CAAGCAA
TACACAA
TGCTGCG
TTGGGCT
CTCCATC
TGAGGAC
CCTCCAG
CAACAAC
ACAAGAT
TCTGTGG
AGATTGG
TGCTCAA
GAGCAAT
TCATCAT
CAAAATG
TCAAACC

AGATCAC
ACGCGTA
TTACCAA
TCCCAA
ATCTGTC
ATAGCTC
AGTCTC
TATCCC
CTGGCAG
TTCCAAG
ACGCAGC
CGACGTG
TGGCGGG
GGTTGGA
GTTGAAT
TTAAAT
TCAGGAT
TGGGGAA
TGAAGATC
TAAAGAA
CAAGCTA
TGGTGGC
TCTAAT
ATTATG
AGACAA
TGAAGGA
TTCAGCC
GTCAAGC
TTTAGAT
GACTACG
CAAGCT
CTTCAAG
AATACTC
TTTCTCGT
AGTCATG
TGCAGGC
TCTGCT
GCCCTTT
TGGGTGG
TCTGTCC

[illegible]

				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	TAGTAGC ATTGTGGTA ATTCGTGTT AGTTTCTTC TTTAGNCAT GGATGGGT TTCTGTATG CC'TGTTACT GGCCGCCG TAGGCGCCT CTTTGTGTT TCTTCTGTA ACTGTCCTT	MTIVWQEGAGFILSPFGLSLRDFWRLFRWPGLFLQLCSV LLQRLNSKLKATARGLQAWSAKKVGHISASLARVL HLEIAQDSRGLSPAELWFLLCLKKHSLSLSQLRTIAR LRSRIGWLREGDANTKLFMHHSRFRKKKNFVPKLVNNG AVLTSHDAKADLVNEFYGNLLGQCSDRREQJTISLESIGIP CHDLSDALDPFTEKEVWDITLHLPSDKAPGPDGFTGKFY RVWCWEIKADVMAVESAVGSRRFANFQILNTAYTLIPKF EGADQVKNFPRPISLVHSVAKLLTKLLANRLAGRLEDHV SPNQSAFIKGRFIQDNMYMLVQQTRYLYHQKQKQPRMILLK LDISKAFDSVSWPPELLEVQLHLFGQVWRDIISGLLWTS SSQVLLNGLPGQPILHRRGLRQGDPLSPFLFILVMDTLGF LINKAAQDGLLQPLAPRALQHRVSLYADDVVIFLRPA MDINVILDCLQLFGDASGLRTNVQKSSVYPICGDDLA	cd016 50	263	526
--	--	--	--	---	--	---	-------------	-----	-----

				AGAGTTC TGCCTCN AATGACT CAGACCA CTGCCCC CTCCTTTT GGGCCTG A	CTTCTTTAA TATAATATC GCGCAGTTC TCCTGCGCT TTCGAGAA AAAAAAA	LLQNMMLPCELAEPFCRYLGLPLAIKKLSKIQQVQPIIDKIA NQLSGWKAELLTKAGRKVHVQHVLTMVVIYLAAMAVD LPAWALKAIDKIRRGFLWRGRDVKGGHCHVAVGWIVC RPRELGLGISSMKELAWSLRMRWLWLAKTEPSRPWA SLAIQVKKVKDFFSMAAQTELGNGATTFWTDRLWS GHRIAEVPRLFHPPKRRVSKRTVVDAISNQSWSDIQG ALSVGIINEFLILWDLIDSIVLRPQVEDKHFERLAANGKF SSKEAYRGFFLGSSEFEFFHRIWKSAPPKTKFFMWLV AHRKVWTADRLQKRGMDHPERCPLCDQDQETLDHMLI GCVFAREFWFKLLQVNLQFLAPOSEDSAFLEWWRITLC TKVSGIARDGLNSLVILGVWTLWKQRNGCVFDNKSPSI ADAIRRVELEVHLWEMAGAKKLSLLTAPIGPAS					
					TGATGATA GGCCTCAAT TGGAATTGT CAGGGTTTA GGGAATAG TTCTAAGGT TCAGTTTCT GAAAGAAC TCATTAGAG ATGAGAAA GTTGACTTT ATTGGTTTG CAAGAAAC CCAAAAA GTCAATTTA GTGATTTCT GGTTGACTT CGCTGGCTG GTAGTAAA CCTTTTGCC TGGTTTCT GCCCGCCT AATGGTAG ATCTGGTGG TTTATTGGT GGGTTTTAA TACTGATGT GTTTGATGT	MAKKRRKAPWAEAPSSPRAPVAHGPVGVRSLVG LPSTVGFAAQGGKCAPAFASSVSWTGWSSQISPLAPTLA HSSRVLLPAATSDAAVSKRLGFPDFVGSHPHCAAF LSRLWSLPVSSRDFHGGWCVTSLRPVPEEWPRLSGS VVRSSKNLGEPMKKNIPNPSWDRWGQKDSAQGSQS WNRNRLSWRLKPQIGKSDPQEGLMSSSSGDGIGSGNPI LSPSSLKKEIDLARPAFCQKCGVEGHHARNCFNALWCDI CRKIDHTVARTCVLPKQSKPNMIPVGMADGLGFYLSHF AKPISKPKRVFIGLVKVIEGLVSAEDLAKDFGFHPWG KTWKATKCHSGFLMQFPQERLDEMISFPELKMMSG KIVVPWSSRAKPKSRLHTAWIVAEVPEELQNYQSICE IGSMIGAVEEVDLMSLSDSDIVRFKVHIKSVAMIPPVVE VAVKPFLYDIYFRIESISDEGWNETINLGKRASVDIHGI NDPVFDKSGKPRNDDDEVGDEDPXPNFKVGESSQKKE SFKKSEVFMDSKDKSGDGMRLDLSDEKVDFDPNEDD LLSSQELEEFADKMEEMHTDFQAKVGEVLSIPLTKDEK QDXVGKKSSSEILMSTSEGIRKSSRLEKNEDIKVADKAISR AEAKDAFLNKGMMNNPFSLLNASNEDLFDISNKLGVCL GSSDSSAMANLDLIRDLEMSRKILAIQSCGSKKNEEPIVD DINEVDSSVHADSEKDIDYDLLDVMVLRKGRKIKHRKK QQKKASPKIRYTPNKKREAENFXLPPDPF					
L1	LINE 1- 15_Z M	.	Zea mays						NA	NA	NA

GAGAGAAA
ATGAAATT
GGTGAATTT
ATGATTCGC
ACCTTGATT
TTTCATAGA
GAAAAAAA
TCTCATCTG
GAACTTTGT
TAATGTGTA
TGGTGCTGC
TCAGAAAG
AGAAACAA
TGCAGATTG
TTGAGTGA
GCTTCGTC
TTTTTGCTC
GAGAACTC
GTGCTCCTT
TGCTTGTTG
GTGGTGATT
TTAATATTA
TTAGAAAA
GCAGARGA
GAAAAACA
AGCTTGGT
GGTGTTAAAT
AAGTGGAG
CTTCCCTTT
TAACAGTAT
CATTGAGC
AAAAATGG
CTTGTGGAG
TTTGACTTG
AACCAATAG
ATTGTTAC
TTGGTCTAA
TAATCGGAT
TGACCTTAC
TTTTGAAAA

GTITGGATA
GATTTCIGG
CTAGTCCTG
AGTGGGAT
CTTGCTTAT
ATAATAATC
AATGTTCTG
GGTTTGAAT
AGATCCCTT
TCTGATCAT
GTTCCTTTA
GTCCTTCAA
ACGGAATGT
ATCCCTACT
GGCAGTAG
AGTTTCAA
ATATGAGTT
AGTTTGA
AGTTTGA
CCAGATTTT
CACAAAAA
AGTGTATAG
ACAATTTGT
CTATCCCTG
TGAAGGCT
AAGGTAG
TATTGACAT
TTTGAAAA
CGAAAAAT
AAACGTTTG
AAAAAAAT
GCTCAAG
TTTGGAA
ATTAATGTT
GAAGGCA
TTTACAAA
AATGAAAA
AAGAAATG
TTTGGCAA
AATTGACAT

CTTTGACAA
AATCAGTG
AAGTGGTT
GGCCTTCA
GAAATGA
CAGGATGG
AGAAACTT
GACCTGGA
ATTGACTCT
CAAAAAAT
TAGTATG
AGGAGGGA
ATTAAAGCTC
AAACAAG
AGCTCGTG
ATAAATCA
TTTGGATG
GCGATGAG
AATTCCTCAA
TAATTCAT
CTTTAGCC
AAGTGTA
GAGAGA
AGCTTAAA
ATTGACT
CTTACTCAT
GAAGACAA
AGTTGCTCA
AGATGATG
ATGAAAT
AATCATTTG
CGCATCTCT
CAAGAAATC
TAGTATTTC
TTTGAATCA
TTTGAATC
GAAAATGT

TTGGATGAT
ACTGACAG
AGATTTCCT
GACACAGC
CGTTTGACA
TGGAGGAA
ATTAGAGA
AGTGGTGT
TTCTATGAA
ACATAATA
GAGCTCTG
GCTCTGATA
GCATCCCA
GCTCAATT
TTTTCAGGAG
TTTTTGGAC
ACGATTAA
AAATGATCT
TTTTCACCT
CTTCAAGA
TTTTTCATGA
TTGGCACTCT
TAACATTGA
GAGACCTA
ATTTCGAA
TTGTTGACTC
TTTATTCCTA
AAGTTCCTA
ATGCTATTCG
ACATTAAA
GCCTTCAGA
CCCTTATTCG
TTTCTTAAT
TGTTGTTAC
AAGATTAT
ACCAAAAGT
TTTATGCCAA
TTGATTCGA
CCAGTTTGG
TCAATAATG

TGATTAGTG
ATTTTCAGT
ATGGTTTA
TCAAAGGT
AAATACAT
CATGGATG
GAGTGGTCT
CTCTCCATG
AGATTATTC
ATGAGGTG
AAAAAGAA
GAAGCAGA
GTGGGGTG
ATTTTAAG
GTGGATTTT
GAGAAAGC
TTATGACAA
GGTTAATTG
GAAC TTCT
TCATCATAT
GATGATTA
AAAAAGGT
TTTGGTGAC
AAATGGTG
TGACTGGGT
TATGAGAA
CTGTTAGAG
GTGGTAAA
GTGGCCATC
AAGACAAA
TGACAGGG
TTGGTCCCT
ATTTACCA
CTCACAAA
GGTGTCAG
ACAAGGTG
ATCCTTCT
CCCCCTGCT
TTTTAATAT
TGCTGCTGA

TGGGCTTGC
TTGTATGAT
TCAAAAAG
CCAAAAGAT
GAAGGTAT
TATTAGGG
GACTCATTC
CTCATATCA
TCCCTAATG
GTTGTGTT
GTTTGCAAT
ATGCTGATG
ATACCATTT
TTTTGCTTC
AAGATGAT
CTGGAGGG
AGCCAGAA
ATCTAAGT
TTATCTTT
GCTTATTG
AACAGATG
TCTGGTCTC
AAAATCAA
TTTTATAA
AAGTGAA
TTTTCTGT
TAGGTGAA
GCTGTGAT
AAGGAAAT
GTGATATG
TGACATTT
TACCTGCC
TTATAATG
TCTTCTAT
GAAATATT
GGGGGTTC
CGATGATG
ACAAAAG
CTTGAAG
TCTTGIGG

TTACCCACA
GTAGAGAA
AGTTGAAA
AATGGGTC
TTGGCAAG
GAAAATTC
CTTAGTTG
GGTGGTCG
GTTAGTCTT
GATTAATA
CCAGTCTTA
CCAATGTCC
CTTTATACA
TGCTCTCGA
TGTATAAA
GCTCCCAA
GAATATTT
GAAGAAAA
TGGATCTCT
TTAGAAAA
AGGCTTTTG
TGGCAAGG
TGGCCATAC
TACAAAAA
AATAATAT
TGGCTGATT
GGAAATG
GTTTGTCT
CCGAGAGG
TCAGGGTG
GCCTTGGG
GTCTTGGAT
CTGCACAA
GATGAATG
AGGCGCTTC
TTGCTAAAT
GGATTGG
AAATTGGA
AAATACGA
ATGGGCTW

TGGCAGAC
CATTATCAG
ACATAAAT
ATGTAAA
GGAAAGGC
TATTATTTC
ATTTAAAA
AAGACAAA
GTGATCTC
ACTTTGGA
AAGGATTTC
TGGATGCTC
GGGACAAT
TTTTACAAA
TATTGTAAA
AAAAGGTA
GGCAATGG
CAATAATA
CCAGCTTTT
GGCACAAT
AATTGGTGT
GATTTAGA
ACCTCTGTC
TTCTAAATA
TCCTAGTTT
GTTTGATAT
TGCTTAAGA
TAAAGATA
TCACAGTG
GAAAAAGC
CTTTCTTC
TGATTTTAG
AGCTGTGA
AATTTAGAA
GAAGGATT
TATTGGTGT
TGTGAAGA
TTGATTAAA
TAACCTTAT
TACACAGT

GTATAAAT
ATTACTATT
TCTGAGGA
GGATCATATA
GGCAGCTATT
AGCTCTCTGG
GTAAFAAA
GGTTTTCT
GTTAAATCT
TTTTACAAA
GAAATCAA
ATGTGCTCA
GATTCCTTT
TCTCTGTG
GCCCTCTAG
TTTCTGTG
GAAGACTA
GATCTCCAC
ATAAAATC
AAAGCTTT
TTGTGGCTT
GTTATGCAT
AAAAAGAT
TTTAAACCA
GGTCAATTT
GTTCAAAA
AAAAATGG
AAGGGCAA
TCTGTGATTG
TACTTTTG
AGAAATTC
TGAATCCA
TAGATCAAT
TTGTTTTCA
GTGCTCGGT
TGCTAGATT
TACTTGGAG
AATAGTCTA
CTCGCTCT
TGGCTTAAT
TCTATTCC

CAACAGTGT
TGGATGATC
TGTTTGGCT
CTTGGATTA
ATAGTTTAA
ACAAAAC
GAAAAGAA
TCTGTCCT
TTTTGGTTG
TGGAGCTGT
CATTGGGC
GATCTGGC
GATGTCGC
AATGATTTG
TGCTTTGGA
ACTGACAG
AATTGATG
ATCCGACTA
ATGTGATTT
TCTCTGTT
GCTATTGGC
TTGATACTT
GGGCTATCC
GTCGAAA
AAGAAGGA
AAGAAAT
TGGTGGAA
CAAGGAAG
CCTCAAAAT
CAGAAAGA
TGGCTGTG
AGATCTAC
AACAAGAC
TCGTGGTTG
GAAGCCAG
TCGACAGA
CGGATCCA
ATGAAGT
TCTGCTGTG
TTTGGCTAT

TTTGGTTCA
TCAACTTGG
TGTTTGCTG
TAAATCCCC
CTGCCTCG
GTAGCTTCA
GGTACTG
AAAACTTG
GGCTATATC
GGGYTGA
AGTTGTTA
GTTTCGTGC
TACTCTAGT
TTGATCTTT
GCCATCTTT
GTTTTTGGT
CAGCTATGT
GCTGATCTT
CTGCTGCTT
GTGCTAGTT
TGTTTCGTT
CAGTTTATG
TTCCGTCGT
AGCTGATA
CTTGTTGTT
GTTGTCGGC
TTTTAAATT
CTGGTGGC
TCTATGGTC
TTGTAAGA
ACTTGTGTC
TCTTCTAT
TTCGTAATG
AAATAGGG
GGTATCCC
CTCTGTCAA
AAAA

Table 44 provides Retroviral reverse transcriptase domains for use in Gene Writer polypeptides. Wild-type reverse transcriptase enzymes were collected and prioritized as according to the descriptions herein (see Example 33). The Type column indicates whether the sequence corresponds to a wild-type sequence (“root”) or comprises mutations that may improve the activity of the enzyme (“derivative”).

5

Table 44. Retroviral reverse transcriptase domains for use in Gene Writer polypeptides. In some embodiments, a system or method described herein involves a reverse transcriptase domain having an amino acid sequence according to a reverse transcriptase domain of Table 44, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or functional fragment thereof.

10

name	virus name	unit ID	type	peptide
AVIRE_P03360	AVIRE	P03360	root	TAPLEEEYRLFI EAPIQNVTLL EQWKREIPKVV AEINPPGLASTQAPIHVQLLSTALPVRVRQYPITL EAKRSLRETIRKFR AAGILRPVHSPWNITPLL PVRKSGTSEYRMVQDLREVNKRVEITHPTVPNPYTL SLLPPDRIWYSVL DLDKDAFFCPLAPESQLFAFEWADAEEGESQLTWTRL PQGFKNSPTL FDEALNRDLQGHRLDHPSVSLQYVDDLLJAAADTQAACL SATRDLLMTLAEI GYRVS GKKAKQLCOEEV TYLGFKIHGSRSLSNSRTQAIL QIPVPKTKRQVREFLGTTGYCRLWIPGFAELAQPL YAA TRGNDPL VWGEKEEAFQSLKAL TQPPALAPSLDKPFQFVEETSGAAGVLTQALGPWKRVPVAYLSKRLLDPVAAAGWPRCLRAIAAAALL TREASKLTFGQDIETSSINLESLLRSPDPKWL TNARITQYQVILLDPVRVRKQTAAALNPATLLPETDDTLPIHICLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDVSVWAEPLPIGTS AQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMVYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVA VMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS
AVIRE_P03360_3mut	AVIRE	P03360	derivative	TAPLEEEYRLFI EAPIQNVTLL EQWKREIPKVV AEINPPGLASTQAPIHVQLLSTALPVRVRQYPITL EAKRSLRETIRKFR AAGILRPVHSPWNITPLL PVRKSGTSEYRMVQDLREVNKRVEITHPTVPNPYTL SLLPPDRIWYSVL DLDKDAFFCPLAPESQLFAFEWADAEEGESQLTWTRL PQGFKNSPTL FNEALNRDLQGHRLDHPSVSLQYVDDLLJAAADTQAACL SATRDLLMTLAEI GYRVS GKKAKQLCOEEV TYLGFKIHGSRSLSNSRTQAIL QIPVPKTKRQVREFLGTTGYCRLWIPGFAELAQPL YAA TRGNDPL VWGEKEEAFQSLKAL TQPPALAPSLDKPFQFVEETSGAAGVLTQALGPWKRVPVAYLSKRLLDPVAAAGWPRCLRAIAAAALL TREASKLTFGQDIETSSINLESLLRSPDPKWL TNARITQYQVILLDPVRVRKQTAAALNPATLLPETDDTLPIHICLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDVSVWAEPLPIGTS AQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMVYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVA VMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS

[illegible]

					SEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVLTPQFSPDPPIPAALCLFSDGATGRGAYCLWKDHLLDFQA VPAPESAQKGELAGLGLAAAPP EPVNIWVDSKYLISLLRTLVGLAWLPDPVPVSALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDOL
BLVJ_P0 3361_2mu t	BL VJ	P03 361	der iva tiv e		GVLDTPPSHIGLEHLPPPEVPOFPLNLERLQALQDLVHRSLEAGYISPDGNGNNPVVPKNGAWRFVHDLRAITNALTKPIPALSPGPPDLTAIPTHP HIICLDLKDADFQIPVEDRFRFYLSTLPSGGLQPHRRFAWRVLPOGFINSFQSLVSYMDIDILYASPTTEEQRSCYQALAA RLRLDGFQVASEKTSQTPSPVPLGQMVHEQIVTYSQSLPTLQISSPSILHQLQAVLGDILQVWSRGTPPTRPLQLLYSSLRHHDPRAIQLSPELOQIAE LRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQGAQFPLAYFQPTLTDNQASPGWGLLLGCQYLQTOALSSYAKPILKYHNHLPKTSLDNWIQS SEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVLTPQFSPDPPIPAALCLFSDGATGRGAYCLWKDHLLDFQA VPAPESAQKGELAGLGLAAAPP EPVNIWVDSKYLISLLRTLVGLAWLPDPVPVSALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDOL
BLVJ_P0 3361_2mu tB	BL VJ	P03 361	der iva tiv e		GVLDTPPSHIGLEHLPPPEVPOFPLNLERLQALQDLVHRSLEAGYISPDGNGNNPVVPKNGAWRFVHDLRAITNALTKPIPALSPGPPDLTAIPTHP PHIICLDLKDADFQIPVEDRFRFYLSTLPSGGLQPHRRFAWRVLPOGFINSFQSLVSYMDIDILYASPTTEEQRSCYQALAA ARLRLDGFQVASEKTSQTPSPVPLGQMVHEQIVTYSQSLPTLQISSPSILHQLQAVLGDILQVWSRGTPPTRPLQLLYSSLRHHDPRAIQLSPELOQIAE ELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQGAQFPLAYFQPTLTDNQASPGWGLLLGCQYLQTOALSSYAKPILKYHNHLPKTSLDNWIQ SSDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVLTPQFSPDPPIPAALCLFSDGATGRGAYCLWKDHLLDFQA VPAPESAQKGELAGLGLAAAPP PEPVNIWVDSKYLISLLRTLVGLAWLPDPVPVSALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDOL
FFV_O93 209	FF V	O9 320 9	ro ot		MDLLKPLTVRKGVKIKGYWNSQADITCPKDLQGEPEVRQNNVTTHGTQEGDVYVNLKIDGRRINTEVIGTTLDYAITPDGVPVWILKKPLELTIK LDLEEQGTLLNNSILSKGKEELKQLEKYSALWQSWENOVGHRIRPHIKIATGTVKPTQKQYHINPKAKPDQIIVINDLLKQGV PVPKNGRWRMVL DYRAVNKVTPLIAVQNHQSYGILGSLFKGRYKTITDLSNGFWAHPVPEDYWTFTWQKGQYCWTVLPQGF DILLQGINVEVYVDDVYISHDSEKEHLEYLDLNFRLKEAGYISLKKSNANSIVDFLGFQITNEGRGLDTITFEKLENTAAPTTLKQL PDFTELIAPLYALIPKSTKNYPWQIEHSTLTITLTKNGAEYLOQRKGDKTLIMKVNASYTTGTIRYVYNEGEKKPISYVSVFSKTELEKLLTTV HKGLLKALDLSMGQNIHVSPIVSMQNIQKTPQAKKALASRWLSWLSYLEDPRIREFYDQMPALKDLPAVDTGKDNKKHPSNFQ TKEGHLNAGMGIVYFINKDGNLQKQEWISLGNHTAQFAEIAFEFALKKCLPLGGNLLVVTDSNYSYVAKAYNEELDVMWASNGFVN KSVADLKRRLRPDVVVTHEPGHKLDSPPHAYGNLLADQLATQASEKVVH
FFV_O93 209-Pro	FF V	O9 320 9	der iva tiv e		VPWILKKPLELTIKDLLEEQGTLLNNSILSKGKEELKQLEKYSALWQSWENOVGHRIRPHIKIATGTVKPTQKQYHINPKAKPDQIIVINDLLKQGV LIQKESTMNTPVVPKPNGRWRMVL DYRAVNKVTPLIAVQNHQSYGILGSLFKGRYKTITDLSNGFWAHPVPEDYWTFTWQKGQYCWTVLPQGF LNSPGLFNGDVVDLLQGINVEVYVDDVYISHDSEKEHLEYLDLNFRLKEAGYISLKKSNANSIVDFLGFQITNEGRGLDTITFEKLENTAAPTTLKQL QSLGGLNFARNIPDFTELIAPLYALIPKSTKNYPWQIEHSTLTITLTKNGAEYLOQRKGDKTLIMKVNASYTTGTIRYVYNEGEKKPISYVSVFSKTE LKFTLEKLLTTVHKGLLKALDLSMGQNIHVSPIVSMQNIQKTPQAKKALASRWLSWLSYLEDPRIREFYDQMPALKDLPAVDTGKDNKKHPSNFQ HIFYTDGSAITSPTEGHLNAGMGIVYFINKDGNLQKQEWISLGNHTAQFAEIAFEFALKKCLPLGGNLLVVTDSNYSYVAKAYNEELDVMWASNGFVN NRKKPLKHISKWKSVAADLKRRLRPDVVVTHEPGHKLDSPPHAYGNLLADQLATQASEKVVH
FFV_O93 209-Pro_2mut	FF V	O9 320 9	der iva tiv e		VPWILKKPLELTIKDLLEEQGTLLNNSILSKGKEELKQLEKYSALWQSWENOVGHRIRPHIKIATGTVKPTQKQYHINPKAKPDQIIVINDLLKQGV LIQKESTMNTPVVPKPNGRWRMVL DYRAVNKVTPLIAVQNHQSYGILGSLFKGRYKTITDLSNGFWAHPVPEDYWTFTWQKGQYCWTVLPQGF LNSPGLFNGDVVDLLQGINVEVYVDDVYISHDSEKEHLEYLDLNFRLKEAGYISLKKSNANSIVDFLGFQITNEGRGLDTITFEKLENTAAPTTLKQL QSLGGLNFARNIPDFTELIAPLYALIPKSTKNYPWQIEHSTLTITLTKNGAEYLOQRKGDKTLIMKVNASYTTGTIRYVYNEGEKKPISYVSVFSKTE LKFTLEKLLTTVHKGLLKALDLSMGQNIHVSPIVSMQNIQKTPQAKKALASRWLSWLSYLEDPRIREFYDQMPALKDLPAVDTGKDNKKHPSNFQ HIFYTDGSAITSPTEGHLNAGMGIVYFINKDGNLQKQEWISLGNHTAQFAEIAFEFALKKCLPLGGNLLVVTDSNYSYVAKAYNEELDVMWASNGFVN NRKKPLKHISKWKSVAADLKRRLRPDVVVTHEPGHKLDSPPHAYGNLLADQLATQASEKVVH
FFV_O93 209-Pro_2mut A	FF V	O9 320 9	der iva tiv e		VPWILKKPLELTIKDLLEEQGTLLNNSILSKGKEELKQLEKYSALWQSWENOVGHRIRPHIKIATGTVKPTQKQYHINPKAKPDQIIVINDLLKQGV LIQKESTMNTPVVPKPNGRWRMVL DYRAVNKVTPLIAVQNHQSYGILGSLFKGRYKTITDLSNGFWAHPVPEDYWTFTWQKGQYCWTVLPQGF LNSPGLFNGDVVDLLQGINVEVYVDDVYISHDSEKEHLEYLDLNFRLKEAGYISLKKSNANSIVDFLGFQITNEGRGLDTITFEKLENTAAPTTLKQL QSLGGLNFARNIPDFTELIAPLYALIPKSTKNYPWQIEHSTLTITLTKNGAEYLOQRKGDKTLIMKVNASYTTGTIRYVYNEGEKKPISYVSVFSKTE LKFTLEKLLTTVHKGLLKALDLSMGQNIHVSPIVSMQNIQKTPQAKKALASRWLSWLSYLEDPRIREFYDQMPALKDLPAVDTGKDNKKHPSNFQ HIFYTDGSAITSPTEGHLNAGMGIVYFINKDGNLQKQEWISLGNHTAQFAEIAFEFALKKCLPLGGNLLVVTDSNYSYVAKAYNEELDVMWASNGFVN NRKKPLKHISKWKSVAADLKRRLRPDVVVTHEPGHKLDSPPHAYGNLLADQLATQASEKVVH
FFV_O93 209_2mut	FF V	O9 320 9	der iva tiv e		MDLLKPLTVRKGVKIKGYWNSQADITCPKDLQGEPEVRQNNVTTHGTQEGDVYVNLKIDGRRINTEVIGTTLDYAITPDGVPVWILKKPLELTIK LDLEEQGTLLNNSILSKGKEELKQLEKYSALWQSWENOVGHRIRPHIKIATGTVKPTQKQYHINPKAKPDQIIVINDLLKQGV PVPKNGRWRMVL DYRAVNKVTPLIAVQNHQSYGILGSLFKGRYKTITDLSNGFWAHPVPEDYWTFTWQKGQYCWTVLPQGF DILLQGINVEVYVDDVYISHDSEKEHLEYLDLNFRLKEAGYISLKKSNANSIVDFLGFQITNEGRGLDTITFEKLENTAAPTTLKQL PDFTELIAPLYALIPKSTKNYPWQIEHSTLTITLTKNGAEYLOQRKGDKTLIMKVNASYTTGTIRYVYNEGEKKPISYVSVFSKTELEKLLTTV HKGLLKALDLSMGQNIHVSPIVSMQNIQKTPQAKKALASRWLSWLSYLEDPRIREFYDQMPALKDLPAVDTGKDNKKHPSNFQ HIFYTDGSAITSPTEGHLNAGMGIVYFINKDGNLQKQEWISLGNHTAQFAEIAFEFALKKCLPLGGNLLVVTDSNYSYVAKAYNEELDVMWASNGFVN NRKKPLKHISKWKSVAADLKRRLRPDVVVTHEPGHKLDSPPHAYGNLLADQLATQASEKVVH

[illegible]

						VFSK AELKFSM L E K L L T T M H K A L I K A M D L A M G Q E I L V Y S P I V S M T K I Q K T P L P E R K A L P I R W I T W M T Y L E D P R I Q I H Y D K T L P E L K H I P D V Y T S S Q S P V K H P S Q Y E G V F Y T D G S A I K S P D P T K S N N A G M G I V H A T Y K P E Y Q V L N Q W S I P L G N H T A Q M A E I A A V E F A C K K A L K I P G P V L V I T D S F Y V A E S A N K E L P Y W K S N G F V N N K K P L K H I S K W K S I A E C L S M K P D J T I Q H E K G I S I Q I P V E I L K G N A L A D K L A T Q G S Y V V N
FOAMV_P14350_Pro.2mutA	FOAMV_P14350_2mutA	AM	P14	350	der iva tiv e	V P W L T Q Q P L Q L T I L V P L Q E Y Q E K I L S K T A L P E D Q K Q Q L K T L F V K Y D N L W Q H W E N Q V G H R K I R P H N I A T G D Y P R P Q K Q Y P I N P K A K S I Q I V I D D L L K Q G V L T P Q N S T M I N T P Y P V P K P D G R W M V L D Y R E V N K T I P L T A A Q N Q H S A G I L A T I V R Q K Y K T I L D L A N G F W A H P I T P E S Y W L T A F T W Q G K Q Y C W T R L P Q G H L N S P A L E N A D V D L L K E I P N V Q V Y V D D I Y L S H D D P K E H V Q Q L E K V F Q I L L Q A G Y V V S L K S E I G Q K T V E H L F G N I T K E G R G L T D T F K T K L L N I T P P K D L K Q L Q S I L G K L N F A R N F I P N F A E L V Q P L Y N L I A P A K G K Y E W S E E N T K Q L N M V I E A L N T A S N L E E R L P E Q R L V I K V N T S P S A G Y V R Y N E T G K K P I M Y L N Y V F S K A E L K F S M L E K L L T T M H K A L I K A M D L A M G Q E I L V Y S P I V S M T K I Q K T P L P E R K A L P I R W I T W M T Y L E D P R I Q I H Y D K T L P E L K H I P D V Y T S S Q S P V K H P S Q Y E G V F Y T D G S A I K S P D P T K S N N A G M G I V H A T Y K P E Y Q V L N Q W S I P L G N H T A Q M A E I A A V E F A C K K A L K I P G P V L V I T D S F Y V A E S A N K E L P Y W K S N G F V N N K K P L K H I S K W K S I A E C L S M K P D J T I Q H E K G I S I Q I P V E I L K G N A L A D K L A T Q G S Y V V N
FOAMV_P14350_2mut	FOAMV_P14350_2mut	AM	P14	350	der iva tiv e	M N P L Q L L Q P L A E I K G T K L L A H W N S G A T I T C I P E S H E D E Q P I K K T L I K T I H G E K Q Q N V Y V T F K V K G R K V E A E V I A S P Y E Y I L L S P T D V P W L T Q Q P L Q L T I L V P L Q E Y Q E K I L S K T A L P E D Q K Q Q L K T L F V K Y D N L W Q H W E N Q V G H R K I R P H N I A T G D Y P R P Q K Q Y P I N P K A K S I Q I V I D D L L K Q G V L T P Q N S T M I N T P Y P V P K P D G R W M V L D Y R E V N K T I P L T A A Q N Q H S A G I L A T I V R Q K Y K T I L D L A N G F W A H P I T P E S Y W L T A F T W Q G K Q Y C W T R L P Q G H L N S P A L E N A D V D L L K E I P N V Q V Y V D D I Y L S H D D P K E H V Q Q L E K V F Q I L L Q A G Y V V S L K S E I G Q K T V E H L F G N I T K E G R G L T D T F K T K L L N I T P P K D L K Q L Q S I L G K L N F A R N F I P N F A E L V Q P L Y N L I A P A K G K Y E W S E E N T K Q L N M V I E A L N T A S N L E E R L P E Q R L V I K V N T S P S A G Y V R Y N E T G K K P I M Y L N Y V F S K A E L K F S M L E K L L T T M H K A L I K A M D L A M G Q E I L V Y S P I V S M T K I Q K T P L P E R K A L P I R W I T W M T Y L E D P R I Q I H Y D K T L P E L K H I P D V Y T S S Q S P V K H P S Q Y E G V F Y T D G S A I K S P D P T K S N N A G M G I V H A T Y K P E Y Q V L N Q W S I P L G N H T A Q M A E I A A V E F A C K K A L K I P G P V L V I T D S F Y V A E S A N K E L P Y W K S N G F V N N K K P L K H I S K W K S I A E C L S M K P D J T I Q H E K G I S I Q I P V E I L K G N A L A D K L A T Q G S Y V V N
FOAMV_P14350_2mutA	FOAMV_P14350_2mutA	AM	P14	350	der iva tiv e	M N P L Q L L Q P L A E I K G T K L L A H W N S G A T I T C I P E S H E D E Q P I K K T L I K T I H G E K Q Q N V Y V T F K V K G R K V E A E V I A S P Y E Y I L L S P T D V P W L T Q Q P L Q L T I L V P L Q E Y Q E K I L S K T A L P E D Q K Q Q L K T L F V K Y D N L W Q H W E N Q V G H R K I R P H N I A T G D Y P R P Q K Q Y P I N P K A K S I Q I V I D D L L K Q G V L T P Q N S T M I N T P Y P V P K P D G R W M V L D Y R E V N K T I P L T A A Q N Q H S A G I L A T I V R Q K Y K T I L D L A N G F W A H P I T P E S Y W L T A F T W Q G K Q Y C W T R L P Q G H L N S P A L E N A D V D L L K E I P N V Q V Y V D D I Y L S H D D P K E H V Q Q L E K V F Q I L L Q A G Y V V S L K S E I G Q K T V E H L F G N I T K E G R G L T D T F K T K L L N I T P P K D L K Q L Q S I L G K L N F A R N F I P N F A E L V Q P L Y N L I A P A K G K Y E W S E E N T K Q L N M V I E A L N T A S N L E E R L P E Q R L V I K V N T S P S A G Y V R Y N E T G K K P I M Y L N Y V F S K A E L K F S M L E K L L T T M H K A L I K A M D L A M G Q E I L V Y S P I V S M T K I Q K T P L P E R K A L P I R W I T W M T Y L E D P R I Q I H Y D K T L P E L K H I P D V Y T S S Q S P V K H P S Q Y E G V F Y T D G S A I K S P D P T K S N N A G M G I V H A T Y K P E Y Q V L N Q W S I P L G N H T A Q M A E I A A V E F A C K K A L K I P G P V L V I T D S F Y V A E S A N K E L P Y W K S N G F V N N K K P L K H I S K W K S I A E C L S M K P D J T I Q H E K G I S I Q I P V E I L K G N A L A D K L A T Q G S Y V V N
GALV_P21414	GALV_P21414	GA	P21	414	ro ot	V L N L E E Y R L H E K P V P S S I D P S W L Q L F P T V W A E R A G M G L A N Q V P P V V E L R S G A S P V A V R Q Y P M S K A R E G I R P H I Q K F L D L G V L V P C R S P W N T P L L P V K K P G T N D Y R P V Q D L R E I N K R V Q D I H P T V P N P Y N L L S S L P S Y T W Y S V L D K D A F C L R L H P N S O P L F A F E W K D P E K G N T G Q L T W T R L P Q G F K N S P T L F D E A L H R D L A P R A L N P Q V V I L Q Y V D D L L V A A P T Y E D C K G T Q K L Q E L S K L G Y R V S A K K A Q L C Q R E V T Y L G Y L L K E G K R W L T P A R K A T V M K I P V P T P R Q V R E F L G T A G F C R L W I P G F A S L A A P L Y P L T K E S I P F I W T E E H Q A F D H I K K A L S A P A L A P D L T K P T L Y I D E R A G V A R G V L T Q T L G P W R R P V A Y L S K K L D P V A S G W P T C L K A V A A V A L L K D A D K L T L G Q N V T V I A S H S L E S I V R Q P P D R W M T N A R M T H Y Q S L L N E R V S F A P P A V L N P A T L L P V E S A T P V H R C S E I L A E E T G T R R D L E D Q P L G V P T W Y T D G S S F T E G K R R A G P I V D G K R T V W A S S L P E G T S A Q K A E L V A L T Q A L R L A E G K N I N Y T D S R Y A F A T A H I H G A I Y K Q R G L L T S A G K D I K N K E I L A L L E A I H L P R R V A I I H C P H Q R G S N P V A T G N R R A D E A A K Q A A L S T R V L A G T T K P
GALV_P21414_3mut	GALV_P21414_3mut	GA	P21	414	der iva tiv e	V L N L E E Y R L H E K P V P S S I D P S W L Q L F P T V W A E R A G M G L A N Q V P P V V E L R S G A S P V A V R Q Y P M S K A R E G I R P H I Q K F L D L G V L V P C R S P W N T P L L P V K K P G T N D Y R P V Q D L R E I N K R V Q D I H P T V P N P Y N L L S S L P S Y T W Y S V L D K D A F C L R L H P N S O P L F A F E W K D P E K G N T G Q L T W T R L P Q G F K N S P T L F N E A L H R D L A P R A L N P Q V V I L Q Y V D D L L V A A P T Y E D C K G T Q K L Q E L S K L G Y R V S A K K A Q L C Q R E V T Y L G Y L L K E G K R W L T P A R K A T V M K I P V P T P R Q V R E F L G T A G F C R L W I P G F A S L A A P L Y P L T K E S I P F I W T E E H Q A F D H I K K A L S A P A L A P D L T K P T L Y I D E R A G V A R G V L T Q T L G P W R R P V A Y L S K K L D P V A S G W P T C L K A V A A V A L L K D A D K L T L G Q N V T V I A S H S L E S I V R Q P P D R W M T N A R M T H Y Q S L L N E R V S F A P P A V L N P A T L L P V E S A T P V H R C S E I L A E E T G T R R D L E D Q P L G V P T W Y T D G S S F T E G K R R A G P I V D G K R T V W A S S L P E G T S A Q K A E L V A L T Q A L R L A E G K N I N Y T D S R Y A F A T A H I H G A I Y K Q R G L L T S A G K D I K N K E I L A L L E A I H L P R R V A I I H C P H Q R G S N P V A T G N R R A D E A A K Q A A L S T R V L A G T T K P
GALV_P21414_3mutA	GALV_P21414_3mutA	GA	P21	414	der iva tiv e	V L N L E E Y R L H E K P V P S S I D P S W L Q L F P T V W A E R A G M G L A N Q V P P V V E L R S G A S P V A V R Q Y P M S K A R E G I R P H I Q K F L D L G V L V P C R S P W N T P L L P V K K P G T N D Y R P V Q D L R E I N K R V Q D I H P T V P N P Y N L L S S L P S Y T W Y S V L D K D A F C L R L H P N S O P L F A F E W K D P E K G N T G Q L T W T R L P Q G F K N S P T L F N E A L H R D L A P R A L N P Q V V I L Q Y V D D L L V A A P T Y E D C K G T Q K L Q E L S K L G Y R V S A K K A Q L C Q R E V T Y L G Y L L K E G K R W L T P A R K A T V M K I P V P T P R Q V R E F L G T A G F C R L W I P G F A S L A A P L Y P L T K E S I P F I W T E E H Q A F D H I K K A L S A P A L A P D L T K P T L Y I D E R A G V A R G V L T Q T L G P W R R P V A Y L S K K L D P V A S G W P T C L K A V A A V A L L K D A D K L T L G Q N V T V I A S H S L E S I V R Q P P D R W M T N A R M T H Y Q S L L N E R V S F A P P A V L N P A T L L P V E S A T P V H R C S E I L A E E T G T R R D L E D Q P L G V P T W Y T D G S S F T E G K R R A G P I V D G K R T V W A S S L P E G T S A Q K A E L V A L T Q A L R L A E G K N I N Y T D S R Y A F A T A H I H G A I Y K Q R G L L T S A G K D I K N K E I L A L L E A I H L P R R V A I I H C P H Q R G S N P V A T G N R R A D E A A K Q A A L S T R V L A G T T K P
HTL1A_P0362	HTL1A_P0362	HT	P03	362	ro ot	A V L G L E H L P R P Q I S P L N P E R I Q A L Q H L V R K A L E A G H I E P Y T G P G N S P T L F E M Q L A H I L Q P R Q A P Q C T I Q Y M D D I L L A S P S H E D I L L S E A T M A S L S H G R D A F T Q P L P K Q F Q P Y A F T V P Q C N Y G P G T R Y A W K V L P Q G F K N S P T L F E M Q L A H I L Q P R Q A P Q C T I Q Y M D D I L L A S P S H E D I L L S E A T M A S L S H G L P V S E N K T Q Q I P G T I K F L Q G I S P N H L Y D A V P T P R S R W A L P E I Q A L L G E I Q W V S K G T P T L R Q P L H S L Y C A L Q R I H T D P R D Q I Y L N P S Q V S L V Q L R Q A L

				SONCRSLVQITPLPLGAIMLTLTGTTTVFQSKQWPLVWLHAPLPHITSQCPWGQLASAVLLLDKTYTLQSYGLLCCQTHHNISTQTFNQFIQTSIDHPSV PILLHSHRFKNLGAQTGELWNTHLKTAAPLAPVKALMPVFTLSPVINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSA RSWRCLNIFLDSKYLHYHLRTLALGTFQGRSSQAPEQALLPRLSRKVYLHIVRSHNTLPDISRLNALTADALLITPVLQL
HTL1A_P 03362_2m ut	HT L1 A	P03 362	der iva tiv e	AVLGLEHLPRPEISOFPPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVPVKANGTWRFHIDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQITIDL RDAFTQIPLPKQFQPYFAFTVPOQCNYGPGTRYAWKVLPOGFKNSPTLFQMQLAHILQPIRQAFPOCTILQYMDILLASPSHEDILLSEATMASLISHG LPVSENKTQQTPTGTHKELGQISPNHLYTDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHIDPRDQIYLNPSQVQSLVQLRQAL SONCRSLVQITPLPLGAIMLTLTGTTTVFQSKQWPLVWLHAPLPHITSQCPWGQLASAVLLLDKTYTLQSYGLLCCQTHHNISTQTFNQFIQTSIDHPSV PILLHSHRFKNLGAQTGELWNTHLKTAAPLAPVKALMPVFTLSPVINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSA RSWRCLNIFLDSKYLHYHLRTLALGTFQGRSSQAPEQALLPRLSRKVYLHIVRSHNTLPDISRLNALTADALLITPVLQL
HTL1A_P 03362_2m utB	HT L1 A	P03 362	der iva tiv e	AVLGLEHLPRPEISOFPPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVPVKANGTWRFHIDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQITIDL RDAFTQIPLPKQFQPYFAFTVPOQCNYGPGTRYAWKVLPOGFKNSPTLFQMQLAHILQPIRQAFPOCTILQYMDILLASPSHEDILLSEATMASLISHG LPVSENKTQQTPTGTHKELGQISPNHLYTDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHIDPRDQIYLNPSQVQSLVQLRQAL SONCRSLVQITPLPLGAIMLTLTGTTTVFQSKQWPLVWLHAPLPHITSQCPWGQLASAVLLLDKTYTLQSYGLLCCQTHHNISTQTFNQFIQTSIDHPSV PILLHSHRFKNLGAQTGELWNTHLKTAAPLAPVKALMPVFTLSPVINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSA RSWRCLNIFLDSKYLHYHLRTLALGTFQGRSSQAPEQALLPRLSRKVYLHIVRSHNTLPDISRLNALTADALLITPVLQL
HTL1C_P 14078	HT L1 C	P14 078	ro ot	AVLGLEHLPRPEISOFPPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVPVKANGTWRFHIDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQITIDL KDAFTQIPLPKQFQPYFAFTVPOQCNYGPGTRYAWKVLPOGFKNSPTLFQMQLAHILQPIRQAFPOCTILQYMDILLASPSHEDILLSEATMASLISHG LPVSENKTQQTPTGTHKELGQISPNHLYTDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHIDPRDQIYLNPSQVQSLVQLRQAL SONCRSLVQITPLPLGAIMLTLTGTTTVFQSKQWPLVWLHAPLPHITSQCPWGQLASAVLLLDKTYTLQSYGLLCCQTHHNISTQTFNQFIQTSIDHPSV PILLHSHRFKNLGAQTGELWNTHLKTAAPLAPVKALMPVFTLSPVINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSA RSWRCLNIFLDSKYLHYHLRTLALGTFQGRSSQAPEQALLPRLSRKVYLHIVRSHNTLPDISRLNALTADALLITPVLQL
HTL1C_P 14078_2m ut	HT L1 C	P14 078	der iva tiv e	AVLGLEHLPRPEISOFPPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVPVKANGTWRFHIDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQITIDL KDAFTQIPLPKQFQPYFAFTVPOQCNYGPGTRYAWKVLPOGFKNSPTLFQMQLAHILQPIRQAFPOCTILQYMDILLASPSHEDILLSEATMASLISHG LPVSENKTQQTPTGTHKELGQISPNHLYTDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHIDPRDQIYLNPSQVQSLVQLRQAL SONCRSLVQITPLPLGAIMLTLTGTTTVFQSKQWPLVWLHAPLPHITSQCPWGQLASAVLLLDKTYTLQSYGLLCCQTHHNISTQTFNQFIQTSIDHPSV PILLHSHRFKNLGAQTGELWNTHLKTAAPLAPVKALMPVFTLSPVINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSA RSWRCLNIFLDSKYLHYHLRTLALGTFQGRSSQAPEQALLPRLSRKVYLHIVRSHNTLPDISRLNALTADALLITPVLQL
HTL1C_P 14078_2m utB	HT L1 C	P14 078	der iva tiv e	AVLGLEHLPRPEISOFPPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVPVKANGTWRFHIDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQITIDL KDAFTQIPLPKQFQPYFAFTVPOQCNYGPGTRYAWKVLPOGFKNSPTLFQMQLAHILQPIRQAFPOCTILQYMDILLASPSHEDILLSEATMASLISHG LPVSENKTQQTPTGTHKELGQISPNHLYTDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHIDPRDQIYLNPSQVQSLVQLRQAL SONCRSLVQITPLPLGAIMLTLTGTTTVFQSKQWPLVWLHAPLPHITSQCPWGQLASAVLLLDKTYTLQSYGLLCCQTHHNISTQTFNQFIQTSIDHPSV PILLHSHRFKNLGAQTGELWNTHLKTAAPLAPVKALMPVFTLSPVINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSA RSWRCLNIFLDSKYLHYHLRTLALGTFQGRSSQAPEQALLPRLSRKVYLHIVRSHNTLPDISRLNALTADALLITPVLQL
HTL1L_P 0C211	HT L1 L	P0 C2 11	ro ot	GLEHLPRPEISOFPPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVPVKANGTWRFHIDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQITIDLKDA FQIPLPKQFQPYFAFTVPOQCNYGPGTRYAWKVLPOGFKNSPTLFQMQLAHILQPIRQAFPOCTILQYMDILLASPSHEDILLSEATMASLISHGLPVS QDKTQQTPTGTHKELGQISPNHLYTDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHIDPRDQIYLNPSQVQSLVQLRQAL RSRLAQITPLPLGAIMLTLTGTTTVFQSKQWPLVWLHAPLPHITSQCPWGQLASAVLLLDKTYTLQSYGLLCCQTHHNISTQTFNQFIQTSIDHPSV HSHRFKNLGAQTGELWNTHLKTAAPLAPVKALTPVFTLSPVINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWH CLNIFLDSKYLHYHLRTLALGTFQGRSSQAPEQALLPRLAHKVIYLVHVRSHNTLPDISRLNALTADALLITPIL
HTL1L_P 0C211_2 mut	HT L1 L	P0 C2 11	der iva tiv e	GLEHLPRPEISOFPPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVPVKANGTWRFHIDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQITIDLKDA FQIPLPKQFQPYFAFTVPOQCNYGPGTRYAWKVLPOGFKNSPTLFQMQLAHILQPIRQAFPOCTILQYMDILLASPSHEDILLSEATMASLISHGLPVS QDKTQQTPTGTHKELGQISPNHLYTDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHIDPRDQIYLNPSQVQSLVQLRQAL RSRLAQITPLPLGAIMLTLTGTTTVFQSKQWPLVWLHAPLPHITSQCPWGQLASAVLLLDKTYTLQSYGLLCCQTHHNISTQTFNQFIQTSIDHPSV HSHRFKNLGAQTGELWNTHLKTAAPLAPVKALTPVFTLSPVINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWH CLNIFLDSKYLHYHLRTLALGTFQGRSSQAPEQALLPRLAHKVIYLVHVRSHNTLPDISRLNALTADALLITPIL
HTL1L_P 0C211_2 mutB	HT L1 L	P0 C2 11	der iva tiv e	GLEHLPRPEISOFPPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVPVKANGTWRFHIDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQITIDLKDA FQIPLPKQFQPYFAFTVPOQCNYGPGTRYAWKVLPOGFKNSPTLFQMQLAHILQPIRQAFPOCTILQYMDILLASPSHEDILLSEATMASLISHGLPVS QDKTQQTPTGTHKELGQISPNHLYTDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHIDPRDQIYLNPSQVQSLVQLRQAL RSRLAQITPLPLGAIMLTLTGTTTVFQSKQWPLVWLHAPLPHITSQCPWGQLASAVLLLDKTYTLQSYGLLCCQTHHNISTQTFNQFIQTSIDHPSV HSHRFKNLGAQTGELWNTHLKTAAPLAPVKALTPVFTLSPVINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWH CLNIFLDSKYLHYHLRTLALGTFQGRSSQAPEQALLPRLAHKVIYLVHVRSHNTLPDISRLNALTADALLITPIL

[illegible]

HTLV2_P 03363_2m ntB	HT LV 2	P03 363	der iva tiv e	HL PPPQVVDQFPLNLPERLQALNDLVSKALEAGHIEPYSGPNPNVPVKKPNKGKWRHHDJLRATNAITTLTSPSPGPPDLTSPPTALPHLQITDLTDAFF KIPLPKQYQFYAFITPCNCTYAWTVLPQGFKNSTPLFQOOLAAVLPNRMKMTSTIVQYMDILLASPTNEELQQLSQTALQALTTHTGLPISQ EKTQOITPGOIRFLGOVISPNNHITYESTIPPIK SQWTLTELQVILGEIQWVSKGTIPILKKHLQSLXSALHPYRDRPRACTITLTPQQLHALHAIOQALQHNCGR LNPALPLJGLISLTSIGTTSVHPQKNWPLAWLTHPPPTSLCPWGHLLACTITLTKYQLQGHQSFHNNMSKQALCDHLNPHSPVGLJHHMG RFHNLGSPSGPWKTLLHPLTLQEPRLRPITLSPVLDLTPCLFSDGSPQKAAAYLVWDQITLQDITJPLSPHSTHSAQKGELLALICGLRAAKPWPSL NIFLDSKYLIKYLHSLAIGAFGLTSAHOTLQAALPPLQCKTLYLHHVRSHTNLDPSTENEYTDLSILAPLVPL
JSRV_P31 623	JSR V	P31 623	ro ot	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQRLGHIEPSTSAWNSPIFVKKSGKWRLQLDKRKVNTMMHMGALQPLPPT SAIPDKSVIIVIDLKDCFYTIPLAPQDCKRFAESLPVNFKEPMQRYQWRVLPQGMNTSPILCQKFVATAIAPVRQRPQLYLVHYMDILLAHTEDEHLLY QAFSILKQHLNLGLVIADEKIQTHFPNYLGFSLYPRVYNTQLVKLQTDHILKTLNDFQKLLGDNWIRPYLKLPTYTLQPLFDILKGDSDPASPRITLSLE GRTALQSIEEAIHQOQITYCDYQRSWGLYILPTAPRTGYLVYQDKPLRWIYLSATPTKHLPLYELVAKIAGRIHEAIOYFGMEPPFICVPYALEQQDWL FQFSDNWSIAFANYPGQITHYPSDKLLQFASSHAFIEPKVRRQPIPEATLFTDGSSNGTAAALINHQITYYAQITSESSAQVVELFAVHQALLTVPITSENL TDSSYVVGALQMIETVPIIGTTSPEVLNFTLQOVLHCRQHPCHGHIRAHSTLPGALVQGNHTADVLTKQVFFQS
JSRV_P31 623_2mut B	JSR V	P31 623	der iva tiv e	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQRLGHIEPSTSAWNSPIFVKKSGKWRLQLDKRKVNTMMHMGALQPLPPT SPIDKSVIIVIDLKDCFYTIPLAPQDCKRFAESLPVNFKEPMQRYQWRVLPQGMNTSPILCQKFVATAIAPVRQRPQLYLVHYMDILLAHTEDEHLLY QAFSILKQHLNLGLVIADEKIQTHFPNYLGFSLYPRVYNTQLVKLQTDHILKTLNDFQKLLGDNWIRPYLKLPTYTLQPLFDILKGDSDPASPRITLSLE GRTALQSIEEAIHQOQITYCDYQRSWGLYILPTAPRTGYLVYQDKPLRWIYLSATPTKHLPLYELVAKIAGRIHEAIOYFGMEPPFICVPYALEQQDWL FQFSDNWSIAFANYPGQITHYPSDKLLQFASSHAFIEPKVRRQPIPEATLFTDGSSNGTAAALINHQITYYAQITSESSAQVVELFAVHQALLTVPITSENL TDSSYVVGALQMIETVPIIGTTSPEVLNFTLQOVLHCRQHPCHGHIRAHSTLPGALVQGNHTADVLTKQVFFQS
KORV_Q 9TTC1	KO RV	Q9 TT C1	ro ot	TLGDQSGRSDPLPEPRVYTLTVEGIPTIEFLVNTIGAEHSVLTKPMGKMGSKRTVVAGATGSKVYVPWTTKRLKIGQKQVTHSEFLVIEPCAPILLGRDILLT KLKAQIQFSTEGPQVTVWEDRPAMCLVLNLEEEYRLHEKVPSPIDPSWLQHPMVAAEKAGMGLANQVPPVVVELKSDASPAVRQYPMISKEAREGIR PHIQRFDLGILVPCQSPWNITPLPVKKPGTNDYRPVQDLREVNRKRVQDIEHTVPNPNYLLSSPPSHTWYSLVDLKDFAFCLKLPNSQPLFAFEWRDP EKGNTGQLTWTRLPQGFKNSTPLFDEALHRLDASFRALNPQVVMQYVDDLLVAAPTYRDCKEGTRRLQELSKLYRVSAKKAQLCREEVYLYG LKGGKRWLTARKATVMKIPTTTPRQVREHFGTAGFCRLWIPGFASLAAPLYPLTREKVPVPTWTEAHQEAFGRIKEALLSAPALAPDLTK KEGARGVLQTILGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLIKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLNERY SEAPPAILNPATLTPVESDDTPHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFMDGRRQAGAAIVDNKRITVWASNLPEGTSQAKAELIALQALR LAEGKSINIYTDSTRYAFATAHVHGAIVKQRGWLTSAKDKNKEEILALLEAHLPKRVAIHICPGHQGTDPVATGNRKADAEEAAKQAAQSTR LIGRDLLTKLKAQIQFSTEGPQVTVWEDRPAMCLVLNLEEEYRLHEKVPSPIDPSWLQHPMVAAEKAGMGLANQVPPVVVELKSDASPAVRQYPM
KORV_Q 9TTC1- Pro	KO RV	Q9 TT C1	der iva tiv e	SKEAREGIRPHIQRFDLGILVPCQSPWNITPLPVKKPGTNDYRPVQDLREVNRKRVQDIEHTVPNPNYLLSSPPSHTWYSLVDLKDFAFCLKLPNSQPL FAFEWRDPEKNGTQTLTWTRLPQGFKNSTPLFNEALHRLDASFRALNPQVVMQYVDDLLVAAPTYRDCKEGTRRLQELSKLYRVSAKKAQLCRE EVTYLYGKGGKRWLTARKATVMKIPTTTPRQVREHFGTAGFCRLWIPGFASLAAPLYPLTREKVPVPTWTEAHQEAFGRIKEALLSAPALAPDLTK PFALVYDEKEGVARGVLQTILGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLIKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQ SLLNERYSEAPPAILNPATLTPVESDDTPHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFMDGRRQAGAAIVDNKRITVWASNLPEGTSQAKAELI ALTOALRLAEGKSINIYTDSTRYAFATAHVHGAIVKQRGWLTSAKDKNKEEILALLEAHLPKRVAIHICPGHQGTDPVATGNRKADAEEAAKQAAQSTR LITETTKN
KORV_Q 9TTC1- Pro_3mut A	KO RV	Q9 TT C1	der iva tiv e	LIGRDLLTKLKAQIQFSTEGPQVTVWEDRPAMCLVLNLEEEYRLHEKVPSPIDPSWLQHPMVAAEKAGMGLANQVPPVVVELKSDASPAVRQYPM SKEAREGIRPHIQRFDLGILVPCQSPWNITPLPVKKPGTNDYRPVQDLREVNRKRVQDIEHTVPNPNYLLSSPPSHTWYSLVDLKDFAFCLKLPNSQPL FAFEWRDPEKNGTQTLTWTRLPQGFKNSTPLFNEALHRLDASFRALNPQVVMQYVDDLLVAAPTYRDCKEGTRRLQELSKLYRVSAKKAQLCRE EVTYLYGKGGKRWLTARKATVMKIPTTTPRQVREHFGTAGFCRLWIPGFASLAAPLYPLTREKVPVPTWTEAHQEAFGRIKEALLSAPALAPDLTK PFALVYDEKEGVARGVLQTILGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLIKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQ SLLNERYSEAPPAILNPATLTPVESDDTPHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFMDGRRQAGAAIVDNKRITVWASNLPEGTSQAKAELI ALTOALRLAEGKSINIYTDSTRYAFATAHVHGAIVKQRGWLTSAKDKNKEEILALLEAHLPKRVAIHICPGHQGTDPVATGNRKADAEEAAKQAAQSTR LITETTKN

KORV_Q 9TTC1_3 mut	KO RV	Q9 TT CI	der iva tiv e	TLGDQSGRSDPLPEPRVLTVEGIPTTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSHFVIEPCAPLLGRDLLT KLKAQIQSFTGEPQVTVWEDRPAMCLVLNLEEEYRLHEKVPSPIDPSWLQJLPMVAAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMKSKEAREGIR PHIQRFDLGILVPCQSPWNITPLL PVKKPGTNDYRPVQDLREVNRKQDIHPTVNPYNLLSSLPSTHTWYSVLDLKDFAFCLKLHPNSQPLFAFEWRDP EKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMQYVDDLLVAAPTYRDCKEGTRRLQELSKLGYRVSARKAQLCREEVTVLGYL LKGGKRWLTARKATVMKIPITPTTQVREFLGTAGFCRLWPGFASLAPLYLTPKVPFTWTEAHQEAFCRIKELLSAPALAPDLTKPFALYVDE KEGVARGVLTQILGPWRPVAIYLSKKLDPVAGSWPTCLKAIAAVALLLKADKLLTGQNVLVIAPIINLESVROPDRWMTNARMTHYQSLLLNERV SFAPPAILNPATLLPVEDDTPHICSEILAEETGRPDLRDPQVPAWYTDGSSFMIDGRROAGAIVDNKRITVWASNLPEGTSAQKAEILALTOALR LAEGKSINIYTDSDRYAFATAHVHGAIYKQRGWLTSAAGDKIKNKEILALLEAHLPKRVAIHCHPGHQGTDPVATGNRKADEAAKQAAQSTRILTETTK N
				TLGDQSGRSDPLPEPRVLTVEGIPTTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSHFVIEPCAPLLGRDLLT KLKAQIQSFTGEPQVTVWEDRPAMCLVLNLEEEYRLHEKVPSPIDPSWLQJLPMVAAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMKSKEAREGIR PHIQRFDLGILVPCQSPWNITPLL PVKKPGTNDYRPVQDLREVNRKQDIHPTVNPYNLLSSLPSTHTWYSVLDLKDFAFCLKLHPNSQPLFAFEWRDP EKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMQYVDDLLVAAPTYRDCKEGTRRLQELSKLGYRVSARKAQLCREEVTVLGYL LKGGKRWLTARKATVMKIPITPTTQVREFLGTAGFCRLWPGFASLAPLYLTPKVPFTWTEAHQEAFCRIKELLSAPALAPDLTKPFALYVDE KEGVARGVLTQILGPWRPVAIYLSKKLDPVAGSWPTCLKAIAAVALLLKADKLLTGQNVLVIAPIINLESVROPDRWMTNARMTHYQSLLLNERV SFAPPAILNPATLLPVEDDTPHICSEILAEETGRPDLRDPQVPAWYTDGSSFMIDGRROAGAIVDNKRITVWASNLPEGTSAQKAEILALTOALR LAEGKSINIYTDSDRYAFATAHVHGAIYKQRGWLTSAAGDKIKNKEILALLEAHLPKRVAIHCHPGHQGTDPVATGNRKADEAAKQAAQSTRILTETTK N
KORV_Q 9TTC1_3 mutA	KO RV	Q9 TT CI	der iva tiv e	TLNLEDEYRLYEISAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL
				TLNLEDEYRLYEISAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL
MLVAV_ P03356_ mut	ML VAV_ V	P03 356	der iva tiv e	TLNLEDEYRLYEISAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL
				TLNLEDEYRLYEISAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL
MLVAV_ P03356_3 mutA	ML VAV_ V	P03 356	der iva tiv e	TLNLEDEYRLYEISAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL
				TLNLEDEYRLYEISAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL
MLVBM_ Q7SVK7_ 3mutA_W S	ML VBM_ V	Q7 SV K7	ro ot	TLGIEDEYRLHEISTEPDVSIGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL
				TLGIEDEYRLHEISTEPDVSIGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL
MLVBM_ Q7SVK7_ 3mutA_W S	ML VBM_ V	Q7 SV K7	ro ot	TLGIEDEYRLHEISTEPDVSIGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL
				TLGIEDEYRLHEISTEPDVSIGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSQYPMQSEAKLGIKPHIQRLLDQGILVPCQSPWNITPLL PV KKPGTNDYRPVQDLREVNRKVEDIHTVNPYNLLSGLPSSHRYTVLVDLKDFAFCLRLHPTSQPLFAFEWRDPGMSGQQLTWTRLPQGFKNSPTLFD EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGTTRALLTLGNLGYRASAKKAQKQKQKYLGYLLKEGQWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWPGFAEMAAPLYPLTKTGTLENWGPQQAQYQEIQAALLTAPALGLPDLTKPELFDVDEKQYAKGVLTQKLGPRWRPVAIYLS KKLDPVAAAGWPPCLRMVAALVLRKDAAGLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDITDRVQFGPVVALNPATLLPPEEGAP HDCLEILAEETHGTRPDLTDQIPDADHTWYTDGSSFLQEGQKAGAAVTETEVIWALPAGTSAQRAELIALTOALKMAEGKRLNVYTDSDRYAFAT AHHIGEYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIHCLGHQKGDSEARGNRLADQAAAREAAIKTPPDSTILL

KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
HDCLEIL	AETHGTRPDL	TQDQIP	ADHWTWY	TDGSSEFL	QEGQKAGAA	VTTET	VEI	WAGAL	PAGTSA	QRAELIAL	TQAL	KMAE	GKRL	NV	YTD	SR
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLGIEDEY	RLHET	STEPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL	GIKPHIQRL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
HDCLEIL	AETHGTRPDL	TQDQIP	ADHWTWY	TDGSSEFL	QEGQKAGAA	VTTET	VEI	WAGAL	PAGTSA	QRAELIAL	TQAL	KMAE	GKRL	NV	YTD	SR
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
LGIEDEY	RLHET	STEPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL	GIKPHIQRL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
HDCLEIL	AETHGTRPDL	TQDQIP	ADHWTWY	TDGSSEFL	QEGQKAGAA	VTTET	VEI	WAGAL	PAGTSA	QRAELIAL	TQAL	KMAE	GKRL	NV	YTD	SR
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLGIEDEY	RLHET	STEPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL	GIKPHIQRL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
HDCLEIL	AETHGTRPDL	TQDQIP	ADHWTWY	TDGSSEFL	QEGQKAGAA	VTTET	VEI	WAGAL	PAGTSA	QRAELIAL	TQAL	KMAE	GKRL	NV	YTD	SR
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVRQAP	LIPL	KAT	STP	SIQ	YPM	SHAE	ARL
KKPGTND	YRPVQD	RL	REVNKR	VEDI	HTVP	NPYN	LL	SGL	PPSHQWY	TV	DL	KDA	FCL	RL	HTPS	QSQL
EAL	HRDL	AGFR	IQH	PD	LILL	Q	YVDDILL	AAT	SELDC	COQ	TR	AL	Q	ITL	GL	GYR
QL	REFL	G	AG	FC	R	L	IPG	FAE	MAA	P	Y	L	T	K	P	GT
KKLDPV	AAGWPPCL	RMV	AAIA	VLTKDAGKL	TMGQPL	VIL	APHAVEAL	VKQPPDRWL	SNARMTHY	QAMILL	DTDR	VQFGPVV	AL	NPATILL	PL	PEEGAP
DCLDL	AE	AH	GT	R	SD	M	Q	P	L	D	H	W	T	W	T	D
AIHUEI	YRRRGWLT	SEGREIK	NKSEIL	ALL	KAL	FLPKRL	SIHICL	GHQK	GDSAE	ARGNRL	ADQAA	AREAA	AIK	TPD	ST	STLL
TLNIEDEY	RLHET	SK	EPDV	SL	GSTW	SDFPQAW	AET	GGMGL	AVR							

					CLDLAEAHGTRSDLMQDPLPADHTWYTDGSSFLQEGQRKAGAAVTTETEVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVTDSRYAFATA HHHGEYRRRGWLTSEGEIKNKDEILALLKALFLPKRLSIHCPGHQKGNRAEARGNRMADQAAAREVATRETPETSTILL
MLVF5_P 26810	VF 5	P26 810	ro ot		TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIPLKATSTPVSQKQYPMQSEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNRVEDIHTVPNPYNLLSGLPSSHQWYTVLIDLKDAFFCLRLHPTSSQSLFAFEWKPDMGSGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGRTRALLQTLGDLGYRASAKKQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFWGPDQQAQYQKQVYGLLKEGQWLTARKETVMGQPTKTPR KKLDPAAGWPPCLRMVAALVTKDAGKLMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTRVQFGPIVALNPATLLPLPEGLQH DCLDLAEAHGTRPDLTDQPLPADHTWYTDGSSFLQEGQRKAGAAVTTETEVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVTDSRYAFATA HHHGEYRRRGWLTSEGEIKNKDEILALLKALFLPKRLSIHCPGHQKGNRAEARGNRMADQAAAREVATRETPETSTILL
MLVF5_P 26810_3m ut	VF 5	P26 810	der iva tiv e		TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIPLKATSTPVSQKQYPMQSEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNRVEDIHTVPNPYNLLSGLPSSHQWYTVLIDLKDAFFCLRLHPTSSQSLFAFEWKPDMGSGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGRTRALLQTLGDLGYRASAKKQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFWGPDQQAQYQKQVYGLLKEGQWLTARKETVMGQPTKTPR KKLDPAAGWPPCLRMVAALVTKDAGKLMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTRVQFGPIVALNPATLLPLPEGLQH DCLDLAEAHGTRPDLTDQPLPADHTWYTDGSSFLQEGQRKAGAAVTTETEVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVTDSRYAFATA HHHGEYRRRGWLTSEGEIKNKDEILALLKALFLPKRLSIHCPGHQKGNRAEARGNRMADQAAAREVATRETPETSTILL
MLVF5_P 26810_3m utA	VF 5	P26 810	der iva tiv e		TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIPLKATSTPVSQKQYPMQSEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNRVEDIHTVPNPYNLLSGLPSSHQWYTVLIDLKDAFFCLRLHPTSSQSLFAFEWKPDMGSGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGRTRALLQTLGDLGYRASAKKQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFWGPDQQAQYQKQVYGLLKEGQWLTARKETVMGQPTKTPR KKLDPAAGWPPCLRMVAALVTKDAGKLMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTRVQFGPIVALNPATLLPLPEGLQH DCLDLAEAHGTRPDLTDQPLPADHTWYTDGSSFLQEGQRKAGAAVTTETEVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVTDSRYAFATA HHHGEYRRRGWLTSEGEIKNKDEILALLKALFLPKRLSIHCPGHQKGNRAEARGNRMADQAAAREVATRETPETSTILL
MLVFF_P 26809	VF F	P26 809	ro ot		TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIPLKATSTPVSQKQYPMQSEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNRVEDIHTVPNPYNLLSGLPSSHQWYTVLIDLKDAFFCLRLHPTSSQSLFAFEWKPDMGSGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGRTRALLQTLGDLGYRASAKKQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFWGPDQQAQYQKQVYGLLKEGQWLTARKETVMGQPTKTPR KKLDPAAGWPPCLRMVAALVTKDAGKLMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTRVQFGPIVALNPATLLPLPEGLQH DCLDLAEAHGTRPDLTDQPLPADHTWYTDGSSFLQEGQRKAGAAVTTETEVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVTDSRYAFATA HHHGEYRRRGWLTSEGEIKNKDEILALLKALFLPKRLSIHCPGHQKGNRAEARGNRMADQAAAREVATRETPETSTILL
MLVFF_P 26809_3m ut	VF F	P26 809	der iva tiv e		TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIPLKATSTPVSQKQYPMQSEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNRVEDIHTVPNPYNLLSGLPSSHQWYTVLIDLKDAFFCLRLHPTSSQSLFAFEWKPDMGSGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGRTRALLQTLGDLGYRASAKKQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFWGPDQQAQYQKQVYGLLKEGQWLTARKETVMGQPTKTPR KKLDPAAGWPPCLRMVAALVTKDAGKLMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTRVQFGPIVALNPATLLPLPEGLQH DCLDLAEAHGTRPDLTDQPLPADHTWYTDGSSFLQEGQRKAGAAVTTETEVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVTDSRYAFATA HHHGEYRRRGWLTSEGEIKNKDEILALLKALFLPKRLSIHCPGHQKGNRAEARGNRMADQAAAREVATRETPETSTILL
MLVFF_P 26809_3m utA	VF F	P26 809	der iva tiv e		TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIPLKATSTPVSQKQYPMQSEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNRVEDIHTVPNPYNLLSGLPSSHQWYTVLIDLKDAFFCLRLHPTSSQSLFAFEWKPDMGSGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGRTRALLQTLGDLGYRASAKKQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFWGPDQQAQYQKQVYGLLKEGQWLTARKETVMGQPTKTPR KKLDPAAGWPPCLRMVAALVTKDAGKLMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTRVQFGPIVALNPATLLPLPEGLQH DCLDLAEAHGTRPDLTDQPLPADHTWYTDGSSFLQEGQRKAGAAVTTETEVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVTDSRYAFATA HHHGEYRRRGWLTSEGEIKNKDEILALLKALFLPKRLSIHCPGHQKGNRAEARGNRMADQAAAREVATRETPETSTILL
MLVMS_ P P03355_P LV919	VF S	P03 355	ro ot		TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIPLKATSTPVSQKQYPMQSEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNRVEDIHTVPNPYNLLSGLPSSHQWYTVLIDLKDAFFCLRLHPTSSQSLFAFEWKPDMGSGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLAATSELDCCQGRTRALLQTLGDLGYRASAKKQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFWGPDQQAQYQKQVYGLLKEGQWLTARKETVMGQPTKTPR KKLDPAAGWPPCLRMVAALVTKDAGKLMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTRVQFGPIVALNPATLLPLPEGLQH DCLDLAEAHGTRPDLTDQPLPADHTWYTDGSSFLQEGQRKAGAAVTTETEVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVTDSRYAFATA HHHGEYRRRGWLTSEGEIKNKDEILALLKALFLPKRLSIHCPGHQKGNRAEARGNRMADQAAAREVATRETPETSTILL

					NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLVVYTDSTRYAFAT AHHGHEYRRRGWLTSECKEIKNKDEILALLKALFLPKRLSIHCPGHQKHSAEARGNRMADQAAARKAAITE:TPDTSITLL
				der iva tiv e	TLNIEDEHRLHETSKPEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYPMQSQEARLGIKPHIQRLLDQGLVPCQSPWNITLLPV KKPGTNDYRPVQDLREVNKRVEDIHTVPNPYNLLSGLPSSHQWYTVLDLKDADFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLLAATSELDCCQGTALITLQNLGYRASAKKAQICQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTILFNWGPQQAQYQEIQAALLTAPALGLPDLTKPFELFVDEKQYAGKVLTKLGPWRRPVAAYLS KKLDPAAGWPPCLRMVAALVLTDAAGKLTMGQPLVILAPHAVALVKQPPDRWLSNARMTHYQALLDTRVQFGPVALNPATLLPLPEGLQH NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLVVYTDSTRYAFAT AHHGHEYRRRGWLTSECKEIKNKDEILALLKALFLPKRLSIHCPGHQKHSAEARGNRMADQAAARKAAITE:TPDTSITLL
		ML VMS_	VM P03 355		TLNIEDEHRLHETSKPEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYPMQSQEARLGIKPHIQRLLDQGLVPCQSPWNITLLPV KKPGTNDYRPVQDLREVNKRVEDIHTVPNPYNLLSGLPSSHQWYTVLDLKDADFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLLAATSELDCCQGTALITLQNLGYRASAKKAQICQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTILFNWGPQQAQYQEIQAALLTAPALGLPDLTKPFELFVDEKQYAGKVLTKLGPWRRPVAAYLS KKLDPAAGWPPCLRMVAALVLTDAAGKLTMGQPLVILAPHAVALVKQPPDRWLSNARMTHYQALLDTRVQFGPVALNPATLLPLPEGLQH NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLVVYTDSTRYAFAT AHHGHEYRRRGWLTSECKEIKNKDEILALLKALFLPKRLSIHCPGHQKHSAEARGNRMADQAAARKAAITE:TPDTSITLL
		ML VMS_	VM P03 355	der iva tiv e	TLNIEDEHRLHETSKPEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYPMQSQEARLGIKPHIQRLLDQGLVPCQSPWNITLLPV KKPGTNDYRPVQDLREVNKRVEDIHTVPNPYNLLSGLPSSHQWYTVLDLKDADFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLLAATSELDCCQGTALITLQNLGYRASAKKAQICQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTILFNWGPQQAQYQEIQAALLTAPALGLPDLTKPFELFVDEKQYAGKVLTKLGPWRRPVAAYLS KKLDPAAGWPPCLRMVAALVLTDAAGKLTMGQPLVILAPHAVALVKQPPDRWLSNARMTHYQALLDTRVQFGPVALNPATLLPLPEGLQH NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLVVYTDSTRYAFAT AHHGHEYRRRGWLTSECKEIKNKDEILALLKALFLPKRLSIHCPGHQKHSAEARGNRMADQAAARKAAITE:TPDTSITLL
		ML VMS_	VM P03 355	der iva tiv e	TLNIEDEHRLHETSKPEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYPMQSQEARLGIKPHIQRLLDQGLVPCQSPWNITLLPV KKPGTNDYRPVQDLREVNKRVEDIHTVPNPYNLLSGLPSSHQWYTVLDLKDADFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLLAATSELDCCQGTALITLQNLGYRASAKKAQICQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTILFNWGPQQAQYQEIQAALLTAPALGLPDLTKPFELFVDEKQYAGKVLTKLGPWRRPVAAYLS KKLDPAAGWPPCLRMVAALVLTDAAGKLTMGQPLVILAPHAVALVKQPPDRWLSNARMTHYQALLDTRVQFGPVALNPATLLPLPEGLQH NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLVVYTDSTRYAFAT AHHGHEYRRRGWLTSECKEIKNKDEILALLKALFLPKRLSIHCPGHQKHSAEARGNRMADQAAARKAAITE:TPDTSITLL
		ML VMS_	VM P03 355	ro ot	TLNIEDEHRLHETSKPEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYPMQSQEARLGIKPHIQRLLDQGLVPCQSPWNITLLPV KKPGTNDYRPVQDLREVNKRVEDIHTVPNPYNLLSGLPSSHQWYTVLDLKDADFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLLAATSELDCCQGTALITLQNLGYRASAKKAQICQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTILFNWGPQQAQYQEIQAALLTAPALGLPDLTKPFELFVDEKQYAGKVLTKLGPWRRPVAAYLS KKLDPAAGWPPCLRMVAALVLTDAAGKLTMGQPLVILAPHAVALVKQPPDRWLSNARMTHYQALLDTRVQFGPVALNPATLLPLPEGLQH NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLVVYTDSTRYAFAT AHHGHEYRRRGWLTSECKEIKNKDEILALLKALFLPKRLSIHCPGHQKHSAEARGNRMADQAAARKAAITE:TPDTSITLL
		ML VMS_	VM P03 355	der iva tiv e	TLNIEDEHRLHETSKPEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYPMQSQEARLGIKPHIQRLLDQGLVPCQSPWNITLLPV KKPGTNDYRPVQDLREVNKRVEDIHTVPNPYNLLSGLPSSHQWYTVLDLKDADFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLLAATSELDCCQGTALITLQNLGYRASAKKAQICQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTILFNWGPQQAQYQEIQAALLTAPALGLPDLTKPFELFVDEKQYAGKVLTKLGPWRRPVAAYLS KKLDPAAGWPPCLRMVAALVLTDAAGKLTMGQPLVILAPHAVALVKQPPDRWLSNARMTHYQALLDTRVQFGPVALNPATLLPLPEGLQH NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLVVYTDSTRYAFAT AHHGHEYRRRGWLTSECKEIKNKDEILALLKALFLPKRLSIHCPGHQKHSAEARGNRMADQAAARKAAITE:TPDTSITLL
		ML VMS_	VM P03 355	der iva tiv e	TLNIEDEHRLHETSKPEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYPMQSQEARLGIKPHIQRLLDQGLVPCQSPWNITLLPV KKPGTNDYRPVQDLREVNKRVEDIHTVPNPYNLLSGLPSSHQWYTVLDLKDADFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLLAATSELDCCQGTALITLQNLGYRASAKKAQICQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTILFNWGPQQAQYQEIQAALLTAPALGLPDLTKPFELFVDEKQYAGKVLTKLGPWRRPVAAYLS KKLDPAAGWPPCLRMVAALVLTDAAGKLTMGQPLVILAPHAVALVKQPPDRWLSNARMTHYQALLDTRVQFGPVALNPATLLPLPEGLQH NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLVVYTDSTRYAFAT AHHGHEYRRRGWLTSECKEIKNKDEILALLKALFLPKRLSIHCPGHQKHSAEARGNRMADQAAARKAAITE:TPDTSITLL
		ML VMS_	VM P03 355	der iva tiv e	TLNIEDEHRLHETSKPEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYPMQSQEARLGIKPHIQRLLDQGLVPCQSPWNITLLPV KKPGTNDYRPVQDLREVNKRVEDIHTVPNPYNLLSGLPSSHQWYTVLDLKDADFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSTPLFN EALHRDLADFRIOHPDLILLQYVDDLLLAATSELDCCQGTALITLQNLGYRASAKKAQICQKQVYGLLKEGQWLTARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTILFNWGPQQAQYQEIQAALLTAPALGLPDLTKPFELFVDEKQYAGKVLTKLGPWRRPVAAYLS KKLDPAAGWPPCLRMVAALVLTDAAGKLTMGQPLVILAPHAVALVKQPPDRWLSNARMTHYQALLDTRVQFGPVALNPATLLPLPEGLQH NCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLVVYTDSTRYAFAT AHHGHEYRRRGWLTSECKEIKNKDEILALLKALFLPKRLSIHCPGHQKHSAEARGNRMADQAAARKAAITE:TPDTSITLL

