

CO-DELIVERY OF A GENE EDITOR CONSTRUCT AND A DONOR TEMPLATE**1. CROSS-REFERENCE TO RELATED APPLICATION**

[0001] This application claims the benefit of U.S. Provisional Application No. 63/292,698, filed December 22, 2021; U.S. Provisional Application No. 63/318,343, filed March 9, 2022; and U.S. Provisional Application No. 63/355,235, filed on June 24, 2022, each of which is hereby incorporated in its entirety by reference.

2. SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing with 577, which has been submitted via Patent Center and is hereby incorporated by reference in its entirety. Said XML copy, created on December 22, 2022, is named 50699PCT-SequenceListing.xml, and is 780,344 bytes in size.

3. BACKGROUND

[0003] Programmable, efficient, and multiplexed genome integration of large, diverse DNA cargo independent of DNA repair remains an unsolved challenge of genome editing. Current gene integration approaches require double strand breaks that evoke DNA damage responses and rely on repair pathways that are inactive in terminally differentiated cells. Furthermore, CRISPR-based approaches that bypass double stranded breaks, such as Prime editing, are limited to modification or insertion of short sequences.

[0004] There is a need in the art for techniques which address and overcome these shortcomings and enable the co-delivery of gene editor constructs and associated donor templates for the insertion and/or deletion of large sequences into cells for therapeutic and circuit-based uses for broad purposes, across eukaryotic as well as prokaryotic systems.

4. SUMMARY

[0005] The present disclosure describes co-delivering (i.e., “dual delivery”) to a cell a (i) gene editor construct and a (ii) donor (i.e., “cargo” or “payload”) template that enables in vivo beacon placement and in vivo integration of a template polynucleotide. In typical embodiments, the gene editor construct is comprised of a polynucleotide sequence that encodes the gene editor construct. In typical embodiments, the gene editor construct, upon polynucleotide expression or direct delivery of the gene editor protein and associated guide

RNAs (gRNAs (e.g., atgRNA), can incorporate an integrase target recognition site (i.e., “beacon” or “landing pad”) or a recombinase target recognition site at a DNA locus. The gene editor polynucleotide construct is packaged within a lipid nanoparticle (LNP) that is capable of localizing the gene editor polynucleotide construct to a cell cytoplasm. The gene editor polynucleotide construct packaged in a LNP is co-delivered with a donor template (i.e., “cargo” or “payload”) polynucleotide construct packaged into a separate vector that is capable of localizing the donor template to a cell nucleus. In certain embodiments, the donor template vector is AAV, helper dependent adenovirus, or integration deficient lentivirus. In typical embodiments, the donor template is integrated into the genomic integrase target recognition site by an integrase, optionally by an integrase fused/linked to a gene editor protein. Also provided herein are methods using LNP mixtures, including a split LNP approach to deliver precise ratios of mRNA encoding the gene editor protein to atgRNAs. These ratios enable robust in vivo beacon placement in both neonatal and adult mice model systems.

[0006] The present disclosure provides a co-delivery platform for site-specific genetic engineering using Programmable Addition via Site-Specific Targeting Elements (PASTE) (see *Ionnidi et al.*; doi: 10.1101/2021.11.01.466786; the entirety of *Ionnidi et al.* is incorporated by reference), transposon-mediated gene editing, or other suitable gene editing or gene incorporation technology.

[0007] Described herein is a method of co-delivering (i.e., “dual delivery”) to a cell a (i) gene editor construct and a (ii) template polynucleotide (i.e., “cargo” or “payload”) . In typical embodiments, the gene editor construct is comprised of a polynucleotide sequence that encodes the gene editor construct. In typical embodiments, the gene editor construct, upon polynucleotide expression or direct delivery of the gene editor protein and associated guide RNAs, can incorporate an integrase target recognition site (i.e., “beacon” or “landing pad”) or a recombinase target recognition site at a DNA locus. The gene editor polynucleotide construct is packaged within a lipid nanoparticle (LNP) that is capable of localizing the gene editor polynucleotide construct to a cell cytoplasm. The gene editor can be packaged into the LNP as a protein along with associated guide RNAs and delivered to the cell cytoplasm or to cell nucleus. The gene editor polynucleotide construct packaged in a LNP is co-delivered with a donor template (i.e., “cargo” or “payload”) polynucleotide construct packaged into a separate vector that is capable of localizing the donor template to a cell nucleus. In certain embodiments, the donor template vector is AAV, helper dependent adenovirus, or integration deficient

lentivirus. In typical embodiments, the donor template is integrated into the genomic integrase target recognition site by an integrase, optionally by an integrase fused/linked to a gene editor protein.

[0008] The present disclosure provides a co-delivery platform for site-specific genetic engineering using Programmable Addition via Site-Specific Targeting Elements (PASTE) (see *Ionnidi et al.*; doi: 10.1101/2021.11.01.466786; U.S. Application No. 17/649,308; PCT Publication No. WO 2022/087235A; each of which is herein incorporated by reference in its entirety), transposon-mediated gene editing, or other suitable gene editing or gene incorporation technology.

[0009] In one aspect, this disclosure features a method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the method comprising:

delivering to a cell:

(a) a lipid nanoparticle (LNP) comprising:

(i) a gene editor polynucleotide; and

(b) a vector comprising:

(i) a template polynucleotide, and

(ii) at least a first attachment site-containing guide RNA (atgRNA).

[0010] In some embodiments, the gene editor polynucleotide is capable of localizing to a cell cytoplasm.

[0011] In some embodiments, the template polynucleotide is capable of localizing to a cell nucleus.

[0012] In some embodiments, the gene editor polynucleotide comprises: a polynucleotide sequence encoding a prime editor system.

[0013] In some embodiments, the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.

[0014] In some embodiments, the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.

[0015] In some embodiments, the nickase is linked to the reverse transcriptase by in-frame fusion. In some embodiments, the nickase is linked to the reverse transcriptase by a linker. In some embodiments, the linker is a peptide fused in-frame between the nickase and reverse transcriptase.

[0016] In some embodiments, the gene editor polynucleotide further comprises: a polynucleotide sequence encoding at least a first integrase.

[0017] In some embodiments, the linked nickase-reverse transcriptase are further linked to the first integrase.

[0018] In some embodiments, the method also includes co-delivering a second vector.

[0019] In some embodiments, the second vector comprises a polynucleotide sequence encoding at least a first integrase.

[0020] In some embodiments, the first integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.

[0021] In some embodiments, the gene editor polynucleotide further comprises a polynucleotide sequence encoding a recombinase. In some embodiments, the recombinase is FLP or Cre.

[0022] In some embodiments, the first atgRNA comprises: (i) a domain that is capable of guiding the prime editor system to a target sequence; and (ii) a reverse transcriptase (RT) template that comprises at least a portion of an at least first integration recognition site.

[0023] In some embodiments, the RT template comprises the entirety of the first integration recognition site.

[0024] In some embodiments, the vector further comprises a second atgRNA.

[0025] In some embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence; the first atgRNA further includes a first RT template that comprises at least a portion of an at least first integration recognition site; the second atgRNA further includes a second RT template that comprises at least a portion of

the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

[0026] In some embodiments, the vector further comprises a nicking gRNA.

[0027] In some embodiments, the LNPs further comprises a nicking gRNA.

[0028] In some embodiments, the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.

[0029] In some embodiments, the template polynucleotide comprises a second integration recognition site.

[0030] In some embodiments, the second integration recognition site is a cognate pair with the first integration recognition site.

[0031] In some embodiments, the template polynucleotide comprises at least a third integration recognition site.

[0032] In some embodiments, the template polynucleotide further comprises at least a fourth integration recognition site.

[0033] In some embodiments, the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.

[0034] In some embodiments, the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.

[0035] In some embodiments, the sub-sequence of the vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.

[0036] In some embodiments, the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.

[0037] In some embodiments, self-circularizing is mediated by recombination of the third integration recognition site and a fourth integration recognition site by the integrase.

[0038] In some embodiments, the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.

[0039] In some embodiments, the vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an anneovirus, a retrovirus, a DoggyboneTM DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid.

[0040] In some embodiments, the LNP and the vector are concurrently delivered.

[0041] In some embodiments, the LNP and the vector are delivered separately.

[0042] In some embodiments, the LNP and the vector are delivered at least 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 weeks apart.

[0043] In some embodiments, the cell is in vivo.

[0044] In another aspect, this disclosure features a method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the method comprising:

delivering to a cell:

(a) a lipid nanoparticle (LNP) comprising:

(i) a gene editor polynucleotide, and

(ii) a first attachment site-containing guide RNA (atgRNA); and

(b) a vector comprising:

(i) a template polynucleotide, and

(ii) a second atgRNA.

[0045] In another aspect, this disclosure features a method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the method comprising:

delivering:

(a) a lipid nanoparticle (LNP) comprising:

(i) a gene editor polynucleotide,

(ii) a first attachment site-containing guide RNA (atgRNA), and

(iii) a second atgRNA; and

- (b) a vector comprising:
 - (i) a template polynucleotide.

[0046] In another aspect, this disclosure features a method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the method comprising:

delivering:

- (a) a lipid nanoparticle (LNP) comprising:
 - (i) a gene editor polynucleotide, and
 - (ii) a first attachment site-containing guide RNA (atgRNA); and
- (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) a nicking atgRNA.

[0047] In some embodiments, the gene editor polynucleotide comprises:

[0048] a polynucleotide sequence encoding a prime editor system.

[0049] In some embodiments, the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.

[0050] In some embodiments, the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.

[0051] In some embodiments, the nickase is linked to the reverse transcriptase by in-frame fusion.

[0052] In some embodiments, the nickase is linked to the reverse transcriptase by a linker.

[0053] In some embodiments, the linker is a peptide fused in-frame between the nickase and reverse transcriptase.

[0054] In some embodiments, the gene editor polynucleotide construct further comprises: a polynucleotide sequence encoding at least a first integrase.

[0055] In some embodiments, the linked nickase-reverse transcriptase are further linked to the integrase.

[0056] In some embodiments, the method also includes delivering a second vector.

[0057] In some embodiments, the second vector comprises a polynucleotide sequence encoding at least a first integrase.

[0058] In some embodiments, the first integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.

[0059] In some embodiments, the gene editor polynucleotide construct further comprises a polynucleotide sequence encoding a recombinase.

[0060] In some embodiments, the recombinase is FLP or Cre.

[0061] In some embodiments, the first atgRNA comprises: (i) a domain that is capable of guiding the prime editor system to a target sequence; and (ii) a reverse transcriptase (RT) template that comprises at least a portion of an at least first integration recognition site.

[0062] In some embodiments, the RT template comprises the entirety of the first integration recognition site.

[0063] In some embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence; the first atgRNA further includes a first RT template that comprises at least a portion of the first integration recognition site; the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

[0064] In some embodiments, the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.

[0065] In some embodiments, the template polynucleotide comprises a second integration recognition site.

[0066] In some embodiments, the second integration recognition site is a cognate pair with the first integration recognition site.

[0067] In some embodiments, the template polynucleotide comprises at least a third integration recognition site.

[0068] In some embodiments, the template polynucleotide further comprises at least a fourth integration recognition site.

[0069] In some embodiments, the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.

[0070] In some embodiments, the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.

[0071] In some embodiments, the sub-sequence of the vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.

[0072] In some embodiments, the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.

[0073] In some embodiments, self-circularizing is mediated by recombination of the third integration recognition site and the fourth integration recognition site by the integrase.

[0074] In some embodiments, the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.

[0075] In some embodiments, the vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an anneovirus, a retrovirus, a Doggybone™ DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid.

[0076] In some embodiments, the LNP and the vector are concurrently delivered.

[0077] In some embodiments, the LNP and the vector are delivered separately.

[0078] In some embodiments, the LNP and the vector are delivered at least 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 weeks apart.

[0079] In some embodiments, the cell is in vivo.

[0080] In another aspect, this disclosure features a method of co-delivering a system capable of site-specifically integrating at least a first integration recognition site into the genome of a cell, the method comprising:

co-delivering to a cell:

(a) a first lipid nanoparticle (LNP) comprising:

(i) a first gene editor polynucleotide, and

(ii) a first attachment site-containing guide RNA (atgRNA); and

(b) a second lipid nanoparticle (LNP) comprising:

(i) a second gene editor polynucleotide, and

(ii) a second attachment site-containing guide RNA (atgRNA),

wherein the first atgRNA and the second atgRNA are an at least first pair of atgRNAs.

[0081] In some embodiments, the method also includes mixing the first LNP and the second LNP prior to co-delivering to the cell.

[0082] In some embodiments, the first LNP and the second LNP are mixed at a ratio of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.

[0083] In some embodiments, the first gene editor polynucleotide construct, the second gene editor polynucleotide construct, or both comprise: a polynucleotide sequence encoding a prime editor system.

[0084] In some embodiments, the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.

[0085] In some embodiments, the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.

[0086] In some embodiments, the nickase is linked to the reverse transcriptase by in-frame fusion.

[0087] In some embodiments, the nickase is linked to the reverse transcriptase by a linker.

[0088] In some embodiments, the linker is a peptide fused in-frame between the nickase and reverse transcriptase.

[0089] In some embodiments, the first gene editor polynucleotide, construct, the second gene editor polynucleotide construct, or both, further comprise:

[0090] a polynucleotide sequence encoding an integrase.

[0091] In some embodiments, the linked nickase-reverse transcriptase are further linked to the integrase.

[0092] In some embodiments, the first gene editor polynucleotide, the second gene editor polynucleotide, or both, further comprise: a polynucleotide sequence encoding a recombinase.

[0093] In some embodiments, the linked nickase-reverse transcriptase are further linked to the recombinase.

[0094] In some embodiments, the first gene editor polynucleotide and the second gene editor polynucleotide are the same.

[0095] In some embodiments, the first gene editor polynucleotide is mRNA, the second gene editor polynucleotide is mRNA, or both the first and second gene editor polynucleotides are mRNA.

[0096] In some embodiments, the first LNP comprises a ratio of mRNA to atgRNA of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.

[0097] In some embodiments, the second LNP comprises a ratio of mRNA to atgRNA of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.

[0098] In some embodiments, the method also includes delivering an integrase.

[0099] In some embodiments, delivering the integrase comprises co-delivering the integrase with (a) and (b).

[0100] In some embodiments, the method comprises delivering a polynucleotide sequence encoding the integrase.

[0101] In some embodiments, the polynucleotide sequence is encoded in a first vector.

[0102] In some embodiments, the first vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an annelovirus, a retrovirus, a Doggybone™ DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid..

[0103] In some embodiments, the first vector further comprises a template polynucleotide and a sequence that is an integration cognate with the first integration recognition site.

[0104] In some embodiments, the method also includes delivering a recombinase.

[0105] In some embodiments, delivering the recombinase comprises co-delivering the recombinase with (a) and (b).

[0106] In some embodiments, the method comprises delivering a polynucleotide sequence encoding the recombinase.

[0107] In some embodiments, the polynucleotide sequence is encoded in the first vector.

[0108] In some embodiments, the method also includes delivering a second vector.

[0109] In some embodiments, the second vector comprises a template polynucleotide and a sequence that is an integration cognate with the first integration recognition site.

[0110] In some embodiments, the second vector is a vector selected from: an adenovirus, an AAV, a lentivirus, an HSV, an annelovirus, a retrovirus, Doggybone™ DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid.

[0111] In some embodiments, the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.

[0112] In some embodiments, the template polynucleotide comprises a second integration recognition site.

[0113] In some embodiments, the second integration recognition site is a cognate pair with the first integration recognition site.

[0114] In some embodiments, the template polynucleotide comprises at least a third integration recognition site.

[0115] In some embodiments, the template polynucleotide further comprises at least a fourth integration recognition site.

[0116] In some embodiments, the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.

[0117] In some embodiments, the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.

[0118] In some embodiments, the sub-sequence of the vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.

[0119] In some embodiments, the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.

[0120] In some embodiments, self-circularizing is mediated by recombination of the third integration recognition site and a fourth integration recognition site by the integrase.

[0121] In some embodiments, the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.

[0122] In some embodiments, the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence; the first atgRNA further includes a first RT template that comprises at least a portion of a first integration recognition site; the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

[0123] In some embodiments, the first integration site is an AttB sequence, a FRT sequence, or a VOX sequence.

[0124] In some embodiments, the first atgRNA, the second atgRNA or both are synthetic.

[0125] In some embodiments, the integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.

[0126] In some embodiments, the cell is *in vivo*.

[0127] In another aspect, this disclosure features a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising:

- (a) a lipid nanoparticle (LNP) comprising:
 - (i) a gene editor polynucleotide construct; and
- (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) at least a first attachment site-containing guide RNA (atgRNA).

[0128] In some embodiments of the system, the gene editor polynucleotide construct comprises a polynucleotide sequence encoding a prime editor system.

[0129] In some embodiments of the system, the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.

[0130] In some embodiments of the system, the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase. In some embodiments of the system, the nickase is linked to the reverse transcriptase by in-frame fusion. In some embodiments of the system, the nickase is linked to the reverse transcriptase by a linker.

[0131] In some embodiments of the system, the linker is a peptide fused in-frame between the nickase and reverse transcriptase.

[0132] In some embodiments of the system, the gene editor polynucleotide construct further comprises: a polynucleotide sequence encoding at least a first integrase.

[0133] In some embodiments of the system, the linked nickase-reverse transcriptase are further linked to the first integrase.

[0134] In some embodiments of the system, the system also includes a second vector.

[0135] In some embodiments of the system, the second vector comprises a polynucleotide sequence encoding at least a first integrase. In some embodiments of the system, the first integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.

[0136] In some embodiments of the system, the gene editor polynucleotide construct further comprises a polynucleotide sequence encoding a recombinase. In some embodiments of the system, the recombinase is FLP or Cre.

[0137] In some embodiments of the system, the first atgRNA comprises: (i) a domain that is capable of guiding the prime editor system to a target sequence; and (ii) a reverse transcriptase (RT) template that comprises at least a portion of an at least first integration recognition site.

[0138] In some embodiments of the system, the RT template comprises the entirety of the first integration recognition site.

[0139] In some embodiments of the system, the vector further comprises a second atgRNA.

[0140] In some embodiments of the system, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence; the first atgRNA further includes a first RT template that comprises at least a portion of the first integration recognition site; the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

[0141] In some embodiments of the system, the vector further comprises a nicking gRNA.

[0142] In some embodiments of the system, the LNP further comprises a nicking gRNA.

[0143] In some embodiments of the system, the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.

[0144] In some embodiments of the system, the template polynucleotide comprises a second integration recognition site.

[0145] In some embodiments of the system, the second integration recognition site is a cognate pair with the first integration recognition site.

[0146] In some embodiments of the system, the template polynucleotide comprises at least a third integration recognition site.

[0147] In some embodiments of the system, the template polynucleotide construct further comprises at least a fourth integration recognition site.

[0148] In some embodiments of the system, the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.

[0149] In some embodiments of the system, the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.

[0150] In some embodiments of the system, the sub-sequence of vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.

[0151] In some embodiments of the system, the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.

[0152] In some embodiments of the system, self-circularizing is mediated by recombination of the third integration recognition site and the fourth integration recognition site by the integrase.

[0153] In some embodiments of the system, the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.

[0154] In some embodiments of the system, the vector is a recombinant adenovirus, a helper dependent adenovirus, or an adeno-associated virus.

[0155] In another aspect, this disclosure features a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising:

- (a) a lipid nanoparticle (LNP) comprising:
 - (i) a gene editor polynucleotide, and
 - (ii) a first attachment site-containing guide RNA (atgRNA); and
- (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) a second atgRNA.

[0156] In another aspect, this disclosure features a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising:

- (a) a lipid nanoparticle (LNP) comprising
 - (i) a gene editor polynucleotide,
 - (ii) a first attachment site-containing guide RNA (atgRNA), and
 - (iii) a second atgRNA; and
- (b) a vector comprising:
 - (i) a template polynucleotide.

[0157] In another aspect, this disclosure features a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising:

- (a) a lipid nanoparticle (LNP) comprising
 - (i) a gene editor polynucleotide, and
 - (ii) a first attachment site-containing guide RNA (atgRNA); and
- (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) a nicking gRNA.

[0158] In some embodiments of the system, the gene editor polynucleotide comprises: a polynucleotide sequence encoding a prime editor system.

[0159] In some embodiments of the system, the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.

[0160] In some embodiments of the system, the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase. In some embodiments of the system, the nickase is linked to the reverse transcriptase by in-frame fusion. In some embodiments of the system, the nickase is linked to the reverse transcriptase by a linker.

[0161] In some embodiments of the system, the linker is a peptide fused in-frame between the nickase and reverse transcriptase.

[0162] In some embodiments of the system, the gene editor polynucleotide further comprises:

[0163] a polynucleotide sequence encoding at least a first integrase.

[0164] In some embodiments of the system, the linked nickase-reverse transcriptase are further linked to the first integrase.

[0165] In some embodiments of the system, the system also includes a second vector.

[0166] In some embodiments of the system, the second vector comprises a polynucleotide sequence encoding at least a first integrase.

[0167] In some embodiments of the system, the first integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.

[0168] In some embodiments of the system, the gene editor polynucleotide further comprises a polynucleotide sequence encoding a recombinase.

[0169] In some embodiments of the system, the recombinase is FLP or Cre.

[0170] In some embodiments of the system, the first atgRNA comprises: (i) a domain that is capable of guiding the prime editor system to a target sequence; and (ii) a reverse transcriptase (RT) template that comprises at least a portion of an at least first integration recognition site.

[0171] In some embodiments of the system, the RT template comprises the entirety of the first integration recognition site.

[0172] In some embodiments of the system, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence; the first atgRNA further includes a first RT template that comprises at least a portion of the first integration recognition site; the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

[0173] In some embodiments of the system, the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.

[0174] In some embodiments of the system, the template polynucleotide comprises a second integration recognition site.

[0175] In some embodiments of the system, the second integration recognition site is a cognate pair with the first integration recognition site.

[0176] In some embodiments of the system, the template polynucleotide comprises at least a third integration recognition site.

[0177] In some embodiments of the system, the template polynucleotide construct further comprises at least a fourth integration recognition site.

[0178] In some embodiments of the system, the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.

[0179] In some embodiments of the system, the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.

[0180] In some embodiments of the system, the sub-sequence of vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.

[0181] In some embodiments of the system, the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.

[0182] In some embodiments of the system, self-circularizing is mediated by recombination of the third integration recognition site and the fourth integration recognition site by the integrase.

[0183] In some embodiments of the system, the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.

[0184] In some embodiments of the system, the vector is recombinant adenovirus, helper dependent adenovirus, or an adeno-associated virus.

[0185] In another aspect, this disclosure features a system capable of site-specifically integrating at least a first integration recognition site into the genome of a cell, the system comprising:

- (a) a first lipid nanoparticle (LNP) comprising:
 - (i) a first gene editor polynucleotide, and
 - (ii) a first attachment site-containing guide RNA (atgRNA); and
- (b) a second lipid nanoparticle (LNP) comprising:
 - (i) a second gene editor polynucleotide, and
 - (ii) a second attachment site-containing guide RNA (atgRNA).

[0186] In some embodiments of the system, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs.