						NCLDILAEAHGTRPDLTDQPLPDAHDHTWYTDGSSLLQEGQORKAGAAVTTETEVWAKALPAGTSAQRAEALLATQALKMAEGKKLVNVTDSRYAFAT AHHGEIYRRGGLTSEGREIKNKSEILALLKALFLPKRLSIHCHQKGDSAEARGNRMDQAARAAITETPDISTILL
						TLNIEDEYRLHEISTEPDVSFGSTWLSDDHPQAAWAEATGGMGLAVRQAPLIPLKATISTPVSIOKQPMISQAEAKGIKPHIQRLLDQGIILVPCQSPWNITPLPVK KPGTNDYRYPQGLREVNRKRVEDIHPTVPNPYNILSGLPTSHRWYTVLDLKDFAHFCRLHPTSQPIFASERWDPDGMGISQGLTWTRLPOGKNSPTLFDIE ALHRGLADRIQHPDILLQYVDDILLAAATSELDCOQGTTRALLKTILNIGYRASAKKAQICQKVXYLLGYLLREGQRLWTEARKETVMGQOPTKTPRQ LREELGTAGFCRLWIPRAEAMAPLYPLTKGTLENWGPDOQKAYHEIKQALLTAPALGIDPLTKPEELVFDEKQYAGKVLTKLGPWRPVPAYLSK KLDPAAGWPCLRMVAAIALVTKDAGKLTMGQPLVILAPHAVEALVQKPPDRWLSNARMTHYQAMLLDTRVQFGPVALNPATLLPLPEEGAPH DCLEILAEATHGTEPDLTDQPIPDADHTWYTDGSSLLQEGQORKAGAAVTTETEVWAKALPAGTSAQRAEALLATQALKMAEGKKRLNVYTDSDRYAFATA HHIGEYKRKGWLTSSEGREIKNKSEILALLKALFLPKRLSIHCHQKGDSAEARGNRMDQAARAAITETPDISTILL
ML_VRD_P11227_D	P11_VR_D	227	ro	ot		TLNIEDEYRLHEISTEPDVSFGSTWLSDDHPQAAWAEATGGMGLAVRQAPLIPLKATISTPVSIOKQPMISQAEAKGIKPHIQRLLDQGIILVPCQSPWNITPLPVK KPGTNDYRYPQGLREVNRKRVEDIHPTVPNPYNILSGLPTSHRWYTVLDLKDFAHFCRLHPTSQPIFASERWDPDGMGISQGLTWTRLPOGKNSPTLFDIE ALHRGLADRIQHPDILLQYVDDILLAAATSELDCOQGTTRALLKTILNIGYRASAKKAQICQKVXYLLGYLLREGQRLWTEARKETVMGQOPTKTPRQ LREELGTAGFCRLWIPRAEAMAPLYPLTKGTLENWGPDOQKAYHEIKQALLTAPALGIDPLTKPEELVFDEKQYAGKVLTKLGPWRPVPAYLSK KLDPAAGWPCLRMVAAIALVTKDAGKLTMGQPLVILAPHAVEALVQKPPDRWLSNARMTHYQAMLLDTRVQFGPVALNPATLLPLPEEGAPH DCLEILAEATHGTEPDLTDQPIPDADHTWYTDGSSLLQEGQORKAGAAVTTETEVWAKALPAGTSAQRAEALLATQALKMAEGKKRLNVYTDSDRYAFATA HHIGEYKRKGWLTSSEGREIKNKSEILALLKALFLPKRLSIHCHQKGDSAEARGNRMDQAARAAITETPDISTILL
ML_VRD_P11227_3_mut	ML_VRD_P11227_3_VR	P11_VR_D	227	der iva tiv	e	TLNIEDEYRLHEISTEPDVSFGSTWLSDDHPQAAWAEATGGMGLAVRQAPLIPLKATISTPVSIOKQPMISQAEAKGIKPHIQRLLDQGIILVPCQSPWNITPLPVK KPGTNDYRYPQGLREVNRKRVEDIHPTVPNPYNILSGLPTSHRWYTVLDLKDFAHFCRLHPTSQPIFASERWDPDGMGISQGLTWTRLPOGKNSPTLFDIE ALHRGLADRIQHPDILLQYVDDILLAAATSELDCOQGTTRALLKTILNIGYRASAKKAQICQKVXYLLGYLLREGQRLWTEARKETVMGQOPTKTPRQ LREELGTAGFCRLWIPRAEAMAPLYPLTKGTLENWGPDOQKAYHEIKQALLTAPALGIDPLTKPEELVFDEKQYAGKVLTKLGPWRPVPAYLSK KLDPAAGWPCLRMVAAIALVTKDAGKLTMGQPLVILAPHAVEALVQKPPDRWLSNARMTHYQAMLLDTRVQFGPVALNPATLLPLPEEGAPH DCLEILAEATHGTEPDLTDQPIPDADHTWYTDGSSLLQEGQORKAGAAVTTETEVWAKALPAGTSAQRAEALLATQALKMAEGKKRLNVYTDSDRYAFATA HHIGEYKRKGWLTSSEGREIKNKSEILALLKALFLPKRLSIHCHQKGDSAEARGNRMDQAARAAITETPDISTILL
ML_VRD_P11227_3_mutA	ML_VRD_P11227_3_VR	P11_VR_D	227	der iva tiv	e	TLNIEDEYRLHEISTEPDVSFGSTWLSDDHPQAAWAEATGGMGLAVRQAPLIPLKATISTPVSIOKQPMISQAEAKGIKPHIQRLLDQGIILVPCQSPWNITPLPVK KPGTNDYRYPQGLREVNRKRVEDIHPTVPNPYNILSGLPTSHRWYTVLDLKDFAHFCRLHPTSQPIFASERWDPDGMGISQGLTWTRLPOGKNSPTLFDIE ALHRGLADRIQHPDILLQYVDDILLAAATSELDCOQGTTRALLKTILNIGYRASAKKAQICQKVXYLLGYLLREGQRLWTEARKETVMGQOPTKTPRQ LREELGTAGFCRLWIPRAEAMAPLYPLTKGTLENWGPDOQKAYHEIKQALLTAPALGIDPLTKPEELVFDEKQYAGKVLTKLGPWRPVPAYLSK KLDPAAGWPCLRMVAAIALVTKDAGKLTMGQPLVILAPHAVEALVQKPPDRWLSNARMTHYQAMLLDTRVQFGPVALNPATLLPLPEEGAPH DCLEILAEATHGTEPDLTDQPIPDADHTWYTDGSSLLQEGQORKAGAAVTTETEVWAKALPAGTSAQRAEALLATQALKMAEGKKRLNVYTDSDRYAFATA HHIGEYKRKGWLTSSEGREIKNKSEILALLKALFLPKRLSIHCHQKGDSAEARGNRMDQAARAAITETPDISTILL
MMTVB_P03365_WS	MMTVB_P03365_VB	P03_VB	365	ro	ot	WVQESIDSRPMHLHYLNGRRFLGLNTGADKTCIAGRDWPAWNPWHQITESSLOGLGMACVARSQPLRWQHEDKSGHHIPFVITPLPTLWGRDIMKDKI VRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKIQAALQOLVTEQOLGHLFEESNPWNITPVFVKKKSGKWRLLQDLRAVNAT MHDMGALQGLPSPVAVPKGWEIIMDLQDCHFNKILHPEDCKRFAFSVPNSFKRPYQRFQWKVLPOGMKNSPTLCOKFVDAKILTVRDKYQDSYIVHY MDDILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQIDYDNLYKYLGTTHQIGDSVSYQKLQURTDKLRTLNDFOKLLGNINWIRPELKLTTGELKPLFEILN GDSNPISTRKLTPAEACKALQLMNERLSTARVKRDLQLSQPSWSCILKTEYTPACLWQDQGVVEWHLPHISPKVITPPTYDICTQILJKGRHSKELFSKDPDY IVVPYTKVQFDILLQEKEDWPISLGHLEGVHFHLPKDPDLTFTLQTLAIHPMTSTITPLEKGIVFTDGSANGRSVITYIQGREPIKENTQNTAQQAEIVAVI TAFEESVQSPENLYTDSKSYVTGLPPEIETATLSPRTKYITELKHLQRLJHKKRQEKFYIGHIRGHTGLPGLAQGNAYADSLTRILTA
MMTVB_P03365_Pro	MMTVB_P03365_VB	P03_VB	365	der iva tiv	e	WVQESIDSRPMHLHYLNGRRFLGLNTGADKTCIAGRDWPAWNPWHQITESSLOGLGMACVARSQPLRWQHEDKSGHHIPFVITPLPTLWGRDIMKDKI KVRMLTSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKIQAALQOLVTEQOLGHLFEESNPWNITPVFVKKKSGKWRLLQDLRAVNAT TMHDMGALQGLPSPVAVPKGWEIIMDLQDCHFNKILHPEDCKRFAFSVPNSFKRPYQRFQWKVLPOGMKNSPTLCOKFVDAKILTVRDKYQDSYIVHY YMDMILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQIDYDNLYKYLGTTHQIGDSVSYQKLQURTDKLRTLNDFOKLLGNINWIRPELKLTTGELKPLFEIL NGDSNPISTRKLTPAEACKALQLMNERLSTARVKRDLQLSQPSWSCILKTEYTPACLWQDQGVVEWHLPHISPKVITPPTYDICTQILJKGRHSKELFSKDPDY IVVPYTKVQFDILLQEKEDWPISLGHLEGVHFHLPKDPDLTFTLQTLAIHPMTSTITPLEKGIVFTDGSANGRSVITYIQGREPIKENTQNTAQQAEIVAVI TAFEESVQSPENLYTDSKSYVTGLPPEIETATLSPRTKYITELKHLQRLJHKKRQEKFYIGHIRGHTGLPGLAQGNAYADSLTRILTA
MMTVB_P03365-Pro	MMTVB_P03365_VB	P03_VB	365	der iva tiv	e	GRDIMKDKI VRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKIQAALQOLVTEQOLGHLFEESNPWNITPVFVKKKSGKWRLLQ DLRAVNATMHDMGALQGLPSPVAVPKGWEIIMDLQDCHFNKILHPEDCKRFAFSVPNSFKRPYQRFQWKVLPOGMKNSPTLCOKFVDAKILTVRDKY QDSYIVHYMDMILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQIDYDNLYKYLGTTHQIGDSVSYQKLQURTDKLRTLNDFOKLLGNINWIRPELKLTTG ELKPLFEILNGDSNPISTRKLTPAEACKALQLMNERLSTARVKRDLQLSQPSWSCILKTEYTPACLWQDQGVVEWHLPHISPKVITPPTYDICTQILJKGRHS KELFSKDPDYIVVPYTKVQFDILLQEKEDWPISLGHLEGVHFHLPKDPDLTFTLQTLAIHPMTSTITPLEKGIVFTDGSANGRSVITYIQGREPIKENTQNT AQQAEIVAVITAFEESVQSPENLYTDSKSYVTGLPPEIETATLSPRTKYITELKHLQRLJHKKRQEKFYIGHIRGHTGLPGLAQGNAYADSLTRILTA
MMTVB_P03365-Pro	MMTVB_P03365_VB	P03_VB	365	der iva tiv	e	GRDIMKDKI VRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKIQAALQOLVTEQOLGHLFEESNPWNITPVFVKKKSGKWRLLQ DLRAVNATMHDMGALQGLPSPVAVPKGWEIIMDLQDCHFNKILHPEDCKRFAFSVPNSFKRPYQRFQWKVLPOGMKNSPTLCOKFVDAKILTVRDKY QDSYIVHYMDMILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQIDYDNLYKYLGTTHQIGDSVSYQKLQURTDKLRTLNDFOKLLGNINWIRPELKLTTG ELKPLFEILNGDSNPISTRKLTPAEACKALQLMNERLSTARVKRDLQLSQPSWSCILKTEYTPACLWQDQGVVEWHLPHISPKVITPPTYDICTQILJKGRHS ELFSKDPDYIVVPYTKVQFDILLQEKEDWPISLGHLEGVHFHLPKDPDLTFTLQTLAIHPMTSTITPLEKGIVFTDGSANGRSVITYIQGREPIKENTQNT AQQAEIVAVITAFEESVQSPENLYTDSKSYVTGLPPEIETATLSPRTKYITELKHLQRLJHKKRQEKFYIGHIRGHTGLPGLAQGNAYADSLTRILTA
MMTVB_P03365-Pro	MMTVB_P03365_VB	P03_VB	365	der iva tiv	e	GRDIMKDKI VRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKIQAALQOLVTEQOLGHLFEESNPWNITPVFVKKKSGKWRLLQ DLRAVNATMHDMGALQGLPSPVAVPKGWEIIMDLQDCHFNKILHPEDCKRFAFSVPNSFKRPYQRFQWKVLPOGMKNSPTLCOKFVDAKILTVRDKY QDSYIVHYMDMILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQIDYDNLYKYLGTTHQIGDSVSYQKLQURTDKLRTLNDFOKLLGNINWIRPELKLTTG ELKPLFEILNGDSNPISTRKLTPAEACKALQLMNERLSTARVKRDLQLSQPSWSCILKTEYTPACLWQDQGVVEWHLPHISPKVITPPTYDICTQILJKGRHS ELFSKDPDYIVVPYTKVQFDILLQEKEDWPISLGHLEGVHFHLPKDPDLTFTLQTLAIHPMTST

MMTVB_	M		der		GRDIMDKIVRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKALQALQOVLTEQLQGHLEESNPWNTPVFVKKSGKWRLLQ
P03365-	MT	P03	iva		DLRAVNA TMHDMGALQPGLPSPVAPPKGWEIIMDLQDCFFNKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDK
Pro_2mut	VB	365	tiv	e	YQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQKYDNLKYLGTHIQGDSVSYQKIQRTDKLRITLNDFOKLLGNINWRPFLKLTG
B					ELKPLFEILNPDSPNSTRKLTPEACKALQLMNERLSTARVKRLDLSQPSWLSCLKTEYTPACLVWDQGVVVEWHLPHISPKVITPYDIFCTQLIKGRHSK
					ELFSKDPDYIVVPYTKVQFDLLI QEKEDWPISLLGFLGEVHFHLPKDPDLTFTLTQTALFPHMTSTTTPLEKGVIFTDGSANGRSVYTIQGREPIKENTQNT
					AQQAIEVAVITAFEEVSQPNLYTDSKYVVTGLPEIETATILSPRTKIYTELKHLQRLIHKRQEKFYIGHRGHTGLPGPLAQGNAYADSLTRIL.T
					WVQEISDSRPMLHIYLNRRRFLGLNTGADKTCIAGRDWPANWPHIQTESLQGLGMACGVARSQPLRWQHEDKSGIHPFVITPFTLWGRDIMKDI
					KVRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKALQALQOVLTEQLQGHLEESNPWNTPVFVKKSGKWRLLQDLRAVNA
MMTVB_	M		der		TMHDMGALQPGLPSPVAPPKGWEIIMDLQDCFFNKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVH
P03365_2	MT	P03	iva		YMDILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQKYDNLKYLGTHIQGDSVSYQKIQRTDKLRITLNDFOKLLGNINWRPFLKLTGELKPLFEIL
mut	VB	365	tiv	e	NPDSPNSTRKLTPEACKALQLMNERLSTARVKRLDLSQPSWLSCLKTEYTPACLVWDQGVVVEWHLPHISPKVITPYDIFCTQLIKGRHSKELFSKDPD
					YIVVPYTKVQFDLLI QEKEDWPISLLGFLGEVHFHLPKDPDLTFTLTQTALFPHMTSTTTPLEKGVIFTDGSANGRSVYTIQGREPIKENTQNTAQQAIEVA
					VITAFEEVSQPNLYTDSKYVVTGLPEIETATILSPRTKIYTELKHLQRLIHKRQEKFYIGHRGHTGLPGPLAQGNAYADSLTRIL.T
					WVQEISDSRPMLHIYLNRRRFLGLNTGADKTCIAGRDWPANWPHIQTESLQGLGMACGVARSQPLRWQHEDKSGIHPFVITPFTLWGRDIMKDI
					KVRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKALQALQOVLTEQLQGHLEESNPWNTPVFVKKSGKWRLLQDLRAVNA
MMTVB_	M		der		TMHDMGALQPGLPSPVAPPKGWEIIMDLQDCFFNKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVH
P03365_2	MT	P03	iva		YMDILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQKYDNLKYLGTHIQGDSVSYQKIQRTDKLRITLNDFOKLLGNINWRPFLKLTGELKPLFEIL
mutB	VB	365	tiv	e	NPDSPNSTRKLTPEACKALQLMNERLSTARVKRLDLSQPSWLSCLKTEYTPACLVWDQGVVVEWHLPHISPKVITPYDIFCTQLIKGRHSKELFSKDPD
					YIVVPYTKVQFDLLI QEKEDWPISLLGFLGEVHFHLPKDPDLTFTLTQTALFPHMTSTTTPLEKGVIFTDGSANGRSVYTIQGREPIKENTQNTAQQAIEVA
					VITAFEEVSQPNLYTDSKYVVTGLPEIETATILSPRTKIYTELKHLQRLIHKRQEKFYIGHRGHTGLPGPLAQGNAYADSLTRIL.T
					WVQEISDSRPMLHIYLNRRRFLGLNTGADKTCIAGRDWPANWPHIQTESLQGLGMACGVARSQPLRWQHEDKSGIHPFVITPFTLWGRDIMKDI
					KVRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKALQALQOVLTEQLQGHLEESNPWNTPVFVKKSGKWRLLQDLRAVNA
MMTVB_	M		der		TMHDMGALQPGLPSPVAPPKGWEIIMDLQDCFFNKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVH
P03365_2	MT	P03	iva		YMDILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQKYDNLKYLGTHIQGDSVSYQKIQRTDKLRITLNDFOKLLGNINWRPFLKLTGELKPLFEIL
mutB WS	VB	365	tiv	e	NPDSPNSTRKLTPEACKALQLMNERLSTARVKRLDLSQPSWLSCLKTEYTPACLVWDQGVVVEWHLPHISPKVITPYDIFCTQLIKGRHSKELFSKDPD
					YIVVPYTKVQFDLLI QEKEDWPISLLGFLGEVHFHLPKDPDLTFTLTQTALFPHMTSTTTPLEKGVIFTDGSANGRSVYTIQGREPIKENTQNTAQQAIEVA
					VITAFEEVSQPNLYTDSKYVVTGLPEIETATILSPRTKIYTELKHLQRLIHKRQEKFYIGHRGHTGLPGPLAQGNAYADSLTRIL.T
					WVQEISDSRPMLHIYLNRRRFLGLNTGADKTCIAGRDWPANWPHIQTESLQGLGMACGVARSQPLRWQHEDKSGIHPFVITPFTLWGRDIMKDI
					KVRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKALQALQOVLTEQLQGHLEESNPWNTPVFVKKSGKWRLLQDLRAVNA
MMTVB_	M		der		TMHDMGALQPGLPSPVAPPKGWEIIMDLQDCFFNKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVH
P03365_2	MT	P03	iva		YMDILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQKYDNLKYLGTHIQGDSVSYQKIQRTDKLRITLNDFOKLLGNINWRPFLKLTGELKPLFEIL
mut WS	VB	365	tiv	e	NPDSPNSTRKLTPEACKALQLMNERLSTARVKRLDLSQPSWLSCLKTEYTPACLVWDQGVVVEWHLPHISPKVITPYDIFCTQLIKGRHSKELFSKDPD
					YIVVPYTKVQFDLLI QEKEDWPISLLGFLGEVHFHLPKDPDLTFTLTQTALFPHMTSTTTPLEKGVIFTDGSANGRSVYTIQGREPIKENTQNTAQQAIEVA
					VITAFEEVSQPNLYTDSKYVVTGLPEIETATILSPRTKIYTELKHLQRLIHKRQEKFYIGHRGHTGLPGPLAQGNAYADSLTRIL.T
					WVQEISDSRPMLHIYLNRRRFLGLNTGADKTCIAGRDWPANWPHIQTESLQGLGMACGVARSQPLRWQHEDKSGIHPFVITPFTLWGRDIMKDI
					KVRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKALQALQOVLTEQLQGHLEESNPWNTPVFVKKSGKWRLLQDLRAVNA
MMTVB_	M		der		TMHDMGALQPGLPSPVAPPKGWEIIMDLQDCFFNKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVH
P03365_2	MT	P03	iva		YMDILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQKYDNLKYLGTHIQGDSVSYQKIQRTDKLRITLNDFOKLLGNINWRPFLKLTGELKPLFEIL
mut WS	VB	365	tiv	e	NPDSPNSTRKLTPEACKALQLMNERLSTARVKRLDLSQPSWLSCLKTEYTPACLVWDQGVVVEWHLPHISPKVITPYDIFCTQLIKGRHSKELFSKDPD
					YIVVPYTKVQFDLLI QEKEDWPISLLGFLGEVHFHLPKDPDLTFTLTQTALFPHMTSTTTPLEKGVIFTDGSANGRSVYTIQGREPIKENTQNTAQQAIEVA
					VITAFEEVSQPNLYTDSKYVVTGLPEIETATILSPRTKIYTELKHLQRLIHKRQEKFYIGHRGHTGLPGPLAQGNAYADSLTRIL.T
					WVQEISDSRPMLHIYLNRRRFLGLNTGADKTCIAGRDWPANWPHIQTESLQGLGMACGVARSQPLRWQHEDKSGIHPFVITPFTLWGRDIMKDI
					KVRLMTDSDPDSQDLMIGAESNLFADQISWKSQDPVWLNQWPLKQEKALQALQOVLTEQLQGHLEESNPWNTPVFVKKSGKWRLLQDLRAVNA
MMTVB_	M		der		TMHDMGALQPGLPSPVAPPKGWEIIMDLQDCFFNKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVH
P03365_2	MT	P03	iva		YMDILLAHPSRSIVDEILTSMIQALNKHGLVNSTEKIQKYDNLKYLGTHIQGDSVSYQKIQRTDKLRITLNDFOKLLGNINWRPFLKLTGELKPLFEIL
mut WS	VB	365	tiv	e	NPDSPNSTRKLTPEACKALQLMNERLSTARVKRLDLSQPSWLSCLKTEYTPACLVWDQGVVVEWHLPHISPKVITPYDIFCTQLIKGRHSKELFSKDPD
					YIVVPYTKVQFDLLI QEKEDWPISLLGFLGEVHFHLPKDPDLTFTLTQTALFPHMTSTTTPLEKGVIFTDGSANGRSVYTIQGREPIKENTQNTAQQAIEVA
					VITAFEEVSQPNLYTDSKYVVTGLPEIETATILSPRTKIYTELKHLQRLIHKRQEKFYIGHRGHTGLPGPLAQGNAYADSLTRIL.T

[illegible]

					WLMQNTMWPACASFVGLDNHYPPNKLQFCFKLHTVFPFQJISKTLNNALLVFTDGSSTGMAAYTLTDITIKFQTNLNSAQLVELQALJAVLSAFPNO PLNIYTDAYSAYLAHSIPLLETVAQIKHSEITAKLFLQCOOLYNSRPFYIGHVRAHSGLPGPIAQGNORADLAKIVASNNIT
MPMV_P 07572_2m ut	MP MV	P07 572	der iva tiv e		LTAADILAPQCAEPTWKSDEPVWVQWPLTNDKLAQAQQLVQEQLEAGHITESSPNTPIVIKKSGKWRLLQDLRANVATMVLMGALQPGLP SPVAIPQGYLKIIDLKDCFHSPILHPSDQKRFASFPLSTNFKPMQRFQWVLPQGMANSPTLCKQYVATAHKVRAHWKQMYIHHYMDILLJAGKDGQ QVLCQFDQKQELTAAGLHIAPEKVQLODPYTYLGFELNGPKITNOKAVIRKDKLQTLNDFQKLGDNWLPYLLKLTGDLKPLFDLTKPDSDFNSHR SLSKEALASLEKETAIAEQVTHINYSILPLIFLJENTALTPTGLFWQDNPMWIHLPASPKKVLPPYDAIAIDLILGRDHSSKYFGIEFSTIQTQPYSKSQID WLMQNTMWPACASFVGLDNHYPPNKLQFCFKLHTVFPFQJISKTLNNALLVFTDGSSTGMAAYTLTDITIKFQTNLNSAQLVELQALJAVLSAFPNO PLNIYTDAYSAYLAHSIPLLETVAQIKHSEITAKLFLQCOOLYNSRPFYIGHVRAHSGLPGPIAQGNORADLAKIVASNNIT
MPMV_P 07572_2m utB	MP MV	P07 572	der iva tiv e		LTAADILAPQCAEPTWKSDEPVWVQWPLTNDKLAQAQQLVQEQLEAGHITESSPNTPIVIKKSGKWRLLQDLRANVATMVLMGALQPGLP SPVAIPQGYLKIIDLKDCFHSPILHPSDQKRFASFPLSTNFKPMQRFQWVLPQGMANSPTLCKQYVATAHKVRAHWKQMYIHHYMDILLJAGKDGQ QVLCQFDQKQELTAAGLHIAPEKVQLODPYTYLGFELNGPKITNOKAVIRKDKLQTLNDFQKLGDNWLPYLLKLTGDLKPLFDLTKPDSDFNSHR SLSKEALASLEKETAIAEQVTHINYSILPLIFLJENTALTPTGLFWQDNPMWIHLPASPKKVLPPYDAIAIDLILGRDHSSKYFGIEFSTIQTQPYSKSQID WLMQNTMWPACASFVGLDNHYPPNKLQFCFKLHTVFPFQJISKTLNNALLVFTDGSSTGMAAYTLTDITIKFQTNLNSAQLVELQALJAVLSAFPNO PLNIYTDAYSAYLAHSIPLLETVAQIKHSEITAKLFLQCOOLYNSRPFYIGHVRAHSGLPGPIAQGNORADLAKIVASNNIT
PERV_Q4 VF22_3m utA_ws	PE RV	Q4 VF Z2	ro ot		LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVQIKASATPVSVRQYPLSKEAQEGIRPHVQRLIQOQILVPVQSPWNTPLPVRK PGTNDYRPVQDLREVNKRQVQDIHTVPNPYNLLCALPQRSWYTVLIDLKDAFFCLRHPTSOPLFAFEWRDPGTGRGQLTWTRLPOGFKNSPTIFNEA LHRDLANFRIQHPQVTLQYVDDLLLAGATKQDCLEGTKALLELSDLYGRASAKKAQICRREVTYLGYSLRDGQWLTARKKTVVQIPAPTTAKQV REFLGKAGFCRLFIPIGFAITLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALPDVTKPFTLYYDERKGVARGVLTQTLGPWRRPVA YLSKLL DPV ASGWVPVCLKAIAA VAIL VKDADKLTJLQGNITVIAPHALENVRQPPDRWMTNARMTHYQSLLLTERTVTFAPPAALNPATLLPEETDEPVTHDCHQ LIEETGVRRKDLTDIPLTGEVLTWFTDGSYVVEGKRMAGAAVVDGTRTWASSLPEGTSQAQKAEALMALTQALRLAEGKSINITYDSRYAFATAHVHGA YKQRGWLTSAGREIKNKEEILSLEALHLPKRLAIHCPGHQKADPISRGNQMDRVAKQAAQGVNLL
PERV_Q4 VF22	PE RV	Q4 VF Z2	der iva tiv e		TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVQIKASATPVSVRQYPLSKEAQEGIRPHVQRLIQOQILVPVQSPWNTPLP VRKPGTNDYRPVQDLREVNKRQVQDIHTVPNPYNLLCALPQRSWYTVLIDLKDAFFCLRHPTSOPLFAFEWRDPGTGRGQLTWTRLPOGFKNSPTIF DEALHRLANFRIQHPQVTLQYVDDLLLAGATKQDCLEGTKALLELSDLYGRASAKKAQICRREVTYLGYSLRDGQWLTARKKTVVQIPAPTTAK QVREFLGTAGFCRLFIPIGFAITLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALPDVTKPFTLYYDERKGVARGVLTQTLGPWRRPVA YL SKLLDPV ASGWVPVCLKAIAA VAIL VKDADKLTJLQGNITVIAPHALENVRQPPDRWMTNARMTHYQSLLLTERTVTFAPPAALNPATLLPEETDEPVTHD CHOLLEETGVRRKDLTDIPLTGEVLTWFTDGSYVVEGKRMAGAAVVDGTRTWASSLPEGTSQAQKAEALMALTQALRLAEGKSINITYDSRYAFATAHV HGA YKQRGWLTSAGREIKNKEEILSLEALHLPKRLAIHCPGHQKADPISRGNQMDRVAKQAAQGVNLL
PERV_Q4 VF22_3m ut	PE RV	Q4 VF Z2	der iva tiv e		TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVQIKASATPVSVRQYPLSKEAQEGIRPHVQRLIQOQILVPVQSPWNTPLP VRKPGTNDYRPVQDLREVNKRQVQDIHTVPNPYNLLCALPQRSWYTVLIDLKDAFFCLRHPTSOPLFAFEWRDPGTGRGQLTWTRLPOGFKNSPTIF NEALHRLANFRIQHPQVTLQYVDDLLLAGATKQDCLEGTKALLELSDLYGRASAKKAQICRREVTYLGYSLRDGQWLTARKKTVVQIPAPTTAK QVREFLGTAGFCRLFIPIGFAITLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALPDVTKPFTLYYDERKGVARGVLTQTLGPWRRPVA YLS KKLDPV ASGWVPVCLKAIAA VAIL VKDADKLTJLQGNITVIAPHALENVRQPPDRWMTNARMTHYQSLLLTERTVTFAPPAALNPATLLPEETDEPVTHD HQLLEETGVRRKDLTDIPLTGEVLTWFTDGSYVVEGKRMAGAAVVDGTRTWASSLPEGTSQAQKAEALMALTQALRLAEGKSINITYDSRYAFATAHV GAIYKQRGWLTSAGREIKNKEEILSLEALHLPKRLAIHCPGHQKADPISRGNQMDRVAKQAAQGVNLL
PERV_Q4 VF22_3m utA_ws	PE RV	Q4 VF Z2	ro ot		LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVQIKASATPVSVRQYPLSKEAQEGIRPHVQRLIQOQILVPVQSPWNTPLPVRK PGTNDYRPVQDLREVNKRQVQDIHTVPNPYNLLCALPQRSWYTVLIDLKDAFFCLRHPTSOPLFAFEWRDPGTGRGQLTWTRLPOGFKNSPTIFNEA LHRDLANFRIQHPQVTLQYVDDLLLAGATKQDCLEGTKALLELSDLYGRASAKKAQICRREVTYLGYSLRDGQWLTARKKTVVQIPAPTTAKQV REFLGKAGFCRLFIPIGFAITLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALPDVTKPFTLYYDERKGVARGVLTQTLGPWRRPVA YLSKLL DPV ASGWVPVCLKAIAA VAIL VKDADKLTJLQGNITVIAPHALENVRQPPDRWMTNARMTHYQSLLLTERTVTFAPPAALNPATLLPEETDEPVTHDCHQ LIEETGVRRKDLTDIPLTGEVLTWFTDGSYVVEGKRMAGAAVVDGTRTWASSLPEGTSQAQKAEALMALTQALRLAEGKSINITYDSRYAFATAHVHGA YKQRGWLTSAGREIKNKEEILSLEALHLPKRLAIHCPGHQKADPISRGNQMDRVAKQAAQGVNLL
PERV_Q4 VF22	PE RV	Q4 VF Z2	der iva tiv e		TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVQIKASATPVSVRQYPLSKEAQEGIRPHVQRLIQOQILVPVQSPWNTPLP VRKPGTNDYRPVQDLREVNKRQVQDIHTVPNPYNLLCALPQRSWYTVLIDLKDAFFCLRHPTSOPLFAFEWRDPGTGRGQLTWTRLPOGFKNSPTIF DEALHRLANFRIQHPQVTLQYVDDLLLAGATKQDCLEGTKALLELSDLYGRASAKKAQICRREVTYLGYSLRDGQWLTARKKTVVQIPAPTTAK QVREFLGTAGFCRLFIPIGFAITLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALPDVTKPFTLYYDERKGVARGVLTQTLGPWRRPVA YL SKLLDPV ASGWVPVCLKAIAA VAIL VKDADKLTJLQGNITVIAPHALENVRQPPDRWMTNARMTHYQSLLLTERTVTFAPPAALNPATLLPEETDEPVTHD CHOLLEETGVRRKDLTDIPLTGEVLTWFTDGSYVVEGKRMAGAAVVDGTRTWASSLPEGTSQAQKAEALMALTQALRLAEGKSINITYDSRYAFATAHV HGA YKQRGWLTSAGREIKNKEEILSLEALHLPKRLAIHCPGHQKADPISRGNQMDRVAKQAAQGVNLL

PERV_Q4_VF22_3mut	PE_VF_Z2	der iva tiv e	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKAQVPPQVLIQKASATPVSVRQYPLSKEAQEGIRPHIVQRLLIQOGILVPVQSPWNTPLLPVRKPGINDYRPVQDLREVNKRVRQDIIHTVPNPNYLLCALPPQRSWYTVLIDLKDAFFCLRLHPTSQPLFAFEWFRDPTGTRTQQLTWTRLPQOGKNSPTIFNEALHRDLANERHQHPQVTLQYVDDLLAGATKQDCLEGTKALLELSDLYRASAKKAQICRREVTVLYGYSLRDQQRWLTEARKKTVVQIPAPITAKQVREHLGTAGFCRLWPGFATLAPLTKPKGEWSAHEPHQKAFDAKALLSAPALPDVITKPTLYYDERKKGARGVLTQILGPWRPVAAYLSKKLDPVASGWPVYCLKAIAAAVAILVKPDADKLTQGNITVIAPHALENVRQDPWMTNARTMTHYOSLTLTERTVTEAPPAALNATLLPEETDEPVTDC HQLLIETGVRKDLTDIPLTGEVLWTFDTGSSYVVEGKRMAGAAVVDGTRITWASSLPEGTSQAQAEALMALTQALRLAEGKSNITYTDSRYAFATAHVHGAIKQKQWLTASAGREIKNEEILSLEALPKRLAIHCPCHQKQADPSRSGNQMAQDRVAKQAQGVNLL
			MDPLQLLQPLEAEIKGTIKLKAHWSGATITCPEAFLEDERPIQTMLIKTHGEKQDDVYVLTFTKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQ LTVL VPLHEYQERLLQQTALPKEQKELLQKFLKYDALWQHWNQVGHRRIKPHINATGTLAPRQKQYPINPKAKPSIQIIVDDLLKQGVLIQONSTM NTPVYPKPDGKWRMVL DYREVNTKIPIIAAQNHQSAGILSSYRGKYKTITLDLTNGFWAHPITPESYWLTAFTWQKGQYCWTRLPQGLNSPALFT ADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFISILLNAGYVVSLLKSEIAQREVEFLGNITKEGRGLTDTFKQKLLNITPPKDLK QILQSILGILLNEARNEIPNYSELVKPLYTIVANANGKFSWTEDESNQLOHISVLNQADNLEERNPEITRIJKVNSSPSAGYIRYNEGSKRPMYVNYIESK FARNIPNYSELVKPLYTIVANANGKFSWTEDESNQLOHISVLNQADNLEERNPEITRIJKVNSSPSAGYIRYNEGSKRPMYVNYIESKAEAKFTQTEK KILLTTMHKGLIKAMDLAMGOEILVYSPVSMITKIQRTIPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSE SAIKHPDVNKSAGMGIAQVQFPEYKIVHQSWSIPLGDHTAQLAEIAAVEFAECKALKISGPNVITDTSFYVAESANKELPYWKSNGFLNNKXKPLRHV SKWSIAECLQLKPDIIIMHEKGHOQPMITLHTEGNNLADKLATQGSYVVH
SFV1_P23_074	SF_V1	ro ot	VPWLMMKKPLQLTVL VPLHEYQERLLQQTALPKEQKELLQKFLKYDALWQHWNQVGHRRIKPHINATGTLAPRQKQYPINPKAKPSIQIIVDDLLKQ GVLQONSTMNTPVYPKPDGKWRMVL DYREVNTKIPIIAAQNHQSAGILSSYRGKYKTITLDLTNGFWAHPITPESYWLTAFTWQKGQYCWTRLPQ GFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFISILLNAGYVVSLLKSEIAQREVEFLGNITKEGRGLTDTFKQKLLNITPPKDLK QILQSILGILLNEARNEIPNYSELVKPLYTIVANANGKFSWTEDESNQLOHISVLNQADNLEERNPEITRIJKVNSSPSAGYIRYNEGSKRPMYVNYIESK AEAKFTQTEKLLTTMHKGLIKAMDLAMGOEILVYSPVSMITKIQRTIPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSE FAMVFTDGSAAIKHPDVNKSAGMGIAQVQFPEYKIVHQSWSIPLGDHTAQLAEIAAVEFAECKALKISGPNVITDTSFYVAESANKELPYWKSNGFLN NKKKPLRHVSKWSIAECLQLKPDIIIMHEKGHOQPMITLHTEGNNLADKLATQGSYVVH
SFV1_P23_074-Pro	SF_V1	der iva tiv e	VPWLMMKKPLQLTVL VPLHEYQERLLQQTALPKEQKELLQKFLKYDALWQHWNQVGHRRIKPHINATGTLAPRQKQYPINPKAKPSIQIIVDDLLKQ GVLQONSTMNTPVYPKPDGKWRMVL DYREVNTKIPIIAAQNHQSAGILSSYRGKYKTITLDLTNGFWAHPITPESYWLTAFTWQKGQYCWTRLPQ GFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFISILLNAGYVVSLLKSEIAQREVEFLGNITKEGRGLTDTFKQKLLNITPPKDLK QILQSILGILLNEARNEIPNYSELVKPLYTIVANANGKFSWTEDESNQLOHISVLNQADNLEERNPEITRIJKVNSSPSAGYIRYNEGSKRPMYVNYIESK AEAKFTQTEKLLTTMHKGLIKAMDLAMGOEILVYSPVSMITKIQRTIPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSE FAMVFTDGSAAIKHPDVNKSAGMGIAQVQFPEYKIVHQSWSIPLGDHTAQLAEIAAVEFAECKALKISGPNVITDTSFYVAESANKELPYWKSNGFLN NKKKPLRHVSKWSIAECLQLKPDIIIMHEKGHOQPMITLHTEGNNLADKLATQGSYVVH
SFV1_P23_074-Pro_2mut	SF_V1	der iva tiv e	VPWLMMKKPLQLTVL VPLHEYQERLLQQTALPKEQKELLQKFLKYDALWQHWNQVGHRRIKPHINATGTLAPRQKQYPINPKAKPSIQIIVDDLLKQ GVLQONSTMNTPVYPKPDGKWRMVL DYREVNTKIPIIAAQNHQSAGILSSYRGKYKTITLDLTNGFWAHPITPESYWLTAFTWQKGQYCWTRLPQ GFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFISILLNAGYVVSLLKSEIAQREVEFLGNITKEGRGLTDTFKQKLLNITPPKDLK QILQSILGILLNEARNEIPNYSELVKPLYTIVANANGKFSWTEDESNQLOHISVLNQADNLEERNPEITRIJKVNSSPSAGYIRYNEGSKRPMYVNYIESK AEAKFTQTEKLLTTMHKGLIKAMDLAMGOEILVYSPVSMITKIQRTIPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSE FAMVFTDGSAAIKHPDVNKSAGMGIAQVQFPEYKIVHQSWSIPLGDHTAQLAEIAAVEFAECKALKISGPNVITDTSFYVAESANKELPYWKSNGFLN NKKKPLRHVSKWSIAECLQLKPDIIIMHEKGHOQPMITLHTEGNNLADKLATQGSYVVH
SFV1_P23_074-Pro_A	SF_V1	der iva tiv e	MDPLQLLQPLEAEIKGTIKLKAHWSGATITCPEAFLEDERPIQTMLIKTHGEKQDDVYVLTFTKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQ LTVL VPLHEYQERLLQQTALPKEQKELLQKFLKYDALWQHWNQVGHRRIKPHINATGTLAPRQKQYPINPKAKPSIQIIVDDLLKQGVLIQONSTM NTPVYPKPDGKWRMVL DYREVNTKIPIIAAQNHQSAGILSSYRGKYKTITLDLTNGFWAHPITPESYWLTAFTWQKGQYCWTRLPQGLNSPALFT ADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFISILLNAGYVVSLLKSEIAQREVEFLGNITKEGRGLTDTFKQKLLNITPPKDLK QILQSILGILLN FARNIPNYSELVKPLYTIVANANGKFSWTEDESNQLOHISVLNQADNLEERNPEITRIJKVNSSPSAGYIRYNEGSKRPMYVNYIESKAEAKFTQTEK LLTTMHKGLIKAMDLAMGOEILVYSPVSMITKIQRTIPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSE FAMVFTDGS AIKHPDVNKSAGMGIAQVQFPEYKIVHQSWSIPLGDHTAQLAEIAAVEFAECKALKISGPNVITDTSFYVAESANKELPYWKSNGFLNNKXKPLRHVS KWSIAECLQLKPDIIIMHEKGHOQPMITLHTEGNNLADKLATQGSYVVH
SFV1_P23_074_2mut	SF_V1	der iva tiv e	MDPLQLLQPLEAEIKGTIKLKAHWSGATITCPEAFLEDERPIQTMLIKTHGEKQDDVYVLTFTKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQ LTVL VPLHEYQERLLQQTALPKEQKELLQKFLKYDALWQHWNQVGHRRIKPHINATGTLAPRQKQYPINPKAKPSIQIIVDDLLKQGVLIQONSTM NTPVYPKPDGKWRMVL DYREVNTKIPIIAAQNHQSAGILSSYRGKYKTITLDLTNGFWAHPITPESYWLTAFTWQKGQYCWTRLPQGLNSPALFT ADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFISILLNAGYVVSLLKSEIAQREVEFLGNITKEGRGLTDTFKQKLLNITPPKDLK QILQSILGILLN FARNIPNYSELVKPLYTIVANANGKFSWTEDESNQLOHISVLNQADNLEERNPEITRIJKVNSSPSAGYIRYNEGSKRPMYVNYIESKAEAKFTQTEK LLTTMHKGLIKAMDLAMGOEILVYSPVSMITKIQRTIPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSE FAMVFTDGS AIKHPDVNKSAGMGIAQVQFPEYKIVHQSWSIPLGDHTAQLAEIAAVEFAECKALKISGPNVITDTSFYVAESANKELPYWKSNGFLNNKXKPLRHVS KWSIAECLQLKPDIIIMHEKGHOQPMITLHTEGNNLADKLATQGSYVVH
SFV1_P23_074_2mut_A	SF_V1	der iva tiv e	MDPLQLLQPLEAEIKGTIKLKAHWSGATITCPEAFLEDERPIQTMLIKTHGEKQDDVYVLTFTKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQ LTVL VPLHEYQERLLQQTALPKEQKELLQKFLKYDALWQHWNQVGHRRIKPHINATGTLAPRQKQYPINPKAKPSIQIIVDDLLKQGVLIQONSTM NTPVYPKPDGKWRMVL DYREVNTKIPIIAAQNHQSAGILSSYRGKYKTITLDLTNGFWAHPITPESYWLTAFTWQKGQYCWTRLPQGLNSPALFT ADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFISILLNAGYVVSLLKSEIAQREVEFLGNITKEGRGLTDTFKQKLLNITPPKDLK QILQSILGILLN FARNIPNYSELVKPLYTIVANANGKFSWTEDESNQLOHISVLNQADNLEERNPEITRIJKVNSSPSAGYIRYNEGSKRPMYVNYIESKAEAKFTQTEK LLTTMHKGLIKAMDLAMGOEILVYSPVSMITKIQRTIPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSE FAMVFTDGS AIKHPDVNKSAGMGIAQVQFPEYKIVHQSWSIPLGDHTAQLAEIAAVEFAECKALKISGPNVITDTSFYVAESANKELPYWKSNGFLNNKXKPLRHVS KWSIAECLQLKPDIIIMHEKGHOQPMITLHTEGNNLADKLATQGSYVVH

						AIKHPDVNKSISAGMGIAQVQFIPEYKIVHQQWSIPLGDHTAQLAEIAAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWQSNKGFLNNKKKPLRHVHS KWKSIAECLQLKPDIIIMHEKGHQPMITLTHTEGNNLADKLATQGSYVNH
						MDPQLQLQPLEAIEIKGTKLKAHWSNGATTCVPOAFLEEEVPIKNWIKTHGEKEQPYYLITFKIQGRKVEAEVISSPYDYILVSPSDIPWLKKKPLQLTIT LVPLOQYEERILLKQTMLTGSYKEKLSQLFKYDALWQHWEVNGVGHRRIKPHHIIATGTVNRPQKQYPINPKAKASIQIVINDILLKQGVLIQQNSIMNITP VYVPKPDGKWRMVLVDYREVNKTPIJAAQNQHSAGILSSIFRGKYKTITDLDSNGFWAHSITPESYWLTAFTWLGQYCWTRLPQGLNSPALFTADVV DLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLNAGYVVSLLKKSEIAQHEVEFGNITKEGRGLTEFTKQKLLNITPPRDLKQLOQSILGILLNFAR NFIPISELVKPLYNIIATAPGYTITWTDNSQQLQNIISMLNSAENLEERNPEVRLIMKNTSPSAGYIRFYNEFAKRPMILYNVYVTKAEVKFTNTEKLL LTTTHKGLIKALDLGMGOEILVYSPVSMTKIOKTPIPERKALPIRWITWMSYLEDPRIOHFHYDKTILPELQQVPTVTDIIAKIKHPSEFSMVFTDGSAAUKH PNNVNSHNAAGMGIAQVQFKPEFTVINTWISPLGDHTAQLAEVAAVEFACKKALKIDGPPVLIVTDSFYVAESVNKELPYWQSNKGFFNNKKKPLKHVSKW KSIADCIQLKPDIIIMHEKGHQPTASTHTEGNNLADKLATQGSYVNH
SFV3L_P 27401_P	SF V3 L	P27 401	ro	ot		IPWLMKPLQLTTLVPLQPEYERILLKQTMLTGSYKEKLSQLFKYDALWQHWEVNGVGHRRIKPHHIIATGTVNRPQKQYPINPKAKASIQIVINDILLKQ GVLIQQNSIMNITPVYVPKPDGKWRMVLVDYREVNKTPIJAAQNQHSAGILSSIFRGKYKTITDLDSNGFWAHSITPESYWLTAFTWLGQYCWTRLPQGL FLNSPALFTADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLNAGYVVSLLKKSEIAQHEVEFGNITKEGRGLTEFTKQKLLNITPPRDLK QLOQSILGILLNFARNFIPISELVKPLYNIIATAPGYTITWTDNSQQLQNIISMLNSAENLEERNPEVRLIMKNTSPSAGYIRFYNEFAKRPMILYNVYV KAEVKFTNTEKLLTTHHGLIKALDLGMGOEILVYSPVSMTKIOKTPIPERKALPIRWITWMSYLEDPRIOHFHYDKTILPELQQVPTVTDIIAKIKHPSEFS MVFTDGSAAUKHPNVNKSNAAGMGIAQVQFKPEFTVINTWISPLGDHTAQLAEVAAVEFACKKALKIDGPPVLIVTDSFYVAESVNKELPYWQSNKGFFN NKKKPLKHVSKWKSIAECLQLKPDIIIMHEKGHQPTASTHTEGNNLADKLATQGSYVNH
SFV3L_P 27401_P Pro_2mut	SF V3 L	P27 401	der iva	tiv e		IPWLMKPLQLTTLVPLQPEYERILLKQTMLTGSYKEKLSQLFKYDALWQHWEVNGVGHRRIKPHHIIATGTVNRPQKQYPINPKAKASIQIVINDILLKQ GVLIQQNSIMNITPVYVPKPDGKWRMVLVDYREVNKTPIJAAQNQHSAGILSSIFRGKYKTITDLDSNGFWAHSITPESYWLTAFTWLGQYCWTRLPQGL FLNSPALFTADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLNAGYVVSLLKKSEIAQHEVEFGNITKEGRGLTEFTKQKLLNITPPRDLK QLOQSILGILLNFARNFIPISELVKPLYNIIATAPGYTITWTDNSQQLQNIISMLNSAENLEERNPEVRLIMKNTSPSAGYIRFYNEFAKRPMILYNVYV KAEVKFTNTEKLLTTHHGLIKALDLGMGOEILVYSPVSMTKIOKTPIPERKALPIRWITWMSYLEDPRIOHFHYDKTILPELQQVPTVTDIIAKIKHPSEFS MVFTDGSAAUKHPNVNKSNAAGMGIAQVQFKPEFTVINTWISPLGDHTAQLAEVAAVEFACKKALKIDGPPVLIVTDSFYVAESVNKELPYWQSNKGFFN NKKKPLKHVSKWKSIAECLQLKPDIIIMHEKGHQPTASTHTEGNNLADKLATQGSYVNH
SFV3L_P 27401_P Pro_2mut A	SF V3 L	P27 401	der iva	tiv e		IPWLMKPLQLTTLVPLQPEYERILLKQTMLTGSYKEKLSQLFKYDALWQHWEVNGVGHRRIKPHHIIATGTVNRPQKQYPINPKAKASIQIVINDILLKQ GVLIQQNSIMNITPVYVPKPDGKWRMVLVDYREVNKTPIJAAQNQHSAGILSSIFRGKYKTITDLDSNGFWAHSITPESYWLTAFTWLGQYCWTRLPQGL FLNSPALFTADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLNAGYVVSLLKKSEIAQHEVEFGNITKEGRGLTEFTKQKLLNITPPRDLK QLOQSILGILLNFARNFIPISELVKPLYNIIATAPGYTITWTDNSQQLQNIISMLNSAENLEERNPEVRLIMKNTSPSAGYIRFYNEFAKRPMILYNVYV KAEVKFTNTEKLLTTHHGLIKALDLGMGOEILVYSPVSMTKIOKTPIPERKALPIRWITWMSYLEDPRIOHFHYDKTILPELQQVPTVTDIIAKIKHPSEFS MVFTDGSAAUKHPNVNKSNAAGMGIAQVQFKPEFTVINTWISPLGDHTAQLAEVAAVEFACKKALKIDGPPVLIVTDSFYVAESVNKELPYWQSNKGFFN NKKKPLKHVSKWKSIAECLQLKPDIIIMHEKGHQPTASTHTEGNNLADKLATQGSYVNH
SFV3L_P 27401_2mut	SF V3 L	P27 401	der iva	tiv e		MDPQLQLQPLEAIEIKGTKLKAHWSNGATTCVPOAFLEEEVPIKNWIKTHGEKEQPYYLITFKIQGRKVEAEVISSPYDYILVSPSDIPWLKKKPLQLTIT LVPLOQYEERILLKQTMLTGSYKEKLSQLFKYDALWQHWEVNGVGHRRIKPHHIIATGTVNRPQKQYPINPKAKASIQIVINDILLKQGVLIQQNSIMNITP VYVPKPDGKWRMVLVDYREVNKTPIJAAQNQHSAGILSSIFRGKYKTITDLDSNGFWAHSITPESYWLTAFTWLGQYCWTRLPQGLNSPALFTADVV DLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLNAGYVVSLLKKSEIAQHEVEFGNITKEGRGLTEFTKQKLLNITPPRDLKQLOQSILGILLNFAR NFIPISELVKPLYNIIATAPGYTITWTDNSQQLQNIISMLNSAENLEERNPEVRLIMKNTSPSAGYIRFYNEFAKRPMILYNVYVTKAEVKFTNTEKLL LTTTHKGLIKALDLGMGOEILVYSPVSMTKIOKTPIPERKALPIRWITWMSYLEDPRIOHFHYDKTILPELQQVPTVTDIIAKIKHPSEFSMVFTDGSAAUKH PNNVNSHNAAGMGIAQVQFKPEFTVINTWISPLGDHTAQLAEVAAVEFACKKALKIDGPPVLIVTDSFYVAESVNKELPYWQSNKGFFNNKKKPLKHVSKW KSIADCIQLKPDIIIMHEKGHQPTASTHTEGNNLADKLATQGSYVNH
SFV3L_P 27401_2mut ut	SF V3 L	P27 401	der iva	tiv e		MDPQLQLQPLEAIEIKGTKLKAHWSNGATTCVPOAFLEEEVPIKNWIKTHGEKEQPYYLITFKIQGRKVEAEVISSPYDYILVSPSDIPWLKKKPLQLTIT LVPLOQYEERILLKQTMLTGSYKEKLSQLFKYDALWQHWEVNGVGHRRIKPHHIIATGTVNRPQKQYPINPKAKASIQIVINDILLKQGVLIQQNSIMNITP VYVPKPDGKWRMVLVDYREVNKTPIJAAQNQHSAGILSSIFRGKYKTITDLDSNGFWAHSITPESYWLTAFTWLGQYCWTRLPQGLNSPALFTADVV DLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLNAGYVVSLLKKSEIAQHEVEFGNITKEGRGLTEFTKQKLLNITPPRDLKQLOQSILGILLNFAR NFIPISELVKPLYNIIATAPGYTITWTDNSQQLQNIISMLNSAENLEERNPEVRLIMKNTSPSAGYIRFYNEFAKRPMILYNVYVTKAEVKFTNTEKLL LTTTHKGLIKALDLGMGOEILVYSPVSMTKIOKTPIPERKALPIRWITWMSYLEDPRIOHFHYDKTILPELQQVPTVTDIIAKIKHPSEFSMVFTDGSAAUKH PNNVNSHNAAGMGIAQVQFKPEFTVINTWISPLGDHTAQLAEVAAVEFACKKALKIDGPPVLIVTDSFYVAESVNKELPYWQSNKGFFNNKKKPLKHVSKW KSIADCIQLKPDIIIMHEKGHQPTASTHTEGNNLADKLATQGSYVNH
SFV3L_P 27401_2mut ut	SF V3 L	P27 401	der iva	tiv e		MDPQLQLQPLEAIEIKGTKLKAHWSNGATTCVPOAFLEEEVPIKNWIKTHGEKEQPYYLITFKIQGRKVEAEVISSPYDYILVSPSDIPWLKKKPLQLTIT LVPLOQYEERILLKQTMLTGSYKEKLSQLFKYDALWQHWEVNGVGHRRIKPHHIIATGTVNRPQKQYPINPKAKASIQIVINDILLKQGVLIQQNSIMNITP VYVPKPDGKWRMVLVDYREVNKTPIJAAQNQHSAGILSSIFRGKYKTITDLDSNGFWAHSITPESYWLTAFTWLGQYCWTRLPQGLNSPALFTADVV DLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLNAGYVVSLLKKSEIAQHEVEFGNITKEGRGLTEFTKQKLLNITPPRDLKQLOQSILGILLNFAR NFIPISELVKPLYNIIATAPGYTITWTDNSQQLQNIISMLNSAENLEERNPEVRLIMKNTSPSAGYIRFYNEFAKRPMILYNVYVTKAEVKFTNTEKLL LTTTHKGLIKALDLGMGOEILVYSPVSMTKIOKTPIPERKALPIRWITWMSYLEDPRIOHFHYDKTILPELQQVPTVTDIIAKIKHPSEFSMVFTDGSAAUKH PNNVNSHNAAGMGIAQVQFKPEFTVINTWISPLGDHTAQLAEVAAVEFACKKALKIDGPPVLIVTDSFYVAESVNKELPYWQSNKGFFNNKKKPLKHVSKW KSIADCIQLKPDIIIMHEKGHQPTASTHTEGNNLADKLATQGSYVNH
SFV3L_P 27401_2mut ut	SF V3 L	P27 401	der iva	tiv e		MDPQLQLQPLEAIEIKGTKLKAHWSNGATTCVPOAFLEEEVPIKNWIKTHGEKEQPYYLITFKIQGRKVEAEVISSPYDYILVSPSDIPWLKKKPLQLTIT LVPLOQYEERILLKQTMLTGSYKEKLSQLFKYDALWQHWEVNGVGHRRIKPHHIIATGTVNRPQKQYPINPKAKASIQIVINDILLKQGVLIQQNSIMNITP VYVPKPDGKWRMVLVDYREVNKTPIJAAQNQHSAGILSSIFRGKYKTITDLDSNGFWAHSITPESYWLTAFTWLGQYCWTRLPQGLNSPALFTADVV DLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLNAGYVVSLLKKSEIAQHEVEFGNITKEGRGLTEFTKQKLLNITPPRDLKQLOQSILGILLNFAR NFIPISELVKPLYNIIATAPGYTITWTDNSQQLQNIISMLNSAENLEERNPEVRLIMKNTSPSAGYIRFYNEFAKRPMILYNVYVTKAEVKFTNTEKLL LTTTHKGLIKALDLGMGOEILVYSPVSMTKIOKTPIPERKALPIRWITWMSYLEDPRIOHFHYDKTILPELQQVPTVTDIIAKIKHPSEFSMVFTDGSAAUKH PNNVNSHNAAGMGIAQVQFKPEFTVINTWISPLGDHTAQLAEVAAVEFACKKALKIDGPPVLIVTDSFYVAESVNKELPYWQSNKGFFNNKKKPLKHVSKW KSIADCIQLKPDIIIMHEKGHQPTASTHTEGNNLADKLATQGSYVNH

				AVDLLKEVPNVQVYVDDIYL SHDNPHIEHQLEKVFQILLQAGYVYVSLKKSEIGQRTVEFLGFNITKEGRGLTDITFKTKLLNVTPPKDLKQLQSILGLLN FARNFIPNFAELVQTLNLIASSKGKYEWTEDNTKQNKVIEALNTASNEERLPDQRLVIKVNTPSAGYVRYVYVNESGKKPMYLYNVFSKAELKFSM LEKLLTMMHKALIKAMDLMAGQEILVYSPVSMTKIOKTPLPERKALPIRWITWMTYLEDPRQHFHYDKTLPELKHIPDVTYSSIPPLKHPQSQYEGVECTD GSAIKSPDPTKSNAGMIGVHAIYNPEYKILNQWSIPLGHHTAQMAEIAA VEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLK HISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLAATQGSYVVN
SFVCP_Q 87040-Pro	SF VC P	Q8 704 0	der iva tiv e	VPWL'TQOPLQLTILVPLEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWEQVGHKIRPHINIAITGDYPPRQKQYPINPKAKPSIQIVIDDILLKQG VLTPQNSTMTPVYPKPDGRWRMVLIDYREVNKTPIPLTAQONQHSAGILATIVRQKYKTITLDLANGFWAHPITPDSYWLTAFTWQKGQYCWTRLPO GFLNSPALFTADA VDLLKEVPNVQVYVDDIYL SHDNPHIEHQLEKVFQILLQAGYVYVSLKKSEIGQRTVEFLGFNITKEGRGLTDITFKTKLLNVTPPKD LKQLQSILGLLNFARNFIPNFAELVQTLNLIASSKGKYEWTEDNTKQNKVIEALNTASNEERLPDQRLVIKVNTPSAGYVRYVYVNESGKKPMYLYNV VFSKAELKFSMLEKLLTMMHKALIKAMDLMAGQEILVYSPVSMTKIOKTPLPERKALPIRWITWMTYLEDPRQHFHYDKTLPELKHIPDVTYSSIPPLK PSQYEGVCTDGS AIKSPDPTKSNAGMIGVHAIYNPEYKILNQWSIPLGHHTAQMAEIAA VEFACKKALKVPGPVLVITDSFYVAESANKELPYWKS GFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLAATQGSYVVN
SFVCP_Q 87040- Pro_2mut	SF VC P	Q8 704 0	der iva tiv e	VPWL'TQOPLQLTILVPLEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWEQVGHKIRPHINIAITGDYPPRQKQYPINPKAKPSIQIVIDDILLKQG VLTPQNSTMTPVYPKPDGRWRMVLIDYREVNKTPIPLTAQONQHSAGILATIVRQKYKTITLDLANGFWAHPITPDSYWLTAFTWQKGQYCWTRLPO GFLNSPALFNADA VDLLKEVPNVQVYVDDIYL SHDNPHIEHQLEKVFQILLQAGYVYVSLKKSEIGQRTVEFLGFNITKEGRGLTDITFKTKLLNVTPPKD LKQLQSILGLLNFARNFIPNFAELVQTLNLIASSKGKYEWTEDNTKQNKVIEALNTASNEERLPDQRLVIKVNTPSAGYVRYVYVNESGKKPMYLYNV VFSKAELKFSMLEKLLTMMHKALIKAMDLMAGQEILVYSPVSMTKIOKTPLPERKALPIRWITWMTYLEDPRQHFHYDKTLPELKHIPDVTYSSIPPLK PSQYEGVCTDGS AIKSPDPTKSNAGMIGVHAIYNPEYKILNQWSIPLGHHTAQMAEIAA VEFACKKALKVPGPVLVITDSFYVAESANKELPYWKS GFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLAATQGSYVVN
SFVCP_Q 87040- Pro_2mut A	SF VC P	Q8 704 0	der iva tiv e	VPWL'TQOPLQLTILVPLEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWEQVGHKIRPHINIAITGDYPPRQKQYPINPKAKPSIQIVIDDILLKQG VLTPQNSTMTPVYPKPDGRWRMVLIDYREVNKTPIPLTAQONQHSAGILATIVRQKYKTITLDLANGFWAHPITPDSYWLTAFTWQKGQYCWTRLPO GFLNSPALFNADA VDLLKEVPNVQVYVDDIYL SHDNPHIEHQLEKVFQILLQAGYVYVSLKKSEIGQRTVEFLGFNITKEGRGLTDITFKTKLLNVTPPKD LKQLQSILGLLNFARNFIPNFAELVQTLNLIASSKGKYEWTEDNTKQNKVIEALNTASNEERLPDQRLVIKVNTPSAGYVRYVYVNESGKKPMYLYNV VFSKAELKFSMLEKLLTMMHKALIKAMDLMAGQEILVYSPVSMTKIOKTPLPERKALPIRWITWMTYLEDPRQHFHYDKTLPELKHIPDVTYSSIPPLK PSQYEGVCTDGS AIKSPDPTKSNAGMIGVHAIYNPEYKILNQWSIPLGHHTAQMAEIAA VEFACKKALKVPGPVLVITDSFYVAESANKELPYWKS GFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLAATQGSYVVN
SFVCP_Q 87040_2m ut	SF VC P	Q8 704 0	der iva tiv e	MNPLQLIQLPLAEVKGTKLLAHWSGATITCIPESFLDEQPIKOTILKTIHGEKQONVYVLTFTKVKGRKVEAEVIASPYEYILLSPITDVPWL'TQOPLQLTIL LVPLQYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWEQVGHKIRPHINIAITGDYPPRQKQYPINPKAKPSIQIVIDDILLKQGVLTPQNSTMTP VYPKPDGRWRMVLIDYREVNKTPIPLTAQONQHSAGILATIVRQKYKTITLDLANGFWAHPITPDSYWLTAFTWQKGQYCWTRLPOGFLNSPALFNAD AVDLLKEVPNVQVYVDDIYL SHDNPHIEHQLEKVFQILLQAGYVYVSLKKSEIGQRTVEFLGFNITKEGRGLTDITFKTKLLNVTPPKDLKQLQSILGLLN FARNFIPNFAELVQTLNLIASSKGKYEWTEDNTKQNKVIEALNTASNEERLPDQRLVIKVNTPSAGYVRYVYVNESGKKPMYLYNVFSKAELKFSM LEKLLTMMHKALIKAMDLMAGQEILVYSPVSMTKIOKTPLPERKALPIRWITWMTYLEDPRQHFHYDKTLPELKHIPDVTYSSIPPLKHPQSQYEGVECTD GSAIKSPDPTKSNAGMIGVHAIYNPEYKILNQWSIPLGHHTAQMAEIAA VEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLK HISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLAATQGSYVVN
SFVCP_Q 87040_2m utA	SF VC P	Q8 704 0	der iva tiv e	MNPLQLIQLPLAEVKGTKLLAHWSGATITCIPESFLDEQPIKOTILKTIHGEKQONVYVLTFTKVKGRKVEAEVIASPYEYILLSPITDVPWL'TQOPLQLTIL LVPLQYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWEQVGHKIRPHINIAITGDYPPRQKQYPINPKAKPSIQIVIDDILLKQGVLTPQNSTMTP VYPKPDGRWRMVLIDYREVNKTPIPLTAQONQHSAGILATIVRQKYKTITLDLANGFWAHPITPDSYWLTAFTWQKGQYCWTRLPOGFLNSPALFNAD AVDLLKEVPNVQVYVDDIYL SHDNPHIEHQLEKVFQILLQAGYVYVSLKKSEIGQRTVEFLGFNITKEGRGLTDITFKTKLLNVTPPKDLKQLQSILGLLN FARNFIPNFAELVQTLNLIASSKGKYEWTEDNTKQNKVIEALNTASNEERLPDQRLVIKVNTPSAGYVRYVYVNESGKKPMYLYNVFSKAELKFSM LEKLLTMMHKALIKAMDLMAGQEILVYSPVSMTKIOKTPLPERKALPIRWITWMTYLEDPRQHFHYDKTLPELKHIPDVTYSSIPPLKHPQSQYEGVECTD GSAIKSPDPTKSNAGMIGVHAIYNPEYKILNQWSIPLGHHTAQMAEIAA VEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLK HISKWKSIAECLSIKPDITIQHEKGHPINTSIHTEGNALADKLAATQGSYVVN
SMRVH_ P03364	SM RV H	P03 364	ro ot	PRSRALDIPVPHADKISWKITDPVWQWPLTYEKTILAAIALVQEOALAGHIEPTNSPWNTPFIUKKSGSWRLLODLRAVNVKVMVPMGALQPLGSPV AIPLYHKIVIDLKDCFTIPLHEDRPFYAFVQINFQSPMPRYQWKVL PQGMANSTPLCQKFVAAAIAPVRSQWPEAYILHYVMDILLACDSAEAAK ACYAHHSCLTSYGLKIAPDKVQVSEPSYGLFELHHQVFTPRVCLTDHLKTLVDLQGLDQVLPRLYKLPSTALVPLNNILKDPNPLSVRALTPE AKQSLALINKAQNSVQOQSYNLPLVILLTPHTPTA VFWQPNCTDPTKNGSPLWLHL PASPSKVLLTYPSSL AMLIKGRYITGROLFRDPHSUIPYT QDQLTWWLQTSDEWALALSSFTGDIDNHYPSDPV IQF AKLHQHPKTKCAPQATLVITDGGSSNGIAAYVIDNQPSIKSPYLAQLVELYAILQVFTVL AHQPFNLVITDSAYTAQSVPLLETVPEHKSSSTNATPLFSKLQQLILNRQHPHFHIGHLRAHLNIPGLAEGNALADAAATQFIPIISD

SMRVH_	SM		der	iva	364	e	PRSRADIPVPHADKISWKITDPVVDQWPLTYEKTLLAALVQEQLAAGHIEPTNSPWNTPIFIKKSGSWRLLODLRAVNKVMVPMGALQPGLPSPV
P03364_2	RV		iva				AIPLNYHKIVIDLKDCFTTIPLHPEDRPYEAFSPQINQSPMPRYQWKVLPQGMANSPTLCQKFAAALAPVRSQWPEAYILHYMDILLACDSAEAAK
mut	H						ACVYAHISCLTSYGLKIADPKVQSEPEPSYLGELHHQOFTPRVCLKTIDHLKTLNDFQKLLGDIQWLRPYLKLPTALSALVPLNNILKPDNPISVRLTPE
							AKQSLALINKAIQNQSVQISYNLPLVILLPTHTPTAVFWQNGDTPDKNGSPJLWLHPASPSKVLLTYPSSLLAMLIJKGRYTGRLFGDRPHSHPYT
							QDQLTWLLQTSDEWAIALSSTFGDIDNHYPSDPVQIEAKLHQEPKTKCAPIQOATLVTDDGSSNGIAAYVIDNQPSIKSPYLQAQVELYAILQVETVL
							AHQPNLYTDSAYIAQSVPLLETVPKISSSTNATPLFSKLOQLLNQRHPFHGHLRAHLNLPGLAEGNALADAATQIPIISD
							PRSRADIPVPHADKISWKITDPVVDQWPLTYEKTLLAALVQEQLAAGHIEPTNSPWNTPIFIKKSGSWRLLODLRAVNKVMVPMGALQPGLPSPV
							APPLNYHKIVIDLKDCFTTIPLHPEDRPYEAFSPQINQSPMPRYQWKVLPQGMANSPTLCQKFAAALAPVRSQWPEAYILHYMDILLACDSAEAAK
							ACVYAHISCLTSYGLKIADPKVQSEPEPSYLGELHHQOFTPRVCLKTIDHLKTLNDFQKLLGDIQWLRPYLKLPTALSALVPLNNILKPDNPISVRLTPE
							AKQSLALINKAIQNQSVQISYNLPLVILLPTHTPTAVFWQNGDTPDKNGSPJLWLHPASPSKVLLTYPSSLLAMLIJKGRYTGRLFGDRPHSHPYT
							QDQLTWLLQTSDEWAIALSSTFGDIDNHYPSDPVQIEAKLHQEPKTKCAPIQOATLVTDDGSSNGIAAYVIDNQPSIKSPYLQAQVELYAILQVETVL
							AHQPNLYTDSAYIAQSVPLLETVPKISSSTNATPLFSKLOQLLNQRHPFHGHLRAHLNLPGLAEGNALADAATQIPIISD
							LTAAIDMLAPOQC AEPTWKSDPEVWVDQWPLTSEKLAQAQQLVQEQLEAGHIESNPWNTPIFIKKSGSWRLLODLRAVNATMVLGALQPGGL
							PSVVAIPQGYLKIIDLKDCFTTIPLHPEDRPYEAFSPQINQSPMPRYQWKVLPQGMANSPTLCQKFAAALAPVRSQWPEAYILHYMDILLACDSAEAAK
							QVLCQFDQLKQELTIAGLHIAPEKIQLODPYTYLGFELNGPKITNQKA VIRKDKLQTLNDFQKLLGDIQWLRPYLKLTTADLPKPLFDLTKGDSNPNSHR
							LSKEALALLDKVETAAEQVTHINYSPLMELJENTALPTTGLFWQNNPMVWHLPAKPKVLLPYDAIADLILGRDHSKKYFEGIEPSVIOQPSKQID
							WLMQNTMWPACASYVGLDNHYPNKLIQFCKLHAHPHQISKTPLNNALLVFTDGSSTGMAAYTLADTTIKFQTNLNSAQLVELQALIAVLSAFPNQ
							PLNIYTDSAYLAHSIPLLETVAQKHISETAKLFLQCOQLYNSRISPIYGHVRAHSGLPGIAHGQKQADLTKTVASNTT
							LTAAIDMLAPOQC AEPTWKSDPEVWVDQWPLTSEKLAQAQQLVQEQLEAGHIESNPWNTPIFIKKSGSWRLLODLRAVNATMVLGALQPGGL
							PSVVAIPQGYLKIIDLKDCFTTIPLHPEDRPYEAFSPQINQSPMPRYQWKVLPQGMANSPTLCQKFAAALAPVRSQWPEAYILHYMDILLACDSAEAAK
							QVLCQFDQLKQELTIAGLHIAPEKIQLODPYTYLGFELNGPKITNQKA VIRKDKLQTLNDFQKLLGDIQWLRPYLKLTTADLPKPLFDLTKGDSNPNSHR
							LSKEALALLDKVETAAEQVTHINYSPLMELJENTALPTTGLFWQNNPMVWHLPAKPKVLLPYDAIADLILGRDHSKKYFEGIEPSVIOQPSKQID
							DWLMQNTMWPACASYVGLDNHYPNKLIQFCKLHAHPHQISKTPLNNALLVFTDGSSTGMAAYTLADTTIKFQTNLNSAQLVELQALIAVLSAFPN
							QPLNIYTDSAYLAHSIPLLETVAQKHISETAKLFLQCOQLYNSRISPIYGHVRAHSGLPGIAHGQKQADLTKTVASNTT
							LATAVDILAPQRYADPTWKSDPEVWVDQWPLTSEKLAQAQQLVQEQLEAGHIESNPWNTPIFIKKSGSWRLLODLRAVNATMVLGALQPGGL
							PSVVAIPQGYLKIIDLKDCFTTIPLHPEDRPYEAFSPQINQSPMPRYQWKVLPQGMANSPTLCQKFAAALAPVRSQWPEAYILHYMDILLACDSAEAAK
							EQVLCQFDQLKQELTIAGLHIAPEKIQLODPYTYLGFELNGPKITNQKA VIRKDKLQTLNDFQKLLGDIQWLRPYLKLTTADLPKPLFDLTKGDSNPNSHR
							SLSEALALLDKVETAAEQVTHINYSPLMELJENTALPTTGLFWQNNPMVWHLPAKPKVLLPYDAIADLILGRDHSKKYFEGIEPSVIOQPSKQID
							SLSEALALLDKVETAAEQVTHINYSPLMELJENTALPTTGLFWQNNPMVWHLPAKPKVLLPYDAIADLILGRDHSKKYFEGIEPSVIOQPSKQID
							HWLMQNTMWPACASYVGLDNHYPNKLIQFCKLHAHPHQISKTPLNNALLVFTDGSSTGMAAYTLADTTIKFQTNLNSAQLVELQALIAVLSAFPN
							RALNVYTDSAYLAHSIPLLETVAQKHISETAKLFLQCOQLYNSRISPIYGHVRAHSGLPGIAHGQKQADLTKTVASNTT
							LATAVDILAPQRYADPTWKSDPEVWVDQWPLTSEKLAQAQQLVQEQLEAGHIESNPWNTPIFIKKSGSWRLLODLRAVNATMVLGALQPGGL
							PSVVAIPQGYLKIIDLKDCFTTIPLHPEDRPYEAFSPQINQSPMPRYQWKVLPQGMANSPTLCQKFAAALAPVRSQWPEAYILHYMDILLACDSAEAAK
							EQVLCQFDQLKQELTIAGLHIAPEKIQLODPYTYLGFELNGPKITNQKA VIRKDKLQTLNDFQKLLGDIQWLRPYLKLTTADLPKPLFDLTKGDSNPNSHR
							SLSEALALLDKVETAAEQVTHINYSPLMELJENTALPTTGLFWQNNPMVWHLPAKPKVLLPYDAIADLILGRDHSKKYFEGIEPSVIOQPSKQID
							HWLMQNTMWPACASYVGLDNHYPNKLIQFCKLHAHPHQISKTPLNNALLVFTDGSSTGMAAYTLADTTIKFQTNLNSAQLVELQALIAVLSAFPN
							RALNVYTDSAYLAHSIPLLETVAQKHISETAKLFLQCOQLYNSRISPIYGHVRAHSGLPGIAHGQKQADLTKTVASNTT
							SCQTKNTLNIDEYLLQFPDQLWASLPTDGRMLVPPITIKIDNASLPSIRQYPLPKDKTEGLRPLISSLENQGLIKCHSPCNTPIPIPKKAGRDEYRMIHDL
							RAINNVAPLTAVASPTTVLSNLAPSLHFWTVIDLSNAFSPVPHKDSQYLFAFTEGHQYTWTVLPQGFHSPTLSQALYQSLHKIKFKISSEICYMDD
							VLIASKDRDITNLKDTAVMLQHLASEGHKVSCKKLQLOCOQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTIRQIRAFGLVGYCRHWIPEFSHSHKLE
							KQLKDDTAEPFQLDQOQVEAFENKLHAIATAPVLVVPDPKPFQLYTSHSEHAIAPLQKHAGRTPIAFVLSKFDASGLPPCLKACASHRSLSLQAD
							SFILGAPLIYTHAICTLLQDRSQVLTASRSKWEADLLRPELTVACSAVSPAHLYMQSCENNPDPHDCVLLTHITSRPRPDLSDLPIPPDPMITLSDGS
							YTTGRGGAADVMMHPRVTDFFIHHQPPGASQAQTEALLALAAACHLADTKTVNYTDSRYAYGVVHDFGHLWMHRRGFTVSAAGTPIKNIHKEIEYLLKQI
							MPKQVSVIKIEAHTKGVSMVRGNAADAEAKNAHVLVQR
							SCQTKNTLNIDEYLLQFPDQLWASLPTDGRMLVPPITIKIDNASLPSIRQYPLPKDKTEGLRPLISSLENQGLIKCHSPCNTPIPIPKKAGRDEYRMIHDL
							RAINNVAPLTAVASPTTVLSNLAPSLHFWTVIDLSNAFSPVPHKDSQYLFAFTEGHQYTWTVLPQGFHSPTLSQALYQSLHKIKFKISSEICYMDD
							VLIASKDRDITNLKDTAVMLQHLASEGHKVSCKKLQLOCOQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTIRQIRAFGLVGYCRHWIPEFSHSHKLE
							KQLKDDTAEPFQLDQOQVEAFENKLHAIATAPVLVVPDPKPFQLYTSHSEHAIAPLQKHAGRTPIAFVLSKFDASGLPPCLKACASHRSLSLQAD
							SFILGAPLIYTHAICTLLQDRSQVLTASRSKWEADLLRPELTVACSAVSPAHLYMQSCENNPDPHDCVLLTHITSRPRPDLSDLPIPPDPMITLSDGS
							YTTGRGGAADVMMHPRVTDFFIHHQPPGASQAQTEALLALAAACHLADTKTVNYTDSRYAYGVVHDFGHLWMHRRGFTVSAAGTPIKNIHKEIEYLLKQI
							MPKQVSVIKIEAHTKGVSMVRGNAADAEAKNAHVLVQR

WDSV_O 92815_2m utA	W DS V	O9 281 5	der iva tiv e	SCQTKNTLNIDEYLLQFPDQLWASLPTDGRMLVPPITIKIKDNASIPSRQYPLPKDKTEGLRPLISSLENOGILJKHSPCNTPIPIKAKGRDEYRMHDL RAINNVAPLTAVASPTTVLSNLAPSLHFWFTVIDLSNAFSPYHIKDSQYLAFTTEHQYTWTVLPGQTHSPLEFQALYQSLHKKFKISSEICICYMDD VLASKDRDINLKDIAVMLQHLASEGHKYSKKLQLOCOQEVVYLQQLLTPEGRKILPDRKVTVSQFOQPTTIRQIRAFGLKGYCRHEFEFSHSKLE KQLKPDIAEPYLLDQDQVEAFNKLHATITAPVLVVPDPAKPTQLYTSHSEHSAIVLTKHAGRTPIAHLSKSDAIESGLPCLKACASHRSLLTQAD SFLGAPLIYTTTHAICTLLQRDSQLTAFSFKWEADLLRPELTFVACSAVSPAHILYMQSCENNPIDHCVLLTHITSRPRDLSLPIPDMDTLFSDGS YTTGRGGAADVVMHRPVTDFFHQQPGGASQTAELALAAACHLATDKTNTVNTDSRYAYGVVHDFGHLWMHRGFVTSAGTIKNIKEIEYLLKQI MRPKQVSVIKIEAHTKGVSMVEVGNAADEAAVNLVQR
WMSV_P 03359	W MS V	P03 359	ro ot	VLNLEERYLHEKVPSSIDPSWILQFPFTVAERAGMGLANQVPPVVELRSGASPVAVROYPMKSKEAREGIRPHIQRIHLDLGVLPQSPWNTPLLPV KKPGTNDYRPVQDRLREINRQVQDIHPTVPNPYNLLSSLPSTWYSVLDLKDADFCLKLHNSQBLFAFEWRDPEKNGTQGLTWTRLPQGFKNSPILFD EALHRLAPRALNPQVLLQYVDDLLVAAPTYRDCKEGTQKLQELSKLGYRVSAKKAQQLCOKEVTVLGYLLKEGRWLTPARKATVMKPPPTTP RQVREFLGTAGFCRLWIPGFASLAPLYPLTKPSIPFIWTEEHQKAFRIKELLSAPALAPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSK KLDPVAGSWPTCLKAAVAALLKDADKLTLGQNVTVIASHSLESIVROPDRWMTNARMTHYQSLLNERSVAPPVNLNPAITLLPVESEATPVHRC SELAETGTTRDLKDQPLPGVPAWYTDGSSFAIEGKRRAAGAAVDRKRTVWASSLPEGTSQAQKAEALVALTQALRLAEGKDINVTDSRYAFATAHHHG AIYKQRGLLTSAGDKIKNEEILALLEAHLPKRVAIIHCPCGHQKGNPDVATGNRRADAQAALSTRVLAETTKP
WMSV_P 03359_3m ut	W MS V	P03 359	der iva tiv e	VLNLEERYLHEKVPSSIDPSWILQFPFTVAERAGMGLANQVPPVVELRSGASPVAVROYPMKSKEAREGIRPHIQRIHLDLGVLPQSPWNTPLLPV KKPGTNDYRPVQDRLREINRQVQDIHPTVPNPYNLLSSLPSTWYSVLDLKDADFCLKLHNSQBLFAFEWRDPEKNGTQGLTWTRLPQGFKNSPILFD EALHRLAPRALNPQVLLQYVDDLLVAAPTYRDCKEGTQKLQELSKLGYRVSAKKAQQLCOKEVTVLGYLLKEGRWLTPARKATVMKPPPTTP RQVREFLGTAGFCRLWIPGFASLAPLYPLTKPSIPFIWTEEHQKAFRIKELLSAPALAPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSK KLDPVAGSWPTCLKAAVAALLKDADKLTLGQNVTVIASHSLESIVROPDRWMTNARMTHYQSLLNERSVAPPVNLNPAITLLPVESEATPVHRC SELAETGTTRDLKDQPLPGVPAWYTDGSSFAIEGKRRAAGAAVDRKRTVWASSLPEGTSQAQKAEALVALTQALRLAEGKDINVTDSRYAFATAHHHG AIYKQRGWLTSAGDKIKNEEILALLEAHLPKRVAIIHCPCGHQKGNPDVATGNRRADAQAALSTRVLAETTKP
WMSV_P 03359_3m utA	W MS V	P03 359	der iva tiv e	VLNLEERYLHEKVPSSIDPSWILQFPFTVAERAGMGLANQVPPVVELRSGASPVAVROYPMKSKEAREGIRPHIQRIHLDLGVLPQSPWNTPLLPV KKPGTNDYRPVQDRLREINRQVQDIHPTVPNPYNLLSSLPSTWYSVLDLKDADFCLKLHNSQBLFAFEWRDPEKNGTQGLTWTRLPQGFKNSPILFD EALHRLAPRALNPQVLLQYVDDLLVAAPTYRDCKEGTQKLQELSKLGYRVSAKKAQQLCOKEVTVLGYLLKEGRWLTPARKATVMKPPPTTP RQVREFLGTAGFCRLWIPGFASLAPLYPLTKPSIPFIWTEEHQKAFRIKELLSAPALAPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSK KLDPVAGSWPTCLKAAVAALLKDADKLTLGQNVTVIASHSLESIVROPDRWMTNARMTHYQSLLNERSVAPPVNLNPAITLLPVESEATPVHRC SELAETGTTRDLKDQPLPGVPAWYTDGSSFAIEGKRRAAGAAVDRKRTVWASSLPEGTSQAQKAEALVALTQALRLAEGKDINVTDSRYAFATAHHHG AIYKQRGWLTSAGDKIKNEEILALLEAHLPKRVAIIHCPCGHQKGNPDVATGNRRADAQAALSTRVLAETTKP
XMRV6_ A1Z651	XM RV 6	A1 Z6 51	ro ot	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPKATSTPVSIIKQYPMQSEARLGKPHIQRIHLDQGLVPCQSPWNTPLLPV KKPGTNDYRPVQDRLREINRQVQDIHPTVPNPYNLLSGLPSSHQWYTVLVDLKDADFCLRLHPTSQPLFAFEWRDPEMISGQLTWTRLPQGFKNSPILFD EALHRLADFRIQHPIHLLIQQYVDDLLAATSEQDCQGRTRALLQTLGNLGYRASAKKAQICQKQVYLYLLKEGORWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFASLAPLYPLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLS KLDPVAGSWPTCLKAAVAALLKDADKLTLGQNVTVIASHSLESIVROPDRWMTNARMTHYQSLLNERSVAPPVNLNPAITLLPVESEATPVHRC SELAETGTTRDLKDQPLPGVPAWYTDGSSFAIEGKRRAAGAAVDRKRTVWASSLPEGTSQAQKAEALVALTQALRLAEGKDINVTDSRYAFATAHHHG AIYKQRGWLTSAGDKIKNEEILALLEAHLPKRVAIIHCPCGHQKGNPDVATGNRRADAQAALSTRVLAETTKP
XMRV6_ A1Z651_3 mut	XM RV 6	A1 Z6 51	der iva tiv e	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPKATSTPVSIIKQYPMQSEARLGKPHIQRIHLDQGLVPCQSPWNTPLLPV KKPGTNDYRPVQDRLREINRQVQDIHPTVPNPYNLLSGLPSSHQWYTVLVDLKDADFCLRLHPTSQPLFAFEWRDPEMISGQLTWTRLPQGFKNSPILFD EALHRLADFRIQHPIHLLIQQYVDDLLAATSEQDCQGRTRALLQTLGNLGYRASAKKAQICQKQVYLYLLKEGORWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFASLAPLYPLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLS KLDPVAGSWPTCLKAAVAALLKDADKLTLGQNVTVIASHSLESIVROPDRWMTNARMTHYQSLLNERSVAPPVNLNPAITLLPVESEATPVHRC SELAETGTTRDLKDQPLPGVPAWYTDGSSFAIEGKRRAAGAAVDRKRTVWASSLPEGTSQAQKAEALVALTQALRLAEGKDINVTDSRYAFATAHHHG AIYKQRGWLTSAGDKIKNEEILALLEAHLPKRVAIIHCPCGHQKGNPDVATGNRRADAQAALSTRVLAETTKP
XMRV6_ A1Z651_3 mutA	XM RV 6	A1 Z6 51	der iva tiv e	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPKATSTPVSIIKQYPMQSEARLGKPHIQRIHLDQGLVPCQSPWNTPLLPV KKPGTNDYRPVQDRLREINRQVQDIHPTVPNPYNLLSGLPSSHQWYTVLVDLKDADFCLRLHPTSQPLFAFEWRDPEMISGQLTWTRLPQGFKNSPILFD EALHRLADFRIQHPIHLLIQQYVDDLLAATSEQDCQGRTRALLQTLGNLGYRASAKKAQICQKQVYLYLLKEGORWLTEARKETVMGQPTKTPR QLREFLGTAGFCRLWIPGFASLAPLYPLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLS KLDPVAGSWPTCLKAAVAALLKDADKLTLGQNVTVIASHSLESIVROPDRWMTNARMTHYQSLLNERSVAPPVNLNPAITLLPVESEATPVHRC SELAETGTTRDLKDQPLPGVPAWYTDGSSFAIEGKRRAAGAAVDRKRTVWASSLPEGTSQAQKAEALVALTQALRLAEGKDINVTDSRYAFATAHHHG AIYKQRGWLTSAGDKIKNEEILALLEAHLPKRVAIIHCPCGHQKGNPDVATGNRRADAQAALSTRVLAETTKP

In some embodiments, reverse transcriptase domains are modified, for example by site-specific mutation. In some embodiments, reverse transcriptase domains are engineered to have improved properties, e.g. SuperScript IV (SSIV) reverse transcriptase derived from the MMLV RT. In some embodiments, the reverse transcriptase domain may be engineered to have lower error rates, e.g., as described in WO2001068895, incorporated herein by reference. In some embodiments, the reverse transcriptase domain may be engineered to be more thermostable. In some embodiments, the reverse transcriptase domain may be engineered to be more processive.

10 In some embodiments, the reverse transcriptase domain may be engineered to have tolerance to inhibitors. In some embodiments, the reverse transcriptase domain may be engineered to be faster. In some embodiments, the reverse transcriptase domain may be engineered to better tolerate modified nucleotides in the RNA template. In some embodiments, the reverse transcriptase domain may be engineered to insert modified DNA nucleotides. In some

15 embodiments, the reverse transcriptase domain is engineered to bind a template RNA. In some embodiments, one or more mutations are chosen from D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K, or D653N in the RT domain of murine leukemia virus reverse transcriptase or a corresponding mutation at a corresponding position of another RT domain. In some

20 embodiments, one or more mutations are chosen as described in WO2018089860A1, incorporated herein by reference (e.g., a C952S, and/or C956S, and/or C952S, C956S (double mutant), and/or C969S, and/or H970Y, and/or R979Q, and/or R976Q, and/or R1071S, and/or R328A, and/or R329A, and/or Q336A, and/or R328A, R329A, Q336A (triple mutant), and/or G426A, and/or D428A, and/or G426A,D428A (double mutant) mutation, and/or any

25 combination thereof; positions relative to WO2018089860A1 SEQ ID NO: 52), in the RT domain of R2Bm retrotransposase or a corresponding mutation at a corresponding position of another RT domain.

In some embodiments, the RT domain possesses proofreading activity. In some embodiments, the RT domain is evolved from a DNA-dependent DNA polymerase and has

30 gained RNA-dependent DNA polymerase activity. The synthetic evolved proofreading RT

known as reverse transcription xenopolymerase (RTX, Genbank: QFN49000.1) was previously generated by taking a DNA-dependent DNA polymerase (KOD, Genbank: ABN15964.1) and selecting for RNA-dependent DNA polymerase activity (Ellefson et al 2016). In some embodiments, the engineered RT may comprise DNA-dependent DNA polymerase signatures as the result of the wild-type enzyme, e.g., IPR006134, PF00136, cd05536.

In some embodiments, the reverse transcription domain only recognizes and reverse transcribes a specific template. In some embodiments, the template comprises of specific sequences. In some embodiments, the template comprises inclusion of a UTR that associates the nucleic acid with the reverse transcriptase domain (e.g. an untranslated region (UTR) from a retrotransposon, e.g. the 3' UTR of an R2 retrotransposon).

The writing domain may also comprise DNA-dependent DNA polymerase activity, e.g., comprise enzymatic activity capable of writing DNA into the genome from a template DNA sequence. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a DNA polymerase domain in the polypeptide. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a reverse transcriptase domain that is also capable of DNA-dependent DNA polymerization, e.g., second strand synthesis.

In some embodiments, a writing domain (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 of the writing domain.

In contrast to other types of reverse transcription machines, e.g., retroviral RTs and LTR retrotransposons, reverse transcription in non-LTR retrotransposons like R2 is performed only on RNA templates containing specific recognition sequences. The R2 retrotransposase requires its template to contain a minimal 3' UTR region in order to initiate TPRT (Luan and Eickbush Mol Cell Biol 15, 3882-91 (1995)). In some embodiments, the Gene Writer polypeptide is derived from a retrotransposase with a required binding motif and the template RNA is designed to contain said binding motif, such that there is specific retrotransposition of only the desired template. In some embodiments, the Gene Writer polypeptide is derived from a retrotransposon selected from Table 3 and the 3' UTR on the RNA template comprises the 3' UTR from the same retrotransposon in Table 3.

Template nucleic acid binding domain:

The 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
5 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) is contained within the reverse transcription domain, e.g., the reverse
10 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)
15 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
20 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
25 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
30 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.

10

Endonuclease domain:

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 I-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%

30

identical to the amino acid sequence of an endonuclease domain of a retrotransposon whose DNA sequence is referenced in Table 1 or 3. A person having ordinary skill in the art is capable of identifying endonuclease domains based upon homology to other known endonuclease domains using tools such as Basic Local Alignment Search Tool (BLAST). In certain
5 embodiments, the heterologous endonuclease is FokI 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*—SsoI 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,
10 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
15 and an endonuclease domain that comprises FokI 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
20 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 endonuclease does not form
25 detectable levels of double stranded breaks.

In some embodiments, the endonuclease domain has nickase activity that nicks the target
site DNA of the first 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
30 nicks the target site DNA of the first strand and does not nick the target site DNA of the second
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). As a further example, when a polypeptide comprises a CRISPR-associated endonuclease domain
5 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 not containing the PAM site (e.g., and does not nick the target site DNA strand that contains the PAM site).

In some other embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the first strand and the second strand. Without wishing to be bound by theory,
10 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 first DNA strand. The target site DNA now contains two different sequences for the first DNA strand: one
15 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 second 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, it is thought that introducing an additional nick to the second strand may bias the cellular DNA repair machinery to adopt the heterologous object sequence-
20 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 first strand.

Alternatively or additionally, without wishing to be bound by theory, it is thought that an
25 additional nick to the second strand may promote second strand synthesis. In some embodiments, where the Gene Writer™ has inserted or substituted a portion of the first strand, synthesis of a new sequence corresponding to the insertion/substitution in the second 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 first strand and the
30 second 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 first strand and an additional gRNA that directs nicking of the second strand. In some embodiments, the polypeptide comprises a plurality of domains having endonuclease activity, and a first endonuclease domain nicks the first strand and a second endonuclease domain nicks the second strand (optionally, the first endonuclease domain does not (e.g., cannot) nick the second strand and the second endonuclease domain does not (e.g., cannot) nick the first 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, 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 some 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 some embodiments, the endonuclease domain comprises a FokI domain.

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 (SEQ ID NO: 1577), 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 (SEQ ID NO: 1577) meganucleases with a single copy of the LAGLIDADG motif (SEQ ID NO: 1577) generally form homodimers, whereas members with two copies of the LAGLIDADG motif (SEQ ID NO: 1577) 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
5 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)).