[0187] In some embodiments of the system, the first LNP and the second LNP are mixed at a ratio of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.

[0188] In some embodiments of the system, the first gene editor polynucleotide, the second gene editor polynucleotide, or both comprise:

[0189] a polynucleotide sequence encoding a prime editor system.

[0190] In some embodiments of the system, the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.

[0191] In some embodiments of the system, the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.

[0192] In some embodiments of the system, the nickase is linked to the reverse transcriptase by in-frame fusion.

[0193] In some embodiments of the system, the nickase is linked to the reverse transcriptase by a linker.

[0194] In some embodiments of the system, the linker is a peptide fused in-frame between the nickase and reverse transcriptase.

[0195] In some embodiments of the system, the first gene editor polynucleotide, the second gene editor polynucleotide, or both, further comprise: a polynucleotide sequence encoding an integrase.

[0196] In some embodiments of the system, the linked nickase-reverse transcriptase are further linked to the integrase.

[0197] In some embodiments of the system, the first gene editor polynucleotide, the second gene editor polynucleotide, or both, further comprise: a polynucleotide sequence encoding a recombinase.

[0198] In some embodiments of the system, the nickase-reverse transcriptase are further linked to the recombinase.

[0199] In some embodiments of the system, the first gene editor polynucleotide and the second gene editor polynucleotide are the same.

[0200] In some embodiments of the system, the first gene editor polynucleotide is mRNA, the second gene editor polynucleotide is mRNA, or both the first and second gene editor polynucleotides are mRNA.

[0201] In some embodiments of the system, the first LNP comprises a ratio of mRNA to atgRNA of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.

[0202] In some embodiments of the system, the second LNP comprises a ratio of mRNA to atgRNA of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.

[0203] In some embodiments of the system, the system also includes an integrase.

[0204] In some embodiments of the system, the system comprises a polynucleotide sequence encoding the integrase.

[0205] In some embodiments of the system, the polynucleotide sequence is encoded in a first vector.

[0206] In some embodiments of the system, the first vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an anneovirus, a retrovirus, a DoggyboneTM

DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid.

[0207] In some embodiments of the system, the first vector further comprises a template polynucleotide and a sequence that is an integration cognate with the first integration recognition site.

[0208] In some embodiments of the system, the system also includes delivering a recombinase.

[0209] In some embodiments of the system, delivering the recombinase comprises co-delivering the recombinase with (a) and (b).

[0210] In some embodiments of the system, the system comprises delivering a polynucleotide sequence encoding the recombinase.

[0211] In some embodiments of the system, the polynucleotide sequence is encoded in the first vector.

[0212] In some embodiments of the system, the system also includes co-delivering a second vector.

[0213] In some embodiments of the system, the second vector comprises a template polynucleotide and a sequence that is an integration cognate with the first integration recognition site.

[0214] In some embodiments of the system, the second vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an annelovirus, a retrovirus, a Doggybone™ DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid.

[0215] In some embodiments of the system, the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.

[0216] In some embodiments of the system, the template polynucleotide comprises a second integration recognition site.

[0217] In some embodiments of the system, the second integration recognition site is a cognate pair with the first integration recognition site.

[0218] In some embodiments of the system, the template polynucleotide comprises at least a third integration recognition site.

[0219] In some embodiments of the system, the template polynucleotide further comprises at least a fourth integration recognition site.

[0220] In some embodiments of the system, the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.

[0221] In some embodiments of the system, the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.

[0222] In some embodiments of the system, the sub-sequence of the vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.

[0223] In some embodiments of the system, the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.

[0224] In some embodiments of the system, self-circularizing is mediated by recombination of the third integration recognition site and the fourth integration recognition site by the integrase.

[0225] In some embodiments of the system, the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.

[0226] In some embodiments of the system, the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence; the first atgRNA further includes a first RT template that comprises at least a portion of a first integration recognition site; the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

[0227] In some embodiments, the first integration site is an AttB sequence, a FRT sequence, or a VOX sequence.

[0228] In some embodiments of the system, the first atgRNA, the second atgRNA or both are synthetic.

[0229] In some embodiments of the system, the integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.

[0230] In some embodiments of the system, the recombinase is FLP or Cre.

[0231] In another aspect, this disclosure features a cell comprising any of the delivery systems or any of the co-delivery systems described herein.

[0232] In another aspect, this disclosure features a pharmaceutical composition comprising the any of the delivery systems described herein or any of the co-delivery systems described herein.

[0233] In another aspect, this disclosure features a method of treating a patient in need thereof, the method comprising administering an effective amount of any of the systems described herein, any of the cells described herein, or any of the pharmaceutical compositions described herein.

[0234] In another aspect, this disclosure features a method of treating a patient in need thereof, the method comprising: administering an effective amount of any of the LNPs described herein, any of the first vectors described herein, or any of the second vectors described herein as a first dose and an effective amount of any of the LNPs described herein, any of the first vectors described herein, or any of the second vectors described herein as a second dose.

[0235] In some embodiments, the first dose and the second dose are separately administered by multiple administrations.

[0236] In some embodiments, the first dose and the second dose are administered at least 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, or 7 days apart.

[0237] In some embodiments, the first dose and the second dose are administered at least 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 weeks apart.

5. BRIEF DESCRIPTION OF THE DRAWINGS

[0238] These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, and accompanying drawings, where:

[0239] **FIG. 1** shows a non-limiting illustration of a gene editor construct packaged within a lipid nanoparticle (LNP).

[0240] **FIG. 2** illustrates the donor template (i.e., “cargo” or “payload” or “template polynucleotide”)) packaged within a vector.

[0241] **FIG. 3** illustrates integrase-mediated self-circularization of the donor template (template polynucleotide) within viral genome. The circularized donor template is capable of being genomically incorporated into an orthogonal integrase target recognition site (i.e., “beacon”).

[0242] **FIG. 4** shows non-limiting illustrations of a gene editor construct packaged within a lipid nanoparticle and an atgRNA, ngRNA, and donor template (i.e., template polynucleotide encoding a gene of interest) packaged within a vector. GOI = gene of interest. PGI = programmable gene insertion. U6 = U6 promoter. atgRNA = attachment site-containing guide RNA (atgRNA).

[0243] **FIG. 5** shows non-limiting illustrations of a gene editor construct (e.g., mRNA encoding PE2-BxB1) and a nicking guide RNA (ngRNA) packaged within a lipid nanoparticle (LNP) and an atgRNA and donor template (i.e., template polynucleotide encoding a gene of interest) packaged within a vector.

[0244] **FIGs. 6A-6B** show non-limiting illustrations of three self-complementary AAV (scAAV) genomes capable of recombinase/integrase-mediated self-circularization. **FIG. 6A** shows the structure of the three self-complementary AAV (scAAV) genomes capable of recombinase/integrase-mediated self-circularization. **FIG. 6B** shows non-limiting examples of sequences that enable self-circularization (e.g., LoxP AttP GT (SEQ ID NO: 568 and SEQ ID NO: 569); FRT AttP GT (SEQ ID NO: 570 and SEQ ID NO: 571); and AttB CC AttP GT (SEQ ID NO: 572 and SEQ ID NO: 573)). GT indicates an AttP site with a GT dinucleotide. AttB CC indicates an AttB site with a CC dinucleotide. LoxP = a LoxP recombinase recognition site. FRT = a FRT recombinase recognition site.

[0245] **FIG. 7** shows a non-limiting illustration of recombinase/integrase-mediated intramolecular circularization products.

[0246] **FIGs. 8A-8B** show non-limiting illustrations of a ddPCR assay and intramolecular circularization ddPCR detection probes. **FIG. 8A** shows a non-limiting illustration of the ddPCR strategy. **FIG. 8B** shows non-limiting examples of the universal probe (SEQ ID NO: 574 and SEQ ID NO: 575) and an AttR probe (SEQ ID NO: 576 and SEQ ID NO: 577) that can be used in the assay shown in **FIG. 8A**.

[0247] **FIG. 9** shows a non-limiting illustration of a pDNA genome and AAV transfection and screening protocol.

[0248] **FIG. 10** shows data for circularization of AAV pDNA and packaged AAV genomic DNA with Bxb1.

[0249] **FIG. 11** shows data for Cre-, FLPe-, and Bxb1-mediated circularization of AAV pDNA confirmed by ddPCR.

[0250] **FIG. 12** shows Cre-, FLPe-, and Bxb1-mediated circularization of packaged AAV confirmed by ddPCR

[0251] **FIG. 13** shows percent circularization between the Bxb1-mediated attR scar ddPCR probe ("attR probe" described in **FIG. 8B**) and the universal ddPCR probe ("universal probe" described in **FIG. 8B**).

[0252] **FIGs. 14A-14E** shows analysis of AttP variants. **FIG. 14A** shows a non-limiting schematic of AttP mutations tested for improving integration efficiency (SEQ ID NOS: 394 and 540-542, respectively, in order of appearance). **FIG. 14B** shows integration efficiencies of wildtype and mutant AttP sites across a panel of AttB lengths. **FIG. 14C** shows a non-limiting schematic of multiplexed integration of different cargo sets at specific genomic loci. Three fluorescent cargos (GFP, mCherry, and YFP) are inserted orthogonally at three different loci (ACTB, LMNB1, NOLC1) for in-frame gene tagging. **FIG. 14D** shows orthogonality of top 4 AttB/AttP dinucleotide pairs evaluated for GFP integration with PASTE at the ACTB locus. **FIG. 14E** shows efficiency of multiplexed PASTE insertion of combinations of fluorophores at ACTB, LMNB1, and NOLC1 loci. Data are mean (n= 3) $\pm$ s.e.m.

[0253] **FIG. 15** illustrates a schematic of single atgRNA and dual atgRNA approaches for beacon placement (“integration recognition site”).

[0254] **FIG. 16** shows percent beacon placement in primary mouse hepatocytes (PMH) following delivery of mRNA to deliver a polynucleotide encoding a gene editor polynucleotide construct and an AAV to deliver the first and second atgRNA according to the following conditions: (i) concurrent delivery (“co-dose”), (ii) AAV delivery followed by a “1-day delay” before delivery of the mRNA, or (iii) AAV delivery followed by a “2-day delay” before delivery of the mRNA.

[0255] **FIG. 17** shows percent beacon placement in primary human hepatocytes (PHH) following delivering of mRNA to deliver a polynucleotide encoding a gene editor polynucleotide construct and an AAV to deliver the first and second atgRNA. The mRNA and AAV were delivered concurrently.

[0256] **FIG. 18** shows percent in vivo beacon placement in the *Nolc1* locus of mice following delivery of a polynucleotide encoding a gene editor polynucleotide construct using a lipid nanoparticle (LNP) and a first atgRNA and second atgRNA using an AAV. %BP = % beacon placement. LNP were administered at doses of 0.5 mg/kg, 1.5 mg/kg, 3 mg/kg, and 5 mg/kg. AAV was administered at 1E11, 3E11, or 1E12 viral genomes (vg) per animal. LNP #F1 = LNP formulation #1. LNP #F2 = LNP formulation #F2. LNP #F3 = LNP formulation #F3.

[0257] **FIG. 19** show percent in vivo integration of a template polynucleotide in AttP mice following delivering of the Bxb1 using adenovirus (AdV) and the template polynucleotide using an AAV (“AAV Cargo”). Bxb1 Adv was administered to the mice at a dose of either 3E10 or 1E11 vector genomes (vg) per animal. AAV Cargo was administered to the mice at a dose of 1E12.