10 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 comprises a CRISPR/Cas domain (also referred to herein as a CRISPR-associated protein). In some embodiments, a DNA-binding
15 domain comprises a CRISPR/Cas domain. In some embodiments, a CRISPR/Cas domain comprises a protein involved in the clustered regulatory interspaced short palindromic repeat (CRISPR) system, e.g., a Cas protein, and optionally binds a guide RNA, e.g., single guide RNA (sgRNA).

CRISPR systems are adaptive defense systems originally discovered in bacteria and
20 archaea. CRISPR systems use RNA-guided nucleases termed CRISPR-associated or “Cas” endonucleases (e. g., Cas9 or Cpf1) to cleave foreign DNA. For example, in a typical CRISPR/Cas system, an endonuclease is directed to a target nucleotide sequence (e. g., a site in the genome that is to be sequence-edited) by sequence-specific, non-coding “guide RNAs” that target single- or double-stranded DNA sequences. Three classes (I-III) of CRISPR systems have
25 been identified. The class II CRISPR systems use a single Cas endonuclease (rather than multiple Cas proteins). One class II CRISPR system includes a type II Cas endonuclease such as Cas9, a CRISPR RNA (“crRNA”), and a trans-activating crRNA (“tracrRNA”). The crRNA contains a “guide RNA”, typically about 20-nucleotide RNA sequence that corresponds to a target DNA sequence. In the wild-type system, and in some engineered systems, crRNA also
30 contains a region that binds to the tracrRNA to form a partially double-stranded structure which is cleaved by RNase III, resulting in a crRNA/tracrRNA hybrid. A crRNA/tracrRNA hybrid then

directs Cas9 endonuclease to recognize and cleave a target DNA sequence. A target DNA sequence is generally adjacent to a “protospacer adjacent motif” (“PAM”) that is specific for a given Cas endonuclease; however, PAM sequences appear throughout a given genome. CRISPR endonucleases identified from various prokaryotic species have unique PAM sequence requirements; examples of PAM sequences include 5'-NGG (*Streptococcus pyogenes*), 5'-NNAGAA (*Streptococcus thermophilus* CRISPR1), 5'-NGGNG (*Streptococcus thermophilus* CRISPR3), and 5'-NNNGATT (*Neisseria meningitidis*). Some endonucleases, e.g., Cas9 endonucleases, are associated with G-rich PAM sites, e. g., 5'-NGG, and perform blunt-end cleaving of the target DNA at a location 3 nucleotides upstream from (5' from) the PAM site.

Another class II CRISPR system includes the type V endonuclease Cpf1, which is smaller than Cas9; examples include AsCpf1 (from *Acidaminococcus* sp.) and LbCpf1 (from *Lachnospiraceae* sp.). Cpf1-associated CRISPR arrays are processed into mature crRNAs without the requirement of a tracrRNA; in other words, a Cpf1 system, in some embodiments, comprises only Cpf1 nuclease and a crRNA to cleave a target DNA sequence. Cpf1 endonucleases, are typically associated with T-rich PAM sites, e. g., 5'-TTN. Cpf1 can also recognize a 5'-CTA PAM motif. Cpf1 typically cleaves a target DNA by introducing an offset or staggered double-strand break with a 4- or 5-nucleotide 5' overhang, for example, cleaving a target DNA with a 5-nucleotide offset or staggered cut located 18 nucleotides downstream from (3' from) from a PAM site on the coding strand and 23 nucleotides downstream from the PAM site on the complimentary strand; the 5-nucleotide overhang that results from such offset cleavage allows more precise genome editing by DNA insertion by homologous recombination than by insertion at blunt-end cleaved DNA. See, e.g., Zetsche et al. (2015) *Cell*, 163:759 – 771.

A variety of CRISPR associated (Cas) genes or proteins can be used in the technologies provided by the present disclosure and the choice of Cas protein will depend upon the particular conditions of the method. Specific examples of Cas proteins include class II systems including Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cpf1, C2C1, or C2C3. In some embodiments, a Cas protein, e.g., a Cas9 protein, may be from any of a variety of prokaryotic species. In some embodiments a particular Cas protein, e.g., a particular Cas9 protein, is selected to recognize a particular protospacer-adjacent motif (PAM) sequence. In some embodiments, a DNA-binding domain or endonuclease domain includes a sequence targeting polypeptide, such as a Cas protein, e.g., Cas9. In certain embodiments a Cas protein, e.g., a Cas9 protein, may be

obtained from a bacteria or archaea or synthesized using known methods. In certain embodiments, a Cas protein may be from a gram positive bacteria or a gram negative bacteria. In certain embodiments, a Cas protein may be from a Streptococcus (e.g., a *S. pyogenes*, or a *S. thermophilus*), a Francisella (e.g., an *F. novicida*), a Staphylococcus (e.g., an *S. aureus*), an
 5 Acidaminococcus (e.g., an Acidaminococcus sp. BV3L6), a Neisseria (e.g., an *N. meningitidis*), a Cryptococcus, a Corynebacterium, a Haemophilus, a Eubacterium, a Pasteurella, a Prevotella, a Veillonella, or a Marinobacter.

In some embodiments, a Cas protein requires a protospacer adjacent motif (PAM) to be present in or adjacent to a target DNA sequence for the Cas protein to bind and/or function. In
 10 some embodiments, the PAM is or comprises, from 5' to 3', NGG, YG, NNGRRT, NNNRRT, NGA, TYCV, TATV, NTTN, or NNNGATT, where N stands for any nucleotide, Y stands for C or T, R stands for A or G, and V stands for A or C or G. In some embodiments, a Cas protein is a protein listed in Table 4. In some embodiments, a Cas protein comprises one or more mutations altering its PAM. In some embodiments, a Cas protein comprises E1369R, E1449H, and
 15 R1556A mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises E782K, N968K, and R1015H mutations or analogous substitutions to the amino acids corresponding to said positions. In some
 20 embodiments, a Cas protein comprises D1135V, R1335Q, and T1337R mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises S542R and K607R mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises S542R, K548V, and N552R mutations or analogous substitutions to the amino acids corresponding to said positions.

25 *Table 4 CRISPR/Cas Proteins, Species, and Mutations*

Name	Enzyme	Species	# of AAs	PAM	Mutations to alter PAM recognition	Mutations to make catalytically dead
<i>FnCas9</i>	Cas9	<i>Francisella novicida</i>	1629	5'-NGG-3'	Wt	D11A/H969A/N995A
<i>FnCas9 RHA</i>	Cas9	<i>Francisella novicida</i>	1629	5'-YG-3'	E1369R/E1449H/R1556A	D11A/H969A/N995A

<i>SaCas</i> 9	Cas9	<i>Staphylococcus aureus</i>	1053	5'- NNGRR T-3'	Wt	D10A/H557A
<i>SaCas</i> 9 KKH	Cas9	<i>Staphylococcus aureus</i>	1053	5'- NNNRR T-3'	E782K/N968K/R10 15H	D10A/H557A
<i>SpCas</i> 9	Cas9	<i>Streptococcus pyogenes</i>	1368	5'-NGG- 3'	Wt	D10A/D839A/H840A/ N863A
<i>SpCas</i> 9 VQR	Cas9	<i>Streptococcus pyogenes</i>	1368	5'-NGA- 3'	D1135V/R1335Q/T 1337R	D10A/D839A/H840A/ N863A
<i>AsCpf</i> I RR	Cpf1	<i>Acidaminococcus</i> sp. BV3L6	1307	5'-TYCV- 3'	S542R/K607R	E993A
<i>AsCpf</i> I RVR	Cpf1	<i>Acidaminococcus</i> sp. BV3L6	1307	5'-TATV- 3'	S542R/K548V/N552 R	E993A
<i>FnCpf</i> I	Cpf1	<i>Francisella novicida</i>	1300	5'-NTTN- 3'	Wt	D917A/E1006A/D1255 A
<i>NmCas</i> 9	Cas9	<i>Neisseria meningitidis</i>	1082	5'- NNNGA TT-3'	Wt	D16A/D587A/H588A/ N611A

In some embodiments, the Cas protein is catalytically active and cuts one or both strands of the target DNA site. In some embodiments, cutting the target DNA site is followed by formation of an alteration, e.g., an insertion or deletion, e.g., by the cellular repair machinery.

- 5 In some embodiments, the Cas protein is modified to deactivate or partially deactivate the nuclease, e.g., nuclease-deficient Cas9. Whereas wild-type Cas9 generates double-strand breaks (DSBs) at specific DNA sequences targeted by a gRNA, a number of CRISPR endonucleases having modified functionalities are available, for example: a “nickase” version of Cas9 that has been partially deactivated generates only a single-strand break; a catalytically inactive Cas9
- 10 (“dCas9”) does not cut target DNA. In some embodiments, dCas9 binding to a DNA sequence may interfere with transcription at that site by steric hindrance. In some embodiments, dCas9 binding to an anchor sequence may interfere with (e.g., decrease or prevent) genomic complex

(e.g., ASMC) formation and/or maintenance. In some embodiments, a DNA-binding domain comprises a catalytically inactive Cas9, e.g., dCas9. Many catalytically inactive Cas9 proteins are known in the art. In some embodiments, dCas9 comprises mutations in each endonuclease domain of the Cas protein, e.g., D10A and H840A or N863A mutations. In some embodiments, a
5 catalytically inactive or partially inactive CRISPR/Cas domain comprises a Cas protein comprising one or more mutations, e.g., one or more of the mutations listed in Table 4. In some embodiments, a Cas protein described on a given row of Table 4 comprises one, two, three, or all of the mutations listed in the same row of Table 4. In some embodiments, a Cas protein, e.g., not described in Table 4, comprises one, two, three, or all of the mutations listed in a row of Table 4
10 or a corresponding mutation at a corresponding site in that Cas protein.

In some embodiments, a catalytically inactive, e.g., dCas9, or partially deactivated Cas9 protein comprises a D11 mutation (e.g., D11A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H969 mutation (e.g.,
15 H969A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a N995 mutation (e.g., N995A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises mutations at one, two, or three of positions D11, H969, and
20 N995 (e.g., D11A, H969A, and N995A mutations) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D10 mutation (e.g., a D10A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a
25 catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H557 mutation (e.g., a H557A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D10 mutation (e.g., a D10A mutation) and a H557 mutation (e.g., a H557A mutation) or analogous substitutions to the amino acids corresponding to said positions.

30 In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D839 mutation (e.g., a D839A mutation) or an analogous

substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H840 mutation (e.g., a H840A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g.,
5 dCas9, or partially deactivated Cas9 protein comprises a N863 mutation (e.g., a N863A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D10 mutation (e.g., D10A), a D839 mutation (e.g., D839A), a H840 mutation (e.g., H840A), and a N863 mutation (e.g., N863A) or analogous substitutions to the amino acids corresponding to said positions.

10 In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a E993 mutation (e.g., a E993A mutation) or an analogous substitution to the amino acid corresponding to said position.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a a D917 mutation (e.g., a D917A mutation) or an analogous
15 substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a a E1006 mutation (e.g., a E1006A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D1255 mutation (e.g., a D1255A
20 mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D917 mutation (e.g., D917A), a E1006 mutation (e.g., E1006A), and a D1255 mutation (e.g., D1255A) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially
25 deactivated Cas9 protein comprises a a D16 mutation (e.g., a D16A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D587 mutation (e.g., a D587A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g.,
30 dCas9, or partially deactivated Cas9 protein comprises a H588 mutation (e.g., a H588A mutation) or an analogous substitution to the amino acid corresponding to said position. In some

embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a N611 mutation (e.g., a N611A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D16 mutation (e.g., D16A), a D587 mutation (e.g., D587A), a
 5 H588 mutation (e.g., H588A), and a N611 mutation (e.g., N611A) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a DNA-binding domain or endonuclease domain may comprise a Cas molecule comprising or linked (e.g., covalently) to a gRNA (e.g., a template nucleic acid, e.g., template RNA, comprising a gRNA).

10 In some embodiments, an endonuclease domain or DNA binding domain 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
 15 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,
 20 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
 25 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, a Gene Writer may comprise a Cas protein as listed in Table 40A. The predicted or validated nickase mutations for installing Nickase activity in the Cas protein as
 30 shown in Table 40A, are based on the signature of the SpCas9(N863A) mutation. In some embodiments, system described herein comprises a GeneWriter protein of Table 3 and a Cas

protein of Table 40A. In some embodiments, a protein or domain of Table 3, 41, or 44 is fused to a Cas protein of Table 40A.

Table 40A. CRISPR/Cas Proteins, Species, and Mutations

Variant	Parental Host	Protein Sequence	Nickase Mutation
Nme2Cas9	<i>Neisseria meningitidis</i>	MAAFKPNPYNILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRFERAEPKTDGSLAMARRLAR SVRRLTRRRRAHRLRLARRLLKREGVLQAADFENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVL LHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGHIRNQR GDYSHTFSRKDLQAEILLFEKQKEFGNPHVSGGLKEGIEITLLMTQRPALSGDAVQKMLGHCTFE PAEPKAAKNTYTAEFIWLTKLNNLRILEQGSERPLTDTERTALMDEPYRKSKLTAYAQARKLLGL EDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIGTAFSLFKT DEDITGRDKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGGKRYDEACAIEYGDHYGKKN TEEKIYLPPIPADEIRNPVLRALASQARKVINGVVRVYGSAPRIIETAREVVGSKFDRKEIEKRQEE NRKDREKAAAKFREYFPNFVGEPSKDILKLRLYEQQHGKCLYSYGKEINLVRLENEKGYVEIDHAL PFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQ KFDEDFGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFAFASNGQITNLLRGFWGLRKVR AEN DRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQWFEFAQE VMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSRPEAVHEYYTPLFVSRAPNRKMSGAHKDTLRS AKRFVKHNEKISVKRWLTTEIKLADLENMNVYKNGREIELYEALKARLEAYGGNAKQAFDPKDN PFYKKGQQLVKA VRVEKTQESGVLLNKKNAytiADNGDMVRVDVFCVDDKGGKNQYFVPIYA WQVAENILPDIDCKGYRIDDSYTFCSLHKYDLIAFQKDEKSKVEFAYYINCDSNNGRFLAWHD KGSKEQQFRISTQNLVLIQYQVNLGKEIRPCRLKKRPPVR	N611A
PpuCas9	<i>Pasteurella pneumotropica</i>	MQNNPLNYILGLDIGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVAKTGESLALSRRLARSSRR LKRRRAERLKKAKRLKAEKILHSIDEKLPINVWQLRVKGLKEKLERQEWAAVLLHLKSHRGYLS QRKNEGKSDNKELGALLSGIASNHQMLQSSEYRTPAELAVKKFQVEEGHIRNQGSYTHTFSRLLDL LAEMELLFQRQAELGNSYTSITLLENLTALLMWQKPALAGDAILKMLGKCTFEPSEYKAAKNSY SAERFVWLTKLNNLRILENGTERALNDNERFALLEQPYEKSCLTYAQVRAMLALSDNAIFKGVRY LGEDKKTVESKTTLEMKFYHQIRKTLGSALKEWNEKGNLSDLLDEIGTAFSLYKTDDEICRYL EGKLPERVLNALLNENLNFDFIQLSLKALHQLPLMLQGQRYDEAVSAIYGDHYGKKSTETTRLLP TIPADEIRNPVLRITLQARKVINAVVRVYGSAPRIIETAREVVGKSYQDRKKLEKQEDNRKQRE SAVKKFKEMFPHFVGEPSKDILKMLRYELQQAELCLYSYGKSLHLRLEKGYVEVDHALPFSRT WDDSFNNKVLVLANENQNKGNLTPYEWLDGKNNSEWQHFFVVRVQTSFGSYAKKQRLNHLKLD EKGFIERNLNDTRYVARFLCNFIADNMLLVGKGKRNVFASNGQITALLRHRWGLQKVRQNDRH HALDAVVVACSTVAMQQKITRFVRYNEGNVFSGERIDRETGEIPLHFPSPWAFKENVIRIFSEN PKLELENRLPDYQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSARKLNEGLSVLKVPLTQLKL SDLERMNVNRDEIALYESLKARLEQFGNDPAKAFAPFYKGGALVKA VRLEQTKSGVLRVDG NGVADNASMVRVDVFTKGKGYFLVPIYTQVAKGILPNRAATQGKDENDWDIMDEMATTFQFSL CQNDLKLVTKKKTIFGYFNGLNRAATSNINIKHEHDLKSKGKLGILEVGVKLALISLEKYQVDFELG KNIRPCRPRTKRQHVR	N605A
SauCas9	<i>Staphylococcus aureus</i>	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELGINPYEARVKGSLQKLSSEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYIEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYEFQIENVFKQKKKPTLTKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIENAEALLDQIAKILTYQSSEDIQEELTNLSELQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVKPKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIELAREKNSKDAQKMINEMQKRNQRTNERIEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNPNFYEVDDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFYQLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNNLRSYFRVN NLDVVKVSINGGFTSFLRRKWKFKKERNKGYKHAEADALIANADFIKWKLDKAKKVMENQ MFEEKQAESMPETEYEQYKEIFITPHQIKHIDFKDYKYSHRVDKKNRELINDTLYSTRKDDKGN TLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPDQTYQKLKLIMEQYGDENPLYKYEEETG NYLTKYSKKDNGPVIKIKIYGNKLNALHDITDDYPNSRNKVKLSLKPYPFDVYLDNGVYKFV TVKNLDVIKENYEVNSKYEEAKKLLKISNQAEFIASFYNNDLIKINGELYRVIGVNDLNLNRI EVNMIDITYREYLENMNDKRPPIRTASKTSQIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
SauCas9-KKH	<i>Staphylococcus aureus</i>	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR VKKLLFDYNLLTDHSELGINPYEARVKGSLQKLSSEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYIEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYEFQIENVFKQKKKPTLTKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIENAEALLDQIAKILTYQSSEDIQEELTNLSELQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVKPKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIELAREKNSKDAQKMINEMQKRNQRTNERIEIIRTTGKENAKYLIEKIKLHDMQEGK	N580A

		CLYSLEAIPLEDLLNPNFNYEVDHIIPRSVSFDNSFNKVLVKQEENSCKGNRTPFYQLSSSDSKISY ETFKKHILNLAKGGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLVVKVSINGGFTSFLRRKWKFKKERNKGKHHAEADALIANADFIKEWKLDKAKKVMENQ MFEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPQTYQKLLIMEQYGDENPLYKYEET GNYLTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLKLPYRFVDVYLDNGVYKF VTVKNLDVIKKENYEVNSKCYEEAKLLKISNQAEFASFYKNDLIKINGELYRVIGVNNDDLNR IEVNMIDITYREYLENMNDKRPPIHKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
SauriCas9	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEDSENNSNRRSKRGARRLKRRI HRLNRVKDLLADYQMDLNNVPKSTDPYTIRVKGLEPLTKEEFAIALHIAKRRGLHNISVSMGD EEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFTEDFVKEVKQLCETQRQ YHNIDDQFIQQYIDLVSTRREYFEGPGNGSPYGWDGDLKWKYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVTRDDNPKELEYEYKHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKS GKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEAKILTYQDEISIKKALDQPELLTESEKQIAQL TGYTGTHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKK VEMSEIDSIPITLVDFELSPVVKRAFI QSIKVINAVINRFGLPEDIIIEAREKNSKDRRKFNKLQKQNEATRKKIEQLLAKYGNNTAKYMIE KIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVVDHIIPRSVSFDNSLNKVLVKQSENSCKGNRTP YQYLSNESKISYNQFKQHILNLSKAKDRISKKRDMLLEERDINKFEVQKEFINRNLVDTRYATR ELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEADALVIANADFLFKTH KALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDDKPNRQLIN DTLYSTREIDGETYVVQTLKDLAKDNEKVKKLFTERPQKILMYQHPDKTEFKLMTILNQYAEAK NPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSKSRFDI YKCEQGYKMVSIGYLDVLKKNYYPKDKYAEKQKKKIKESDLFVGSFYNDLIMYEDELFR VIGVNSDINNVLNLMVDITYKDFCEVNNVTGEKRIKKTIGKRVVLEKYTTDILGNLYKTPLPKK QLIFKRGE	N588A
SauriCas9-KKH	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEDSENNSNRRSKRGARRLKRRI HRLNRVKDLLADYQMDLNNVPKSTDPYTIRVKGLEPLTKEEFAIALHIAKRRGLHNISVSMGD EEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFTEDFVKEVKQLCETQRQ YHNIDDQFIQQYIDLVSTRREYFEGPGNGSPYGWDGDLKWKYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVTRDDNPKELEYEYKHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKS GKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEAKILTYQDEISIKKALDQPELLTESEKQIAQL TGYTGTHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKK VEMSEIDSIPITLVDFELSPVVKRAFI QSIKVINAVINRFGLPEDIIIEAREKNSKDRRKFNKLQKQNEATRKKIEQLLAKYGNNTAKYMIE KIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVVDHIIPRSVSFDNSLNKVLVKQSENSCKGNRTP YQYLSNESKISYNQFKQHILNLSKAKDRISKKRDMLLEERDINKFEVQKEFINRNLVDTRYATR ELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEADALVIANADFLFKTH KALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDDKPNRQLIN DTLYSTREIDGETYVVQTLKDLAKDNEKVKKLFTERPQKILMYQHPDKTEFKLMTILNQYAEAK NPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSKSRFDI YKCEQGYKMVSIGYLDVLKKNYYPKDKYAEKQKKKIKESDLFVGSFYNDLIMYEDELFR VIGVNSDINNVLNLMVDITYKDFCEVNNVTGEKHIKKTIGKRVVLEKYTTDILGNLYKTPLPKK QLIFKRGE	N588A
ScaCas9-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPKSKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRRYTRRKNRIRYLQEIFANEMAKLDDSFQRLEESFLVEEDKKNERHPFGNLADEVAYHRN YPTIYHLRKKLADSPEKADRLIYLALAHIIKFRGHFLIEGKLNANSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAAKILSARLSKSRLEKLIIVFPNEKKNGLFNGNIALALGLTPNFKSNFDLADAKLQ SKDTYDDDLDELLGQIGDQYADLFSAAKNLSAAILSDILRSNSEVTKAPLSASMVKRYDEHHQD LALLKTLVRQPFPEKYAEIFKDDTKNGYAGYVADKKLRKRSGLATEEFYKFIKPILEKMDGA EELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPLKENREKIEKILTRIPYVGP LARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPHKSHLLYEYFT VYNELTKVKYVTERMRKPEFLSGEQKKAIVDILLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVED RFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTFEDREMIEERLKYAHLFDDKVMKQL KRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDFGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQG DSLHEQIADLAGSPAIIKKGILQTVKIVDELVKVMGHKPNIVNEMARENQTTKGLQSSREKRRKRI EEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYVDVHIVPQSFIKDD SIDNKVLRSVENRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKA GFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSCLVSDFRKDFQLYKVRDINN YHHAHDAYLNAVVGTAIIKKYPKLESEFYVDYKVDYVRKMKIAKSEQIGKATAKRRFFYSNIMN FFKTEVVKLANGEIRKPLIETNGETGEVWNKEKDFATVRKVLAMPQVNVKKTVEQTGGFSKES ILSKRESAKLIPRKGWDRKYGFGSPTVAYSILVVAKEVGKAKKLSVKVLVGITIMEKGSY EKDPIGFLEAKGYKDIKKELIFKLPKYSLEFELNGRRRMLASAKELQKANELVLVPHLVRLLYYTQ NISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNLSKSSFDEQFAVSDSILLSNSFVSL KYTSFGASGGFTFLDLVKQGRRLRYQTVEVLDTATLIYQSITGLYETRTDLSQLGGD	N872A
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPKSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRIRYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPFGNIVDEVAYHEKY PTIYHLRKKLVDSIDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVVKLFLQVLTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIALQLPGEKKNGLFNGNIALSLGLTPNFKSNFDLADAKLQ LSKDTYDDDLNLLAQIGDQYADLFSAAKNLSAAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTRIPYVGPVARGNSRFAW	N863A

		<p>MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIIEGKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSDINKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFSDPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS EDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLKVLSAYNKHRRDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDAIHLQHSITGLYETRIDLSQLGGD</p>	
SpyCas9-NG	Streptococcus pyogenes	<p>MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFGNIVDEVAYHEKY PTYHLRKKLVDSITDKADRLIYLAALAHMIKFRGHFLIEGDLNPDNSDVDFLQVLTQYNQLFEE NPINASGVDAKAILSAKLSRRLLENLIAQLPGEKKNGFLGNLIALSLGLTPNFKNFDLAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYVVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIIEGKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSDINKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFSDPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLKVLSAYNKHRRDKPIREQAENIHLFTLT NLGAPRAFKYFDTTIDRKVYRSTKEVLDAIHLQHSITGLYETRIDLSQLGGD</p>	N863A
SpyCas9-SpRY	Streptococcus pyogenes	<p>MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFGNIVDEVAYHEKY PTYHLRKKLVDSITDKADRLIYLAALAHMIKFRGHFLIEGDLNPDNSDVDFLQVLTQYNQLFEE NPINASGVDAKAILSAKLSRRLLENLIAQLPGEKKNGFLGNLIALSLGLTPNFKNFDLAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYVVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIIEGKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSDINKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFSDPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLKVLSAYNKHRRDKPIREQAENIHLFTLT RLGAPRAFKYFDTTIDPKQYRSTKEVLDAIHLQHSITGLYETRIDLSQLGGD</p>	N863A
StlCas9	Streptococcus thermophilus	<p>MSDLVLGLDIGIGSVGILNKTGTEIHKNSRIFFAAQAENNLVRRTRNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSEELFIALKNMVKHKGISYLDASDDGNSSVG DYAQIVKENSQKLETKTPGQIQLERYQTYGQLRGDFTEKDGKKHRLINVFPTSAIRSEALRILQT QQEFNPQITDEFINRYLEILTGRKRYHHPGNGEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDFR AAKASYTAQEFNLLNDLNNLTVPTEKLSKEQKNQINYNVNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKLTLETLDIEQMDRETLDKLAAYVLTNTEREGIEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSNK TSYIDEKLLTEIYNPVAKSVRQAIVNAAIKEYGDFDNIVIEAMARETNEDEKKAQKIQKANK DEKDAAMLKAANQYNGKAEPLHSVFHGHKLATKIRLWHQQGERCLYTGKTSIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKQRTPYQALDSMDDAWSFRELKAFVRESKTLSSNK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQTSQLRRH WGIEKTRDTYHHHADVADALIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFDVTLKSKEFEDSILFSYQVDSKFNKISDATIYATRQAKVGGDKADETYVLGKIKDIYTDQD YDAFMKIYKKDKSKFLMYRHPDQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY</p>	N622A

		SKKNGNGPEIKSLKYYDSKLGHNHIDITPKDSNNKVVLSQSVSPWRADVFNKTTGKYEILGLKYADLQFEKGTGTYSQEKYNDIKKKEGVDSSEFKFTLYKNDLLL VKDTEKEQQLFRFLSRTPMPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISYKVRTDVLGNQHIIKNEGDKPKLDF	
BlatCas9	Brevibacillus laterosporus	MAYTMGIDVGIASCGWAIVDLERQRIIDIGVRTFEKAENPKNGEALAVPREARSSRRRLRRKKHRIERLKHMFVRNGLAVDIQHLEQTLRSQNEIDVWQLRVDGLDRMLTQKEWLRVLIHLAQRGFGQS NRKTDGSSSEDGQVLVNVVTENDRLMEEKDYRTVAEMMVVDEKFSHDKRNKNGNYHGVVSRSSL LVEIHITLFTQRQHNSLASKDFELEYVNIWSAQRPVATKDQIEKMIGTCTFLPKEKRAPKASWHFQYFMLLQTINHIRTNVQGRSLNKEEIEQVVMALTKSKVSYHDTRKILDLSEYQFVGLDYGKEDEKKKVESKETIILDDYHKLNKIFNEVELAKGETWEADDYDTVAAYALTFKDDDEDIRDYLQNKYKDSKNRLVKNLANKETYNELIGKVSTLSFRKVGHLSLKALRKIIPFLEQGMITYDKACQAAGFDFQGISKKKRSVVL PVIDQISNPVVRALQTTRKVINALLIKKYGSPETHIETARELSKTFDERKNITKDYKENRDKNEHAKKHLSELGIINPTGLDIVKYKLWCEQQGRCMYSNQPISEFRLKESGYTEVDHIIPYSRSMNDSYNNRVLVMTRENREKGNQTPFEYMGNDTQRWYEFQQRVTTPNQIKKEKRQNLRLKGFTNRRELEMLERNLNDTRYITKYLSHFISTNLEFSPSDKKKKVVNTSGRITSILRSRWGLEKNRGQNDLHHAMDAIVAVTSDSFIQVVTNYKRRERRELNGDDKFLPWKFFREEVIARLSPNPKEQIEALPNHFYSEDELADLQPIFVSRMPKRSITGEAHQAQFRRVVGKTKEGKNITAKKTALVDISYDKNGDNFMYGRETDPATYEAIKERYLEFGGNVKKAFSTDHLHKPKKDGTGGLIKSVRMENKTLVHPVNGKGKGVVYNSSIVRTDVFQRKEKYLLPVVYTDVTKGKLPNKVIVAKKGYHDWIEVDSSFTFLFSLYPNDLIFIRQNPKKKISLKKRIESHSDSKVEQEIHAAYKGVDSSTAIEFIHHDGSYYAKGVGVQNLDCFEKYQVDILGNFYKVKGEKRLLETSDSNHKGKDVNSIKSTSR	N607A
cCas9-v16	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSSEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKEAKQLLKVKQAYHQLDQS FIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYKFIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTS TGKPEFTNLK VYHDIKDITARKEIENAEALLDQIAKILTYQSSEDIQEELTNLSEL TQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQKKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIEELAREKNSKDAQKMINEMQKRNQRTNERIEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNPNFNYEVDHIIPRSVSFDNSFNKVLVKQEENSCKGNRTPPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLDVVKVSINGGFTSFLRRKWKFKKERNKGKHHAEADALIANADFIKWKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDENKPLYKYEEET GNYLTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKF VTVKNLDVIKKENYEVNSKCYEEAKKLLKISNQAEFIASFYKNDLIKINGELYRVIGVNSDKNNL IELNMIDITYREYLENMNDKRPPIIHTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
cCas9-v17	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSSEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKEAKQLLKVKQAYHQLDQS FIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYKFIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTS TGKPEFTNLK VYHDIKDITARKEIENAEALLDQIAKILTYQSSEDIQEELTNLSEL TQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQKKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIEELAREKNSKDAQKMINEMQKRNQRTNERIEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNPNFNYEVDHIIPRSVSFDNSFNKVLVKQEENSCKGNRTPPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLDVVKVSINGGFTSFLRRKWKFKKERNKGKHHAEADALIANADFIKWKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDENKPLYKYEEET GNYLTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKF VTVKNLDVIKKENYEVNSKCYEEAKKLLKISNQAEFIASFYKNDLIKINGELYRVIGVNSDNRNI VELNMIDITYREYLENMNDKRPPIIHTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
cCas9-v21	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSSEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKEAKQLLKVKQAYHQLDQS FIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYKFIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTS TGKPEFTNLK VYHDIKDITARKEIENAEALLDQIAKILTYQSSEDIQEELTNLSEL TQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQKKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIEELAREKNSKDAQKMINEMQKRNQRTNERIEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNPNFNYEVDHIIPRSVSFDNSFNKVLVKQEENSCKGNRTPPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLDVVKVSINGGFTSFLRRKWKFKKERNKGKHHAEADALIANADFIKWKLDKAKKVMENQ MFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDENKPLYKYEEET GNYLTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKF VTVKNLDVIKKENYEVNSKCYEEAKKLLKISNQAEFIASFYKNDLIKINGELYRVIGVNSDDNRNI ELNMIDITYREYLENMNDKRPPIIHTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A

cCas9-v42	Staphylococcus aureus	MKRNILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLRKRRRRHRIQR VKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEDTGN LSTKEQISRNKSALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVQKAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYKFIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVSTGKPEFTNLK VYHDIKDITARKEIENAEALLDQIAKILTIYQSSEDIQEELTNLSELQTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVKKVVDLSQKKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIEELAREKNSKDAQKMINEMQKRRNRQTNERIEIIRITGKENAKYLIIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLDVVKVSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIANADFIKWKLDKAKKVMENQ MFEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTYSTRKDDKG NTLIVNNLNGLYDKDNDKLLKLKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDENPLYKYEEET GNYLTKYSKKDNGPVIKIKIYYGNKLNALHDITDDYPNSRNKVVKLSLKPFRFVDVYLDNGVYKF VTVKNLVDIKENYEVNSKCYEEAKKLKKSINQAEFIASFYKNDLIKINGELYRVIGVNNRNLNK IELNMIDITYREYLENMNDKRPHHIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
CdiCas9	Corynebacterium diphtheriae	MKYHVIGIDVGTFSVGLAAIEVDDAGMPIKTLVSLVSHIHDSGLDPDEIKSAVTRLASSGIARRTRRL YRRKRRRLQLQDKFIQRQGWPIVELEDYSDPLYPWKVRAELAASYIADEKERGEKLSVALRHRAR HRGWRNPYAKVSSLYLPDGPDAFKAIREEIKRASGQVPVETATVQGMVTLCELGTLLKRGEGGV LSARLQQSDYAREIQEICRMQEIQELRYKIIDVFAAESPKGSASSRVGKDPLOPQGNRLKASD AFQRYIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIAELGIDRGQLIGTAT MTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWTASALEQHAMVKALSNAEVDFFDSPEGAK VQAFADLDDDDVHAKLDSLHLVPGRAAYSEDTLVRLTRRMLSDGVLDLYTARLQEFGEPSWTPPT PRIGEPVGNPAVDRVLKTVSRWLESATKTWGAPERVIEHVREGEVTEKAREMDGDMRRRAAR NAKLQEMQEKLNVQKPSRADLWRYQSVQRQNCQACAYCGSPITFSNSEMDHIVPRAGQGSTNT RENLVAVCHRCNQSKGNTPFIAWAKNTSIEGVSVEKAEVERTRHVVTDGTMRSTDFKFKFTKAVVE RFQRATMDDEIDARSMESVAWMANELRSRV AQHFASHGTTVRVYRGSALTAEARRASGISGKLKF FDGVGKSRLDRRHHAIDA AVIAFTSDYVAETLAVRSNLKQSQAHQAEAPQWREFTGKDAEHRAA WRVWCQKMEKLSALLTEDLRDDR VVVMNSVRLRLGNGSAHKETIGKLSKVLKSSQLSVSDIDKA SSEALWCALTREPFGDPKEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGY AELGSSFH HARVYKITSGKKPAFAMLRVYITDILLPYRNQDLFSVELKPQTMMSRQAEKKLRDALATGNAEYL GWLVDDELVDVDTSKIATDQVKAELGTIRRWVVDGFFSPSKLRLRPLQMSKEGIIKESAPEL SKIIDRPGWLPAVNKLFSDBGVNTVVRDLSLGRVRLSTAHLPVTWKVQ	H573A (Alternate)
CjeCas9	Campylobacter jejuni	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLN HLKHLIANEFKLNIEDYQSFDES LAKAYKGLSLPYELRFRALNELLSSKQDFARVILHIAKRRGYD DIKNSDDKEKGAILKAIKQNEEKLANYSQSVGEYLKEYEFQKFENSKFTNVRNKKESYERCIQAQ SFLKDELKLIFFKKQREFGFSFSKKFEEVLSVAFYKRALKDFSHLVGNCSSFTDEKRAPKNSPLAFM FVALTRIINLLNNLKNTEGILYTKDDL NALLNEVLKNGTLTYKQTKLLGLSDDYEFKGEKGTYFI EFKKYKEFIKALGEHNSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNIS FKALKLVTPLMLEGKKYDEACNELNLKVAINEDKDDFLPAFNETYKDEVTNPVVLRAIKERYK VLNALLKKYGVHKNINELAREVGKNHSSQAKIEKEQENENYKAKKDAELECEKLGKINSKNILK LRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRFSDDSYMNVKLVFTKQNEKLNQTPFE AFGNSAKWQKIEVLAKNLP TKKQKRILDKNYKDKKEQKNFKDRNLNDTRYARLVNTYTKDYLD FLPLSDDENTKLNQDQKSKVHV EAKSGMLTSALRHTWGFSAKDRNNHLLHHAIDAVIIAYANNSI VKAFSDFKKEQESNSAELYAKKISELDYKNNKRFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALH EETFRKEEFYQSYGGKEGVSKALELGKIRK VNGKIVKNGMDMFRVDIFKHKTNFYAVPIYTM FALKVLPNKAVARSKKGEIKDWILMDENYEFCSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLI VSKHDKNFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFFK	N582A
GeoCas9	Geobacillus stearothermophilus	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRANPQTGESLALPRRLARSARRRLRRRKHR LERIRRLVIREGILTKEELDKLFEKHEIDVWQLRVEALDRKLNDELARVLLHLAKRRGFKSNRK SERNKENSTMLKHIEENRAILSSYRTVGEMIVKDPKFALHKNKNGENYNTIARDDLEREIRLIFS KQREFGNMSTEEFENEYITIWASQRPVASKDDIEKKVGFCFTEPKEKRAPKATYTFQSFIWEHIN KLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYDRGESRKQENIRF LELDAYHQIRKAVDKVYGGKSSSFLPIDFDTFGYALTFLKDDADIHSYLRNEYEQNGKRMPLA NKVYDNLIEELLNLSFTKFGHLSLALRSILPYMEQGEVYSSACERAGYFTTGPKKKQKTMLLPN IPPIANPVVMRALTQARKVVNAIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIR QLMEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRLDDSYTNK VLVLTRENREKGNRIAEYLGVGTERWQQFETVLTNKFQSKKKRDRLLRLHYDENEEETEFKNRN LNDTRYISRFANFIREHLKFAESDDKQKVYTVNGRVTAHLRSRWEFNKNREESDLHHAVIDAVIV ACTTPSDIAKVTAFYQRREQNKELAKKTEPHFPQWPWHFADELRLRSKHPKESIKALNLGNYDD QKLESLQPVFVSRMPKRSVYGAHQETLRRYVGIDERSGKIQT VVKTKLSEIKLDASGHFPMYGG ESDPRTYEAIRQLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIIDTKNQVIPLNDGKTVAYN SNIVRVDVFEKDGKYYCVPVYTMIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLRIE LPREKTVKTAAGEENVKDVVYVYKTDANGGLELISHDRFSLRGVGSRTLKRFKQYQVDVLG NIYKVRGEKRVGLASSAHSPGKTIRPLQSTRD	N605A
iSpyMac Cas9	Streptococcus spp.	MDKKYSIGLDIGITSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTYHLRKKLVDSDDKADRLRLYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRKLENLIALQPGKKNGLFGNLIALSLGLTPNFKNFNLDAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD	N863A

		<p>LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLKRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGLPARGNSRFAW MTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDILLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFGANRNFQMLIHDDSLTFKEDIQKAQVSGGQDLSLHEIANL AGSPAICKGILQTVKVVDELVKVMGRHKPENIVIEARENQTTQKQKNSRERMKRIIEGKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSDNKNV TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGISELDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTAIICKYPKLESEFVYGDYKVVYDVRKMIKSEQEIQKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKTETIQTQVQNGGLFDDNPKSPL EVTPSKLVPLKKNLNPKKYGGYQKPTTAYPVLLITDTKQIPISVMNKKQFEQNPVKFLDRGRYQQ VGKNDFIKLPKYTLVDIGDGIRKLWASSKEIHKGNQLVVSQKSSQILLYHAHLDSLSDNYLQNH NQQDFVLFNEIISFSKKCKLGEKHIQKIENVYSNKKNSASIEELAESFIKLLGTFQKSDSPFNGLGV KLNQKQYKGGKDYILPCTEGTLIRQSITGLYETRVLDLSKIGEDSGSGSGSKRTADGSEFES</p>	
NmeCas9	Neisseria meningitidis	<p>MAAFKPNISINYLGLDIGIASVGWAMVEIDEEENPRLIDLGVRFERAEVPTGDSLAMARRLAR SVRRLTRRRARHLRTRRLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDRKLTLEWSAVL LHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQR SDYSHTFSRKDLQAEILLFEKQKEFGNPHVSGGLKEGIELTMTQRPALSGDAVQKMLGHCTFEP AEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSLQTYAQARKLLGLE DTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLDKKSPLNLSPELQDEIGTAFSLFKTD EDITGRLKDRIQPEILEALLKHISFDKVFQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTE EKIYLPPIPADEIRNPVVLRLALSQARKVINGVRRYVGPARIHIETAREVGSKFDRKIEKRQEENR KDREKAAAKFREYFPNFVGEPPKSKDILKLRLYEQQHKGKCLYSGKEINLGRLEKGYVEIDHALPFS RTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKF DEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGGKRVFASNGQITNLLRGFWGLRKRVAEND RHHAHDVVVACSTVAMQQKITRFRVRYKEMNAFDGKTIDKETGEVLHQKTHFPQVEFFAQEV MIRVFGKPDGKPEFEEADTLEKLRITLLAEKLSRPEAVHEYVTPLFVSRAPNRKMSQGGHMETVK SAKRLDEGVSVLRVPLTQLKLKLEKMMVNREREPKLYEALKARLEAHKDDPAKAFAPFYYKYDK AGNRTOQVKAVRVEQVQKTGVVWRNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGIL PDRAVVOGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFYFASCHRGTGNNIRIHDLDH KIGKNGILEGIGVKTALSFOKYQIDELGKEIRPCLKKRPPVR</p>	N611A
ScaCas9	Streptococcus canis	<p>MEKKYSIGLDIGTNSVGWAVITDDYKVPSSKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRRYTRRKNRIRYLQEIFANEMAKLDDSFQRLSEESFLVEEDKKNERHPFIGNLADEVAYHRN YPTIYHLRKKLADSPEKADRLIYLALAHIIKFRGHFLIEGKLNANSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAGKILSARLSKSKRLEKLIIVFPNEKKNGLFGNIALALGLTPNFKSNFDLTEDAKLQL SKDTYDDDLDELGGQIGDQYADLFSAAKNLSDAILSDILRSNSEVTKAPLSASVMKRYDEHHQD LALLKTLVRQPFPEKYAEIFKDDTKNGYAGYVIGIKHRRKRTTKLATQEEFYKFIKPILEKMDGAE ELLAKLNRDILLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYFPLKENREKIEKILTFRIPYYVGLP ARGNSRFAWLTRKSEEAITPWNFEEVVDKGASQSFIERMTNFDQLPNKKVLPHKSHLLYEYFTV YNELTKVKYVTERMRKPEFLSGEQKKAIVDILLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDR FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLK RRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDFGSNRFQMLIHDDSLTFKEEIEKAQVSGQGD SLHEQIADLAGSPAICKGILQTVKIVDELVKVMGHKPENIVIEARENQTTTGLQSSRERKKRIE GKIELESQILKENPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVHIVPQSFIKDDSI DNKVLTRSVENRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGISEADKAG FIKRQLVETROITKHVARILDSRMNTKRDKNDKPIREVKVITLKSCLVSDFRKDFQLYKVRDINNY HHAHDAYLNAVVGTAIICKYPKLESEFVYGDYKVVYDVRKMIKSEQEIQKATAKRFFYSNIMN FKTEVKLANGEIRKRPLIETNGETGEVWNKEKDFATVRKVLAMPQVNVKKTETVQTGGFSKESI LSKRESAKLIPRKKGWDRKYGGFGSPVAYSILVVAKEVGKAKKLKSVKVLVGITIMEKGSYE KDPGFLEAKGYKDIKKELIFKLPKYSLEFENGRRRMLASATELQKANELVLPQHLVRLLYYTQN ISATTGSNNLGYIEQHREEFKEIFEKIIFSEKYLKNKVNLSLKSSFDEQFVSDSILLSNSFVSLLK YTSFGASGGFITLDDLVDKQGRRLRYQTVEVLDTLIYQSITGLYETRTDLSQLGGD</p>	N872A
ScaCas9-HiFi-Sc++	Streptococcus canis	<p>MEKKYSIGLDIGTNSVGWAVITDDYKVPSSKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRRYTRRKNRIRYLQEIFANEMAKLDDSFQRLSEESFLVEEDKKNERHPFIGNLADEVAYHRN YPTIYHLRKKLADSPEKADRLIYLALAHIIKFRGHFLIEGKLNANSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAGKILSARLSKSKRLEKLIIVFPNEKKNGLFGNIALALGLTPNFKSNFDLTEDAKLQL SKDTYDDDLDELGGQIGDQYADLFSAAKNLSDAILSDILRSNSEVTKAPLSASVMKRYDEHHQD LALLKTLVRQPFPEKYAEIFKDDTKNGYAGYVADKKLKRSGKLATEEFYKFIKPILEKMDGA EELLAKLNRDILLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYFPLKENREKIEKILTFRIPYYVGLP LARGNSRFAWLTRKSEEAITPWNFEEVVDKGASQSFIERMTNFDQLPNKKVLPHKSHLLYEYFTV RYNASTKYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLK KRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDFGSNANFQMLIHDDSLTFKEEIEKAQVSGQ GDSLHEQIADLAGSPAICKGILQTVKIVDELVKVMGHKPENIVIEARENQTTTGLQSSRERKKR IEGKIELESQILKENPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVHIVPQSFIK DSDNKNLTRSVENRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGISEADK AGFIKRQLVETROITKHVARILDSRMNTKRDKNDKPIREVKVITLKSCLVSDFRKDFQLYKVRDIN NYHHAHDAYLNAVVGTAIICKYPKLESEFVYGDYKVVYDVRKMIKSEQEIQKATAKRFFYSNIM</p>	N872A

		NFFKTEVKLANGAIRKRPLIETNGETGEVWVWNEKDFATVRKVLAMPQVNVKKTEVQTGGFSKE SILSKRESAKLIPRKKGWDRTRYGGFGSPVAVAYSILVVAKEVGKAKKLKSVKVLVGITIMEKGSY EKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQHLVRLLYYTQ NISATTGSNNLYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSL KYTSFGASGGFTFLDLVDYKQGRRLRYQTVTEVLDATLIYQSITGLYETRIDLSQLGGD	
SpyCas9- 3var- NRRH	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLLK TARRRYTTRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTYHLRKKLVDSSTDKADLRLLIYALAHMIKFRGHFLIEGDLNPDNSDVVKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEVVDKGASASQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDILLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRLRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDLSHEHIANL AGSPAIIKKGILQTVKVVDELVKVMGGHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVFREINNYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQIEGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTAAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKFNPIGFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLHKGNELALPSKYVNFLYLASHYEKLKGS EDNEQKQLFVEQHKHYLDEIIQISEFSKRVLADANLKDVLSAYNKHRDKPIREQAENIHLFTLT NLGVAAPAFKYFDTTIDKKRYTSTKEVLDATLIHQSIITGLYETRIDLSQLGGD	N863A
SpyCas9- 3var- NRTH	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLLK TARRRYTTRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTYHLRKKLVDSSTDKADLRLLIYALAHMIKFRGHFLIEGDLNPDNSDVVKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEVVDKGASASQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDILLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRLRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDLSHEHIANL AGSPAIIKKGILQTVKVVDELVKVMGGHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVFREINNYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQIEGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTAAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKFNPIGFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASASVLHKGNELALPSKYVNFLYLASHYEKLKGS EDNKQQLFVEQHKHYLDEIIQISEFSKRVLADANLKDVLSAYNKHRDKPIREQAENIHLFTLT NLGASAAAFKYFDTTIGRKLYTSTKEVLDATLIHQSIITGLYETRIDLSQLGGD	N863A
SpyCas9- 3var- NRCH	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLLK TARRRYTTRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTYHLRKKLVDSSTDKADLRLLIYALAHMIKFRGHFLIEGDLNPDNSDVVKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEVVDKGASASQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDILLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRLRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDLSHEHIANL AGSPAIIKKGILQTVKVVDELVKVMGGHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVFREINNYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQIEGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTAAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKFNPIGFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGS EDNEQKQLFVEQHKHYLDEIIQISEFSKRVLADANLKDVLSAYNKHRDKPIREQAENIHLFTLT NLGAPAAAFKYFDTTINRKQYNTTKEVLDATLIHQSIITGLYETRIDLSQLGGD	N863A

SpyCas9-HF1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTTYHLRKKLVDSDDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQVLVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFSDLAEADAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSHLLYEFYTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANL AGSPAIAKKGILQTVKVVDDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVIDHIVPQSFLKDDSDNKVL TRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSLLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLSVKELLGITIMERSSEFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLLKGS PEDNEQKQLFVEQHKHYLDEIIQISEFSKRVLADANLKDVLKLSAYNKHDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGGD	N863A
SpyCas9-QQR1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTTYHLRKKLVDSDDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQVLVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFSDLAEADAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSHLLYEFYTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANL AGSPAIAKKGILQTVKVVDDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVIDHIVPQSFLKDDSDNKVL TRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSLLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLSVKELLGITIMERSSEFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLLKGS PEDNEQKQLFVEQHKHYLDEIIQISEFSKRVLADAQDLKVLKLSAYNKHDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTTFKQKQYRSTKEVLDTLHQSITGLYETRIDLSQLGGD	N863A
SpyCas9-SpG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTTYHLRKKLVDSDDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQVLVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFSDLAEADAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSHLLYEFYTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANL AGSPAIAKKGILQTVKVVDDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVIDHIVPQSFLKDDSDNKVL TRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSLLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFWPTVAYSVLVVAKEVGKSKKLSVKELLGITIMERSSEFEKNPIDFLEA AKGYKEVKKDLIIKLPKYSLEFELNGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLLKGS PEDNEQKQLFVEQHKHYLDEIIQISEFSKRVLADANLKDVLKLSAYNKHDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKQYRSTKEVLDTLHQSITGLYETRIDLSQLGGD	N863A
SpyCas9-VQR	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTTYHLRKKLVDSDDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQVLVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFSDLAEADAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLLKDNREKIEKILTFRIPYYVGPLARGNSRFAW	N863A

		<p>MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQLIHDDSLTFKEDIQKAQVSGQGSLSHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKQKNSRERMKRIIEGKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKEVQTTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFVSPVAVSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASAGELQKGNELALPSKYVNFYLAASHYEKLKGGSP EDNEQKQLFVEQHKHYLDEIIQISEFSKRVILADANLKDVL SAYNKHHRDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSIITGLYETRIDLSQLGGD</p>	
SpyCas9- VRER	Streptoco ccus pyogenes	<p>MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTYHLRKKLVDSITDKADRLIYLAALAHMIKFRGHFLIEGDLNPDNSDVVKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFDLAEDTKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTTEELLVKLNRED LLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQLIHDDSLTFKEDIQKAQVSGQGSLSHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKQKNSRERMKRIIEGKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKEVQTTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFVSPVAVSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASARELQKGNELALPSKYVNFYLAASHYEKLKGGSP EDNEQKQLFVEQHKHYLDEIIQISEFSKRVILADANLKDVL SAYNKHHRDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSIITGLYETRIDLSQLGGD</p>	N863A
SpyCas9- xCas	Streptoco ccus pyogenes	<p>MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTYHLRKKLVDSITDKADRLIYLAALAHMIKFRGHFLIEGDLNPDNSDVVKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFDLAEDTKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTTEELLVKLNRED LLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQLIHDDSLTFKEDIQKAQVSGQGSLSHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKQKNSRERMKRIIEGKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKEVQTTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAVSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASAGVLQKGNELALPSKYVNFYLAASHYEKLKGGSP EDNEQKQLFVEQHKHYLDEIIQISEFSKRVILADANLKDVL SAYNKHHRDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSIITGLYETRIDLSQLGGD</p>	N863A
SpyCas9- xCas-NG	Streptoco ccus pyogenes	<p>MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTYHLRKKLVDSITDKADRLIYLAALAHMIKFRGHFLIEGDLNPDNSDVVKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFDLAEDTKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTTEELLVKLNRED LLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQLIHDDSLTFKEDIQKAQVSGQGSLSHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKQKNSRERMKRIIEGKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKRQL</p>	N863A

1077

	KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFVDTLKSKEFEDSILFSYQVDSKFNKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY SKKGNPGEIKSLKYYDSKLGHNHIDITPKDSNNKVVLSLKPWRDVTYFNKNTGKYEILGLKYSDM QFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSNGEQILLRFTSRNDTSKHY VELKPYNRQKFEGSEYLIKSLGTVVKGGRCIKGLGKSNISYKVRTDVLGNQHIIKNEGDKPKLDF
--	---

In some embodiments, the endonuclease domain or DNA binding domain 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/Cpf1, 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/Cpf1, 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/Cpf1, 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/Cpf1, 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 substitutions, e.g., selected from H840A, D10A, P475A, W476A, N477A, D1125A, W1126A, and D1127A. In 5
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, 10
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 syphilidicola*, *Prevotella intermedia*, *Spiroplasma taiwanense*, *Streptococcus iniae*, *Belliella baltica*, *Psychroflexus torquis*, *Streptococcus thermophilus*, *Listeria innocua*, *Campylobacter jejuni*, *Neisseria meningitidis*, *Streptococcus* 15
pyogenes, or *Staphylococcus aureus*, or a fragment or variant thereof.

In some embodiments, the endonuclease domain or DNA binding domain 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, the endonuclease domain or DNA binding domain comprises 20
spCas9, spCas9-VRQR (SEQ ID NO: 1578), spCas9-VRER (SEQ ID NO: 1579), xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER (SEQ ID NO: 1580), spCas9-LRKIQK (SEQ ID NO: 1581), or spCas9-LRVSQL (SEQ ID NO: 1582).

In some embodiments, the endonuclease domain or DNA-binding domain comprises an 25
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 30
differences (e.g., mutations) relative to any of the amino acid sequences described herein.

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 (SEQ ID NO: 1023)
Exemplary Linker Motif	(SGGS) _n (SEQ ID NO: 1583)
Exemplary Linker Motif	(GGGS) _n (SEQ ID NO: 1584)
Exemplary Linker Motif	(GGGGS) _n (SEQ ID NO: 1535)
Exemplary Linker Motif	(G) _n
Exemplary Linker Motif	(EAAAK) _n (SEQ ID NO: 1534)
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 <i>Streptococcus thermophilus</i>	NCBI Ref: YP_820832.1
Cas9 from <i>Listeria innocua</i>	NCBI Ref: NP_472073.1
Cas9 from <i>Campylobacter jejuni</i>	NCBI Ref: YP_002344900.1
Cas9 from <i>Neisseria meningitidis</i>	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 <i>Parcubacteria</i> group bacterium])
CasX	uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53
CasX	>trF0NH53IF0NH53_SULIR CRISPR associated protein, Casx OS = <i>Sulfolobus islandicus</i> (strain REY15A) GN=SiRe_0771 PE=4 SV=1
Deltaproteobacteria CasX	
Cas12b/C2c1	((uniprot.org/uniprot/T0D7A2#2) splT0D7A2IC2C1_ALIAG CRISPR- associated endonuclease C2c1 OS = <i>Alicyclobacillus acido-terrestris</i> (strain ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1)
BhCas12b (<i>Bacillus hisashii</i>)	NCBI Reference Sequence: WP_095142515
BvCas12b (<i>Bacillus</i> sp. V3-13)	NCBI Reference Sequence: WP_101661451.1
Wild-type <i>Francisella novicida</i> 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:

5

Cas9 nickase (H840A):

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
 TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGN
 IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDV
 5 DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI
 ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAIL
 LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
 YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
 10 DKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLS
 GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
 KDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG
 RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSL
 HEHIANLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRE
 15 RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV
 DAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRK
 FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
 TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYK
 VYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
 20 KGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
 FDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKK
 DLIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED
 NEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHARDKPIREQAENIIHL
 FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD (SEQ
 25 ID NO: 1585)

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:

30

M-MLV (WT):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS
 KQYPMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
 RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
 5 GQLTWTRLPPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
 TRALLQTLGNLGYRASAKKAQICQKQVKYLGILLKEGQRWLTEARKETVMGQPTPKT
 PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIQALLTAP
 ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
 VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTR
 10 VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
 LQEGQRKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSTRY
 AFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIHCPGHQKGHSAEAR
 GNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 1586)

15 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:
 TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS
 KQYPMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
 RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
 20 GQLTWTRLPPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
 TRALLQTLGNLGYRASAKKAQICQKQVKYLGILLKEGQRWLTEARKETVMGQPTPKT
 PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIQALLTAP
 ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
 VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLDTR
 25 VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
 LQEGQRKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSTRY
 AFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIHCPGHQKGHSAEAR
 GNRMADQAARKAAITETPDTSTLL (SEQ ID NO: 1548)

30 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:

5 TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS
 KQYPMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVN
KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKD**AFFCLRLHPTSQPLFAFEWRDP**
EMGISGQLTWTRL**PQGFKN****SPTLFDEALHRDLAD****FRIQHPDLILLQYVDDLLLAAT**
SELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLG**YLLKEGQRWL****TEARKE**
 TVMGQPTPKTPRQLREFLGTAFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAY
 10 QEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPV
 AAGWPPCLRMVAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTH
 YQALLDTRVQFGPVVALNPATLLPLPEEGLQHNCILDILAEAHGTRPDLTDQPLPDAD
 HTWYTDGSSLLQEGQRKAGAAVTTETEVIAKALPAGTSAQRAELIALTQALKMAEGK
 15 KLNVYTDSRYAFATAHIGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIHCPG
HQKGHSAEARGNRMADQAARKAA (SEQ ID NO: 1587)

Core RT (bold), annotated per above
RNAseH (underlined), annotated per above

20 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
 25 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,
 30 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:

35 *M-MLV (PE2):*
 TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVS

KQYPMSQEARLGKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
 RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
 GQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCCQQG
 TRALLQTLGNLGYRASAKKAQICQKQVKYLGILLKEGQRWLTEARKETVMGQPTPKT
 5 PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAP
 ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
 VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTR
 VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
 LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVTDSRY
 10 AFATAHIGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
 GNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 1588)

In some embodiments, a Gene Writer polypeptide may comprise a linker, e.g., a peptide
 linker, e.g., a linker as described in Table 38. In some embodiments, a Gene Writer polypeptide
 15 comprises a flexible linker between the endonuclease and the RT domain, e.g., a linker
 comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ
 ID NO: 1589). 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.

20

Table 38. Exemplary linker sequences

Amino Acid Sequence
GGS
GGSGGS
GGSGGSGGS
GGSGGSGGSGGS
GGSGGSGGSGGSGGS
GGSGGSGGSGGSGGSGGS
GGGGS
GGGGSGGGGS
GGGGSGGGGSGGGGS
GGGGSGGGGSGGGSGGGGS
GGGGSGGGGSGGGSGGGGS

GGGSGGGSGGGSGGGSGGGSGGGSGGGGS
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
GGGGSGSS
GGGGSSGGS
GSSGSGGG
GSSGGGGGS
GGSGGGEAAAK
GGSEAAAKGGG
GGGGGSEAAAK
GGGEAAAKGGS
EAAAKGGSGGG
EAAAKGGGGGS
GGSGGGPAP
GGSPAPGGG
GGGGGSPAP
GGGPAPGGS
PAPGSGGG
PAPGGGGGS
GGSGSSEAAAK
GGSEAAAKGSS
GSSGGSEAAAK
GSSEAAAKGGS
EAAAKGGSGSS
EAAAKGSSGGS
GGSGSSPAP
GGSPAPGSS
GSSGGSPAP
GSSPAPGGS
PAPGSGSS
PAPGSSGGS
GGSEAAAKPAP
GGSPAPEAAAK
EAAAKGGSPAP

EAAAKPAPGGS
PAPGGSEAAAK
PAPEAAAKGGS
GGGSSEAAAK
GGGEAAAKGSS
GSSGGGEAAAK
GSSEAAAKGGG
EAAAKGGGGSS
EAAAKGSSGGG
GGGGSSPAP
GGGPAPGSS
GSSGGGPAP
GSSPAPGGG
PAPGGGGSS
PAPGSSGGG
GGGEAAAKPAP
GGGPAPEAAAK
EAAAKGGGPAP
EAAAKPAPGGG
PAPGGGEAAAK
PAPEAAAKGGG
GSSEAAAKPAP
GSSPAPEAAAK
EAAAKGSSPAP
EAAAKPAPGSS
PAPGSSEAAAK
PAPEAAAKGSS
AEAAAKEAAAKEAAAKEAAAKALEAAAKEAAAKEAAAKEAAAKA
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
 AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPI
 FGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
 SDVDKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLF
 5 GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLS
 DAILLSILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
 YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
 LHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFE
 EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
 10 AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
 LKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
 GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQ
 GDSLHEHIANLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENIVIAMARENQTTQKGQK
 NSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
 15 DYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLI
 TQRKFDNLTKAERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIR
 EVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVY
 GDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRLIETNGETG
 EIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
 20 KYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYK
 EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
 SPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAE
 NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLGGD
 (SEQ ID NO: 1590)

25

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
- 30 (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 (SEQ ID NO: 1591).
- 35 (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.

5 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.

10 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.

20 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.

25 In some embodiments, the RT and endonuclease domains are joined by a flexible linker, e.g., comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 1589).

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.

30 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
5 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.

In some embodiments, an endonuclease domain or DNA-binding domain comprises a TAL effector molecule. A TAL effector molecule, e.g., a TAL effector molecule that specifically
10 binds a DNA sequence, typically comprises a plurality of TAL effector domains or fragments thereof, and optionally one or more additional portions of naturally occurring TAL effectors (e.g., N- and/or C-terminal of the plurality of TAL effector domains). Many TAL effectors are known to those of skill in the art and are commercially available, e.g., from Thermo Fisher Scientific.

15 Naturally occurring TALEs are natural effector proteins secreted by numerous species of bacterial pathogens including the plant pathogen *Xanthomonas* which modulates gene expression in host plants and facilitates bacterial colonization and survival. The specific binding of TAL effectors is based on a central repeat domain of tandemly arranged nearly identical repeats of typically 33 or 34 amino acids (the repeat-variable di-residues, RVD domain).

20 Members of the TAL effectors family differ mainly in the number and order of their repeats. The number of repeats typically ranges from 1.5 to 33.5 repeats and the C-terminal repeat is usually shorter in length (e.g., about 20 amino acids) and is generally referred to as a “half-repeat”. Each repeat of the TAL effector generally features a one-repeat-to-one-base-pair correlation with different repeat types exhibiting different base-pair specificity (one repeat
25 recognizes one base-pair on the target gene sequence). Generally, the smaller the number of repeats, the weaker the protein-DNA interactions. A number of 6.5 repeats has been shown to be sufficient to activate transcription of a reporter gene (Scholze et al., 2010).

Repeat to repeat variations occur predominantly at amino acid positions 12 and 13, which have therefore been termed “hypervariable” and which are responsible for the specificity of the
30 interaction with the target DNA promoter sequence, as shown in Table 5 listing exemplary repeat variable diresidues (RVD) and their correspondence to nucleic acid base targets.

Table 5 – RVDs and Nucleic Acid Base Specificity

Target	Possible RVD Amino Acid Combinations												
A	NI	NN	CI	HI	KI								
G	NN	GN	SN	VN	LN	DN	QN	EN	HN	RH	NK	AN	FN
C	HD	RD	KD	ND	AD								
T	NG	HG	VG	IG	EG	MG	YG	AA	EP	VA	QG	KG	RG

5 Accordingly, it is possible to modify the repeats of a TAL effector to target specific DNA sequences. Further studies have shown that the RVD NK can target G. Target sites of TAL effectors also tend to include a T flanking the 5' base targeted by the first repeat, but the exact mechanism of this recognition is not known. More than 113 TAL effector sequences are known to date. Non-limiting examples of TAL effectors from *Xanthomonas* include, Hax2, Hax3, Hax4, 10 AvrXa7, AvrXa10 and AvrBs3.

Accordingly, the TAL effector domain of a TAL effector molecule described herein may be derived from a TAL effector from any bacterial species (e.g., *Xanthomonas* species such as the African strain of *Xanthomonas oryzae* pv. *Oryzae* (Yu et al. 2011), *Xanthomonas campestris* pv. *raphani* strain 756C and *Xanthomonas oryzae* pv. *oryzicola* strain BLS256 15 (Bogdanove et al. 2011)). In some embodiments, the TAL effector domain comprises an RVD domain as well as flanking sequence(s) (sequences on the N-terminal and/or C-terminal side of the RVD domain) also from the naturally occurring TAL effector. It may comprise more or fewer repeats than the RVD of the naturally occurring TAL effector. The TAL effector molecule can be designed to target a given DNA sequence based on the above code and others known in 20 the art. The number of TAL effector domains (e.g., repeats (monomers or modules)) and their specific sequence can be selected based on the desired DNA target sequence. For example, TAL effector domains, e.g., repeats, may be removed or added in order to suit a specific target sequence. In an embodiment, the TAL effector molecule of the present invention comprises between 6.5 and 33.5 TAL effector domains, e.g., repeats. In an embodiment, TAL effector 25 molecule of the present invention comprises between 8 and 33.5 TAL effector domains, e.g., repeats, e.g., between 10 and 25 TAL effector domains, e.g., repeats, e.g., between 10 and 14 TAL effector domains, e.g., repeats.

In some embodiments, the TAL effector molecule comprises TAL effector domains that correspond to a perfect match to the DNA target sequence. In some embodiments, a mismatch between a repeat and a target base-pair on the DNA target sequence is permitted as long as it allows for the function of the polypeptide comprising the TAL effector molecule. In general, TALE binding is inversely correlated with the number of mismatches. In some embodiments, the TAL effector molecule of a polypeptide of the present invention comprises no more than 7 mismatches, 6 mismatches, 5 mismatches, 4 mismatches, 3 mismatches, 2 mismatches, or 1 mismatch, and optionally no mismatch, with the target DNA sequence. Without wishing to be bound by theory, in general the smaller the number of TAL effector domains in the TAL effector molecule, the smaller the number of mismatches will be tolerated and still allow for the function of the polypeptide comprising the TAL effector molecule. The binding affinity is thought to depend on the sum of matching repeat-DNA combinations. For example, TAL effector molecules having 25 TAL effector domains or more may be able to tolerate up to 7 mismatches.

In addition to the TAL effector domains, the TAL effector molecule of the present invention may comprise additional sequences derived from a naturally occurring TAL effector. The length of the C-terminal and/or N-terminal sequence(s) included on each side of the TAL effector domain portion of the TAL effector molecule can vary and be selected by one skilled in the art, for example based on the studies of Zhang et al. (2011). Zhang et al., have characterized a number of C-terminal and N-terminal truncation mutants in Hax3 derived TAL-effector based proteins and have identified key elements, which contribute to optimal binding to the target sequence and thus activation of transcription. Generally, it was found that transcriptional activity is inversely correlated with the length of N-terminus. Regarding the C-terminus, an important element for DNA binding residues within the first 68 amino acids of the Hax 3 sequence was identified. Accordingly, in some embodiments, the first 68 amino acids on the C-terminal side of the TAL effector domains of the naturally occurring TAL effector is included in the TAL effector molecule. Accordingly, in an embodiment, a TAL effector molecule comprises 1) one or more TAL effector domains derived from a naturally occurring TAL effector; 2) at least 70, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200, 220, 230, 240, 250, 260, 270, 280 or more amino acids from the naturally occurring TAL effector on the N-terminal side of the TAL effector domains; and/or 3) at least 68, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200,

220, 230, 240, 250, 260 or more amino acids from the naturally occurring TAL effector on the C-terminal side of the TAL effector domains.

5 In some embodiments, an endonuclease domain or DNA-binding domain is or comprises a Zn finger molecule. A Zn finger molecule comprises a Zn finger protein, e.g., a naturally occurring Zn finger protein or engineered Zn finger protein, or fragment thereof. Many Zn finger proteins are known to those of skill in the art and are commercially available, e.g., from Sigma-Aldrich.

10 In some embodiments, a Zn finger molecule comprises a non-naturally occurring Zn finger protein that is engineered to bind to a target DNA sequence of choice. See, for example, Beerli, et al. (2002) *Nature Biotechnol.* 20:135-141; Pabo, et al. (2001) *Ann. Rev. Biochem.* 70:313-340; Isalan, et al. (2001) *Nature Biotechnol.* 19:656-660; Segal, et al. (2001) *Curr. Opin. Biotechnol.* 12:632-637; Choo, et al. (2000) *Curr. Opin. Struct. Biol.* 10:411-416; U.S. Pat. Nos. 6,453,242; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,030,215; 6,794,136; 7,067,317; 15 7,262,054; 7,070,934; 7,361,635; 7,253,273; and U.S. Patent Publication Nos. 2005/0064474; 2007/0218528; 2005/0267061, all incorporated herein by reference in their entireties.

An engineered Zn finger protein may have a novel binding specificity, compared to a naturally-occurring Zn finger protein. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using 20 databases comprising triplet (or quadruplet) nucleotide sequences and individual Zn finger amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers which bind the particular triplet or quadruplet sequence. See, for example, U.S. Pat. Nos. 6,453,242 and 6,534,261, incorporated by reference herein in their entireties.

25 Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in U.S. Pat. Nos. 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as International Patent Publication Nos. WO 98/37186; WO 98/53057; WO 00/27878; and WO 01/88197 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger proteins has been described, for example, in International 30 Patent Publication No. WO 02/077227.

In addition, as disclosed in these and other references, zinc finger domains and/or multi-

fingering zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Pat. Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned International Patent Publication No. WO 02/077227.

Zn finger proteins and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Pat. Nos. 6,140,081; 789,538; 6,453,242; 6,534,261; 5,925,523; 6,007,988; 6,013,453; and 6,200,759; International Patent Publication Nos. WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536; and WO 03/016496.

In addition, as disclosed in these and other references, Zn finger proteins and/or multi-fingering Zn finger proteins may be linked together, e.g., as a fusion protein, using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Pat. Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The Zn finger molecules described herein may include any combination of suitable linkers between the individual zinc finger proteins and/or multi-fingering Zn finger proteins of the Zn finger molecule.

In certain embodiments, the DNA-binding domain or endonuclease domain comprises a Zn finger molecule comprising an engineered zinc finger protein that binds (in a sequence-specific manner) to a target DNA sequence. In some embodiments, the Zn finger molecule comprises one Zn finger protein or fragment thereof. In other embodiments, the Zn finger molecule comprises a plurality of Zn finger proteins (or fragments thereof), e.g., 2, 3, 4, 5, 6 or more Zn finger proteins (and optionally no more than 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 Zn finger proteins). In some embodiments, the Zn finger molecule comprises at least three Zn finger proteins. In some embodiments, the Zn finger molecule comprises four, five or six fingers. In some embodiments, the Zn finger molecule comprises 8, 9, 10, 11 or 12 fingers. In some embodiments, a Zn finger molecule comprising three Zn finger proteins recognizes a target DNA sequence comprising 9 or 10 nucleotides. In some embodiments, a Zn finger molecule

comprising four Zn finger proteins recognizes a target DNA sequence comprising 12 to 14 nucleotides. In some embodiments, a Zn finger molecule comprising six Zn finger proteins recognizes a target DNA sequence comprising 18 to 21 nucleotides.

In some embodiments, a Zn finger molecule comprises a two-handed Zn finger protein.

- 5 Two handed zinc finger proteins are those proteins in which two clusters of zinc finger proteins are separated by intervening amino acids so that the two zinc finger domains bind to two discontinuous target DNA sequences. An example of a two handed type of zinc finger binding protein is SIP1, where a cluster of four zinc finger proteins is located at the amino terminus of the protein and a cluster of three Zn finger proteins is located at the carboxyl terminus (see
- 10 Remade, et al. (1999) EMBO Journal 18(18):5073-5084). Each cluster of zinc fingers in these proteins is able to bind to a unique target sequence and the spacing between the two target sequences can comprise many nucleotides.

DNA binding domain:

- 15 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 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
- 20 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
- 25 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
- 30 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 certain embodiments, the DNA-binding domain of the polypeptide is a heterologous DNA-binding protein or domain relative to a native retrotransposon sequence. In some
5 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
10 endonuclease activity. In some embodiments, the heterologous DNA binding element retains
partial endonuclease activity to cleave ssDNA, e.g., possesses nickase 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, TAL domain, ZF domain, Myb domain, combinations thereof, or multiples
15 thereof. In certain embodiments, the heterologous DNA-binding domain is a DNA binding
domain of a retrotransposon or virus described in Table 1 or Table 3. A person having ordinary
skill in the art is capable of identifying DNA binding domains 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
20 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.,
25 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
30 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).

5 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 polypeptide may bind to one or more than one host DNA sequence.

10 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.

15 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
20 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
25 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
30 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
5 than 5, 10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or 1000 bp in length. 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
10 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
15 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
20 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). 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
25 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 the neighboring domain of the Gene Writer.

In some embodiments, an NLS comprises the amino acid sequence
30 MDSLLMNRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 1592),
PKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 1593),

RKSGKIAAIWKRPKPKKKRKV (SEQ ID NO: 1594), KRTADGSEFESPKKKRKV (SEQ ID NO: 1595), KKTELQTTNAENKTKKL (SEQ ID NO: 1596), or KRGINDRNFWRGENGRKTR (SEQ ID NO: 1597), KRPAATKKAGQAKKKK (SEQ ID NO: 1598), 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 39. 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 Bioinform 13:157 (2012), incorporated herein by reference in its entirety).

Table 39. Exemplary nuclear localization signals for use in Gene Writing systems

Sequence	Sequence References
AHFKISGEKRPSTDPGKKAK NPKKKKKKKDP	Q76IQ7
AHRAKKMSKTHA	P21827
ASPEYVNLPIGNG	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
IINGRKLKLLKSSRRSSQTS 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
KKSRRKGS	B4FG96
KKSTALSRELKGIMRRR	SeqNLS, P32354
KKSYPQDEIIAHSRPRK	Q9U7C9
KKTGKNRKLKSKRVKTR	Q9Z301, O54943, Q8K3T2
KKVSIAGQSGKLWRWKR	Q6YUL8
KKYENVVIKRSRKRGRPRK	SeqNLS
KNKKRK	SeqNLS
KPKKKR	SeqNLS
KRAMKDDSHGNSTSPKRRK	Q0E671
KRANSNLVAAAYEKAKKK	P23508
KRASEDTTSGSPPKSSAGPKR	Q9BZZ5, Q5R644
KRFKRRWMVRKMKTCK	SeqNLS
KRGLNSSFETSPKKVK	Q8IV63
KRGNSSIGPNDLSKRKQRK	SeqNLS
KRIHSVLSQSQIDPSKKVK	SeqNLS
RAK	SeqNLS
KRKGKLKNKGSKRKK	O15381
KRRRRRRRREKRKR	Q96GM8
KRSNDRTYSPEEEKQRRRA	Q91ZF2
KRTVATNGDASGAHRAKMSK	SeqNLS
KRVYNKGEDEQEHLPGGKKR	SeqNLS
KSGKAPRRRAVSMDNSNK	Q9WVH4, O43524
KVNFLDMSLDDIIHYKELE	Q9P127
KVQHRIAKKTTRRRR	Q9DXE6
LSPSLSP	Q9Y261, P32182, P35583

MDSLLMNRKFLYQFKNVR WAKGRRETYLC	Q9GZX7
MPQNEYIELHRKRYGYRLD YHEKKRKKESREAHERSKK AKKMIGLKAKLYHK	SeqNLS
MVQLRPRASR	SeqNLS
NNKLLAKRRKGGASPKDDP MDDIK	Q965G5
NYKRPMDGTYGPPAKRHEG E	O14497, A2BH40
PDTKRAKLDSETTMVKKK	SeqNLS
PEKRTKI	SeqNLS
PGGRGKKK	Q719N1, Q9UBP0, A2VDN5
PGKMDKGEHRQERRDRPY	Q01844, Q61545
PKKGDYDKTD	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
PPNKRMMKVKH	Q8BN78
PPRIYPQLPSAPT	P0C799
PQRSPFPKSSVKR	SeqNLS
PRPRKVPR	P0C799
PRRRVQRKR	SeqNLS, Q5R448, Q5TAQ9
PRRVRLK	Q58DJ0, P56477, Q13568
PSRKRP	Q62315, Q5F363, Q92833
PSSKKRKV	SeqNLS
PTKKRVK	P07664
QRPGPYDRP	SeqNLS
RGKGGKGLGKGGAKRHRK	SeqNLS
RKAGKGGGGHKTTKKRSA KDEKVP	B4FG96
RKIKLKRAK	A1L3G9
RKIKRKRAK	B9X187

RKKEAPGPREELESRGR	O35126, P54258, Q5IS70, P54259
RKKRKKGK	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
RKKSIPLSIKLNLRKHKRKK NKITR	Q9C0C9
RKLVKPKNTKMKTCLRTP Y	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,
RKRSPKDKKEKDLGAGKR RKT	Q7RTP6
RKRTPRVDGQTGENDMNK RRRK	O94851
RLPVRRRRRR	P04499, P12541, P03269, P48313, P03270
RLRFRKPKSK	P69469
RQQRKR	Q14980
RRDLNSSFETSPKKVK	Q8K3G5
RRDRAKLR	Q9SLB8
RRGDGRRR	Q80WE1, Q5R9B4, Q06787, P35922
RRGRKRKAQEKQ	Q812D1, Q5XXA9, Q99JF8, Q8MJG1, Q66T72, O75475
RRKKRR	Q0VD86, Q58DS6, Q5R6G2, Q9ERI5, Q6AYK2, Q6NYC1
RRKRKSEDMDSVESKRRR	Q7TT18
RRKRSR	Q99PU7, D3ZHS6, Q92560, A2VDM8
RRPKGKTLQKRKPK	Q6ZN17
RRRGFERFGPDNMGRKRK	Q63014, Q9DBR0
RRRGKNKVAAQNCRK	SeqNLS
RRRKRR	Q5FVH8, Q6MZT1, Q08DH5, Q8BQP9
RRRQKQKGGASRRR	SeqNLS
RRRREGPRARRR	P08313, P10231

RRTIRLKLVDKCDRSCKIQ KKNRNKCQYCRFHKCLSVG MSHNAIRFGRMPRSEKAKL KAE	SeqNLS
RRVPQRKEVSRCKCRK	Q5RJN4, Q32L09, Q8CAK3, Q9NUL5
RVGGRRQAVECIEDLLNEP GQPLDLSCKRPRP	P03255
RVVKLRIAP	P52639, Q8JMN0
RVVRRR	P70278
SKRKTKISRKTR	Q5RAY1, O00443
SYVKTVPNRTRTYIKL	P21935
TGKNEAKKRKIA	P52739, Q8K3J5, Q5RAU9
TLSPASSPSSVSCPVIPASTD ESPGSALNI	SeqNLS
VSKKQRTGKKIH	P52739, Q8K3J5, Q5RAU9
SPKKKRKVE	
KRTAD GSEFE SPKKKRKVE	
PAAKRVKLD	
PKKKRKV	
MDSLLMNRKFLYQFKNVR WAKGRRETYLC	
SPKKKRKVEAS	
MAPKKRKVGHRGVP	

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 (SEQ ID NO: 1598), wherein the spacer is bracketed. Another exemplary bipartite NLS has the sequence PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 1600). 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 sequence 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 or 3 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 KRASSQALGTIPKRRSSSRFIKRRK (SEQ ID NO: 1530). In some embodiments, the nucleolar localization signal may be derived from nuclear factor- κ B-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 Gene Writer™ 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 Gene Writer™ 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

5 RNA encoding the Gene Writer™ polypeptide is present in a non-target cell, it would be bound by the miRNA, and when the RNA encoding the Gene Writer™ 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 Gene Writer™ polypeptide may reduce production of the Gene Writer™

10 polypeptide, e.g., by degrading the mRNA encoding the polypeptide or by interfering with translation. Accordingly, in such embodiments the Gene Writer would add to/edit the genome of target cells more efficiently than it edits the genome of non-target cells, e.g., the heterologous object sequence would be inserted into the genome of target cells more efficiently than into the genome of non-target cells, or an insertion or deletion is produced more efficiently in target cells

15 than in non-target cells. A system having a microRNA binding site in the RNA encoding the Gene Writer™ 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.” In some embodiments, e.g., for liver indications, a miRNA is selected from Table 4 of

20 WO2020014209, incorporated herein by reference.

In some embodiments, the DNA encoding a Gene Writer polypeptide 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 Gene Writer™ system. For

25 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. A system having a tissue-specific promoter sequence in the DNA of the polypeptide may also be used in combination with a microRNA binding site, e.g., in the template RNA or a nucleic acid encoding a Gene Writer™ protein, e.g., as described herein. A system having a tissue-specific promoter sequence in the

30 DNA encoding the Gene Writer polypeptide may also be used in combination with a DNA encoding the RNA template driven by a tissue-specific promoter, e.g., to achieve higher levels of

RNA template in target cells than in non-target cells. In some embodiments, e.g., for liver indications, a tissue-specific promoter is selected from Table 3 of WO2020014209, incorporated herein by reference.

A skilled artisan can, based on the Accession numbers and/or sequences provided in
 5 Tables 1 and 3, determine the nucleic acid and corresponding polypeptide sequences of each retrotransposon or virus, and domains thereof, 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>.

10 Tables 1 and 3 herein provide the sequences of exemplary transposons or viruses, including the amino acid sequence(s) of the retrotransposase, reverse transcriptase, DNA-binding domain, and/or endonuclease domain; sequences of 5' and 3' untranslated regions to allow a polypeptide, e.g., the retrotransposase to bind the template RNA; and/or the full transposon nucleic acid sequence. In some embodiments, a 5' UTR contained in or referenced by Tables 1
 15 or 3 allows a polypeptide, e.g., the retrotransposase, to bind the template RNA. In some embodiments, a 3' UTR contained in or referenced by Tables 1 or 3 allows a polypeptide, e.g., 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 1 or 3 herein, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity
 20 thereto. In some embodiments, the system further comprises one or both of a 5' or 3' untranslated region contained in or referenced by Tables 1 or 3 (or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto), e.g., from the same transposon 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'
 25 untranslated region contained in or referenced by Tables 1 or 3, e.g., a segment of the full transposon 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 system or method described herein involves a 3' UTR, 5' UTR, or both from a retrotransposon of Table 3. In some embodiments, the 3' UTR, 5' UTR, or both,
 30 has a sequence comprising a portion of the full retrotransposon DNA sequence shown in column 5 of Table 3 of International Application PCT/US2019/048607, which is incorporated by

reference herein in its entirety, including Table 3. In some embodiments, the nucleic acid sequence or amino acid sequence has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to the sequence in Table 3 of PCT/US2019/048607.

In some embodiments, a system or method described herein involves a nucleic acid sequence or amino acid sequence of a retrotransposon described in Table 1 or Table 2 of International Application PCT/US2019/048607, which is incorporated by reference herein in its entirety, including Tables 1 and 2. In some embodiments, the nucleic acid sequence or amino acid sequence has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to the sequence of a retrotransposon described in said Table 1 or Table 2 of PCT/US2019/048607.

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. A skilled artisan can, based on the Accession numbers provided herein, align polypeptides or nucleic acid sequences, 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.

Thermostable Gene Writer™ Systems

While not wishing to be bound by theory, in some embodiments, retrotransposases that evolved in cold environments may not function as well at human body temperature. This application provides a number of thermostable Gene Writer™ systems, including proteins derived from avian retrotransposases. Exemplary avian transposase sequences in Table 3 include those of *Taeniopygia guttata* (zebra finch; transposon name R2-1_TG), *Geospiza fortis* (medium ground finch; transposon name R2-1_Gfo), *Zonotrichia albicollis* (white-throated sparrow;

transposon name R2-1_ZA), and *Tinamus guttatus* (white-throated tinamou; transposon name R2-1_TGut).

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 temperature (e.g., 25°C).

5 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
10 activity at 25°C under otherwise similar conditions.

In some embodiments, a Gene Writer™ polypeptide (e.g., a sequence of Table 1 or 3 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
15 embodiments, a Gene Writer™ polypeptide described herein is functional at 37°C. In some embodiments, a Gene Writer™ 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 Gene Writer™ polypeptide described herein has greater activity in a human cell than in a zebrafish cell.

In some embodiments, a Gene Writer™ polypeptide is active in a human cell cultured at
20 37°C, e.g., using an assay of Example 6 or Example 7 of PCT/US2019/048607 which are hereby incorporated by reference.

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 of FuGENE® HD transfection
25 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 Gene Writer™ protein, (d) sequence encoding the Gene Writer™ protein, (e) sequence encoding a 3' untranslated region that binds the Gene Writer™ protein (f) 100 bp of sequence homologous to the 100 bp downstream of the target site,
30 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/US2019/048607 which is hereby incorporated by reference.

In some embodiments, the Gene Writer™ polypeptide results in insertion of the
5 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,
10 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 Gene Writer™ 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-
15 truncated, as measured by an assay described herein, e.g., an assay of Example 6 and Figure 8 of PCT/US2019/048607 which are hereby incorporated by reference. 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-
20 truncated using an assay comprising amplification with a forward primer situated 565bp from the end of the element (e.g., a wild-type transposon sequence, e.g., of *Taeniopygia guttata*) 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
25 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
30 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/US2019/048607 which is hereby incorporated by

reference. 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/US2019/048607.

5 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.

10 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

15 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 Gene Writer™ 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

20 PCT/US2019/048607 which is hereby incorporated by reference. In some embodiments, the Gene Writer™ 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/US2019/048607, the target site represents more than 50%, 60%, 70%, 80%, 90%, or 95% of binding events of the Gene Writer™ or DNA-binding domain thereof to human

25 genomic DNA.

Genetically engineered, e.g., dimerized Gene Writer™ Systems

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

30 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
5 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).

Additional domains:

10 The Gene Writer™ polypeptide comprises the functions necessary to bind a target DNA sequence and template nucleic acid (e.g., template RNA), nick the target site, and write (e.g., reverse transcribe) the template into DNA, resulting in a modification of the target site. In some embodiments, additional domains may be added to the polypeptide to enhance the efficiency of the process. In some embodiments, the Gene Writer™ polypeptide may contain an additional
15 DNA ligation domain to join reverse transcribed DNA to the DNA of the target site. In some embodiments, the polypeptide may comprise a heterologous RNA-binding domain. In some embodiments, the polypeptide may comprise a domain having 5' to 3' exonuclease activity (e.g., wherein the 5' to 3' exonuclease activity increases repair of the alteration of the target site, e.g., in favor of alteration over the original genomic sequence). In some embodiments, the
20 polypeptide may comprise a domain having 3' to 5' exonuclease activity, e.g., proof-reading activity. In some embodiments, the writing domain, e.g., RT domain, has 3' to 5' exonuclease activity, e.g., proof-reading activity.

In some embodiments, the polypeptide does not comprise an RNase H domain. In some embodiments, the polypeptide comprises an RNaseH domain endogenous to one of the other
25 domains. In some embodiments, the polypeptide comprises an RNase H domain that is heterologous to the other domains. In some embodiments, the polypeptide comprises an inactivated endogenous RNaseH domain.

In some embodiments, a Gene Writer as described herein comprises a polypeptide associated with a guide RNA (gRNA). In certain embodiments, the gRNA is comprised in the
30 template nucleic acid molecule. In other embodiments, the gRNA is separate from the template nucleic acid molecule. In some embodiments wherein the gRNA is comprised in the template

nucleic acid molecule, the template nucleic acid molecule further 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). In embodiments, the gRNA spacer comprises a sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a nucleic acid sequence comprised in the target nucleic acid molecule. In embodiments, the gRNA spacer directs Cas domain (e.g., Cas9) activity at the nucleic acid sequence comprised in the target nucleic acid molecule. In some embodiments wherein the gRNA is comprised in the template nucleic acid molecule, the template nucleic acid molecule further comprises a primer binding site (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 3' end). In embodiments, the primer binding site comprises a nucleic acid sequence comprising at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a nucleic acid sequence positioned at the 5' end (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, or 50 nucleotides) of a nick site on the target nucleic acid molecule. In embodiments, binding of the primer binding site to the target nucleic acid molecule operates to prime TPRT.

15

Template Nucleic Acid Component of Gene Writer™ gene editor system

The Gene Writer™ systems described herein can modify a host target DNA site using a template nucleic acid sequence. In some embodiments, the Gene Writer™ systems described herein transcribe an RNA sequence template into host target DNA sites by target-primed reverse transcription (TPRT). 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. The Gene Writer™ system can also delete a sequence from the target genome or introduce a substitution using an object sequence. 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, a Gene Writer system comprises a template nucleic acid (e.g., RNA or DNA) molecule. In some embodiments, the template nucleic acid molecule comprises a 5' homology region and/or a 3' homology region. In some embodiments, the 5' homology region comprises a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%,

30

or 100% sequence similarity 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 region 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 region is heterologous to the remainder of the template nucleic acid molecule. In some embodiments, the 3' homology region 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 region 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 region to the target nucleic acid molecule permits initiation of target-primed reverse transcription (TPRT), e.g., with the 3' homology region acting as a primer for TPRT. 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, a template nucleic acid (e.g., template RNA) comprises a heterologous object sequence. In some embodiments, the heterologous object sequence may be transcribed by the RT domain of a Gene Writer™ polypeptide, e.g., thereby introducing an alteration into a target site in genomic DNA. In some embodiments, the heterologous object sequence is at least 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 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 nucleotides (nts) in length, or at least 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 in length. In some embodiments, the heterologous object sequence is no more than 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 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, 1,000, or 2000 nucleotides (nts) in length, or no more than 20, 15, 10, 9, 8, 7, 6, 5, 4, or 3 kilobases in length. In some embodiments, the heterologous object sequence is 30-1000, 40-1000, 50-1000, 60-1000, 70-1000, 74-1000, 75-1000, 76-1000, 77-1000, 78-1000, 79-1000, 80-1000, 85-1000, 90-1000, 100-1000, 120-1000, 140-1000, 160-1000, 180-1000, 200-1000, 500-1000, 30-500, 40-500, 50-500, 60-500, 70-500, 74-500, 75-500, 76-500, 77-500, 78-500, 79-500, 80-500, 85-500, 90-500, 100-500, 120-500, 140-500, 160-500, 180-500, 200-500, 30-200, 40-200, 50-200, 60-200, 70-200, 74-200, 75-200, 76-200, 77-200, 78-200, 79-200, 80-200, 85-200, 90-200, 100-200, 120-200, 140-200, 160-200, 180-200, 30-100, 40-100, 50-100, 60-100, 70-100, 74-100, 75-100, 76-100, 77-100, 78-100, 79-100, 80-100, 85-100, or 90-100 nucleotides (nts) in length, or 1-20, 1-15, 1-10, 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, 2-20, 2-15, 2-10, 2-9, 2-8, 2-7, 2-6, 2-5, 2-4, 2-3, 3-20, 3-15, 3-10, 3-9, 3-8, 3-7, 3-6, 3-5, 3-4, 4-20, 4-15, 4-10, 4-9, 4-8, 4-7, 4-6, 4-5, 5-20, 5-15, 5-10, 5-9, 5-8, 5-7, 5-6, 6-20, 6-15, 6-10, 6-9, 6-8, 6-7, 7-20, 7-15, 7-10, 7-9, 7-8, 8-20, 8-15, 8-10, 8-9, 9-20, 9-15, 9-10, 10-15, 10-20, or 15-20 kilobases in length. In some embodiments, the heterologous object sequence is 10-100, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, or 10-20 nt in length, e.g., 10-80, 10-50, or 10-20 nt in length, e.g., about 10-20 nt in length. In some embodiments, a template RNA comprises a

sequence as listed in Table 43, or a sequence with at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

In certain embodiments, the template nucleic acid comprises a customized RNA sequence
5 template which 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 linked
10 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
15 homologous to the RLE retrotransposase, be engineered to contain heterologous coding
sequences, or combinations thereof.

The template nucleic acid (e.g., template RNA) may have some homology to the target
DNA. In some embodiments, the template nucleic acid (e.g., template RNA) 3' target homology
domain may serve as an annealing region to the target DNA, such that the target DNA is
20 positioned to prime the reverse transcription of the template nucleic acid (e.g., template RNA). In
some embodiments the template nucleic acid (e.g., 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 nucleic acid (e.g., template RNA) has at least 2, 3, 4, 5, 6, 7, 8, 9, 10,
25 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 nucleic acid (e.g.,
template RNA). In some embodiments the template nucleic acid (e.g., 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
30 at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% homology to the 3'

sequence of a non-LTR retrotransposon, e.g., a non-LTR retrotransposon described herein, e.g. a non-LTR retrotransposon in Table 1 or 3.

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 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 nucleic acid (e.g., 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 non-LTR retrotransposon, e.g., a non-LTR retrotransposon described herein.

The template nucleic acid (e.g., template RNA) of the system typically comprises an object sequence (e.g., a heterologous object sequence) for insertion into a target DNA. The
5 object sequence may be coding or non-coding.

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
10 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 second 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' target homology domain. In some embodiments, when
15 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. For example, in some embodiments a first RNA comprises a first conjugating domain and a second RNA comprises a second conjugating domain, and the first and second conjugating domains are capable of
20 hybridizing to one another, e.g., under stringent conditions. In some embodiments, the stringent conditions for hybridization include hybridization in 4x sodium chloride/sodium citrate (SSC), at about 65°C, followed by a wash in 1xSSC, at about 65°C.

In some embodiments, the object sequence may contain an open reading frame. In some embodiments the template nucleic acid (e.g., template RNA) has a Kozak sequence. In some
25 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
30 entirety. Exemplary splice acceptor site sequences are known to those of skill in the art and include, by way of example only, CTGACCCTTCTCTCTCTCCCCAGAG (SEQ ID NO:

1601) (from human HBB gene) and TTTCTCTCCCACAAG (SEQ ID NO: 1602) (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, 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 Gene Writer™ system. For instance, the microRNA binding site can be chosen on the basis that it 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 its activity, e.g., may interfere with insertion of the heterologous object sequence into the genome. Accordingly, the system would edit the genome of target cells more efficiently than it edits the genome of non-target cells, e.g., the heterologous object sequence would be inserted into the genome of target cells more efficiently than into the genome of non-target cells, or an insertion or deletion is produced more efficiently in target cells than in 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 Gene Writer™ polypeptide, wherein expression of the Gene Writer™ 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, e.g., for liver indications, a miRNA is selected from Table 4 of WO2020014209, incorporated herein by reference.

In some embodiments, the object sequence may contain a non-coding sequence. For example, the template nucleic acid (e.g., template RNA) may comprise a regulatory element, e.g., a promoter or enhancer sequence or miRNA binding site. In some embodiments, integration of the object sequence at a target site will result in upregulation of an endogenous gene. In some embodiments, integration of the object sequence at a target site will result in downregulation of an endogenous gene. In some embodiments the template nucleic acid (e.g., 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, 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 Gene Writer™ 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 Gene Writer™ 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 Gene Writer™ polypeptide, driven by a tissue-specific promoter, e.g., to achieve higher levels of Gene Writer™ protein in target cells than in non-target cells. In some embodiments, e.g., for liver indications, a tissue-specific promoter is selected from Table 3 of WO2020014209, incorporated herein by reference.

In some embodiments, a Gene Writer system, e.g., DNA encoding a Gene Writer polypeptide, DNA encoding a template RNA, or DNA or RNA encoding a heterologous object sequence, is designed such that one or more elements is operably linked to a tissue-specific promoter, e.g., a promoter that is active in T-cells. In further embodiments, the T-cell active promoter is inactive in other cell types, e.g., B-cells, NK cells. In some embodiments, the T-cell active promoter is derived from a promoter for a gene encoding a component of the T-cell receptor, e.g., TRAC, TRBC, TRGC, TRDC. In some embodiments, the T-cell active promoter is derived from a promoter for a gene encoding a component of a T-cell-specific cluster of differentiation protein, e.g., CD3, e.g., CD3D, CD3E, CD3G, CD3Z. In some embodiments, T-cell-specific promoters in Gene Writer systems are discovered by comparing publicly available gene expression data across cell types and selecting promoters from the genes with enhanced expression in T-cells. In some embodiments, promoters may be selected depending on the desired expression breadth, e.g., promoters that are active in T-cells only, promoters that are active in NK cells only, promoters that are active in both T-cells and NK cells.

15 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 nucleic acid (e.g., template RNA) comprises a site that coordinates epigenetic modification. In some embodiments the template nucleic acid (e.g., template RNA) comprises a chromatin insulator. For example, the template nucleic acid (e.g.,
20 template RNA) comprises a CTCF site or a site targeted for DNA methylation.

In some embodiments the template nucleic acid (e.g., 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).

25 In some embodiments the object sequence of the template nucleic acid (e.g., template RNA) is inserted into a target genome in an endogenous intron. In some embodiments the object sequence of the template nucleic acid (e.g., 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.

30 In some embodiments, the object sequence of the template nucleic acid (e.g., template RNA) is inserted into the target genome in a genomic safe harbor site, such as AAVS1, CCR5,

ROSA26, or albumin locus. In some embodiments, a Gene Writer is used to integrate a CAR into the T-cell receptor α constant (*TRAC*) locus (Eyquem et al Nature 543, 113-117 (2017)). In some embodiments, a Gene Writer is used to integrate a CAR into a T-cell receptor β constant (*TRBC*) locus. Many other safe harbors have been identified by computational approaches (Pellenz et al Hum Gen Ther 30, 814-828 (2019)) and could be used for Gene Writer-mediated integration. In some embodiments, the object sequence of the template nucleic acid (e.g., template RNA) is added to the genome in an intergenic or intragenic region. In some embodiments, the object sequence of the template nucleic acid (e.g., 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 nucleic acid (e.g., 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 nucleic acid (e.g., 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).

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, 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. gene expression unit without UTRs used to bind the Gene Writer™ protein).

Alternative orientations of the template RNA motifs can be employed, e.g., to limit target site integration to the desired genetic payload. In some embodiments, the polypeptide association domains may be located 5' of the desired template sequence. For example, the heterologous object sequence may be located downstream of the 5' UTR and 3' UTR, giving the 5'-3' orientation 5'UTR-3'UTR-(heterologous object sequence). In other embodiments, only the 3' UTR is added upstream of the heterologous object sequence. For example, giving the 5'-3' orientation 3'UTR-(heterologous object sequence). In certain embodiments, the polypeptide coding region and the heterologous object sequence may be encoded on the same molecule, but where the 5' UTR (e.g., 5' UTR from R2 retrotransposon) occurs between the two regions, e.g., giving the 5'-3' orientation (polypeptide coding sequence)-5'UTR-(heterologous object sequence).

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 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 conformation where the blocking domain is not hybridized or or not substantially hybridized to the gRNA. In some embodiments, in the first conformation the 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 conformation the 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 conformation can 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.,

5 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 40 or Table 37 or a functional fragment thereof, or a sequence having at least 70%, 75%, 80%, 85%, 90%,

10 95%, 96%, 97%, 98%, 99% identity thereto.

Table 40 CRISPR/Cas Proteins, Species, and Mutations

Variant	Parental Host	Protein Sequence	Nickase Mutation
Nme2Cas9	Neisseria meningitidis	MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRFERAEVPKTGDSLAMARRLAR SVRRLTRRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAADRLKLTPLWSAVL LHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFESGHIRNQR GDYSHTFSRKDLQAEILLFEKQKEFGNPHVSGGLKEGIELTLLMTQRPALSGDAVQKMLGHCTFE PAEPKAAKNTYTAERFIWLTCLNLRILEQGSERPLTDERATLMDEPYRKSKLTYAQARKLLGL EDTAFFKGLRYGKDNAAEASTLMMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIGTAFSLFKT DEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKN TEEKIYLPPIPADEIRNPVLRALSQARKVINGVRRYGSPIRIHETAREVGKSFKDRKEIEKRQEE NRKDREKAAAKFREYFPNFVGEPSKSDILKLRLYEQQHGKCLYSGKEINLVRLEKGYVEIDHAL PFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQ KFDEDFGKECNLNDTRYVNRFLCQFVADHILLTGKGRRVFASNGQITNLLRGFWGLRKVRAEN DRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQE VMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRS AKRFVKHNEKISVKRVWLTEIKLADLENMVMNYKNGREIELYEALKARLEAYGGNAQAFDPKDN PFYKGGQLVKAVRVEKTQESGVLLNKKNAITYADNGDMVRVDVFCVDDKKGKQYFIVPIYA WQVAENILPDIDCKGYRIDDSYTFCSLHKYDLIAFQKDEKSKVEFAYYINCDSNGRFFYLAWHD KGSKEQQFRISTQNLVLIQKYQVNLGKEIRPCRLKKRPPVR	N611A
PpnCas9	Pasteurella pneumotrica	MQNNPLNYILGLDIGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVAKTGESLALSRRLARSSRL LIKRRRAERLKKAKRLLKAEKILHSIDEKLPINVWQLRVKGLKEKLERQEWAAVLLHLKSHRGYLS QRKNEGKSDNKELGALLSGIASNHQMLQSSEYRTPAEIAVKKFQVEEGHIRNQRGSYTHTFSRLLD LAEMELLFQRQAEALGNSYTSITLLENLTALLMWQKPALAGDAILKMLGKCTFEPSEYKAAKNSY SAERFVWLTCLNLRILENGTERALNDNERFALLEQPYEKSCLTYAQVRAMLALSNAIFKGVRY LGEDKKTVESKTTLIEMKFYHQIRKTLGSALKEWNEKLGNSDLLDEIGTAFSLYKTDICRYL EGKLPVRLNALLNENLNFDFKTLQSLKALHQILPLMLQGQRYDEAVSAIYGDHYGKKSTETTRLLP TIPADEIRNPVLRITLQARKVINAVVRLYGSPIRIHETAREVGKSYQDRKKLEKQEDNRKQRE SAVKKFKEMFPHFVGEPSKSDILKMLRYELQAKCLYSGKSLELHRLLEKGYVEVDHALPFSRT WDDSFNNKVLVLANENQNKGNLTPYEWLDGKNNSEWRQHFVVRVQTSFGSYAKKQRIILNHLKD EKGFIERNLNDTRYVARFLCNFIADNMLLVGKGKRVFASNGQITALLRHRWGLQKVREQNDRH HALDAVVVACSTVAMQQKITRFVRYNEGNVFSGERIDRETGEIPLHFPSPWAFKENVIRFSEN PKLELENRLPDYPQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSARKLNEGLSVLKVPLTQLKL SDLERMVNRDREIALYESLKARLEQFGNDPAKFAEPFYKKGALVKAVRLEQTQKSGVLVRDG NGVADNASMVRVDVFTKGGKYFLVPIYTWQVAKGILPNRAATQGKDENDWDIMDEMATFQFSL CQNDLIKLVTKKKTIFGYFNGLNRAATSNINIKHEHDLKSKGKLGILEVGVKLALISLEKYQVDELG KNIRPCRPTKRQHVR	N605A
SauCas9	Staphylococcus aureus	MKKNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKLEANVENNEGRRSKRGARRLKRHHRIQR VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLESEEFSAALLHLAKRRGVHNVNEVEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRSINRFKTSDYVKEAKQLLKVKQAYHQLDQS	N580A

		FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYKFIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIENAEELLDQIAKILTYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVKKVVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIELAREKNSKDAQKMINEMQKRNRTNERIEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSCKGNRTPPFQYLSSSDSKISY ETFKKHILNLAAGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLDVVKVSINGGFTSFLRRKWKFKKERNKGKHHAEADALIANADFIKWKLDKAKKVMENQ MFEKQAESMPEIETEYEQYKEIFITPHQIKHIKDFKDYKYSHRVDKPNRELINDTLYSTRKDDKGN TLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDENPLYKYEEETG NYLTKYSKKDNGPVIKKIKYYGNKLNLAHLDDITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFV TVKNLDVIKKENYEVNSKCYEEAKLLKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLNRI EVNMIDITYREYLENMNDKRPPRIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
SauCas9- KKH	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRIQR VKKLLFDYNLLTDHSELGINPYEARVKGSLQKLSSEEFSAALLHLAKRRGVHNVNEVEDTNE LSTKEQISRNKSALEEKYVAELQLERLKKDGEVRGSINRFTSDYVKEAKQLLKVKQAYHQLDQS FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYKFIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIENAEELLDQIAKILTYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVKKVVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIELAREKNSKDAQKMINEMQKRNRTNERIEIIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSCKGNRTPPFQYLSSSDSKISY ETFKKHILNLAAGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLDVVKVSINGGFTSFLRRKWKFKKERNKGKHHAEADALIANADFIKWKLDKAKKVMENQ MFEKQAESMPEIETEYEQYKEIFITPHQIKHIKDFKDYKYSHRVDKPNRELINDTLYSTRKDDKGN TLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDENPLYKYEEETG NYLTKYSKKDNGPVIKKIKYYGNKLNLAHLDDITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFV TVKNLDVIKKENYEVNSKCYEEAKLLKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLNRI EVNMIDITYREYLENMNDKRPPRIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
SauriCas9	Staphylococcus aureus	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNRRSKRGARRLKRRI HRLNRVKDLLADYQMDLNNVPKSTDPYTIRVKGLEPLTKEEFAIALHIAKRRGLHNISVSMGD EEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFTEDFVKEVKQLCETQQRQ YHNIDDOFIQYIDL VSTRREYFEGPGNGSPYGDGDLKWKYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKELEYEYKHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRTKS GKQPQTSFKLYHDLKNIFEQAKYLEDVEMLDLAKILTYQDEISIKKALDQPELLTESEKQIAQL TGTYTGHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKKVEMSEIDSIPITLVDFILSPVVKRAFI QSIKVINAVINRFGLPEDIIELAREKNSKDRRKFNKLQKQNEATRKKIEQLLAKYGNNAKYMIE KIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVVDHIIPRSVSFDNSLNKVLVKQSENSCKGNRTP YQYLSNESKISYNQFKQHILNLSAKDRISKKKRDMLEERDINKFEVQKEFINRNLVDTRYATR ELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHNRHGYKHHAEADALVIANADFLKTH KALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNKQFRDFKYSHRVDKPNRQLIN DTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFETERPKILMYQHDPKTFEKLMTILNQYAEAK NPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSKSRFDI YKCEQGYKMVSIGYLDVLKKNYYPKDKYEAQKQKKIKESDLFVGSFYNDLIMYEDELFR VIGVNSDINNVELNMVDITYKDFCEVNNVTGEKRIKKTIGKRVVLIEKYTTDILGNLYKTPLPKK QLIFKRGEL	N588A
SauriCas9 -KKH	Staphylococcus aureus	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNRRSKRGARRLKRRI HRLNRVKDLLADYQMDLNNVPKSTDPYTIRVKGLEPLTKEEFAIALHIAKRRGLHNISVSMGD EEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFTEDFVKEVKQLCETQQRQ YHNIDDOFIQYIDL VSTRREYFEGPGNGSPYGDGDLKWKYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKELEYEYKHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRTKS GKQPQTSFKLYHDLKNIFEQAKYLEDVEMLDLAKILTYQDEISIKKALDQPELLTESEKQIAQL TGTYTGHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKKVEMSEIDSIPITLVDFILSPVVKRAFI QSIKVINAVINRFGLPEDIIELAREKNSKDRRKFNKLQKQNEATRKKIEQLLAKYGNNAKYMIE KIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVVDHIIPRSVSFDNSLNKVLVKQSENSCKGNRTP YQYLSNESKISYNQFKQHILNLSAKDRISKKKRDMLEERDINKFEVQKEFINRNLVDTRYATR ELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHNRHGYKHHAEADALVIANADFLKTH KALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNKQFRDFKYSHRVDKPNRQLIN DTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFETERPKILMYQHDPKTFEKLMTILNQYAEAK NPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSKSRFDI YKCEQGYKMVSIGYLDVLKKNYYPKDKYEAQKQKKIKESDLFVGSFYNDLIMYEDELFR VIGVNSDINNVELNMVDITYKDFCEVNNVTGEKHIKKTIGKRVVLIEKYTTDILGNLYKTPLPKK PQLIFKRGEL	N588A
ScaCas9- Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPKSKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRRYTRRKNRIRYLQEIFANEMAKLDDSFQRLSEESFLVEEDKKNERHPFIGNLADEVAYHRN YPTIYHLRKKLADSPEKADRLIYALAHIIKFRGHFLIEGKLNANESDVAKLFYQLIQTYNQLFEE SPLDEIEVDAGILSARLSKSKRLEKLIJAVFPNEKKNGLFGNIALALGLTPNFKSNFDLTEDAKLQL SKDTYDDDLDELLGQIGDQYADLFSAKKNLSDAILSDILRSNSEVTAPLASMVKRYDEHHQD LALLKTLVRQFPPEKYAEIFKDDTKNGYAGYVAGDKKLRKRSGLATEEEFYKFKPILEKMDGA EELLAKLNRDILLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYFPLKENREKIEKILTFRIPYYVGP	N872A

		LARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDQELPNKKVLPKHSLLYEFYTVYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIGVEDRFNASLGTYYHDLKKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIIKKGILQTVKIVDELVKVMGHPENIVIAMARENQTTTKGLQQSRERKKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLRVSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGSEADKA GFIRKQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSCLVSDFRKDFQLYKVRDINN YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKRFFYSNIMN FFKTEVKLANGEIRKRPLIETNGETGEVWNKEKDFATVRKVLAMPQVNVKKTVEVQTGGFSKES ILSKRESAKLIPRKKGWDRTRKYGGFGSPTVAYSILVVAKEVGKAKKLLKSVKVLVGITIMEKGSY EKDPGFLEAKGYKDIKKELIFKLPKYSLEFELNGRRRMLASAKELQKANELVLPQHILVRLLYYTQ NISATTGSNNLYGIEQHREEFKEIFEKIIDFSEKYILKNKVNLSNLKSSFDEQFAVSDSILLSNSFVSL LKYTSFGASGGFTFLDLVDVKQGRRLRYQTVTEVLDATLIHQSIITGLYETRIDLSQLGGD	
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTIYHLRKKLVDSITDKADLRLLIYALAHMIKFRGHFLIEGDLNPDNSDVDFLQVLTQYNNQLFEE NPINASGVDAKAILSAKLSRRLLENLIAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKIQ LSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYVVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKHSLLYEFYTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYY HDLLKKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYYAHLFDDKVMKQLKRRRYTGWGR LRSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIIKKGILQTVKIVDELVKVMGRHKPENIVIAMARENQTTQKQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFILKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKQRL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLAMPQVNVKKTVEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFGSPTVAYSILVVAKEVGKSKKLLKSVKELLGITIMERSSEKFNPIDFLEA KGYKEVKKDLIILPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLEKGGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDDKVL SAYNKHRRDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKRYTSKEVLDAITLIHQSIITGLYETRIDLSQLGGD	N863A
SpyCas9-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTIYHLRKKLVDSITDKADLRLLIYALAHMIKFRGHFLIEGDLNPDNSDVDFLQVLTQYNNQLFEE NPINASGVDAKAILSAKLSRRLLENLIAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKIQ LSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYVVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKHSLLYEFYTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYY HDLLKKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYYAHLFDDKVMKQLKRRRYTGWGR LRSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIIKKGILQTVKIVDELVKVMGRHKPENIVIAMARENQTTQKQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFILKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKQRL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLAMPQVNVKKTVEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFGSPTVAYSILVVAKEVGKSKKLLKSVKELLGITIMERSSEKFNPIDFLEA KGYKEVKKDLIILPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLEKGGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDDKVL SAYNKHRRDKPIREQAENIHLFTLT NLGAPRAFKYFDTTIDRKVYRSTKEVLDAITLIHQSIITGLYETRIDLSQLGGD	N863A
SpyCas9-SpRY	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTIYHLRKKLVDSITDKADLRLLIYALAHMIKFRGHFLIEGDLNPDNSDVDFLQVLTQYNNQLFEE NPINASGVDAKAILSAKLSRRLLENLIAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKIQ LSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYVVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKHSLLYEFYTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYY HDLLKKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYYAHLFDDKVMKQLKRRRYTGWGR LRSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIIKKGILQTVKIVDELVKVMGRHKPENIVIAMARENQTTQKQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFILKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAGFIKQRL	N863A

		VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTTEVQTGGFSKESIRPKRNSDK LIARKKDWDPKKYGGFLWPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEKPNIDFLE AKGYKEVKKDLIILPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFYLYASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLKDVLASYNKHRDKPIREQAENIHLFTLT RLGAPRAFKYFDDTTIDPKQYRSTKEVLDAITLHQISITGLYETRIDLSQLGGD	
St1Cas9	Streptococcus thermophilus	MSDLVLGLDIGIGSVGILNKTGTEIHKNSRIFFAAQAENNLVRRTRNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHKGISYLDASDDGNSSVG DYAQIVKENSQKLETKTPGQIQLERYQTYGQLRGDFTVEKDGKXHLINVFPTSAIRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYHHPGNEKSRDYGRYRTSGETLDNIFGILIGKCTFYDEFER AAKASYTAQEFNLLNDLNNLTVPETKKSKEQKNQIINYVKNKEAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTFEAYRKMKLTLETLDIEQMDRETLDKLAYVLTNTEREGIEALEHEFAD GSFSQKQVDELQVRKANSSIFGKGWHNFSVKLMMEIPELYETSEEQMTILTRLGKQKTTSSSNK TKYIDEKLLTEIYNPVVAKSVRQAIKIVNAAIKYGDGDNVIMARETNEDEKKAQIKQKANK DEKDAAMLKAANQYNGKAEIHPHVFHGHKQLATKIRLWHQOGERCLYTGKTSIHDLINNSNQF EVDHILPLSTIFDDSLANKVLVYATANQEKQGRTPYQALDSMDDAWSFRELKAFVRESKTLSSNK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQTSQLRRH WGIEKTRDTYHHHAVDALIAASSQLNLWKKQKNTLVSYSEDQLDIETGELISDDEYKESVFKAP YQHFDVTLKSKEFEDSILFSYQVDSKFNKISDATIYATRQAKVKGDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHDPTFEKVIPEILENYPNKQINEKGKEVPCNPLFKYIEEHGYKY SKKGNPGEIKSLKYDSKLGNIHIDITPKDSNNKVVLSQSVPRADVYFNKTTGKYEILGLKYADL QFEKGTGTGYISQEKYNDIKKKEGVDSSEFKFTLYKNDLLL VKDTETKEQQLFRFLSRTMPKQK HYVELKPYDKQKFEQGEALIKVLGNVANSQCCKGLGKSNISYKVRTDVLGNQHIKNEGDKPK LDF	N622A
BlatCas9	Brevibacillus laterosporus	MAYTMGIDVGIASCWVAIVDLERQRIIDIGVRTFEKAENPKNGEALAVPRREARSSRRRLRRKKHR IERLKHMFVRNGLAVDIQHLEQLTRSQNEIDVWQLRVDGLDRMLTQKEWLRVLIIHLAQRGRFQS NRKTDGSSDDGQVLNVVTENDRLMEEDYRTVAEMMVKDEKFSCHKRNKNGNYHGVSRSLSL LVEIHTLFETQRQHNSLASKDFELEYVNIWSAQRPVATKDKQIEKMIGTCTFLPKEKRAPKASWHF QYFMLLQTHIRITNVQGTSLNKEEIEQVNMALTKSKVSYHDTRKILDLSEYQFVGLDYGKE DEKKVESKETIHLDDYHKLNFNEVELAKGETWEADDYDTVAAYALTFKDDDEDIRDLTKNY KDSKNRLVKNLANKETNELIGKVSTLSFRKVGHLKALRKIIPLEQGMITYDKACQAAGDFDQ GISKKKRSVVLVIDQISNPVNRALTQTRKVINAIKKYSGPETIHIETARELSKTFDERKNITKDY KENRDKNEHAKKHLSELGIINPTGLDIVKYKLWCEQQGRCMYNSQPISFERLKESGYTEVDHIIPY SRSMNDSYNNRVLMVTRENREKGNQTPFEYMGNDTQRWYEFQRTVTNPQIKKEKQNLKLG FTNRRELEMLERNLNDTRYITKYLSHFISTNLEFSPSDKKKKVVNTSGRITSHLSRWGLEKNRGQ NDLHHAMDIAIVATSDFSQVNTNYKRRERRELNGDDKFLPLPWKFFREEVIARLSPNPKQIEA LPNHFYSEDELADLQPIFVSRMPKRSITGEAHQAQFRRVVGKTKEGKNITAKKTALVDISYDKNGD FNMYGRETDPATYEAIKERYLEFGGNVKKAFSTDHLKPKKDGTKGPLIKSVRIMENKTLVHPVNK GKGVVYNSIVRTDVFQRKEKYLLPVYVTDVTKGLPNKVIVAKKGYHDWIEVDSDFTFLFSLY PNDLIFIRQNPKKKISLKKRIESHSISDSKEVQEHAYYKGVDSSTAIEFIHHDGSYYAKGVGVQNL DCFKEYQVDILGNFYKVKGEKRLELETSDSNHKGKDVNSIKSTSR	N607A
cCas9-v16	Staphylococcus aureus	MKRNILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRSKRGARRLRKRRRHRIQR VKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSSEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNKSALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVKQAYHQLDQS FIDTYIDLETRRTYIEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNAL DLNNLVITRDENEKLEYEYKFIENVFQKQKPTLQKIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIENAEILLDQIAKILTYQSSEDIQEELTNLSELQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIVINAIK KYGLPNDIIEELAREKNSKDAQKMINEMQKRNQRTNERIEEIRTTGKENAKYLIEIKLHDMQEGK CLYSLEAIPLEDLLNNPFIYEVDDHIIPRSVSDNSFNKVLVKQEENSCKGNRTPFYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLVDVKVSINGGFTSFLRRKWKFKKERNKGYKHHAEADALIANADFIKWKKLDKAKKVMENQ MFEKQAESMPETEYQYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDENPLYKYEEET GNYLTKYSKKDNPGVIKKIKYYGNKLNALHLDITDDYPNSRNKVVKLSLKPFRDVFYLDNGVYKF VTVKNLDDVIKENYEVNSKCYEEAKLKKISNQAEFIASFYKNDLIKINGELRYRIGVNSDKNNL IEVNMDITYREYLENMNDKRPPIHIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A
cCas9-v17	Staphylococcus aureus	MKRNILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRSKRGARRLRKRRRHRIQR VKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSSEEFSAALLHLAKRRGVHNVNEVEEDTGNE LSTKEQISRNKSALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVKQAYHQLDQS FIDTYIDLETRRTYIEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNAL DLNNLVITRDENEKLEYEYKFIENVFQKQKPTLQKIAKEILVNEEDIKGYRVTSTGKPEFTNLK VYHDIKDITARKEIENAEILLDQIAKILTYQSSEDIQEELTNLSELQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIVINAIK KYGLPNDIIEELAREKNSKDAQKMINEMQKRNQRTNERIEEIRTTGKENAKYLIEIKLHDMQEGK CLYSLEAIPLEDLLNNPFIYEVDDHIIPRSVSDNSFNKVLVKQEENSCKGNRTPFYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLVDVKVSINGGFTSFLRRKWKFKKERNKGYKHHAEADALIANADFIKWKKLDKAKKVMENQ MFEKQAESMPETEYQYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDDKG	N580A

		NTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDENPLYKYEEET GNYLTYSKKDNGPVIKKIKYYGNKLNALHLDITDDYPNSRNKVVKLKSLKPYRFDVYLDNGVYKF VTVKNLVDVIKENYEVNSKCYEEAKLLKISNQAEIFASFYKNDLIKINGELYRVIGVNNSTRNI VELNMIDITYREYLENMNDKRPPHIKTIASKTQSIKKYSTDILGNLYEVKSKKHPPHIKKG	
cCas9- v21	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRHHRIQR VKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEFSAALLHLAKRRGVHNVNEVEDTGNEL LSTKEQISRNKSALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVQKAYHQLDQS FIDTYIDLETRRTYYEGPGEPSFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYKFQIENVFVKQKKPTLKQIAKEILVNEEDIKGYRVTSYTGKPEFTNLK VYHDIKDITARKEIENAEELLDQIAKILTYQSSEDIQEELTNLSELQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVKPKVDLSQKKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIEELAREKNSKDAQKMINEMQKRNRQTNERIEEIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSCKGNRTPFYQLSSSDSKISY ETFKKHILNLAQKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLDVVKVSINGGFTSFLRRKWKFKERNKGKHHAEADALIANADFIKEWKLDKAKKVMENQ MFEKQAESMPETEYQYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDENPLYKYEEET GNYLTYSKKDNGPVIKKIKYYGNKLNALHLDITDDYPNSRNKVVKLKSLKPYRFDVYLDNGVYKF VTVKNLVDVIKENYEVNSKCYEEAKLLKISNQAEIFASFYKNDLIKINGELYRVIGVNSDDRNII ELNMIDITYREYLENMNDKRPPHIKTIASKTQSIKKYSTDILGNLYEVKSKKHPPHIKKG	N580A
cCas9- v42	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRHHRIQR VKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEFSAALLHLAKRRGVHNVNEVEDTGNEL LSTKEQISRNKSALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVQKAYHQLDQS FIDTYIDLETRRTYYEGPGEPSFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN DLNNLVITRDENEKLEYEYKFQIENVFVKQKKPTLKQIAKEILVNEEDIKGYRVTSYTGKPEFTNLK VYHDIKDITARKEIENAEELLDQIAKILTYQSSEDIQEELTNLSELQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVKPKVDLSQKKEIPTTLVDDFILSPVVKRSFIQSIKVINAIK KYGLPNDIIEELAREKNSKDAQKMINEMQKRNRQTNERIEEIRTTGKENAKYLIEKIKLHDMQEGK CLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSCKGNRTPFYQLSSSDSKISY ETFKKHILNLAQKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVN NLDVVKVSINGGFTSFLRRKWKFKERNKGKHHAEADALIANADFIKEWKLDKAKKVMENQ MFEKQAESMPETEYQYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDDKG NTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDENPLYKYEEET GNYLTYSKKDNGPVIKKIKYYGNKLNALHLDITDDYPNSRNKVVKLKSLKPYRFDVYLDNGVYKF VTVKNLVDVIKENYEVNSKCYEEAKLLKISNQAEIFASFYKNDLIKINGELYRVIGVNNRNLNK IELNMIDITYREYLENMNDKRPPHIKTIASKTQSIKKYSTDILGNLYEVKSKKHPPHIKKG	N580A
CdiCas9	Corynebacterium diphtheriae	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLVSHIHDSGLDPEIKSAVTRLASSGIARRTRRL YRRKRRRLQQLDKFIQRQGWVPIELEDYSDPLYPWKVRAELAAASYIADEKERGEKLSVALRHAR HRGWRNPYAKVSSLYLPDGPDSAFKAIREIKRASGQPVPETATVQGQMVTLCELGTLLKRGEGGV LSARLQSSDYAREIQEICRMQEIQGLYRKIIDVFAAESPKGSASSRVGKDPQPGKNRALKASD AFQRYRIALIGNLVRVDGEKRLSVEEKNLVFDHLVNLTPKKEPEWVTIAEILGIDRGQLIGTAT MTDDGERAGARPPTHDTNRSIVNSRIAPLDVWWTASALEQHAMVKALSNAEVDFFDSPEGAK VQAFFADLDDDVHAKLDSLHLPVGRAAAYSEDITLVRLLTRRMLSDGVLDLYTARLQEFGEIPEWTPPT PRIGEPVGNPAVDRVLKTVSRWLESATKTWGAPERVIIHVREGFVTEKRAREMDGDMRRRAAR NAKLFQEMQEKLNVQKPSRADLWRYQSVQRQNCQACGSPITFSNSEMDHIVPRAGQGSINT RENLAACHRCNQSKGNTPFIAWAKNTSIEGVSVEKAVERTRHWVTDTGMRSTDFKFKFTKAVVE RFQRATMDEEIDARSMESVAWMANELRSRVAQHFAHSGTTVRVYRGSLEAEARRASGISGKLKF FDGVGKSRLLDRRHADAIAVIAFTSDYVAETLAVRSNLKQSAHRQEAPQWREFTGKDAEHRAA WRVWCQKMEKLSALLTEDLRDDR VVVMNSVRLRLGNGSAHKETIGKLSKVKLSQSLVSVDIDKA SSEALWCALTREPFGDPKEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGY AELGSSFH HARVYKITSGKKPAFAMLRVYITDILLPYRNQDLFSVELKPQTMMSMRQAEKKLRDALATGNAEYL GWLVDDELVDVTSKIATDQKAVEAELGTIRRWVVDGFFSPSKLRLRPLQMSKEGIKKESAPEL SKIIDRPGWLPAVNKLFSNGVNTVVRDLSLGRVRLSTAHLPTWKVQ	H573A (Alternate)
CjeCas9	Campylobacter jejuni	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLN HLKHLIANEFKLYEDYQSFDES LAKAYKGLSIPYELRFRALNELLSKQDFARVILHIAKRRGYD DIKNSDDKEKGAILKAIKQNEEKLANYSQSVGEYLKEYYFQKFKENSKEFTNVNRNKKESYERCIQ SFLKDELKLIFFKKQREFGFSFKKFEELVSVAFYKRALKDFSHLVGNCSEFTDEKRAPKENSPLAFM FVALTRIINLLNNLKNTEGILYTKDDL NALLNEVLKNGTLTYKQTKLLGLSDDYEFKGEKTYFI EFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNIS FKALKLVTPLMLEGGKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKERYK VLNALLKKYGVHKNIELAREVGKNHSQRAKIEQENENYKAKKDAELECGLGLKINSKNILK LRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRFSDDSYMKNKVLVFTKQKQEKLNQTPFE AFGNDSAKWQKIEVLAKNLP TKKQKRILDKNYKDKEQKNFKDRNLNDTRYARLVNLYTKDYLD FLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFS AKDRNNHLLHDAVIAIYANNSI VKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIPVSKPERKKPSGALH EETFRKEEFYQSYGGKEGVLKALELGKIRKVNKGIVKNGDMFRVDIFKHKTNKFYAVPIYTMID FALKVLPNKAVARSKKGEIKDWILMDENYEFCSLYKDSLILIQTKDMQEPEFVYNAFTSSTVSIL VSKHDNKFETLSKNQKILFKNANEKEVIAKSIGQNLKVFEKYIVSALGEVTKAEFRQREDFKK	N582A
GeoCas9	Geobacillus	MRYKIGLDIGITSVGWAVMNLDPRIEDLGVRIFDRAENPQTGESLALPRRLARSARRLRRRKH LERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDRKLNDELARVLLHLAKRRGFKSNRK	N605A

	stearothermophilus	SERSNKENSTMLKHIEENRAILSSYRTV GEMIVKDPKFALHKRNKGENTNTIARDDLEREIRLIFS KQREFGNMSTEEFENEYITIWASQRPVASKDDIEKKVGFCITFEPKEKRAPKATYTFQSFIAWEHIN KLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYDRGESRKQENIRF LELDAYHQIRKAVDK VYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPLA NKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLLPN IPPIANPVVMRALTQARKVVNAIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIR QLMEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLOPIEIERLLEPGYVEVDHVIPYSRSLDSDSYTNK VVLTLTRENREKGNRIPEAYLGVGTERWQQFETVLTNKKQFSKKKRDRLRLHYDENEETEFKNRN LNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVT AHLRSRWEFNKNREESDLHHAVDIV ACTTPSDIAKVTA FYQRREQNELAKKTEPHFPQWPWFAD ELRARLSKHPKESIKALNLGNYDD QKLESLQPVFVSRMPKRSVTGAHQETLRRYVGIDERSGKIQT VVKTLSEIKLDSAGHFPMYKG ESDPRTYEAIRQRLLEHNNDPKAFQEPL YKPKKNGEPPVIRTVKIIDTKNQVIPLNDGKTVAYN SNIVRVDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDYTRFSLYPNDLRIE LPREKTVKTAAGEEINVKDVVYYKTIDSANGGELISHDHRFSLRGVGSRTLKRFEKYQVDVLG NIYKVRGEKRVGLASSAHSPGKTIPLQSTRD	
iSpyMac Cas9	Streptococcus spp.	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRIRYLQEIFANEMAKLDDSFQRLEESFLVEEDKKNERHPIFGNLAD E VAYHRN YPTIYHLRKKLADSPKADRLIY LALAHMIKFRGHFLIEGDLNPDNSDVKLFVLIQTYNQLFEE NPINASGVDAKAILSARLSKSRKLENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTIELLVKLRRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYVVGPLARGNSRFAW MTRKSEETITPWNFEFVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGD SLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE MARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVVDHIVPQSFLKDDSDNKV TRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHD AYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMIKSEQIEGKATAKYFFYSNMIMNFKEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEIQT VQNGGLFDDNPKSPL EVTPSKL VPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQ VGKNDFIKLPKYTLVDIGDGIRLWASSKEIHKGNQLVVSQKSLLYHAHLHDSLDLSDNYLQNH NQQDFVLFNEIISFSKKCKLGEKHIQKIENYSSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGV KLNQKQYKGGKDYILPCTEGTLIRQSITGLYETRVLDLSKIGEDSGSGSGSKRTADGSEFES	N863A
NmeCas9	Neisseria meningitidis	MAAFKPSNINYLGLDIGASVGWAMVEIDEEENPRLIDLGVRFERAEVPKTDGSLAMARRLAR SVRRLTRRRRAHRLRTRRLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDRKLTPLFWSAVL LHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFTPAELALNKFEKESGHIRNQR SDYSHTFSRKDLQAEILLFEKQKEFGNPHVSGGLKEGIELTLLMTQRPALSGDAVQKMLGHCTFEP AEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDPEYRKSKLTYAQARKLLGLE DTAFFKGLRYGKDNAAEASTLMEKAYHAISRALEKEGLDKKSPNLNLSPELQDEIGTAFSLFKTD EDITGRLKDRIQPEILEALLKHISFDKFVQISL KALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTE EKIYLPPIADEIRNPVVLRLALSQARKVINGVVRRYGSPARIHIETAREVGKSFDRKEIEKQREENR KDREKAAAKFREYFPNFVGEPPSKDILKLRLYEQQHGKCLYSGKEINLGRLENEKGYVEIDHALPFS RTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSFRPNSKMGQGHMETVK DEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAEND RHHALDAVVVACSTVAMQQKITRFRVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEV MIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSRPEAVHEYVTPLFVSRAPNRKMSQGQGHMETVK SAKRLDEGVSVLRVPLTQLKLDLEKVMNREREPPLYEALKARLEAHKDDPAKAFAPFYYKYDK AGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDYLYLVPIYSWQVAKGIL PDRAVVQKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFYFASCHRGTGNNIRIHDLDH KIGKNGILEGIGVKTALSFOKYQIDELGKEIRPCRLKKRPPVR	N611A
ScaCas9	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSSKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRRYTRRKNRIRYLQEIFANEMAKLDDSFQRLEESFLVEEDKKNERHPIFGNLAD E VAYHRN YPTIYHLRKKLADSPKADRLIY LALAHMIKFRGHFLIEGDLNPDNSDVKLFVLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSRKLEKLI A VFPNEKKNGLFGNLIALLSLGLTPNFKSNFDLAEDAKLQ SKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILSDILRSNSEVT KAPLSASMYKRYDEHHQD LALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVIGIKHRKRTTKLATQEEFYKFIKPILEKMDGAE ELLAKLNRDILLRKQRTFDNGSIPHQIHLGELHAILRRQEEFYPLKENREKIEKILTFRIPYVVGPL ARGNSRFAWLTRKSEEAITPWNFEFVVDKGASQSFIERMTNFDQELPNKKVLPHKSLLYEYFTV YNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDR FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLK RRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDFSNRNFQMLIHDDSLTFKEIEKAQVSGQGD SLHEQIADLAGSPAIKKGILQTVKVVDELVKVMGHKPENIVIE MARENQTTTKGLQSSRERKKRIEE GIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVVDHIVPQSFIKDDSI DNKVLTRSVENRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAG FIKRLVETROITKHVARILDSRMNTKRDKNDKPIREVKVITLKSCLVSDFRKDFQLYKVRDINNY HHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMIKSEQIEGKATAKRFFYSNMIMNF FKTEVKLANGAIRKRPLIETNGETGEVWVNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESI	N872A

		LSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKEVGKAKKLKSVKVLVGITIMEKGSYE KDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQHLVRLLYYTQN ISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLK YTSFGASGGFTFLDLDVKQGRRLRYQTVTEVLDTLIYQSITGLYETRTDLSQLGGD	
ScaCas9- HiFi- Sc++	Streptoco ccus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSSKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLK RTARRRYTRRKNRIRYLQEIFANEMAKLDDSFQRLEESFLVEEDKKNERHPFIGNLADEVAYHRN YPTIYHLRKKLADSPEKADRLRIYLALAHMIKFRGHFLIEGKLNENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDKILSARLSKSKRLEKLIIVFPNEKKNGLFGNIALALGLTPNFKSNFIDLTEADAKLQ SKDITYDDDLDELLGQIGDQYADLFSAAKNLSAAILSDILRSNSEVTAKPLSASMVKRYDEHHQD LALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRSGLATEEEFYKFIKPILEKMDGA EELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYFPLKENREKIEKILTFRIPYYVGP LARGNSRFAWLTRKSEEAITPWNFEFVVDKGASQSFIERMTNFEQLPNKKVLPKHSLLYEYFT VYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQK KRHHYTGWGRLSRKMINGRDKQSGKTILDFLSDGFSNANFMQLIHDDSLTFKEEIEKAQVSGQ GDSLHEQIADLAGSPAIIKGGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQSSRERKKR IEEGIKELESQILKENPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVIDHIVPOSFLK DSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDAKAGFIK AGFIKRLVETROITKHVARILDSRMNTKRDKNKPIREVKVITLKSCLVSDFRKDFQYKVRDIN NYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKRFYFSNIM NFFKTEVKLANGEIRKRLPIETNGETGEIVWDKGRDFATVRKVLSPMPQVNIVKKTEVQTGGFSKES SILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKEVGKAKKLKSVKVLVGITIMEKGSY EKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQHLVRLLYYTQN NISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLK YTSFGASGGFTFLDLDVKQGRRLRYQTVTEVLDTLIYQSITGLYETRTDLSQLGGD	N872A
SpyCas9- 3var- NRRH	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRIRYLQEIFANEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQILVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFIDLAEADAKLQ LSKDTYDDDLNLLAQIGDQYADLFSAAKNLSAAILSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEFVVDKGASQSFIERMTNFDKNLPNEKVLPHKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVDFRNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQKRLRYTGW RLSRKLINGIRDKQSGKTILDFLSDGFSNANFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIIKGGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEIGIKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVIDHIVPOSFLKDDSDNKVLR TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDAKAGFIKRL VETROITKHVARILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQYKVRDINNYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYFSNIMNFFKTEITL ANGEIRKRLPIETNGETGEIVWDKGRDFATVRKVLSPMPQVNIVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTAAYSILVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIGFLEA KGYKEVKKDLIIKLPKYSLFELENGRRRMLASAGVLHKGNELALPSKYVNFLYLASHYEKLKGP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLKDVL SAYNKHHRDKPIREQAENIHLFTLT NLGVAFAFKYFDTTIDKKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGD	N863A
SpyCas9- 3var- NRTH	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRIRYLQEIFANEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQILVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFIDLAEADAKLQ LSKDTYDDDLNLLAQIGDQYADLFSAAKNLSAAILSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGIIPHQIHLGELHAILRRQGDYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEFVVDKGASQSFIERMTNFDKNLPNEKVLPHKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVDFRNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQKRLRYTGW RLSRKLINGIRDKQSGKTILDFLSDGFSNANFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIIKGGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEIGIKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVIDHIVPOSFLKDDSDNKVLR TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDAKAGFIKRL VETROITKHVARILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQYKVRDINNYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYFSNIMNFFKTEITL ANGEIRKRLPIETNGETGEIVWDKGRDFATVRKVLSPMPQVNIVKKTEVQTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTVAYSILVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIGFLEA KGYKEVKKDLIIKLPKYSLFELENGRRRMLASASVLHKGNELALPSKYVNFLYLASHYEKLKGS EDNKQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLKDVL SAYNKHHRDKPIREQAENIHLFTLT NLGASAAFKYFDTTIGRKLTYSTKEVLDTLIHQSIITGLYETRIDLSQLGGD	N863A
SpyCas9- 3var- NRCH	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRIRYLQEIFANEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQILVQTYNQLFEE	N863A

		NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLLAAKNLSDAILSDILRVNTEITKAPLSASMVKRYDEHHQ DLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLRKQRTFDNGIIPHQIHLGELHAILRRQGDYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGHHKPNIVIEAMARENQTTQKGQKNSRERMKRIEKGKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVIDHVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKEVQTTGGFSKESILPKGNSDK LIARKKDWDPKKYGGFNSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGVLQKGNELALPSKYVNFYLYASHYEKLGKSP EDNEQKQLFVEQHKHYLDEIIQISEFSKRVLADANLKDVLASAYNKHDKPIREQAENIHLFTLT NLGAPAAFKYFDTTNRKQYNTTKEVLDTLIRQSTGLYETRIDLSQLGGD	
SpyCas9- HF1	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTTYHLRKKLVDSITDKADLRLLIYLAALAHMIKFRGHFLIEGDLNPDNSDVKLFQVLQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLLAAKNLSDAILSDILRVNTEITKAPLSASMVKRYDEHHQD DLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKNIVIEAMARENQTTQKGQKNSRERMKRIEKGKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVIDHVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKEVQTTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFYLYASHYEKLGKSP EDNEQKQLFVEQHKHYLDEIIQISEFSKRVLADANLKDVLASAYNKHDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDTLHQSTGLYETRIDLSQLGGD	N863A
SpyCas9- QQR1	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTTYHLRKKLVDSITDKADLRLLIYLAALAHMIKFRGHFLIEGDLNPDNSDVKLFQVLQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLLAAKNLSDAILSDILRVNTEITKAPLSASMVKRYDEHHQD DLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKNIVIEAMARENQTTQKGQKNSRERMKRIEKGKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVIDHVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHD AYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKEVQTTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIDFLEA KGYKEVKKDLIIKLPKYSLEFENGRKRMLASARELQKGNELALPSKYVNFYLYASHYEKLGKSP EDNEQKQLFVEQHKHYLDEIIQISEFSKRVLADAQLDKVLASAYNKHDKPIREQAENIHLFTLT NLGAPAAFKYFDTTFKQKQYRSTKEVLDTLHQSTGLYETRIDLSQLGGD	N863A
SpyCas9- SpG	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKY PTTYHLRKKLVDSITDKADLRLLIYLAALAHMIKFRGHFLIEGDLNPDNSDVKLFQVLQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLLAAKNLSDAILSDILRVNTEITKAPLSASMVKRYDEHHQD DLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRYTGWG	N863A

		RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNIKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFWPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLE AKGYKEVKKDLIHLKPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLKDVL SAYNKHDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSIITGLYETRIDLSQLGGD	
SpyCas9- VQR	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLV DSTDKADRLRIYLAALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFDLAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEFVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQKLRRTYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNIKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEA KGYKEVKKDLIHLKPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS EDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLKDVL SAYNKHDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSIITGLYETRIDLSQLGGD	N863A
SpyCas9- VRER	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLV DSTDKADRLRIYLAALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFDLAEDAKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEFVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQKLRRTYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNIKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEA KGYKEVKKDLIHLKPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGS EDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLKDVL SAYNKHDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKEYRSTKEVL DATLIHQSIITGLYETRIDLSQLGGD	N863A
SpyCas9- xCas	Streptoco ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLV DSTDKADRLRIYLAALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFDLAEDTKLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTTEELLVKLNRED LLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETTPWNFEKVVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMQKLRRTYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHD AYLNAVVGTAALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNIKKTEVQTGGFSKESILPKRNSDK	N863A

		LIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLSVKELLGITIMERSSSFENPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASAGVLQKGNELALPSKYVNFYLAASHYEKLKGSF EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLKDVLASYNKHRDKPIREQAENIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDTLHQISITGLYETRIDLSQLGGD	
SpyCas9- xCas-NG	Streptoco- ccus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFGNIVDEVAYHEKY PTIYHLRKKLVSDTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIALQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQ LSKDYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKLYADEHHQD LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNEKVVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHANL AGSPAIIKGIQTQVKKVDELKVMGRHKPENIVIAMARENQTTQKQKNSRERMKRIIEGKELG SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSDNFKVL TSRDKNRGKSDNPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGSELDKAGIFKRL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTAIIKYPKLESEFVYGDYKVDVRKMIKSEQEGKATAKYFFYSNIMNFFKTEITL ANGEIKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNVKKTETVQTGGFSKESRPNRNSDK LIARKKDWDPKKYGGFVSPTVAYSVLVVAKEVGKSKKLSVKELLGITIMERSSSFENPIDFLEA KGYKEVKKDLIIKLPKYSLEFELNGRKRMLASARFLQKGNELALPSKYVNFYLAASHYEKLKGSF EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLKDVLASYNKHRDKPIREQAENIHLFTLT NLGAPRAFKYFDTTIDRKVYRSTKEVLDTLHQISITGLYETRIDLSQLGGD	N863A
St1Cas9- CNRZ106 6	Streptoco- ccus thermophi- lus	MSDLVLGLDIGISVGVGILNKVTGEIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLLDASDDGNSSVG DYAQIVKENSQLETKTPGQIQLERYQTYGQLRGDFTVEKDGGKHHRLINVPFPTSAIRSEALRILQT QQEFNPQITDEFINRYLEILTGRKRYHGGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDFE AAKASYTAQEFNLLNDLNNLTVPETETKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA YKGYRIDKSGKAEIHTFEAYRKMKLTLETLDIEQMDRETLDKLAYVLTNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTSSSNK TKYIDEKLLTEIYNPVPVAKSVRQAIKIVNAAIKEYGDFDNIVIAMARETNEDDEKKAQIKQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKQORTPYQALDSMDDAWSFRELKAFVRESKTLSSNK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAP YQHFDVTLKSKEFEDSILFSYQVDSKFNKISDATIYATRQAKVGKDKKDETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHPDPTFEKVIPIENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRK YSKKGNGPEIKSLKYDYSKLLGNPIDITPENSNNKVVLSLKPWRTDYFNKNTGKYELLGLKYA DLQFEKGTGTYSKQEKYNDIKKKEGVDSSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRLPKQ KHYYVELKPYDKQKFEKGGEALIKVLGNVANGGQCIKGLAKSNISYKVRTDVLGNQHIIKNEGDKP KLDF	N622A
St1Cas9- LMG1831	Streptoco- ccus thermophi- lus	MSDLVLGLDIGISVGVGILNKVTGEIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLLDASDDGNSSVG DYAQIVKENSQLETKTPGQIQLERYQTYGQLRGDFTVEKDGGKHHRLINVPFPTSAIRSEALRILQT QQEFNPQITDEFINRYLEILTGRKRYHGGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDFE AAKASYTAQEFNLLNDLNNLTVPETETKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA YKGYRIDKSGKAEIHTFEAYRKMKLTLETLDIEQMDRETLDKLAYVLTNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTSSSNK TKYIDEKLLTEIYNPVPVAKSVRQAIKIVNAAIKEYGDFDNIVIAMARETNEDDEKKAQIKQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKQORTPYQALDSMDDAWSFRELKAFVRESKTLSSNK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAP YQHFDVTLKSKEFEDSILFSYQVDSKFNKISDATIYATRQAKVGKDKKDETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHPDPTFEKVIPIENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRK YSKKGNGPEIKSLKYDYSKLLGNPIDITPENSNNKVVLSLKPWRTDYFNKNTGKYELLGLKYA DLQFEKGTGTYSKQEKYNGIMKEGVDSSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPN VKYYVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLGKSNISYKVRTDVLGNQHIIKNEGDKPK LDF	N622A
St1Cas9- MTH17C L396	Streptoco- ccus thermophi- lus	MSDLVLGLDIGISVGVGILNKVTGEIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLLDASDDGNSSVG DYAQIVKENSQLETKTPGQIQLERYQTYGQLRGDFTVEKDGGKHHRLINVPFPTSAIRSEALRILQT QQEFNPQITDEFINRYLEILTGRKRYHGGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDFE AAKASYTAQEFNLLNDLNNLTVPETETKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVA YKGYRIDKSGKAEIHTFEAYRKMKLTLETLDIEQMDRETLDKLAYVLTNTEREGIQEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTSSSNK TKYIDEKLLTEIYNPVPVAKSVRQAIKIVNAAIKEYGDFDNIVIAMARETNEDDEKKAQIKQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTIHDLINNSNQF	N622A

		EVDHILPLSITFDDSLANKVLVYATANQEKQORTPYQALDSMDDAWSFRELKAFVRESKTLNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFDVTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHPDPTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY SKKGNNGPEIKSLKYDYSKLGHNHIDITPKDSNNKVVLQSLKPWRDVTYFNKNTGKYEILGLKYSDM QFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSengeQILLRFTSRNDTSKHY VELKPYNRQKFEGSEYLIKSLGTVAKGQCKIKGLGKSNISYKVRDVLGNQHIIKNEGDKPKLDF	
StlCas9- TH1477	Streptococcus thermophilus	MSDLVLGLDIGSVGVGILNKVTGEIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRL NRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDASDDGNSSVG DYAQIVKENSQKLETKTPGQIQLERYQTYGQLRGDFIVEKDGKKHRLINVFPTSAIRSEALRILQT QQEFNPQITDEFINRYLEILTGKRKYHGPNGEKSRTDYGRYRTSGE.TLDNIFGILIGKCTFYDDEF AAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVNEKAMGPAKLFKYIAKLLSCDVA DIKGYRIDKSGKAEIHTEAYRKMKLTLETLDIEQMDRETLDKLAYVLTLNTEREGIEALEHEFAD GSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILRLGKQKTTSSNK TKYIDEKLLTEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVEMARETNEDDEKKAQIKQKANK DEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTIHDLINNSNQF EVDHILPLSITFDDSLANKVLVYATANQEKQORTPYQALDSMDDAWSFRELKAFVRESKTLNKK KEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRH WGIEKTRDTYHHHAVDALIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAP YQHFDVTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDG YDAFMKIYKKDKSKFLMYRHPDPTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKY SKKGNNGPEIKSLKYDYSKLGHNHIDITPKDSNNKVVLQSLKPWRDVTYFNKNTGKYEILGLKYSDM QFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSengeQILLRFTSRNDTSKHY VELKPYNRQKFEGSEYLIKSLGTVVGGRCKIKGLGKSNISYKVRDVLGNQHIIKNEGDKPKLDF	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	GTTGTAGCTC CCTTTCTCAT TTCG	GAAA	CGAAATGAGA ACCGTTGCTA CAATAAGGCC GTCTGAAAAG ATGTGCCGCA ACGCTCTGCC

								CCTTAAAGCT TCTGCTTTAA GGGGCATCGT TTA
PpnCas9	NNNNR TT		1	21	24	GTTGTAGCTC CCTTTTTCAT TTCGC	GAAA	GCGAAATGAA AAACGTTGTT ACAATAAGAG ATGAATTCT CGCAAAGCTC TGCCTCTTGA AATTTTCGTT TCAAGAGGCA TCTTTTT
SauCas9	NNGRR; NNGRRT	-3	1	21	23	GTTTTAGTAC TCTG	GAAA	CAGAATCTAC TAAAACAAGG CAAAATGCCG TGTTTATCTC GTCAACTTGT TGGCGAGA
SauCas9 -KKH	NNNRR; NNNRRT	-3	1	21	21	GTTTTAGTAC TCTGTAAT	GAAA	ATTACAGAAT CTACTAAAAC AAGGCAAAAT GCCGTGTTTA TCTCGTCAAC TTGTTGGCGA GA
SauriCas 9	NNGG	-3	1	21	21	GTTTTAGTAC TCTG	GAAA	CAGAATCTAC TAAAACAAGG CAAAATGCCG TGTTTATCTC GTCAACTTGT TGGCGAGATT TTT
SauriCas 9-KKH	NNRG	-3	1	21	21	GTTTTAGTAC TCTG	GAAA	CAGAATCTAC TAAAACAAGG CAAAATGCCG TGTTTATCTC GTCAACTTGT TGGCGAGATT TTT
ScaCas9 -Sc++	NNG	-3	1	20	20	GTTTTAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC
SpyCas9	NGG	-3	1	20	20	GTTTTAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC

SpyCas9 -NG	NG (NGG=N GA=NGT >NGC)	-3	1	20	20	GTTTAAGAG CTATGCTG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACT TGAAAAAGTG GCACCGAGTC GGTGC
SpyCas9 -SpRY	NRN>N YN	-3	1	20	20	GTTT TAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC
StlCas9	NNAGA AW>NN AGGAW =NNGG AAW	-3	1	20	20	GTCTTTGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT
BlatCas 9	NNNC NAA>N NNNCN DD>NN NNC	-3	1	19	23	GCTATAGTTC CTTACT	GAAA	GGTAAGTTGC TATAGTAAGG GCAACAGACC CGAGGCGTTG GGGATCGCCT AGCCCGTGTT TACGGGCTCT CCCCATATTC AAAATAATGA CAGACGAGCA CCTTGGAGCA TTTATCTCCG AGGTGCT
cCas9- v16	NNVAC T;NNVA TGM;NN VATT;N NVGCT; NNVGT G;NNVG TT	-3	2	21	21	GUCUUAGUA CUCUG	GAAA	CAGAAUCUAC UAAGACAAG GCAAAAUGCC GUGUUUAUC UCGUCAACUU GUUGGCGAG AUUUUUUU
cCas9- v17	NNVRR N	-3	2	21	21	GUCUUAGUA CUCUG	GAAA	CAGAAUCUAC UAAGACAAG GCAAAAUGCC GUGUUUAUC UCGUCAACUU GUUGGCGAG AUUUUUUU
cCas9- v21	NNVAC T;NNVA TGM;NN VATT;N NVGCT;	-3	2	21	21	GUCUUAGUA CUCUG	GAAA	CAGAAUCUAC UAAGACAAG GCAAAAUGCC GUGUUUAUC UCGUCAACUU

	NNVGT G;NNVG TT							GUUGGCGAG AUUUUUUU
cCas9- v42	NNVRR N	-3	2	21	21	GUCUUAGUA CUCUG	GAAA	CAGAAUCUAC UAAGACAAG GCAAAAUGCC GUGUUUAUC UCGUCAACUU GUUGGCGAG AUUUUUUU
CdiCas9	NNRHH HY;NNR AAAY		2	22	22	ACUGGGGUU CAG	GAAA	CUGAACCUC GUAAGCAUU GGCUCGUUUC CAAUGUUGA UUGCUCGCGC GGUGCUCCUU AUUUUUAAG GGCGCCGGC
CjeCas9	NNNNR YAC	-3	2	21	23	GTTTTAGTCC CT	GAAA	AGGGACTAAA ATAAAGAGTT TGCGGGACTC TGCGGGGTTA CAATCCCCTA AAACCGCTTT TTT
GeoCas 9	NNNNC RAA		2	21	23	GUCAUAGUU CCCCUGA	GAAA	UCAGGGUUA UAUGAUAA GGCUUUCUG CUAAGGCAGA CUGACCCGCG GCGUUGGGG AUCGCCUGUC GCCCCUUUU GGCGGGCAUU CCCCAUCCUU
iSpyMac Cas9	NAAN	-3	2	19	21	GTTTTAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC
NmeCas 9	NNNNG AYT;NN NNGYTT ;NNNNG AYA;NN NNGTCT	-3	2	20	24	GTTGTAGCTC CCTTCTCAT TTCG	GAAA	CGAAATGAGA ACCGTTGCTA CAATAAGGCC GTCTGAAAAG ATGTGCCGCA ACGCTCTGCC CCTTAAAGCT TCTGCTTTAA GGGGCATCGT TTA
ScaCas9	NNG	-3	2	20	20	GTTTTAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT

								AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC
ScaCas9 -HiFi- Sc++	NNG	-3	2	20	20	GTTTTAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC
SpyCas9 -3var- NRRH	NRRH	-3	2	20	20	GTTTAAGAG CTATGCTG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACT TGAAAAAGTG GCACCGAGTC GGTGC
SpyCas9 -3var- NRTH	NRTH	-3	2	20	20	GTTTAAGAG CTATGCTG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACT TGAAAAAGTG GCACCGAGTC GGTGC
SpyCas9 -3var- NRCH	NRCH	-3	2	20	20	GTTTAAGAG CTATGCTG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACT TGAAAAAGTG GCACCGAGTC GGTGC
SpyCas9 -HF1	NGG	-3	2	20	20	GTTTTAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC
SpyCas9 -QQR1	NAAG	-3	2	20	20	GTTTTAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC
SpyCas9 -SpG	NGN	-3	2	20	20	GTTTTAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC
SpyCas9 -VQR	NGAN	-3	2	20	20	GTTTTAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA

								AAGTGGCACC GAGTCGGTGC
SpyCas9 -VRER	NGCG	-3	2	20	20	GTTT TAGAG CTA	GAAA	TAGCAAGTTA AAATAAGGCT AGTCCGTTAT CAACTTGAAA AAGTGGCACC GAGTCGGTGC
SpyCas9 -xCas	NG;GAA ;GAT	-3	2	20	20	GTTTAAGAG CTATGCTG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACT TGAAAAAGTG GCACCGAGTC GGTGC
SpyCas9 -xCas- NG	NG	-3	2	20	20	GTTTAAGAG CTATGCTG	GAAA	CAGCATAGCA AGTTTAAATA AGGCTAGTCC GTTATCAACT TGAAAAAGTG GCACCGAGTC GGTGC
St1Cas9 - CNRZ1 066	NNACA A	-3	2	20	20	GTCTTTGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT
St1Cas9 - LMG18 31	NNGCA A	-3	2	20	20	GTCTTTGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT
St1Cas9 - MTH17 CL396	NNAAA A	-3	2	20	20	GTCTTTGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT
St1Cas9 - TH1477	NNGAA A	-3	2	20	20	GTCTTTGTAC TCTG	GTAC	CAGAAGCTAC AAAGATAAGG CTTCATGCCG AAATCAACAC CCTGTCATTTT ATGGCAGGGT GTTTT

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
5 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
10 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
15 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
20 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
25 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
30 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' (SEQ ID NO: 1591)), 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).

Methods and Compositions for Modified RNA (e.g., gRNA or template RNA)

5 In some embodiments, an RNA component of the system (e.g., a template RNA or a gRNA) 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
10 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'-O-(2-methoxyethyl) (2'-O-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
15 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
20 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
25 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 RNAs (e.g., at the portion thereof that binds a target site) or guide RNAs described herein comprises any of the sequences shown in Table 4 of WO2018107028A1, incorporated herein by reference in its entirety. In some embodiments,
30 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

5 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

10 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%,

15 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

20 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%,

25 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%,

30 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 RNAs (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
5 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 described 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*
10 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
15 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
20 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
25 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
30 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.

5 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
10 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
15 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,
20 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,
25 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
30 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
5 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
10 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,
15 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
20 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
25 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
30 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.,

5 GGGAAAUAGAGAGAGAAAAGAAGAGUAAGAAGAAAUUAAGAGCCACC (SEQ ID NO: 1603) and

UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCC
AGCCCCUCCUCCCCUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGA

10 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
15 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
20 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
25 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
30 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 be initiated by the addition of required Cu(I) ions. Other exemplary mechanisms by which RNA segments may be connected include, without limitation, 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 illustration, 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
5 US20160102322A1, which is incorporated herein by reference in its entirety.

Template nucleic acid composition

In some embodiments, the template nucleic acid is a template RNA. In some
embodiments, the template RNA comprises one or more modified nucleotides. For example, in
10 some embodiments, the template RNA comprises one or more deoxyribonucleotides. In some
embodiments, regions of the template RNA are replaced by DNA nucleotides, e.g., to enhance
stability of the molecule. For example, the 3' end of the template may comprise DNA
nucleotides, while the rest of the template comprises RNA nucleotides that can be reverse
transcribed. For instance, in some embodiments, the heterologous object sequence is primarily
15 or wholly made up of RNA nucleotides (e.g., at least 90%, 95%, 98%, or 99% RNA
nucleotides). In some embodiments, one or both of the 3' UTR and the 3' target homology
domain are primarily or wholly made up of DNA nucleotides (e.g., at least 90%, 95%, 98%, or
99% DNA nucleotides). In other embodiments, the template region for writing into the genome
may comprise DNA nucleotides. In some embodiments, the DNA nucleotides in the template are
20 copied into the genome by a domain capable of DNA-dependent DNA polymerase activity. In
some embodiments, the DNA-dependent DNA polymerase activity is provided by a DNA
polymerase domain in the polypeptide. In some embodiments, the DNA-dependent DNA
polymerase activity is provided by a reverse transcriptase domain that is also capable of DNA-
dependent DNA polymerization, e.g., second strand synthesis. In some embodiments, the DNA-
25 dependent DNA polymerase activity is provided by a DNA polymerase. In some embodiments,
the DNA-dependent DNA polymerase activity provided by a DNA polymerase domain in the
polypeptide is not capable of reverse transcription activity. In some embodiments, the template
molecule is composed of only DNA nucleotides. In some embodiments, the DNA template is
polymerized into the genome by a DNA polymerase. In some embodiments, the template
30 composed of DNA nucleotides comprises modified DNA nucleotides. In some embodiments, the
template composed of DNA nucleotides comprises a modified backbone.

The nucleotides comprising the template of the Gene Writer™ system can be natural or modified bases, or a combination thereof. For example, the template may contain pseudouridine, dihydrouridine, inosine, 7-methylguanosine, or other modified bases. In some embodiments, the template may contain locked nucleic acid nucleotides. In some embodiments, the modified bases
5 used in the template do not inhibit the reverse transcription of the template. In some embodiments, the modified bases used in the template may improve reverse transcription, e.g., specificity or fidelity.

Additional Functional Characteristics for Gene Writers™

10 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
15 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.

20

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
25 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.,
30 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.

5 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
10 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*.

RNA Binding Domain

15 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
20 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
25 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
30 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

5 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
10 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
15 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
20 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
25 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
30 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 et 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 $1 \times 10^{-3} - 1 \times 10^{-5} \text{ min}^{-1}$ as measured by such methods.

In some embodiments, the endonuclease domain has a catalytic efficiency (k_{cat}/K_m) greater than about $1 \times 10^8 \text{ s}^{-1} \text{ M}^{-1}$ in vitro. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1×10^5 , 1×10^6 , 1×10^7 , or $1 \times 10^8 \text{ s}^{-1} \text{ M}^{-1}$ 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 $1 \times 10^8 \text{ s}^{-1} \text{ M}^{-1}$ in cells. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1×10^5 , 1×10^6 , 1×10^7 , or $1 \times 10^8 \text{ s}^{-1} \text{ M}^{-1}$ 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 $5 \times 10^{-3}/\text{nt}$, $5 \times 10^{-4}/\text{nt}$, or $5 \times 10^{-6}/\text{nt}$, 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

5 (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
10 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
15 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 1×10^{-3} – 1×10^{-4} or 1×10^{-4} – 1×10^{-5} substitutions/nt, e.g., as described in Yasukawa et al. (2017) *Biochem Biophys Res Commun*
20 492(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 1×10^{-3} – 1×10^{-4} or 1×10^{-4} – 1×10^{-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).

25 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,
30 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
 5 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
 10 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).

In some embodiments, a reverse transcriptase domain may comprise a mutation, e.g., as listed in Table 45. In embodiments, the mutation modifies, e.g., increases the stability and
 15 functionality of the RT domain. In some embodiments, the mutation modifies, e.g., increases processivity and template affinity of the RT domain. In some embodiments, the mutated RT domain may show at least 5 fold, at least 10 fold, at least 15 fold, at least 20 fold, at least 25 fold, at least 30 fold, at least 40 fold, at least 45 fold, at least 50 fold, at least 55 fold, at least 60 fold, at least 65 fold, at least 70 fold, at least 80 fold, at least 100 fold increase to processivity
 20 compared to an unmutated RT domain. In embodiments, a mutated RT domain may show at least at least 5 fold, at least 10 fold, at least 15 fold, at least 20 fold, at least 25 fold, at least 30 fold, at least 40 fold, at least 45 fold, at least 50 fold, at least 55 fold, at least 60 fold, at least 65 fold, at least 70 fold, at least 80 fold, at least 100 fold increase in template affinity compared to an unmutated RT domain. In some embodiments, a mutant RT domain may comprise one or more
 25 mutations selected from D200N/T330P/L603W, T306K, W313F, L139P, E607K.

Table 45 discloses mutations improve the properties of various reverse transcriptases. Core mutations expected to be the most impactful were applied across groups of retroviruses. Conservation of sequence across a group of viruses at one of these core mutations led to the
 30 installation of the mutation across that group (see Example 33, Figures 36 A and B). Sequence

positions refer to the positions in MMLV RT. In some embodiments, a RT domain described herein comprises a mutation as described in Table 45.

Table 45. List of exemplary RT domain mutations

Group	L139	D200	T306	W313	T330	L603	E607
Gamma		D200N	T306K	W313F	T330P	L603W	
Epsilon		D200N	T306K	W313F	T330P	L603W	
Delta	L139P	D200N	X	X	T330P	L603W*	X
Beta	L139P	X	X	X	T330P	X	X
Spuma		D200N	T306K	X	T330P	L603W	

5

Cas-RT fusions

In some embodiments, a GeneWriter polypeptide comprises a RT domain fused to a Cas molecule. In some embodiments, the Cas molecule is the DBD and/or the endonuclease domain of the GeneWriter polypeptide. In some embodiments, the an RT domain comprises Cas9. In some embodiments, the Cas9 may comprise a mutation, e.g., a disclosed in Table 40A. Table 46 discloses a list of exemplary Cas-RT fusion proteins.

In some embodiments, a Cas molecule in a GeneWriter polypeptide has a similar activity to an otherwise similar Cas molecule that is not fused to a RT domain. In some embodiments, the activity is at least 40%, 50%, 60%, 70%, 80%, or 90% of that of the otherwise similar Cas molecule. In some embodiments, the Cas molecule in the GeneWriter polypeptide may have an indel formation activity at least 40%, 50%, 60%, 70%, 80%, or 90% of that of an otherwise similar Cas molecule that is not fused to a RT domain, e.g., in an assay accrodign to Example 32.