[0258] **FIG. 20A** shows ddPCR data for percent in vivo beacon placement in the *Nolc1* locus of neonatal mice at eight days post-delivery of a single dose of a mixture of two LNPs. First LNP contained mRNA encoding a prime editing system and a first synthetic atgRNA (atgRNA1) at a 1:1 ratio. Second LNP contained mRNA encoding a prime editing system and a second synthetic atgRNA (atgRNA2) at a 1:1 ratio. Each of the first and second atgRNAs targeted the mouse *Nolc1* locus, encoded a portion of an integration recognition site (“beacon”), and together included a 6bp overlap. The first and second LNPs were combined

1:1 as mixture and administered at either 1 mg/kg or 3 mg/kg. LNP #F2 = LNP formulation #F2.

[0259] **FIG. 20B** show NGS data for percent in vivo beacon placement in the *Nolc1* locus of the same neonatal mice and treatment conditions as described in **FIG. 20A**. NGS data shows beacon placement eight days after administration of the LNP mixture. LNP #F2 = LNP formulation #F2.

[0260] **FIG. 20C** shows NGS data for percentage of in vivo beacons placed in the *Nolc1* NGS data is for the same mice with the same treatment conditions as described in **FIG. 20A**. NGS data shows data for eight days after administration of the LNP mixture. LNP #F2 = LNP formulation #F2.

[0261] **FIG. 21A** shows ddPCR data for percent in vivo beacon placement in the *Nolc1* locus of neonatal mice at 6 weeks post-delivery of a single dose of a mixture of two LNPs. First LNP contained mRNA encoding a prime editing system and a first synthetic atgRNA (atgRNA1) at a 1:1 ratio. Second LNP contained mRNA encoding a prime editing system and a second synthetic atgRNA (atgRNA2) at a 1:1 ratio. Each of the first and second atgRNAs targeted the mouse *Nolc1* locus, encoded a portion of an integration recognition site (“beacon”), and together included a 6bp overlap. The first and second LNPs were combined 1:1 as mixture and administered at either 1 mg/kg or 3 mg/kg. LNP #F2 = LNP formulation #F2.

[0262] **FIG. 21B** shows NGS data for percent in vivo beacon placement in the *Nolc1* locus of the same neonatal mice and treatment conditions as described in **FIG. 21A**. NGS data shows beacon placement 6 weeks after administration of the LNP mixture. LNP #F2 = LNP formulation #F2.

[0263] **FIG. 21C** shows NGS data for percentage of in vivo beacons placed in the *Nolc1* locus that included the expected integration recognition site. Data is from the same mice with the same treatment conditions as described in **FIG. 22A**. NGS data shows data at 6 weeks after administration of the LNP mixture. LNP #F2 = LNP formulation #F2.

[0264] **FIG. 22A** shows ddPCR data for percent in vivo beacon placement in the *Factor IX* (“mF9”) locus of 6-8 week old mice at day 8 post-delivery of a single dose of a mixture of two LNPs. First LNP contained mRNA encoding a prime editing system and a first synthetic atgRNA (atgRNA1) at a ratio of 1:0.5, 1:1, or 1:2. Second LNP contained mRNA encoding a

prime editing system and a second synthetic atgRNA (atgRNA2) at a ratio of 1:1, 1:0.5, or 1:2. Each of the first and second atgRNAs targeted the mouse *Factor IX* locus, encoded a portion of an integration recognition site (“beacon”), and together included a 6bp overlap. The first and second LNPs were combined 1:1 as mixture with the final ratio of mRNA:atgRNA1:atgRNA2 at 1:0.25:0.25; 1:0.5:0.5, or 1:1:1. LNP #F2 = LNP formulation #F2.

[0265] **FIG. 22B** shows NGS data for percent in vivo beacon placement in the *mF9* locus of the same neonatal mice and treatment conditions as described in **FIG. 22A**. NGS data shows beacon placement 8 days after administration of the LNP mixture. LNP #F2 = LNP formulation #F2.

[0266] **FIG. 22C** shows NGS data for percent of in vivo beacons placed in the *mF9* locus that included the expected integration recognition site. Data is from the same mice with the same treatment conditions as described in **FIG. 22A**. NGS data shows data at 8 days after administration of the LNP mixture. LNP #F2 = LNP formulation #F2.

6. DETAILED DESCRIPTION

[0267] Described herein is a method of co-delivering (i.e., “dual delivery”) to a cell a (i) gene editor construct and a (ii) donor (i.e., “cargo” or “payload”) template. The gene editor construct is comprised of a polynucleotide sequence that encodes the gene editor construct. In typical embodiments, the gene editor construct, upon polynucleotide expression or direct delivery of the gene editor protein and associated guide RNAs, can incorporate an integrase target recognition site (i.e., “beacon” or “landing pad”) or a recombinase target recognition site at a DNA locus. The gene editor polynucleotide construct is packaged within a lipid nanoparticle (LNP) that is capable of localizing the gene editor polynucleotide construct to a cell cytoplasm. The gene editor polynucleotide construct packaged in a LNP is co-delivered with a donor template (i.e., “cargo” or “payload”) polynucleotide construct packaged into a separate vector that is capable of localizing the donor template to a cell nucleus. In certain embodiments, the donor template vector is AAV, helper dependent adenovirus, or integration deficient lentivirus. In typical embodiments, the donor template is integrated into the genomic integrase target recognition site by an integrase, optionally by an integrase fused/linked to a gene editor protein.

6.1. Terminology

[0268] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. As used herein, the following terms have the meanings ascribed to them below.

[0269] “**Gene editor**” as used herein, is a protein that that can be used to perform gene editing, gene modification, gene insertion, gene deletion, or gene inversion. As used herein, the terms “gene editor polynucleotide” refers to polynucleotide sequence encoding the gene editor protein. Such an enzyme or enzyme fusion may contain DNA or RNA targetable nuclease protein (i.e., Cas protein, ADAR, or ADAT), wherein target specificity is mediated by a complexed nucleic acid (i.e., guide RNA). Such an enzyme or enzyme fusion may be a DNA/RNA targetable protein, wherein target specificity is mediated by internal, conjugated, fused, or linked amino acids, such as within TALENs, ZFNs, or meganucleases. The skilled person in the art would appreciate that the gene editor can demonstrate targeted nuclease activity, targeted binding with no nuclease activity, or targeted nickase activity (or cleavase activity). A gene editor comprising a targetable protein may be fused, linked, complexed, operate in cis or trans to one or more proteins or protein fragment motifs. Gene editors may be fused or linked to one or more integrase, recombinase, polymerase, telomerase, reverse transcriptase, or invertase. A gene editor can be a prime editor fusion protein or a gene writer fusion protein.

[0270] “**Prime editor fusion protein**” as used herein, describes a protein that is used in prime editing. “Prime editor system” as used herein describes the components used in prime editing. Prime editing uses CRISPR enzyme that nicks or cuts only single strand of double stranded DNA, i.e., a nickase; the nickase can occur either naturally or by mutation or modification of a nuclease that makes double stranded cuts. The nickase is programmed (directed) with a prime-editing guide RNA (pegRNA). The skilled person in the art would appreciate that the pegRNA both specifies the target site and encodes the desired edit. Described herein are attachment site containing guide RNA (atgRNA) that both specifies the target and encodes for the desired integrase target recognition site. The nickase may be programmed (directed) with an atgRNA. Advantageously the nickase is a catalytically impaired Cas9 endonuclease, a Cas9 nickase, that is fused to the reverse transcriptase. During genetic editing, the Cas9 nickase part of the protein is guided to the DNA target site by the atgRNA (or pegRNA), whereby a nick or single stranded cut occurs. The reverse transcriptase

domain then uses the atgRNA (or pegRNA) to template reverse transcription of the desired edit, directly polymerizing DNA onto the nicked target DNA strand. The edited DNA strand replaces the original DNA strand, creating a heteroduplex containing one edited strand and one unedited strand. Afterward, optionally, the prime editor (PE) guides resolution of the heteroduplex to favor copying the edit onto the unedited strand, completing the process (typically achieved with a nickase gRNA). Other enzymes that can be used to nick or cut only a single strand of double stranded DNA includes a cleavase (e.g., cleavase I enzyme).

[0271] In some embodiments, an additional agent or agents may be added that improve the efficiency and outcome purity of the prime edit. In some embodiments, the agent may be chemical or biological and disrupt DNA mismatch repair (MMR) processes at or near the edit site (i.e., PE4 and PE5 and PEmax architecture by *Chen et al.* Cell, 184, 1-18, October 28, 2021; *Chen et al.* is incorporated herein by reference). In typical embodiments, the agent is a MMR-inhibiting protein. In certain embodiments, the MMR-inhibiting protein is dominant negative MMR protein. In certain embodiments, the dominant negative MMR protein is MLH1dn. In particular embodiments, the MMR-inhibiting agent is incorporated into the co-delivery method described herein. In some embodiments, the MMR-inhibiting agent is linked or fused to the prime editor protein fusion, which may or may not have a linked or fused integrase. In some embodiments, the MMR-inhibiting agent is linked or fused to the Gene Writer™ protein, which may or may not have a linked or fused integrase.

[0272] The prime editor or gene editor system can be used to achieve DNA deletion and replacement. In some embodiments, the DNA deletion replacement is induced using a pair of atgRNAs or pegRNA that target opposite DNA strands, programming not only the sites that are nicked but also the outcome of the repair (i.e., PrimeDel by *Choi et al.* Nat. Biotechnology, October 14, 2021; *Choi et al.* is incorporated herein by reference and TwinPE by *Anzalone et al.* BioRxiv, November 2, 2021; *Anzalone et al.* is incorporated herein by reference). In some embodiments described herein, the DNA deletion is induced using a single atgRNA. In some embodiments, the DNA deletion and replacement is induced using a wild type Cas9 prime editor (PE-Cas9) system (i.e., PEDAR by *Jiang et al.* Nat. Biotechnology, October 14, 2021; *Jiang et al.* is incorporated herein by reference in its entirety). In some embodiments, the DNA replacement is an integrase target recognition site or recombinase target recognition site. In certain embodiments, the constructs and methods described herein may be utilized to incorporate the pair of pegRNAs (or atgRNAs) used in PrimeDel, TwinPE (WO2021226558

incorporated by reference herein in its entirety), or PEDAR, the prime editor fusion protein or Gene Writer protein, optionally a nickase guide RNA (ngRNA), an integrase, a nucleic acid cargo, and optionally a recombinase into a LNP delivery system or vector delivery system (e.g., AAV or Adenovirus). The integrase may be directly linked, for example by a peptide linker, to the prime editor fusion or gene writer protein.

[0273] In some embodiments, the prime editors can refer to a retrovirus or lentivirus reverse transcriptase such as a Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase (RT) fused to a CRISPR enzyme nickase such as a Cas9 H840A nickase, a Cas9nickase. In some embodiments, the prime editors can refer to a retrovirus or lentivirus reverse transcriptase such as a Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase (RT) fused to a cleavase. In some embodiments the RT can be fused at, near or to the C-terminus of a Cas9nickase, e.g., Cas9 H840A. Fusing the RT to the C-terminus region, e.g., to the C-terminus, of the Cas9 nickase may result in higher editing efficiency. Such a complex is called PE1. In some embodiments, the CRISPR enzyme nickase, e.g., Cas9(H840A), i.e., a Cas9nickase, can be linked to a non-M-MLV reverse transcriptase such as an AMV-RT or XRT (Cas9(H840A)-AMV-RT or XRT). In some embodiments, instead of the CRISPR enzyme nickase being a Cas9 (H840A), i.e., instead of being a Cas9 nickase, the CRISPR enzyme nickase instead can be a CRISPR enzyme that naturally is a nickase or cuts a single strand of double stranded DNA; for instance, the CRISPR enzyme nickase can be Cas12a/b. Alternatively, the CRISPR enzyme nickase can be another mutation of Cas9, such as Cas9(D10A). A CRISPR enzyme, such as a CRISPR enzyme nickase, such as Cas9 (wild type), Cas9(H840A), Cas9(D10A) or Cas 12a/b nickase can be fused in some embodiments to a pentamutant of M-MLV RT (D200N/ L603W/ T330P/ T306K/ W313F), whereby there can be up to about 45-fold higher efficiency, and this is called PE2. In some embodiments, the M-MLV RT comprise one or more of the mutations Y8H, P51L, S56A, S67R, E69K, V129P, L139P, T197A, H204R, V223H, T246E, N249D, E286R, Q2911, E302K, E302R, F309N, M320L, P330E, L435G, L435R, N454K, D524A, D524G, D524N, E562Q, D583N, H594Q, E607K, D653N, and L671P. Specific M-MLV RT mutations are shown in Table 1.