In some embodiments, a GeneWriter polypeptide comprises an amino acid sequence according to Table 46 below, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a nucleic acid encoding a GeneWriter polypeptide comprises a nucleic acid sequence according to Table 47, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

Table 46. List of exemplary Gene Writer polypeptides comprising Cas-RT fusions

Name	Cas domain	Linker	RT retroviral source	RT source polypeptide	Gene Writer polypeptide sequence
Cas-RT(MMLV)	Cas9 (N863A)	SGSS GGSSG SETPG TSEA TPSS GGSSG GSS	Moloney murine leukemia virus (MMLV) or MLVMS)	P03355	<p> MKRTADGSEFESPKKRVKDKKYSIGLDIGTNSVGWAVITDEYKVPKSKFKVLG NTDHRSIKKNLJGALLFDSGETAEATRLKRTARRRYTRRNKNRCYLQEIFSNEMAK VDDSFHRLLESFLVEEDKKHERHPFGNIVDEVAYHEKYPTTYHLRKKLVDSTDK ADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVVDKLFQLVQTYNQLFEENPINAS GVDAILSAKLSKRRLENLAQLPGEKKNGLFNLIALSLGLTPNFKSNFLAE DAKLQSKDTYDDDLNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE EFYKFIKPILEKMDGTTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE DFYPLKDNREKIEKILTFRIPIYVYVGLARGNSRFVWTRKSEETITPWNFEVVD KGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNKVTVKQLKEDYFKIECFDSVEISGVEDRFNASLGT YHDLKIIKDKDFLDNEENEDILEDVLTLLFEDREMIERLKTYYAHLFDDDKVMK QLKRRRYTGWRGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNMQLIHDDSLTF KEDIQKQVSGQGSLSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGKNSRERMRKIEEGIKELGSQLKEHPVENTQLQNEKLYLY YLQGRDMYVDQELDNRLSDYDVHIVPQSFLKDDSDIDNKVLTRSDKARGKSD NVPSEEVKKMKNYWRQLNNAKLITQRKFDNLTKAERGGLSLDDKAGFIKRQLV ETRQITKHVAQILD SRMNTKYDENDKLJREVKVITLKSCLVSDFRKDFQYKVRER NNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKYVDVRKMKIAKSEQIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS VLVAKVKEGKSKLKSVKELGITIMERSSEKKNPIDFLEAKGYKVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLAHYEKLKGSPEDN EQQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVLASYNKHRDKPIREQAEN IHLFTLNLGAPAAFKYFDTTIDRRRYTSTKEVDATLIHQSIITGLYETRIDLSQLG GDSGSGSGSGSETPGTSESA TPSSGSGSGSGSTLNIEDEYRLHETSKPEPDVSLG STWLSDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIKQYPMSEQEARLGIKPHI QRLLDQGLVPCQSPWNTPLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPY NLLSGLPPSHQWYTVLCLKDAFFCLRLHPTSQLPFQAFWRDPEMGISGQLTWTRL PQGKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLAATSELDCCQCGTRALL QTLGNLGYRASAKKAQICQKQVKYGLYLLKEGQWLTEARKETVMGQPTPKTP RQLREFLGKAGFCRLFIPIGFAEMAAPLYPLTKPGITLFWGPDQQAQYQEIQAALL </p>

Cas-RT(PERV)	Cas9 (N863A)	SGSS GGSSG SETPG TSEA TPSS GGSSG GSS	Porcine endogenous retrovirus (PERV)	Q4VFZ2	<p>TAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGWRRPVAYLSKKLDPVAAG WPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMT HYQALLDTRVQFGPVVALNPATLLPLPEGLQHNCLDILAEAHGTRPDLTDQP LPDADHTWYTDGSSLLQEQGRKAGAAVTTEVIAWAKALPAGTSAQRAELIALT QALMAEGKKLVYTDTRYAFATAHHGEIYRRGWLTSEGKEIKNKDEILALL KALFLPKRLSIHCPGHQGHSAEARGNRMAQARKAAITETPDTSTLLIENSSP SGSKRTADSGSEFEPKKKKV</p> <p>MAPKKRKYGHGVPAAADKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNT DRHSIKNLGALLFSDSETAEATRLKRTARRRYTRRNKRICYLQEFSNEMAKVD DSFFHRLSEESFLVEEDKKHERHPFIGNIVDEVAYHEKYPTIYHLRKKLVDSITDKAD LRLIYLALAHMIKFRGHFIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGV DAKAILSARLSKRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA KLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLSDAILSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKNREDLRLKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYVVGPLARGNSRFAWMTRKSEETITPWNTEEVVDKGA SAQSFIERTMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVYVTEGMRKKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIHKDKDFLDNEENEDILEIVLTLTFEDREMIEERLKYAHLFDDDKVMKQLK RRRYTGWGRLSRKLINGIRDQSGKTILDFLKSDFANRNFQMLHDDSLTFKEDI QKAQVSGQGDLSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVEM ARENQTTQKGQKNSRERMKRIEIGIKELQSILKEHPVENTQLQNEKLYLYLQN GRDMYVDQELDINRLSDYDVVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPS EVVKMKMKNYWRQLNNAKLJTQRKFDNLTKAERGGLSELDKAGFIKRLVETROI TKHVAQILD SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQYKVRINNYYH HAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVVRKMKIAKSEQIEGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWWDKGRDFATVRKVLSMPQ VNIVKKTVEQIGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGSKKLLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFYLASHYKELKGSPEDEQKQ LFVEQHKHYLDEHIEQISEFSKRIVLADANLDDKLSAYNKHDKPIREQAENIHLF TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSTIGLYETRIDLSQLGGDS GGSSGGSSGSETPGTSEATPSSGGSSGSSLDDEYRLYSPLVKPDQNIQFWLEQ FPQAWAETAGMGLAKVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLLQ QGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNRKRVQDIHPTVNPYNLLCA LPPQRSWYTVLCLKDAFFCLRLHPTSQLPFAFEWRDPGTGTGRTGQLTWTRLPQGFK NSPTIFNEALHRDLANFRIQHPQVTLQYVDDLLAGATKQDCLEGTAKLLELS DLGYRASAKKAQICRREVITYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVRE FLGKAGFCRLFPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALA LPDVTKPFITLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPV ASGWVPVCLK</p>
--------------	--------------	--	--------------------------------------	--------	---

Cas- RT(MLVB M)	Cas9 (N863A)	SGSS GGSSG SETPG TSEA TPSS GGSSG GSS	Murine leukemia virus (MLVBM)	Q7SVK7	<p>AIAAVAILVKDADKLTGQNTVIAPHALENIVRQPPDRWMTNARMTHYQSLLT ERVTFAPPAALNPATLLPEETDEPVTDHCHQLLIEETGVRKDLTDIPLTGEVLTWF TDGSSVVEGKRMAGAAVVDGTRTIWASSLPEGTSQAQKAELMALTAQLRLAEG KSINITYDSRYAFATAHVHGAIYKQKRWLTSAGREIKNKEEILSLLEALHLPKRLA IIPCQGHQAKADPSIRGNQMAADRVAKQAAQGVNLLP</p> <p>MAPKKRKVGIHGVPAADKKYSIGLDIGTNSVGA VITDEYKVPSKKFKVLGNT DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLOEIFSNEMAKVD DSFFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKAD LRLIYLALAHMIKFRHFLIEGDLNPDNSDVDFLFIQLVQTYNQLFEENPINASGV DAKAILSARLSKSRLENLIALQLPGEKKNGLFNLIALSLGLTPNFKSNFDLAEDA KLQSKDTYDDDLNLLAQIGDOYADFLAAKNLSDAILSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKNREDLLRKQRTDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIYVYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGA SAQSFIERMTNFDKNLPNEKVLPHKSHLLYFYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEIVLTLTFEDREMEERLKTYYAHLFDDKVMKQLK RRRYTGWRRLSRKLINGIRDKQSGKTILDFLKSDFANRNMQLIHDDSLTFKEDI QKAQVSGQDLSLHEHIANLAGSPAICKGILQTVKVVDDELVKVMGRHKKPENIVEM ARENQTTQKGQKNSRMRMKRIEGIKELGSQILKEHPVENTQLQNEKLYLYLQN GRDMYVDQELDNRLSDYVDHVPQSFLKDDSDNKVLTNRSDKARGKSDNVFSE EVVKMKNYWRQLNNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLQVETRQI TKHV AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQYKVRINNHYH HAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKASEQIEGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQ VNIVKKTVEQTGGFSKESILPKRNSDKLJARKKDWDPKKGFFDSTVAYSVLV AKVEGKSKLKS VKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIILPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA SHYEKLKGPEDNEQKQ LFVEQHKHYLDEIEQISEFSKRVILADANLKVLSAYNKHQDKPIREQAENIHLF TLTNLGAPAAFKYFDTIDRKRYTSTKEVLDATLIHQSTGLYETRIDLSQLGGDS GGSSGSGSETPGTSESA TPSSGSGSSGLGIEDEYRLHETSTEPDVSLGSTWL SDFPQAWAETGGMGLAVRQAPLIPLKATSTPVSIOQYPMSEARLGKPHIQRL DQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNRKRVEDIHPTVPNPNLLS GLPSSHQWYTVLDLKDFAFFCLRLHPTSQPLFAFEWRDPMGSGQLTWTRLPQGF KNSPITLNEALHRDLADRIQHPLDILLQYVDDILLAATSELDCCQQTALLQTLG DLGYRASAKKAQICQKQVKYLYLLREGQRLWTEARKETVMGQVPKTPRQLR EFLGKAGFCRLIFPGFAEMAAPLYPLTKPGTLFSGWPDQKQAYQEKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTKLGPWRPVA YLSKKLDPVAAGWPPCL RMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQA MLLDITDRVQVFGPVVALNPATLLPPEEGAPHDCLLEILAEITHGTRPDLTDQPIPDAD</p>
-----------------------	-----------------	--	--	--------	---

Cas-RT(MMTV B)	Cas9 (N863A)	SGSS GGSSG SETPG TSEA TPSS GGSSG GSS	Mouse mammary tumor virus (MMTVB)	P03365	HTWYTDGSSFLQEQQRKAGAAVTTETEVWAGALPAGTSAQRAELIALTQALKM AEGKRLNVYTDTRYAFATAHIGEIYRRRGWLTSEGREIKNKSEILALLKALFLPK RLSIHCLGHQKGDSEARGNRLADQAAEAIAIKTPDPSTILLI MAPKKRKVGJHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPKSKFKVLGNT DRHSIKKNLGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVD DSFHRLEESFLVEEDKKHERHPFGNIVDEVAHEKYPTIYHLRKKLVDSITDKAD LRLIYALAHMIKFRGHFLEGDLNPDNSVDKLFQLVQTYNQLFEENPINASGV DAKAILSARLSKSRLENLJALPGEKKNGFLGNLJALSLGLTPNFKNFDLAEDA KLQLSKDYDDDDLLLAQIGDQYADLFLAAKAILSDAILSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFTKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTHRIPIYVVGPLARGNSRFAMTRKSEETITTPWNFEEVVDKGA SAQSFIERTMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEIVLTLTFEDREMIEERLKYAHLFDDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNMQLIHDDSLTFKEDI QKAQVSGQGDLSHEHIANLAGSPAICKGILQTVKVVDDELVKVMGRHKPENIVEM ARENQTTQKGQKNRERMKRIEIGELGSQLKEHPVENTQLQNEKLYLYLQN GRDMMYVDQELDNRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPS EVVKMKNYWRQLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLQVETRQI TKHV AQILDSRMNTKYDENDKLIREVKVITLKSFLVDFRKFQFYKVRINNHYH HAHDAYLNAVGTALIKKYPKLESEFYVDYKVYDVRKMKIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQ VNIVKTEVQTGGSFESILPKRNSDKLIARCKDWDPKKYGGFDSPTVAYSVLV AKVEKGSKKLKSVELLGTIMERSFEKNPIDFLEAKGYKEVKKDLIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFYLAHYEKLKGSPEDEQKQ LFVEQHKHYLDEHIEQISEFSKRIVILADANLDKVL SAYNKHDKPIREQAENIHLF TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLHQSTGLYETRIDLSQLGGDS GGSSGGSGSETPGTSESATPSSGGSSGSSVQESDSRPMHLHYLNGRRFLGLD TGADKTCIAGRDWPANWPIHQTESLQGLGMACGVARSQPLRWQHEDKSGIHH PFVITPLFTLWGRDIMDKIVRLMTDSDPDDSDQLMIGAESNLFADQISWKSQDP VWLNQWPLKQEKQLQALQQLVTEQLQLGHLEESNPWTPVFIKKSGKWRLL QDLRAVNATMHDMGALQPGLSPVAVPKGWEIIDLQDCFFNKLHPEDCKRFA FSVPSNFKRYPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRKYQDSYIVHY MDDILLAHPSRSIVDEILTSMIQLNKHGLVSTEKIQKYDNLYLGLTHIQGDSVS YQKLQIRTDKRLTLNDFQKLLGNINWIRPFLKLTGELKPLFEILNGDSNPSTRKL TPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLVQDGVVEWIIH LPHISPKVITPYDIFCTQLIJKGRHRSKELFSKDPDYIVVYPTKVQFDLLQEKEDW PISLLGFLGELVEHFLPKDPDLLTFTLQTAIIPPHMTSTTPLEKGIVIFIDGSANGRSVT
----------------	--------------	--	-----------------------------------	--------	---

Cas-RT(MPMV)	Cas9 (N863A)	SGSS GGSSG SETPG TSEA TPSS GGSSG GSS	Mason-Pfizer monkey virus (MPMV)	P07572	<p>YIQGREPIKENTQNTAQAEIVAVITAFEEVSQPNLYTDSKYVVTGLFPEJETATLS PRTKIYTELKHQLRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA MAPKKRKVGIGHVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNT DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRNRCYVLOEIFSNEMAKVD DSFFHLEESFLVEEDKKHERHPFGNIVDEVAHYEKYPTIYHLRKKLVDSIDKAD LRLIYALAHMIKFRGHFLIEGLDNPNSDVDFLQFLVQTYNQLFEENPINASGV DAKAILSARLSKSRLENIAQLPGEKKNGLFNGLIALSLGLTPNFKSNFDLAEDA KQLSKDITYDDDLNLLAQIGDQYADFLAAKNSDAILSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLALKALVRQQLPEKYKEIFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLRRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTRIPYVYGPLARGNSRFAMTRKSEETITPWNFEFVVDKGA SAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTFEDREMEERLKTYYAHLFDDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNMQLIHDDSLTFKEDI QKAQVSGQDLSLHEHIANLAGSPAIKKILQTVKVVDELVKVMGRHKPENIVEM ARENQTTQKQKNSRERMKRIEIGELGSKILKEHPVENTQLQNEKLYLYLQN GRDMYVDQELDINRLSDYDVHVPSQSLKDDSDNKNVLTNRSDKARGKSDNVPS EVVKMKNYWRQLNKLITQRKFDNLTKAERGGLSELDKAGFIKRLQVETRQI TKHV AQILDSRMNTKYDENDKLIREVKVITLKSLSVDFRDKDFYKVREINNYH HAHDAYLNAVVGITALIKKYPKLESEFYGYDYKVYDVRKMIKAEQIEGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQ VNIVKKTVEQTGGFSKESILPKRNSDKLARKKDWDPKKGFGDPTVAYSVLV AKVEGSKKLKSVKELLGITIMERSSEKKNPIDFLEAKGYKEVKKDLIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAHYEKLKGSPEDEKQK LFVEQHKHYLDEIEQISEFSKRVILADANLKVLSAYNKHRRDKPIREQAENIHLF TLTNLGAPAAAFKYFDTTIDRKRYTSTKEVLDATLHQSIITGLYETRIDLSQLGGDS GGSSGGSGSETPGTSESA TPESGGSSGGSSSTA AIDILAPQQCAEPTWKSDEPVW VDQWPLTNDKLA A A QQLVQEQLA GHITESSPWNTPIFVIKKKSGKWRLQLDL RAVNATMVLMGALQPGLSPVAIPQGYLKIIDLKDCFFSIPLHPSDQKRFASFSLPS TNFKPEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIHHYMDDI LIAGKDGQVLCQFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQK AVIRKDKLQTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDLTKGSDSPNSHRSL KEALASLEKVEIAEAEQVTHINYSPLIFLJFNLTALPTGLFWQDNPMWIHLPAS PKKVLPPYYDAIADLILGRDHSKKYFGIEPSTIQPYSKSQIDWLMQNTMWWPIAC ASFVGILDNHYPPNKLQFCCKLHTFVFPQIISKPTLNNALLVFTDGSSTGMAAYTLT DTTIKFQTNLSAQLVELQALIAVLSAFPNQPLNIYTD SAYLAHSIPLLETVAQIKH ISETAKLFLQCQQLIYNRSIPIFYIGHVRAHSGPLGPPIAQGNQRADLATKIVASNINT N</p>
--------------	--------------	--	----------------------------------	--------	--

Table 47. Exemplary Gene Writer polypeptide coding mRNAs sequences

Name	mRNA (5' to 3')	Tail
Cas9-RT(MMLV)	AGGAAAU AAGAGAGAAAAGAAGAGUAAGAAGAAAUUAAGAGCCACCAU GAAACGGACAGCCGACGGAAGCGAGUUCGAGUCACCAAAGAAGAAGCGGA AAGUCGACAAGAAGUACAGCAUCGGCCUGGACAUCCGGACCAACUCUGUG GGCUGGGCCGUGAUCACCGACGAGUACAAGGUGCCCAGCAAGAAAUUCA GGUGCUGGGCAACACCGACCGGCACAGCAUCAAGAAGAACCUGAUCGGAG CCCUGCUGUUCGACAGCGGCGAAACAGCCGAGGCCACCCGGCUGAAGAGA ACCGCCAGAAGAAGAUACACCAGACGGAAGAACCAGGUAUCUCUAUCUGCA AGAGAUUCUACAGCAACGAGAUUGGCCAAGGUGGACGACAGCUUCUCCACA GACUGGAAGAGUCCUUCUGGUGGAAGAGGAUAAGAAGCACGAGCGGCAC CCCAUCUUCGGCAACAUCGUGGACGAGGUGGCCUACCACGAGAAGUACCC CACCUCUACCAACCUGAGAAAGAAACUGGUGGACAGCACCGACAAGGCCG ACCUGCGGCUGAUCUAUCUGGCCCGUGGCCACAUCAAGUUCGGGGC CACUUCUGAUCGAGGGCGACCUAACCACGACAACAGCGACGUGGACAA GCUGUUCAUCCAGCUGGUGCAGACCUACAACCAGCUGUUCGAGGAAAACC CCAUCAACGCCAGCGGCGUGGACGCCAAGGCCAUCCUGUCUGCCAGACUGA GCAAGAGCAGACGGCUGGAAAAUCUGAUCGCCACGUGCCCGGCGAGAAG AAGAAUGGCCUGUUCGGAACCCUGAUUGCCUGAGCCUGGGCCUGACCCC CAACUUC AAGAGCAACUUCGACCUGGCCGAGGAUGCCAAACUGCAGCUGA GCAAGGACACCUACGACGACGACCUGGACAACCUGCUGGCCCAGAUCCGGC GACCAGUACGCCGACCUGUUCUGGCCGCCAAGAACCUGUCCGACGCCAUC CUGCUGAGCGACAUCUGAGAGUGAACACCGAGAUACCAAGGCCCCCCCU GAGCGCCUCUAUGAUCAAGAGAUACGACGAGCACCACCAGGACCUGACCC UGCUGAAAGCUCUCGUGCGGCAGCAGCUGCCUGAGAAGUACAAGAGAUU UUCUUCGACCAAGAGCAAGAACGGCUACGCCGGCUACAUCGAGCGGAGC CAGCCAGGAAGAGUUCUACAAGUUAUCAAGCCCAUCCUGGAAAAAGUUG ACGGCACCGAGGAACUGCUCGUGAAGCUGAACAGAGAGGACCUGCUGCGG AAGCAGCGGACCUUCGACAACGGCAGCAUCCCCACCAGAUCCACCUGGGA GAGCUGCACGCCAUUCGCGGCGGAGGAAGAUUUUACCAUUCUGAA GGACAACGGGAAAAGAUAGAGAAUCCUGACCUUCCGCAUCCCUACU ACGUGGGCCUCUGGCCAGGGGAAACAGCAGAUUCGCCUGGAUAGCAGA AAGAGCGAGGAAACCAUACCCCCUGGAACUUCGAGGAAGUGGUGGACAA GGGCGCUUCCGCCAGAGCUUCAUCGAGCGGAUGACCAACUUCGAUAAGA ACCUGCCCAACGAGAAGGUGCUGCCCAAGCACAGCCUGCUGUACGAGUAC UUCACCGUGUAUAACGAGCUGACCAAGUGAAAAUACGUGACCGAGGGAU GAGAAAGCCCGCCUUCUGAGCGGCGAGCAGAAAAAGGCCAUCGUGGACC UGCUGUUAAGACCAACCGGAAAGUGACCGUGAAGCAGCUGAAAGAGGAC UACUUAAGAAAAUCGAGUGCUUCGACUCCGUGGAAAAUCUCCGGCGUGGA AGAUCGGUUAACGCCUCCUGGGCACAUACCACGAUCUGCUGAAAAUUA UCAAGGACAAGGACUUCUGGACAAUGAGGAAAAACGAGGACAUUCUGGAA GAUAUCGUGCUGACCCUGACACUGUUUGAGGACAGAGAGAUGAUCGAGGA ACGGCUGAAAACCUAUGCCCACCUGUUCGACGACAAAGUGAUGAAGCAGC UGAAGCGGCGGAGAUACACCGGCUGGGGCAGGCUGAGCCGGAAGCUGAUC AACGGCAUCCGGGACAAGCAGUCCGGCAAGACAAUCCUGGAUUUCUGAA GUCCGACGGCUUCGCCAACAGAAACUUAUGCAGCUGAUCCACGACGACA GCCUGACCUUUAAGAGGACAUCCAGAAAGCCCAGGUGUCCGGCCAGGGC GAUAGCCUGCACGAGCACAUCGCCAUCUGGCCGGCAGCCCCGCCAUUAA GAAGGGCAUCCUGCAGACAGUGAAGGUGGUGGACGAGCUCUGAAGUGA UGGGCCGGCACAAGCCCGAGAACAUUGAUCGUAUGGAAUUGGCCAGAGAAC CAGACCACCCAGAAGGGACAGAAAGAACAGCCGCGAGAGAAUGAAGCGGAU CGAAGAGGGCAUCAAGAGCUGGGCAGCCAGAUCCUGAAAGAACACCCCG UGGAAAACACCCAGCUGCAGAACGAGAAGCUGUACCUGUACUACCUGCAG AAUGGGCGGGAUAUGUACGUGGACCAGGAACUGGACAUAACCGGCUGUC	(A)80

CGACUACGAUGUGGACCAUAUCGUGCCUCAGAGCUUUCUGAAGGACGACU
 CCAUCGACAACAAGGUGCUGACCAGAAGCGACAAGGCCCGGGGCAAGAGC
 GACAACGUGCCCUCGGAAGAGGUCGUGAAGAAGAUGAAGAAUCUGGGC
 GCAGCUGCUGAACGCCAAGCUGAUUACCCAGAGAAAAGUUCGACAAUCUGA
 CCAAGGCCGAGAGAGGCCGCCUGAGCGAACUGGAUAAGGCCGGCUUCAUC
 AAGAGACAGCUGGUGGAAACCCGGCAGAUACAAAAGCACGUGGCACAGAU
 CCUGGACUCCCGGAUGAACACUAAGUACGACGAGAAUGACAAGCUGAUCC
 GGAAGUGAAAAGUGAUCACCCUGAAGUCCAAGCUGGUGUCCGAUUUCCGG
 AAGGAUUUCCAGUUUUACAAAGUGCGCGAGAUCAACAACUACCACCACGC
 CCACGACGCCUACCUGAACGCCGUCGUGGGAACCGCCCUGAUCAAAAAGU
 ACCCUAAGCUGGAAAAGCGAGUUCGUGUACGGCGACUACAAGGUGUACGAC
 GUGCGGAAGAUGAUCGCCAAGAGCGAGCAGGAAAUCGGCAAGGCUACCGC
 CAAGUACUUCUUCUACAGCAACAUCGAACUUUUCAAGACCGAGAUUA
 CCCUGGCCAACGGCGAGAUCCGGAAGCGGCCUCUGAUCGAGACAAACGGC
 GAAACCGGGGAGAUUCGUGUGGGAUAAGGGCCGGGAUUUUGCCACCGUGCG
 GAAAGUGCUGAGCAUGCCCCAAGUGAAUAUCGUGAAAAAGACCGAGGUGC
 AGACAGGGCCCUUCAGCAAAGAGUCUAUCCUGCCCCAAGAGGAACAGCGAU
 AAGCUGAUCGCCAGAAAAGGACUGGGACCCUAAGAAGUACGGCGGCUU
 CGACAGCCCCACCGUGGCCUAUUCUGUGCUGGUGGUGGCCAAAAGUGGAAA
 AGGGCAAGUCCAAGAAACUGAAGAGUGUGAAAAGAGCUGCUGGGGAUCACC
 AUCAUGGAAAGAAGCAGCUUCGAGAAGAAUCCCAUCGACUUUCUGGAAGC
 CAAGGGCUACAAAGAAGUGAAAAAGGACCUGAUCAUCAAGCUGCCUAAGU
 ACUCCUGUUCGAGCUGGAAAACGGCCGGAAGAGAAUGCUGGCCUCUGCC
 GGCGAACUGCAGAAAGGGAACGAACUGGGCCUGCCCUCCAAAUUUGUGAA
 CUUCCUGUACCGGCCAGCCACUAUGAGAAGCUGAAGGGCUCCCCGAGG
 AUAAUGAGCAGAAACAGCUGUUUGUGGAACAGCACAAAGCACUACCUGGAC
 GAGAUCAUCGAGCAGAUACAGCGAGUUCUCCAAGAGAGUGAUCCUGGCCGA
 CGCUAAUCUGGACAAAGUGCUGUCCGCCUACAACAAGCACCGGGGAUAAGC
 CCAUCAGAGAGCAGGCCGAGAAUAUCAUCCACCGUUAUACCCUGACCAAU
 CUGGGAGCCCCUGCCGCCUUAAGUACUUUGACACCACCAUCGACCGGAAG
 AGGUACACAGCACAAAGAGGUGCUGGACGCCACCCUGAUCCACCAAGAG
 CAUACCCGGCCUGUACGAGACACGGAUCGACCUGUCUCAGCUGGGAGGUG
 ACUCUGGAGGAUCUAGCGGAGGAUCCUCUGGCAGCGAGACACCAGGAACA
 AGCGAGUCAGCAACACCAGAGAGCAGUGGCGGCAGCAGCGGCGGCAGCAG
 CACCCUAAAUUAGAAGAUGAGUAUCGGCUACAUGAGACCUCAAAAGAGC
 CAGAUUUUCUAGGGUCCACAUGGCUGUCUGAUUUUCCUCAGGCCUGG
 GCGGAAACCGGGGGCAUGGGACUGGCAGUUCGCCAAGCUCCUCUGAUCAU
 ACCUCUGAAAGCAACCUCUACCCCCGUGUCCAUAUAAACAUAACCCAUUGC
 ACAAGAAGCCAGACUGGGGAUCAAGCCCCACAUACAGAGACUGUUGGACC
 AGGGAUACUGGUACCCUGCCAGUCCCCUGGAACACGCCCCUGCUACCCG
 UUAAGAAACCAGGGACUAAUGAUUAUAGGCCUGUCCAGGAUCUGAGAGAA
 GUCAACAAGCGGGUGGAAGACAUCCACCCACCGUGCCCAACCCUUAACAAC
 CUCUUGAGCGGGCUCACCCGUGCCACCCAGUGGUACACUGUGCUUGAUUU
 AAAGGAUGCCUUUUUCUGCCUGAGACUCCACCCACCCAGUCAGCCUCUCU
 CGCCUUUGAGUGGAGAGAUCCAGAGAUGGGAAUCUCAGGACAAUUGACCU
 GGACCAGACUCCACAGGGUUUCAAAAACAGUCCACCCUGUUUAAUGAG
 GCACUGCACAGAGACCUAGCAGACUUCGGAUCCAGCACCCAGACUUGAU
 CCUGCUACAGUACGUGGAUGACUACUGCUGGCCGCCACUUCUGAGCUAG
 ACUGCCAACAAGGUACUCGGGGCCUGUUACAAACCCUAGGGAACCUCCGG
 UAUCGGGCCUCGGCCAAGAAAGGCCAAAUUUGCCAGAAACAGGUCAAGUA
 UCUGGGGUACUUCUAAAAGAGGGUCAGAGAUGGCUGACUGAGGCCAGAA
 AAGAGACUGUGAUGGGGCAGCCUACUCCGAAGACCCUCGACAACUAAGG
 GAGUCCUAGGGAAGGCAGGCUUCUGUCGCCUCUUCAUCCCUUGGUUUGC
 AGAAAUGGCAGCCCCCUGUACCCUCUACCAAACCGGGGACUCUGUUUA
 AUUGGGGCCAGACCAACAAAGGCCUAUCAAGAAUCAAGCAAGCUCU
 CUAACUGCCCCAGCCCUGGGGUUGCCAGAUUUGACUAAGCCCUUUGAACU

	<p>CUUUGUCGACGAGAAGCAGGGCUACGCCAAAGGUGUCCUAAACGCAAAAAC UGGGACCUUGGCGUCGGCCGGUGGCCUACCUGUCCAAAAAGCUAGACCCA GUAGCAGCUGGGUGGCCCCCUUGCCUACGGAUGGUAGCAGCCAUUGCCGU ACUGACAAAGGAUGCAGGCAAGCUAACCAUGGGACAGCCACUAGUCAUUC UGGCCCCCAUGCAGUAGAGGCACUAGUCAAAACAACCCCCGACCGCUGGC UUUCCAACGCCCCGAUGACUCACUAUCAGGCCUUGCUUUUGGACACGGAC CGGGUCCAGUUCGGACCGGUGGUAGCCCUGAACCCGGCUACGCUGCUCCA CUGCCUGAGGAAGGGCUGCAACACAACUGCCUUGAUUCCUGGCCGAAGC CCACGGAACCCGACCCGACCUAACGGACCAGCCGCUCCCAGACGCCGACCA CACCUGGUACACGGAUGGAAGCAGUCUCUACAAGAGGGACAGCGUAAGG CGGGAGCUGCGGUGACCACCGAGACCGAGGUAUUCUGGGCUAAAGCCCUG CCAGCCGGGACAUCCGCUCAGCGGGCUGAACUGAUAGCACUCACCCAGGCC CUAAGAUGGCAGAAGGUAAGAAGCUAAAUGUUUAUACUGAUAGCCGUUA UGCUUUUGCUACUGCCCAUAUCCAUGGAGAAAUAUACAGAAGGCGUGGGU GGCUCACAUCAGAAGGCCAAAGAGAUCAAAAUAAGACGAGAUUCUUGGCC CUACUAAAAGCCCUCUUUCUGCCCAAAAGACUUAGCAUAAUCCAUUGUCC AGGACAUCAAAAGGGACACAGCGCCGAGGCUAGAGGCAACCGGAUGGCUG ACCAAGCGGCCCGAAAGGCAGCCAUACAGAGACUCCAGACACCUCAACCC UCCUCAUAGAAAAUUAUCACCCUCUGGCGGCUCAAAAAGAACCGCCGAC GGCAGCGAAUUCGAGCCCAAGAAGAAGAGGAAAGUCUGAUUAAUUAAGCU GCCUUCUGCGGGGCUUGCCUUCUGGCCAUGCCCUUCUUCUCCCUUGCAC CUGUACCUCUUGGUCUUUGAAUAAAGCCUGAGUAGGAAGUCUAG</p>	
Cas9- RT(PERV)	<p>AGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACCAU GGCUCCCAAAAAGAAAAGGAAGGUGGGCAUUCACGGCGUGCCUGCGGCCG ACAAAAAGUACAGCAUCGGCCUUGAUUUCGGCACCAAUAGCGUGGGCUGG GCCGUUAUCACAGACGAAUACAAGGUACCCAGCAAGAAGUUAAGGUGCU GGGGAUACAGACAGGCACUCUAUCAAGAAAAACCUUUCGGGGCUCUGC UGUUUGACUCAGGCGAGACCGCCGAGGCCACCAGGUUGAAGAGGACCGCA AGGCGAAGGUACACCCGGAGGAAGAACAGGAUCUGCUAUCUGCAGGAGAU CUUCAGCAACGAGAUUGGCCAAGGUGGACGACAGCUUCUCCACAGGCUGG AGGAGAGCUUCCUUGUCGAGGAGGAUAAGAAGCACGAACGACACCCCAUC UUCGGCAACAUAGUCGACGAGGUCGCUUAUCACGAGAAGUACCCCAACCAU CUACCACUUCGAAAGAAAUUGGUGGAUAGCACCAGAUAAAGCCGACUUGC GACUUAUCUACUUGGCUCUGGCGCACAUGAUUAAGUUCAGGGGCCACUUC CUGAUCGAGGGCGACCUUAACCCCGACAACAGUGACGUAGACAAAUUGUU CAUCCAGCUUGUACAGACCUUAUACCAGCUGUUCGAGGAAAACCCUAUUA ACGCCAGCGGGGUGGAUGCGAAGGCCAUACUUAAGCGCCAGGCUGAGCAAA AGCAGGCGCUUGGAGAACCUGAUAGCCAGCUGCCCGGUGAAAAGAAGAA CGGCCUCUUCGGUAAUCUGAUUGCCUGAGCCUGGGCCUGACCCCCAACUU CAAGAGCAACUUCGACCUUGGCAGAAAGAUAGCCAAGCUGCAGUUGAGUAAGG ACACCUAUGACGACGACUUGGACAAUCUGCUCGCCCAAAUCGGCGACCAG UACGCUGACCUGUCCUCGCCGCCAAGAACCUUUCUGACGCAAUCCUGCUU</p>	(A)80

AGCGAUAUCCUAGGGUGAACACAGAGAUACCAAGGCCCCCUGAGCGC
CAGCAUGAUCAAGAGGUACGACGAGCACCAUCAGGACCUGACCCUUCUGA
AGGCCUUGGUGAGGCAGCAACUGCCCCGAGAAGUACAAGGAGAUUUUUUC
GACCAGAGCAAGAACGGCUACGCCGGCUACAUCGACGGCGGAGCCAGCCA
AGAGGAGUUCUACAAGUUCAUCAAGCCCAUCCUGGAGAAGAUUGGAGGCA
CCGAGGAGCUGCUGGUGAAGCUGAACAGGGAAGAUUUGCUCCGGAAGCAG
AGGACCUUUGACAACGGUAGCAUCCCCACCAGAUCACCUGGGCGAGCU
GCACGCAAUACUGAGGCGACAGGAGGAUUUCUACCCCUUCCUCAAGGACA
AUAGGGAGAAAAUCGAAAAGAUUCUGACCUUCAGGAUCCCCUACUACGUG
GGCCCUUUGCCAGGGGCAACAGCCGAUUCGCUUGGAUGACAAGAAAGAG
CGAGGAGACCAUACCCCCUGGAACUUCGAGGAAGUGGUGGACAAAGGAG
CAAGCGCGCAGUCUUCAUCGAACGGGAUGACCAAUUUCGACAAAAACCUG
CCUAACGAGAAGGUGCUGCCCAAGCACAGCCUGCUUUACGAGUACUUCAC
CGUGUACAACGAGCUCACCAAGGUGAAAUAUGUGACCGAGGGCAUGCGAA
AACCCGCUUCCUGAGCGGCGAGCAGAAAGAAGGCCAUCGUGGACCUGCUG
UUCAAGACCAACAGGAAGGUGACCGUGAAGCAGCUGAAGGAGGACUACUU
CAAGAAGAU CGAGUGCUUUGAUAGCGUGGAAAUAAGCGGCGUGGAGGACA
GGUUC AACGCCAGCCUGGGCACCUACCACGACUUGUUGAAGAUAAUCAA
GACAAGGAUUUCCUGGAUAAUGAGGAGAACGAGGAUUAUCUGGAGGACU
CGUGCUGACUUUGACCCUGUUUGAGGACCGAGAGAUGAUUGAAGAAAGGC
UCAAACCUACGCCACCUGUUCGACGACAAAGUGAUGAAACAACUGAAG
AGACGAAGAUACACCGGCUUGGGGCGAGCUGUCCAGGAAGCUCAUCAACGG
CAUUAAGGGACAAGCAGAGCGGCAAGACCAUCCUGGAUUUCCUGAAGUCCG
ACGGCUUCGCCAACCGAAACUUC AUGCAGCUGAUUCACGAUGACGAACU
ACCUUCAAGGAGGACAUCAGAAAGGCCAGGUUAGCGGCCAGGGCGACUC
CCUGCACGAACAUUUGCAAACCUUGGCAGGCUCUCCUGCGAUCAAGAAGG
GCAUACUGCAGACCGUUAAGGUUGUGGACGAAUUGGUCAAGGUCAUGGGC
AGGCACAAGCCCCGAAAACAUAGUUUAUGAGAUGGCCAGAGAGAACCAGAC
CACCCAAAAGGGCCAGAAGAACAGCCGGGAGCGCAUGAAAAGGAUCGAGG
AGGGUAUCAAGGAACUCGGAAGCCAGAUCCUCAAAGAGCACCCCGUGGAG
AAUACCCAGCUCAGAACGAGAAGCUGUACCUUGUACUACCUAGCAGAACCG
CAGGGACAUGUACGUUGACCAGGAGUUGGACAUC AACAGGCUUUCAGACU
AUGACGUGGAUCACAUAGUGCCCCAGAGCUUUCUAAAAGACGAUAGCAUC
GACAACAAGGUCCUGACCCGCUCCGACAAAGCCAGGGGCAAAAGCGACAA
CGUGCCAAGCGAAGAGGUGGUUAAAAAGAUGAAGAACUACUGGAGGCAAC
UGCUC AACGCGAAAUUGAUCACCCAGAGAAAGUUCGAUAACCUAGACCAAG
GCCGAGAGGGGCGGACUCUCCGAACUUGACAAAGCGGGCUUCAUAAAGAG
GCAGCUGGUCGAGACCCGACAGAUACGAAAGCAGUGGCCCAAAUCCUCG
ACAGCAGAAUGAAUACCAAGUACGAUGAGAAUGACAAACUCAUCAGGGAA
GUGAAAGUGAUUACCCUGAAGAGCAAGUUGGUGUCCGACUUUCGCAAAGA
UUUCCAGUUCUACAAGGUGAGGGGAGAUCAACAACUACCACCAUGCCCACG
ACGCAUACCUGAACGCCGUGGUCGGCACCGCCUGAUU AAGAAGUAUCCA
AAGCUGGAGUCCGAAUUGUCUACGGCGACUACAAAGUUUACGAUGUGAG
GAAGAUGAUCGCUAAGAGCGAACAGGAGAUCCGCAAGGCCACCGCUAAGU
AUUUCUUCUACAGCAACAUCAUGAACUUUUUCAAGACCGAGAUACACUU
GCCAACGGCGAAAUCAGGAAGAGGCCGCUU AUCGAGACCAACGGUGAGAC
CGGCGAGAUUCGUGGGACAAGGGCAGGGACUUCGCCACCGUGAGGAAAG
UCCUGAGCAUGCCCCAGGUGAAUAUUGUGAAAAAAACUGAGGUGCAGACA
GGCGGCUUAGCAAGGAUCCAUCCUGCCCCAAGAGGAACAGCGACAAGCU
GAUCGCCCGGAAGAAGGACUGGGACCCUAAGAAGUAUGGAGGCUUCGACA
GCCCCACCGUAGCCUACAGCGUGCUGGUGGUCGCGAAGGUAGAGAAGGGG
AAGAGCAAGAAACUGAAGAGCGUGAAGGAGCUGCUCGGCAUAACCAUCAU
GGAGAGGUCCAGCUUUGAGAAGAACCCCAUUGACUUUUUGGAAGCCAAGG
GCUACAAAGAGGUCAAAAAGGACCUGAUCAUCAAAACUCCCCAAGUACUCC
CUGUUUGAAUUGGAGAACGGCAGAAAGAGGAUGCUGGCGAGCGCUGGGGA
ACUGCAAAAGGGCAACGAACUGGCGCUGCCCAGCAAGUACGUGAAUUUUC

UGUACCUGGGCGUCCACUACGAAAAGCUGAAAGGCAGCCCCGAGGACAAC
GAGCAGAAGCAGCUGUUCGUGGAGCAGCACAAGCAUUACCUGGACGAGAU
AAUCGAGCAAAUCAGCGAGUUCAGCAAGAGGGUGAUUCUGGGCCGACGCGA
ACCUGGAUAAGGUCCUCAGCGCCUACAACAAGCACCGAGACAAACCCAUC
AGGGAGCAGGCCGAGAAUAUCAUACACCUGUUCACCCUGACAAAACUGGG
CGCACCUGCGGCAUUAACAAUACUUCGAUACCACCAUCGACAGGAAAAAGGU
ACACUAGCACUAAGGAGGUGCUGGAUGCCACCUUGAUCCACCAGUCCAUI
ACCGGCCUGUAUGAGACCAGGAUCGACCUGAGCCAGCUUGGAGGCGACUC
UGGAGGAUCUAGCGGAGGAUCCUCUGGCAGCGAGACACCAGGAACAAGCG
AGUCAGCAACACCAGAGAGCAGUGGCGGCAGCAGCGGGCGGCAGCAGCCUG
GACGACGAGUACAGACUGUAUAGCCUCUGGUGAAGCCAGAUACAACAUA
UCAGUUCUGGCUGGAACAGUUUCCACAGGCCUGGGCCGAAACAGCCGAA
UGGGCCUGGCCAAGCAGGUGCCUCCUCAGGUGAUCCAGCUGAAGGCCAGC
GCCACACCUGUGUCCGUGCGGCAGUACCCUCUGUCCAAGGAGGCUCAGGA
GGGCAUCAGACCUCACGUCCAGCGGCUGAUCCAGCAGGGGAUCCUGGUGC
CCGUGCAAAGCCCUUGGAACACCCCUUCUGGCCGUGAGAAAACCCGGCA
CAAACGACUACCGGCCUGUGCAGGACCUGAGAGAAGUGAACAAGCGGGUG
CAGGACAUCCACCCACAGUGCCAAAUCCUACAACCUGCUUUGUGCCUG
CCCCCCCAGCGCAGCUGGUACACCGUUCUGGACCUGAAAGAUGCCUUUUC
UGUCUGAGACUUAUCCUACAAGCCAGCCCCUGUUCGCCUUCGAGUGGCG
GGAUCCUGGCACCGGCCGACAGGCCAGCUGACAUGGACCAGACUGCCUC
AGGGCUUAAGAACAGCCUACCAUCUUAACGAGGCCUGCACAGAGAC
CUUGCCAACUUCAGAAUCCAACACCCACAGGUGACCCUGCUCCAGUACGUG
GAUGACCUGCUGUGGCCGGCGCCACAAAACAAGAUUGCCUGGAAGGCG
CAAGGCCUUCUGCUGGAGCUGAGCGACCUUGGGAUAUCGGGCCUCUGCUA
AGAAAGCUCAGAUUCGACGAGAGAGGUGACCUACCUGGGCUACUCUCUG
AGAGAUGGCCAAAGAUGGCUGACCGAGGCCAGAAAGAAAACCGUGGUGCA
AAUCCCCGCUCCUACAACAGCCAAGCAGGUUAGAGAGUUCUGGGAAAGG
CUGGAUUUUGCAGACUGUUAUCCACAGGCUUUGCCACCCUGGCCGCCUC
UGUACCCUGACCAAAACCUAAGGGCGAGUUCAGCUGGGCCCCAGAGCACC
AGAAGGCAUUCGACGCGAUCAAGAAGGCUCUGCUGUCCGCCUGGCCUG
GCUCUGCCCGACGUGACAAAGCCCUACCCUGUACGUGGACGAACGGAA
GGGCGUGGCUAGAGGCGUUCUGACCCAGACCCUGGGUCCUUGGAGAAGGC
CUGUGGCCUACCUCAGUAAGAAGCUGGAUCCUGUGGCCUCUGGCUGGCCU
GUGUGCCUGAAGGCCAUCGCCGCCGUGGCCAUUCUGGUCAAGGAUGCCGA
UAAGCUGACCCUAGGCCAGAAUAUCACCGUGAUCGCCCCUCACGCCUCGA
GAACAUCGUGCGGCAGCCUCCCGACAGAUUGGAUGACCAACGCCAGAAUGA
CCCACUACCAGAGCCUGUUGCUGACCGAGAGAGUGACCUUCGCCCCUCCAG
CUGCCCUGAAUCCCGCCACUCUGCUGCCCGAGGAAACCGACGAGCCUGUGA
CCCACGACUGCCACCAGCUGCUGAUCCGAGGAAACCGGCGUCAGAAAGGAC
CUGACAGAUUCCCCUCUGACCGGAGAGGUGCUGACAUGGUUACCCGACGG
CAGCAGCUACGUCGUGGAAGGCAAGCGGAUGGCCGGCGCCGUGUGGUGC
ACGGCACAGAACCAUCUGGGCUUCCAGCCUGCCUGAGGGCACCAGCGCCC
AGAAGGCCGAGCUGAUGGCCUCACACAGGCCUGCGGCUGGCUGAGGGC
AAAAGCAUCAACAUCUAUACAGACAGCCGUUACGCCUUCGCCACAGCGCA
CGUGCACGGCGCCAUCUACAAGCAGAGAGGAUGGCUGACCUUGCCGGAA
GAGAAAUAAGAACAAGGAAGAAUCCUGAGCCUGCUGGAAGCCUGCAU
CUCCCAAAGAGACUGGCCAUCAUCCACUGCCCCGGCCACCAGAAGGCCAAA
GACCCUAUCAGCAGAGGCAACCAGAUUGGCCGACCGGGUGGCCAAGCAAGC
CGCCAAAGGCGUGAAUCUGCUGCCUAGUUAUUAAGCUGCCUUCUGCGG
GGCUUGCCUUCUGGCCAUGCCCUUCUUCUCUCCCUUGCACCUGUACCUCU
GGUCUUUGAAUAAAGCCUGAGUAGGAAGUCUAG

In some embodiments, a fusion protein may comprise a Cas molecule, e.g., a mutated Cas9, e.g., a Cas-nuclease containing a mutation inhibiting (e.g., inactivating) one endonuclease active site, e.g., the Cas9 nickase Cas9(N863A). In some embodiments, the fusion protein comprises a peptide linker, e.g., a glycine serine rich flexible peptide linker, e.g., a linker as disclosed in Tables 38 and/or 42, e.g., linker 10, in Table 42. In some embodiments, the fusion protein comprises a RT domain, e.g., a RT domain comprising a sequence from Table 1, Table 3, Table 30, Table 31, Table 41, Table 44, Table 50, or a fragment or variant thereof. In some embodiments, the Cas-RT fusion protein (or nucleic acid encoding the same) is formulated with a gRNA. In some embodiments, the linker length is between 2-40 amino acids, between 5-30 amino acids, between 5-20 amino acids, between 10-20 amino acids, or between 10-15 amino acids. In some embodiments, the Cas-RT fusion proteins has similar DNA binding activity to a Cas molecule that is not fused with a RT domain. In some embodiments, a Cas-RT may comprise a RT domain comprising a mutation. In embodiments, the mutant RT domain shows increased processivity and template affinity compared to an unmutated RT domain. 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% or 10% 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% or 10% 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.

Second Strand Nicking

In some embodiments, a Gene Writer system described herein comprises nickase activity that nicks the first strand and the second strand of target DNA. As discussed herein, without wishing to be bound by theory, nicking of the first strand of the target site DNA is thought to provide a 3' OH that can be used by an RT domain to reverse transcribe a sequence of a template RNA, e.g., a heterologous object sequence. Without wishing to be bound by theory, it is thought that introducing an additional nick to the second 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 to the second strand is made by the same endonuclease domain (e.g., nickase domain) as the nick to the first strand. In some embodiments, the same Gene Writer polypeptide performs both the nick to the first strand and the nick to the second strand. In some embodiments, the Gene Writer polypeptide comprises a CRISPR/Cas domain and the additional nick to the second strand is directed by an additional nucleic acid, e.g., comprising a second gRNA directing the CRISPR/Cas domain to nick the second strand. In other embodiments, the additional second strand nick is made by a different endonuclease domain (e.g., nickase domain) than the nick to the first strand. In some embodiments, that different endonuclease domain is situated in an additional polypeptide (e.g., a system of the invention further comprises the additional polypeptide), separate from the Gene Writer polypeptide. In some embodiments, the additional polypeptide comprises an endonuclease domain (e.g., nickase domain) described herein. In some embodiments, the additional polypeptide comprises a DNA binding domain, e.g., described herein.

It is contemplated herein that the position at which the second strand nick occurs relative to the first strand nick may influence the extent to which one or more of: desired Gene Writing DNA modifications are obtained, undesired double-strand breaks (DSBs) occur, undesired insertions occur, or undesired deletions occur. Without wishing to be bound by theory, second strand nicking may occur in two general orientations: inward nicks and outward nicks.

In some embodiments, in the inward nick orientation, the RT domain polymerizes (e.g., using the template RNA (e.g., the heterologous object sequence)) away the second strand nick. In some embodiments, in the inward nick orientation, the location of the nick to the first strand and the location of the nick to the second strand are positioned between the first PAM site and second PAM site (e.g., in a scenario wherein both nicks are made by a polypeptide (e.g., a Gene Writer polypeptide) comprising a CRISPR/Cas domain). In some embodiments, in the inward

nick orientation, the location of the nick to the first strand and the location of the nick to the second strand are between the sites where the polypeptide and the additional polypeptide bind to the target DNA. In some embodiments, in the inward nick orientation, the location of the nick to the second strand is positioned on the same side of the binding sites of the polypeptide and additional polypeptide relative to the location of the nick to the first strand. In some embodiments, in the inward nick orientation, the location of the nick to the first strand and the location of the nick to the second strand are positioned between the PAM site and the site at a distance from the target site.

An example of a Gene Writer system that provides an inward nick orientation comprises a Gene Writer polypeptide comprising a CRISPR/Cas domain, a template RNA comprising a gRNA that directs nicking of the target site DNA on the first strand, and an additional nucleic acid comprising an additional gRNA that directs nicking at a site a distance from the location of the first nick, wherein the location of the first nick and the location of the second nick are between the PAM sites of the sites to which the two gRNAs direct the Gene Writer polypeptide. As a further example, another Gene Writer system that provides an inward nick orientation comprises a Gene Writer polypeptide comprising a zinc finger molecule and a first nickase domain wherein the zinc finger molecule binds to the target DNA in a manner that directs the first nickase domain to nick the first strand of the target site; an additional polypeptide comprising a CRISPR/Cas domain, and an additional nucleic acid comprising a gRNA that directs the additional polypeptide to nick a site a distance from the target site DNA on the second strand, wherein the location of the first nick and the location of the second nick are between the PAM site and the site to which the zinc finger molecule binds. As a further example, another Gene Writer system that provides an inward nick orientation comprises a Gene Writer polypeptide comprising a zinc finger molecule and a first nickase domain wherein the zinc finger molecule binds to the target DNA in a manner that directs the first nickase domain to nick the first strand of the target site; an additional polypeptide comprising a TAL effector molecule and a second nickase domain wherein the TAL effector molecule binds to a site a distance from the target site in a manner that directs the additional polypeptide to nick the second strand, wherein the location of the first nick and the location of the second nick are between the site to which the TAL effector molecule binds and the site to which the zinc finger molecule binds.

In some embodiments, in the outward nick orientation, the RT domain polymerizes (e.g., using the template RNA (e.g., the heterologous object sequence)) toward the second strand nick. In some embodiments, in the inward nick orientation when both the first and second nicks are made by a polypeptide comprising a CRISPR/Cas domain (e.g., a Gene Writer polypeptide), the first PAM site and second PAM site are positioned between the location of the nick to the first strand and the location of the nick to the second strand. In some embodiments, in the inward nick orientation, the polypeptide (e.g., the Gene Writer polypeptide) and the additional polypeptide bind to sites on the target DNA between the location of the nick to the first strand and the location of the nick to the second. In some embodiments, in the inward nick orientation, the location of the nick to the second strand is positioned on the opposite side of the binding sites of the polypeptide and additional polypeptide relative to the location of the nick to the first strand. In some embodiments, in the inward orientation, the PAM site and the site at a distance from the target site are positioned between the location of the nick to the first strand and the location of the nick to the second strand.

An example of a Gene Writer system that provides an outward nick orientation comprises a Gene Writer polypeptide comprising a CRISPR/Cas domain, a template RNA comprising a gRNA that directs nicking of the target site DNA on the first strand, and an additional nucleic acid comprising an additional gRNA that directs nicking at a site a distance from the location of the first nick, wherein the location of the first nick and the location of the second nick are outside of the PAM sites of the sites to which the two gRNAs direct the Gene Writer polypeptide (i.e., the PAM sites are between the the location of the first nick and the location of the second nick). As a further example, another Gene Writer system that provides an outward nick orientation comprises a Gene Writer polypeptide comprising a zinc finger molecule and a first nickase domain wherein the zinc finger molecule binds to the target DNA in a manner that directs the first nickase domain to nick the first strand of the target site; an additional polypeptide comprising a CRISPR/Cas domain, and an additional nucleic acid comprising a gRNA that directs the additional polypeptide to nick a site a distance from the target site DNA on the second strand, wherein the location of the first nick and the location of the second nick are outside the PAM site and the site to which the zinc finger molecule binds (i.e., the PAM site and the site to which the zinc finger molecule binds are between the the location of the first nick and the location of the second nick). As a further example, another Gene Writer system that provides an

outward nick orientation comprises a Gene Writer polypeptide comprising a zinc finger molecule and a first nickase domain wherein the zinc finger molecule binds to the target DNA in a manner that directs the first nickase domain to nick the first strand of the target site; an additional polypeptide comprising a TAL effector molecule and a second nickase domain wherein the TAL effector molecule binds to a site a distance from the target site in a manner that directs the additional polypeptide to nick the second strand, wherein the location of the first nick and the location of the second nick are outside the site to which the TAL effector molecule binds and the site to which the zinc finger molecule binds (i.e., the site to which the TAL effector molecule binds and the site to which the zinc finger molecule binds are between the location of the first nick and the location of the second nick).

Without wishing to be bound by theory, it is thought that, for Gene Writer systems where a second strand nick is provided, an outward nick orientation is preferred in some embodiments. As is described herein, an inward nick may produce a higher number of double-strand breaks (DSBs) than an outward nick orientation. DSBs may be recognized by the DSB repair pathways in the nucleus of a cell, which can result in undesired insertions and deletions. An outward nick orientation may provide a decreased risk of DSB formation, and a corresponding lower amount of undesired insertions and deletions. In some embodiments, undesired insertions and deletions are insertions and deletions not encoded by the heterologous object sequence, e.g., an insertion or deletion produced by the double-strand break repair pathway unrelated to the modification encoded by the heterologous object sequence. In some embodiments, a desired Gene Writing modification comprises a change to the target DNA (e.g., a substitution, insertion, or deletion) encoded by the heterologous object sequence (e.g., and achieved by the Gene Writer writing the heterologous object sequence into the target site). In some embodiments, the first strand nick and the second strand nick are in an outward orientation.

In addition, the distance between the first strand nick and second strand nick may influence the extent to which one or more of: desired Gene Writing DNA modifications are obtained, undesired double-strand breaks (DSBs) occur, undesired insertions occur, or undesired deletions occur. Without wishing to be bound by theory, it is thought the second strand nick benefit, the biasing of DNA repair toward incorporation of the heterologous object sequence into the target DNA, increases as the distance between the first strand nick and second strand nick decreases. However, it is thought that the risk of DSB formation also increases as the distance

between the first strand nick and second strand nick decreases. Correspondingly, it is thought that the number of undesired insertions and/or deletions may increase as the distance between the first strand nick and second strand nick decreases. In some embodiments, the distance between the first strand nick and second strand nick is chosen to balance the benefit of biasing DNA repair toward incorporation of the heterologous object sequence into the target DNA and the risk of DSB formation and of undesired deletions and/or insertions. In some embodiments, a system where the first strand nick and the second strand nick are at least a threshold distance apart has an increased level of desired Gene Writing modification outcomes, a decreased level of undesired deletions, and/or a decreased level of undesired insertions relative to an otherwise similar inward nick orientation system where the first nick and the second nick are less than the a threshold distance apart. In some embodiments the threshold distance(s) is given below.

In some embodiments, the first nick and the second nick are at least 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 apart. In some embodiments, the first nick and the second nick are no more than 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, 200, or 250 nucleotides apart. In some embodiments, the first nick and the second nick are 20-200, 30-200, 40-200, 50-200, 60-200, 70-200, 80-200, 90-200, 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 20-190, 30-190, 40-190, 50-190, 60-190, 70-190, 80-190, 90-190, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 20-180, 30-180, 40-180, 50-180, 60-180, 70-180, 80-180, 90-180, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 170-180, 20-170, 30-170, 40-170, 50-170, 60-170, 70-170, 80-170, 90-170, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 20-160, 30-160, 40-160, 50-160, 60-160, 70-160, 80-160, 90-160, 100-160, 110-160, 120-160, 130-160, 140-160, 150-160, 20-150, 30-150, 40-150, 50-150, 60-150, 70-150, 80-150, 90-150, 100-150, 110-150, 120-150, 130-150, 140-150, 20-140, 30-140, 40-140, 50-140, 60-140, 70-140, 80-140, 90-140, 100-140, 110-140, 120-140, 130-140, 20-130, 30-130, 40-130, 50-130, 60-130, 70-130, 80-130, 90-130, 100-130, 110-130, 120-130, 20-120, 30-120, 40-120, 50-120, 60-120, 70-120, 80-120, 90-120, 100-120, 110-120, 20-110, 30-110, 40-110, 50-110, 60-110, 70-110, 80-110, 90-110, 100-110, 20-100, 30-100, 40-100, 50-100, 60-100, 70-100, 80-100, 90-100, 20-90, 30-90, 40-90, 50-90, 60-90, 70-90, 80-90, 20-80, 30-80, 40-80, 50-80, 60-80, 70-80, 20-70, 30-70, 40-70, 50-70, 60-70, 20-60, 30-60, 40-60, 50-60, 20-50, 30-

50, 40-50, 20-40, 30-40, or 20-30 nucleotides apart. In some embodiments, the first nick and the second nick are 40-100 nucleotides apart.

Without wishing to be bound by theory, it is thought that, for Gene Writer systems where a second strand nick is provided and an inward nick orientation is selected, increasing the distance between the first strand nick and second strand nick may be preferred. As is described herein, an inward nick orientation may produce a higher number of DSBs than an outward nick orientation, and may result in a higher amount of undesired insertions and deletions than an outward nick orientation, but increasing the distance between the nicks may mitigate that increase in DSBs, undesired deletions, and/or undesired insertions. In some embodiments, an inward nick orientation wherein the first nick and the second nick are at least a threshold distance apart has an increased level of desired Gene Writing modification outcomes, a decreased level of undesired deletions, and/or a decreased level of undesired insertions relative to an otherwise similar inward nick orientation system where the first nick and the second nick are less than the a threshold distance apart. In some embodiments the threshold distance is given below.

In some embodiments, the first strand nick and the second strand nick are in an inward orientation. In some embodiments, the first strand nick and the second strand nick are in an inward orientation and the first strand nick and second strand nick are at least 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280, 300, 350, 400, 450, or 500 nucleotides apart, e.g., at least 100 nucleotides apart, (and optionally no more than 500, 400, 300, 200, 190, 180, 170, 160, 150, 140, 130, or 120 nucleotides apart). In some embodiments, the first strand nick and the second strand nick are in an inward orientation and the first strand nick and second strand nick are 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 170-180, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 100-160, 110-160, 120-160, 130-160, 140-160, 150-160, 100-150, 110-150, 120-150, 130-150, 140-150, 100-140, 110-140, 120-140, 130-140, 100-130, 110-130, 120-130, 100-120, 110-120, or 100-110 nucleotides apart.

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-continuous 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 $5 \cdot 10^5$ cells/ml, about 10^6 cells/ml, about $5 \cdot 10^6$ cells/ml, about 10^7 cells/ml, about $5 \cdot 10^7$ cells/ml, about 10^8 cells/ml, about $5 \cdot 10^8$ cells/ml, about 10^9 cells/ml, about $5 \cdot 10^9$ cells/ml, about 10^{10} cells/ml, or about $5 \cdot 10^{10}$ 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 transcarbamylase promoter and enhancer may be used to control expression of the ornithine transcarbamylase 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](http://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	Hematopoietic 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	Hematopoietic cells
Mb promoter	muscle cells
Nphs1 promoter	podocytes
OG-2 promoter	Osteoblasts, Odonblasts
SP-B promoter	Lung
Syn1 promoter	Neurons
WASP promoter	Hematopoietic cells
SV40/bAlb promoter	Liver
SV40/bAlb promoter	Liver
SV40/Cd3 promoter	Leukocytes and platelets
SV40/CD45 promoter	hematopoietic 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)
SERPINA1 (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)
ITNNI3 (cTnl)	Troponin I type 3 (cardiac)	Cardiomyocytes (from immature state)
ITNNI3 (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	Internexin 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)	Integrin, alpha M (complement component 3 receptor 3 subunit)	Monocytes, macrophages , granulocytes , natural killer cells
Urogenital cell specific promoters		
Promoter	Gene Description	Gene Specificity
Pbsn	Probasin	Prostatic epithelium
Upk2	Uroplakin 2	Bladder
Sbp	Spermine binding protein	Prostate
Fer114	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- α (CamKII α) 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 SM22 α 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 SM22 α 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 sequence. A “**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 thyroxine 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 (α -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, Jezard 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 transgene 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	uguaguguuuccuacuuuauugga
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
Dorsal root ganglion neurons	miR-183	hsa-miR-183-3p	gugaauuaccgaaggccauaa
Hepatocytes	miR-122	hsa-miR-122-5p	uggagugugacaauagguguuug
Hepatocytes	miR-122	hsa-miR-122-3p	aacgccauuauacacuaaaaua

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 BCL-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, pages 12–17 (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 II Volume 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, Fv₃₆ or Fv). 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, 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, 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.

Resolution of Gene Writing™ events

After writing of the template nucleic acid into the target site, additional activities may be performed to increase the overall efficiency of incorporation. In some embodiments, a nick may be initiated in the genome on the non-written DNA strand to encourage copying of the newly written DNA onto the second strand. In some embodiments, the nick may be within at least 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 bases of the target site. In some embodiments, this second nick is performed by the same polypeptide performing the writing. In other embodiments, the second nick may be performed by an additional polypeptide encoding nickase activity, e.g. a Cas9 nickase.

For some Gene Writer™ systems, the writing process may leave a 3' flap containing the newly written DNA that must displace the flanking target sequence to anneal to the second

genomic strand to complete the edit. In some embodiments, the 3' flap is designed to have enhanced strand invasion capability. In some embodiments, 5'-3' exonuclease activity is supplemented to chew back the exposed 5' end of the displaced strand. In some embodiments, DNA ligase activity is supplemented to complete the reaction. In some embodiments, the exonuclease and/or ligase activities are optionally provided on the Gene Writer™ polypeptide. In some embodiments, the exonuclease and/or ligase activities are optionally provided separately from the Gene Writer™ polypeptide.

Based on the published mechanism of non-LTR retrotransposons, Gene Writing™ systems derived therefrom may not require supplementation of additional functions for resolution of the writing event. In some embodiments, the system may result in complete writing without requiring endogenous host factors. In some embodiments, the system may result in complete writing without the need for DNA repair. In some embodiments, the system may result in complete writing without eliciting a DNA damage response.

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; or a gRNA) 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., Biochimica 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 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 6, one or more chemical backbone modifications of Table 7, one or more chemically modified caps of Table 7. 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 6. 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 7. Alternatively or in combination, the nucleic acid may comprise one or more modified cap, e.g., as described herein, e.g., in Table 8. 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 6. 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-Thiophosphate)-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-
4-methoxy-pseudouridine	hydroxyisopentenyl)adenosine
4-methoxy-2-thio-pseudouridine	N6-methyl-N6-threonylcarbamoyladenosine

5-aza-cytidine	N6-hydroxynorvalylcarbamoyladenine
pseudoisocytidine	2-methylthio-N6-hydroxynorvalyl
3-methyl-cytidine	carbamoyladenine
N4-acetylcytidine	2'-O-ribosyladenine (phosphate)
5-formylcytidine	1,2'-O-dimethyladenine
N4-methylcytidine	5,2'-O-dimethylcytidine
5-hydroxymethylcytidine	N4-acetyl-2'-O-methylcytidine
1-methyl-pseudoisocytidine	Lysidine
pyrrolo-cytidine	7-methylguanosine
pyrrolo-pseudoisocytidine	N2,2'-O-dimethylguanosine
2-thio-cytidine	N2,N2,2'-O-trimethylguanosine
2-thio-5-methyl-cytidine	2'-O-ribosylguanosine (phosphate)
4-thio-pseudoisocytidine	Wybutosine
4-thio-1-methyl-pseudoisocytidine	Peroxywybutosine
4-thio-1-methyl-1-deaza-pseudoisocytidine	Hydroxywybutosine
1-methyl-1-deaza-pseudoisocytidine	undermodified hydroxywybutosine
zebularine	methylwyosine
5-aza-zebularine	queuosine
5-methyl-zebularine	epoxyqueuosine
5-aza-2-thio-zebularine	galactosyl-queuosine
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-diaminopurine	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)	5-methoxycarbonylmethyl-2-thiouridine
adenosine	5-aminomethyl-2-thiouridine
N6-glycylcarbamoyladenine	5-methylaminomethyluridine
N6-threonylcarbamoyladenine	5-methylaminomethyl-2-thiouridine
2-methylthio-N6-threonyl	5-methylaminomethyl-2-selenouridine
carbamoyladenine	5-carbamoylmethyluridine
N6,N6-dimethyladenosine	5-carbamoylmethyl-2'-O-methyluridine
7-methyladenine	5-carboxymethylaminomethyluridine
2-methylthio-adenine	5-carboxymethylaminomethyl-2'-O-
2-methoxy-adenine	methyluridine

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 7. 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
methylenimine backbone
methylenediazine backbone
amide backbone

Table 8. 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

As will be appreciated by one of skill, methods of designing and constructing nucleic acid constructs and proteins or polypeptides (such as the systems, constructs and polypeptides described herein) are routine in the art. 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, 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 part 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 further directed 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:

(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, glypyatyon, 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, 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.

Circular RNAs in Gene Writing System

Circular RNAs (circRNA) 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). 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 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) prior to translation. Circular RNAs (circRNA) 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 some embodiments, the Gene Writer™ polypeptide is encoded as circRNA. While in certain embodiments the template nucleic acid is a DNA, such as a ssDNA, in some embodiments it can be provided as an RNA, e.g., with a reverse transcriptase.

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 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, 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 some 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.

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 some embodiments, nucleic acid (e.g., encoding a polypeptide, or a template nucleic acid, or both) delivered to cells is designed as minicircles, where plasmid backbone sequences not pertaining to Gene Writing™ are removed before administration to cells. For example, a minicircle may lack a bacterial origin of replication and a selectable marker. In some embodiments, the minicircle does not comprise any bacterial sequence. Minicircles have been shown to result in higher transfection efficiencies and gene expression as compared to plasmids with backbones containing bacterial parts (e.g., bacterial origin of replication, antibiotic selection cassette) and have been used to improve the efficiency of transposition (Sharma et al Mol Ther Nucleic Acids 2013). In some embodiments, the DNA vector encoding the Gene Writer™ polypeptide is delivered as a minicircle. In some embodiments, the DNA vector containing the Gene Writer™ template nucleic acid (e.g., template RNA) is delivered as a minicircle. In some embodiments, the bacterial parts are flanked by recombination sites, e.g., attP/attB, loxP, FRT sites. In some embodiments, the addition of a cognate recombinase results in intramolecular recombination and excision of the bacterial parts. In some embodiments, the recombinase sites are recognized by phiC31 recombinase. In some embodiments, the recombinase sites are recognized by Cre recombinase. In some embodiments, the recombinase sites are recognized by FLP recombinase. In addition to plasmid DNA, minicircles can be generated by excising the desired construct, e.g., transposase expression cassettes or therapeutic expression cassette, from a viral backbone. Previously, it has been shown that excision and circularization of the donor sequence from a viral backbone may be important for transposase-mediated integration efficiency (Yant et al Nat Biotechnol 2002). In some embodiments, minicircles are first formulated and then delivered to target cells. In other embodiments, minicircles are formed from a DNA vector (e.g., plasmid DNA, rAAV, scAAV, ceDNA, doggybone DNA) intracellularly by co-delivery of a recombinase, resulting in excision and circularization of the recombinase recognition site-flanked nucleic acid, e.g., a nucleic acid encoding the Gene Writer™ polypeptide, template nucleic acid (e.g., template RNA) or nucleic acid encoding same, or both.

For optimizing protein expression, it can be helpful to provide tunable controls that can be used to modulate protein activity. In some embodiments, a tunable system may comprise at least one effector module that is responsive to at least one stimulus. The system may be, but is not limited to, a destabilizing domain (DD) system. This system is further taught in PCT/US2018/020704, as well as U.S. Provisional Patent Application No. 62/320,864 filed April

11, 2016, 62/466,596 filed March 3, 2017 and the International Publication WO2017/180587 (the contents each of which are herein incorporated by reference in their entirety). In some embodiments, the tunable system may comprise a first effector module. In some embodiments, the effector module may comprise a first stimulus response element (SRE) operably linked to at least one payload. In one aspect, the payload may be an immunotherapeutic agent. In one aspect, the first SRE of the composition may be responsive to or interact with at least one stimulus. In some embodiments, the first SRE may comprise a destabilizing domain (DD). The DD may be derived from a parent protein or from a mutant protein having one, two, three, or more amino acid mutations compared to the parent protein. In some embodiments, the parent protein may be selected from, but is not limited to, human protein FKBP, comprising the amino acid sequence of PCT/US2018/020704 SEQ. ID NO. 3; human DHFR (hDHFR), comprising the amino acid sequence of PCT/US2018/020704 SEQ. ID NO. 2; *E. coli* DHFR, comprising the amino acid sequence of PCT/US2018/020704 SEQ. ID NO. 1; PDE5, comprising the amino acid sequence of PCT/US2018/020704 SEQ. ID NO. 4; PPAR, gamma comprising the amino acid sequence of PCT/US2018/020704 SEQ. ID NO. 5; CA2, comprising the amino acid sequence of PCT/US2018/020704 SEQ. ID NO. 6; or NQO2, comprising the amino acid sequence of PCT/US2018/020704 SEQ. ID NO. 7. In some embodiments, the tunable controls are applied to the Gene Writer polypeptide, such that, e.g., a DD and stimulus can be used to modulate template integration efficiency. In some embodiments, the tunable controls are applied to one or more peptides encoded within the heterologous object sequence of the template, such that, e.g., a DD and stimulus can be used to modulate activity of a genomically integrated payload. In certain embodiments, the payload comprising the DD may be a therapeutic protein, e.g., a functional copy of an endogenously mutated gene. In certain embodiments, the payload comprising the DD may be a heterologous protein, e.g., a CAR.

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, glypyatyon, 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 proteins include an 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.

A Gene Writing™ system may be used to modify immune cells. In some embodiments, a Gene Writing™ system may be used to modify T cells. In some embodiments, T-cells may include any subpopulation of T-cells, e.g., CD4+, CD8+, gamma-delta, naïve T cells, stem cell memory T cells, central memory T cells, or a mixture of subpopulations. In some embodiments, a Gene Writing™ system may be used to deliver or modify a T-cell receptor (TCR) in a T cell. In some embodiments, a Gene Writing™ system may be used to deliver at least one chimeric antigen receptor (CAR) to T-cells. In some embodiments, a Gene Writing™ system may be used to deliver at least one CAR to natural killer (NK) cells. In some embodiments, a Gene Writing™ system may be used to deliver at least one CAR to natural killer T (NKT) cells. In some embodiments, a Gene Writing™ system may be used to deliver at least one CAR to a progenitor cell, e.g., a progenitor cell of T, NK, or NKT cells. In some embodiments, cells modified with at least one CAR (e.g., CAR-T cells, CAR-NK cells, CAR-NKT cells), or a combination of cells modified with at least one CAR (e.g., a mixture of CAR-NK/T cells) are used to treat a condition as identified in the targetable landscape of CAR therapies in MacKay, et al. Nat Biotechnol 38, 233-244 (2020), incorporated by reference herein in its entirety. In some embodiments, the immune cells comprise a CAR specific to a tumor or a pathogen antigen selected from a group consisting of AChR (fetal acetylcholine receptor), ADGRE2, AFP (alpha fetoprotein), BAFF-R, BCMA, CAIX (carbonic anhydrase IX), CCR1, CCR4, CEA (carcinoembryonic antigen), CD3, CD5, CD8, CD7, CD10, CD13, CD14, CD15, CD19, CD20, CD22, CD30, CD33, CLLI, CD34,

CD38, CD41, CD44, CD49f, CD56, CD61, CD64, CD68, CD70, CD74, CD99, CD117, CD123, CD133, CD138, CD44v6, CD267, CD269, CDS, CLEC12A, CS1, EGP-2 (epithelial glycoprotein-2), EGP-40 (epithelial glycoprotein-40), EGFR(HER1), EGFR-VIII, EpCAM (epithelial cell adhesion molecule), EphA2, ERBB2 (HER2, human epidermal growth factor receptor 2), ERBB3, ERBB4, FBP (folate-binding protein), Flt3 receptor, folate receptor-a, GD2 (ganglioside G2), GD3 (ganglioside G3), GPC3 (glypican-3), GPI00, hTERT (human telomerase reverse transcriptase), ICAM-1, integrin B7, interleukin 6 receptor, IL13Ra2 (interleukin-13 receptor 30 subunit alpha-2), kappa-light chain, KDR (kinase insert domain receptor), LeY (Lewis Y), L1CAM (LI cell adhesion molecule), LILRB2 (leukocyte immunoglobulin like receptor B2), MARTI, MAGE-A1 (melanoma associated antigen A1), MAGE- A3, MSLN (mesothelin), MUC16 (mucin 16), MUCI (mucin I), KG2D ligands, NY-ESO-1 (cancer-testis antigen), PRI (proteinase 3), TRBC1, TRBC2, TFM-3, TACI, tyrosinase, survivin, hTERT, oncofetal antigen (h5T4), p53, PSCA (prostate stem cell antigen), PSMA (prostate-specific membrane antigen), hRORI, TAG-72 (tumor- associated glycoprotein 72), VEGF-R2 (vascular endothelial growth factor R2), WT-1 (Wilms tumor protein), and antigens of HIV (human immunodeficiency virus), hepatitis B, hepatitis C, CMV (cytomegalovirus), EBV (Epstein-Barr virus), HPV (human papilloma virus).

In some embodiments, immune cells, e.g., T-cells, NK cells, NKT cells, or progenitor cells are modified ex vivo and then delivered to a patient. In some embodiments, a Gene Writer™ system is delivered by one of the methods mentioned herein, and immune cells, e.g., T-cells, NK cells, NKT cells, or progenitor cells are modified in vivo in the patient.

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 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 Writing system can be used to make an insertion, deletion, substitution, or combination thereof in a cell, tissue, or subject. 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, a Gene Writing system can be used to make multiple modifications (e.g., multiple insertions, deletions, or substitutions, and all combinations thereof) to a target cell, either simultaneously or sequentially. In some embodiments, a Gene Writing system can be used to further modify an already modified cell. In some embodiments, a Gene Writing system can be used to modify a cell edited by a complementary technology, e.g., a gene edited cell, e.g., a cell with one or more CRISPR knockouts. In some embodiments, the previously edited cell is a T-cell. In some embodiments, the previous modifications comprise gene knockouts in a T-cell, e.g., endogenous TCR (e.g., TRAC, TRBC), HLA Class I (B2M), PD1, CD52, CTLA-4, TIM-3, LAG-3, DGK. In some embodiments, a Gene Writing system is used to insert a TCR or CAR into a T-cell that has been previously modified.

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, be 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 some 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 characteristic 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 some 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 Gene Writer system described herein is used to treat an indication of any of Tables 9-12. For instance, 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 9-12, e.g., in a subject having the corresponding indication listed in any of Tables 9-12. In some embodiments, the cell is a liver cell and the target site is in a gene of Table 9, e.g., in a subject having the corresponding indication listed in Table 9. In some embodiments, the cell is an HSC and the target site is in a gene of Table 10, e.g., in a subject having the corresponding indication listed in Table 10. In some embodiments, the cell is a CNS cell and the target site is in a gene of Table 11, e.g., in a subject having the corresponding indication listed in Table 11. In some embodiments, the cell is a cell of the eye and the target site is in a gene of Table 12, e.g., in a subject having the corresponding indication listed in Table 12. In some embodiments, the target site is in a coding region in the gene. In some embodiments, the target site is in a promoter. In some embodiments, the target site is in a 5' UTR or a 3' UTR of the gene of any of Tables 9-12. In some embodiments, the target site is in an intron or exon of the gene. 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 9: Indications and genetic targets, e.g., in the liver

Disease	Gene Affected
Acute intermittent porphyria	HMBS
Alpha-1-antitrypsin deficiency (AAT)	SERPINA1

Arginase deficiency	ARG1
Argininosuccinate lyase deficiency	ASL
Carbamoyl phosphate synthetase I deficiency	CPS1
Citrin deficiency	SLC25A13
Citrullinemia type I	ASS1
Crigler-Najjar syndrome (Hyperbilirubinemia)	UGT1A1
Fabry disease	GLA
Familial hypercholesterolemia 4 (homozygous familial cholesterolemia)	LDLRAP1
Glutaric aciduria I	GCDH
Glutaric aciduria II (multiple acyl-CoA dehydrogenase deficiency)	GA IIA: ETFA GA IIB: ETFB GA IIC: ETFDH
Glycogen storage disease type IV	GBE1
Hemophilia A	F8
Hemophilia B	F9
Hereditary hemochromatosis	HFE
Homocystinuria	CBS
Maple syrup urine disease (MSUD)	Type Ia: BCKDHA Type Ib: BCKDHB Type II: DBT
Methylmalonic acidemia (methylmalonyl-CoA mutase deficiency)	MMUT
MPS 1S (Scheie syndrome)	IDUA
MPS2	IDS
MPS3 (San Filippo Syndrome)	Type IIIa: SGSH Type IIIb: NAGLU Type IIIc: HGSNAT Type IIId: GNS
MPS4	Type IVA: GALNS Type IVB: GLB1
MPS6	ARSB
MPS7	GUSB
Ornithine transcarbamylase deficiency	OTC
Phenylketonuria (phenylalanine hydroxylase deficiency)	PAH
Polycystic Liver Disease	PCLD1: PRKCSH PCLD2: SEC63 PLCD3: ALG8 PCLD4: LRP5
Pompe disease	GAA
Primary Hyperoxaluria 1 (oxalosis)	AGXT

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
Pyruvate carboxylase deficiency	PC
Tyrosinemia type I	FAH
Wilson's disease	ATP7B

Table 10: Indications and genetic targets for HSCs

Disease	Gene Affected
Adrenoleukodystrophy (CALD)	ABCD1
Alpha-mannosidosis	MAN2B1
Blackfan-Diamond Anemia	
Congenital amegakaryocytic thrombocytopenia	MPL
Dyskeratosis Congenita	TERC
Fanconi anemia	FANC
Gaucher disease	GBA
Globoid cell leukodystrophy (Krabbe disease)	GALC
Hemophagocytic lymphohistiocytosis	PRF1; STX11; STXBP2; UNC13D
Malignant infantile osteopetrosis- autosomal recessive osteopetrosis	Many genes implicated
Metachromatic leukodystrophy	PSAP
MPS 1S (Scheie syndrome)	IDUA
MPS2	IDS
MPS7	GUSB
Mucopolidosis II	GNPTAB
Niemann-Pick disease A and B	SMPD1
Niemann-Pick disease C	NPC1
Pompe disease	GAA
Pyruvate kinase deficiency (PKD)	PKLR
Sickle cell disease (SCD)	HBB
Tay Sachs	HEXA
Thalassemia	HBB

Table 11: Indications and genetic targets for the CNS

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
Mucopolipidosis 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 12: Indications and genetic targets for the eye

Disease	Gene Affected
Achromatopsia	CNGB3
Amaurosis Congenita (LCA1)	GUCY2D
Amaurosis Congenita (LCA10)	CEP290
Amaurosis Congenita (LCA2)	RPE65
Amaurosis Congenita (LCA8)	CRB1
Choroideremia	CHM
Cone Rod Dystrophy (ABCA4)	ABCA4
Cone Rod Dystrophy (GUCY2D)	GUCY2D
Cystinosis, Ocular Nonnephropathic	CTNS
Doyne Honeycomb Retinal Dystrophy (DHRD)	EFEMP1
Familial Oculoleptomeningeal Amyloidosis	TTR
Keratitis-ichthyosis-deafness (KID)	GJB2
Lattice corneal dystrophy type I	TGFBI
Macular Corneal Dystrophy (MCD)	CHST6
Meesmann Corneal Dystrophy	KRT12; KRT3
Optic Atrophy	OPA1
Retinitis Pigmentosa (AR)	USH2A
Retinitis Pigmentosa (AD)	RHO
Sorsby Fundus Dystrophy	TIMP3
Stargardt Disease	ABCA4

Additional 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;

Mucopolysaccharidosis 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-Sjogren 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; Supravalvular 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-pyruvoyl-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; Vitelliform dystrophy; type II, type IV, IV (combined hepatic and myopathic), type V, and type VI; Atypical Rett syndrome; Atrioventricular septal defect 4; Papillon-Lefevre 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 IB, 2B2, 2C, 2F, 2I, 2U (axonal), 1C (demyelinating), dominant intermediate C, recessive intermediate A, 2A2, 4C, 4D, 4H, IF, 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 1 and 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 land 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 and 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 Lange 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 ophthalmoplegia 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; Spondylocheiroadysplasia, Ehlers-Danlos syndrome-like, with immune dysregulation, Aggreacan 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, C1, 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, 11m; 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; Coproporphyrria; 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, 1b; 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; Comeal 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-Hemond-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, 1FF, 1HH, 1I, 1KK, 1N, 1S, 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; Chediak-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; Acroerythrodermatosis; 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 1 and 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 1(nonsyndromic ocular) and 4; Corneal fragility keratoglobus, blue sclerae and joint hypermobility; Microspherophakia; Chudley-McCullough syndrome; Epidermolysis 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 syndrome; 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 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 neuropathy 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 1 and 4; Hypotrichosis 8 and 12; Short QT syndrome 3; Ectodermal dysplasia 11b; 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 type 1 (with cardiac involvement) 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; Glutathione 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- α deficiency; Carcinoma of colon; Nance-Horan syndrome; Sandhoff disease, adult and infantile types; Arthrogryposis renal dysfunction cholestasis syndrome; Autosomal recessive hypophosphatemic bone disease; Doyme 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 hypomaturational 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; Reifstein 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 virus 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	SGSH
Mucopolipidosis 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

Disease	Gene Affected
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
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 Pigmentosa (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
Mucopolidosis 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

Disease	Gene Affected
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 (SEQ ID NO: 1605); 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 causative pathogenic mutations.

Disease	Gene	Pathogenic Mutation[#]
Achromatopsia	CNGB3	1148delC
Alpha-1 Antitrypsin Deficiency	SERPINA1	E342K
Alpha-1 Antitrypsin Deficiency	SERPINA1	E342K
Alpha-1 Antitrypsin Deficiency	SERPINA1	R48C (R79C)
Amaurosis Congenita (LCA10)	CEP290	2991+1655A>G
Andersen–Tawil syndrome	KCNJ2	R218W
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

Disease	Gene	Pathogenic Mutation [#]
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
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

Disease	Gene	Pathogenic Mutation [#]
hemophilia B	F9	R75*
hemophilia B	F9	R294*
hemophilia B	F9	R379Q
Hereditary antithrombin deficiency type I	SERPINC1	R48C (R79C)
hereditary chronic pancreatitis	PRSS1	R122H
Hunter syndrome	IDS	R88C
Hunter syndrome	IDS	G374G
Hurler syndrome (MPS I)	IDUA	Q70*
Hurler syndrome (MPS I)	IDUA	W402*
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	R173W
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	HPRT1	R51*
Lesch-Nyhan syndrome	HPRT1	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
Mucopolysaccharidosis IV	MCOLN1	406-2A>G
Mucopolysaccharidosis IV	MCOLN1	511_6943del
Neimann-Pick disease type A	SMPD1	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*

Disease	Gene	Pathogenic Mutation [#]
Pompe disease	GAA	c.-32-13T>G
Primary ciliary dyskinesia	DNAI1	IVS1+2_3insT
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 pigmentosa	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
Stargardt disease	ABCA4	A1038V
Stargardt disease	ABCA4	L541P
Stargardt disease	ABCA4	G1961E
Stargardt disease	ABCA4	G1961E

Disease	Gene	Pathogenic Mutation [#]
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 I	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 causative 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	SERPINA1	F51L
Alpha-1 Antitrypsin Deficiency	SERPINA1	M374I
Alpha-1 Antitrypsin Deficiency	SERPINA1	A348V/A347V
Alpha-1 Antitrypsin Deficiency	SERPINA1	K387R
Alpha-1 Antitrypsin Deficiency	SERPINA1	T59A
Alpha-1 Antitrypsin Deficiency	SERPINA1	T68A
ATTR amyloidosis	TTR	A108V
ATTR amyloidosis	TTR	R104H
ATTR amyloidosis	TTR	T119M
Cystic fibrosis	CFTR	R555K
Cystic fibrosis	CFTR	F409L
Cystic fibrosis	CFTR	F433L
Cystic fibrosis	CFTR	H667R
Cystic fibrosis	CFTR	R1070W
Cystic fibrosis	CFTR	R29K
Cystic fibrosis	CFTR	R553Q
Cystic fibrosis	CFTR	I539T
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 *bc111a*, 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
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

Disease	Gene	Nucleotide Change [#]
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.1165+2T>C
Porphyrias	ALAS1	c.1166-1A>G
Porphyrias	ALAS1	c.1331-2A>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. -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
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

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. -90 BCL11A Binding
sickle cell disease	HBG1/2	c. -90 BCL11A 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 treat 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
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

Disease	Gene	Causal repeat	Repeat location
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	CCCCGCCCGCG (SEQ ID NO: 1606)	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	TCCCCTCCAGGCCGTGCATAGTTTT AGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGGACCGAGTCGGTCTcG TCGATGGTCAGCACAGCCTTATGCA CGGCCTGGA (SEQ ID NO: 1607)	TTTGTT GAACTT GACCTC GG (SEQ ID NO: 1608)
Cystic fibrosis	Lung	CFTR	deltaF508	ACCATTAAAGAAAATATCATGTTTT AGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGGACCGAGTCGGTCCAC CAaagATGATATTTTCTTTA (SEQ ID NO: 1609)	AaagAT GATATT TTCTTT AA (SEQ ID NO: 1610)
Sickle cell	HSC	HBB	HbS	GTAACGGCAGACTTCTCCACGTTTT AGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGGACCGAGTCGGTCCGA CTCCTGaGGAGAAGTCTGCC (SEQ ID NO: 1611)	TGGTGA GGCCCT GGGCA GGT (SEQ ID NO: 1612)
Wilson's Disease	Liver	ATP7B	H1069Q	TTGGTGACTGCCACGCCCAAGTTTT AGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGGACCGAGTCGGTCCAC AcCCCTTGGGCGTGGCAGTC (SEQ ID NO: 1613)	GGCCA GCAGT GAACA cCCCT (SEQ ID NO: 1614)
ARVC	Heart	PKP2	235C>T	ACTCAGGAACACTGCTGGTTGTTTT AGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGGACCGAGTCGGTCTTC ACtGAACCAGCAGTGTTC (SEQ ID NO: 1615)	TTGGTT GAAAA TGATTT TGT (SEQ ID NO: 1616)
Long QT syndrome 1	Heart	KCNQ1	P343S	CCAGGGAAAACGCACCCACGGTTTT AGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGGACCGAGTCGGTCTCc CAGCGGTAGGTGCCCCGTGGGTGC GTTTTT (SEQ ID NO: 1617)	CTCCTT CTTTGC GCTCc AG (SEQ ID NO: 1618)
Mucopolidosis IV	CNS	MCO LN1	406-2A>G	GCCCTCCCCTTCTCTGCCAGTTTTA GAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAA AAAGTGGGACCGAGTCGGTCCGGT ACtGTGGGCAGAGAAGGGG (SEQ ID NO: 1619)	TCAGGC AACGC CAGGT ACtG (SEQ ID NO: 1620)

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 CCCCCagcagcagc agcagcagcagcagcag cagcagcagcagcagca gcagcagcagcagcag CATTCCCGGCTA CAAGGACCCT (SEQ ID NO: 1621)	ATCA CAGG ACTG GAGC TGGG (SEQ ID NO: 1622)

myotonic dystrophy 2	CNBP	NC_000003.12 (129167827..1 29183896, complement)	CCTG	Intron 1	ACCACTGCACT CCAGCCTAGGT TTTAGAGCTAG AAATAGCAAGT TAAAATAAGGC TAGTCCGTTATC AACTTGAAAAA GTGGGACCGAG TCGGTCCGTGTC TGTCTGTCTGTC TGTCTGTCTGTC TGTCTGTCTGTC TGcctgcctgcctgcctg cctgcctgcctgcctggct gcctgtctgcctgtctgcct gcctgcctgcctgcctgcc tgctgTCTGTCTC ACTTTGTCCCCT AGGCTGGAGTG CA (SEQ ID NO: 1623)	GCCT CAGC CTCC TGAG TAGC (SEQ ID NO: 1624)
fragile X mental retardation syndrome	FMR1	NC_000023.11 (147911919..1 47951127)	CGG	5' UTR	GGGGGCGTGCG GCAGCGCGGGT TTTAGAGCTAG AAATAGCAAGT TAAAATAAGGC TAGTCCGTTATC AACTTGAAAAA GTGGGACCGAG TCGGTCCTGCG GGCGCTCGAGG CCCAGcgcgcgcgcgc cgccgcgcgcgcgcgcgcg cctccgcgcgcgcgcgcgcg gcccgcgcgcgcgcgcgcg CGCTGCCGCAC G (SEQ ID NO: 1625)	GCTC AGAG GCGG CCCT CCAC (SEQ ID NO: 1626)

Friedreich's ataxia	FXN	NC_000009.12 (69035752..69079076)	GAA	Intron	CAGGCGCGCGA CACCACGCCGT TTTAGAGCTAG AAATAGCAAGT TAAAATAAGGC TAGTCCGTTATC AACTTGAAAAA GTGGGACCGAG TCGGTCCAACC CAGTATCTACTA AAAAATACAAA AAAAAAAAAAAA AAgaagaagaaga agaaAATAAAGA AAAGTTAGCCG GGCGTGGTGTC GCGC (SEQ ID NO: 1627)	CGCT TGAG CCCG GGAG GCAG (SEQ ID NO: 1628)
Huntington disease	HTT	NC_000004.12 (3074681..3243960)	CAG	Coding	GGCGGCTGAGG AAGCTGAGGGT TTTAGAGCTAG AAATAGCAAGT TAAAATAAGGC TAGTCCGTTATC AACTTGAAAAA GTGGGACCGAG TCGGTCCAGTCC CTCAAGTCCTTC cagcagcagcagcagc gcagcagcagcagcagc agcagcagcagcagcagc cagcagcaacagccgcc accgccgccgccgccgc cgccgcctctCAGCT TCCTCAG (SEQ ID NO: 1629)	CGCT GCAC CGAC CGTG AGTT (SEQ ID NO: 1630)

spinocerebellar ataxia 1	ATXN 1	NC_000006.12 (16299112..16761490, complement)	CAG	Coding	TGAGCCCCGGA GCCCTGCTGGTT TTAGAGCTAGA AATAGCAAGTT AAAATAAGGCT AGTCCGTTATCA ACTTGAAAAAG TGGGACCGAGT CGGTCCACAAG GCTGAGcagcagca gcagcagcagcagcagc agcagcagcagcatcag catcagcagcagcagca gcagcagcagcagcagc agcagcagcagCACC TCAGCAGGGCT CCGGG (SEQ ID NO: 1631)	TCCA GTTC TCCG CAGA ACAC (SEQ ID NO: 1632)
-----------------------------	-----------	---	-----	--------	---	---

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	

Therapeutic peptide	Category	Patent Number
Interferon gamma-1b	Immunosuppressive Agents	US6936695
Interferon alpha-2a, Recombi nt		CA2172664
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 alphascon-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

Therapeutic peptide	Category	Patent Number
Muromo b	Immunologic Factors and Immunosuppressive Agents	
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	Gastrointestinal Agents and Enzyme Replacement Agents	
Streptokinase	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
Choriogonadotropin alpha	Fertility Agents and Gonadotropins	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		

Therapeutic peptide	Category	Patent Number
Ranibizumab	Ophthalmics	CA2286330
Idursulfase	Enzyme Replacement Agents	
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 Monoclonal antibodies	
Ipilimumab	Antineoplastic Agents and Monoclonal antibodies	CA2381770
Sulodexide	Antithrombins and Fibrinolytic Agents and Hypoglycemic Agents and Anticoagulants and Hypolipidemic Agents	
Tocilizumab		CA2201781
Teriparatide	Bone Density Conservation Agents	US6977077
Pertuzumab	Monoclonal antibodies	CA2376596
Rilotriptan	Immunosuppressive Agents	US5844099
Denosumab	Bone Density Conservation Agents and Monoclonal antibodies	CA2257247
Liraglutide		US6268343

Therapeutic peptide	Category	Patent Number
Golimumab	Antipsoriatic Agents and Monoclo l antibodies and TNF inhibitor	
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 Immunosuppressive 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

Therapeutic peptide	Category	Patent Number
Satumomab Pendetide	Diagnostic Agents	
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 Monoclonal antibodies Antineoplastic and Immunomodulating Agents	US20120328618
C1 Esterase Inhibitor (Human)		
Coagulation Factor XIII A-Subunit (Recombinant)		
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 Recombinant)		US7767429
Idarucizumab	Anticoagulant	
Immune Globulin Human	Immunologic Factors; Immunosuppressive Agents; Anti-Infective Agents	
Vedolizumab	Immunosuppressive agent, Antineoplastic agent	US2012151248
Ustekinumab	Deramptologic agent, Immunosuppressive agent, antineoplastic agent	
Turoctocog alpha		
Tuberculin Purified Protein Derivative		
Simoctocog alpha	Antihemorrhagics: 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	Pulmonary 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, Immunosuppressive 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	TNF α	Crohn's disease
Trastuzumab	HER2	Breast cancer
Alemtuzumab	CD52	Chronic myeloid leukemia
Adalimumab	TNF α	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	TNF α	Crohn's disease
Ustekinumab	IL-12/23	Psoriasis
Canakinumab	IL-1 β	Muckle-Wells syndrome
Golimumab	TNF α	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	<i>B. anthracis</i> PA	Anthrax infection
Obinutuzumab	CD20	Chronic lymphocytic leukemia
Siltuximab	IL-6	Castleman disease
Ramucirumab	VEGFR2	Gastric cancer
Vedolizumab	$\alpha 4\beta 7$ 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-17 α	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-17 α	Psoriasis
Reslizumab	IL-5	Asthma
Olaratumab	PDGFR α	Soft tissue sarcoma
Bezlotoxumab	<i>Clostridium difficile</i> enterotoxin B	Prevention of <i>Clostridium difficile</i> infection recurrence
Atezolizumab	PD-L1	Bladder cancer
Obiltoxaximab	<i>B. anthracis</i> PA	Prevention of inhalational anthrax
Inotuzumab ozogamicin	CD22	Acute lymphoblastic leukemia
Brodalumab	IL-17R	Plaque psoriasis
Guselkumab	IL-23 p19	Plaque psoriasis
Dupilumab	IL-4R α	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-5R α	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	IFN γ	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	CD79 β	Diffuse large B-cell lymphoma
Brolicizumab	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.

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 composition 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, poplar; 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 esculentum, 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 rimpai, 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.

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/US2019/048607 which is hereby incorporated by reference. The skilled artisan will understand that the components of the Gene Writer™ system may 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 embodiment 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 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.

A variety of nanoparticles can be used for delivery, such as a liposome, a lipid nanoparticle, a cationic lipid nanoparticle, an ionizable lipid nanoparticle, a polymeric nanoparticle, a gold nanoparticle, a dendrimer, a cyclodextrin nanoparticle, a micelle, or a combination of the foregoing.

Lipid nanoparticles are an 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 Patent Application WO2020014209, the teachings of which relating to fusosome design, preparation, and usage are incorporated herein by reference).

In some embodiments, the protein component(s) of the Gene Writing™ system may be pre-associated with the template nucleic acid (e.g., template RNA). For example, in some embodiments, the Gene Writer™ polypeptide may be first combined with the template nucleic acid (e.g., template RNA) to form a ribonucleoprotein (RNP) complex. In some embodiments, the RNP may be delivered to cells via, e.g., transfection, nucleofection, virus, vesicle, LNP, exosome, fusosome.

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 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)

(cactctccccatcctctccctctgtccctctgtccctctgacctgcactgtcccagcacc (SEQ ID NO: 1633)) or orosomucoid 1 (ORM1)

(caggacacagccttggatcaggacagagactgggggccatcctgccccccaacccgacatgtgtacctcagcttttccctcacttgcataataaagcttctgtgtttggaacagctaa (SEQ ID NO: 1634)) (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 (SEQ ID NO: 1603) and/or the 3' UTR comprising UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGA (SEQ ID NO: 1604), 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 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' (SEQ ID NO: 1603). In some embodiments, the 3' UTR of an mRNA component (or transcript produced from a DNA component) of the system comprises the sequence 5'-UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGA-3' (SEQ ID NO: 1604). 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.

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.

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 some 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.64R1, AAVhu.37, AAVrh.8, AAVrh.32.33, AAV8, AAV9, AAV-DJ, AAV2/8, AAVrh10, AAVLK03, AV10, AAV11, AAV 12, rh10, 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 <i>et al.</i> , <i>Mol. Ther.</i> 18 , 118–25 (2010) 2. Ginn <i>et al.</i> , <i>JHEP Reports</i> , 100065 (2019) 3. Paulk <i>et al.</i> , <i>Mol. Ther.</i> 26 , 289–303 (2018). 4. L. Lisowski <i>et al.</i> , <i>Nature</i> . 506 , 382–6 (2014). 5. L. Wang <i>et al.</i> , <i>Mol. Ther.</i> 23 , 1877–87 (2015). 6. Hausl <i>Mol Ther</i> (2010)
Lung	AAV (AAV4, AAV5, AAV6 ¹ , AAV9, H22 ²) Adenovirus (Ad5, Ad3, Ad21, Ad14) ³	1. Duncan <i>et al.</i> , <i>Mol Ther Methods Clin Dev</i> (2018) 2. Cooney <i>et al.</i> , <i>Am J Respir Cell Mol Biol</i> (2019) 3. Li <i>et al.</i> , <i>Mol Ther Methods Clin Dev</i> (2019)
Skin	AAV (AAV6 ¹ , AAV-LK19 ²)	1. Petek <i>et al.</i> , <i>Mol. Ther.</i> (2010) 2. L. Lisowski <i>et al.</i> , <i>Nature</i> . 506 , 382–6 (2014).
HSCs	Adenovirus (HAd5/35 ⁺⁺)	Wang <i>et al.</i> <i>Blood Adv</i> (2019)

In some embodiments, a pharmaceutical composition (e.g., comprising an AAV as described 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 1×10^{13} vg/ml, e.g., less than or equal to 40 ng/ml rHCP per 1×10^{13} vg/ml or 1-50 ng/ml rHCP per 1×10^{13} vg/ml. In some embodiments, the pharmaceutical composition comprises less than 10 ng rHCP per 1.0×10^{13} vg, or less than 5 ng rHCP per 1.0×10^{13} vg, less than 4 ng rHCP per 1.0×10^{13} vg, or less than 3 ng rHCP per 1.0×10^{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 5×10^6 pg/ml hcDNA per 1×10^{13} vg/ml, less than or equal to 1.2×10^6 pg/ml hcDNA per 1×10^{13} vg/ml, or 1×10^5 pg/ml hcDNA per 1×10^{13} vg/ml. In some embodiments, the residual host cell DNA in said pharmaceutical composition is less than 5.0×10^5 pg per 1×10^{13} vg, less than 2.0×10^5 pg per 1.0×10^{13} vg, less than 1.1×10^5 pg per 1.0×10^{13} vg, less than 1.0×10^5 pg hcDNA per 1.0×10^{13} vg, less than 0.9×10^5 pg hcDNA per 1.0×10^{13} vg, less than 0.8×10^5 pg hcDNA per 1.0×10^{13} vg, or any concentration in between.

In some embodiments, the residual plasmid DNA in the pharmaceutical composition is less than or equal to 1.7×10^5 pg/ml per 1.0×10^{13} vg/ml, or 1×10^5 pg/ml per $1 \times 1.0 \times 10^{13}$ vg/ml, or 1.7×10^6 pg/ml per 1.0×10^{13} vg/ml. In some embodiments, the residual DNA plasmid in the pharmaceutical composition is less than 10.0×10^5 pg by 1.0×10^{13} vg, less than 8.0×10^5 pg by 1.0×10^{13} vg or less than 6.8×10^5 pg by 1.0×10^{13} vg. In embodiments, the pharmaceutical composition comprises less than 0.5 ng per 1.0×10^{13} vg, less than 0.3 ng per 1.0×10^{13} vg, less than 0.22 ng per 1.0×10^{13} vg or less than 0.2 ng per 1.0×10^{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 1.0×10^{13} vg, less than 0.1 ng by $1.0 \times$

10^{13} vg, less than 0.09 ng by 1.0×10^{13} vg, less than 0.08 ng by 1.0×10^{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 5.0×10^{13} vg / mL, 1.2 to 3.0×10^{13} vg / mL or 1.7 to 2.3×10^{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 μ m 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 μm per container.

In one embodiment, the pharmaceutical composition has a genomic titer of 0.5 to 5.0 $\times 10^{13}$ vg / mL, 1.0 to 4.0 $\times 10^{13}$ vg / mL, 1.5 to 3.0 $\times 10^{13}$ vg / ml or 1.7 to 2.3 $\times 10^{13}$ 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 $\times 10^{13}$ 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 $\times 10^{13}$ vg, less than about 6.8 $\times 10^5$ pg of residual DNA plasmid per 1.0 $\times 10^{13}$ vg, less than about 1.1 $\times 10^5$ pg of residual hcDNA per 1.0 $\times 10^{13}$ vg, less than about 4 ng of rHCP per 1.0 $\times 10^{13}$ 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 $\times 10^{13}$ - 2.3 $\times 10^{13}$ vg / mL genomic titer, infectious titer of about 3.9 $\times 10^8$ to 8.4 $\times 10^{10}$ IU per 1.0 $\times 10^{13}$ vg, total protein of about 100-300 pg per 1.0 $\times 10^{13}$ vg, mean survival of >24 days in A7SMA mice with about 7.5 $\times 10^{13}$ 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 $\pm 20\%$, between $\pm 15\%$, between $\pm 10\%$ or within $\pm 5\%$ 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](https://doi.org/10.1038/s41573-019-0012-9), including Table 1 thereof, which is incorporated by reference in its entirety.

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 independently 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 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 to as "intein-N." The intein encoded by the dnaE-c gene may be herein referred to 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(1):446-461, 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
CATTTATACTCAGCCAGTTGCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTTCG
AATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATG
ACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTC
ATGCGAGTTGACAACCTTCCTAAT (SEQ ID NO: 1637)

DnaE Intein-N Protein:

CLSYETEILTVEYGLLPKIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL
EDGSLIRATKDHKFMFTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 1638)

DnaE Intein-C DNA:

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGG
AGTCGAAAGAGATCACAACCTTGCTCTGAAGAACGGATTCATAGCTTCTAAT (SEQ
ID NO: 1639)

Intein-C:

MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO: 1640)

Cfa-N DNA:

TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAA
AGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTC
GTTTACACACAGCCCATTTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGA
GTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGA
CCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTC
AAACAAGTGGATGGATTG CCA (SEQ ID NO: 1641)

Cfa-N Protein:

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCL
EDGSIIRATKDHKFMTTDQGMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 1642)

Cfa-C DNA:

ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGT
AAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGA
GAAAGATCACAACCTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC (SEQ ID NO:
1643)

Cfa-C Protein:

MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID
NO: 1644)

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 1-(monomethoxy-polyethyleneglycol)-2,3- dimyristoylglycerol (PEG-DMG)), PEG-dialkyloxypropyl (DAA), PEG-phospholipid, PEG- ceramide (Cer), a pegylated phosphatidylethanolamine (PEG-PE), PEG succinate diacylglycerol (PEGS-DAG) (such as 4-0-(2',3'-di(tetradecanoyloxy)propyl)-1-0-(w- methoxy(polyethoxy)ethyl) butanedioate

(PEG-S-DMG)), PEG dialkoxypopylcarbarn, 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](https://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 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, a 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), (iii), (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.

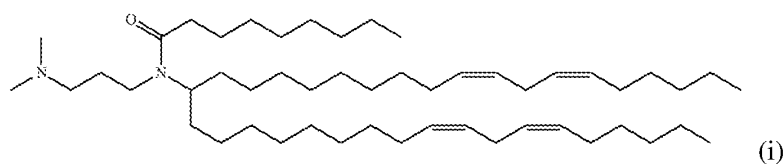
Exemplary ionizable lipids that can be used in lipid nanoparticle formulations include, without limitation, those listed in Table 1 of WO2019051289, incorporated herein by reference. Additional exemplary lipids include, without limitation, one or more of the following formulae: X of US2016/0311759; I 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 WO2013/016058; A of

WO2012/162210; 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 WO2009/132131; 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.

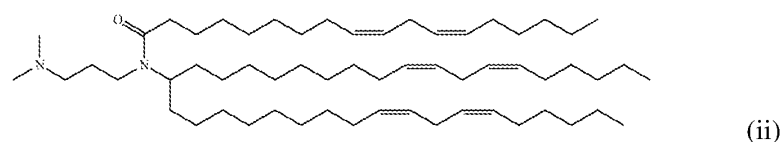
In some embodiments, the ionizable lipid is MC3 (6Z,9Z,28Z,3 IZ)-heptatriaconta-6,9,28,3 I-tetraen-19-yl-4-(dimethylamino) butanoate (DLin-MC3-DMA or MC3), e.g., as described in Example 9 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is 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 (13Z,16Z)-A,A-dimethyl-3- nonyldocosa-13, 16-dien-1-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-1H-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).

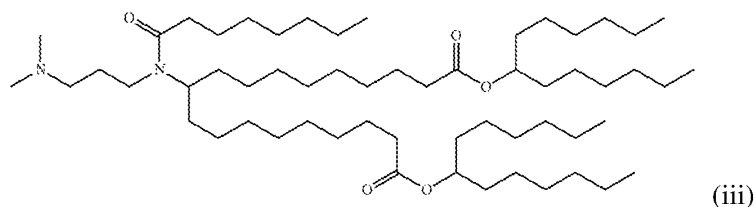
Some non-limiting example of lipid compounds that may be used (e.g., in combination with other lipid components) to form lipid nanoparticles for the delivery of compositions described herein, e.g., nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a GeneWriter) includes,



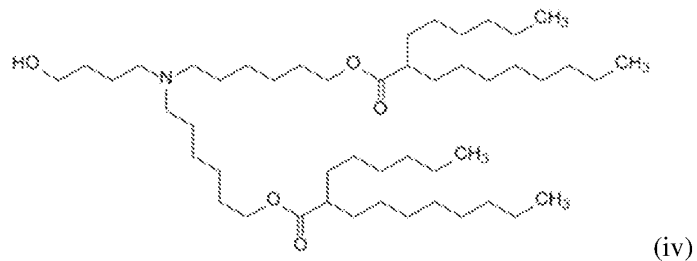
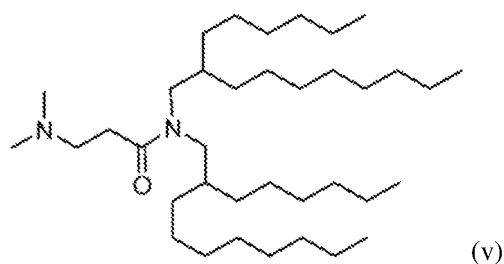
In some embodiments an LNP comprising Formula (i) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



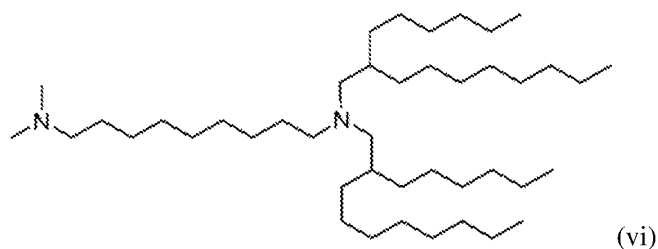
In some embodiments an LNP comprising Formula (ii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



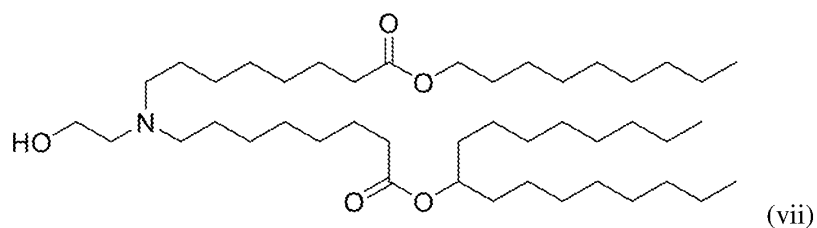
In some embodiments an LNP comprising Formula (iii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

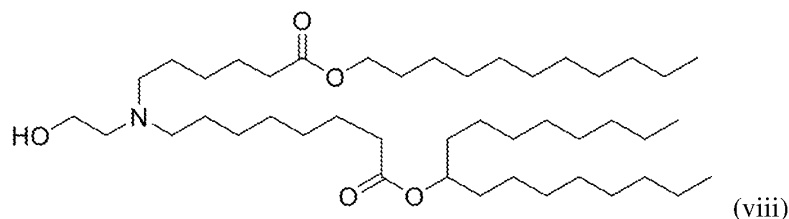


In some embodiments an LNP comprising Formula (v) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

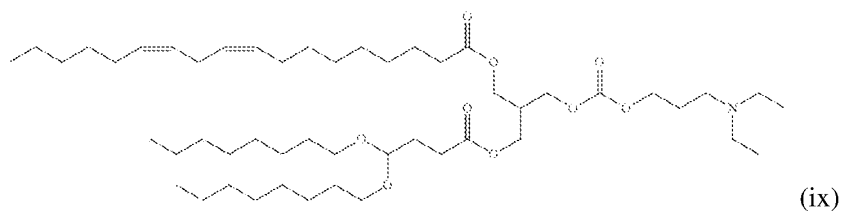


In some embodiments an LNP comprising Formula (vi) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

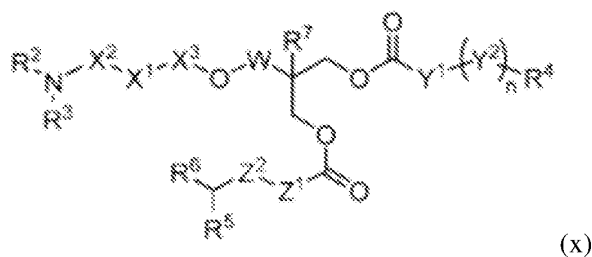




In some embodiments an LNP comprising Formula (viii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

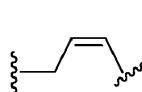
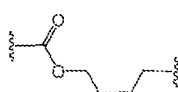
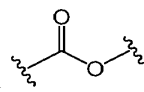


In some embodiments an LNP comprising Formula (ix) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.



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 C1-3 alkyl, R^2 is C1-3 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

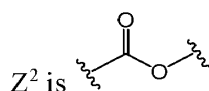


(in either orientation),
orientation),

(in either orientation),

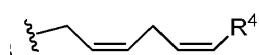
(in either

n is 0 to 3, R⁴ is C1-15 alkyl, Z¹ is C1-6 alkylene or a direct bond,



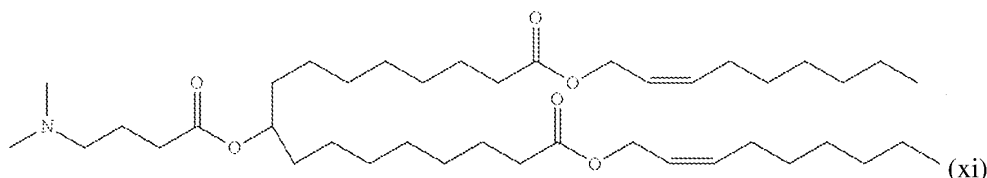
(in either orientation) or absent, provided that if Z¹ is a direct bond, Z² is absent;

R⁵ is C5-9 alkyl or C6-10 alkoxy, R⁶ is C5-9 alkyl or C6-10 alkoxy, W is methylene or a direct bond, and R⁷ is H or Me, or a salt thereof, provided that if R³ and R² are C2 alkyls, X¹ is O, X² is linear C3 alkylene, X³ is C(=O), Y¹ is linear C_n alkylene, (Y²)_n-R⁴ is

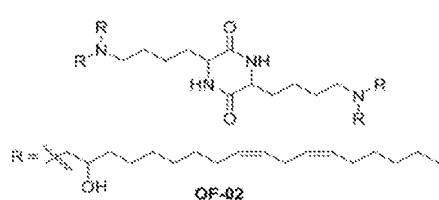


, R⁴ is linear C5 alkyl, Z¹ is C2 alkylene, Z² is absent, W is methylene, and R⁷ is H, then R⁵ and R⁶ are not C_x alkoxy.

In some embodiments an LNP comprising Formula (xii) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

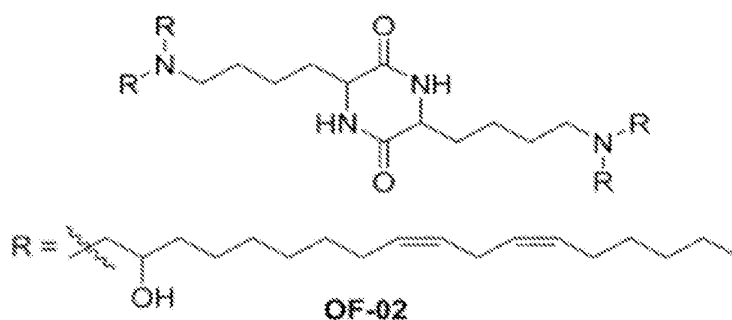


In some embodiments an LNP comprising Formula (xi) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

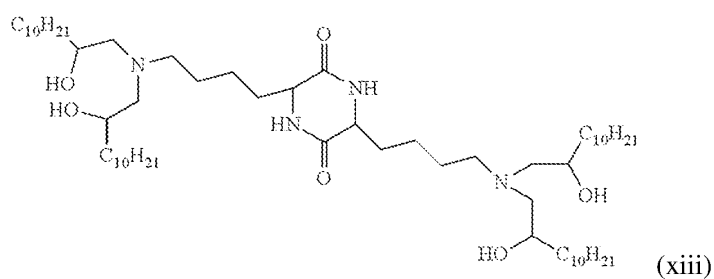
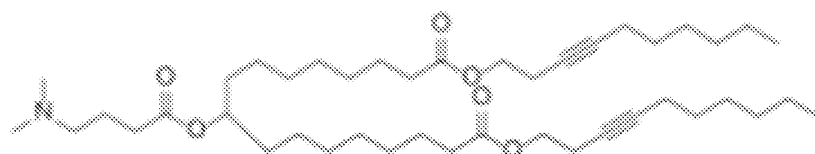


where R = (xii)

In some embodiments, a lipid of Formula (xii) can be represented by the following structure

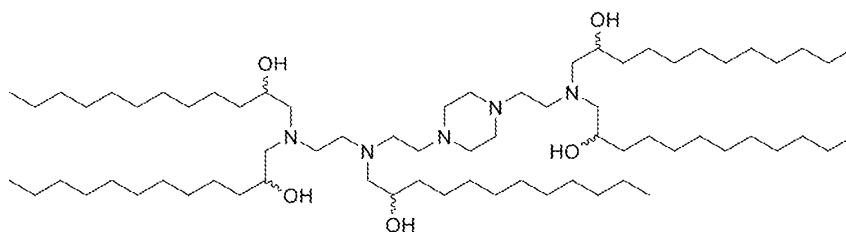


(xii) (a)



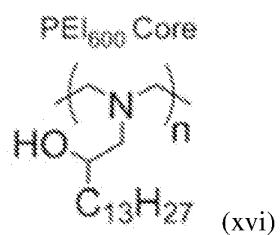
(xiv)

In some embodiments an LNP comprises a compound of Formula (xiii) and a compound of Formula (xiv).

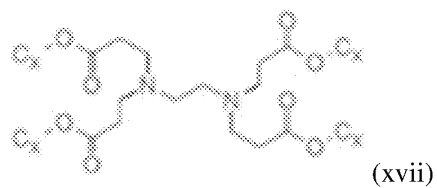
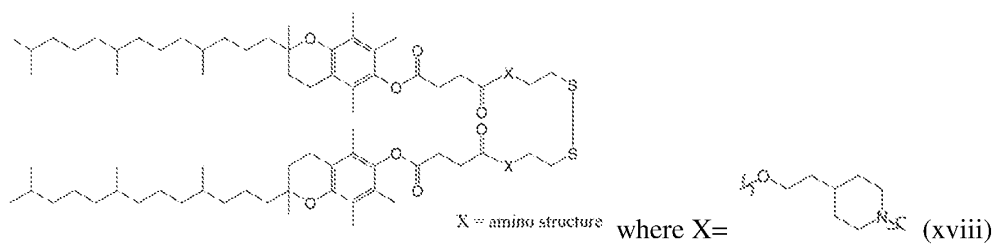


(xv)

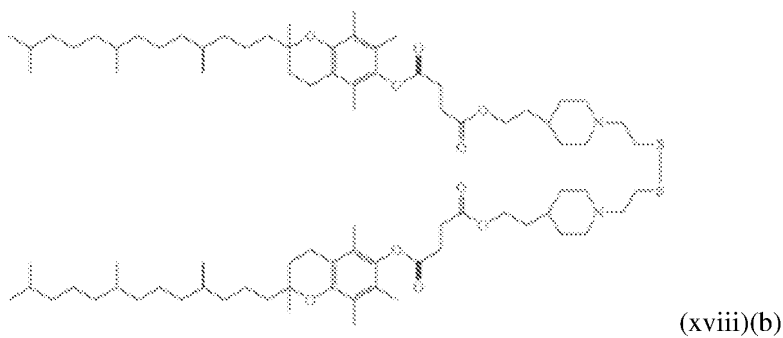
In some embodiments an LNP comprising Formula (xv) is used to deliver a GeneWriter composition described herein to the liver and/or hepatocyte cells.

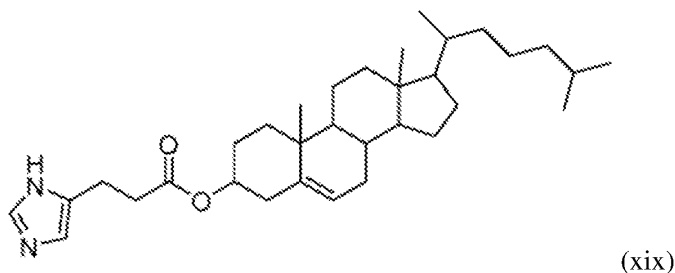


In some embodiments an LNP comprising a formulation of Formula (xvi) is used to deliver a GeneWriter composition described herein to the lung endothelial cells.



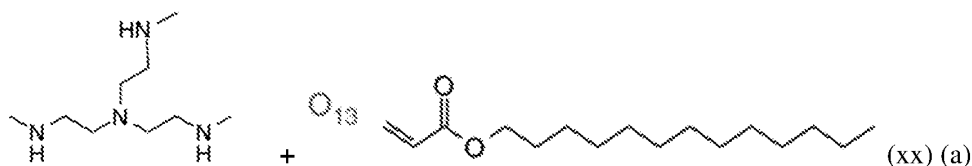
(a)



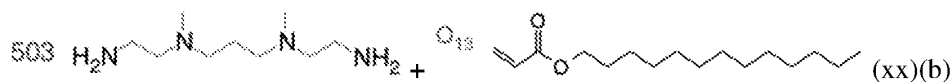


(xix)

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)

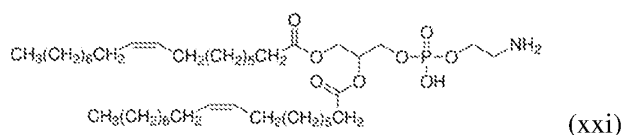


(xx)(b)

Exemplary non-cationic lipids include, but are not limited to, distearoyl-sn-glycero-phosphoethanolamine, distearoylphosphatidylcholine (DSPC), dioleoylphosphatidylcholine (DOPC), dipalmitoylphosphatidylcholine (DPPC), dioleoylphosphatidylglycerol (DOPG), dipalmitoylphosphatidylglycerol (DPPG), dioleoyl-phosphatidylethanolamine (DOPE), palmitoyl-oleoylphosphatidylcholine (POPC), palmitoyl-oleoylphosphatidylethanolamine (POPE), 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), dioleoyl-phosphatidylethanolamine 4-(N-maleimidomethyl)-cyclohexane-1-carboxylate (DOPE-mal), dipalmitoyl phosphatidylethanolamine (DPPE), dimyristoylphosphoethanolamine (DMPE), distearoyl-phosphatidylethanolamine (DSPE), monomethyl-phosphatidylethanolamine (such as 16-O-monomethyl PE), dimethyl-phosphatidylethanolamine (such as 16-O-dimethyl PE), 18-1-trans PE, 1-stearoyl-2-oleoyl-phosphatidylethanolamine (SOPE), hydrogenated soy phosphatidylcholine (HSPC), egg phosphatidylcholine (EPC), dioleoylphosphatidylserine (DOPS), sphingomyelin (SM), dimyristoyl phosphatidylcholine (DMPC), dimyristoyl phosphatidylglycerol (DMPG),

distearoylphosphatidylglycerol (DSPG), dieryucoylphosphatidylcholine (DEPC), palmitoyloleoylphosphatidylglycerol (POPG), dielaidoyl-phosphatidylethanolamine (DEPE), lecithin, phosphatidylethanolamine, lysolecithin, lysophosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, sphingomyelin, egg sphingomyelin (ESM), cephalin, cardiolipin, phosphatidic acid, 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, palmitoyl, 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](https://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



Other examples of non-cationic lipids suitable for use in the lipid nanoparticles include, without limitation, nonphosphorous lipids such as, e.g., stearylamine, dodecylamine, hexadecylamine, acetyl palmitate, glycerol ricinoleate, hexadecyl stearate, isopropyl myristate, amphoteric acrylic polymers, triethanolamine-lauryl sulfate, alkyl-aryl sulfate polyethoxylated 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.

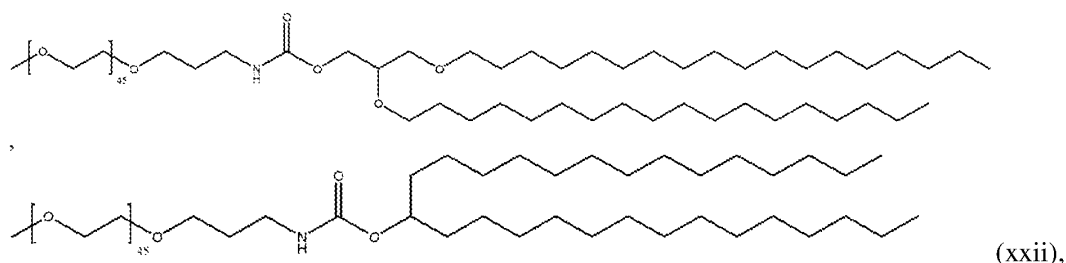
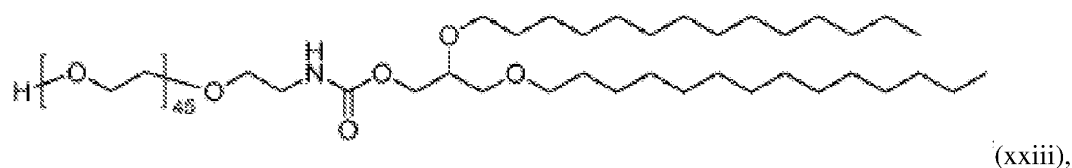
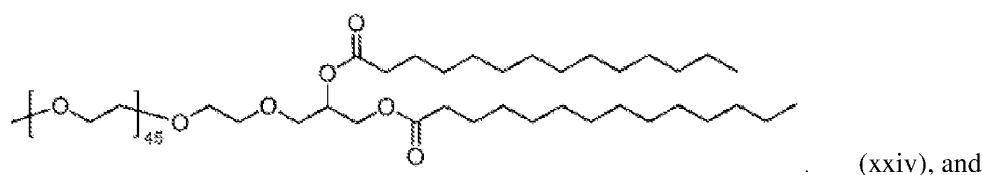
In some aspects, the lipid nanoparticle can further comprise a component, such as a sterol, to provide membrane integrity. One exemplary sterol that can be used in the lipid nanoparticle is cholesterol and derivatives thereof. Non-limiting examples of cholesterol derivatives include polar analogues such as 5 α -cholestanol, 5 β -coprostanol, choiesterolyl-(2'-hydroxy)-ethyl ether, choiesterolyl-(4'-hydroxy)-butyl ether, and 6-ketocholestanol; non-polar analogues such as 5 α -cholestane, cholestenone, 5 α -cholestanone, 5 β -cholestanone, and choiesterolyl decanoate; and mixtures thereof. In some embodiments, the cholesterol derivative is a polar analogue, e.g., choiesterolyl-(4'-hydroxy)-butyl 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.

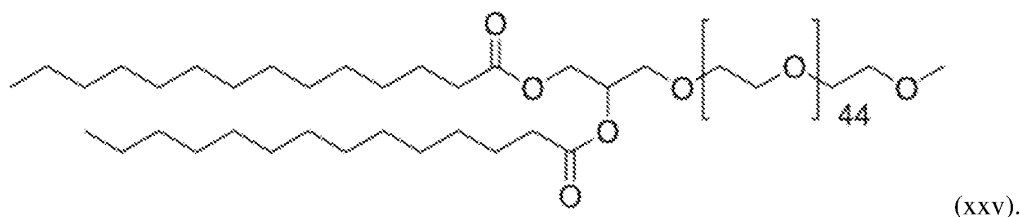
In some embodiments, the component providing membrane integrity, such as a sterol, can comprise 0-50% (mol) (e.g., 0-10%, 10-20%, 20-30%, 30-40%, or 40-50%) of the total lipid present in the lipid nanoparticle. In some embodiments, such a component is 20-50% (mol) 30-40% (mol) of the total lipid content 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 1-(monomethoxy-polyethyleneglycol)-2,3-dimyristoylglycerol (PEG-DMG)), PEG-dialkoxypentyl (DAA), PEG-phospholipid, PEG-ceramide (Cer), a pegylated phosphatidylethanolamine (PEG-PE), PEG succinate diacylglycerol (PEGS-DAG) (such as 4-0-(2',3'-di(tetradecanoyloxy)propyl-1-0-(w-methoxy(polyethoxy)ethyl) butanedioate (PEG-S-DMG)), PEG dialkoxypentylcarbam, N-(carbonyl-methoxypolyethylene glycol 2000)-1,2-distearoyl-sn-glycero-3-phosphoethanolamine sodium salt, 1,2-dimyristoyl-sn-glycerol, methoxypoly ethylene glycol (DMG-PEG-2K), or a mixture thereof. Additional exemplary PEG-lipid conjugates are described, for example, in US5,885,613, US6,287,591,

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-dipalmitoyloxypropyl, or PEG-distearoyloxypropyl. The PEG-lipid can be one or more of PEG-DMG, PEG-dilaurylglycerol, PEG-dipalmitoylglycerol, PEG- disteryl glycerol, PEG-dilaurylglycamide, PEG-dimyristylglycamide, PEG- dipalmitoylglycamide, PEG-disteryl glycamide, PEG-cholesterol (1-[8'-(Cholest-5-en-3[beta]- oxy)carboxamido-3',6'-dioxaoctanyl] carbamoyl-[omega]-methyl-poly(ethylene glycol), PEG- DMB (3,4-Ditetradecoxybenzyl- [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:





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 reagent 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

40. 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 41. 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 41.

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,12Z)-3-(((4,4-bis(octyloxy)butanoyl)oxy)-2-(((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9,12-dienoate, also called 3-(((4,4-bis(octyloxy)butanoyl)oxy)-2-(((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl (9Z,12Z)-octadeca-9,12-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 1mm 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 tablet 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 ONPATTRA.

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.

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 sequence accession

numbers specified herein, 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: Gene Writer™ enabling nucleotide substitution in genomic DNA to correct alpha-1 antitrypsin deficiency mutation in human cells

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 (E342K), 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 nucleic acid, which encodes a gRNA scaffold for polypeptide binding, a spacer for polypeptide homing, target homology domain to set up TPRT, and a template sequence for reverse transcription that includes the required substitution. An exemplary template RNA carries the sequence

(1)TCCCCTCCAGGCCGTGCATA(2)GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC(3)TcGTCGATGGTCAGCACAGCCTTAT(4)GCACGGCCTGGA (SEQ ID NO: 1607), where numbers are used to delineate the modules of the template in the order (5'-3') (1) gRNA spacer, (2) gRNA scaffold, (3) heterologous object sequence, (4) 3' homology priming domain, and the lowercase "c"

indicates the position in the template carrying the nucleotide substitution to be written into the target site to correct the E342K mutation. An exemplary gRNA for providing a second nick as described in embodiments of this system comprises the spacer sequence TTTGTTGAACTTGACCTCGG (SEQ ID NO: 1608) and directs a Cas9 nickase to nick the second strand of the target site within the homologous region. In some embodiments, this second nick improves the efficiency of the edit.

After transfection, cells are incubated for three days to allow for expression of the Gene Writing™ system and conversion of the genomic DNA target, and 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 2: Gene Writer™ enabling short insertion in genomic DNA to correct CFTR

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence by insertion of a short string of nucleotides.

In this example, the Gene Writer™ polypeptide and writing template are provided as DNA transfected into HEK293T cells that possess the CFTR delta-F508 mutation, a common allele associated with cystic fibrosis. 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 short insertion.

To create a short insertion in the affected human CFTR locus that restores the TTT triplet coding for phenylalanine 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 3-nt insertion.

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 3: Gene Writer™ enabling deletion of genomic DNA to correct Duchenne Muscular Dystrophy (DMD)

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence by deletion of nucleotides.

One of the most common mutations found in patients with DMD is a deletion that eliminates exon 50 in the rod domain of dystrophin, which places exon 51 out of frame with preceding exons. Such a mutation results in production of truncated dystrophin, leading to the pathological effects of the disease. In order to ameliorate disease, the remainder of the 79 total exons, the splice acceptor site is deleted from exon 51, resulting in restoration of the full-length protein, an approach known as exon skipping.

In this example, the Gene Writer™ polypeptide and writing template are provided as RNA nucleofected into cells containing a deletion in exon 50 that results in a truncated dystrophin product, as described above. Target cells are either patient-derived iPSCs containing the mutation or are synthetically engineered using CRISPR-Cas to generate the deletion. 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 a deletion at the desired position, such that reverse transcription of the template RNA results in the generation of a new DNA strand lacking the deleted nucleotides.

To create a short deletion that removes the exon 51 5' splice acceptor site, the Gene Writer™ polypeptide is used with a specific template that encodes a spacer for polypeptide homing, target homology domain to set up TPRT, and a template sequence for reverse transcription that includes a 5-nt deletion proximal to the Gene Writer™ polypeptide-induced nick, which includes the splice acceptor site.

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. Protein analysis by Western blot is used to further confirm the expression of the restored dystrophin, as compared to the truncated dystrophin produced in non-edited cells.

Example 4: 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 5: Gene Writer™ edits not incorporating binding sequences from the template RNA

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence by insertion of a genetic payload without causing the insertion of additional sequence from the template molecule.

In this example, the Gene Writer™ polypeptide 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 writing template is designed to have homology to the target sequence, while incorporating the genetic payload (e.g. GFP gene expression unit) 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 accomplish specific insertion of a genetic payload without also incorporating extraneous template motifs (e.g. protein binding motif), the layout of the template RNA molecule is such that the protein binding sequences (e.g. UTRs) are terminal to the homology sequences used to write the new payload into the genomic target site.

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 6: Gene Writer™ genome editing in the presence of DNA repair inhibitors

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence by insertion of a genetic payload without causing the insertion of additional sequence from the template molecule.

In this example, experiments will test the effect of different DNA repair pathways on Gene Writing™ via the application of DNA repair pathway inhibitors or DNA repair pathway deficient cell lines. When applying DNA repair pathway inhibitors, PrestoBlue cell viability assay is performed first to determine the toxicity of the inhibitors and whether any normalization should be applied for following assays. SCR7 is an inhibitor for NHEJ, which is applied at a

series of dilutions during Gene Writer™ delivery. PARP protein is a nuclear enzyme that binds as homodimers to both single- and double-strand breaks. Thus, its inhibitors are used in the test of relevant DNA repair pathways, including homologous recombination repair pathway and base excision repair pathway. The experiment procedure is the same with that of SCR7. Cell lines with deficient core proteins of nucleotide excision repair (NER) pathway are used to test the effect of NER on Gene Writing™. After the delivery of the Gene Writer™ system into the cell, ddPCR is used to evaluate the retrotransposition in the context of inhibition of DNA repair pathways. Sequencing analysis is also performed to evaluate whether certain DNA repair pathways play a role in the alteration of the integration junction. In some embodiments, Gene Writing™ into the genome will not be decreased by the knockdown of DNA repair pathways, suggesting that the system does not utilize host cell repair pathways for DNA integration. In some embodiments, Gene Writing™ into the genome will not be decreased by more than 50% by the knockdown of DNA repair pathways, suggesting that the system does not rely on host cell repair pathways for DNA integration.

Example 7: Internal Gene Writer deletions demonstrating protein domain modularity

This example describes deletions in a Gene Writer polypeptide that retain functionality and further demonstrate the modularity of the DNA binding domain.

In this example, a series of experiments were performed to test the activity of various mutant retrotransposases, as well as gaining structural knowledge about the protein modularity. This experiment tested removing a polypeptide stretch after the c-myb motif in the DNA binding domain (DBD) and replacing it with a flexible linker (Figure 8a). The polypeptide stretch removed is referred to as the “natural linker” since it is the intervening region between the DNA binding motifs and the RNA binding domain. The polypeptide region removed spans the following: on the N terminal side at either, location A (predicted random coil following c-myb motif) or location B (end of predicted alpha helix that contains part of the c-myb motif) and the removed region ends at either location v1 (alpha helical region of R2Tg that preceded the predicted -1 RNA binding motif or at location v2 (C-terminal side of an alpha helical region of R2Tg that preceded the predicted -1 RNA binding motif). In place of the polypeptide stretch removed, “natural linker”, is either of two linkers (Linker A, XTEN: SGSETPGTSESATPES

(SEQ ID NO: 1023), and Linker B, 3GS: GGGG (SEQ ID NO: 1024)). For each of these mutant retrotransposases that contain different removed regions (location A – v1, location A – v2, location B – v1, or location B – v2) they were replaced with either linker A or linker B by PCR to a DNA plasmid that expressed R2Tg, thereby yielding these sequences: c-mybA – v1 replaced with 3GS linker (SEQ ID NO: 1024), c-mybA – v2 replaced with 3GS linker (SEQ ID NO: 1024), c-mybA – v1 replaced with XTEN linker, c-mybA – v2 replaced with XTEN linker, c-mybB – v1 replaced with 3GS linker (SEQ ID NO: 1024), c-mybB – v2 replaced with 3GS linker (SEQ ID NO: 1024), c-mybB – v1 replaced with XTEN linker, c-mybB – v2 replaced with XTEN linker, as shown in Table E1 below. The insertion of the linkers was verified by Sanger sequencing and the DNA plasmids were purified for transfection.

Table E1. Amino acid sequences of R2Tg mutants with linkers in place of the “natural linker” region that intervenes the DNA binding domain (DBD) and RNA binding domain.

The N-terminal DNA-binding domain is italicized and the linker connecting to the rest of the protein is in bold and underlined.

R2Tg Mutant	Amino Acid Sequence	SEQ ID NO:
R2Tg with natural linker deletion c-mybA location – v1 replaced with 3GS linker	<p><i>MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNLANSGSDFGGGGL</i> <i>GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV</i> <i>DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH</i> <i>FAAHWEGLSGLPEVYEQ LAPQPCVGETLHSSLPRDSELFVPEEGSSEKE</i> <i>SEDAPKTSPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNL</i> <i>IEHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFPKCRGPETEKAPAGE</i> <i>WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKKR</i> <i>CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRK</i> <i>PAEEPREPGTCHHTRRAAGGGGSCFGCLESISQIRTATRDKDDTVTREK</i> <i>HPKKPFQKWKMDRAIKKGNLYRFQRLFYLDGRGLAKIILDDIECLSDI</i> <i>PLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNV</i> <i>QEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGC</i> <i>RTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNP</i> <i>RQRGFIRAAGCSENKLLQTI IWSAKREHRPLGVVFVDIAKAFDVTSHQ</i> <i>HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQ</i> <i>GDPMSPLLFNLAMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDS</i> <i>WENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTTINDCAAWTI</i> <i>NGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDLFWLQRIDQAPLK</i> <i>PLQKTDILKTYTIPRLIYIADHSEVKTALETLDQKIRTAVKEWLHLPP</i> <i>CTCDAILYSSTRDGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFME</i> <i>KEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ</i></p>	1646

	KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPH RKLLLTALQLRANVYPTREFLARGRQDQYIKACRHCADIESCAHIIGNC PVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILIF VKDARALVVDVTVRYEAAKSSLEEAAA EKVRKYKHLETEVRHLTNAKDV TFVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDI VHMFASRARKSMVM	
R2Tg with natural linker deletion c- mybA location – v2 replaced with 3GS linker	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH FAAHWEGLSGLPEVYEQ LAPQPCVGETLHSSLPRDSELFVPEEGSSEKE SEDAPKTSPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNL IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVFCHFPKCRGPETEKAPAGE WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR CWTKEEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRK PAEEPREEPGTCHHTTRRAAGGGSATATRDKKDVTREKHPKPPFQKWMKD RAIKKGNLYRFQRLFYLDRGKLA KIILDDIECLSCDIPLSEIYSVFKTR WETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPD GITLGDVVKMDPEFSRTMEIFNLWLT TGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGFIRAAGCS ENLKLQTI IWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVD PHIVGLVSNMYENISTYITTKRNTHTDKI QIRVGKVGQGDPMSPLLFNLA MDPLLCLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILE TFCNL TGLKTQGGKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGE SEKYLGLQFDPWIGIARSGSLSTKLDLFWLQRIDQAPLKPLQKTDILKTYT IPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSTR DGG LGITKLAGLIPSVQARRLHRIAQSSDTMKCFMEKEKMEQLHKKLW IQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRK NEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPH RKLLLTALQLRAN VYPTREFLARGRQDQYIKACRHCADIESCAHIIGNCPVTQDARIKRHN YICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILIFVKDARALVVDV TVRYEAAKSSLEEAAA EKVRKYKHLETEVRHLTNAKDVTFVGFP LGARGK WHQDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFASRARKSM VM	1647
R2Tg with natural linker deletion c- mybA location – v1 replaced with XTEN linker	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVH FAAHWEGLSGLPEVYEQ LAPQPCVGETLHSSLPRDSELFVPEEGSSEKE SEDAPKTSPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNL IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVFCHFPKCRGPETEKAPAGE WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR CWTKEEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRK PAEEPREEPGTCHHTTRRAASGSETPGTSESATPESCFGCLESISQIRTA TRDKKDTV TREKHPKPPFQKWMKDRAIKKGNLYRFQRLFYLDRGKLA KI ILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFR ELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLT TGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIV TARLSKACPLNPRQGRGFIRAAGCSENKLLQTI IWSAKREHRPLGVVFV DIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTH TDKI QIRVGKVGQGDPMSPLLFNLA MDPLLCLEESGKGYHRGQSSITAM AFADDLVLLSDSWENMNTNISILETFCNL TGLKTQGGKCHGFYIKPTK SYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGSLSTKLD	1648

	FWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKI RTAVKEWLHLPPCTCDAILYSTRDGGGLITKLAGLIPSVQARRLHRIA QSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNN VSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKIS NHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDA DIESCAHIIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRD SNKELYKPDILFVKDARALVVDVTVRYEAAKSSLEEAAA EKVRKYKHLE TEVRHLTNAKDVTFVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAET FSTVALFSSVDIVHMFASRARKSMVM	
R2Tg with natural linker deletion c- mybA location – v2 replaced with XTEN linker	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNLANSGSDFGGGGL GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVECVH FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVEEGSSEKE SEDAPKTSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNL IEHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFPKCRGPETEKAPAGE WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKKR CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRK PAEPPREPGTCHHTRRAA SGSETPGTSESATPES TATRDKKDVTVTREK HPKKPFQKWMKDRAIKKGNLYRFLFYLDGRGLAKIILDDIECLSDI PLSEIYSVFKTRWETTGSKSLGDFKTYGKADNTAFRELITAKEIEKNV QEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGC RTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNP RQRGFIRAAGCSENKLLQTI IWSAKREHRPLGVVFDIAKAFDTVSHQ HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQ GDPMSPLLFNLAMDPLLCLEESGKGYHRGQSSITAMAFADDLVLLSDS WENMNTNISILETFCNLTGLKTQGGKCHGFYIKPTKDSYITINDCAAWTI NGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDLFWLQRIDQAPLK PLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPP CTCDAILYSTRDGGGLITKLAGLIPSVQARRLHRIAQSSDDTMKCFME KEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIP RKLTLTALQLRANVYPTREFLARGRQDQYIKACRHCDA DIESCAHIIIGNC PVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILF VKDARALVVDVTVRYEAAKSSLEEAAA EKVRKYKHLETEVRHLTNAKDV TFVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDI VHMFASRARKSMVM	1649
R2Tg with natural linker deletion c- mybB location – v1 replaced with 3GS linker	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNLANSGSDFGGGGL GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVECVH FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVEEGSSEKE SEDAPKTSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNL IEHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFPKCRGPETEKAPAGE WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKKR CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLL GGGS CFGCLESISQIRTATRDKKDVTVTREKHPKKPFQKWMKDRAIKKGNLYRFL QRLFYLDGRGLAKIILDDIECLSDIPLSEIYSVFKTRWETTGSKSLG DFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRP I TIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTI IW SAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPPLLFNLAMDPLLCLEES	1650

	GKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQ GQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDP WIGIARSGSLSTKLDFWLQRIDQAPLKPQKTDILKTYTIPRLIYIADHS EVKTALLETLQKIRTAVKEWLHLPCTCDAILYSSSTRDGGGLGITKLAG LIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIP SIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLAS QGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARG RQDQYIKACRHCADADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAK KKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLE EAAAEEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTE LGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM	
R2Tg with natural linker deletion c- mybB location – v2 replaced with 3GS linker	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNLANSGSDFGGGGL GLPLRLLRVSVGTQTSRSDWDLVSWSHPGPTSKSQQVDLVSLFPKHRV DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVEYCVH FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE SEDAPKTSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNL IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHEFPKCRGPETEKAPAGE WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR CWTKEEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL GGGS TATRDKKDVTREKHPKPFQKWMKDRAIKKGNLRFQRLFYLDRGKLA KIIILDDIECLSCDIPLSEIYSVFKTRWETTGSEKSLGDFKTYGKADNTA FRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNL WLTTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSR IVTARLSKACPLNPRQGRFIRAAGCSENKLLQTI IWSAKREHRPLGVV FVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRN THTDKIQIRVGKQGDPMSPLLFNLMADP LLCKLEESGKGYHRGQSSIT AMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGGQKCHGFYIKPT KDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGSLSTK LDFWLQRIDQAPLKPQKTDILKTYTIPRLIYIADHSEVKTALLETLQ KIRTAVKEWLHLPCTCDAILYSSSTRDGGGLGITKLAGLIPSVQARRLHR IAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDK ISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHC DADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHI RDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEAAAEEKVRKYKH LETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMA ETFSTVALFSSVDIVHMFASRARKSMVM	1651
R2Tg with natural linker deletion c- mybB location – v1 replaced with XTEN linker	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNLANSGSDFGGGGL GLPLRLLRVSVGTQTSRSDWDLVSWSHPGPTSKSQQVDLVSLFPKHRV DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVEYCVH FAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPRDSELFVPEEGSSEKE SEDAPKTSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNL IEHLKVSHGKRGVCFRCAKCGKENSNYHSVVCHEFPKCRGPETEKAPAGE WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKR CWTKEEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRL SGSE TPGTSESATPES CFGCLESISQIRTATRDKKDVTREKHPKPFQKWMK DRAIKKGNLRFQRLFYLDRGKLAKIIILDDIECLSCDIPLSEIYSVFKT RWETTGSEKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSK DRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRFIRAAGC SENKLLQTI IWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREV	1652

	DPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISIL ETFCNLTLGLKTQGGKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPG ESEKYLGLQFDPWIGIARSGSLSTKLDLFWLQRIDQAPLKPLQKTDILKTY TIPRLIYIADHSEVKTALLETLQKIRTAVKEWLHLPPCTCDAILYSSST RDGGLGITKLGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKL WIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWR KNEFKKWKLASQGRGIVNFERDKISNHWIQYRRIPHRKLLTALQLRA NVYPTREFLARGRQDQYIKACRHCDADIESCAHIIIGNCPVTQDARIKRH NYICELLLEEAKKDWVVFKEPHIRDSNKELYKPDLI FVKDARALVVDV TVRYEAAKSSLEAAAAEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARG KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKS MVM	
R2Tg with natural linker deletion c- mybB location – v2 replaced with XTEN linker	MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGL GLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRV DLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVEYCVH FAAHWEGLSGLPEVYEQ LAPQPCVGETLHSSLP RDSELFVPEEGSSEKE SEDAPKTSPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLNL IEHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFPKCRGPETEKAPAGE WICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKKR CWTKEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRRLSGSE TPGTSESATPES TATRDKKDVTVTREKHPKKPFQKWMKDRAIKKGNYLRF QRLFYLD RGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTG SFKSLG DFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPI TIGSILLRLFSRIVTARLSKACPLNPRQGRFIRAAGCSENKLLQTI IW SAKREHRPLGVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEES GKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQ GQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDP WIGIARSGSLSTKLDLFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHS EVKTALLETLQKIRTAVKEWLHLPPCTCDAILYSSSTRDGGLGITKLGL LIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIP SIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWKLAS QGRGIVNFERDKISNHWIQYRRIPHRKLLTALQLRANVYPTREFLARG RQDQYIKACRHCDADIESCAHIIIGNCPVTQDARIKRHNYICELLLEEAK KKDWVVFKEPHIRDSNKELYKPDLI FVKDARALVVDVTVRYEAAKSSLE EAAAAEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTE LGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM	1653

HEK293T cells were plated in 96-well plates and grown overnight at 37°C, 5% CO₂. The HEK293T cells were transfected with plasmids that expressed R2Tg (wild-type), R2 endonuclease mutant, and natural linker mutants. The transfection was carried out using the Fugene HD transfection reagent according to the manufacturer recommendations, where each well received 80 ng of plasmid DNA and 0.5 µL of transfection reagent. All transfections were performed in duplicate and the cells were incubated for 72 h prior to genomic DNA extraction.

Activity of the mutants was measured by a ddPCR assay that quantified the copy number of R2Tg integrations by measuring the number of 3' junction amplicons (Figure 8b).

Deletions that begin after the random coil following the c-myb DNA binding motif (location A, c-mybA) are well-tolerated with integration activity near that of wild-type R2Tg. The natural linker region deletion end point is nearly the same for either location v1 (N-terminal to the alpha helix preceding the -1 RNA binding motif) or v2 (C-terminal to the alpha helix preceding the -1 RNA binding motif). For the deletion beginning at location A and ending at location v1 or v2, replacement of this polypeptide stretch with the XTEN linker (SEQ ID NO: 1023) seems to retain the most amount of activity whereas replacement with the 3GS linker (SEQ ID NO: 1024) has approximately a 50% reduction in integration activity. For natural linker deletions that begin at location B (c-mybB), these configurations show a more marked reduction in integration activity when compared to wild-type or location A (c-mybA). The difference in activity may be related to the structure of the protein based on the position of the deletion that creates a non-optimal three dimensional structure of the retrotransposase through the location of the linker, length of the linker, or amino acid combination of the linker that is not optimal to connect location B to locations v1 or v2. Even though the N-terminal natural linker deletion start location mybB is a sub-optimal, the C-terminal end of the deletion was most tolerated at v2 with either the 3GS (SEQ ID NO: 1024) or XTEN linker and appears to be the preferential location for having a polypeptide preceding the RBD -1 region.

Example 8: Determination of target specificity of a Gene Writer endonuclease domain

This example describes using a custom genomic landing pad in human cells to determine whether there is a sequence requirement for target cleavage and subsequent integration by a Gene Writer system.

In this example, cell lines were created to have "landing pads" or stable integrations that mimic a region of rDNA that contain the R2 position to which R2 retrotransposases target for retrotransposition (see Fig 9). The integrants or landing pads were designed to either have the wild-type region sequence in and around the R2 site found in rDNA, 12-bp of sequence mutation at and around the R2 cleavage site, or 75-bp of sequence mutation at and around the R2 cleavage site (Table E2). The DNA for these different landing pads was chemically synthesized and

cloned into the pLenti-N-tGFP vector. The cloned landing pads into the lentiviral expression vector were confirmed and sequence verified by Sanger sequencing of the landing pad. The sequence verified plasmids (9µg) along with the lentiviral packaging mix (9µg, obtained from Biosettia) were transfected using Lipofectamine2000™ according the manufacturer instructions into a packaging cell line, LentiX-293T (Takara Bio). The transfected cells were incubated at 37°C, 10% CO₂ for 48 hours (including one medium change at 24hrs) and the viral particle containing medium was collected from the cell culture dish. The collected medium was filtered through a 0.2 µm filter to remove cell debris and prepared for transduction of U2OS cells. The viral containing medium was diluted in DMEM and mixed with polybrene to prepare a dilution series for transduction of U2OS cells where the final concentration of polybrene was 8 µg/ml. The U2OS cells were grown in viral containing medium for 48 hour and then split with fresh medium. The split cells were grown to confluence and transduction efficiency of the different dilutions of virus were measured by GFP expression via flow cytometry and ddPCR detection of the genomic integrated lentivirus that contained GFP and the different rDNA landing pads (WT, 12-bp mutation, or 75-bp mutation). The GFP positive cell line from the 1:10 viral medium dilution (>99% GFP+) was chosen for subsequent experiments and cryopreserved.

To test if mutations in and around the R2 cleavage position can impact the Gene Writer system activity, the R2Tg Gene Writer Driver along with a plasmid that expressed a Gene Writer transgene molecule were electroporated into the different landing pad cell lines. In order to test if the sequence in and around the cleavage site impacted the Gene Writer polypeptide sequence activity to integrate, the homology arms for the Gene Writer template molecule were designed to have 100% homology 100 bp to the left (Gene Writer molecule module A) and 100 bp to the right (Gene Writer molecule module F) of the cleavage position for each of the landing pads. The changes to the homology arms of the Gene Writer template molecule expression plasmid were introduced by PCR and were confirmed by Sanger Sequencing. Either 73 ng of the WT R2Tg Gene Writer Driver or the Endonuclease domain mutant R2Tg Gene Writer Driver expression plasmids were co-nucleofected) using nucleofection program DN100 into each of the different U2OS landing pad cell lines (WT, 12-bp mutant, or 75-bp mutant) with 177 ng of plasmids that expressed the Gene Writer template molecules that had 100% homology to either the WT landing pad, 12-bp mutant landing pad, or 75-bp mutant landing pad. After nucleofection, cells were grown at 37°C, 10% CO₂ for 3 days prior to cell lysis and genomic DNA extraction. The

extracted gDNA was measured for Gene Writer template molecule integration at the landing pad site by ddPCR. The DNA nicking activity was measured by detection of insertions, deletions, and/or a combination of both insertions and deletions at the landing pad through next-generation sequence analysis of an amplicon that was generated from the landing pad found in the gDNA.

The integration activity of the R2Tg Gene Writer is greatly reduced when the cleavage region is mutated where there is no integration of a Gene Writer template molecule in either of the 12-bp or 75-bp landing pad cell lines (Figure 10a). Furthermore, integration is not detected with Gene Writer template molecules that have homology arms that correspond to either the 12-bp or 75-bp mutant landing pads. To rule out that the lost integration activity is due to incompatible homology arms, DNA nicking activity was measured by NGS analysis of the landing pad. The nicking activity is independent of the Gene Writer template molecule as the WT R2Tg Gene Writer driver had comparable indels at the WT landing pad with the WT, 12-bp mutant, or 75-bp mutant Gene Writer template molecule (Figure 10b). The 12-bp and 75-bp landing pads, regardless of Gene Writer template molecule co-nucleofected with the WT R2Tg Gene Writer did not show any reads above background that contained indels. The level of indels was similar to the Gene Writer template driver containing endonuclease mutations.

Table E2: Exemplary Landing Pads

Landing Pad Sequence Name	Sequence 5'-> 3' (<u>rDNA</u> , underline; cleavage region , bold; <i>mutated sequence</i> , bold-italic)
WT	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTGTTGACG CGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCA ATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTT AAGGTAG CCAAA TGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGA GATTCCCACTGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGA AATTCAGTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGC GTTACCCAACCTTAATCGCCTTGCAGCACATCC (SEQ ID NO: 1654)
12-bp Mutant	GCTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTGTTGACG CGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCA ATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTT TCCAATA TGATT TGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGA GATTCCCACTGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGA AATTCAGTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGC GTTACCCAACCTTAATCGCCTTGCAGCACATCC (SEQ ID NO: 1655)
75-bp Mutant	gctcacacaggaaacagctatgaccatgattacgccaaagctggtgacg cgatgtgatttctgcccagtgctctgaatgtcaaagtgaagaaattca atgaagcgcgggtaaaccggcgaggtaactatgactctcttt ccaata tgattccaccatggcaattccatggcaccgtcaaggctgagaacgg gaagcttgatcatcaatggaa actatccagcgaaaccacagccaaggga aattcactggccgtcgTTTTACAACGTCGTGACTGGGAAAACCTGGC gttacccaacttaatcgcccttgcagcacatcc (SEQ ID NO: 1656)

In some embodiments, a Gene Writer is derived from a retrotransposase with some level of target sequence specificity in the endonuclease domain. Thus, it may be desirable to retarget the Gene Writer to a location in the genome that possesses homology to the natural target sequence recognized by an endonuclease domain, referred to as the endonuclease recognition motif (ERM). In some embodiments, this sub-target sequence may be contained in the region surrounding the nick site. In specific embodiments, a 13 nt sequence (TAAGGTAGCCAAA (SEQ ID NO: 1657)) based on the nick site of an R2 element, e.g., R2Tg, is used to search the human genome for suitable locations for retargeting the Gene Writer, wherein a heterologous DNA-binding domain is designed to localize the Gene Writer to an endogenous ERM to direct endonuclease activity and subsequent retrotransposition of a template RNA. In some embodiments, the human genome site possesses 100% identity to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13 nucleotides in the 13 nt motif. In further embodiments, the human genome site

containing the ERM is selected from Table E3, and a DNA-binding domain fusion, e.g., ZF, TAL, or dCas9 with a custom gRNA, is designed to localize the polypeptide to the site (e.g., see Example 9). In preferred embodiments, the genome site possesses a safe harbor score of at least 5, 6, 7, 8 as defined in Pellenz et al Hum Gene Ther 30, 814-282 (2019) and shown in Table E3. In some embodiments, the template RNA (or DNA encoding the template RNA) is designed such that the homology arms match the flanking genomic sequences surrounding the expected nick site at the new target.

Table E3: Human genome sites matching a 13 nt stretch around the nicking site of R2 elements.

The human genome was searched for 100% identity to the full 13 nt match or 12 consecutive nucleotides ("Match"). Chromosomal location and start and end coordinates are provided for each match. Score ("Score") is a metric evaluating each site for eight desirable safe harbor characteristics.

Chromosome	Start	End	Source	Match	Score
chr06	123749082	123749094	NC_000006.12	13	8
chr02	5035294	5035305	NC_000002.12	12	8
chr02	145760352	145760341	NC_000002.12	12	8
chr02	147034635	147034624	NC_000002.12	12	8
chr02	181792104	181792115	NC_000002.12	12	8
chr03	34017171	34017182	NC_000003.12	12	8
chr03	74784684	74784695	NC_000003.12	12	8
chr03	110093351	110093362	NC_000003.12	12	8
chr06	14459104	14459093	NC_000006.12	12	8
chr06	119620936	119620947	NC_000006.12	12	8
chr06	145123473	145123462	NC_000006.12	12	8
chr07	12024654	12024665	NC_000007.14	12	8
chr07	52001436	52001447	NC_000007.14	12	8
chr07	115339421	115339410	NC_000007.14	12	8
chr08	126384299	126384310	NC_000008.11	12	8
chr12	84083562	84083573	NC_000012.12	12	8
chrX	117646432	117646421	NC_000023.11	12	8
chr02	106547509	106547521	NC_000002.12	13	7
chr02	226038592	226038604	NC_000002.12	13	7

chr03	102522532	102522520	NC_000003.12	13	7
chr03	110933592	110933604	NC_000003.12	13	7
chr03	119752575	119752563	NC_000003.12	13	7
chr03	172868603	172868615	NC_000003.12	13	7
chr03	191985222	191985210	NC_000003.12	13	7
chr05	6213503	6213515	NC_000005.10	13	7
chr05	58295578	58295566	NC_000005.10	13	7
chr05	129844500	129844512	NC_000005.10	13	7
chr06	1454372	1454360	NC_000006.12	13	7
chr06	48973921	48973909	NC_000006.12	13	7
chr08	18663054	18663066	NC_000008.11	13	7
chr08	93499020	93499032	NC_000008.11	13	7
chr08	119753973	119753985	NC_000008.11	13	7
chr09	86856907	86856919	NC_000009.12	13	7
chr12	29955571	29955583	NC_000012.12	13	7
chr12	118529104	118529092	NC_000012.12	13	7
chr13	65656029	65656041	NC_000013.11	13	7
chr22	34266611	34266623	NC_000022.11	13	7
chrX	26651640	26651628	NC_000023.11	13	7
chrX	119194351	119194363	NC_000023.11	13	7
chrX	139180620	139180608	NC_000023.11	13	7
chr01	106465846	106465857	NC_000001.11	12	7
chr02	964160	964171	NC_000002.12	12	7
chr02	40018947	40018936	NC_000002.12	12	7
chr02	62845403	62845392	NC_000002.12	12	7
chr02	64834920	64834909	NC_000002.12	12	7
chr02	67969608	67969619	NC_000002.12	12	7
chr02	76183118	76183129	NC_000002.12	12	7
chr02	81819286	81819297	NC_000002.12	12	7
chr02	119597238	119597249	NC_000002.12	12	7
chr02	122897376	122897365	NC_000002.12	12	7
chr02	123603423	123603412	NC_000002.12	12	7
chr02	144644206	144644217	NC_000002.12	12	7
chr02	145221757	145221746	NC_000002.12	12	7
chr02	158367531	158367520	NC_000002.12	12	7
chr02	160092083	160092072	NC_000002.12	12	7
chr02	192245037	192245048	NC_000002.12	12	7
chr02	195223552	195223563	NC_000002.12	12	7
chr02	200351999	200351988	NC_000002.12	12	7

chr02	237068525	237068514	NC_000002.12	12	7
chr03	18724351	18724340	NC_000003.12	12	7
chr03	23969399	23969388	NC_000003.12	12	7
chr03	25177339	25177350	NC_000003.12	12	7
chr03	34880863	34880852	NC_000003.12	12	7
chr03	66233879	66233890	NC_000003.12	12	7
chr03	74527939	74527950	NC_000003.12	12	7
chr03	98583025	98583014	NC_000003.12	12	7
chr03	99278452	99278463	NC_000003.12	12	7
chr03	116060228	116060239	NC_000003.12	12	7
chr03	139468578	139468589	NC_000003.12	12	7
chr03	140064054	140064043	NC_000003.12	12	7
chr03	140438138	140438127	NC_000003.12	12	7
chr03	152457330	152457341	NC_000003.12	12	7
chr03	160950736	160950725	NC_000003.12	12	7
chr03	167207758	167207769	NC_000003.12	12	7
chr03	167722472	167722483	NC_000003.12	12	7
chr03	180475661	180475672	NC_000003.12	12	7
chr04	121590786	121590775	NC_000004.12	12	7
chr04	133719599	133719588	NC_000004.12	12	7
chr05	11564132	11564121	NC_000005.10	12	7
chr05	11970221	11970210	NC_000005.10	12	7
chr05	32814431	32814420	NC_000005.10	12	7
chr05	38003029	38003018	NC_000005.10	12	7
chr05	39758118	39758129	NC_000005.10	12	7
chr05	41221615	41221604	NC_000005.10	12	7
chr05	74838717	74838728	NC_000005.10	12	7
chr05	86444529	86444518	NC_000005.10	12	7
chr05	86617117	86617106	NC_000005.10	12	7
chr05	89438360	89438349	NC_000005.10	12	7
chr05	108102395	108102406	NC_000005.10	12	7
chr05	110231750	110231761	NC_000005.10	12	7
chr05	113996496	113996485	NC_000005.10	12	7
chr05	117233050	117233039	NC_000005.10	12	7
chr05	121622921	121622932	NC_000005.10	12	7
chr05	122520704	122520693	NC_000005.10	12	7
chr05	142330490	142330479	NC_000005.10	12	7
chr05	156359105	156359094	NC_000005.10	12	7
chr06	19187842	19187831	NC_000006.12	12	7

chr06	41103469	41103458	NC_000006.12	12	7
chr06	49856872	49856883	NC_000006.12	12	7
chr06	54896309	54896298	NC_000006.12	12	7
chr06	64416107	64416096	NC_000006.12	12	7
chr06	104438997	104439008	NC_000006.12	12	7
chr06	109349688	109349677	NC_000006.12	12	7
chr06	110631149	110631160	NC_000006.12	12	7
chr06	114751383	114751372	NC_000006.12	12	7
chr06	116514339	116514350	NC_000006.12	12	7
chr06	121606126	121606137	NC_000006.12	12	7
chr06	126139788	126139777	NC_000006.12	12	7
chr06	130852449	130852460	NC_000006.12	12	7
chr06	136843057	136843068	NC_000006.12	12	7
chr06	155559692	155559681	NC_000006.12	12	7
chr06	158099232	158099221	NC_000006.12	12	7
chr07	24339208	24339219	NC_000007.14	12	7
chr07	39519736	39519725	NC_000007.14	12	7
chr07	72228733	72228722	NC_000007.14	12	7
chr07	82955239	82955228	NC_000007.14	12	7
chr07	104990180	104990191	NC_000007.14	12	7
chr07	107698186	107698197	NC_000007.14	12	7
chr07	111002449	111002438	NC_000007.14	12	7
chr07	115191852	115191841	NC_000007.14	12	7
chr07	115572755	115572766	NC_000007.14	12	7
chr08	21126162	21126151	NC_000008.11	12	7
chr08	25928055	25928066	NC_000008.11	12	7
chr08	45036535	45036524	NC_000008.11	12	7
chr08	45248484	45248473	NC_000008.11	12	7
chr08	45502096	45502085	NC_000008.11	12	7
chr08	45763984	45763973	NC_000008.11	12	7
chr08	53335054	53335043	NC_000008.11	12	7
chr08	55581238	55581227	NC_000008.11	12	7
chr08	63169546	63169557	NC_000008.11	12	7
chr08	66553887	66553898	NC_000008.11	12	7
chr08	86283378	86283367	NC_000008.11	12	7
chr08	113060704	113060715	NC_000008.11	12	7
chr08	114537195	114537206	NC_000008.11	12	7
chr08	114886082	114886071	NC_000008.11	12	7
chr08	127206415	127206426	NC_000008.11	12	7

chr08	133590421	133590410	NC_000008.11	12	7
chr08	135425161	135425150	NC_000008.11	12	7
chr10	7034863	7034852	NC_000010.11	12	7
chr10	124304797	124304786	NC_000010.11	12	7
chr10	131502495	131502506	NC_000010.11	12	7
chr11	5557203	5557192	NC_000011.10	12	7
chr11	31242576	31242565	NC_000011.10	12	7
chr11	31419537	31419526	NC_000011.10	12	7
chr11	33169689	33169678	NC_000011.10	12	7
chr11	55948947	55948958	NC_000011.10	12	7
chr11	85654901	85654890	NC_000011.10	12	7
chr11	92588724	92588735	NC_000011.10	12	7
chr11	105227927	105227938	NC_000011.10	12	7
chr11	106302916	106302927	NC_000011.10	12	7
chr11	110594096	110594085	NC_000011.10	12	7
chr11	125228337	125228326	NC_000011.10	12	7
chr12	8769205	8769194	NC_000012.12	12	7
chr12	30438984	30438973	NC_000012.12	12	7
chr12	33556205	33556216	NC_000012.12	12	7
chr12	39419629	39419640	NC_000012.12	12	7
chr12	88509246	88509257	NC_000012.12	12	7
chr12	92878719	92878708	NC_000012.12	12	7
chr12	107034133	107034144	NC_000012.12	12	7
chr12	109455187	109455176	NC_000012.12	12	7
chr13	88827671	88827660	NC_000013.11	12	7
chr14	48126959	48126948	NC_000014.9	12	7
chr20	851155	851166	NC_000020.11	12	7
chr22	17367227	17367238	NC_000022.11	12	7
chrX	20420571	20420582	NC_000023.11	12	7
chrX	22545830	22545819	NC_000023.11	12	7
chrX	28899719	28899708	NC_000023.11	12	7
chrX	33013952	33013963	NC_000023.11	12	7
chrX	103880419	103880408	NC_000023.11	12	7
chrX	105612448	105612459	NC_000023.11	12	7
chrX	107929443	107929454	NC_000023.11	12	7
chrX	112571297	112571286	NC_000023.11	12	7
chrX	127584278	127584267	NC_000023.11	12	7
chrX	130836800	130836811	NC_000023.11	12	7
chrX	140554200	140554211	NC_000023.11	12	7

chrX	146697583	146697572	NC_000023.11	12	7
chr01	97461481	97461469	NC_000001.11	13	6
chr01	104600535	104600547	NC_000001.11	13	6
chr02	12589473	12589461	NC_000002.12	13	6
chr02	187173643	187173631	NC_000002.12	13	6
chr03	29181713	29181701	NC_000003.12	13	6
chr04	32171549	32171537	NC_000004.12	13	6
chr04	116646904	116646916	NC_000004.12	13	6
chr04	164205821	164205833	NC_000004.12	13	6
chr04	170244792	170244780	NC_000004.12	13	6
chr05	15818439	15818451	NC_000005.10	13	6
chr05	174059501	174059489	NC_000005.10	13	6
chr06	94936128	94936140	NC_000006.12	13	6
chr06	98142018	98142006	NC_000006.12	13	6
chr06	151814731	151814719	NC_000006.12	13	6
chr07	14490189	14490201	NC_000007.14	13	6
chr07	53075165	53075153	NC_000007.14	13	6
chr07	87815318	87815306	NC_000007.14	13	6
chr07	103485572	103485584	NC_000007.14	13	6
chr08	35951572	35951560	NC_000008.11	13	6
chr08	39327231	39327219	NC_000008.11	13	6
chr08	69690270	69690258	NC_000008.11	13	6
chr08	117816166	117816154	NC_000008.11	13	6
chr08	123134654	123134666	NC_000008.11	13	6
chr09	68817454	68817466	NC_000009.12	13	6
chr09	68894040	68894028	NC_000009.12	13	6
chr09	80470190	80470178	NC_000009.12	13	6
chr10	1642234	1642222	NC_000010.11	13	6
chr10	73077072	73077060	NC_000010.11	13	6
chr10	110134589	110134601	NC_000010.11	13	6
chr11	9150979	9150991	NC_000011.10	13	6
chr11	9153635	9153623	NC_000011.10	13	6
chr11	13413693	13413705	NC_000011.10	13	6
chr11	41773900	41773912	NC_000011.10	13	6
chr11	77886545	77886557	NC_000011.10	13	6
chr11	79988166	79988154	NC_000011.10	13	6
chr11	108008162	108008150	NC_000011.10	13	6
chr12	30847723	30847735	NC_000012.12	13	6
chr12	86693014	86693026	NC_000012.12	13	6

chr12	122128926	122128914	NC_000012.12	13	6
chr13	29622714	29622726	NC_000013.11	13	6
chr14	39336522	39336510	NC_000014.9	13	6
chr15	94819443	94819431	NC_000015.10	13	6
chr17	10951262	10951250	NC_000017.11	13	6
chr19	30854506	30854518	NC_000019.10	13	6
chr20	42688485	42688473	NC_000020.11	13	6
chrX	38138789	38138801	NC_000023.11	13	6
chrX	86361231	86361243	NC_000023.11	13	6
chrX	107051786	107051798	NC_000023.11	13	6
chrX	109054235	109054247	NC_000023.11	13	6
chr01	32830533	32830522	NC_000001.11	12	6
chr01	56714138	56714127	NC_000001.11	12	6
chr01	79950536	79950547	NC_000001.11	12	6
chr01	81600862	81600851	NC_000001.11	12	6
chr01	88351333	88351344	NC_000001.11	12	6
chr01	100720346	100720335	NC_000001.11	12	6
chr01	103153587	103153598	NC_000001.11	12	6
chr01	163679268	163679279	NC_000001.11	12	6
chr01	178138239	178138228	NC_000001.11	12	6
chr01	202443386	202443375	NC_000001.11	12	6
chr01	214381798	214381787	NC_000001.11	12	6
chr01	239483920	239483909	NC_000001.11	12	6
chr02	5995932	5995921	NC_000002.12	12	6
chr02	14869774	14869785	NC_000002.12	12	6
chr02	37466261	37466250	NC_000002.12	12	6
chr02	38845623	38845634	NC_000002.12	12	6
chr02	38849877	38849866	NC_000002.12	12	6
chr02	52660534	52660523	NC_000002.12	12	6
chr02	55372861	55372872	NC_000002.12	12	6
chr02	62005199	62005210	NC_000002.12	12	6
chr02	70287567	70287556	NC_000002.12	12	6
chr02	79359701	79359712	NC_000002.12	12	6
chr02	84655638	84655627	NC_000002.12	12	6
chr02	126324776	126324787	NC_000002.12	12	6
chr02	149537132	149537143	NC_000002.12	12	6
chr02	169529510	169529521	NC_000002.12	12	6
chr02	175817135	175817124	NC_000002.12	12	6
chr02	180079693	180079682	NC_000002.12	12	6

chr02	206324011	206324000	NC_000002.12	12	6
chr02	206814054	206814043	NC_000002.12	12	6
chr02	224807794	224807805	NC_000002.12	12	6
chr02	229238864	229238853	NC_000002.12	12	6
chr02	236280053	236280064	NC_000002.12	12	6
chr03	154343	154354	NC_000003.12	12	6
chr03	8511973	8511984	NC_000003.12	12	6
chr03	16880365	16880376	NC_000003.12	12	6
chr03	18087857	18087846	NC_000003.12	12	6
chr03	47168148	47168137	NC_000003.12	12	6
chr03	47937628	47937617	NC_000003.12	12	6
chr03	48992978	48992989	NC_000003.12	12	6
chr03	82163078	82163067	NC_000003.12	12	6
chr03	103449909	103449898	NC_000003.12	12	6
chr03	120049593	120049604	NC_000003.12	12	6
chr03	143783076	143783065	NC_000003.12	12	6
chr03	149601763	149601752	NC_000003.12	12	6
chr03	167891194	167891183	NC_000003.12	12	6
chr03	181054638	181054627	NC_000003.12	12	6
chr03	191545181	191545170	NC_000003.12	12	6
chr03	197899144	197899155	NC_000003.12	12	6
chr04	668375	668364	NC_000004.12	12	6
chr04	19382020	19382031	NC_000004.12	12	6
chr04	19484541	19484552	NC_000004.12	12	6
chr04	26997338	26997349	NC_000004.12	12	6
chr04	55658608	55658619	NC_000004.12	12	6
chr04	70437852	70437841	NC_000004.12	12	6
chr04	79981798	79981809	NC_000004.12	12	6
chr04	94968197	94968208	NC_000004.12	12	6
chr04	102674459	102674470	NC_000004.12	12	6
chr04	124485434	124485445	NC_000004.12	12	6
chr04	126123159	126123148	NC_000004.12	12	6
chr04	137124764	137124753	NC_000004.12	12	6
chr04	160702860	160702849	NC_000004.12	12	6
chr04	167052375	167052386	NC_000004.12	12	6
chr04	179139043	179139032	NC_000004.12	12	6
chr04	179161408	179161397	NC_000004.12	12	6
chr04	187143772	187143761	NC_000004.12	12	6
chr05	10200709	10200720	NC_000005.10	12	6

chr05	33225853	33225842	NC_000005.10	12	6
chr05	76255175	76255186	NC_000005.10	12	6
chr05	82855245	82855256	NC_000005.10	12	6
chr05	84139572	84139561	NC_000005.10	12	6
chr05	88198462	88198473	NC_000005.10	12	6
chr05	102501084	102501073	NC_000005.10	12	6
chr05	109583817	109583806	NC_000005.10	12	6
chr05	128180682	128180671	NC_000005.10	12	6
chr05	136190403	136190392	NC_000005.10	12	6
chr05	154189555	154189566	NC_000005.10	12	6
chr05	171957271	171957282	NC_000005.10	12	6
chr05	175317578	175317567	NC_000005.10	12	6
chr06	4853151	4853162	NC_000006.12	12	6
chr06	16133021	16133032	NC_000006.12	12	6
chr06	26103447	26103436	NC_000006.12	12	6
chr06	35947570	35947581	NC_000006.12	12	6
chr06	68279419	68279430	NC_000006.12	12	6
chr06	79806546	79806557	NC_000006.12	12	6
chr06	85260407	85260418	NC_000006.12	12	6
chr06	136633874	136633885	NC_000006.12	12	6
chr06	137931054	137931043	NC_000006.12	12	6
chr06	139739984	139739973	NC_000006.12	12	6
chr06	140341418	140341429	NC_000006.12	12	6
chr06	145869806	145869795	NC_000006.12	12	6
chr06	146731539	146731528	NC_000006.12	12	6
chr06	168728425	168728436	NC_000006.12	12	6
chr07	51771646	51771635	NC_000007.14	12	6
chr07	137082304	137082315	NC_000007.14	12	6
chr07	141052267	141052278	NC_000007.14	12	6
chr08	17556548	17556537	NC_000008.11	12	6
chr08	30097319	30097308	NC_000008.11	12	6
chr08	68502659	68502670	NC_000008.11	12	6
chr08	86697209	86697198	NC_000008.11	12	6
chr08	91622182	91622171	NC_000008.11	12	6
chr08	92498179	92498168	NC_000008.11	12	6
chr08	124481608	124481597	NC_000008.11	12	6
chr08	129563081	129563092	NC_000008.11	12	6
chr08	131305462	131305451	NC_000008.11	12	6
chr09	14627274	14627285	NC_000009.12	12	6

chr09	15151836	15151847	NC_000009.12	12	6
chr09	22322306	22322295	NC_000009.12	12	6
chr09	23783142	23783153	NC_000009.12	12	6
chr09	26318093	26318104	NC_000009.12	12	6
chr09	31054959	31054970	NC_000009.12	12	6
chr09	79007585	79007596	NC_000009.12	12	6
chr09	88239264	88239253	NC_000009.12	12	6
chr09	96543680	96543669	NC_000009.12	12	6
chr09	99112802	99112813	NC_000009.12	12	6
chr09	123836553	123836564	NC_000009.12	12	6
chr10	33633573	33633562	NC_000010.11	12	6
chr10	65551995	65551984	NC_000010.11	12	6
chr10	66717930	66717941	NC_000010.11	12	6
chr10	74291798	74291787	NC_000010.11	12	6
chr10	82621770	82621781	NC_000010.11	12	6
chr10	91090519	91090530	NC_000010.11	12	6
chr10	99682921	99682910	NC_000010.11	12	6
chr10	107653284	107653273	NC_000010.11	12	6
chr10	127387876	127387887	NC_000010.11	12	6
chr11	10330421	10330410	NC_000011.10	12	6
chr11	21052051	21052062	NC_000011.10	12	6
chr11	56948810	56948799	NC_000011.10	12	6
chr11	91992913	91992902	NC_000011.10	12	6
chr11	96712150	96712139	NC_000011.10	12	6
chr11	99478699	99478710	NC_000011.10	12	6
chr11	103284503	103284514	NC_000011.10	12	6
chr11	110624774	110624763	NC_000011.10	12	6
chr11	118226686	118226697	NC_000011.10	12	6
chr11	121927186	121927175	NC_000011.10	12	6
chr11	127371998	127372009	NC_000011.10	12	6
chr12	21742376	21742387	NC_000012.12	12	6
chr12	33375091	33375102	NC_000012.12	12	6
chr12	79305333	79305322	NC_000012.12	12	6
chr12	87018030	87018041	NC_000012.12	12	6
chr12	97027085	97027074	NC_000012.12	12	6
chr12	97030674	97030685	NC_000012.12	12	6
chr12	97794786	97794775	NC_000012.12	12	6
chr12	99326334	99326345	NC_000012.12	12	6
chr12	100617295	100617284	NC_000012.12	12	6

chr12	106997614	106997603	NC_000012.12	12	6
chr12	114419769	114419758	NC_000012.12	12	6
chr13	29428703	29428714	NC_000013.11	12	6
chr13	34838980	34838991	NC_000013.11	12	6
chr13	68672648	68672637	NC_000013.11	12	6
chr13	68677576	68677565	NC_000013.11	12	6
chr13	79534292	79534303	NC_000013.11	12	6
chr13	83374368	83374357	NC_000013.11	12	6
chr13	91208120	91208131	NC_000013.11	12	6
chr13	92057240	92057251	NC_000013.11	12	6
chr13	105912154	105912165	NC_000013.11	12	6
chr14	37970959	37970948	NC_000014.9	12	6
chr14	40492006	40491995	NC_000014.9	12	6
chr14	44782915	44782926	NC_000014.9	12	6
chr14	48758306	48758317	NC_000014.9	12	6
chr14	88004548	88004537	NC_000014.9	12	6
chr15	56610753	56610764	NC_000015.10	12	6
chr15	70757589	70757578	NC_000015.10	12	6
chr15	96964230	96964219	NC_000015.10	12	6
chr16	66442829	66442818	NC_000016.10	12	6
chr16	74623964	74623975	NC_000016.10	12	6
chr16	75189302	75189291	NC_000016.10	12	6
chr17	9332911	9332900	NC_000017.11	12	6
chr18	32474384	32474373	NC_000018.10	12	6
chr18	34128952	34128963	NC_000018.10	12	6
chr18	55039826	55039815	NC_000018.10	12	6
chr18	78931519	78931508	NC_000018.10	12	6
chr19	31065225	31065236	NC_000019.10	12	6
chr19	32434028	32434017	NC_000019.10	12	6
chr19	51221292	51221303	NC_000019.10	12	6
chr20	1361969	1361958	NC_000020.11	12	6
chr20	4448895	4448906	NC_000020.11	12	6
chr20	13696489	13696478	NC_000020.11	12	6
chr20	20275384	20275395	NC_000020.11	12	6
chr20	26367536	26367525	NC_000020.11	12	6
chr21	37223237	37223248	NC_000021.9	12	6
chr21	46496495	46496484	NC_000021.9	12	6
chr22	39560335	39560346	NC_000022.11	12	6
chrX	986645	986656	NC_000023.11	12	6

chrX	5921242	5921253	NC_000023.11	12	6
chrX	6765829	6765840	NC_000023.11	12	6
chrX	15504137	15504126	NC_000023.11	12	6
chrX	22546280	22546269	NC_000023.11	12	6
chrX	41199361	41199372	NC_000023.11	12	6
chrX	43885293	43885282	NC_000023.11	12	6
chrX	67874307	67874296	NC_000023.11	12	6
chrX	110216026	110216037	NC_000023.11	12	6
chrX	110566890	110566879	NC_000023.11	12	6
chrX	111357390	111357379	NC_000023.11	12	6
chrX	150589443	150589454	NC_000023.11	12	6
chr01	23207589	23207577	NC_000001.11	13	5
chr01	25897408	25897420	NC_000001.11	13	5
chr01	65491478	65491490	NC_000001.11	13	5
chr01	154831168	154831180	NC_000001.11	13	5
chr02	35254361	35254349	NC_000002.12	13	5
chr02	207969171	207969159	NC_000002.12	13	5
chr03	185371630	185371642	NC_000003.12	13	5
chr04	46469891	46469879	NC_000004.12	13	5
chr04	105058847	105058835	NC_000004.12	13	5
chr04	124730032	124730044	NC_000004.12	13	5
chr04	158619352	158619364	NC_000004.12	13	5
chr06	85949972	85949960	NC_000006.12	13	5
chr06	109604972	109604960	NC_000006.12	13	5
chr10	59089285	59089273	NC_000010.11	13	5
chr10	99263586	99263598	NC_000010.11	13	5
chr11	96315922	96315934	NC_000011.10	13	5
chr15	33186727	33186715	NC_000015.10	13	5
chr15	87091718	87091706	NC_000015.10	13	5
chr16	16972153	16972165	NC_000016.10	13	5
chr16	59986446	59986458	NC_000016.10	13	5
chr18	12587445	12587457	NC_000018.10	13	5
chr18	78691060	78691048	NC_000018.10	13	5
chr19	39627504	39627492	NC_000019.10	13	5
chr19	54674561	54674573	NC_000019.10	13	5
chr20	30512867	30512855	NC_000020.11	13	5
chr20	45173430	45173442	NC_000020.11	13	5
chr21	35062647	35062659	NC_000021.9	13	5
chrX	77412877	77412889	NC_000023.11	13	5

chrX	130349739	130349727	NC_000023.11	13	5
chr01	8663054	8663065	NC_000001.11	12	5
chr01	26335998	26336009	NC_000001.11	12	5
chr01	42582606	42582595	NC_000001.11	12	5
chr01	47032830	47032819	NC_000001.11	12	5
chr01	69196253	69196264	NC_000001.11	12	5
chr01	70300023	70300034	NC_000001.11	12	5
chr01	82771042	82771053	NC_000001.11	12	5
chr01	100102957	100102946	NC_000001.11	12	5
chr01	107996202	107996213	NC_000001.11	12	5
chr01	162211653	162211642	NC_000001.11	12	5
chr01	208646365	208646354	NC_000001.11	12	5
chr01	215734460	215734449	NC_000001.11	12	5
chr01	234143991	234144002	NC_000001.11	12	5
chr01	241045297	241045286	NC_000001.11	12	5
chr02	140780861	140780872	NC_000002.12	12	5
chr02	149162575	149162586	NC_000002.12	12	5
chr02	162692841	162692852	NC_000002.12	12	5
chr02	222738270	222738259	NC_000002.12	12	5
chr03	67248099	67248110	NC_000003.12	12	5
chr03	174292637	174292648	NC_000003.12	12	5
chr04	12331297	12331308	NC_000004.12	12	5
chr04	21504937	21504948	NC_000004.12	12	5
chr04	43962965	43962976	NC_000004.12	12	5
chr04	57433948	57433937	NC_000004.12	12	5
chr04	85682861	85682872	NC_000004.12	12	5
chr04	106114290	106114301	NC_000004.12	12	5
chr04	113028283	113028294	NC_000004.12	12	5
chr04	151151805	151151794	NC_000004.12	12	5
chr04	152051162	152051173	NC_000004.12	12	5
chr04	179052931	179052920	NC_000004.12	12	5
chr05	6661409	6661420	NC_000005.10	12	5
chr05	93549147	93549158	NC_000005.10	12	5
chr05	148916732	148916721	NC_000005.10	12	5
chr05	153193520	153193531	NC_000005.10	12	5
chr05	169165696	169165685	NC_000005.10	12	5
chr06	99056822	99056833	NC_000006.12	12	5
chr07	21203640	21203651	NC_000007.14	12	5
chr07	27364344	27364355	NC_000007.14	12	5

chr07	45331667	45331656	NC_000007.14	12	5
chr08	28102047	28102036	NC_000008.11	12	5
chr08	64148089	64148078	NC_000008.11	12	5
chr08	121058238	121058249	NC_000008.11	12	5
chr08	134902692	134902681	NC_000008.11	12	5
chr09	26814924	26814935	NC_000009.12	12	5
chr09	35739632	35739643	NC_000009.12	12	5
chr09	77017601	77017612	NC_000009.12	12	5
chr09	83041777	83041788	NC_000009.12	12	5
chr09	87072669	87072658	NC_000009.12	12	5
chr09	134613617	134613628	NC_000009.12	12	5
chr10	7938397	7938408	NC_000010.11	12	5
chr10	59688277	59688266	NC_000010.11	12	5
chr10	91834373	91834384	NC_000010.11	12	5
chr10	106036664	106036653	NC_000010.11	12	5
chr11	1648239	1648228	NC_000011.10	12	5
chr11	28286474	28286485	NC_000011.10	12	5
chr11	59609982	59609993	NC_000011.10	12	5
chr11	82154773	82154784	NC_000011.10	12	5
chr12	56884436	56884425	NC_000012.12	12	5
chr12	65309897	65309908	NC_000012.12	12	5
chr12	70312802	70312791	NC_000012.12	12	5
chr12	108169798	108169809	NC_000012.12	12	5
chr13	41643771	41643782	NC_000013.11	12	5
chr13	43730188	43730177	NC_000013.11	12	5
chr13	66772070	66772081	NC_000013.11	12	5
chr13	67266239	67266250	NC_000013.11	12	5
chr13	70438394	70438405	NC_000013.11	12	5
chr13	72462904	72462915	NC_000013.11	12	5
chr13	73589220	73589209	NC_000013.11	12	5
chr13	114256981	114256970	NC_000013.11	12	5
chr14	53548116	53548105	NC_000014.9	12	5
chr14	91128016	91128005	NC_000014.9	12	5
chr15	55623598	55623609	NC_000015.10	12	5
chr15	59650410	59650421	NC_000015.10	12	5
chr15	67895787	67895798	NC_000015.10	12	5
chr15	75030887	75030898	NC_000015.10	12	5
chr15	80376611	80376600	NC_000015.10	12	5
chr17	2259971	2259960	NC_000017.11	12	5

chr17	13599804	13599793	NC_000017.11	12	5
chr17	49970374	49970385	NC_000017.11	12	5
chr17	74411987	74411998	NC_000017.11	12	5
chr18	6692184	6692173	NC_000018.10	12	5
chr18	26936361	26936372	NC_000018.10	12	5
chr18	32164785	32164796	NC_000018.10	12	5
chr18	57372141	57372152	NC_000018.10	12	5
chr18	76028676	76028665	NC_000018.10	12	5
chr18	79860251	79860240	NC_000018.10	12	5
chr20	2767508	2767497	NC_000020.11	12	5
chr20	32334864	32334853	NC_000020.11	12	5
chr20	42969400	42969411	NC_000020.11	12	5
chr21	15405882	15405871	NC_000021.9	12	5
chr21	27128817	27128828	NC_000021.9	12	5
chr21	27724878	27724889	NC_000021.9	12	5
chr21	33775512	33775523	NC_000021.9	12	5
chr22	40201219	40201208	NC_000022.11	12	5
chrX	24583713	24583724	NC_000023.11	12	5
chrX	53003928	53003939	NC_000023.11	12	5
chrX	75537169	75537180	NC_000023.11	12	5
chrX	91187582	91187593	NC_000023.11	12	5
chr01	237603124	237603136	NC_000001.11	13	4
chr02	132279864	132279852	NC_000002.12	13	4
chr02	176672291	176672279	NC_000002.12	13	4
chr04	47096940	47096952	NC_000004.12	13	4
chr05	170123837	170123825	NC_000005.10	13	4
chr10	97944808	97944796	NC_000010.11	13	4
chr10	114226626	114226614	NC_000010.11	13	4
chr13	67884795	67884783	NC_000013.11	13	4
chr14	59591410	59591398	NC_000014.9	13	4
chr16	3659076	3659088	NC_000016.10	13	4
chr18	25418784	25418772	NC_000018.10	13	4
chrX	45634061	45634049	NC_000023.11	13	4
chr01	3217976	3217987	NC_000001.11	12	4
chr01	92837827	92837816	NC_000001.11	12	4
chr01	112701651	112701662	NC_000001.11	12	4
chr01	166000671	166000660	NC_000001.11	12	4
chr01	178801277	178801288	NC_000001.11	12	4
chr02	177290177	177290166	NC_000002.12	12	4

chr02	218084695	218084706	NC_000002.12	12	4
chr02	236494650	236494639	NC_000002.12	12	4
chr04	42894460	42894471	NC_000004.12	12	4
chr04	66200304	66200315	NC_000004.12	12	4
chr06	35644009	35643998	NC_000006.12	12	4
chr06	35671520	35671531	NC_000006.12	12	4
chr09	95179956	95179945	NC_000009.12	12	4
chr09	122078420	122078431	NC_000009.12	12	4
chr09	132891241	132891252	NC_000009.12	12	4
chr09	134244101	134244112	NC_000009.12	12	4
chr10	46934395	46934384	NC_000010.11	12	4
chr10	48117437	48117448	NC_000010.11	12	4
chr10	102716315	102716304	NC_000010.11	12	4
chr12	31614069	31614080	NC_000012.12	12	4
chr13	18693215	18693204	NC_000013.11	12	4
chr14	30845671	30845660	NC_000014.9	12	4
chr14	94062711	94062722	NC_000014.9	12	4
chr17	10363532	10363543	NC_000017.11	12	4
chr17	59667014	59667025	NC_000017.11	12	4
chr17	68278027	68278038	NC_000017.11	12	4
chr18	44686796	44686785	NC_000018.10	12	4
chr18	55570049	55570060	NC_000018.10	12	4
chr20	37099530	37099519	NC_000020.11	12	4
chr21	14473970	14473981	NC_000021.9	12	4
chr21	28191101	28191112	NC_000021.9	12	4
chr01	85362838	85362827	NC_000001.11	12	3
chr14	106817445	106817434	NC_000014.9	12	3
chr17	12074729	12074718	NC_000017.11	12	3
chr21	8217645	8217657	NC_000021.9	13	2
chr21	8400683	8400695	NC_000021.9	13	2
chr21	8444915	8444927	NC_000021.9	13	2
chr01	65227883	65227872	NC_000001.11	12	2
chr17	31347890	31347879	NC_000017.11	12	1

Example 9: Retargeting of a Gene Writer to a genomic safe harbor site

This example describes a Gene Writer comprising a heterologous DNA binding domain that redirects its activity to a genomic safe harbor site.