Table 1		
SEQ ID NO	Description	Forward Sequence (5'-3')
SEQ ID NO: 01	RT_mut_L139P	ttgagcgggCCCccaccgt

SEQ ID NO: 02	RT_mut_E562Q	cagcgggctCAGctgatagca
SEQ ID NO: 03	RT_mut_D653N	cggatggctAACcaagcggcc

[0274] In some embodiments, the reverse transcriptase can also be a wild-type or modified transcription xenopolymerase (RTX), avian myeloblastosis virus reverse transcriptase (AMV RT), Feline Immunodeficiency Virus reverse transcriptase (FIV-RT), FeLV-RT (Feline leukemia virus reverse transcriptase), HIV-RT (Human Immunodeficiency Virus reverse transcriptase). In some embodiments, the reverse transcriptase can be a fusion of MMuLV to the Sto7d DNA binding domain (see Ionnidi et al.; <https://doi.org/10.1101/2021.11.01.466786>). The fusion of MMuLV to the Sto7d DNA binding domain sequence is given in Table 2.

Table 2		
Description	Forward Sequence (5'-3')	SEQ ID NO:
RT(1-478)_Sto7d fusion [MMulv sequence (in bold), Sto7d sequence]	atgactcactatcaggccttgcttttggacacggaccgggtccagttcggaccgggtgtagccctgaaccc ggctacgctgctcccactgcctgaggaagggtgcaacacaactgccttgatGGGACAGGTGGC GGTGGTGTCAACGTCAAGTTCAAGTACAAGGGTGAGGAACTTGA AGTTGATATTAGCAAAATCAAGAAGGTTTGGCGCGTTGGTAAAA TGATATCTTTTACTTATGACGACAACGGCAAGACAGGTAGAGGG GCAGTGTCTGAGAAAGACGCCCCAAGGAGCTGTTGCAAATGTT GGAAAAGTCTGGGAAAAAGtctggcgggtcaaaaagaaccgcccagcgcagcgaattc gagcccaagaagaagaggaaagtc	4

[0275] PE3, PE3b, PE4, PE5, and/or PEmax, which a skilled person can incorporate into the co-delivery system described herein, involves nicking the non-edited strand, potentially causing the cell to remake that strand using the edited strand as the template to induce HR. The nicking of the non-edited strand can involve the use of a nicking guide RNA (ngRNA).

[0276] The skilled person can readily incorporate into the co-delivery system described herein described herein a prime editing or CRISPR system. Examples of prime editors can be found in the following: WO2020/191153, WO2020/191171, WO2020/191233, WO2020/191234, WO2020/191239, WO2020/191241, WO2020/191242, WO2020/191243, WO2020/191245, WO2020/191246, WO2020/191248, WO2020/191249, each of which is incorporated by reference herein in its entirety. In addition, mention is made, and can be

used herein, of CRISPR Patent Applications and Patents of the Zhang laboratory and/or Broad Institute, Inc. and Massachusetts Institute of Technology and/or Broad Institute, Inc., Massachusetts Institute of Technology and President and Fellows of Harvard College and/or Editas Medicine, Inc. Broad Institute, Inc., The University of Iowa Research Foundation and Massachusetts Institute of Technology, including those claiming priority to US Application 61/736,527, filed December 12, 2012, including US Patents 11,104,937, 11,091,798, 11,060,115, 11,041,173, 11,021,740, 11,008,588, 11,001,829, 10,968,257, 10,954,514, 10,946,108, 10,930,367, 10,876,100, 10,851,357, 10,781,444, 10,711,285, 10,689,691, 10,648,020, 10,640,788, 10,577,630, 10,550,372, 10,494,621, 10,377,998, 10,266,887, 10,266,886, 10,190,137, 9,840,713, 9,822,372, 9,790,490, 8,999,641, 8,993,233, 8,945,839, 8,932,814, 8,906,616, 8,895,308, 8,889,418, 8,889,356, 8,871,445, 8,865,406, 8,795,965, 8,771,945, and 8,697,359; CRISPR Patent Applications and Patents of the Doudna laboratory and/or of Regents of the University of California, the University of Vienna and Emmanuelle Charpentier, including those claiming priority to US application 61/652,086, filed May 25, 2012, and/or 61/716,256, filed October 19, 2012, and/or 61/757,640, filed January 28, 2013, and/or 61/765,576, filed February 15, 2013 and/or 13/842,859, including US Patents 11,028,412, 11,008,590, 11,008,589, 11,001,863, 10,988,782, 10,988,780, 10,982,231, 10,982,230, 10,900,054, 10,793,878, 10,774,344, 10,752,920, 10,676,759, 10,669,560, 10,640,791, 10,626,419, 10,612,045, 10,597,680, 10,577,631, 10,570,419, 10,563,227, 10,550,407, 10,533,190, 10,526,619, 10,519,467, 10,513,712, 10,487,341, 10,443,076, 10,428,352, 10,421,980, 10,415,061, 10,407,697, 10,400,253, 10,385,360, 10,358,659, 10,358,658, 10,351,878, 10,337,029, 10,308,961, 10,301,651, 10,266,850, 10,227,611, 10,113,167, and 10,000,772; CRISPR Patent Applications and Patents of Vilnius University and/or the Siksnys laboratory, including those claiming priority to US application 62/046384 and/or 61/625,420 and/or 61/613,373 and/or PCT/IB2015/056756, including US Patent 10,385,336; CRISPR Patent Applications and Patents of the President and Fellows of Harvard College, including those of George Church's laboratory and/or claiming priority to US application 61/738,355, filed December 17, 2012, including 11,111,521, 11,085,072, 11,064,684, 10,959,413, 10,925,263, 10,851,369, 10,787,684, 10,767,194, 10,717,990, 10,683,490, 10,640,789, 10,563,225, 10,435,708, 10,435,679, 10,375,938, 10,329,587, 10,273,501, 10,100,291, 9,970,024, 9,914,939, 9,777,262, 9,587,252, 9,267,135, 9,260,723, 9,074,199, 9,023,649; CRISPR Patent Applications and Patents of the President and Fellows of Harvard College, including those of David Liu's laboratory, including 11,111,472, 11,104,967, 11,078,469,

11,071,790, 11,053,481, 11,046,948, 10,954,548, 10,947,530, 10,912,833, 10,858,639, 10,745,677, 10,704,062, 10,682,410, 10,612,011, 10,597,679, 10,508,298, 10,465,176, 10,323,236, 10,227,581, 10,167,457, 10,113,163, 10,077,453, 9,999,671, 9,840,699, 9,737,604, 9,526,784, 9,388,430, 9,359,599, 9,340,800, 9,340,799, 9,322,037, 9,322,006, 9,228,207, 9,163,284, and 9,068,179; and CRISPR Patent Applications and Patents of Toolgen Incorporated and/or the Kim laboratory and/or claiming priority to US application 61/717,324, filed October 23, 2012 and/or 61/803,599, filed March 20, 2013 and/or 61/837,481, filed June 20, 2013 and/or 62/033,852, filed August 6, 2014 and/or PCT/KR2013/009488 and/or PCT/KR2015/008269, including US Patent 10,851,380, and 10,519,454; and CRISPR Patent Applications and Patents of Sigma and/or Millipore and/or the Chen laboratory and/or claiming priority to US application 61/734,256, filed December 6, 2012 and/or 61/758,624, filed January 30, 2013 and/or 61/761,046, filed February 5, 2013 and/or 61/794,422, filed March 15, 2013, including US Patent 10,731,181, each of which is hereby incorporated herein by reference, and from the disclosures of the foregoing, the skilled person can readily make and use a prime editing or CRISPR system, and can especially appreciate impaired endonucleases, such as a mutated Cas9 that only nicks a single strand of DNA and is hence a nickase, or a CRISPR enzyme that only makes a single-stranded cut that can be employed in a PASTE system of the invention. Further, from the disclosures of the foregoing, the skilled person can incorporate the selected CRISPR enzyme, as part of the prime editor fusion or gene editor fusion, into the co-delivery method described herein.

[0277] Prior to RT-mediated edit incorporation, the prime editor protein (or system) (1) site-specifically targets a genomic locus and (2) performs a catalytic cut or nick. These steps are typically performed by a CRISPR-Cas. However, in some embodiments the Cas protein may be substituted by other nucleic acid programmable DNA binding proteins (napDNAbp) such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), or meganucleases. In addition, to the extent the “targeting rules” of other napDNAbp are known or are newly determined, it becomes possible to use new napDNAbp, beyond Cas9, to site specifically target and modify genomic sites of interest.

[0278] Similar to a prime editor protein, a Gene Writer can introduce novel DNA elements, such as an integration target site, into a DNA locus. A Gene Writer protein comprises: (A) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a

reverse transcriptase domain, and either (x) an endonuclease domain that contains DNA binding functionality or (y) an endonuclease domain and separate DNA binding domain; and (B) a template RNA comprising (i) a sequence that binds the polypeptide and (ii) a heterologous insert sequence. Examples of such Gene Writer™ proteins and related systems can be found in US20200109398, which is incorporated by reference herein in its entirety.

[0279] In some embodiments, the prime editor or Gene Writer protein fusion or prime editor protein linked or fused to an integrase is expressed as a split construct. In typical embodiments, the split construct is reconstituted in a cell. In some embodiments, the split construct can be fused or ligated via intein protein splicing. In some embodiments, the split construct can be reconstituted via protein-protein inter-molecular bonding and/or interactions. In some embodiments, the split construct can be reconstituted via chemical, biological, or environmental induced oligomerization. In certain embodiments, the split construct can be adapted into one or more delivery vectors described herein.

[0280] In some embodiments, an integrase or recombinase is directly linked or fused, for example by a peptide linker, which may be cleavable or non-cleavable, to the prime editor fusion protein (i.e., fused Cas9 nickase-reverse transcriptase) or Gene Writer protein. Suitable linkers, for example between the Cas9, RT, and integrase, may be selected from Table 3:

Table 3.				
	Sequence (5'-3')	SEQ ID NO:	Amino acid sequence	SEQ ID NO:
A - P2A	GGAAGCGGAGCTACTAACTTC AGCCTGCTGAAGCAGGCTGGC GACGTGGAGG AGAACCCTGGACCT	5	GSGATNFSLLKQAG DVEENPGP	13
B - (GGGS)3	GGGGGAGGAGGTTCTGGAGGC GGAGG CTCCGGAGGCGGAGGGTCA	6	GGGGSGGGSGGGG S	14
C - GGGGS	GGAGGTGGCGGGAGC	7	GGGGS	15
D - PAPAP	CCCGCACCAGCGCCT	8	PAPAP	16
E - (EAAAK)3	GAGGCAGCTGCCAAGGAAGCC GCTGCCAAGGAGGCGGCCGCA AAG	9	EAAAKEAAKEAAA K	17
F - XTEN	AGTGGGAGCGAGACCCCTGGG ACTAGCGAGTCAGCTACACCCG AAAGC	10	SGSETPGTSESATPES	18

Table 3.				
	Sequence (5'-3')	SEQ ID NO:	Amino acid sequence	SEQ ID NO:
G - (GGS)6	GGGGGGTCAGGTGGATCCGGC GGAAGTGGCGGATCCGGTGA TCTGGCGGCAGT	11	GGSGGSGGSGGSGG SGGS	19
H - EAAAK	GAAGCTGCTGCTAAG	12	EAAAK	20
(GGGGS)4	GGCGGCGGCGGCAGCGGCGGC GGCGGCAGCGGCGGCGGCGGC AGCGGCGGCGGCGGCAGC	543	GGGGSGGGSGGGG SGGGGS	551
PAS8	GGCGGCGCGAGCCCGGCGGGC GGC	544	GGASPAGG	552
PAS12	GGCGGCGCGAGCCCGGCGGCG CCGCGCCGCGGCGGC	545	GGASPAAPAPAG	553
A(EAAK)4ALEA(EAAAK)4A	GCGGAAGCGGCGAAAGAAGCG GCGAAAGAAGCGGCGAAAGAA GCGGCGAAAGCGCTGGAAGCG GAAGCGGCGGCGAAAGAAGCG GCGGCGAAAGAAGCGGCGGCG AAAGAAGCGGCGGCGAAAGCG	546	AEAAKEAAKEAAKE AAKALEAEAAAKEA AAKEAAAKEAAAK A	554
Camel	GCGCATCATAGCGAAGATCCG GGCGGCGGCGGCAGCGGCGGC GGCGGCAGCGGCGGCGGCGGC AGC	547	AHHSEDPGGGSGG GGSGGGGS	555
FRF	GGCGGCGGCGGCAGCGAAGCG GCGGCGAAAGGCGGCGGCGGC AGC	548	GGGGSEAAAKGGGG S	556
RFR	GAAGCGGCGGCGAAAGGCGGC GGCGGCAGCGAAGCGGCGGCG AAA	549	EAAAKGGGGSEAAA K	557
Modified XTEN (mXTEN)	AGCGGCGGCAGCAGCGGCGGC AGCAGCGGCAGCGAAACCCCG GGCACCAGCGAAAGCGCGACC CCGGAAAGCAGCGGCGGCAGC AGCGGCGGCAGCAGCACC	550	SGGSSGGSSGSETPG TSESATPESGGSSG GSST	558

[0281] In some embodiments, the prime editor or Gene Writer protein fusion or prime editor protein linked or fused to an integrase is expressed as a split construct. In typical embodiments, the split construct is reconstituted in a cell. In some embodiments, the split construct can be fused or ligated via intein protein splicing. In some embodiments, the split construct can be reconstituted via protein-protein inter-molecular bonding and/or interactions. In some embodiments, the split construct can be reconstituted via chemical, biological, or

environmental induced oligomerization. In certain embodiments, the split construct can be adapted into one or more nucleic acid constructs described herein.

6.2. Type II CRISPR proteins

[0282] The skilled person can incorporate a selected CRISPR enzyme, described below, as part of the prime editor fusion, into the co-delivery method described herein. *Streptococcus pyogenes* Cas9 (SpCas9), the most common enzyme used in genome-editing applications, is a large nuclease of 1368 amino acid residues. Advantages of SpCas9 include its short, 5'-NGG-3' PAM and very high average editing efficiency. SpCas9 consists of two lobes: a recognition (REC) lobe and a nuclease (NUC) lobe. The REC lobe can be divided into three regions, a long α helix referred to as the bridge helix (residues 60–93), the REC1 (residues 94–179 and 308–713) domain, and the REC2 (residues 180–307) domain. The NUC lobe consists of the RuvC (residues 1–59, 718–769, and 909–1098), HNH (residues 775–908), and PAM-interacting (PI) (residues 1099–1368) domains. The negatively charged sgRNA:target DNA heteroduplex is accommodated in a positively charged groove at the interface between the REC and NUC lobes. In the NUC lobe, the RuvC domain is assembled from the three split RuvC motifs (RuvC I–III) and interfaces with the PI domain to form a positively charged surface that interacts with the 30 tail of the sgRNA. The HNH domain lies between the RuvC II–III motifs and forms only a few contacts with the rest of the protein. Structural aspects of SpCas9 are described by Nishimasu et al., Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA, Cell 156, 935-949, February 27, 2014.

[0283] *REC lobe*: The REC lobe includes the REC1 and REC2 domains. The REC2 domain does not contact the bound guide:target heteroduplex, indicating that truncation of REC lobe may be tolerated by SpCas9. Further, SpCas9 mutant lacking the REC2 domain (D175–307) retained ~50% of the wild-type Cas9 activity, indicating that the REC2 domain is not critical for DNA cleavage. In striking contrast, the deletion of either the repeat-interacting region (D97–150) or the anti-repeat-interacting region (D312–409) of the REC1 domain abolished the DNA cleavage activity, indicating that the recognition of the repeat:anti-repeat duplex by the REC1 domain is critical for the Cas9 function.

[0284] *PAM-Interacting domain*: The NUC lobe contains the PAM-interacting (PI) domain that is positioned to recognize the PAM sequence on the noncomplementary DNA strand. The PI domain of SpCas9 is required for the recognition of 5'-NGG-3' PAM, and deletion of the

PI domain (Δ 1099-1368) abolished the cleavage activity, indicating that the PI domain is critical for SpCas9 function and a major determinant for the PAM specificity.

[0285] *RuvC domain:* The RuvC nucleases of SpCas9 have an RNase H fold and four catalytic residues, Asp10 (Ala), Glu762, His983, and Asp986, that are critical for the two-metal cleavage of the noncomplementary strand of the target DNA. In addition to the conserved RNase H fold, the Cas9 RuvC domain has other structural elements involved in interactions with the guide:target heteroduplex (an end-capping loop between α 42 and α 43) and the PI domain/stem loop 3 (β hairpin formed by β 3 and β 4).

[0286] *HNH domain:* SpCas9 HNH nucleases have three catalytic residues, Asp839, His840, and Asn863 and cleave the complementary strand of the target DNA through a single-metal mechanism.

[0287] *sgRNA:DNA recognition:* The sgRNA guide region is primarily recognized by the REC lobe. The backbone phosphate groups of the guide region (nucleotides 2, 4–6, and 13–20) interact with the REC1 domain (Arg165, Gly166, Arg403, Asn407, Lys510, Tyr515, and Arg661) and the bridge helix (Arg63, Arg66, Arg70, Arg71, Arg74, and Arg78). The 2'-hydroxyl groups of G1, C15, U16, and G19 hydrogen bond with Val1009, Tyr450, Arg447/Ile448, and Thr404, respectively.

[0288] A mutational analysis demonstrated that the R66A, R70A, and R74A mutations on the bridge helix markedly reduced the DNA cleavage activities, highlighting the functional significance of the recognition of the sgRNA “seed” region by the bridge helix. Although Arg78 and Arg165 also interact with the “seed” region, the R78A and R165A mutants showed only moderately decreased activities. These results are consistent with the fact that Arg66, Arg70, and Arg74 form multiple salt bridges with the sgRNA backbone, whereas Arg78 and Arg165 form a single salt bridge with the sgRNA backbone. Moreover, the alanine mutations of the repeat:anti-repeat duplex-interacting residues (Arg75 and Lys163) and the stemloop-1-interacting residue (Arg69) resulted in decreased DNA cleavage activity, confirming the functional importance of the recognition of the repeat:anti-repeat duplex and stem loop 1 by Cas9.

[0289] *RNA-guided DNA targeting:* SpCas9 recognizes the guide:target heteroduplex in a sequence-independent manner. The backbone phosphate groups of the target DNA (nucleotides 1, 9–11, 13, and 20) interact with the REC1 (Asn497, Trp659, Arg661, and Gln695), RuvC

(Gln926), and PI (Glu1108) domains. The C2' atoms of the target DNA (nucleotides 5, 7, 8, 11, 19, and 20) form van der Waals interactions with the REC1 domain (Leu169, Tyr450, Met495, Met694, and His698) and the RuvC domain (Ala728). The terminal base pair of the guide:target heteroduplex (G1:C20') is recognized by the RuvC domain via end-capping interactions; the sgRNA G1 and target DNA C20' nucleobases interact with the Tyr1013 and Val1015 side chains, respectively, whereas the 2'-hydroxyl and phosphate groups of sgRNA G1 interact with Val1009 and Gln926, respectively.

[0290] *Repeat:Anti-Repeat duplex recognition:* The nucleobases of U23/A49 and A42/G43 hydrogen bond with the side chain of Arg1122 and the main-chain carbonyl group of Phe351, respectively. The nucleobase of the flipped U44 is sandwiched between Tyr325 and His328, with its N3 atom hydrogen bonded with Tyr325, whereas the nucleobase of the unpaired G43 stacks with Tyr359 and hydrogen bonds with Asp364.

[0291] The nucleobases of G21 and U50 in the G21:U50 wobble pair stack with the terminal C20:G10 pair in the guide:target heteroduplex and Tyr72 on the bridge helix, respectively, with the U50 O4 atom hydrogen bonded with Arg75. Notably, A51 adopts the syn conformation and is oriented in the direction opposite to U50. The nucleobase of A51 is sandwiched between Phe1105 and U63, with its N1, N6, and N7 atoms hydrogen bonded with G62, Gly1103, and Phe1105, respectively.

[0292] *Stem-loop recognition:* Stem loop 1 is primarily recognized by the REC lobe, together with the PI domain. The backbone phosphate groups of stem loop 1 (nucleotides 52, 53, and 59–61) interact with the REC1 domain (Leu455, Ser460, Arg467, Thr472, and Ile473), the PI domain (Lys1123 and Lys1124), and the bridge helix (Arg70 and Arg74), with the 2'-hydroxyl group of G58 hydrogen bonded with Leu455. A52 interacts with Phe1105 through a face-to-edge p-p stacking interaction, and the flipped U59 nucleobase hydrogen bonds with Asn77.

[0293] The single-stranded linker and stem loops 2 and 3 are primarily recognized by the NUC lobe. The backbone phosphate groups of the linker (nucleotides 63–65 and 67) interact with the RuvC domain (Glu57, Lys742, and Lys1097), the PI domain (Thr1102), and the bridge helix (Arg69), with the 2'-hydroxyl groups of U64 and A65 hydrogen bonded with Glu57 and His721, respectively. The C67 nucleobase forms two hydrogen bonds with Val1100.

[0294] Stem loop 2 is recognized by Cas9 via the interactions between the NUC lobe and the non-Watson-Crick A68:G81 pair, which is formed by direct (between the A68 N6 and G81 O6 atoms) and water-mediated (between the A68 N1 and G81 N1 atoms) hydrogen-bonding interactions. The A68 and G81 nucleobases contact Ser1351 and Tyr1356, respectively, whereas the A68:G81 pair interacts with Thr1358 via a water-mediated hydrogen bond. The 20-hydroxyl group of A68 hydrogen bonds with His1349, whereas the G81 nucleobase hydrogen bonds with Lys33.

[0295] Stem loop 3 interacts with the NUC lobe more extensively, as compared to stem loop 2. The backbone phosphate group of G92 interacts with the RuvC domain (Arg40 and Lys44), whereas the G89 and U90 nucleobases hydrogen bond with Gln1272 and Glu1225/Ala1227, respectively. The A88 and C91 nucleobases are recognized by Asn46 via multiple hydrogen-bonding interactions.