In this example, the Gene Writer polypeptide sequence is altered to where its natural DNA binding domain is replaced, mutated/inactivated, and/or joined with another polypeptide sequence that can redirect the Gene Writer system to another genomic location that is not its endogenous or natural binding site. In some instances, the polypeptide sequence that redirects the Gene Writer system to a non-natural genomic location may also be attached and/or inserted to any module of the Gene Writer polypeptide sequence.

In some embodiments, the polypeptide sequence used to redirect the Gene Writer system to a non-natural genomic target encodes for: a zinc finger, a series of adjacent, regularly, or irregularly spaced zinc fingers, a transcription activator-like effector (TALE), a series of adjacent, regularly, or irregularly spaced a transcription activator-like effectors (TALEs), Cas9, Cas9 with mutations to its catalytic residues inactivating the double-stranded DNA endonuclease activity (referred to as catalytically-dead Cas9 (dCas9)), Cas9 with a point mutation or multiple point mutations in a single catalytic domain in order to render Cas9 endonuclease only able to cleave one strand of double-stranded DNA (referred to as Cas9 nickase) (see Fig. 12)..

In some embodiments, the polypeptide sequence used to re-direct the Gene Writer system targets a genomic safe-harbor location (*e.g.*, AAVS1 site on human chromosome 19) (Pellenz, S., et. al., Human Gene Therapy, 30(7), 814–828, 2019), see Figures 11 and 13. In further embodiments, the polypeptide sequence used to re-direct the Gene Writer polypeptide sequence is used in conjunction with a nucleic acid that targets the genomic safe harbor location (*e.g.*, the polypeptide sequence for catalytic dead Cas9 along with a single-guide RNA targeting the AAVS1 site on chromosome 19).

Table E4: Re-targeted Gene Writer constructs. Examples shown are to re-target R2Tg Gene Writer polypeptide sequence to the AAVS1 site using ZF or Cas9 domains.

Gene Writer Polypeptide Name	Polypeptide Sequence (<i>Re-targeting polypeptide sequence, italic; <u>Linker, bold underline</u></i>)
AAVS1 Left ZFP attached at v2 location of DBD of R2Tg with 3GS linker	<p> <i>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGKPFACDICG RKFAQNSTRIGHTKIHRLGSGGGSTATRDKKDVTREKHPKPFQKWKDRAI KKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSF KSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTGKIPDMVRGCRVTLIPKSSKPDRLKDINNWRPITIGS ILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTI IWSAKREHRP LGVVFVDIAKAFDTVSHQHI IHALQQREVDPHIVGLVSNMYENISTYITTKRN</i> </p>

	<p>THTDKIQIRVGVKQGDPMSPLLFNLMADPLLCKLEESGKGYHRGQSSITAMAF ADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFYIKPTKDSYTIND CAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAP LKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCT CDAILYSSSTRDGGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQ LHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNW RKNEFKKWKTKLASQGRGIVNFERDKISNHWIQYYRRIPIHRKLLTALQLRANVY PTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSL EEEAAEKVRKYKHLETEVRHLTNAKDVT FVGFLGARGKWHQDNFKLLTELGL SKSRQVKMAET FSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1658)</p>
AAVS1 Left ZFP attached at v2 location of DBD of R2Tg with XTEN linker	<p>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNF SHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIHLRGSSGSETPGTSESATPESTATRDKKDVT TREKHPK KPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKI ILDDIECLSDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGIRAAGCSENKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLMADPLLCKLEESGKGY HRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFY IKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTK LDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSSSTRDGGGLGITKLAGLIPSVQARRLHRIAQSSDDT MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAP TQKDKFPKPCNWRKNEFKKWKTKLASQGRGIVNFERDKISNHWIQYYRRIPIHRK LLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDA RIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVD VTVRYEAAKSSLEEEAAEKVRKYKHLETEVRHLTNAKDVT FVGFLGARGKWH QDNFKLLTELGLSKSRQVKMAET FSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1659)</p>
AAVS1 Left ZFP attached at v1 location of DBD of R2Tg with 3GS linker	<p>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNF SHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIHLRGSGGGSCFGCLESISQIRTATRDKKDVT TREKHPK KPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKI ILDDIECLSDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGIRAAGCSENKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLMADPLLCKLEESGKGY HRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFY IKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTK LDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSSSTRDGGGLGITKLAGLIPSVQARRLHRIAQSSDDT MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAP TQKDKFPKPCNWRKNEFKKWKTKLASQGRGIVNFERDKISNHWIQYYRRIPIHRK</p>

	LLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDA RIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILIFVKDARALVVD VTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTTFVGFPLGARGKWH QDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1660)
AAVS1 Left ZFP attached at v1 location of DBD of R2Tg with XTEN linker	MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIHRLGSS SGSETPGTSESATPES CFGCLESISQIRTATRD KKDVTVTREKHPKKPFQKWMKDRAIKKGNLYRFQRLFYLDGRGLAKIILDDIEC LSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKN VQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTV LIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGFI AAGCSENKLLQTI IWSAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLMAMP LLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTG LKTQGGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDP WIGIARSGSLSTKLDLFWLQRIDQAPLKPLOKTDILKTYTIPRLIYIADHSEVKT ALLETLDQKIRTAVKEWLHLPCTCDAILYSTRDGGGLGITKLAGLIPSVQAR RLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNH WIIQYYRIRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCA HIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDIL IFVKDARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTTF VGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFAS RARKSMVM (SEQ ID NO: 1661)
AAVS1 Right ZFP attached at v2 location of DBD of R2Tg with 3GS linker	MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA RTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG RKFAQGYNLAGHTKIHRLGSS GGGS TATRDKKDVTVTREKHPKKPFQKWMKDRAI KKGNLYRFQRLFYLDGRGLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSF KSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMD PEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGS ILLRLFSRIVTARLSKACPLNPRQGRGFI AAGCSENKLLQTI IWSAKREHRP LGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNT HTDKIQIRVGVKQGDPMSPLLFNLMAMP LLCKLEESGKGYHRGQSSITAMAF ADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGGQKCHGFYIKPTKDSYTIND CAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGSLSTKLDLFWLQRIDQAP LKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEWLHLPCT CDAILYSTRDGGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQ LHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWR KNEFKKWKLASQGRGIVNFERDKISNHWIIQYYRIRIPHRKLLTALQLRANVY PTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELL LEEAKKKDWVVFKEPHIRDSNKELYKPDILIFVKDARALVVDVTVRYEAAKSSL EEAAAEKVRKYKHLETEVRHLTNAKDVTTFVGFPLGARGKWHQDNFKLLTELGL SKSRQVKMAETFTSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1662)
AAVS1 Right ZFP attached at v2	MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA RTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG

location of DBD of R2Tg with XTEN linker	<p><i>RKFAQGYNLAGHTKIHRLGSS</i>SGSETPGTSESATPESTATRDKKDVTVTREKHPK KPFQKWMKDRAIKKGNLYLRFQRLFYLDGRGKLAKIILDDIECLSCDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGIRAAGCSENKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLAMDPLLCKLEESGKGY HRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFY IKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTK LDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDT MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSEPPNNVSTNSEWEAP TQDKDFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYYRIRPHRK LLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPTQDA RIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILFVKDARALVVD VTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWH QDNFKLLTELGLSKSRQVKMAETFSVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1663)</p>
AAVS1 Right ZFP attached at v1 location of DBD of R2Tg with 3GS linker	<p><i>MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA RTDYLDVHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG RKFAQGYNLAGHTKIHRLGSS</i>GGGSCFGCLESISQIRTATRDKKDVTVTREKHPK KPFQKWMKDRAIKKGNLYLRFQRLFYLDGRGKLAKIILDDIECLSCDIPLSEIYS VFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGP DGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLK DINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGIRAAGCSENKLLQ TIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMY ENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLAMDPLLCKLEESGKGY HRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFY IKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTK LDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRT AVKEWLHLPPCTCDAILYSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDT MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSEPPNNVSTNSEWEAP TQDKDFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYYRIRPHRK LLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPTQDA RIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILFVKDARALVVD VTVRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWH QDNFKLLTELGLSKSRQVKMAETFSVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1664)</p>
AAVS1 Right ZFP attached at v1 location of DBD of R2Tg with XTEN linker	<p><i>MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA RTDYLDVHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG RKFAQGYNLAGHTKIHRLGSS</i>SGSETPGTSESATPESCFGCLESISQIRTATRD KKDVTVTREKHPKPKPFQKWMKDRAIKKGNLYLRFQRLFYLDGRGKLAKIILDDIEC LSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKN VQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTV LIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGIR AAGCSENKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREV DPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLAMDP</p>

	LLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTG LKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDP WIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKT ALLETLDQKIRTAVKEWLHLPPCTCDAILYSSSTRDGGGLGITKLAGLIPSVQAR RLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPP NNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNH WIIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCA HIIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEAAAEEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFAS RARKSMVM (SEQ ID NO: 1665)
AAVS1 Left and Right ZFP (separated by XTEN linker) attached at v2 location of DBD of R2Tg with 3GS linker	MGIHGVPAAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNF SHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIH LRGS SGSETPGTSESATPES GIHGVPAAMAERPFQCR ICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYLVDHTKIHTGSQKPF QCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHTKIH LR GS SGGS TATRDKKDVTREKHPKKPFQKWMKDRAIKKGNLRFQRLFYLD RGK LAKIILDDIECLSDIPLSEIYSVFKTRWETTG SFKSLGDFKTYGKADNTAFR ELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLT TGK IPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSRIVTARLSKAC PLNPRQRGFIRAAGCSENKLLQTI IWSAKREHRPLGVVFDIAKAFDTVSHQ HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVGVKQGDPM SPLLFN LAMDPL LCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNI SILETFCNLTG LKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGE SEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDILKTYTIPRL IYIADHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAILYSSSTRDGGGLGITK LAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPS IWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWTKLASQGRGI VNFERDKISNHWIIQYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKAC RHCDADIESCAHIIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIR DSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEAAAEEKVRKYKHLETEV RHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFTSTVALF SSVDIVHMFASRARKSMVM (SEQ ID NO: 1666)
AAVS1 Left and Right ZFP (separated by XTEN linker) attached at v2 location of DBD of R2Tg with XTEN linker	MGIHGVPAAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNF SHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIH LRGS SGSETPGTSESATPES GIHGVPAAMAERPFQCR ICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYLVDHTKIHTGSQKPF QCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHTKIH LR GS SGSETPGTSESATPES TATRDKKDVTREKHPKKPFQKWMKDRAIKKGNL RFQRLFYLD RGKLAKIILDDIECLSDIPLSEIYSVFKTRWETTG SFKSLGDF KTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRT MEIFNLWLT TGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLF SRIVTARLSKACPLNPRQRGFIRAAGCSENKLLQTI IWSAKREHRPLGVVFD IAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKI QIRVGVKQGDPMSPLLFN LAMDPL LCKLEESGKGYHRGQSSITAMAFADDLV LSDSWENMNTNISILETFCNLTG LKTQGQKCHGFYIKPTKDSYTINDCAAWTI NGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLDFWLQRIDQAPLKPLQK

	TDILKTYTIPRLIYIADHSEVKTALLETLTDQKIRTAVKEWLHLPPCTCDAILY SSTRDGGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLW IQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFK KWTKLASQGRGIVNFERDKISNHWIQYYRRIPIHRKLLTALQLRANVYPTREFL ARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKK KDWVVFKEPHIRDSNKELYKPDILFVKDARALVVDVTVRYEAAKSSLEEAAAE KVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELGLSKSRQV KMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1667)
AAVS1 Left ZFP attached to N- terminus of R2Tg with XTEN linker	MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIHRLGSS SGSETPGTSESATPES ASCPKPGPPVSAGAMSL ESGLTTHSVLAIERGPNLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVD LVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLA NDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQ LAPQPCVGETLHS SLPRDSELFVPEEGSSEKESEDAPKTSPTPGKHGLEQTGEEKVMVTPDKNP PCPCCGTRVNSVLNLIHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFPKCR GPETEKAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS NRGAHKRCWTKEEEELLIRLEAQFEGKNKINKLIAEHITTKAKQISDKRLL SRKPAEEPREEPGTCHHTRAAASLRTEPEMSHHAQAEDRDNGPGRPLPGRA AAGGRMTDEIRRHPDKNGQQRP TKQKSEEQLQAYYKKTLEERLSAGALNTFP RAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDVTREKHPKKP FQKWMKDRAIKKGNLRFQRLFYLDGRGLAKIILDDIECLSCDIPLSEIYSVF KTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDG ITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGIRAAGCSENKLLQTI IWSAKREHRPLGVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYEN ISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNAMDP LLCKLEESGKGYHR GQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFYIK PTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKLQKTDILKTYTIPRLIYIADHSEVKTALLETLTDQKIRTAV KEWLHLPPCTCDAILYSSTRDGGGLGITKLAGLIPSVQARRLHRIAQSSDDTMK CFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPIHRKLL TALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARI KRHNYICELLLEEAKKDWVVFKEPHIRDSNKELYKPDILFVKDARALVVDVT VRYEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1668)
AAVS1 Right ZFP attached to N- terminus of R2Tg with XTEN linker	MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA RTDYLDVHTKIHTGSQKPFQCRICMRNFSYNTHLTHRIRTHTGEKPFACDICG RKFAQGYNLAGHTKIHRLGSS SGSETPGTSESATPES ASCPKPGPPVSAGAMSL ESGLTTHSVLAIERGPNLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVD LVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLA NDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQ LAPQPCVGETLHS SLPRDSELFVPEEGSSEKESEDAPKTSPTPGKHGLEQTGEEKVMVTPDKNP PCPCCGTRVNSVLNLIHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFPKCR GPETEKAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS

	<p>NRGAHKRCWTKEEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAAAGGRTMDEIRRHPDKNGQQRP TKQKSEEQLQAYYKKTLEERLSAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDVTREKHPKPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGFI RAAGCSENKLLQTI IWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVG VKQGDPMSPLLFNLAMDP LLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKP LQKTDILKTYTIPRLIYIADHSEVKTALLETL DQKIRTAVKEWLHLP PCTCDAILYSS TRDGG LGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYRRIPHRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVT VRYEAAKSSLEEA AAEKVRYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1669)</p>
AAVS1 Left and Right ZFP attached to N-terminus of R2Tg with XTEN linker	<p>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNF SHNYARDCHIRTHTGEKPFACDICGRKFAQNSTRIGHTKIH LRGS SGSETPGTSESATPES GIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYLVDHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHTKIH LRGS SGSETPGTSESATPES ASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNSLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLP RDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQTGEEKVMVTPDKNPPCPCGTRVNSVLN LIEHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFPKCRGPETEKAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKRCWTKEEEEELLIRLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNGPGRRPLPGRAAAGGRTMDEIRRHPDKNGQQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDVTREKHPKPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGFI RAAGCSENKLLQTI IWSAKREHRPLGVVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQIRVG VKQGDPMSPLLFNLAMDP LLCKLEESGKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKP LQKTDILKTYTIPRLIYIADHSEVKTALLETL DQKIRTAVKEWLHLP PCTCDAILYSS TRDGG LGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQ</p>

	AGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKW TKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLLTALQLRANVYPTREFLAR GRQDQYIKACRHCDADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKD WVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEAAA EKV RKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKM AETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1670)
AAVS1 Left ZFP attached to N- terminus of R2Tg containing DBD inactivation mutations with XTEN linker	MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNFSHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIH LRGS SGSETPGTSESATPES ASCPKPGPPVSAGAMSL ESGLTTHSVLAIERGPNLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVD LVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAE NDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQ LAPQPCVGETLHS SLPRDSELFVPEEGSSEKESEDAPKTSPTPGKHGLEQTGEEKVMVTPDKNP PSPSSGTRVNSVLN LIEHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFPKCR GPETEKAPAGEWISEVSNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS NRGAHKACATKEEEEELLIRLEAQFEGNKNINKLIAEHITTKAKQISDKRRL SRKPAEEPREEPGTCHHTRRAAASLRTEPEM SHHAQAEDRDNGPGRRLPGRA AAGGRMTDEIRRHDPKGNQQRPTKQKSEEQLQAYYKKTLEERLSAGALNTFP RAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDVTREKHPKKP FQKWMKDRAIKKGNLRFQRLFYLD RGKLAKIILDDIECLSDIPLSEIYSVF KTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDG ITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRGIRAAGCSENKLLQTI IWSAKREHRPLGVFVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYEN ISTYITTKRNTHTDKIQIRVGKQGDPMSP LLLFN LAMDPLLCLEESGKGYHR GQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGGQKCHGFYIK PTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKP LQKTDILKTYTIPRLIYIADHSEVKTALLETLDQKIRTAV KEWLHLPPCTDAILYSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDTMK CFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLL TALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVTQDARI KRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVT VRYEAAKSSLEEAAA EKVRKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQD NFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1671)
AAVS1 Right ZFP attached to N- terminus of R2Tg containing DBD inactivation mutations with XTEN linker	MGIHGVPAAMAERPFQCRICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFA RTDYLDVHTKIHTGSQKPFQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICG RKFAQGYNLAGHTKIH LRGS SGSETPGTSESATPES ASCPKPGPPVSAGAMSL ESGLTTHSVLAIERGPNLANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVD LVSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAE NDLALLVNLEFYRSDLHVYECVHFAAHWEGLSGLPEVYEQ LAPQPCVGETLHS SLPRDSELFVPEEGSSEKESEDAPKTSPTPGKHGLEQTGEEKVMVTPDKNP PSPSSGTRVNSVLN LIEHLKVSHGKRGVCFRCAKCGKENSNYHSVCHFPKCR GPETEKAPAGEWISEVSNRDFTTKIGLGQHKRLAHPAVRNQERIVASQPKETS NRGAHKACATKEEEEELLIRLEAQFEGNKNINKLIAEHITTKAKQISDKRRL SRKPAEEPREEPGTCHHTRRAAASLRTEPEM SHHAQAEDRDNGPGRRLPGRA

	<p>AAGGRTMDEIRRHPDKNGGQORPTKQKSEEQLQAYYKKTLEERLSAGALNTFP RAFKQVMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKKDVTREKHPKKP FQKWMKDRAIKKGNLYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVF KTRWETTGSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDG ITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKDI NNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRFIRAAGCSENKLLQTI IWSAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYEN ISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLAMDP LLCKLEESGKGYHR GQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTLGLKTQGQKCHGFYIK PTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLQKIRTAV KEWLHLPCTCDAILYSTRDGGGLGITKLAGLIPSVQARRLHRIAQSSDDTMK CFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQ KDKFPKPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWIQYYRIPHRKLL TALQLRANVYPTREFLARGRQDQYIKACRHCADIESCAHIIGNCPVTQDARI KRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVT VRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVTFVGFPLGARGKWHQD NFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1672)</p>
AAVS1 Left and Right ZFP attached to N-terminus of R2Tg containing DBD inactivation mutations with XTEN linker	<p>MGIHGVPAAMAERPFQCRICMRNFSYNWHLQRHIRTHTGEKPFACDICGRKFA RSDHLTTHTKIHTGSQKPFQCRICMRNF SHNYARDCHIRTHTGEKPFACDICG RKFAQNSTRIGHTKIH LRGSSGSETPGTSESATPESGIHGVPAAMAERPFQCR ICMRNFSQSSNLARHIRTHTGEKPFACDICGRKFARTDYLVDHTKIHTGSQKP FQCRICMRNFSYNTHLTRHIRTHTGEKPFACDICGRKFAQGYNLAGHTKIH LR GSSGSETPGTSESATPESASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPN LANS GSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDLVSWSHPGPTSKSQQVDL VSLFPKHRVDLLSKNDQVDLVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHV YECVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLPDSELFVPEEGSSEK ESEDAPKTSPPTPGKHGLEQTGEEKVMVTPDKNPPSPSSGTRVNSVLN LIEH LKVSHGKRGVCFRCAKCGKENS NYHSVCHFPKCRGPETEKAPAGEWISEVSN RDTTKIGLGQHKRLAHPAVRNQERIVASQPKETSNRGAHKACATKEEEELL I RLEAQFEGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEEPREPGTCHHT RRAAASLRTEPEM SHHAQAEDRDNGPGRRLPGRAAAGGRTMDEIRRHPDKGN GQORPTKQKSEEQLQAYYKKTLEERLSAGALNTFPRAFQVMEGRDIKLVINQ TAQDCFGCLESISQIRTATRDKKDVTREKHPKKPFQKWMKDRAIKKGNLYLRF QRLFYLDRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETTGSFKSLGDFKT YGKADNTAFRELITAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFSRTME IFNLWLTGKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIGSILLRLFSR IVTARLSKACPLNPRQGRFIRAAGCSENKLLQTI IWSAKREHRPLGVVFDI AKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYENISTYITTKRNTHTDKIQI RVGVKQGDPMSPLLFNLAMDP LLCKLEESGKGYHRGQSSITAMAFADDLVLLS DSWENMNTNISILETFCNLTLGLKTQGQKCHGFYIKPTKDSYTINDCAAWTING TPLNMIDPGESEKYLGLQFDPWIGIARSGLSTKLD FWLQRIDQAPLKPLQKTD ILKTYTIPRLIYIADHSEVKTALLETLQKIRTAVKEWLHLPCTCDAILYSS TRDGGGLGITKLAGLIPSVQARRLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQ AGGDRENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKKW TKLASQGRGIVNFERDKISNHWIQYYRIPHRKLLTALQLRANVYPTREFLAR</p>

	GRQDQYIKACRHCADADIESCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARALVVDVTVRYEAAKSSLEEAAA EKV RKYKHLETEVRHLTNAKDVT FVGFP LGARGKWHQDNFKLLTELGLSKSRQVKM AETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1673)
<i>S. pyogenes</i> Cas9 attached at v2 location of DBD of R2Tg with XTEN33aa linker	MAPKKRKRKVG I HGVPAADKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE M AKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLF EENP INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRE RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVFREINNYHHAHDAYLNA VVG TALIKKYPKLESEFVYGDYKVYDV RKMIKSEQEI GKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKK TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE KGKSKKLKSVKELLGITIMERSSSF EKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASAGELQKGNELALPSKYVNFYLYASHYEKLGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQLGGD SGGSSGGSSGSETPGTSESATPESSGGSSGGSS TATRDKKDVTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKIILDDIECLSCDIPLSE IYSVFKTRWETTG SFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLT TGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENK LLQTI IWSAKREHRPLGVVFDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVG VKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSG STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPCTCDAILYSSSTRDGG LGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLW IQAGGDRENIPSIWEAPP SSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCADADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARAL VVDVTVRYEAAKSSLEEAAA EKV RKYKHLETEVRHLTNAKDVT FVGFP LGARG

	KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1674)
<i>S. pyogenes</i> Cas9 containing catalytic mutations (dCas9) attached at v2 location of DBD of R2Tg with XTEN33aa linker	MAPKKRKRKVGIIHGVPAAADKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE AKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSVDKFLIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADFLAANKLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPHKSL YEFYTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQDLSHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVFREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVVKMIKSEQEI GKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKK TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE KGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKDLIIKLPKYS FELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLKGSPEDEQKQ LFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDL SQLGGD SGGSSGGSSGSETPGTSESATPESGGSSGGSS TATRDKKDVTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE IYSVFKTRWETTGSKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRFIRAAGCSENK LLQTI IWSAKREHRPLGVVFDIAKAFDTVSHQHI IHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLMADPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSG STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPCTCDAILYSSSTRDGGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPP SSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPDILFVKDARAL VVDVTVRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARG KWHQDNFKLLTELGLSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1675)

<p><i>S. pyogenes</i> Cas9 D10A nickase mutant attached at v2 location of DBD of R2Tg with XTEN33aa linker</p>	<p>MAPKKKRKVGIIHGVPAAADKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE AKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSVDKLFQILVQTYNQLF EENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWRSLRKLINGIRDK QSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQDLSHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINR LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKK TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE KGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYS FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQ LFVEQHKHYLDEII EQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQLGGD SGGSSGGSSGSETPGTSESATPESSGGSSGGSS TATRDK KDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLD RGKLAKI ILDDIECLSCDIPLSE IYSVFKTRWETTG SFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLT TGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSENK LLQTI IWSAKREHRPLGVVFDIAKAFDTVSHQHI IHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVG VKQGDPMSPLLFNLAMDPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETF CNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSG STKLDFWLQRIDQAPLKP LQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSSTRDGG LGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLW IQAGGDRENIPSIWEAPP SSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYRRIP HRKLLTALQLRANVYPTREFLAGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARAL VVDVTVRYEAAKSSLEEA AAEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARG KWHQDNFKLLTELGLSKSRQVKMAET FSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1676)</p>
<p><i>S. pyogenes</i> Cas9 N863A nickase mutant attached at</p>	<p>MAPKKKRKVGIIHGVPAAADKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE AKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV</p>

v2 location of DBD of R2Tg with XTEN33aa linker	DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF EENP INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPPYYVGPLARGNSR FAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSHLL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGT YHDLKIIKDKDFLDNEENEDILEDIV LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSHEHIAN LAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR LSDYDVHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKK TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE KGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASAGELQKGNELALPSKYVNFYLYASHYEKLGSPEDNEQKQ LFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDL SQLGGD SGGSSGGSSGSETPGTSESATPESSGGSSGGSS TATRDKKDVTVREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDGRGLAKIILDDIECLSCDIPLSE IYSVFKTRWETTG SFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRFIRAAGCSENK LLQTI IWSAKREHRPLGVVVDIAKAFDTVSHQHI IHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNLMADPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSG STKLDFWLQRIDQAPLKPLOKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSSTRDGGGLITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPP SSEPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPD LIFVKDARAL VVDVTVRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVT FVGFP LGARG KWHQDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1677)
<i>S. pyogenes</i> Cas9 D10A nickase mutant attached at v2 location of DBD of R2Tg containing	MAPKKRKRKVGIVHGVPAADKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF EENP INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILL

EN mutation with XTEN33aa linker	SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSHLL YEFYTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTQKQKED YFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIAN LAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRE RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINR LSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVVKK TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE KGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASAGELQKGNELALPSKYVNFLLASHYEKLKGSPEDEQKQ LFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDL SQLGGD SGGSSGGSSGSETPGTSESATPESGGSSGGSS TATRDKDDTVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDGRGLAKIILDDIECLSCDIPLSE IYSVFKTRWETTGSEFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQGRFIRAAGCSENK LLQTI IWSAKREHRPLGVVVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLMADPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSG STKLDLFWLQRIDQAPLKPLOKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSSTRDGGGLITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIIQAGGDRENIPSIWEAPPSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPALIFVKDARAL VVDVTVRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVTFGFPLGARG KWHQDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1678)
<i>S. pyogenes</i> Cas9 N863A nickase mutant attached at v2 location of DBD of R2Tg containing EN mutation with XTEN33aa linker	MAPKKRKRKVGIIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE AKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGFLGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLLKDNREKIEKILTFRIPYYVGPLARGNSR

	<p>FAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDLSHEHIAN LAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR LSDYDVHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKK TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE KGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKDLIIKLPKYS FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEQNEQKQ LFVEQHKHYLDEIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDL SQLGGDSGGSSGGSSGSETPGTSESATPESGGSSGGSSTATRDKDDTVTREK HPKKPFQKWMKDRAIKKGNLRFQRLFYLDGRGLAKIILDDIECLSCDIPLSE IYSVFKTRWETTSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRFIRAAGCSENK LLQTI IWSAKREHRPLGVVVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLMADPLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSG STKLDLFWLQRIQAPLKPLOKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSSTRDGGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPALIFVKDARAL VVDVTVRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVTFGVPLGARG KWHQDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1679)</p>
<p><i>S. pyogenes</i> Cas9 attached at v2 location of DBD of R2Tg containing EN mutation with XTEN33aa linker</p>	<p>MAPKKRKRKVGIIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGFLGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV</p>

	<p> LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN LAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKK TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKE KGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKDLIIKLPKYSL FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDL SQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTATRDKKDVTREK HPKKPFQKWMKDRAIKKGNYLRFQRLFYLDGRGLAKIILDDIECLSCDIPLSE IYSVFKTRWETTSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEMSKGSA PGPDGITLGDVVKMDPEFSRTMEIFNLWLTGKIPDMVRGCRTVLIPKSSKPD RLKDINNWRPITIGSILLRLFSRIVTARLSKACPLNPRQRFIRAAGCSENK LLQTI IWSAKREHRPLGVVVDIAKAFDTVSHQHIIHALQQREVDPHIVGLVS NMYENISTYITTKRNTHTDKIQIRVGKQGDPMSPLLFNLAMDLLCKLEESG KGYHRGQSSITAMAFADDLVLLSDSWENMNTNISILETFCNLTGLKTQGQKCH GFYIKPTKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQFDPWIGIARSG STKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIADHSEVKTALLETLDQK IRTAVKEWLHLPPCTCDAILYSSSTRDGGGLGITKLAGLIPSVQARRLHRIAQSS DDTMKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPPSEPPNNVSTNSEW EAPTQKDKFPKPCNWRKNEFKKWKLASQGRGIVNFERDKISNHWIQYYRRIP HRKLLTALQLRANVYPTREFLARGRQDQYIKACRHCDADIESCAHIIGNCPVT QDARIKRHNYICELLLEEAKKKDWVVFKEPHIRDSNKELYKPALIFVKDARAL VVDVTVRYEAAKSSLEEAEEKVRKYKHLETEVRHLTNAKDVTFGVFLGARG KWHQDNFKLLTELGLSKSRQVKMAETFTSTVALFSSVDIVHMFASRARKSMVM (SEQ ID NO: 1680) </p>
--	--

Example 10: Inactivation of an endogenous nucleolar localization signal in a Gene Writer

This example describes a Gene Writer in which an endogenous nucleolar localization signal has been inactivated to reduce intracellular targeting of the protein to the nucleolus.

In this example, the nucleolar localization signal (NoLS) of a retrotransposase is computationally predicted using a published algorithm that was trained on validated proteins that localize to the nucleolus (Scott, M. S., et al, Nucleic Acids Research, 38(21), 7388–7399 (2010)). The predicted NoLS sequence is based on both amino acid sequence, amino acid sequence context, and predicted secondary structure of the retrotransposase. The identified sequence is typically rich with basic amino acids (Scott, M. S., et al, Nucleic Acids Research, 38(21), 7388–

7399 (2010)) and when these residues are mutated to a simple side-chain, non-basic, amino acids or removed from the retrotransposase polypeptide chain then it can prevent localization to the nucleolus (Yang, C. P., et. al., Journal of Biomedical Science, 22(1), 1–15. (2015), Martin, R. M., et. al., Nucleus, 6(4), 314–325 (2015)). In some embodiments, the NoLS sequence is located in the amino acid region of a retrotransposase that is between the reverse transcriptase polymerase motif and the restriction-like endonuclease motifs. The predicted NoLS region contains lysine, arginine, histidine, and/or glutamine amino acids where nucleolar localization is inactivated by mutation of one or more of these residues to an alanine amino acid residue and/or one or more of these amino acids are removed from the polypeptide chain of the retrotransposase. In some embodiments, the amino acid sequence of the Gene Writer driver of R2Tg found upstream of the RLE is mutated such that lysines (K) are substituted for alanines (A), e.g., the predicted NoLS of R2Tg (amino acids 1,128-1,154 of polypeptide sequence), (APTQKDKFPKPCNWRKNEFKKWKLAS (SEQ ID NO: 1681)) is mutated at 1, 2, 3, 4, 5, 6, or 7 residues to produce an inactivated NoLS (APTQADAFPAPCNWRANEFAAWTALAS (SEQ ID NO: 1682)).

Example 11: Application of second-strand nicking in a Gene Writer system

This example describes a Gene Writer system in which retrotransposition is paired with targeted second-strand nicking activity in order to increase the efficiency of integration events. The second strand nick can be achieved by (1) a Cas9 nickase fused to a gene writer system, in which the Gene Writer introduces one nick through its endonuclease domain (EN), and the fused nickase Cas9 places another nick on either the top and bottom DNA strands (Fig 7A), or (2) a GeneWriter system in which the active EN domain introduces a nick, and a Cas9 nickase introduces a second nick on either top or bottom strand of the DNA, upstream or downstream of the Gene Writer induced nick (Fig 7B).

In the first part of this example, a Cas9 nickase is fused to a Gene Writer protein (Fig 7A). The Cas9 is targeted to a DNA sequence through a gRNA. The Gene Writer protein introduces a DNA nick through its EN domain, and an additional nick is generated through the nickase Cas9 activity. This additional nick can be targeted to the top or bottom strands of the

DNA surrounding the Gene Writer introduced nick (Fig 8A). Constructs designed and tested include (see schematic Fig 14A):

- Cas9-N863A-R2tg (RBD*, RT, EN)
- Cas9-H840A-R2tg (RBD*, RT, EN)
- Cas9-D10A-R2tg (RBD*, RT, EN)
- dCas9- R2tg (RBD*, RT, EN)

The DNA binding domain is the nickase Cas9 that directs the Gene Writer molecule to a DNA target through a gRNA. The RNA binding domain (RBD) in this set of Gene Writer constructs is inactivated with a point mutation (RBD*). As a donor for insertion, constructs in which the R2Tg RNA binding domain is inactive use a gRNA that is extended at its 3' end to include donor sequence for genome modification (Fig 14B). These modifications include nucleotide substitutions, nucleotide deletions and nucleotide insertions. In this first set of experiments, the above constructs-R2Tg(RBD*, RT, EN) and dCas9- R2Tg(RBD*, RT, EN) fusions with a 3' extended gRNA template targeting the AAVS1 locus are delivered to U2OS cells through nucleofection in SE buffer using program DN100. gRNAs used include gRNAs for each construct that target either the bottom or top strand of DNA. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3, and amplicon sequencing followed by computational analysis using CRISPResso (indel analysis tool) are performed. 3' extended gRNA mediated insertions, deletions or nucleotide substitutions are observed upon delivery of dCas9-R2Tg(RBD*, RT, EN), and increased in frequency when delivering Nickase-Cas9-R2Tg(RBD*, RT, EN) constructs.

In the second part of this example, a Cas9 nickase is fused to a Gene Writer protein (Fig 7A). The Cas9 is targeted to a DNA sequence through a gRNA. The Gene Writer protein introduces a DNA nick through its EN domain, and an additional nick is generated through the nickase Cas9 activity. This additional nick can be targeted to the top or bottom strands of the DNA surrounding the Gene Writer introduced nick (Fig 7A). In contrast to the constructs listed above, the RNA binding domain of R2Tg is active (Fig 15A), and the template used for genome modification is a transgene flanked by UTRs (Fig 15B). Constructs include (see schematic Fig 15A):

- Cas9-N863A-R2tg (RBD, RT, EN)
- Cas9-H840A-R2tg (RBD, RT, EN)

- Cas9-D10A-R2tg (RBD, RT, EN)
- dCas9- R2tg (RBD, RT, EN)

The transgene flanked by UTRs requires homology arms at the site of nicking. To determine the site of nicking for the accurate design of homology arms for the donor transgene DNA, the above listed constructs are nucleofected into 200k U2OS cells with a gRNA targeting the AAVS1 locus using pulse code DN100. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3, and amplicon sequencing followed by computational analysis using CRISPResso as an indel analysis tool are performed. The nicking site of the EN domain is identified from the indels the EN domain produces at the AAVS1 site. Homology arms of 100 bp flanking the EN nicking site are designed and included in the transgene. To achieve genome modification, Cas9-R2Tg fusion constructs listed above are nucleofected into U2OS cells, along with a gRNA targeting either the top or bottom strand of the AAVS1 locus, and the appropriate transgenes harboring homology to the previously determined nicking site. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3, and ddPCR is performed to detect transgene integration at the AAVS1 site. Integrations are observed upon delivery of dCas9-R2Tg(RBD, RT, EN), and increased in frequency when delivering Nickase-Cas9-R2Tg(RBD, RT, EN) constructs.

In another example, a Gene Writer protein is targeted to DNA through its DNA binding domain (Fig 7B). The Gene Writer protein will introduce a DNA nick at a DNA strand. In addition, a Cas9 nickase is used to generate a second nick either on the top or bottom strands of the DNA, upstream or downstream of the first nick. In this example, a Gene Writer plasmid targeting the AAVS1 site (Fig 16A) and with a UTR flanked transgene with homology to the AAVS1 site (Fig 16B) is nucleofected into 200k U2OS cells using pulse code DN100. The following Cas9 constructs are transfected alongside the Gene Writer plasmids (Fig 16C):

- Cas9-N863A
- Cas9-H840A
- Cas9-D10A
- dCas9

All Cas9 constructs are co-nucleofected with gRNAs targeting the AAVS1 locus on either the top or bottom strands, upstream or downstream of the Gene Writer introduced nick. After nucleofection, cells are grown in complete medium for 3 days. gDNA is harvested on day 3, and

ddPCR is performed to detect transgene integration at the AAVS1 site. Integrations are observed upon delivery of dCas9 and increased in frequency when delivering Nickase-Cas9 constructs.

Example 12: Improved expression of Gene Writer polypeptide by heterologous UTRs

This example describes the use of heterologous UTRs to enhance the intracellular expression of the Gene Writer polypeptide.

In this example, the Gene Writer polypeptide was expressed from mRNA (Figure 17). In the plasmid template for the mRNA production, the native retrotransposon UTRs were replaced with UTRs optimized for the protein expression (C3 5'UTR and ORM 3' UTR from Asrani et al., RNA biology 15, 756-762 (2018) or 5' and 3' UTRs from Richter et al., Cell 168, 1114-1125 (2017)). The plasmid included the T7 promoter followed by the 5'UTR, the retrotransposon coding sequence, the 3' UTR, 3GS linker (SEQ ID NO: 1024), SV40 nuclear localization signal (NLS), XTEN linker, HiBit sequence and 96-100 nucleotide long poly(A) tail (SEQ ID NO: 1683). The plasmid was linearized by enzymatic restriction resulting in blunt end or 5' overhang downstream of poly(A) tail and used for in vitro transcription (IVT) using T7 polymerase (NEB). Following the IVT step the RNA was treated with DNase I (NEB). After the buffer exchange step the enzymatic capping reaction was performed using Vaccinia capping enzyme (NEB) and 2'-O-methyltransferase (NEB) in the presence of GTP and SAM (NEB). The capped RNA was concentrated and buffer exchanged. 50,000 HEK293T cells were transfected with 0.5 µg with the Gene Writer mRNA in the presence or in the absence of the RNA template in 1:1 molar ratio using Neon transfection system (1150 V per pulse, 20 msec per pulse, 2 pulses in 10 µL tips in 96 well format). The RNA template was in vitro transcribed from plasmid as described in Example 14 (Improved Gene Writer components for RNA-based delivery).

After transfection HEK293T cells were grown for 5 hours before assaying the Gene Writer expression by probing its HiBit tag expression using standard protocol <https://www.promega.com/-/media/files/resources/protocols/technical-manuals/500/nano-glo-hibit-lytic-detection-system-technical-manual.pdf?la=en>. Protein expression was found to be greatly improved by the use of 5' and 3' UTR_{exp} from C3-ORM as compared to using the native UTRs from R2Tg (Figure 17). The genome integration was assayed 3 days post-transfection using 3' ddPCR (Figure 18).

Example 13: Improved Gene Writer components for mixed RNA and DNA delivery

This example describes improvements to the RNA molecule encoding a Gene Writer polypeptide that enhance expression and allow for increased efficiency of retrotransposition when used with a Gene Writer template encoded on plasmid DNA.

In this example, the polypeptide component of the Gene Writer™ system is expressed from mRNA described in Example 12 (Improved expression of Gene Writer polypeptide by heterologous UTRs). The plasmid template was synthesized such that the reporter gene (eGFP) was flanked by R2Tg untranslated regions (UTRs) and 100 bp of homology to its rDNA target. The template expression was driven by the mammalian CMV promoter. We introduced the plasmid into HEK393T cells using the FuGENE® HD transfection reagent. HEK293T cells were seeded in 96-well plates at 10,000 cells/ well 24 hours before transfection. On the transfection day, 0.5 µl transfection reagent and 80 ng DNA was mixed in 10 µl Opti-MEM and incubated for 15 minutes at room temperature. The transfection mixture was then added to the medium of the seeded cells. Cells were detached and used for the electroporation of 0.5 µg of mRNA per well using Neon transfection system (1150 V per pulse, 20 msec per pulse, 2 pulses in 10 µL tips in 96 well format).

HEK293T cells were transfected with the following test agents:

1. mRNA coding for the polypeptide described above
2. Plasmid encoding template RNA described above
3. Combination of 1 and 2. The plasmid was pre-lipofected 24 hrs before mRNA transfection as described above.

After transfection, HEK293T cells were cultured for 1-3 days and then assayed for site-specific genome editing. Genomic DNA was isolated from each group of HEK293 cells.

ddPCR was performed to confirm integration and assess integration efficiency. Taqman probes and primers were designed as described in PCT/US2019/048607 to amplify the expected product across 5' and 3' ends of integration junctions. The results of the ddPCR copy number analysis (in comparison to reference gene RPP30) are shown in Figure 19. The genome integration in the presence of the mRNA and the template plasmid achieved a mean copy number of 0.683 integrants/genome when targeting 3' junction and of 0.249 integrants/genome when targeting 5'

junction. The mRNA only transfection resulted in a mean copy number of 0.002 integrants/genome, in comparison to 0.0004 integrants/genome for the plasmid only transfection.

Example 14: Improved Gene Writer components for RNA-based delivery

This example describes improvements to the RNA molecule encoding a Gene Writer polypeptide that enhance expression and allow for increased efficiency of retrotransposition when co-delivered with a Gene Writer RNA template.

In this example, the polypeptide component of the Gene Writer™ system is expressed from mRNA described in Example 12 (Improved expression of Gene Writer polypeptide by heterologous UTRs). The plasmid template for the RNA template production included T7 promoter followed by the IRES-expressing reporter gene (eGFP) flanked by R2Tg untranslated regions (UTRs) and 100 bp of homology to its rDNA target. The plasmid template was linearized by enzymatic restriction resulting in blunt end or 5' overhang downstream of the RNA template sequence and used for in vitro transcription (IVT) using T7 RNA polymerase (NEB). Following the IVT step the RNA was treated with DNase I (NEB) and either enzymatically polyadenylated by poly(A) polymerase (NEB) or not. After the buffer exchange step the enzymatic capping reaction was performed using Vaccinia capping enzyme (NEB) and 2'-O-methyltransferase (NEB) in the presence of GTP and SAM (NEB). The capped RNA was concentrated and buffer exchanged. 50,000 HEK293T cells were co-transfected with 0.5 to 1 µg of the GeneWriter mRNA and the RNA template in 1:4 to 1:12 molar ratios. The Neon transfection system was used for the RNA transfection (1150 V per pulse, 20 msec per pulse, 2 pulses in 10 µL tips in 96 well format).

After transfection, HEK293T cells were cultured for at least 1 day and then assayed for site-specific genome editing. Genomic DNA was isolated from each group of HEK293 cells. ddPCR was performed to confirm integration and assess integration efficiency. Taqman probes and primers were designed as described in PCT/US2019/048607 to amplify the expected product across 5' and 3' ends of integration junctions. The mean copy number of 0.498 integrants/genome was achieved in the presence of the 0.5 µg of mRNA and 1:8 molar ratio of Gene Writer mRNA to the RNA template when the RNA template was enzymatically

polyadenylated, in comparison to that of 0.031 integrants/genome when the RNA transgene was not polyadenylated.

Example 15: Gene Writers that deliver genetic cargo containing introns

This example describes the use of a Gene Writer system to integrate genetic cargo that contains introns by using RNA-based delivery to tune expression of the gene of interest from its newly introduced genomic locus.

In this example, Gene writing technology uses an RNA template encoding a protein of interest including its native or non-native introns. For example, intron 6 of the triose phosphate isomerase (TPI) gene (Nott et al., 2003) will be used as one of the non-native introns in these experiments.

The presence of introns in the genomic copy of a gene and their removal by splicing has been reported to affect nearly every aspect of the gene expression, including its transcription rate, the mRNA processing, export, cell localization, translation and decay (reviewed in Shaul International Journal of Biochemistry and Cell Biology 91B, 145-155 (2017)). The introns can be inserted into different parts of the RNA template (Figure 21) and depending on the intron location their role in gene expression can differ.

An intron in the 5' UTR_{exp}, close to the transcription start site, introduces activating chromatin modifications (Bieberstein et al., Cell Reports 2, 62-68 (2012)), improves accuracy of transcription start site recognition and facilitates PolII recruitment (Laxa et al., Plant Physiology 172, 313-327 (2016)), increases rates of transcription initiation (Kwek et al., Nature Structural Biology 9, 800-805 (2002)) and elongation (Lin et al., Nature Structural and Molecular Biology 15, 819-826 (2008)), and improve the productive elongation in the sense relative to the antisense orientation (Almada et al., Nature 499, 360-363 (2013)).

An intron in the 3' UTR_{exp} limits the mRNA expression to one protein molecule per mRNA: the exon junction complex (EJC) left by spliceosome downstream of stop codon is recognized by the nonsense-mediated decay (NMD) machinery and therefore the mRNA is marked for deletion at the end of the pioneering round of translation (Zhang et al., RNA 4, 801-815 (1998)).

The ability to employ introns in a therapeutic gene may, however, be limited by splicing that occurs prior to integration of the template. For example, an intron in the forward orientation

would be spliced out when an RNA template was encoded and delivered on a DNA plasmid, since transcription in the same direction would yield a template RNA that would be spliced prior to integration, thus failing to incorporate the intron in the genome. Additionally, lentivirus constructs designed to deliver a transgene must encode a sequence with an intron in the reverse orientation, since the viral packaging process would result in intron splicing and absence of the intron in packaged viral particles (Miller et al. J Virol 62, 4337-45 (1988)). However, the reverse orientation has also been thought to result in a reduction in viral titer and transduction (Uchida et al., Nat Commun 10, 4479 (2019)). It is worth noting that since the Gene Writer template can be generated through *in vitro* transcription and delivered directly as RNA, the problem of pre-integration splicing of desired introns can be avoided. In some embodiments, the Gene Writer template may thus contain one or more introns in same-sense orientation with the transcript, which is generated by IVT and delivered to the target cell as RNA.

An intron in any location depicted in Figure 21 will recruit U1 snRNP that protects mRNA from the premature cleavage and polyadenylation (Kaida et al., Nature 468 664-681 (2010); Berg et al., Cell 150, 53-64 (2012)). In addition, the EJC interacts with components of the TREX (transcription-export) complex and increases the rate of mRNA export from nucleus to cytoplasm 6-10-fold in comparison to the constructs lacking introns (Valencia et al., PNAS 105, 3386-3391 (2008)). It was also demonstrated that the binding of the polypyrimidine tract-binding protein, a splicing regulator protein, mediates a significant increase in the half-life of the spliced transcripts (Lu & Cullen, RNA 9, 618-630 (2003); Millevoi et al., Nucleic Acid Research 37, 4672-4683 (2009)). The efficiency of the mRNA translation was shown to be increased by the presence of the SR proteins (serine-arginine rich proteins, involved in RNA splicing) (Sanford et al., Genes & Development 18, 755-768 (2004); Sato et al., Molecular Cell 29, 255-262 (2008)) and the EJC proteins and its peripheral factors (Nott et al., Genes & Development 18, 210-222 (2004)).

In this example both the template RNAs harboring an intron or introns and Gene Writer polypeptide are delivered to the cells as *in vitro* transcribed capped RNAs as described in Example 14 (Improved Gene Writer components for RNA-based delivery). One to three days post-transfection the GOI expression and the genomic integration are assayed. In some embodiments, the genome integration and/or protein expression will be higher for the intron-containing RNA template.

Example 16: Engineering of the retrotransposon 5' UTR to improve efficiency of integration

This example describes the deletion, replacement, or mutation of the 5'UTR of a retrotransposon to increase integration efficiency.

The 5'UTR region of non-LTR retrotransposons has multiple functions including self-cleaving ribozyme activity, which has been shown in certain elements and is predicted in additional retrotransposons (see modules B and C of Figure 27-28) (Ruminski et al. J Biol Chem 286, 41286-41295 (2011)). Ribozymal activity is predicted to cleave the RNA within or upstream of the 5'UTR. Either increasing or restricting this activity and structural component of the 5'UTR may benefit retrotransposition efficiency. A prediction of the ribozyme structure of R2Tg is provided in Figure 29.

In order to evaluate engineering of the 5'UTR, constructs were designed to enhance or diminish these activities (Figure 20). In case (A), the natural 5'UTR of R2Tg is used to integrate in trans as in previous experiments. Case (B) illustrates deletion of the 5'UTR. (C) and (D) represent cases in which the 5'UTR from the original species (in this case R2Tg from *T. guttata*) has been replaced by the 5'UTR of a retrotransposon from a distinct species. Case (C) provides an example in which the 5'UTR from *A. maritima* R2 has replaced that of R2Tg. (D) represents the generic case in which UTRs from additional species may be substituted ("Rx"), such as that from *B. mori*, *D. ananasse*, *F. auricularia*, *L. polyphemus*, *N. giraulti*, or *O. latipes*, or from a retrotransposon selected from a Table herein, or any of Tables 1-3 of PCT/US2019/048607, herein incorporated by reference in its entirety. Case (E) represents the substitution of a ribozyme, such as a hammerhead ribozyme, e.g., RiboJ (Lou et al Nat Biotechnol 30, 1137-1142 (2012)). Case (F) represents the inactivation of the 5'UTR of R2Tg through point mutations, e.g., 75C>T in the 5' UTR (Figure 20.B, position indicated by shaded box). 5'UTR sequences are expected to be modular to any insertion sequence mediated by the retrotransposon.

Each case is evaluated as in previous examples by transfection of Gene Writer polypeptide plasmid with template plasmid and evaluation of integration frequency via ddPCR. In some embodiments, substitution or mutation of the 5' UTR results in increased efficiency of integration.

Example 17: Modifying the 5' and 3' ends of Gene Writer RNA components to improve RNA stability

This example describes the addition of non-coding sequences to the 5' and 3' ends of RNA in order to improve stability in a mammalian cell.

The decay of eukaryotic RNAs in cells are mostly carried out by exoribonucleases. In this example, the half-life of RNAs is prolonged by introducing protective sequences and/or modifications at their 5' and 3' ends. The most common natural way of protecting the RNA ends is by introduction of 5' cap structure and 3' poly(A) tail. In this example, the polypeptide component of the Gene Writer™ system is expressed from mRNA described in Example 12 (Improved expression of Gene Writer polypeptide by heterologous UTRs). The plasmid template for the RNA template production included T7 promoter followed by the IRES-expressing reporter gene (eGFP) flanked by R2Tg untranslated regions (UTRs) and 100 bp of homology to its rDNA target. The plasmid template was linearized by enzymatic restriction resulting in blunt end or 5' overhang downstream of the RNA template sequence and used for in vitro transcription (IVT) using T7 polymerase (NEB). Following the IVT step the RNA was treated with DNase I (NEB) and either enzymatically polyadenylated by poly(A) polymerase (NEB) or not. After the buffer exchange step the enzymatic capping reaction resulting in cap 1 structure was performed as described in Example 14 (Improved Gene Writer components for RNA-based delivery) or not performed. The template RNA was concentrated and buffer exchanged. 50,000 HEK293T cells were co-transfected with 0.5 µg with the GeneWriter mRNA and the RNA template in 1:1 to 1:8 molar ratios using Neon transfection system (1150 V per pulse, 20 msec per pulse, 2 pulses in 10 µL tips in 96 well format).

After transfection, HEK293T cells were cultured for 1-3 days and then assayed for site-specific genome editing. Genomic DNA was isolated from each group of HEK293 cells. ddPCR was performed to confirm integration and assess integration efficiency. Taqman probes and primers were designed as described in PCT/US2019/048607 to amplify the expected product across 3' end of integration junctions. The genome integration was improved when the enzymatically capped and poly(A) tailed template was used (Figure 22).

The mean copy number of 0.498 integrants/genome was achieved in the presence of the 0.5 µg of mRNA and 1:8 molar ratio of mRNA:RNA template when the RNA template was

enzymatically polyadenylated, in comparison to that of 0.031 integrants/genome when the RNA transgene was not enzymatically polyadenylated.

3' end modifications of RNAs.

It has been reported that the interactions between poly(A) tail shorter than 15-20 nts and the poly(A) binding protein (PABP) are destabilized resulting in the fast degradation of the RNA (Chang et al., *Molecular Cell* 53, 1044-1052 (2014); Subtelny et al., *Nature* 508, 66-71 (2014)). To determine the suitable lengths of the poly(A) tail of the template RNA we will test its lengths of 30, 40, 50, 60, 70, 80, 90 and 100 nucleotides. The IVT templates will be produced by PCR using reverse primers encoding the poly(A) tails of the abovementioned length. The IVT, DNase I treatment and capping of Gene Writer and the RNA template will be performed as described in Example 14 (Improved Gene Writer components for RNA-based delivery). After one to three days post-transfection the genomic integration will be assayed. In some embodiments, the genome integration will be higher for the RNA template tailed with a poly(A) tail of a suitable length.

In a cell the RNA degradation is initiated by shortening its poly(A) tail by deadenylases. Since the deadenylases are 3'-5' exoribonucleases favoring the poly(A) stretches, the terminal uridine, cytidine and most often guanine detected in the natural poly(A) tails of many mRNA were proposed to protect the poly(A) tail from its shortening (Chang et al., *Molecular Cell* 53, 1044-1052 (2014)). We will assay the Gene Writer and template RNAs with the encoded poly(A) tail with terminal G or C, or intermittent Gs or Cs (similar to that used in Lim et al., *Science* 361, 701-704 (2018)) according as described before.

Some of the RNAs have been described to evolve alternative ways of protections their 3' ends. A specific 16-nucleotide long stem-loop structure flanked with unpaired 5 nucleotides on both sides has been reported to protect the 3' end of mRNA encoding H2a.X histone (Mannironi et al., *Nucleic Acid Research* 17, 9113-9126 (1989)). It has been shown that the heterologous mRNA ending with the histone stem-loop structure is cell cycle-regulated (Harris et al., *Molecular Cellular Biology* 11, 2416-2424 (1991); Stauber et al., *EMBO Journal* 5, 3297-3303 (1986)). The stem-loop structure is recognized and protected by the Stem-Loop Binding Protein (SLBP). The protein accumulates shortly before cells enter S-phase and is rapidly degraded at the end of S-phase (Whifield et al., *Molecular Cellular Biology* 20, 4188-4198 (2000)). The

stem-loop element will be inserted to the 3' end of the Gene Writer mRNA and the RNA templates and tested as described above to induce cell-cycle specific genome integration events.

Some viral and long non-coding RNAs have evolved to protect their 3' ends with triple-helical structures (Brown et al., PNAS 109, 19202-19207 (2012)). Additionally, the structural elements of tRNA, Y RNA and vault RNA (reviewed in Labno et al., Biochimica et Biophysica Acta 1863, 3125-3147 (2016)) have been reported to extend half-life of these non-coding RNAs. We will insert the structures to protect the 3' end of the RNA templates and probe their efficiencies in Gene Writing system as described above.

Finally, we will incorporate dNTP, 2'-O-Methylated NTPs or phosphorothioate-NTP at the 3' of the RNA transgenes to increase the half-life of these molecules by protecting the 3' end of the RNA from exoribonucleases. We will incorporate single modified nucleotides or their stretches by extending the 3' end of the RNA by the DNA polymerases (for example, Klenow fragment) capable of extending an RNA sequence by adding modified nucleotides (Shcherbakova & Brenowitz, Nature Protocols 3, 288-302 (2008)).

A single nucleotide chemical modification of the 3' end of the RNA can be done by first oxidation of 3' terminal end of ribose sugar with sodium periodate to form a reactive aldehyde followed by conjugation of an aldehyde-reactive modified nucleotide. Alternatively, T4 DNA or T4 RNA ligases can be used for the splinted ligation (Moore & Query, Methods in Enzymology 317, 109-123 (2000)) of the stretches of modified nucleotides to the 3' end of the RNAs.

Chemical ligation of two fragments is also possible. The phosphodiester bond linkage between two RNA substrates can be formed either by activating the phosphomonoester group using a reactive imidazolide or by using a condensing reagent such as cyanogen bromide. A disadvantage of chemical ligation is that it can also result in the creation of a 2'-5' phosphodiester linkage, together with the desired 3'-5' phosphodiester linkages.

5' end modifications of RNAs

In addition to the cap 1 structure described in Example 14 (Improved Gene Writer components for RNA-based delivery) other 5' end protection groups will be explored. Particularly, we will use hypermethylated (Wurth et al. Nucleic Acid Res 42, 8663-8677 (2014)), phosphorothioate (Kuhn et al., Gene Therapy 17, 961-971 (2010)), NAD⁺-derived (Kiledjian,

Trends in Cell Biology 28, 454-464 (2018)) and modified (for example, biotinylated: Bednarek et al., Phil Trans R Soc B 373, 20180167 (2018)) cap analogs for co-transcriptional capping.

We will also label the 5' of the RNA with 5'-[γ-thio]triphosphate to create a reactive sulfur group and chemically modify the 5' end with the protective modifications using a haloacetamide derivative of the modified group.

The proposed modifications to protect 3' and 5' end of the RNA will be introduced in RNA templates and/or Gene Writer mRNA (if compatible with translation). The genome integration efficiencies of the RNAs will be tested as described in Example 14 (Improved Gene Writer components for RNA-based delivery).

Example 18: Use of modified RNA bases in a Gene Writer system

This example describes Gene Writer systems comprising modified RNA bases to potentially improve features of the system, e.g., increase efficiency of integration, decrease cellular response to foreign nucleic acids. For the Gene Writer polypeptide, the proposed modifications pertaining to the coding region are compatible with translation. For the RNA template, the proposed modifications are compatible with reverse transcription.

In this example, mRNA encoding the Gene Writer polypeptide was in vitro transcribed with a 100% replacement of the corresponding rNTP with one of the modified rNTPs: pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U) or 5-methylcytidine (5mC). Otherwise, the RNA preparation, purification and cell transfections were performed as described in the Example 14 (Improved Gene Writer components for RNA-based delivery). The gene integration capacity of the modified mRNAs was compared with that of the non-modified mRNA (G0) using ddPCR, with all polypeptide mRNAs being paired with an unmodified template RNA (Figure 23). Integration was detected when the polypeptide was encoded using each modified rNTP, with the highest signal coming from 5-MO-U and the lowest from 5mC. This demonstrates that the Gene Writer polypeptide component is functional when expressed from mRNA containing modified bases.

Further, this example describes the modularity of the Gene Writer template molecule where it is composed of all or a subset of exemplary modules listed in Figure 6 and illustrated in Figure 5. Individual modules can be produced by chemical or in vitro syntheses as a contiguous nucleic acid molecule or in separate pieces that are later combined together. The individual

modules of the Gene Writer template molecule can be chemically modified nucleic acids, be comprised in part or in entirety of non-nucleic acids, re-arranged in order, and/or omitted to form the Gene Writer template molecule.

In some embodiments, the Gene Writer template molecule (all modules, A-F) is synthesized by in vitro transcription where 0-100% replacement of a corresponding rNTP (adenosine, cytidine, guanosine, and/or uridine) is with one or more modified rNTPs (base or ribose modification), e.g., 5' hydroxyl, 5' Phosphate, 2'-O-methyl, 2'-O-ethyl, 2'-fluoro, ribothymidine, C-5 propynyl-dC (pdC), C-5 propynyl-dU (pdU), C-5 propynyl-C (pC), C-5 propynyl-U (pU), 5-methyl C, 5-methyl U, 5-methyl dC, 5-methyl dU methoxy, (2,6-diaminopurine), 5'-Dimethoxytrityl-N4-ethyl-2'-deoxyCytidine, C-5 propynyl-fC (pfC), C-5 propynyl-fU (pfU), 5-methyl fC, 5-methyl fU, C-5 propynyl-mC (pmC), C-5 propynyl-fU (pmU), 5-methyl mC, 5-methyl mU, LNA (locked nucleic acid), MGB (minor groove binder) pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U). The modified nucleotides in this embodiment rely on incorporation through a transcription reaction which utilizes a natural or mutant polypeptide sequence of a RNA polymerase that readily incorporates modified nucleotides into a RNA transcript that is made in vitro (Padilla, R., *Nucleic Acids Research*, 30(24), 138e – 138, 2002; Ibach, J., et. al., *Journal of Biotechnology*, 167(3), 287–295, 2013; Meyer, A. J., et. al., *Nucleic Acids Research*, 43(15), 7480–7488, 2015). The modified Gene Writer template molecule is typically in whole or in part compatible with the reverse transcriptase activity of the Gene Writer polypeptide sequence; for modules or parts of modules of the Gene Writer template molecule used as a template for reverse transcription, preference is given to modifications that are compatible with reverse transcription (Motorin et al., *Methods in Enzymology* 425 21-53, 2007; Mauger et al., *PNAS* 116, 24075-24083, 2019). Gene Writer systems with template molecules containing modified rNTPs are tested as described above and in Example 14 (Improved Gene Writer components for RNA-based delivery).

In some embodiments, individual modules are chemically synthesized containing modified nucleotides, e.g., 5' hydroxyl, 5' Phosphate, 2'-O-methyl, 2'-O-ethyl, 2'-fluoro, ribothymidine, C-5 propynyl-dC (pdC), C-5 propynyl-dU (pdU), C-5 propynyl-C (pC), C-5 propynyl-U (pU), 5-methyl C, 5-methyl U, 5-methyl dC, 5-methyl dU methoxy, (2,6-diaminopurine), 5'-Dimethoxytrityl-N4-ethyl-2'-deoxyCytidine, C-5 propynyl-fC (pfC), C-5 propynyl-fU (pfU), 5-methyl fC, 5-methyl fU, C-5 propynyl-mC (pmC), C-5 propynyl-fU

(pmU), 5-methyl mC, 5-methyl mU, LNA (locked nucleic acid), MGB (minor groove binder) pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U), where the individual modules are then ligated together through enzymatic (e.g., splint ligation using T4 DNA ligase, Moore, M. J., & Query, C. C. *Methods in Enzymology*, 317, 109–123, 2000) or chemical processes (e.g., Fedorova, O. A., et. al., *Nucleosides and Nucleotides*, 15(6), 1137–1147, 1996) to form a complete Gene Writer template molecule.

An example of a modified Gene Writer template molecule is where modules A and F are each 100 nt of chemically synthesized RNA with cytidine and uridine nucleotides containing 2'-*O*-methyl ribose modifications and module A contains (3) phosphorothioate linkages between the first 3 nucleotides on the 5' end and module F contains (3) phosphorothioate linkages between the last 3 nucleotides on the 3' end of the module. Modules B-E are synthesized by in vitro transcription using an RNA polymerase (RNAP), e.g., T7 RNAP, T3 RNAP, or SP6 RNAP (NEB), or derivatives thereof that possess enhanced properties, e.g., increased fidelity, increased processivity, or increased efficiency of incorporating modified nucleotides. Module A is ligated to the 5' end of the in vitro transcribed module B-E molecule and module F is ligated on to the 3' end of the in vitro transcribed module B-E molecule by splint ligation (described by Moore, M. J., & Query, C. C. *Methods in Enzymology*, 317, 109–123, 2000). This fully assembled template RNA (all modules, A-F) is then used with a Gene Writer polypeptide (or nucleic acid encoding the polypeptide) in a target cell to assess genomic integration as in previous examples. In some embodiments, RNA modifications do not decrease the efficiency of integration greater than 50%, e.g., as measured by ddPCR. In some embodiments, RNA modifications improve the efficiency of integration, e.g., as measured by ddPCR. In some embodiments, RNA modifications improve the reverse transcription reaction, e.g., improve the processivity or fidelity as measured by sequencing of integration events.

Example 19: Gene Writer templates that do not incorporate UTRs

This example describes a configuration of the Gene Writer template molecule that results in an exclusion of the UTRs, such that these regions used in retrotransposition are not integrated into the host cell.

In this example, we describe the positioning, omission, and/or substitution of the UTR modules of the Gene Writer template molecule (Figures 5 and 6) to result in the Gene Writer

driver to not incorporate the UTR modules into the genome as a part of retrotransposition. In some embodiments, the Gene Writer template molecule modules for the 5' and 3' UTRs (modules B+C and E of Gene Writer template molecule) are moved to the ends of the molecule so that their function of interacting with the Gene Writer driver does not change but the homology arm is now located adjacent to the heterologous object sequence (module D) where complementarity of the homology arms act as a primer for reverse transcription. In some cases, modules B and/or C are omitted from the Gene Writer template molecule with module E following module F.

Additional examples of not incorporating the UTRs into the genome are removing modules B and C from the Gene Writer template molecule, re-positioning module F (3' homology arm) to follow module D (heterologous object sequence) and have module E be substituted with a binding ligand such as biotin. This Gene Writer template molecule would now consist of module A (5' homology arm) – module D (heterologous object sequence) – module F (3' homology arm) – module E comprised of biotin. The Gene Writer driver polypeptide sequence would be modified to incorporate the amino acid sequence for monomeric streptavidin. This example illustrates how the utility of mediating a non-nucleic acid mediated association of the Gene Writer template molecule with the Gene Writer driver polypeptide sequence.

Example 20: Gene Writers can integrate genetic cargo independently of the homology directed repair pathway

This example describes the use of a Gene Writer system in a human cell wherein the homologous recombination repair pathway is inhibited.

In this example, U2OS cells were treated with 30pmols (1.5 μ M) non-targeting control siRNA (Ctrl) or a siRNA against Rad51, a core component of the homologous recombination repair pathway. SiRNAs were co-delivered with R2Tg driver and transgene plasmid in *trans* (see Fig 24 for driver and transgene configuration schematic). Specifically, Plasmid expressing R2Tg, control R2Tg with a mutation in the RT domain, or control R2Tg with an endonuclease inactivating mutation were used in conjunction with transgene (Fig 25 A, B). A total of 250ng DNA plasmids with a 1:4 molar ratio of driver to transgene, along with 30pmol of siRNAs were nucleofected into 200k U2OS cells resuspended in 20 μ L of nucleofection buffer SE using pulse

code DN100. Protein lysates collected on day 3 showed the absence of Rad51 in the siRad51 treated condition (Fig 25C). gDNA was extracted at day 3 and ddPCR assays to detect transgene integration at the rDNA locus was performed. The results of the ddPCR copy number analysis (in comparison to reference gene RPP30) are shown in Figure 26. The absence of Rad51 leads to a ~20% reduction in R2Tg mediated transgene integration at the rDNA locus both at the 3' and 5' junctions (Fig 26), indicating that R2TG mediated transgene insertion is not wholly dependent on the presence of the homologous recombination pathway, and can occur in the absence of the endogenous HR pathway. In some embodiments, HR independence enables Gene Writing to work in cells and tissues with endogenously low levels of HR, e.g., liver, brain, retina, muscle, bone, nerve, cells in G₀ or G₁ phase, non-dividing cells, senescent cells, terminally differentiated cells. In some embodiments, HR independence enables Gene Writing to work in cells or in patients or tissues containing cells with mutations in genes involved in the HR pathway, e.g., BRCA1, BRCA2, P53, RAD51.

Example 21: 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 (1.5 μ M) 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 22: Gene Writer systems with enhanced activity for target vs non-target cells

This example describes the incorporation of regulatory sequences into Gene Writer systems in order to decrease integration activity in non-target cells.

In this example, genetic regulation is accomplished through (i) using tissue-specific promoters to upregulate component expression and integration in target cells and (ii) using miRNA binding sites to decrease integration in non-target cells that have increased endogenous levels of the corresponding miRNA. Target cells used are human hepatocytes and non-target cells are hematopoietic stem cells (HSCs). The driver of integration here is a plasmid encoding the Gene Writer polypeptide (e.g., R2Tg retrotransposase) driven by different promoters and with scrambled or specific miRNA binding sites after the coding sequence. The template for integration is encoded on plasmid DNA, such that transcription results in a homology- and UTR-flanked heterologous object sequence. The heterologous object sequence may comprise a reporter gene that is driven by different promoters and with scrambled or specific miRNA binding sites after the coding sequence. The control promoter used here is CMV and the control for miRNA binding site is a randomly scrambled version of the binding site for miR-142. The target tissue-specific promoter used here is ApoE.HCR.hAAT, which is expressed in liver cells, and the off-target tissue-specific miRNA binding site is complementary to miR-142 (uguaguguuuuccuacuuuaugga (SEQ ID NO: 1684)), which is expressed in HSCs.

Target cells and non-target cells are nucleofected with a combination of Gene Writer polypeptide (1) and template (2) selected from:

Gene Writer polypeptide constructs (1):

- a. Non-specific driver: CMV-R2Tg
- b. Non-specific inactivated driver: CMV-R2Tg(EN*)
- c. Tissue-specific driver: ApoE.HCR.hAAT-R2Tg-miR142
- d. Tissue-specific inactivated driver: ApoE.HCR.hAAT-R2Tg(EN*)-miR142

Gene Writer template constructs (2):

- a. Non-specific transgene: CMV-gfp
- b. Tissue-specific transgene: ApoE.HCR.hAAT-gfp-miR142

Cells are incubated for at least three days and subsequently evaluated for integration efficiency and reporter expression. For integration efficiency, ddPCR is performed to quantify the average number of integrations per genome for each sample. In some embodiments, the ratio between the integration efficiency in target cells and non-target cells is higher when using a template paired

with the tissue-specific driver (1a) vs a non-specific driver (1c). To assess reporter expression, cells are analyzed by flow cytometry to detect GFP fluorescence and RT-qPCR to detect transcription. In some embodiments, the ratio between fluorescence in target cells and non-target cells is higher when using a driver paired with a tissue-specific transgene cassette (2b) vs a non-specific transgene cassette (2a). In some embodiments, the ratio between transcript levels in target cells and non-target cells is higher when using a driver paired with a tissue-specific transgene cassette (2b) vs a non-specific transgene cassette (2a). In some embodiments, the combination of a tissue-specific driver (1a) with a tissue-specific transgene cassette (2b) results in the highest ratio of transcription or expression between target and non-target cells. Alternatively, a screening assay can be performed in the same cell line artificially expressing or not expressing a given miRNA, e.g., the on-target screening cell is a HEK293T cell and the non-target cell is mimicked by introducing overexpression of miR-142 in HEK293T cells.

Example 23: Correction of alpha-1 antitrypsin deficiency using lipid nanoparticles comprising Gene Writers.

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence at a single nucleotide in vivo. More specifically, the Gene Writer™ polypeptide and writing template are delivered to mouse liver cells via lipid nanoparticles to correct the SERPINA1 PiZ mutation causing alpha-1 antitrypsin deficiency.

Formulation and treatment of murine models with LNPs (LNP-INT01 system) carrying Cas9 and gRNA are taught by Finn et al. *Cell Reports* 22:2227-2235 (2018), the methods of which are incorporated herein by reference.

Capped and polyadenylated Gene Writer polypeptide mRNA containing N1-methyl pseudo-U is generated by in vitro transcription using a linearized plasmid DNA template and T7 RNA polymerase. The polypeptide mRNA is purified from enzyme and nucleotides using a MegaClear Transcription Clean-up Kit, in accordance with the manufacturer's protocol (ThermoFisher). The transcript concentration is determined by measuring the light absorbance at 260 nm (Nanodrop), and the transcript is analyzed by capillary electrophoresis by TapeStation (Agilent). Template RNA comprising the mutation correcting sequence is also prepared by in

vitro transcription and translation using similar methods. In this example, the template RNA comprises the sequence as exemplified in Example 1.

LNPs are formulated with an amine-to-RNA-phosphate (N:P) ratio of 4.5. The lipid nanoparticle components are dissolved in 100% ethanol with the following molar ratios: 45 mol% LP01 lipid, 44 mol% cholesterol, 9 mol% DSPC, and 2 mol% PEG2k-DMG. The RNA cargo (1:40 molar ratio of polypeptide mRNA:template RNA) is dissolved in 50 mM acetate buffer (pH 4.5), resulting in a concentration of RNA cargo of approximately 0.45 mg/mL. LNPs are formed by microfluidic mixing of the lipid and RNA solutions using a Precision Nanosystems NanoAssemblr Benchtop Instrument, in accordance with the manufacturer's protocol. After mixing, the LNPs are collected and diluted in PBS (approximately 1:1), and then the remaining buffer is exchanged into PBS (100-fold excess of sample volume) overnight at 4C under gentle stirring using a 10 kDa Slide-a-Lyzer G2 Dialysis Cassette (ThermoFisher Scientific). The resultant mixture is then filtered using a 0.2-mm sterile filter. The filtrate is stored at 2C–8C. Multi-dose formulations may be formulated using 25 mM citrate, 100 mM NaCl cargo buffer (pH 5), and buffer exchanged by TFF into tris-saline sucrose buffer (TSS) buffer (5% sucrose, 45 mM NaCl, and 50 mM Tris). Formulated LNPs have an average size of 105 nm. Encapsulation efficiencies are determined by ribogreen assay (Leung et al., 2012). Particle size and polydispersity are measured by dynamic light scattering (DLS) using a Malvern Zetasizer DLS instrument.

NSG-PiZ mice carrying the human SERPINA1 PiZ allele (E342K) are acquired from The Jackson Laboratory. To assess the ability of Gene Writing to edit the mutant allele in vivo, LNPs are dosed via the lateral tail vein at 3 mg/kg in a volume of 0.2 mL per animal. Excipient-treated animals are used as negative controls for all studies. Animals are euthanized at various time points by exsanguination via cardiac puncture under isoflurane anesthesia. In some embodiments, animals are euthanized at one week post-treatment to be analyzed for Gene Writing. Liver tissue is collected from the median or left lateral lobe from each animal for DNA extraction and analysis.

For NGS analysis of editing efficiency, PCR primers are designed around the target site, and the region of interest is amplified from extracted genomic DNA. Additional PCR is performed in accordance with the manufacturer's protocols (Illumina) to add the appropriate chemistry for sequencing, and amplicons are then sequenced on an Illumina MiSeq. Sequencing

In some embodiments, this example is repeated with additional groups of mice and a redosing regimen is used to analyze dose-to-effect properties of the system. In these experiments, mice are assigned to groups for weekly dosing up to 4 weeks, with euthanasia and tissue analysis as described herein being performed each week. In some embodiments, mice that receive more doses of the LNP formulation demonstrate higher Gene Writing efficiency by sequencing, e.g., mice receiving 2 doses one week apart that are analyzed at week three show a higher fraction of gene corrected reads by NGS of liver tissue samples as compared to mice receiving a single dose and analyzed at week three. In application, dosing in this manner may allow tuning of therapeutic intervention after evaluating patient response to one or more doses.

This example describes the use of a Gene Writer™ gene editing system to treat a repeat expansion disease by rewriting a normal number of repeats into the locus. More specifically, the Gene Writer™ polypeptide and writing template are delivered to mouse CNS via AAV to reset the CAG repeats in HTT as per the custom template RNA to cure Huntington Disease. Healthy humans tend to carry between 10 and 35 CAG repeats within the huntingtin gene (*HTT*), while those with Huntington Disease may possess between 36 to greater than 120 repeats.

1409

TCAG (SEQ ID NO: 1685), where numbers are used to delineate the modules of the template in the order (5'-3') (1) gRNA spacer, (2) gRNA scaffold, (3) heterologous object sequence, (4) 3' homology priming domain, with the repeat correction being encoded in (3). The CAG repeat region is followed by a short repeat region encoding for 11 proline residues (8 residues being encoded by CCG triplets). Without wishing to be bound by theory, this region is included in (3) to place (4) in a more unique region to prevent mispriming. An exemplary gRNA for providing a second nick as described in embodiments of this system comprises the spacer sequence CGCTGCACCGACCGTGAGTT (SEQ ID NO: 1630) and directs a Cas9 nickase to nick the second strand of the target site within the homologous region. In some embodiments, this second nick improves the efficiency of the edit.

In order to deliver a complete Gene Writing system to the CNS, in this example, the Gene Writer is split across two AAV genomes, with the first encoding the nickase Cas9 domain fused to intein-N of a split intein pair (DnaE Intein-N:

CLSYETEILTVEYGLLPYGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL
EDGSLIRATKDHKFMFTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 1638)) and

the second encoding the RT domain fused to an intein-C of a split intein pair (DnaE Intein-C, MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO: 1640)) and the template RNA. The two polypeptide components are expressed from a polymerase II promoter, e.g. a neuronal cell-specific promoter described herein, and the template RNA and gRNA for providing a second nick are expressed from a polymerase III promoter, e.g. a U6 promoter.

When co-infecting a cell, the two polypeptide components reconstitute a complete Gene Writer polypeptide with N-terminal Cas9 and C-terminal RT and the template RNA is expressed and reverse transcribed into the target locus. To achieve delivery for cells of the CNS (specifically the claudate nucleus and the putamen of the basal ganglia), the pseudotyped system rAAV2/1 is used here, where the AAV2 ITRs are used to package the described nucleic acids into particles with AAV1 capsid. AAV preparation and mouse injection and harvesting protocols used here follow the teachings of Monteys et al. *Mol Ther* 25(1):12-23 (2017).

FVB-Tg(YAC128)53Hay/J mice are acquired from The Jackson Laboratory. These transgenic mice express the full-length human huntingtin protein with ~118 glutamine repeats (CAG trinucleotide repeats) and develop hyperkinesia at three months of age. At 8 weeks of age, mice are treated with a combination 1:1 of rAAV2/1-Cas9 virus and rAAV-

MMLV_RT/hU6templateRNA virus. For rAAV injections, mice are anesthetized with isoflurane and 5 μ L of rAAV mixture injected unilaterally into the right striata at 0.2 μ L/min. After three weeks, mice are sacrificed and brain tissue taken for genomic DNA extraction and NGS analysis.

For NGS analysis of editing efficiency, PCR primers are designed flanking the target site, and the region of interest is amplified from extracted genomic DNA. Additional PCR is performed in accordance with the manufacturer's protocols (Illumina) to add the necessary chemistry for sequencing, and amplicons are then sequenced on an Illumina MiSeq. Sequencing reads are aligned to the mouse reference genome after eliminating those having low quality scores. The resultant files containing the reads are mapped to the reference genome (BAM files), where reads that overlap the target region of interest are selected, and the number of diseased allele (>35 CAG repeats) reads versus the number of repaired allele (10-35 CAG repeats) reads are calculated. The editing percentage (e.g., the "editing efficiency" or "percent editing") is defined as the total number of repaired reads, as defined above, over the total number of sequence reads.

Example 25: Delivery of a Gene Writing system by LNP and AAV vehicles

This example describes the use of a Gene Writer™ gene editing system to alter a genomic sequence at a single nucleotide in vivo. More specifically, the Gene Writer™ polypeptide and writing template are delivered to mouse liver cells via a combination of lipid nanoparticles (mRNA encoding polypeptide) and AAV (DNA encoding the RNA template) to correct the SERPINA1 PiZ mutation causing alpha-1 antitrypsin deficiency.

Capped and tailed mRNA encoding the Gene Writer polypeptide are prepared by in vitro transcription and formulated into LNP-INT01 as described in Example 23, but without template RNA co-formulation.

In this example, the template RNA is encoded as DNA and delivered via AAV. The teachings of Cunningham et al. *Mol Ther* 16(6):1081-1088 (2008) describe the use of rAAV2/8 with the human alpha-1 antitrypsin (hAAT) promoter and two copies of the hepatic control region of the apolipoprotein E enhancer (ApoE) to effectively transduce and drive expression of cargo in juvenile mouse liver. Accordingly, rAAV2/8.ApoE-hAAT.PiZ (rAAV2/8.PiZ) as described here comprises the above described AAV and promoter system driving expression of

an RNA template for correcting the PiZ mutation, in addition to a second nick-directing gRNA being driven by a U6 promoter (RNA sequences previously described in Example 1).

NGS-PiZ mice carrying the human SERPINA1 PiZ allele (E342K) are acquired from The Jackson Laboratory. To assess the activity of Gene Writing to edit the mutant allele in vivo, 8-week-old mice are dosed i.p. with $\sim 10^{11}$ vg of rAAV2/8.PiZ to express the template RNA and via the lateral tail vein with formulated LNPs at 3 mg/kg in a volume of 0.2 mL per animal to express the Gene Writer polypeptide. Animals are euthanized at various time points by exsanguination via cardiac puncture under isoflurane anesthesia. In some embodiments, animals are euthanized at one week post-treatment to be analyzed for Gene Writing. Liver tissue is collected from the median or left lateral lobe from each animal for DNA extraction and analysis.

For NGS analysis of editing efficiency, PCR primers are designed around the target site, and the region of interest is amplified from extracted genomic DNA. Additional PCR is performed in accordance with the manufacturer's protocols (Illumina) to add the necessary chemistry for sequencing, and amplicons are then sequenced on an Illumina MiSeq. Sequencing reads are aligned to the mouse reference genome after eliminating those having low quality scores. The resultant files containing the reads are mapped to the reference genome (BAM files), where reads that overlap the target region of interest are selected, and the number of wild-type reads versus the number of reads that contain the SERPINA1 reversion mutation encoded in the template RNA are calculated. The editing percentage is defined as the total number of reversion sequence reads over the total number of sequence reads.