[0296] Cas9 proteins smaller than SpCas9 allow more efficient packaging of nucleic acids encoding CRISPR systems, e.g., Cas9 and sgRNA into one rAAV (“all-in-one-AAV”) particle. In addition, efficient packaging of CRISPR systems can be achieved in other viral vector systems (i.e., lentiviral, integration deficient lentiviral, hd-AAV, etc.) and non-viral vector systems (i.e., lipid nanoparticle). Small Cas9 proteins can be advantageous for multidomain-Cas-nuclease-based systems for prime editing. Well characterized smaller Cas9 proteins include *Staphylococcus aureus* (SauCas9, 1053 amino acid residues) and *Campylobacter jejuni* (CjCas9, 984 amino residues). However, both recognize longer PAMs, 5'-NNGRRT-3' for SauCas9 (R = A or G) and 5'-NNNNRYAC-3' for CjCas9 (Y = C or T), which reduces the number of uniquely addressable target sites in the genome, in comparison to the NGG SpCas9 PAM. Among smaller Cas9s, Schmidt et al. identified *Staphylococcus lugdunensis* (Slu) Cas9 as having genome-editing activity and provided homology mapping to SpCas9 and SauCas9 to facilitate generation of nickases and inactive (“dead”) enzymes (Schmidt et al., 2021, Improved CRISPR genome editing using small highly active and specific engineered RNA-guided nucleases. *Nat Commun* 12, 4219. doi.org/10.1038/s41467-021-24454-5) and engineered nucleases with higher cleavage activity by fragmenting and shuffling Cas9 DNAs. The small Cas9s and nickases are useful in the instant invention.

[0297] Besides dead Cas9 and Cas9 nickase variants, the Cas9 proteins used herein may also include other “Cas9 variants” having at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical,

at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to any reference Cas9 protein, including any wild type Cas9, or mutant Cas9 (e.g., a dead Cas9 or Cas9 nickase), or fragment Cas9, or circular permutant Cas9, or other variant of Cas9 disclosed herein or known in the art. In some embodiments, a Cas9 variant may have 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, 50 or more amino acid changes compared to a reference Cas9. In some embodiments, the Cas9 variant comprises a fragment of a reference Cas9 (e.g., a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9. In some embodiments, the fragment is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9 (e.g., SEQ ID NO: 18).

[0298] In some embodiments, the disclosure also may utilize Cas9 fragments that retain their functionality and that are fragments of any herein disclosed Cas9 protein. In some embodiments, the Cas9 fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or at least 1300 amino acids in length.

[0299] In various embodiments, the prime editors disclosed herein may comprise one of the Cas9 variants described as follows, or a Cas9 variant thereof having at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to any reference Cas9 variants.

Table 4- Cas9 orthologs						
Streptococcus pyogenes AJN60024.1	MDKKYSIGLD	IGTNSVGWAV	ITDEYKVPSP	KFKVLGNTDR	HSIKKNLIGA	(SEQ ID NO: 21)
	LLFDSGETAE	ATRLKRTARR	RYTRRKNRIC	YLQEIFSNE	AKVDDSFHHR	
	LEESFLVEED	KKHERHPIFG	NIVDEVAYHE	KYPTIYHLRK	KLVDSTDKAD	
	LRLIYLALAH	MIKFRGHFLI	EGDLNPDNSD	VDKLFQQLVQ	TYNQLFEENP	
	INASGVDAKA	ILSARLSKSR	RLENLIAQLP	GEKKNGLFGN	LIALSLGLTP	

Table 4- Cas9 orthologs						
GI: 757015980 WP_010922 251.1	NFKSNFDLAE LLSDILRVNT FFDQSKNGYA KQRTFDNGSI YVGPLARGNS NLPNEKVLPK LLFKTNRKVT IKDKDFLDNE LKRRTYTGWG SLTFKEDIQK MGRHKPENIV VENTQLQNEK SIDNKVLTRS TKAERGGLSE REVKVITLKS YPKLESEFVY TLANGEIRKR QTGGFSKESI KKGSKKLKSV YSLFELENGR DNEQKQLFVE PIREQAENII SITGLYETRI	DAKLQLSKDT EITKAPLSAS GYIDGGASQE PHQIHLGELH RFAWMTRKSE HSLLYEYFTV VKQLKEDYFK ENEDILEDIV RLSRKLINGI AQVSGQGDSL IEMARENQTT LYLYYLQNGR DKNRGKSDNV LDKAGFIKRQ KLVSDFRKDF GDYKVYDVRK PLIETNGETG LPKRNSDKLI KELLGITIME KRMLASAGEL QHKHYLDEII HLFTLTNLGA DLS	YDDDLNLLA MIKRYDEHHQ EFYKFIKPIL AILRRQEDFY ETITPWNFEE YNELTKVKYV KIECFDSVEI LTLTLFEDRE RDKQSGKTIL HEHIANLAGS QKGQKNSRER DMYVDQELDI PSEEVVKMK LVETRQITKH QFYKVVREINN MIAKSEQEIG EIVWDKGRDF ARKKDWDPKK RSSFEKNPID QKGNELALPS EQISEFSKRV PAAFKYFDTT	QIGDQYADLF DLTLLKALVR EKMDGTEELL PFLKDNREKI VVDKGASQAS TEGMRKPAFL SGVEDRFNAS MIEERLKYTA DFLKSDBGAN PAIKKGILQT MKRIEEGIKE NRLSDYDVDH NYWRQLLNAK VAQILDSRMN YHHAHDAYLN KATAKYFFYS ATVRKVLSP YGGFDSPTVA FLEAKGYKEV KYVNFLYLAS ILADANLDKV IDRKRYTSTK	LAANKLSDAI QQLPEKYKEI VKLNREDLLR EKILTFRIPI FIERMTNFDP SGEQKKAIVD LGTYHDLKI HLFDDKVMQ RNFMQLIHDD VKVVDLVKVP LGSQILKEHP IVPQSFLKDD LITQRKFDNL TKYDENDKLI AVVGTALIKK NIMNFFKTEI QVNIVKKTEV YSVLVVAKEV KKDLI KLPK HYEKLKGSPE LSAYNKHARDK EVLDTATLIHQ	
AJN60021.1 GI: 757015977 J7RUA5.1 WP_053019 794.1 Staphylococ cus aureus	MKRNYILGLD RGARRLKRRR SEEFSAALL AELQLERLKK YIDLLETRRT YNADLYNALN KEILVNEEDI IAKILTIYQS NLILDELWHT KRSFIQSIKV TNERIEEIR FNYEVDHIIP YETFKKHILN YATRGLMNL HAEDALIIAN KEIFITPHQI IVNNLNGLYD KNPLYKYEE RNKVVKLSLK KKLKKISNQA YREYLENMND KKG	IGITSVGYGI RHRIQRVKKL HLAKRRGVHN DGEVRGSINR YYEGPGEOSP DLNNLVITRD KGYRVTSTGK SEDIQEELTN NDNQIAIFNR INAIKKYGL TTGKENAKYL RSVSFDNSFN LAKGKGRISK RSYFRVNNLD ADFIKKEWKK KHIDFKDYK KDNDKLKLI TGNYLTKYSK PYRFDVYLDN EFIASFYNN KRPPRIIKTI	IDYETRDVID LFDYNLLTDH VNEVEEDTGN FKTSDYVKEA FGWKDIKEWY ENEKLEYEYK PEFTNLKVYH LNSELTQEEI LKLVPKVDL PNDIIIEELAR IEKIKLHDMQ NKVLVKQEE TKKEYLLEER VKVKSINGGF LDKAKKVMEN YSHRVDKKPN NKSPEKLLMY KDNGPVIKKI GVYKFVTVKN LIKINGELYR ASKTQSIKKY	AGVRLFKKAN SELSGINPYE ELSTKEQISR KQLLKVQKAY EMLMGHCTYF FQIIENVFKQ DIKDITARKE EQISNLKGYT SQQKEIPTTL EKNSKDAQKM EGKCLYSLEA SKGNRTPFQ DINRFSVQKD TSFLRRKWK QMFEEKQAES RELINDTLYS HHDPQTYQKL KYYGNKLN LDVIKENY VIGVNNDLLN STDILGNLYE	VENNEGRRSK ARVKGSLQKL NSKALEEKYV HQLDQSFIDT PEELRSVKYA KKKPTLKQIA I IENAELLDQ GTHNLSLKAI VDDFILSPVV INEMQKRNRQ IPLEDLLNPN YLSSSDSKIS FINRNLVDTR KKERNKGYKH MPEIETEQBY TRKDDKGNTL KLIMEQYGD LDITDDYPNS EVNSKCYEBA RIEVNMIDIT VKSCKHPQII	(SEQ ID NO: 22)
AJN60008.1 GI: 757015964 WP_002864 485.1 Campylobac ter jejuni subsp. jejuni	MARILAFDIG ARSARKRLAR PYELRFRALN QNEEKLANQ KDELKLIFKK DEKRAPKNSP NGTLTYKQTK DLNEIAKDIT LKLVTPLMLE VVLRAIKEYR NYKAKKDAEL	ISSIGWAFSE RKARLNHLKH ELLSKQDFAR SVGEYLYKEY QREFGFSFSK LAFMFVALTR KLLGLSDDYE LIKDEIKLKK GKKYDEACNE KVLNALLKKY ECEKLGLKIN	NDELKDCGVR LIANEFKLN VILHIAKRRG FQKFKENSKE KFEEEVLSVA IINLLNNLKN FKGEKGTYFI ALAKYDLNQN LNLKVAINED GKVHKINIEL SKNILKLRLF	IFTKVENPKT EDYQSFD YDDIKNSDDK FTNVRNKKES FYKRALKDFS TEGILYTKDD EFKKYKEFIK QIDSLSKLEF KKDFLPAFNE AREVGKNHSQ KEQKEFCAYS	GESLALPRRL AKAYKGLSIS EKGAILKAIK YERCIAQSFT HLVGNCSTFL LNALLNEVLK ALGEHNSQD KDHLNISFKA TYYKDEVTNP RAKIEKEQNE GEKIKISDLQ	(SEQ ID NO: 23)