Example 26: Application of a Gene Writer™ system for delivering therapeutic gene to liver in a human chimeric liver mouse model

This example describes a Gene Writer™ genome editing system delivered to the liver in vivo for integration and stable expression of a genetic payload. Specifically, LNPs are used to deliver a Gene Writing system capable of integrating a complete OTC expression cassette to treat a humanized mouse model of OTC-deficiency.

In this example, a Gene Writing system is used to treat a humanized mouse model of OTC deficiency, in which human hepatocytes derived from patients with OTC deficiency are engrafted into a mouse model (Ginn et al JHEP Reports 2019). An exemplary Gene Writing

system for large payload integration comprises a Cas9-directed reverse transcriptase system utilizing a highly processive reverse transcriptase, e.g., MarathonRT. An exemplary template RNA component comprises, from 5' to 3', (1) a gRNA spacer with homology to the AAVS1 safe harbor site, (2) a gRNA scaffold, (3) a heterologous object sequence, and (4) a 3' target homology region for annealing to the genomic DNA immediately upstream of the first strand nick to prime TPRT of the heterologous object sequence. An exemplary sequence for (1) is GGGGCCACTAGGGACAGGAT (SEQ ID NO: 1689). Region (2) carries the gRNA scaffold as described in this application, generally comprising the sequence GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGGACCGAGTCGGTCC (SEQ ID NO: 1591). In this example, (3) comprises a complete OTC expression cassette, where a liver-codon-optimized sequence encoding human OTC (UniProt P00480) is in operable association with the ApoE.hAAT promoter system as described in Example 25. An exemplary sequence for (4) is CTGTCCCTAGTG (SEQ ID NO: 1690). An exemplary sequence of an additional gRNA spacer for generating a second strand nick to improve the efficiency of integration is AGAGAGATGGCTCCAGGAAA (SEQ ID NO: 1691).

Eight to 12-week-old female *Fah^{-/-}Rag2^{-/-}Il2rg^{-/-}* (FRG) mice are engrafted with human hepatocytes, isolated from pediatric donors or purchased from Lonza (Basel, Switzerland), as described previously (Azuma et al Nat Biotechnol 2007). Engrafted mice are cycled on and off 2-(2-nitro-4-trifluoro-methylbenzoyl)-1,3-cyclohexanedione (NTBC) in drinking water to promote liver repopulation. Blood is collected every two weeks and at the end of the experiment to measure the levels of human albumin, used as a marker to estimate the level of engraftment, in serum by enzyme-linked immunosorbent assay (ELISA; Bethyl Laboratories, Inc., Montgomery, TX). Eleven weeks after engraftment, mice are treated with the Gene Writer™s formulated as in Example 23. For treatment, LNPs are delivered via the lateral tail vein at 3 mg/kg in a volume of 0.2 mL per animal.

After vector injection, mice are cycled on NTBC for another 5 weeks before being euthanized. DNA and RNA are subsequently extracted from liver lysates by standard methods. OTC expression is subsequently assayed by performing RT-qPCR on isolated RNA samples using sequence-specific primers. Levels of human OTC are also measured throughout the experiment by using a human OTC ELISA kit (e.g., Aviva Systems Biology OTC ELISA Kit

(Human) (OKCD07437)) on serum at Days -7, 0, 2, 4, 7, 14, 21, 28, and 35 post-injection, following the manufacturer's recommended protocol.

For analysis of editing efficiency, a ddPCR assay is performed using a pair of primers that anneal across either the 5' junction or the 3' junction of integration, with one primer in each set annealing to the heterologous object sequence, and the other to an appropriate region of the AAVS1 site on the genome. The assay is normalized to a reference gene to quantify the number of target site integrations per genome.

To analyze integrations at the target site, long-read sequencing across the integration site is performed. PCR primers are designed flanking the target site, and the region of interest is amplified from extracted genomic DNA. Additional PCR is performed in accordance with the manufacturer's protocols (PacBio) to add the necessary chemistry for sequencing, and amplicons are then sequenced via PacBio. Sequencing reads are aligned to the mouse reference genome after eliminating those having low quality scores. The resultant files containing the reads are mapped to the reference genome (BAM files), where reads that contain an insertion sequence relative to the reference genome are selected for further analysis to determine completeness of integration, defined in this example as containing the complete promoter and coding sequence of OTC.

Example 27: Gene Writers for integration of a CAR in T-cells ex vivo

This example describes delivery of a Gene Writer™ genome editing system to T-cells ex vivo for integration and stable expression of a genetic payload. Specifically, LNPs are used to deliver a Gene Writing system capable of integrating a chimeric antigen receptor (CAR) into the *TRAC* locus to generate CAR-T cells for treating B-cell lymphoma.

In this example, a Gene Writing system comprises a Gene Writing polypeptide, e.g., a nickase Cas9 and R2Tg reverse transcriptase domain, as described herein, a gRNA for directing nickase activity to the target locus, and a template RNA comprising, from 5' to 3':

- (1) 100 nt homology to target site 3' of first strand nick
- (2) 5' UTR from R2Tg
- (3) Heterologous object sequence
- (4) 3' UTR from R2Tg
- (5) 100 nt homology to target site 5' of first strand nick

Wherein (3) comprises the coding sequence for the CD19-specific Hu19-CD828Z (Genbank MN698642; Brudno et al. *Nat Med* 26:270-280 (2020)) CAR molecule. The Gene Writer in this example is guided to the 5' end of the first exon of *TRAC* by using a targeted gRNA, e.g., TCAGGGTTCTGGATATCTGT (SEQ ID NO: 1692), in order to place the cargo under endogenous expression control from that locus while disrupting the endogenous TCR, as taught by Eyquem et al. *Nature* 543:113-117 (2017). These three components (polypeptide, gRNA, and template) all comprise RNA, which is synthesized by in vitro transcription (e.g., polypeptide mRNA, template RNA) or chemical synthesis (gRNA).

The LNP formulation used in this example has been screened and validated for delivery to T-cells ex vivo, being taught in Billingsley et al. *Nano Lett* 20(3):1578-1589 (2020), which is incorporated herein by reference in its entirety. Specifically, the LNP formulation C14-4, comprising cholesterol, phospholipid, lipid-anchored PEG, and the ionizable lipid C14-4 (Figure 2C of Billingsley et al. *Nano Lett* 20(3):1578-1589 (2020)) was used to encapsulate all three RNA components in a molar ratio of polypeptide mRNA:gRNA:template RNA of about 1:40:40.

Additional edits can be performed on T-cells in order to improve activity of the CAR-T cells against their cognate target. In some embodiments, a second LNP formulation of C14-4 as described comprises a Cas9/gRNA preformed RNP complex, wherein the gRNA targets the *Pdcd1* exon 1 for PD-1 inactivation, which can enhance anti-tumor activity of CAR-T cells by disruption of this inhibitory checkpoint that can otherwise trigger suppression of the cells (see Rupp et al. *Sci Rep* 7:737 (2017)). The application of both nanoparticle formulation thus enables lymphoma targeting by providing the anti-CD19 cargo, while simultaneously boosting efficacy by knocking out the PD-1 checkpoint inhibitor. In some embodiments, cells may be treated with the nanoparticles simultaneously. In some embodiments, the cells may be treated with the nanoparticles in separate steps, e.g., first deliver the RNP for generating the PD-1 knockout, and subsequently treat cells with the nanoparticles carrying the anti-CD19 CAR. In some embodiments, the second component of the system that improves T cell efficacy may result in the knockout of PD-1, TCR, CTLA-4, HLA-I, HLA-II, CS1, CD52, B2M, MHC-I, MHC-II, CD3, FAS, PDC1, CISH, TRAC, or a combination thereof. In some embodiments, knockdown of PD-1, TCR, CTLA-4, HLA-I, HLA-II, CS1, CD52, B2M, MHC-I, MHC-II, CD3, FAS, PDC1, CISH, or TRAC may be preferred, e.g., using siRNA targeting PD-1. In some embodiments, siRNA targeting PD-1 may be achieved using self-delivering RNAi as described

by Ligtenberg et al. *Mol Ther* 26(6):1482-1493 (2018) and in WO2010033247, incorporated herein by reference in its entirety, in which extensive chemical modifications of siRNAs, conferring the resulting hydrophobically modified siRNA molecules the ability to penetrate all cell types *ex vivo* and *in vivo* and achieve long-lasting specific target gene knockdown without any additional delivery formulations or techniques. In some embodiments, one or more components of the system may be delivered by other methods, e.g., electroporation. In some embodiments, additional regulators are knocked in to the cells for overexpression to control T cell- and NK cell-mediated immune responses and macrophage engulfment, e.g., PD-L1, HLA-G, CD47 (Han et al. *PNAS* 116(21):10441-10446 (2019)). Knock-in may be accomplished through application of an additional Gene Writing system with a template carrying an expression cassette for one or more such factors (3) with targeting to a safe harbor locus, e.g., AAVS1, e.g., using gRNA GGGGCCACTAGGGACAGGAT (SEQ ID NO: 1689) to target the Gene Writer polypeptide to AAVS1.

LNPs are used to treat primary T cells activated by Dynabeads at a 1:1 CD4⁺:CD8⁺ ratio at 450 ng/ μ L total mRNA concentrations. The resulting T cell populations are analyzed for integration, expression, and effect. For assessing integration, ddPCR is used with primers producing an amplicon extending from within the integrated CAR to the flanking genomic TRAC sequence. Comparing signal to a reference gene (e.g., RPP30), allows quantification of the average copy number per genome and integration efficiency. To analyze expression, flow cytometry with immunological probes is used to assess the level and percent of cells displaying surface CAR expression. To analyze activity of the CAR-T cells, treated cells are assessed via a co-plated cancer cell killing assay. By engineering Nalm6 ALL cells to express luciferase, cancer cell killing can be assessed by change in luminescence after co-culture with CAR-T cells as compared to signal from Nalm6 cells alone Billingsley et al. *Nano Lett* 20(3):1578-1589 (2020). Thus, a Gene Writing system can be used to generate CAR-T cells *ex vivo* with the desired cytotoxic activity.

Example 28: Gene Writers for integration of a CAR in T-cells in vivo

This example describes a Gene Writer™ genome editing system delivered to T-cells in vivo for integration and stable expression of a genetic payload. Specifically, targeted

nanoparticles are used to deliver a Gene Writing system capable of integrating a chimeric antigen receptor (CAR) expression cassette into the murine Rosa26 locus to generate CAR-T cells in a murine model.

In this example, a Gene Writing system comprises a Gene Writing polypeptide, e.g., a nickase Cas9 and R2Tg reverse transcriptase domain, as described herein, a gRNA for directing nickase activity to the target locus, and a template RNA comprising, from 5' to 3':

- (1) 100 nt homology to target site 3' of first strand nick
- (2) 5' UTR from R2Tg
- (3) Heterologous object sequence
- (4) 3' UTR from R2Tg
- (5) 100 nt homology to target site 5' of first strand nick

Wherein (3) comprises the coding sequence for the CD19-specific m194-1BBz CAR driven by the EF1a promoter (Smith et al. *Nat Nanotechnol* 12(8):813-820 (2017)). The Gene Writer in this example is guided to the murine Rosa26 locus using a gRNA, e.g.,

ACTCCAGTCTTTCTAGAAGA (SEQ ID NO: 1693) (Chu et al. *Nat Biotechnol* 33(5):543-548 (2015)). Production of RNA molecules is as according to examples provided herein, e.g., by in vitro transcription (e.g., Gene Writer polypeptide mRNA, template RNA) and by chemical synthesis (e.g., gRNA). Modifications to the RNA components of the system are as described elsewhere. For Gene Writer mRNA, the sequence additionally comprises a 5' UTR (e.g., GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC (SEQ ID NO: 1603)) and a 3' UTR (e.g.,

UGAUAUAAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGA (SEQ ID NO: 1604)) flanking the coding sequence. This combination of 5' UTR and 3' UTR has been shown to result in good expression of an operably linked ORF (Richner et al. *Cell* 168(6): P1114-1125 (2017)).

In order to achieve delivery specifically to T-cells, targeted LNPs (tLNPs) are generated that carry a conjugated mAb against CD4. See, e.g., Ramishetti et al. *ACS Nano* 9(7):6706-6716 (2015). Alternatively, conjugating a mAb against CD3 can be used to target both CD4⁺ and CD8⁺ T-cells (Smith et al. *Nat Nanotechnol* 12(8):813-820 (2017)). In other embodiments, the

nanoparticle used to deliver to T-cells in vivo is a constrained nanoparticle that lacks a targeting ligand, as taught by Lokugamage et al. *Adv Mater* 31(41):e1902251 (2019).

The tLNP can be made by first preparing the nucleic acid mix (e.g., polypeptide mRNA:gRNA:template RNA molar ratio of 1:40:40) with a mixture of lipids (cholesterol, DSPC, PEG-DMG, Dlin-MC3-DMA, and DSPE-PEG-maleimide) and then chemically conjugating the desired DTT-reduced mAb (e.g., anti-CD4, e.g., clone YTS.177) to the maleimide functional group on the LNPs. See Ramishetti et al. *ACS Nano* 9(7):6706-6716 (2015).

Six to 8 week old C57BL6/J mice are injected intravenously with formulated LNP at a dose of 1 mg RNA/kg body weight. Blood is collected at one day and three days post-administration in heparin-coated collection tubes, and the leukocytes are isolated by density centrifugation using Ficoll-Paque PLUS (GE Healthcare). Five days post-administration, animals are euthanized and blood and organs (spleen, lymph nodes, bone marrow cells) are harvested for T-cell analysis. Expression of the anti-CD19 CAR is detected by FACS using specific immunological sorting. Positive cells are confirmed for integration by ddPCR on the sorted population, where primers are used that flank an integration junction, e.g., one primer of the pair annealing to the integrated cargo and the other to genomic DNA from the Rosa26 target site.

Example 29: Assessment of distance and PAM orientation between the first and second nicks to reduce non-templated indel formation during Gene Writing

This examples describes how the placement of a second nick used in a Gene Writing system can be optimized to (1) increase the frequency of installation of a desired edit using a Gene Writer polypeptide with a template RNA, while (2) decreasing undesired insertions and/or deletions that may arise as a byproduct of the second nick.

An exemplary Gene Writing system can install a desired genomic modification (e.g., an insertion, deletion, or point mutation) using 1) a template RNA that comprises a gRNA and a heterologous object sequence comprising the desired genomic modification, and 2) a Gene Writing polypeptide comprising a nickase Cas9 (e.g., Cas9 N863A) fused to a reverse transcriptase (RT) (e.g., an RT domain from MMLV). In said exemplary Gene Writing system, the Cas9-RT fusion introduces a first nick, which exposes an available 3'OH to initiate the

reverse transcriptase reaction using the template RNA as a template for target primed reverse transcription. The placement of a second nick adjacent to, but on the opposite strand as the first nick, enhances the installation of the desired genome modification.

In this experiment, a 3 nt insertion (CTT) is directed to the HEK3 locus. The template RNA for the insertion comprises (1) a gRNA spacer with homology to the HEK3 site, (2) a gRNA scaffold, (3) a heterologous object sequence including the CTT insertion, and (4) a 3' target homology region for annealing to the genomic DNA immediately upstream of the first strand nick to set up target-primed reverse transcription of the heterologous object sequence. The sequence of the template RNA used is (5'-3')

GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCTGCCATCA<AAG>CGTGCTCAGTCTG (SEQ ID NO: 1694), where "<>" is used to denote the insertion sequence.

In addition, a set of second nick gRNAs, targeting a nick to the opposite DNA strand as the first nick, were designed that place a second nick either upstream or downstream of the location of the desired CTT insertion at various distances ranging from 26 to 257 bp. The upstream second nick creates a set of nicks with an inward orientation, with the PAM sites out (PAM-out), while the downstream second nick creates a set of nicks with an outward orientation, with the PAM sites inside the nicks (PAM-in), as described herein. Second nick gRNAs were designed using a web-based tool and are listed in Tables E5 and E6. The distance between dual nicks indicates the distance between the first nick directed by the template RNA and the second nick directed by the second nick gRNA, and the PAM orientation (e.g., "PAM-in" and thus outward orientation, or "PAM-out" and thus inward orientation) is provided with respect to the first nick as depicted in Figure 31.

Table E5: gRNA targeting the second nick upstream of the first nick in "PAM-out" orientation

PAM orientation to first nick	Orientation	Distance between dual nicks (nts)	sgRNA Sequence	PAM Sequence
out	antisense	28	TGGGCCCAAGGATTGACCC (SEQ ID NO: 1695)	AGG
out	antisense	33	CCCAAGGATTGACCCAGGCC (SEQ ID NO: 1696)	AGG

out	antisense	34	CCAAGGATTGACCCAGGCCA (SEQ ID NO: 1697)	GGG
out	antisense	38	GGATTGACCCAGGCCAGGGC (SEQ ID NO: 1698)	TGG
out	antisense	108	GCAGAAATAGACTAATTGCA (SEQ ID NO: 1699)	TGG
out	antisense	109	CAGAAATAGACTAATTGCAT (SEQ ID NO: 1700)	GGG
out	antisense	120	TAATTGCATGGGCGTTTCCC (SEQ ID NO: 1701)	TGG
out	antisense	121	AATTGCATGGGCGTTTCCCT (SEQ ID NO: 1702)	GGG
out	antisense	136	TCCCTGGGATCCCTGTCTCC (SEQ ID NO: 1703)	AGG
out	antisense	161	TCTCTCATCCATGCCTTTCT (SEQ ID NO: 1704)	AGG
out	antisense	197	CCCTTGCTTAAACTCTCCA (SEQ ID NO: 1705)	AGG
out	antisense	222	TCTCATGCCAAGCTCCCTGC (SEQ ID NO: 1706)	AGG
out	antisense	232	AGCTCCCTGCAGGACATCCC (SEQ ID NO: 1707)	AGG
out	antisense	240	GCAGGACATCCCAGGCCCTC (SEQ ID NO: 1708)	TGG
out	antisense	241	CAGGACATCCCAGGCCCTCT (SEQ ID NO: 1709)	GGG
out	antisense	255	CCCTCTGGGACAGCAGCTCA (SEQ ID NO: 1710)	CGG
out	antisense	256	CCTCTGGGACAGCAGCTCAC (SEQ ID NO: 1711)	GGG

Table E6: : gRNA targeting the second nick downstream of the first nick in “PAM-in”
orientation

PAM orientation to first nick	Orientation	Distance between dual nicks (nts)	sgRNA Sequence	PAM Sequence
In	antisense	26	GACGCCCTCTGGAGGAAGCA (SEQ ID NO: 1712)	GGG
In	antisense	27	CGACGCCCTCTGGAGGAAGC (SEQ ID NO: 1713)	AGG

In	antisense	34	TGTCCTGCGACGCCCTCTGG (SEQ ID NO: 1714)	AGG
In	antisense	37	AGCTGTCCTGCGACGCCCTC (SEQ ID NO: 1715)	TGG
In	antisense	63	GCACATACTAGCCCCTGTCT (SEQ ID NO: 1716)	AGG
In	antisense	90	GTCAACCAGTATCCCGGTGC (SEQ ID NO: 1717)	AGG
In	antisense	96	AAACTTGTCAACCAGTATCC (SEQ ID NO: 1718)	CGG
In	antisense	133	CCAGGGACCTCCCTAGGTGC (SEQ ID NO: 1719)	TGG
In	antisense	139	CCCCTCCAGGGACCTCCCT (SEQ ID NO: 1720)	AGG
In	antisense	150	GGTGAGGCTGGCCCCTTCCA (SEQ ID NO: 1721)	GGG
In	antisense	151	TGGTGAGGCTGGCCCCTTCC (SEQ ID NO: 1722)	AGG
In	antisense	162	CCCTCCTCTCCTGGTGAGGC (SEQ ID NO: 1723)	TGG
In	antisense	166	AGGTCCTCCTCTCCTGGTG (SEQ ID NO: 1724)	AGG
In	antisense	171	GGGCCAGGTCCTCCTCTCC (SEQ ID NO: 1725)	TGG
In	antisense	186	GAGCTCGACCCTGAAGGGCC (SEQ ID NO: 1726)	AGG
In	antisense	191	CTGTTGAGCTCGACCCTGAA (SEQ ID NO: 1727)	GGG
In	antisense	192	TCTGTTGAGCTCGACCCTGA (SEQ ID NO: 1728)	AGG
In	antisense	233	GCTGAAAGCCACTGGGCTCT (SEQ ID NO: 1729)	GGG
In	antisense	234	TGCTGAAAGCCACTGGGCTC (SEQ ID NO: 1730)	TGG
In	antisense	240	TGCAGGTGCTGAAAGCCACT (SEQ ID NO: 1731)	GGG
In	antisense	241	ATGCAGGTGCTGAAAGCCAC (SEQ ID NO: 1732)	TGG
In	antisense	257	TTGATCTCTGATTTTCATGC (SEQ ID NO: 1733)	AGG

To conduct the experiment, 200,000 U2OS cells in 20 μ L SE buffer are nucleofected with 800 ng of plasmid encoding the Gene Writer polypeptide (N863ACas9-RT), 200 ng of

template RNA, and 83 ng of a second nick gRNA listed in Tables E5 and E6. The Lonza Amaxa nucleofection system is used with the nucleofection code DN100. After nucleofection, 80 μ L of DMEM+10%FBS medium are added to the cell suspension and the cells are plate in a 24 well plate with 500 μ L of DMEM+10%FBS. Genomic DNA is extracted at day 3 post-nucleofection.

To analyze extracted DNA for the desired CTT insertion, amplicon sequencing is performed as described herein by amplifying the HEK locus using primers surrounding the first nick. The anticipated 300-350 bp amplicon is then sequenced on an Illumina MiSeq. The frequency of the desired CCT insertions is determined using the CRISPResso computational analysis pipeline (Clement et al. Nat Biotechnol 37(3):224-226 (2019)).

To measure undesired insertions and/or deletions arising as byproducts of the reaction, long-range amplification is performed with primers located >1.5 kb upstream and downstream of the first nick site, producing an amplicon >3 kb. This amplicon is sequenced using long-read sequencing (e.g., PacBio) and analyzed for the presence of insertions and deletions resulting from the dual nicking.

In some embodiments, a reaction using a second nick gRNA that cuts downstream of the first nick and provides a “PAM in” or outward orientation results in fewer unintended mutations (e.g., mutations in the target site other than the targeted CTT insertion) as compared to gRNAs placed upstream of the first nick at a similar distance but providing a “PAM-out” or inward orientation, as measured by the methods described herein. In other embodiments, a second nick gRNA that cuts upstream of the first nick and provides a “PAM-out” or inward orientation results in fewer undesired mutations (e.g., mutations in the target site other than the targeted CTT insertion) when the distance between the first and second nick is at least 100 nt as compared to a second nick gRNA providing a distance between the first and second nick of less than 100 nt, as measured by the methods described herein.

Thus, in some embodiments, a preferred design for a second nick gRNA is one resulting in 1) a “PAM-in” or outward orientation, or 2) a “PAM-out” or inward orientation with at least 100 nt separation between the first and second nicks (Figure 32).

Example 30: Design and human cell expression of Gene Writing systems utilizing various Cas-RT fusions

This example describes the construction and expression of Gene Writing polypeptides comprising fusions of Cas and reverse transcriptase domains in mammalian cells. Gene Writing polypeptides with these domains have been shown herein to enable the precise, site-specific modification of a DNA target from an RNA template molecule. Here, we describe the expression of a library of domains to create novel systems that may have diverse functional characteristics. More specifically, described here are fusion proteins comprising 1) a Cas-nuclease containing a mutation inactivating one endonuclease active site, e.g., the Cas9 nickase Cas9(N863A); 2) a peptide-linker to connect the functional protein domains, e.g., a sequence from Table 38 or 42, e.g., SGGSSGGSSGSETPGTSESATPESSGGSSGGSS; and 3) a reverse transcriptase (RT), e.g., an RT domain described in this application, e.g., an RT domain comprising a sequence from Table 1, Table 3, Table 30, Table 31, or Table 41, or a derivative thereof may be used in such an assay, collectively referred to in this Example as Cas-RT. Accordingly, Cas-RT fusion proteins are assembled on a plasmid and co-delivered with a single guide RNA (sgRNA) expression plasmid to validate system expression in human cells.

Gene Writer polypeptides generated by Cas-RT domain fusions assayed here comprised: (1) a Cas9 wild-type or Cas9(N863A) nickase domain; (2) a peptide linker (SGGSSGGSSGSETPGTSESATPESSGGSSGGSS); (3) a selection of RT domains from Table 1 and Table 30 taken from diverse sources; and (4) at least one nuclear localization signal. U2OS or HEK293T cells were transfected by Lonza Amaxa nucleofection of 250,000 cells/well with ~800 ng of Cas9(N863A)-RT fusion plasmid with 200 ng of a sgRNA plasmid. To assess the expression level of Cas9-RT fusions, cell lysates were collected on day 2 post-transfection and analyzed by Western blot using a primary antibody against Cas9. Several Cas9-RT fusions showed appreciable protein expression (Figure 33), suggestive of expression levels sufficient for Gene Writing activity. Notably, a wide range of expression levels is observed for the different Cas9-RT constructs, demonstrating the impact of the fusion design and RT selection on expression level of Cas-RT in cells.

Example 31: Improvement of expression of Cas-RT fusions through linker selection

This example demonstrates the optimization of Cas-RT fusions to improve protein expression in mammalian cells. As described in Example 30, construction of novel Cas-RT fusions by the simple substitution of new functional domains may result in low or moderate expression of the

Gene Writer polypeptide. Thus, it is contemplated here that modified configurations of the fusion may be advantageous in the context of different domains. Without wishing to be limited by the example, one such approach for improving the expression and stability of new fusions is through the use of a linker library. Here, the peptide linker sequence between the Cas and RT domains of the Cas-RT fusion is varied using a library of linker sequences. More specifically, linkers from Table 42 were used to generate new variants of a Cas9-RT fusion construct previously demonstrating low protein expression (see Example 30 and Figure 33) and delivered to human cells to screen for improved Cas-RT protein expression.

A set of 22 peptide linkers (Table 42) with varying degrees of length, flexibility, hydrophobicity, and secondary structure was first used to generate variants of a Cas-RT fusion protein by substitution of the original linker (see Example 30). HEK293T cells were transfected by electroporation of 250,000 cells/well with ~800 ng of each Cas9-RT fusion plasmid along with 200 ng of a single-guide RNA plasmid. To assess the expression level of Cas9-RT fusions, cell lysates were collected on day 2 post-transfection and analyzed by Western blot using a primary antibody against Cas9. linker 10 listed in Table 38 significantly improved Cas-RT fusion expression (Figure 34), demonstrating the potentially profound impact of the peptide linker sequence on Cas-RT expression.

Table 42. Peptide sequences used as linkers between the Cas and RT domains in Gene Writer polypeptides comprising Cas-RT fusions

#	Linker sequence	Notes
1	GGS	Short
2	GGGGS	Flexible, short
3	(GGGGS) ₂	Flexible
4	(GGGGS) ₃	Flexible, long
5	(GGGGS) ₄	Flexible, very long

6	(G) ₆	Flexible
7	(G) ₈	Flexible
8	GSAGSAAGSGEF	Flexible
9	(GSSGSS)	Mid
10	(GSSGSS) ₂	Mid, Flexible
11	(GSSGSS) ₃	Mid
12	SGSETPGTSESATPES	XTEN
13	(EAAAK)	Rigid helix, short
14	(EAAAK) ₂	Rigid helix, mid
15	(EAAAK) ₃	Rigid helix, long
16	PAP	Rigid, short
17	PAPAP	Rigid, short
18	PAPAPAPAP	Rigid, mid
19	A(EAAAK) ₄ ALEA(EAAAK) ₄ A	Rigid, very long with helices
20	GGGGS(EAAAK)GGGGS	Flexible – helix – flex
21	(EAAAK)GGGGS(EAAAK)	Helix – flex – helix
22	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	Flexible – XTEN – flexible

Example 32: Cas-mediated cleavage activity of Gene Writers comprising Cas-RT fusions

This example demonstrates the ability of Cas-RT fusions to retain functionality of the protein domains. Specifically, by assaying cells treated with Gene Writer polypeptides comprising a cleavage-competent Cas domain (cleavase), DNA binding can be read by target site analysis to demonstrate activity of Cas in the context of the fusions. Here, such Cas-RT cleavase fusions in which both nuclease active sites are functional, e.g., Cas9(wild-type)-RT, were co-delivered on plasmid vectors along with a sgRNA-expression plasmid to target the Cas to the AAVS1 site in human cells. Analysis of indel formation at the predicted cleavage site in AAVS1 by Cas-RT cleavase fusions functioned as a readout of both DNA binding activity and endonuclease activity, thereby confirming effective DNA targeting by the Cas-RT fusions.

Cas-RT fusions with fully functional endonuclease domains, e.g., comprising wild-type Cas9 with both nuclease active sites intact, e.g., Cas9(N863), were generated from Cas-RT fusion proteins described herein, e.g., comprising a Cas9 nickase, e.g., Cas9(N863A), in order to increase the sensitivity of detection of DNA binding and cleavage. Since the intact Cas9 nuclease can cut both strands to generate a double-stranded cleavage event in the genome, repair of these sites generates a higher mutation (indel) signal than repair of a single-stranded DNA nick. Thus, the frequency of indel formation of the fusions was compared to that of unfused, wild-type Cas9 in order to assess the maintenance of Cas functionality when placed in the context of the novel Cas-RT fusions.

U2OS or HEK293T cells were transfected by Lonza Amaxa nucleofection of 250,000 cells/well with ~800 ng of Cas9(WT)-RT fusion plasmid along with 200 ng of a sgRNA plasmid to produce the gRNA targeting Cas9 to AAVS1 (Table 43 gRNA P7). To assess the DNA binding and cleavage activity of Cas9-RT cleavase fusions, genomic DNA (gDNA) was collected on day 3 post-transfection. Indel patterns in the gDNA were analyzed by amplicon sequencing at loci targeted by the sgRNA. Sequencing results were analyzed by the CRISPResso2 pipeline (Clement et al Nat Biotechnol 37(3):224-226 (2019)). All tested Cas-RT cleavase fusions showed indel formation commensurate to their respective protein expression levels (Figure 33), indicating that Cas-mediated DNA binding activity is retained in Cas-RT fusions (Figure 35).

Example 33: Gene Writers comprising Cas-RT fusions with various RT domains enable precise editing in human cells

This example demonstrates the ability of multiple tested Cas-RT fusions to programmably install mutations in genomic DNA in human cells. More specifically, the reverse transcriptase domain of Cas-RT fusions, e.g., an RT domain described in this application, was varied to determine the genome editing capacity of Cas-RT fusions employing novel RT combinations. Template RNAs were co-delivered on plasmid vectors along with Cas-RT expression plasmids in human cells to determine the Rewriting activity of Cas-RT fusions.

In order to generate domain libraries for Gene Writer polypeptides, Cas effector proteins were selected; see in Table 37 and Table 40A. Additional Cas9 domains were further selected for use in the Gene Writer polypeptides described herein, as features including PAM requirements of a target sequence, predicted mutations for conferring nickase activity (e.g., D10A, H840A, or N863A for SpCas9), and gRNA features including single-guide composition, e.g., specific spacer parameters and gRNA scaffold sequence for conferring polypeptide binding for the cognate Cas enzyme, were able to be determined (Table 40A). Linker sequences to connect Cas and RT domains were collected based on a search for diversity of length, flexibility, and composition in order to optimize fusion proteins (Tables 38 and 42). Optimization of fusion expression by linker screening is further described in Example 31. Reverse transcriptase domains were mined from a variety of sources using literature and RT protein domain signatures as described in this application, including from non-LTR retrotransposons, LTR retrotransposons, group II introns, diversity-generating elements, retrons, telomerases, retroplasmids, retroviruses, and polymerases with evolved RNA-dependent DNA polymerase activity (e.g., an RT domain comprising a sequence from Table 1, Table 3, Table 30, Table 31, Table 44, or Table N41, or a derivative thereof may be used in such an assay).

Specifically, to assess the use of novel RT domains in the context of a Gene Writer polypeptide to successfully edit the genome, a subset of exemplary RT domains from retroviruses was selected for fusion to a Cas9(N863A) nickase. Briefly, a database of POL proteins from Retroviridae was first generated and then prioritized (see The UniProt Consortium Nucleic Acids Res 47(D1):D506-D515 (2019); Mitchell et al. Nucleic Acids Res 47(D1):D351-

D360 (2019)). Though not wishing to be limited by such example, retroviral RTs from the genera *Betaretrovirus*, *Deltaretrovirus*, *Gammaretrovirus*, and *Spumavirus* may function as monomeric proteins (see, for example, Table 1 from Herschhorn et al Cell Mol Life Sci 67(16):2717-2747 (2010)) and thus may be advantageous for use in a fusion protein, as described herein. A selection of retroviral monomeric RT sequences emerging from the analysis with these criteria is shown in Table 44. Further, mutations that have been shown to stabilize RT domains, as described in this application and in the literature (Table 45) (Anzalone et al Nat Biotechnol 38(7):824-844 (2020); Baranauskas et al Protein Eng Des Sel 25(10):657-668 (2012); Arezi and Hogrefe Nucleic Acids Res 37(2):473-481 (2009); Yasukawa et al J Biotechnol 150(3):299-306 (2010); the findings of which as they relate to improving RT stability and function are incorporated herein in their entirety), were analyzed for application to candidate RT domains (positions provided here based on the MMLV RT amino acid sequence as reference). As examples, MMLV RT with the mutational profile L139P/D200N/T330P/L603W/E607K showed an approximately 65-fold increase in processivity and 48-fold increase in template affinity (Baranauskas et al Protein Eng Des Sel 25(10):657-668 (2012)) and increased efficiency of prime editing of genomic DNA by a range of 1.6-5.1-fold with mutational profile D200N/T306K/W313F/T330P/L603W (Anzalone et al Nat Biotechnol 38(7):824-844 (2020)). From these studies, the core set of D200N/T330P/L603W was identified and an alignment of RT domains from the retroviral genera described here was used to predict the relevant amino acid positions where conserved (Figure 36 A). The additional mutations T306K and/or W313F were also applied where relevant and L139P and/or E607K was used when neither mutation of the T306K/W313F set was able to be applied (Figure 36 B). Cas9 nickase fusions with these wild-type RT domains or mutational variants with potentially improved activity were generated and exemplary fusions are described in Table 46.

To generate precise edits using Gene Writer Cas-RT fusions, Template RNAs were constructed to template reverse transcription of an edit into the genomic target site by the RT domain. Template RNAs were designed to comprise (i) a gRNA spacer sequence for guiding the Cas-RT to the target region, e.g., a sequence complementary to a 20-nucleotide sequence in the HEK3 locus; (ii) a primer-binding sequence capable of complementary base pairing with a single strand of the nicked DNA for target-primed reverse transcription; (iii) a heterologous object sequence providing a template for reverse transcription that further comprises the intended final

target sequence; and (iv) a gRNA scaffold sequence to associate with the Cas9 domain of the Cas9-RT polypeptide fusion. The constructs employed here specifically followed the 5' to 3' orientation (i), (iv), (iii), (ii). Template RNAs encoded on plasmids were cloned such that expression was driven by the U6 promoter and transcription termination controlled by a 7 nt polyT stretch following the primer-binding sequence at the 3' end of the Template RNA cassette. Template compositions are described in Table 43 (Templates P1, P2, P3).

U2OS or HEK293T cells were transfected by electroporation of 250,000 cells/well with ~800 ng of Cas9-RT(MMLV) fusion expression plasmid, 200 ng of a Template RNA expression plasmid, and 83 ng of an additional second-nick gRNA (2gRNA P5) expression plasmid (Table 43). To assess the genome editing capacity of Cas-RT fusions, genomic DNA (gDNA) was collected on day 3 post-transfection. The frequency of intended (exact and scarless edit as designed) versus unintended (any non-intended changes to the target sequence) edits ("Activity ratio") at target loci were analyzed by amplicon sequencing. As used herein, amplicon sequencing of a target site comprises the use of site-specific primers in PCR amplification of the target site, sequencing of amplicons on an Illumina MiSeq, and detection and characterization of editing events using the CRISPResso2 pipeline (Clement et al Nat Biotechnol 37(3):224-226 (2019)). Several Cas-RT fusions showed appreciable genome editing activity, with multiple Cas-RT fusions having Activity Ratios of ~3 (Figure 37), demonstrating that various Cas-RT fusions drawing from reverse transcriptase domains described herein can efficiently and precisely encode edits into the human genome.

Table 43. List of Template RNA and gRNA used in select examples.

Name	Description	spacer	scaffold	RT+ins	PBS	Template RNA
Template P1	HEK3_8PBS_10RT(CTTa t1)	GGCCCAGACTGA GCACGTGA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	TCTGCCATCAA AG	CGTG CTCA	GGCCCAGACTGAG CACGTGAGTTT TAG AGCTAGAAATAGC AAGTTAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAGT GGCACCAGATCGGT GCTCTGCCATCAAA GCGTGCTCA

Template P2	HEK3_13PB S_10RT(CTT at1)	GGCCCAGACTGA GCACGTGA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	TCTGCCATCAA AG	CGTG CTCA GTCT G	GGCCCAGACTGAG CACGTGAGTTTTAG AGCTAGAAATAGC AAGTTAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAGT GGCACCGAGTCGGT GCTCTGCCATCAAA GCGTGCTCAGTCTG
Template P3	HEK3_17PB S_10RT(CTT at1)	GGCCCAGACTGA GCACGTGA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	TCTGCCATCAA AG	CGTG CTCA GTCT GGGC C	GGCCCAGACTGAG CACGTGAGTTTTAG AGCTAGAAATAGC AAGTTAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAGT GGCACCGAGTCGGT GCTCTGCCATCAAA GCGTGCTCAGTCTG GGCC
Template P4	HBB_13PBS _10RT(TtoAa t4)	GCATGGTGCACC TGACTCCTG	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	AGACTTCTCCA CAG	GAGT CAGG TGCA C	GCATGGTGCACCTG ACTCCTGGTTTTAG AGCTAGAAATAGC AAGTTAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAGT GGCACCGAGTCGGT GCAGACTTCTCCAC AGGAGTCAGGTGC AC
2gRNA P5	HEK3_+90	GTCAACCAGTAT CCCGGTGC	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	NA	NA	NA
2gRNA P6	HBB_+72	GCCTTGATACCA ACCTGCCCCA	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	NA	NA	NA

gRNA P7	g19_AAVS1	GTCCCCTCCACC CCACAGTG	GTTTTAGA GCTAGAAA TAGCAAGT TAAAATAA GGCTAGTC CGTTATCA ACTTGAAA AAGTGGCA CCGAGTCG GTGC	NA	NA	NA
------------	-----------	--------------------------	--	----	----	----

Example 34: Multiplexing of a Gene Writer system to simultaneously edit multiple loci in a human cell

This example demonstrates the use of a Gene Writer system to edit multiple sites in the genome. In some applications, it may be of high value to be able to engineer multiple locations in the genome, e.g., to correct multiple genetic mutations or to optimize an engineered cell for cell therapy by performing multiple simultaneous modifications *ex vivo* or *in vivo*. In this example, a 3-plasmid system is utilized comprising: 1) a Gene Writer polypeptide expression plasmid, e.g., a plasmid encoding a Cas9 nickase fused to a reverse transcriptase (Cas-RT); 2) a Template plasmid, e.g., a plasmid encoding an expression cassette for a Template RNA that determines the genome site and the edit to instill at that site; and 3) a second-nick gRNA expression plasmid, e.g., a plasmid encoding an additional gRNA sequence to direct a second-strand nick for Cas9 at a location proximal to the target site.

In this example, two genome loci, the HBB gene and the human HEK3 locus, were targeted using gRNA comprising spacer sequences with identity to these sites to determine the ability to target multiple loci in parallel. To assess targeting of either locus separately or both simultaneously, cells were treated with different compositions of the Template plasmids to enable targeting of: 1) HEK3 alone, 2) HBB alone, or 3) both HBB and the HEK3 locus. Specifically, 800 ng of plasmid encoding the Cas9-RT(MMLV) fusion (Table 46), 200 ng of plasmid encoding the HEK3-modifying Template (Template P2, Table 43) and/or plasmid encoding the HBB-modifying Template (Template P4, Table 43), and 83 ng of plasmid encoding the HEK3 second-nick gRNA (2gRNA P5, Table 43) and/or plasmid encoding the HBB second-nick gRNA (2gRNA P6, Table 43) were nucleofected using nucleofection program DS_150 into HEK293T cells. After nucleofection, cells were grown at 37°C, 5% CO₂ for 3 days prior to cell lysis and genomic DNA extraction. Primers specific to each locus were used to amplify the

region and amplicons were sequenced using an Illumina MiSeq. Perfect correction and indel rates were analyzed using the CRISPResso2 pipeline (Clement et al Nat Biotechnol 37(3):224-226 (2019)) to determine Gene Writing efficacy. Table 51 lists the components of the Gene Writer System used in this example.

Table 51.

Name	Description	spacer	scaffold	RT+ins	PBS	Template RNA
Template P1	HEK3_8P BS_1 0RT(CTTa t1)	GGCCCA GACTGA GCACGT GA	GTTTGTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC	TCTGC CATCA AAG	CGTGCTC A	GGCCCAGACTGAGCACGT GAGTTTTAGAGCTAGAAA TAGCAAGTTAAAATAAGG CTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTC GGTGCTCTGCCATCAAAG CGTGCTCA
Template P2	HEK3_13 PBS_10RT (CTTat1)	GGCCCA GACTGA GCACGT GA	GTTTGTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC	TCTGC CATCA AAG	CGTGCTC AGTCTG	GGCCCAGACTGAGCACGT GAGTTTTAGAGCTAGAAA TAGCAAGTTAAAATAAGG CTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTC GGTGCTCTGCCATCAAAG CGTGCTCAGTCTG
Template P3	HEK3_17 PBS_10RT (CTTat1)	GGCCCA GACTGA GCACGT GA	GTTTGTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC	TCTGC CATCA AAG	CGTGCTC AGTCTG GGCC	GGCCCAGACTGAGCACGT GAGTTTTAGAGCTAGAAA TAGCAAGTTAAAATAAGG CTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTC GGTGCTCTGCCATCAAAG CGTGCTCAGTCTGGGCC
Template P4	HBB_13P BS_1 0RT(CTG	GCATGG TGCACC TGACTC CTG	GTTTGTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC	AGAC TTCTC CACA G	GAGTCA GGTGCA C	GCATGGTGCACCTGACTC CTGGTTTTAGAGCTAGAA ATAGCAAGTTAAAATAAG GCTAGTCCGTTATCAACTT

	TtoA at4)		GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC			GAAAAAGTGGCACCGAGT CGGTGCAGACTTCTCCAC AGGAGTCAGGTGCAC
2gRNA P5	HEK 3_+9 0	GTCAAC CAGTAT CCCGGT GC	GTTTTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC	NA	NA	NA
2gRNA P6	HBB _+72	GCCTTG ATACCA ACCTGC CCA	GTTTTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC	NA	NA	NA
gRNA P7	g19_ AAV S1	GTCCCC TCCACC CCACAG TG	GTTTTAGAGCT AGAAATAGCA AGTTAAAATA AGGCTAGTCC GTTATCAACTT GAAAAAGTGG CACCGAGTCG GTGC	NA	NA	NA

When tested independently, both targets saw a high degree of precise correction, with approximately 36% editing in HEK3 and 23% editing in HBB (Figure 38). Further, when targeted at the same time, approximately 34% editing of HEK3 and 14% editing of HBB target sites was achieved with precise correction conferred by the respective Template RNAs. Additionally, insertions and deletions were observed with low frequency in all conditions, with indels for each locus reaching similar levels when tested alone or in combination. Though not the express intent of this example, the lack of increase in indels during simultaneous editing is a

positive indicator for the potential to increase the number of loci targetable in parallel without compromising the precision of each individual edit.

Example 35: Delivery of DNA-free Gene Writer systems through nucleofection of human cells

This example describes the application of a Gene Writer system to edit the genome in human cells via delivery of RNA components, e.g., mRNA encoding the Gene Writer polypeptide and an RNA template. Without wishing to be bound by theory, the ability to deliver only RNA components in the absence of DNA is expected to confer major advantages to this system, including a reduction in immunogenicity and cellular toxicity linked to the detection of DNA in the cytoplasm and the availability of lipid nanoparticles systems described herein, the majority of which are optimized for RNA delivery, that can circumvent issues associated with viral delivery of nucleic acid therapeutics (e.g., manufacturing challenges, pre-existing immunity, immunogenic response to viral proteins). The reduction in cellular toxicity through use of an RNA system may be especially important for the modification of more sensitive cell types, such as primary cells. Further, nucleofection may be an effective method of delivering these systems to a patient's cells, e.g., for ex vivo cell engineering. Thus, it is of significant value to demonstrate the capacity of a Gene Writing system to function appropriately when delivered as all RNA and in the absence of DNA. Specifically, this example demonstrates delivery of an all-RNA Gene Writing system to modify the genome of HEK293T cells. To demonstrate RNA-based Gene Writing is not limited to a single composition, two versions of a Cas-RT fusion polypeptide are employed that comprise an RT domain derived from either Moloney murine leukemia virus (Cas9-RT(MMLV)) or porcine endogenous retrovirus (Cas9-RT(PERV)) (Table 46).

Gene Writer polypeptide-encoding mRNAs (1) were generated using T7 polymerase-driven in vitro transcription. In general, plasmids encoding the mRNA constructs comprised a transcriptional cassette comprising the following components: T7 promoter, 5'UTR, Gene Writer coding sequence (Cas9 nickase fused with a reverse transcriptase by a peptide linker and further comprising a nuclear localization signal), 3'UTR, and an 80 nt polyA tail. In this example, RNA molecules were prepared using unmodified nucleotides from linearized plasmid template. The

mRNAs encoding Cas9-RT(MMLV) or Cas9-RT(PERV) (Table 47) were co-transcriptionally capped with CleanCap AG (TriLink BioTechnologies).

Gene Writer Template RNAs (2) encoding genomic edits were generated by chemical synthesis and purified by standard desalting. The first and last three bases of each Template RNA comprised 2'-O-methyl phosphorothioate modifications. Template RNAs of varying length were designed to introduce different mutations into the human HEK3 locus (Table 48).

Where indicated, second nick gRNAs (3) were generated by chemical synthesis and comprised the following sequence modifications:

mG*mC*mA*rGrArArArUrArGrArCrUrArArUrUrGrCrArGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCmU*mU*mU*rU.

To assay the RNA Gene Writing systems described herein, HEK293T cells were plated 2 days before nucleofection to obtain 70-80% confluency on the day of nucleofection. RNAs were mixed according to the following combinations: i) Cas9-RT mRNA (1) only; ii) Cas9-RT mRNA (1), template RNA (2), and second nick gRNA (3); or iii) Cas9-RT mRNA (1) and template RNA (2). RNA mixes comprised 4.5 µg of the Cas9-RT mRNA (1), 5 µM final concentration of template RNA (2), and 1.3 µM final concentration of second nick gRNA (3). Mixes were nucleofected into approximately 200,000 cells using the Lonza Amaxa Nucleofector 96 Well Shuttle System, as according to manufacturer's protocols. Cells were then lysed and genomic DNA was collected 72 hours after nucleofection. Amplicon sequencing libraries were prepared using primers to amplify across the target site and Illumina sequencing was performed. Precise correction and indel rates were analyzed using the CRISPResso2 pipeline (Clement et al Nat Biotechnol 37(3):224-226 (2019)).

In these experiments, approximately 20% precise Writing activity was achieved with Cas9-RT(MMLV) using Template 1 (Table 48). A drop in activity was observed for templates that were longer than 120 nt in length; Template 4, which encoded the same edit as Template 1, but with an addition of 20 nt at the 3' end of the RT template, showed an approximately 3.1-fold drop in precise Writing activity and an approximately 2.4-fold drop in the ratio of precise corrections to indels (Figure 39). The use of Cas9-RT-encoding mRNAs with different UTRs and capping approaches produced similar levels of activity, though there was a slight increase with mRNA-5 (Table 49; Figure 40). The all-RNA nucleofection of the Gene Writer Cas9-

RT(PERV) with Template 1 and the second-nick gRNA further resulted in a precise Writing efficiency of approximately 7% (Figure 41). Across the experiments of this example, the addition of the second-nick gRNA resulted in an increase in Writing activity.

Table 48 provides sequences of Template RNA molecules used in all-RNA Gene Writing Examples. The spacer sequence of each Template RNA described here was kept constant and comprised 20 nt (5'-GGCCCAGACTGAGCACGTGA-3') of 100% identity to a target site in the human HEK3 locus (also known as LINC01509) (sequence maps to NC_000009.12:107422339..107422358, assembly GRCh38.p13). A Template RNA will typically comprise the components shown in the table, such that spacer+scaffold+RT+edit+PBS+Tail can yield the complete molecule.

Table 48. Template RNAs used in various Examples disclosed herein

Name	Description	spacer	scaffold	RT+edit	PBS	Tail	TemplateRNA Combined	RNA Sequence	Length
Templ ate 1	HEK3_13P BS_10RT(C TTat1)	GGCC CAGA CTGA GCAC GTGA	GTTTTAGAG CTAGAAAT AGCAAGTT AAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACCG AGTCGGTG C	TCTGCCA TCAAAAG	CGT GCT CA GTC TG	TTTT	GGCCCAGACTGAG CACGTGAGTTTTAG AGCTAGAAAATAGC AAGTTAAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAAGT GGCACCGAGTCGGT GCTCTGCCATCAAA GCGTGCTCAGTCTG TTTT	mG*mG*mC*rCrArGrArCrUr GrArGrCrArCrGrUrGrArGrUrUr UrUrArGrArGrCrUrArGrArArAr UrArGrCrArGrUrUrArArArAr UrArArGrCrUrArGrUrCrCrGr UrArUrCrArArCrUrUrGrArAr ArArArGrUrGrCrArCrCrGrAr GrUrCrGrUrGrCrUrCrUrGrCrC rArUrCrArArGrCrGrUrGrCrUr CrArGrUrCrUrGrUr*mU*mU	126
Templ ate 2	HEK3_13P BS_10RT(1 0nts at1)	GGCC CAGA CTGA GCAC GTGA	GTTTTAGAG CTAGAAAT AGCAAGTT AAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACCG AGTCGGTG C	TCTGCCA TCACATG TAGTTG	CGT GCT CA GTC TG	TTTT	GGCCCAGACTGAG CACGTGAGTTTTAG AGCTAGAAAATAGC AAGTTAAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAAGT GGCACCGAGTCGGT GCTCTGCCATCACA TGTAGTTGCCGTGCT CAGTCTGTTTT	mG*mG*mC*rCrArGrArCrUr GrArGrCrArCrGrUrGrArGrUrUr UrUrArGrArCrUrArGrArArAr UrArGrCrArArGrUrUrArArAr UrArArGrGrCrUrArGrUrCrCrGr UrUrUrCrArArCrUrUrGrArAr ArArArGrUrGrCrArCrCrGrAr GrUrCrGrUrGrCrUrCrUrGrCrC rArUrCrArCrUrGrUrArGrUrUr GrCrGrUrGrCrUrCrArGrUrCrUrG rU*mU*mU*mU	133
Templ ate 3	HEK3_13P BS_10RT(2 0nts at1)	GGCC CAGA CTGA GCAC GTGA	GTTTTAGAG CTAGAAAT AGCAAGTT AAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACCG AGTCGGTG C	TCTGCCA TCACATG TAGTTGA GGTCAA TGA	CGT GCT CA GTC TG	TTTT	GGCCCAGACTGAG CACGTGAGTTTTAG AGCTAGAAAATAGC AAGTTAAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAAGT GGCACCGAGTCGGT GCTCTGCCATCACA TGTAGTTGAGGTCA ATGACGTGCTCAGT CTGTTTT	mG*mG*mC*rCrArGrArCrUr GrArGrCrArCrGrUrGrArGrUrUr UrUrArGrArCrUrArGrArArAr UrArGrCrArArGrUrUrArArAr UrArArGrGrCrUrArGrUrCrCrGr UrUrUrCrArArCrUrUrGrArAr ArArArGrUrGrCrArCrCrGrAr GrUrCrGrUrGrCrUrCrUrGrCrC rArUrCrArCrUrGrUrArGrUrUr GrArGrUrCrArArUrGrArCrGr UrGrCrUrCrArGrUrCrUrGrUr*mU *mU*mU	143

Templ ate 4	HEK3_13P BS_30RT/C TTat1)	GGCC CAGA CTGA GCAC GTGA	GTTTATAGAG CTAGAAAT AGCAAGTT AAAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACC AGTCGGTG C	GGAAGC AGGGCT TCCTTTC CTCTGCC ATCAAA G	CGT GCT CA GTC TG	TTTT	GGCCAGACTGAG CACGTGAGTTTATAG AGCTAGAAATAGC AAGTTAAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAGT GGCACCGAGTCGGT GCGGAAGCAGGGC TTCCTTTCTCTGCCC ATCAAAAGCGTGCTC AGTCTGTTTT	mG*mG*mC*CrArGrArCrUr GrArGrCrArCrGrUrGrArGrUrUr UrUrArGrArGrCrUrArGrArArAr UrArGrCrArGrGrUrUrArArArAr UrArGrGrCrUrArGrUrCrGr UrUrArUrCrArArCrUrUrGrArAr ArArArGrUrGrCrArCrCrGrAr GrUrCrGrUrGrCrGrGrArArGr CrArGrGrCrUrUrCrCrUrUrUrC rCrUrCrUrGrCrArUrCrArCrAr UrGrUrArGrUrUrGrGrUrGrCr UrCrArGrUrCrUrGrUr*mU*mU* mU	146
Templ ate 5	HEK3_13P BS_30RT(1 Onts at1)	GGCC CAGA CTGA GCAC GTGA	GTTTATAGAG CTAGAAAT AGCAAGTT AAAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACC AGTCGGTG C	GGAAGC AGGGCT TCCTTTC CTCTGCC ATCACAT GTAGTTG	CGT GCT CA GTC TG	TTTT	GGCCAGACTGAG CACGTGAGTTTATAG AGCTAGAAATAGC AAGTTAAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAGT GGCACCGAGTCGGT GCGGAAGCAGGGC TTCCTTTCTCTGCCC ATCACATGTAGTTG CGTGCTCAGTCTGT TTT	mG*mG*mC*CrArGrArCrUr GrArGrCrArCrGrUrGrArGrUrUr UrUrArGrArGrCrUrArGrArArAr UrArGrCrArArGrUrUrArArArAr UrArGrGrCrUrArGrUrCrGr UrUrArUrCrArArCrUrUrGrArAr ArArArGrUrGrCrArCrCrGrAr GrUrCrGrUrGrCrGrGrArArGr CrArGrGrCrUrUrCrCrUrUrUrC rCrUrCrUrGrCrArUrCrArCrAr UrGrUrArGrUrUrGrGrUrGrCr UrCrArGrUrCrUrGrUr*mU*mU* mU	153
Templ ate 6	HEK3_13P BS_30RT(2 Onts at1)	GGCC CAGA CTGA GCAC GTGA	GTTTATAGAG CTAGAAAT AGCAAGTT AAAAATAAG GCTAGTCCG TTATCAACT TGAAAAAG TGGCACC AGTCGGTG C	GGAAGC AGGGCT TCCTTTC CTCTGCC ATCACAT GTAGTTG AGGTCA ATGA	CGT GCT CA GTC TG	TTTT	GGCCAGACTGAG CACGTGAGTTTATAG AGCTAGAAATAGC AAGTTAAAAATAAG GCTAGTCCGTTATC AACTTGAAAAAGT GGCACCGAGTCGGT GCGGAAGCAGGGC TTCCTTTCTCTGCCC ATCACATGTAGTTG AGGTCAATGACGTG CTCAGTCTGTTTT	mG*mG*mC*CrArGrArCrUr GrArGrCrArCrGrUrGrArGrUrUr UrUrArGrArGrCrUrArGrArArAr UrArGrCrArArGrUrUrArArArAr UrArGrGrCrUrArGrUrCrGr UrUrArUrCrArArCrUrUrGrArAr ArArArGrUrGrCrArCrCrGrAr GrUrCrGrUrGrCrGrGrArArGr CrArGrGrCrUrUrCrCrUrUrUrC rCrUrCrUrGrCrArUrCrArCrAr UrGrUrArGrUrUrGrGrUrGrCr ArArUrGrArCrGrUrGrUrCrAr GrUrCrUrGrUr*mU*mU*mU	163

Table. 49 Different production and compositions of Gene Writer polypeptide mRNAs used in various Examples

Name	Transcription template	Capping	Poly A	Modified NTPs	5' UTR	3' UTR
mRNA-1	PCR amplicon	CleanCap (AG); co-transcriptional	Added during amplification	None	AGGAAA TAAGAG AGAAAA GAAGAG TAAGAA GAAATA TAAGAG CCACC	GCTGGAGCCTCG GTGGCCATGCTTC TTGCCCCCTTGGGC CTCCCCCCAGCCC CTCCTCCCCCTTCC TGCACCCGTACCC CCGTGGTCTTTGA ATAAAGTCTGA
mRNA-2	PCR amplicon	CleanCap (AG); co-transcriptional	Added during amplification	5moU	AGGAAA TAAGAG AGAAAA GAAGAG TAAGAA GAAATA TAAGAG CCACC	GCTGGAGCCTCG GTGGCCATGCTTC TTGCCCCCTTGGGC CTCCCCCCAGCCC CTCCTCCCCCTTCC TGCACCCGTACCC CCGTGGTCTTTGA ATAAAGTCTGA
mRNA-3	PCR amplicon	Enzymatic, 2'O Methylated (Cap1); post-transcriptional	Added during amplification	None	GGGAAA TAAGAG AGAAAA GAAGAG TAAGAA GAAATA TAAGAG CCACC	GCTGGAGCCTCG GTGGCCATGCTTC TTGCCCCCTTGGGC CTCCCCCCAGCCC CTCCTCCCCCTTCC TGCACCCGTACCC CCGTGGTCTTTGA ATAAAGTCTGA

mRNA-4	PCR amplicon	Enzymatic, 2'-O Methylated (Cap1); post-transcriptional	Added during amplification	5moU	GGGAAA TAAGAG AGAAAA GAAGAG TAAGAA GAAATA TAAGAG CCACC	GCTGGAGCCTCG GTGGCCATGCTTC TTGCCCCCTGGGC CTCCCCCAGCCC CTCCTCCCCCTCC TGCACCCGTACCC CCGTGGTCTTTGA ATAAAGTCTGA
mRNA-5	Linearized plasmid	CleanCap (AG); co-transcriptional	Plasmid - encoded	None	AGGAAA TAAGAG AGAAAA GAAGAG TAAGAA GAAATA TAAGAG CCACC	GCTGCCTTCTGCG GGGCTTGCCTTCT GGCCATGCCCTTC TTCTCTCCCTTGC ACCTGTACCTCTT GGTCTTTGAATAA AGCCTGAGTAGG AAGTCTA

Example 36: Use of modified nucleotides in an all-RNA Gene Writer system

This example describes the application of a Gene Writer system to edit the genome in human cells via delivery of RNA components, e.g., mRNA encoding the Gene Writer polypeptide and an RNA template. Further to the demonstration of the DNA-free system in Example 35, this example describes the incorporation of modified nucleotides, e.g., 5-methoxyuridine, into the mRNA encoding the Gene Writer polypeptide, and the incorporation of modified nucleotides, e.g. 2'-O-methyl phosphorothioate, into the Gene Writer template RNA.

Gene Writer polypeptide-encoding mRNAs (1) were generated using T7 polymerase-driven in vitro transcription of an amplicon generated from a plasmid by PCR. The plasmid encoding the mRNA construct comprised a transcriptional cassette comprising the following components: T7 promoter, 5'UTR, Gene Writer coding sequence (Cas9 nickase fused with a reverse transcriptase by a peptide linker and further comprising a bipartite SV40 NLS), and a 3'UTR. A poly A tail component was added such that it was encoded in the amplicon serving as the template for RNA transcription. In this example, mRNA molecules were prepared by incorporating one modified nucleotide, 5-methoxyuridine (5moU), into the transcription reaction. The mRNA encoding Cas9-RT(MMLV) (Table 47) was capped either co-transcriptionally with CleanCap AG (TriLink BioTechnologies) or post-transcriptionally via enzymatic capping (2'O methylated, Cap1) (Table 49).

Gene Writer Template RNAs (2) encoding genomic edits were generated by chemical synthesis and purified by standard desalting. The first and last three bases of each Template RNA comprised 2'-O-methyl phosphorothioate modifications. Here, Template 1 was used to introduce a CTT insertion into the human HEK3 locus (Table 48).

Where indicated, second nick gRNAs (3) were generated by chemical synthesis and comprised the following sequence modifications:

mG*mC*mA*rGrArArArUrArGrArCrUrArArUrUrGrCrArGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCmU*mU*mU*rU.

To assay the RNA Gene Writing systems described herein, HEK293T cells were plated 2 days before nucleofection to obtain 70-80% confluency on the day of nucleofection. RNAs were

mixed according to the following combinations: i) Cas9-RT mRNA (1) only; ii) Cas9-RT mRNA (1), template RNA (2), and second nick gRNA (3); or iii) Cas9-RT mRNA (1) and template RNA (2). RNA mixes comprised 4.5 µg of the Cas9-RT mRNA (1), 5 µM final concentration of template RNA (2), and 1.3 µM final concentration of second nick gRNA (3). Mixes were nucleofected into approximately 200,000 cells using the Lonza Amaxa Nucleofector 96 Well Shuttle System, as according to manufacturer's protocols. Cells were then lysed and genomic DNA was collected 72 hours after nucleofection. Amplicon sequencing libraries were prepared using primers to amplify across the target site and Illumina sequencing was performed. Precise correction and indel rates were analyzed using the CRISPResso2 pipeline (Clement et al Nat Biotechnol 37(3):224-226 (2019)).

In these experiments, approximately 20% precise Writing activity was achieved using an all RNA Gene Writing system that incorporated modified nucleotides (5moU) in the mRNA encoding the Gene Writer polypeptide (Figure 42). Notably, the incorporation of the modified nucleotide 5moU did not result in an observable inhibitory effect on Writing efficiency. Similar efficiencies resulted from the mRNA capping methods assayed here (see Table 49). A slight decrease in efficiency was observed in the absence of a second nick gRNA (Figure 42).

Example 37: Delivery of DNA-free Gene Writer systems through lipid-based transfection of human cells

This example describes the application of a Gene Writer system to edit the genome in human cells via delivery of RNA components, e.g., mRNA encoding the Gene Writer polypeptide and an RNA template. Without wishing to be bound by theory, the ability to deliver only RNA components in the absence of DNA is expected to confer major advantages to this system, including a reduction in immunogenicity and cellular toxicity linked to the detection of DNA in the cytoplasm and the availability of lipid nanoparticles systems described herein, the majority of which are optimized for RNA delivery, that can circumvent issues associated with viral delivery of nucleic acid therapeutics (e.g., manufacturing challenges, pre-existing immunity, immunogenic response to viral proteins). The reduction in cellular toxicity through use of an RNA system may be especially important for the modification of more sensitive cell types, such as primary cells. Lipid transfection reagents may be utilized directly for ex vivo cell engineering and lipid-based nanoparticles are suitable for in vivo RNA delivery to a patient's

cells. Thus, it is of significant value to demonstrate the capacity of a Gene Writing system to function appropriately when delivered as all RNA and in the absence of DNA. Specifically, this example demonstrates delivery of an all RNA Gene Writing system to modify the genome of HEK293T cells using the lipid-based transfection reagents Lipofectamine 3000 and MessengerMAX (Invitrogen). To demonstrate RNA-based Gene Writing is not limited to a single composition, two versions of a Cas-RT fusion polypeptide are employed that comprise an RT domain derived from either Moloney murine leukemia virus (Cas9-RT(MMLV)) or porcine endogenous retrovirus (Cas9-RT(PERV)) (Table 46).

Gene Writer polypeptide-encoding mRNAs (1) were generated using T7 polymerase-driven in vitro transcription. In general, plasmids encoding the mRNA constructs comprised a transcriptional cassette comprising the following components: T7 promoter, 5'UTR, Gene Writer coding sequence (Cas9 nickase fused with a reverse transcriptase by a peptide linker and further comprising a nuclear localization signal), 3'UTR, and an 80 nt polyA tail. In this example, RNA molecules were prepared using unmodified nucleotides from either linearized plasmid template or using a PCR amplicon of the transcriptional cassette described above. The mRNA encoding Cas9-RT(MMLV) was capped either co-transcriptionally with CleanCap AG (TriLink BioTechnologies) or post-transcriptionally via enzymatic capping (2'O methylated, Cap1) (Table 49). The mRNA encoding Cas9-RT(PERV) was generated from plasmid template and co-transcriptionally capped with CleanCap AG (TriLink BioTechnologies) (Table 47).

Gene Writer Template RNAs (2) encoding genomic edits were generated by chemical synthesis and purified by standard desalting. The first and last three bases of each Template RNA comprised 2'-O-methyl phosphorothioate modifications. Here, Template 1 was used to introduce a CTT insertion into the human HEK3 locus (Table 48).

Where indicated, second nick gRNAs (3) were generated by chemical synthesis and comprised the following sequence modifications:

mG*mC*mA*rGrArArArUrArGrArCrUrArArUrUrGrCrArGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArGrUrGrGrCrArCrGrArGrUrCrGrGrUrGrCmU*mU*mU*rU.

To assay the RNA Gene Writing systems described herein, approximately 50,000 HEK293T cells were plated in 24-well plates 1 day before lipofection. RNAs were mixed according to the following combinations: i) Cas9-RT mRNA (1) only; ii) Cas9-RT mRNA (1),

template RNA (2), and second nick gRNA (3); or iii) Cas9-RT mRNA (1) and template RNA (2). RNA mixes comprised 0.45 µg of the Cas9-RT mRNA (1), 2.5 pM final concentration of template RNA (2), and 1.0 pM final concentration of second nick gRNA (3). RNAs were mixed with Opti-MEM media (Gibco) and Lipofectamine 3000 or MessengerMAX reagent (Invitrogen) and added to cells. Cells were then lysed and genomic DNA was collected 72 hours after nucleofection. Amplicon sequencing libraries were prepared using primers to amplify across the target site and Illumina sequencing was performed. Precise correction and indel rates were analyzed using the CRISPResso2 pipeline (Clement et al Nat Biotechnol 37(3):224-226 (2019)).

In these experiments, up to approximately 17% precise Writing activity was achieved using an all RNA Gene Writing system delivered by lipid-based transfection, approaching efficiencies similar to nucleofection (Figure 43 B; see Example 35 for nucleofection). Lipofectamine 3000 was also used (Figure 43 A). In contrast to nucleofection (Example 35), there was not an observable reduction when utilizing the 20 nt longer Template 4 as compared to Template 1 (Table 48; Figure 43 C). Further, the all-RNA lipofection of the Gene Writer Cas9-RT(PERV) with Template 1 resulted in precise Writing of the desired edit with an efficiency of approximately 3.5% (Figure 44).

Example 38: RNA Gene Writing enables DNA-free precise editing of primary T cells

This example describes the use of a Cas9-RT fusion polypeptide-based Gene Writer system for the genomic editing of target DNA sequences. More specifically, this example describes nucleofection of an all-RNA system into primary CD4+ T cells for Gene Rewriting in primary human cells, e.g., as a means of demonstrating the Gene Rewriter system for ex vivo application.

The all RNA system described here comprised: 1) Gene Writer polypeptide-encoding mRNA, e.g., an RNA encoding the Cas9-RT fusion polypeptide as a driver for programmed gene editing through a targeted nicking and reverse transcription process as described in this invention; 2) a template RNA molecule, e.g., an RNA comprising (i) a gRNA spacer sequence for guiding the driver to the targeted region, e.g., a sequence complementary to a 20-nucleotide sequence in the HEK3 locus; (ii) a primer-binding sequence capable of complementary base pairing with a single strand of the nicked DNA for target-primed reverse transcription; (iii) a heterologous object sequence providing a template for reverse transcription that further comprises the intended final target sequence; and (iv) a gRNA scaffold sequence to associate

with the Cas9 domain of the Cas9-RT polypeptide fusion; and 3) an optional additional gRNA to promote second-strand nicking near the target site, e.g., an RNA comprising (i) a spacer sequence for targeting the driver to induce a second nick, on the opposite strand of the first nick guided by the template RNA, at a site proximal to the target site (e.g., within 50-150 nt from the first nick); and (ii) a gRNA scaffold sequence mediating an association with the Cas9 domain of the driver. In this example, the Cas-RT fusion polypeptide (1) (Table 46) comprises a Cas9(N863A) nickase fused to an MMLV reverse transcriptase domain. The template RNAs (2) employed here specifically follow the 5' to 3' orientation (i), (iv), (iii), (ii), as listed in the description thereof and are detailed in Table 48 and Example 35.

To deliver the RNA Gene Writer system into primary human CD4+ T cells and validate protein expression, 1,000,000 cells (Human Peripheral Blood CD4+ T Cells, Lonza catalog #2W-200) were stimulated by CD3/CD28 for two days and then nucleofected with 0, 2.5, 5, or 10 µg of mRNA encoding the Cas-RT polypeptide using a Nucleofector 96-well Shuttle System (Lonza) with the EO-115 nucleofection program, as according to manufacturer's protocols. One day post-nucleofection, the efficiency of delivery was assessed by immunoblotting with a Cas9 antibody (Cell Signaling) to measure protein expression of the Gene Writer polypeptide from the nucleofected mRNA (Figure 45 A).

Subsequently, primary human CD4+ T cells were nucleofected with either: (1) 5 µg Gene Writer polypeptide mRNA (Writer only control); (2) 5 µg Gene Writer polypeptide mRNA and 5 µM template RNA, e.g., one of six template RNAs from Table 48 that target the same site of the HEK3 locus, but differ in editing result or design; or (3) 5 µg Gene Writer polypeptide mRNA, 5 µM template RNA, e.g., one of six template RNAs from Table 48, and 2.075 µM of an additional gRNA for generating a second-strand nick, e.g., the second-nick gRNA targeting a sequence 108 nt upstream of the HEK3 target site described in Example 35. Three days post-nucleofection, cells were harvested to examine 1) cell viability after RNA delivery of the Gene Writer system, and 2) editing efficiency on the target site of the genome. To assess the cell viability, the percentage of live cells was measured by flow cytometry after staining cells with a fluorescent live/dead dye (BioLegend). Cell viability was comparable in experimental conditions and in the absence of nucleofection (Untreated control) (Figure 45 B). To evaluate efficiency of editing by the Gene Writing systems, genomic DNA was analyzed by PCR-based amplicon sequencing assay, as described in Example 35. The efficiency of the desired editing (Perfect

Write) reached approximately 6.3% using Template 1 (Table 48) with the Gene Writer polypeptide mRNA (Figures 46 A and B). Here, the addition of a second-nick gRNA (Figure 46 B) resulted in similar levels of editing. Thus, this example demonstrates the use of Gene Writing systems for highly specific editing in primary T cells and further shows the successful application of DNA-free, all-RNA Gene Writing in these cells.

Example 39: Detection of Retrotransposase-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. 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 50, 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 48). 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 48).

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 48). 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 50 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 49. 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..

Example 40: 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 stable-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 41: 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 40).

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 42: 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 43: 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.

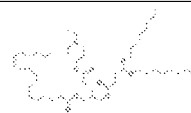
In this example, the SSTR pathway will be inhibited using siRNAs against the core components of the pathway: FANCA, FANCD2, FANCE, USP1. Control siRNAs of a non-target control will also be included. 200k U2OS cells will be nucleofected with 30pmols (1.5μM) siRNAs, as well as R2Tg driver and transgene 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.

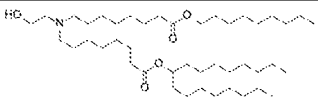
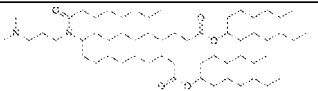
Example 44: Formulation of Lipid Nanoparticles encapsulating Firefly Luciferase mRNA

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 44 (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 spectrophotometer (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 45: 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% CO₂. 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% CO₂. 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 15,000g 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. 50A and B and

indicate successful LNP-mediated delivery of RNA into primary cells in culture. As shown in Fig. 50A, LNPs formulated as according to Example 44 were analyzed for delivery of cargo to primary human (A) and mouse (B) hepatocytes, as according to Example 45. 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 46: 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 44 and tested in vitro prior (Example 45) 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 μ L 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 PierceTM BCA Protein Assay Kit (ThermoFisher Scientific) according to the manufacturer's instructions. Luciferase activity was measured with 200 μ g (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. 51). As shown in FIG. 51, 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 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 the RT domain comprises a sequence of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

2. 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 the RT domain comprises a sequence of Table 1 or 3, or a sequence of a reverse transcriptase domain of Table 2,

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.

3. 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 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 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.

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 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 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.

5. The system of any of the preceding claims, 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.

6. 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 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.

7. 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 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 (a)(ii) and/or (a)(iii) comprises a TAL domain; a zinc finger domain; or a CRISPR/Cas domain chosen from Table 3 or a functional variant (e.g., mutant) thereof.

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 (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 the polypeptide, (iii) a heterologous object sequence, and (iv) a 3' target 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.

9. 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) a sequence that specifically binds the RT domain, (iii) a heterologous object sequence, and (iv) a 3' target homology domain.

10. The system of claim 9, wherein the template RNA further comprises a sequence that binds (a)(ii) and/or (a)(iii).

11. 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' target homology domain.

12. 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 second 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' target homology domain.

13. 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.
14. The template RNA of claim 13, 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).
15. The template RNA of either of claims 13 or 14, wherein the RT domain comprises a sequence selected of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2 or a sequence that has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.
16. The template RNA of either of claims 13 or 14, wherein the RT domain comprises a sequence selected of Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, 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.
17. The template RNA of any of claims 13-16, wherein the sequence of (ii) specifically binds the RT domain.
18. The template RNA of any of claims 13-17, wherein the sequence that specifically binds the RT domain is a sequence, e.g., a UTR sequence, of Table 1 or from a domain of Table 2, or a sequence having at least 70, 75, 80, 85, 90, 95, or 99% identity thereto.
19. 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.

20. 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,.

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

22. 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 second 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) heterologous object sequence, and (iii) a 3' target homology domain.

23. The system of claim 14, wherein the nucleic acid encoding the first template RNA and the nucleic acid encoding the second template RNA are two separate nucleic acids.

24. The system of claim 14, 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.

25. 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 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

26. 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 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, 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.

27. 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 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, 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.

28. 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 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, 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; 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.

29. The system of either of claims 27 or 28, wherein the nucleic acid encoding the first polypeptide and the nucleic acid encoding the second polypeptide are two separate nucleic acids.

30. The system of either of claims 27 or 28, 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.
31. A reaction mixture comprising:
a cell and any system, polypeptide, template RNA, or DNA encoding the same of any preceding claim.
32. A reaction mixture comprising:
a DNA comprising a target site and any system, polypeptide, template RNA, or DNA encoding the same of any preceding claim.
33. A kit comprising:
the system, polypeptide, template RNA, or DNA encoding the same of any preceding claim;
instructions for using the system, polypeptide, template RNA, or DNA encoding the same; and
one or both of a cell or a DNA comprising a target site.
34. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the DBD comprises a TAL domain.
35. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the DBD comprises a zinc finger domain.
36. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the DBD comprises a CRISPR/Cas domain.
37. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the endonuclease domain is a nickase domain.

38. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the endonuclease domain comprises a CRISPR/Cas domain.
39. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the CRISPR/Cas domain comprises a domain or polypeptide from Table 4, or a functional variant (e.g., mutant) thereof.
40. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the CRISPR/Cas domain comprises a domain or polypeptide from genus/species from Table 4.
41. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the endonuclease domain comprises a type II nuclelease (e.g., FokI), a Holliday Junction resolvase, or a double-stranded DNA nuclease comprising an alteration that abrogates its ability to cut one strand (e.g., transforming the double-stranded DNA nuclease into a nickase).
42. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the RT domain comprises a reverse transcriptase or functional fragment or variant thereof chosen from Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2.
43. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the RT domain comprises one or more mutations (e.g., an insertion, deletion, or substitution) relative to a naturally occurring RT domain or an RT domain or functional fragment chosen from Table 1 or 3 or a sequence of a reverse transcriptase domain of Table 2, e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 mutations.
44. The system, kit, polypeptide, or reaction mixture of claim 43, wherein the one or more mutations are chosen from D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K or D653N in the RT domain of murine leukemia virus reverse transcriptase or a corresponding mutation at a corresponding position of another RT domain.

45. The system, kit, polypeptide, or reaction mixture of claim 43, wherein the one or more mutations are chosen from WO2018089860A1, incorporated herein by reference (e.g., a C952S, and/or C956S, and/or C952S, C956S (double mutant), and/or C969S, and/or H970Y, and/or R979Q, and/or R976Q, and/or R1071S, and/or R328A, and/or R329A, and/or Q336A, and/or R328A, R329A, Q336A (triple mutant), and/or G426A, and/or D428A, and/or G426A,D428A (double mutant) mutation, and/or any combination thereof; positions relative to WO2018089860A1 SEQ ID NO: 52), in the RT domain of R2Bm retrotransposase or a corresponding mutation at a corresponding position of another RT domain.
46. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the RT domain is positioned C-terminal of the DBD in the polypeptide.
47. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the RT domain is positioned C-terminal of the nickase domain in the polypeptide.
48. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the RT domain is positioned N-terminal of the DBD in the polypeptide.
49. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the RT domain is positioned N-terminal of the nickase domain in the polypeptide.
50. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the polypeptide comprises a linker, e.g., positioned between the RT domain and the DBD or the RT domain and the nickase domain.
51. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the linker is between 2-50, e.g., 2-30, amino acids in length.
52. The system, kit, polypeptide, or reaction mixture of any preceding claim, wherein the linker is a flexible linker, e.g., comprising Gly and/or Ser residues.

53. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the 3' target homology domain is complementary to a sequence adjacent to a site to be modified by the system, 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.

54. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the 3' target homology domain is more than 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, 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 long, (e.g., 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, or 30 nucleotides long).

55. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the 3' target homology domain is no more than 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, 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 long.

56. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the heterologous object sequence is complementary to a site to be modified by the system except at the position or positions to be modified.

57. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the heterologous object sequence is complementary to a site to be modified by the system except at positions encoding a sequence to be inserted to the site.

58. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the heterologous object sequence is complementary to a site to be modified by the system except the heterologous object sequence does not comprise nucleotides encoding a sequence to be deleted at the site.

59. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the heterologous object sequence is more than 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, or 30 nucleotides long, (e.g., 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, or 30 nucleotides long).

60. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the heterologous object sequence is no more than 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, or 30 nucleotides long.

61. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the heterologous object sequence substitutes at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides for non-target site nucleotides.

62. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the heterologous object sequence inserts 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, or 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 into the target site.

63. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the heterologous object sequence deletes 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, 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides.

64. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the heterologous object sequence is separated from the sequence that binds the polypeptide (e.g., that binds the endonuclease domain and/or DBD domain) by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 nucleotides.

65. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the sequence that binds the polypeptide (e.g., that binds the endonuclease domain and/or DBD

domain) is at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, or 130 nucleotides long (and optionally no more than 150, 140, 130, 120, 110, 100, 90, 85, or 80 nucleotides long).

66. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the sequence that binds the polypeptide binds the endonuclease domain and/or DBD domain.

67. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the sequence that binds the polypeptide (e.g., that binds the endonuclease domain and/or DBD domain) comprises a gRNA.

68. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the sequence that binds a target site (e.g., a second strand of a site in a target genome) is at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, or 130 nucleotides long (and optionally no more than 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, or 20 nucleotides long), e.g., is 17, 18, 19, 20, 21, 22, 23, or 24 nucleotides long.

69. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the sequence that binds a target site is complementary to the second strand of the target site, or comprises no more than 1, 2, 3, 4, or 5 mismatches to a sequence complementary to the second strand of the target site.

70. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the sequence that binds a target site (e.g., a second strand of a site in a target genome) is separated from the sequence that binds the polypeptide (e.g., that binds the endonuclease domain and/or DBD domain) by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 nucleotides.

71. The system, kit, template RNA, or reaction mixture of any preceding claim, further comprising a second strand-targeting gRNA that directs the endonuclease domain (e.g., nickase) domain to nick the second strand (e.g., in the target genome)..

72. The system, kit, template RNA, or reaction mixture of claim 71, wherein the template RNA further comprises the second strand-targeting gRNA.
73. The system, kit, template RNA, or reaction mixture of claim 71, wherein the second strand-targeting gRNA is disposed on a separate nucleic acid from the template RNA.
74. The system, kit, template RNA, or reaction mixture of any of claims 71-73, wherein the gRNA directs the endonuclease domain (e.g., nickase) domain to nick the second strand (e.g., in the target genome) at a site that is 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 nick on the first strand).
75. The system, kit, template RNA, or reaction mixture of any of claims 71-74, wherein the gRNA specifically binds the first strand.
76. A system comprising:
an mRNA encoding the polypeptide or system of any preceding claim, and
a template RNA of any preceding claim.
77. The system of claim 76, wherein the mRNA encoding the polypeptide or system of any preceding claim and the template RNA of any preceding claim are disposed on different nucleic acid molecules.
78. A system comprising an RNA molecule comprising:
a template RNA (or RNA encoding the template RNA) of any preceding claim, and
a sequence encoding the system or polypeptide of any preceding claim.
79. The system of claim 78, wherein the RNA molecule comprises an internal ribosome entry site, e.g., operably linked to the sequence encoding the system or polypeptide.

80. The system of either of claims 78 or 79, wherein the RNA molecule comprises a cleavage site, e.g., situated between the template RNA (or RNA encoding the template RNA) and the sequence encoding the system or polypeptide.
81. The system or polypeptide of any preceding claim, wherein the polypeptide comprises a split intein, e.g., two or more (e.g., all) of the RT domain, DBD, endonuclease (e.g., nickase) domain, or combinations thereof are translated as separate proteins which combine into a single polypeptide by protein splicing.
82. A lipid nanoparticle (LNP) comprising the system, template RNA, polypeptide (or RNA encoding the same), or DNA encoding the system, template RNA, or polypeptide, of any preceding claim.
83. A system comprising:
a first lipid nanoparticle comprising the polypeptide (or DNA or RNA encoding the same) of a system or polypeptide of any preceding claim; and
a second lipid nanoparticle comprising the template RNA (or DNA encoding the same) of a system or template RNA of any preceding claim.
84. A virus, viral-like particle, fusosome, or virosome comprising the system, template RNA, polypeptide (or RNA encoding the same), or DNA encoding the system, template RNA, or polypeptide, of any preceding claim.
85. A system comprising:
a first virus, viral-like particle, fusosome, or virosome comprising the polypeptide (or DNA or RNA encoding the same) of a system or polypeptide of any preceding claim; and
a second virus, viral-like particle, or virosome comprising the template RNA (or DNA encoding the same) of a system or template RNA of any preceding claim.
86. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein 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, or at least 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 long (and optionally less than 15, 10, 5, or 20 kilobases long, or less than 500, 400, 300, or 200 nucleotides long).

87. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein 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, 40, 50, 60, 70, 80, 90, or 100 nucleotides in length).

88. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains:

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)), and/or

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.

89. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein the template RNA 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).

90. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein at least 80, 85, 90, 95, 96, 97, 98, or 99% of the template RNA present contains one or more modified nucleotides.

91. The system, kit, template RNA, or reaction mixture of any preceding claim, wherein 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, or at least 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 long) after a stability test.

92. The system, kit, or reaction mixture of any preceding claim, wherein at least 1 % of target sites are modified after the system is assayed for potency.

93. The system, kit, template RNA, polypeptide, or reaction mixture of any preceding claim, wherein the system, polypeptide, template RNA, and/or DNA encoding the same, is formulated as a lipid nanoparticle (LNP).

94. The system, kit, template RNA, polypeptide, or reaction mixture of any preceding claim, wherein the DNA encoding the system, polypeptide, and/or template RNA are packaged into a virus, viral-like particle, virosome, liposome, vesicle, exosome, or LNP.

95. The system, kit, template RNA, polypeptide, or reaction mixture of claim 94, wherein the DNA encoding the system, template RNA, or polypeptide is packaged into an adeno-associated virus (AAV).

96. The system, kit, template RNA, polypeptide, or reaction mixture of any preceding claim, wherein the system, template RNA, polypeptide, lipid nanoparticle (LNP), virus, viral-like

particle, or virosome is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, and/or host cell protein contamination.

97. A virus, viral-like particle, or virosome comprising:
the system, template RNA, or polypeptide of any preceding claim, or DNA encoding any of the same, and
an adeno-associated virus (AAV) capsid protein.

98. The system, kit, template RNA, polypeptide, virus, viral-like particle, or virosome of any preceding claim, wherein the system, template RNA, and/or polypeptide is active in a target tissue and less active (e.g., not active) in a non-target tissue.

99. The system, kit, template RNA, polypeptide, virus, viral-like particle, or virosome of claim 98, further comprising 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 the template RNA, the polypeptide or nucleic acid encoding the same, or both.

100. A method for manufacturing an template RNA, comprising:

(a) providing an template RNA of any preceding claim, and

(b) assaying one or more of:

(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, or 30 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 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 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; or

(vii) 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 template RNA is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, or host cell protein contamination.

101. A method for manufacturing a system for modifying DNA, comprising:

(a) providing a system for modifying DNA of any preceding claim, and

(b) assaying one or more of:

(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, or 30 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 a selected post-translational modification;

(ix) the presence, absence, and/or type of one or more artificial, synthetic, or non-canonical amino acids 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.

102. 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,

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

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

contacting the cell with a system, polypeptide, template RNA, or DNA encoding the same of any preceding claim,

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

104. 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,

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

wherein the cell is euploid, is not immortalized, is part of a tissue, is part of an organism, is a primary cell, is non-dividing, is haploid (e.g., a germline cell), is a non-cancerous polyploid cell, or is from a subject having a genetic disease.

105. A method for treating a subject having a disease or condition associated with a genetic defect, the method comprising:

administering to the subject:

(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,
thereby treating the subject having a disease or condition associated with a genetic defect.

106. A method for treating a subject having a disease or condition associated with a genetic defect, the method comprising:

administering to the subject a system, polypeptide, template RNA, or DNA encoding the same of any preceding claim,

thereby treating the subject having a disease or condition associated with a genetic defect.

107. The method of either of claims 105 or 106, wherein the disease or condition associated with a genetic defect is an indication listed in any of Tables 9-12, and/or wherein the genetic defect is a defect in a gene listed in any of Tables 9-12.

108. The method of any of claims 105-107, wherein the subject is a human patient.