Table 4- Cas9 orthologs						
NCTC 11168 = ATCC 700819	DEKMLEIDHI WQKIEVLAKN TKDYLDLPL RNNHLHHAID NKRKFFEPFS SYGGKEGVK TMDFAKVL MQEPEFVY SIGIQNLKVF	YPYSRSFDDS LPTKKQKRIL SDDENTKLND AVIIAYANNS GFRQKVLDKI ALELGKIRKV NKAVARSKKG AFTSSTVSLI EKYIVSALGE	YMNKVLVFTK DKNYKDKEQK TQKGSKVHVE IVKAFSDFKK DEIFVSKPER NGKIVKNGDM EIKDWILMDE VSKHDNKFET VTKAEFRQRE	QNQEKLNQTP NFKDRNLNDT AKSGMLTSAL EQESNSAELY KKPSGALHEE FRVDIFKHKK NYEFCFSLYK LSKNQKILFK DFKK	FEAFGNDSAK RYIARLVLNY RHTWGFSAKD AKKISELDYK TFRKEEEFYQ TNKFYAVPIY DSLILIQTKD NANEKEVIAK	
Streptococcus thermophilus LMD-9 AJN60026.1 GI: 757015982 WP_011680957.1	MSDLVLGLDI QGRRLARRKK SNEELFIALK PGQIQLERYQ QEFNPQITDE IFGILIGKCT KNQIINLVKN EAYRKMKTLE FSQKQVDELV GDFDNIVIE LPHSVFHGHK LPLSITFDDS RESKTLNKK HFRAHKIDTK LNLWKKQKNT SKEFEDSILF DIYTDGYDA KGKEVPCNPF KDSNNKVVLQ QEKYNDIKKK KQKHYYVELK TDVLGNQHII	GIGSVGVGIL HRRVRLNRLF NMVKHGRISY TYGQLRGDFT FINRYLEILT FYPDEFRAAK EKAMGPAKLF TLDIEQMDRE QFRKANSSIF ARETNEDDEK QLATKIRLWH LANKVLVYAT KEYLLTEEDI VSVVRGQFTS LVSYSDELQ SYQVDSKFNR FMKIYKKDKS LKYKEEHGYI SVSPWRADV EGVDSSEFK YDKQKFEGGE KNEGDKPKLD	NKVTGEIHK EESGLITDFT LDDASDDGNS VEKDGGKHRL GKRKYHGP ASYTAQEFNL KYIAKLLSCD TLDKLAYVLT GKGWHNFSVK KLLTEEIYNP KAIQKIQKAN QQGERCLYTG ANQEKQRT SKFDVRKKFI QLRRHWGIEK DIETGELISD KISDATIYAT KFLMYRHDPQ RKYSKKGNGP FNKTTGKYEI FTLYKNDLLL ALIKVLGNVA F	NSRIFPAAQA KISINLNPYQ SVGDYQIVK INVFP TSAYR NEKSR TDYGR LNDLNNLTVP VADIKGYRID LNTEREGIQE LMMELIPELY VVAKSVRQAI KDEKDAAMLK KTISIHDLIN YQALDSMDDA ERNLVDTRYA TRDTYHHHAV DEYKESVFKA RQAKVGKDKA TFEKVIEPIL EIKSLKYYS LGLKYADLQF VKDTETKEQQ NSGQCKKGLG	ENNLVRRRTNR LRVKGLTDEL ENSKQLETKT SEALRILQTO YRTSGETLDN TETKKLSKEQ KSGKAEIHTF ALEHEFADGS ETSEEQMTIL KIVNAAIKEY AANQYNGKAE NSNQFEVDHI WSFRELKAFV SRVVLNALQE DALIIAASSQ PYQHFVDTLK DETYVLGKIK ENYPNKQINE KLGNHIDITP EKG GTYKIS LFRFLSRTMP KSNIS IYKVR	(SEQ ID NO: 24)
Parvibaculum lavamentivorans DS-1 AJN60020.1 GI: 757015976 WP_011995013.1	MERIFGFDIG QQRQRKMMR ELRRRGLEEG ANERATLKA VQSKFHPALK SWLSQQRML SALKALYKQ HPRKQEI DFGITGEQAA WEGWRRTNFP RKVVNNLIGL EDLIKNGIAN HIWPRSRSFD QGMVSAKGGT LWPDMPGPEAP DALTVACTHP IVVSHRVRKK ELDEIRDPRI QQLNLMAQTG PIVQRTRADG TDADHSTTR	TTSIGFSVID RQLRRRRIR LSAYEFGRAI LKNEQTTLGA SEEMRARISD EKLNNLAIAG GEPGAEKSLK VHERLWAADY QLQALKLPTG HRNQPTGEIL YGKPDRI PSRDDVEKWI NSPRNKTL GMSPGKVKRF VKVEAVTGQV GMTNKL VSGPLHKETT KEIVAAHVAG NGYADLGSNH ASFVMSLAAG PMPNPILKDD	YSSTQSAGNI KALNETLHEA YHLAQHRHFK WLARRPPSDR TIFAQRPVFW GNARPLDAEE FNLELGGESK GETPDKKRV WEPYSIPALN DKLPSPASKE VGRDVGKSKR LWKEGQERCP KDVNIEKGNR LAKTMPEDFA TAQLRKLWTL QLRDPRAEK YGDGTGDIKT RGGDPKKA HIAIYRLPDG EAIMIPEGSK AKKVSIDPIG	QRLGVRI GFLPAYGSAD GRELEESDTP KRGIAHARNV RKNTLGECRF RDAILSKLQ LLGNALEAKL ILSEKDRKAH LFLAELEKGE ERERISQLRN EREIISGIR YTGDQIGFNA MPFEAFGHDE ARQINDTRYA NNILADDGEK PALTPP KSGTYRQFVT PYPCVSPGGP KADFEIVSLF KGIWIVQGVW RVRPSND	ARDPDGTPLN WPVVMAD DPDVEDKEA VAEEFERLWE MPGEPLCPKG QASMSWPGVR ADMFGPDWPA REAAANSFVA RFGALVNGPD PTVVRTQNEL RNEKQRKKAT LFREGRYEVE DRWSAIQIRL AKQILAQ TRADHRHAI RADA RKKIESLSKG EIRKVR DASRRLAQRN ASGQVVL GLDPDEIKSA LEDYS KVSSLYLPDG	(SEQ ID NO: 25)
Corynebacterium diphtheriae	MKYHVGI VTRLASSGIA WKVRAELAAS	TFSVGLAAIE RRTRRLYRRK YIADEKERGE	VDDAGMPIKT RRRLQQLDKF KLSVALRHIA	LSLVSHIHDS IQRQGW RHRGWRNPYA	GLDPDEIKSA LEDYS KVSSLYLPDG	(SEQ ID NO: 26)

Table 4- Cas9 orthologs						
NCTC 13129 AJN60012.1 GI: 757015968 WP_010933 968.1	PSDAFKAIRE LQQSDYAREI QPGKNRALK NLTPKKEPEW VNSRIAPLVD DDDVAHAKLDS PSWTPTTPRI GFVTEKRARE QRQNCQCA GNTPFPAIWAK FQRATMDEEI ARRASGISGK LKQSQAHQRQ VVVMSNVRLR TREPGEFDPKE ELGSSFHHAR SMRQAEKKLR TIRRWVRDGF KLFSGDNVT	EIKRASGQPV QEICRMQEIG SDFQRYRIA VTIAEILGID WWKTASALEQ LHLPVGRAA GEPVGNPAVD MDGDMRRRA GSPITFSNSE NTSIEGVSVK DARSMESVA LKFFDGVGKS APQWREFTEG LGNGSAHKET GLPANPERHI VYKITSKKP DALATGNAEY FSPSKLRLRP VRRDSLGRV	PETATVGQMV QELYRKIIDV ALIGNLRVRV RGQLIGTATM HAMVKALSNA SEDTLVRRLR RVLKTVSRWL RNAKLFQEMQ MDHIVPRAGQ EAVERTRHVW MANELRSRVA RLDRRHHAID DAEHRAAWRV IGKLSKVKLS RVNGTHVYAG AFAMLRVYTI LGWLVDDEL LQMSKEGIKK LESTAHLPT	TLCELGLTKL VFAAESPKGS DGEKRILSVE TDDGERAGAR EVDDFDSPEG RMLSDGVDLY ESATKTWGAP EKLNVQGKPS GSTNTRENLV TDTGMRSTDF QHFASHGTTV AAVIAFTSDY WCQKMEKLSA SQLSVSDIDK DNIGLFPVSA DLLPYRNQDL VVDTSKIATD ESAPELSKII WKVQ	RGEGGVLSAR ASSRVGKDPL EKNLVFDHLV PPTHDTNRSI AKVQAFFADL TARLQEFEGIE ERVIEHVRE RADLWRYQSV AVCHRCNQSK KKFTKAVVER RVYRGSMTAE VAETLAVRSN LLTEDLRDDR ASSEALWCAL GSIALRGGYA FSVELKQPOTM QVKAVEAELG DRPGWLPVAVN	NO: 26)
Streptococcus pasteurianus WP_013852 048.1	MTNGKILGLD RGSRRNLNRK NEELFAALRT QIQLEKLEKY DYNKKITAEF FGILIGKCNF ESLVEFAKNT PYRKLKFNLE TEEQISEIIEK RLEKFKVNKK GDFDKIVIE LPDEVFHNK LPLSLSFDD LKQKGLGKKK ALRELKDKT LKLWEKQDNP TISSKGFED KIKDIYSQNG KSEDGKNDVK DITPEESRKN YHISQEKYDA RTMPNVNHVY YKVRTDVLGN	IGIASVGVGI KHRVKVRDL ISKRRGISYL GQLRGNFTVY IDDYVEILTQ YPDEYRASKA ATLGPALKLK SINIDDLRSRE VRKSQSTAFN SSKNTKTIDE PRDKNADDEK QLETKIRLWY LANKVLVYAW RDYLLTTENI VSVVRGQFTS MFVDYGNQV ILFSYQVDSK FDTFIKKYNK CNPFEYRRE VILQSINPWR IKEKEGIGKK ELKPYDKEKF KYFVKKKGDK	IEAKTGKVVH FEKYGIVTDF DDAEDDSTGS DENGAEHRLI KRKYHGPNG SYTAQEYNFL EIAKILDCKV VIDKLADILT KGWHSFSAKL KEVTDEIYNP KFIDKRKNEN VDKQTGEILS YNRKVSDATI DKTQFLMYQK NGLICKYSKK ADVYFNPETL SEFKFTLYRN DNVQELVEAL PKLDFKNNK	ANSRLFSAAN RNLNLPYEL TDYAKSIDEN NVFSTSDYK EKSRTDYGRF NDLNNLKVST DEIKGYREDD LNTEREGIED MNELIPELYA VVAKSVRQTI KKEKDDALKR KPISIQLVH YQVIDSMDAA ERNLVDTRYA SRETYHHHAV VSDDEYKELV YSTRKAKIGK DSLWENVIE GKGTPIKSLK KYELMGLKYS DLILIKDIAS GEADKVGRCI K	AENNAERRGF RVKGLTEQLK RRLKNKTPG EARKILETQA RTDGTLENI ETGKLSTEQK KGKPDLTTFE AIKRNLPNQF TSDEQMTILT KIINAADVKKY AAYLYNSSDK NSNNFEIDHI WSFREMMDYV SRVVLNSLQ DALIIAASSQ FQPPYQGFVN DKKEETYVLG VILRDYPTTK YYDKKLGNCI DLSFEKGTGN GEQEIYRFLS KGLNKPNI	(SEQ ID NO: 27)
Neisseria cinerea ATCC 14685 AJN60019.1 GI: 757015975 WP_003676 410.1	MAAFKPNPMN VPKTGDSLAA GLIKSLPNT ADKELGALLK HTFNRKDLQA VQKMLGHCTF ERATLMDEPY KAYHAISRAL DRVQPEILEA DHYGKKNTTE IHETAREVVG KDILKLRLYE NNKVLALGSE RILLQKFDED	ARRLARSVRR WQLRAALDR GVADNTHALQ ELNLLFEKQK EPTEPKAAKN RKSCLTYAQA EKEGLKDKKS LLKHISFDKF KIYLPPIPAD KSFKDRKEIE QQHGKCLYS NQNKGNQTPY GFKERNLNDT	SVGWAIVEID LTRRRARHLL KLTPLEWSAV TGDFRTPAEL EFGNPHVSDG TYTAERFVWL RKLLDLDDTA PLNLSPELQD VQISLKALRR EIRNPVVLRA KRQEENRKDR KEINLGRLE EYFNGKDNR RYINRFLCQF	EEENPIRLID RARRLLKREG LLHLIKHRGY ALNKFEKESG LKEGIETLLM TKLNNLRILE FFKGLRYGKD EIGTAFSLFK IVPLMEQGNR LSQARKVING EKSAAKFREY KGYVEIDHAL EWQEFKARVE VADHMLLTGK	LGVRVFERAE VLQAADFEN LSQRKNEGET HIRNQRGDYS TQRPALSGDA QGSRPLTDT NAEASTLMEM TDEDITGRLE YDEACTEITYG VVRRYGSPAR FPNFVGEPKS PFSRTWDDSF TSRFRPSKKQ GKRRVFASNG	(SEQ ID NO: 28)

Table 4- Cas9 orthologs						
	QITNLLRGFW NAFDGKTIDK DTPEKLRLL KRLDEGISVL KAFAPFPYKY DVFEKGGKY LYANDLIKLT GVKTALSFQK	GLRKVRAEND ETGEVLHQKA AEKLSSRPEA RVPLTQLKLL DKAGNRTQQV LVPIYSWQVA AKKNEFLGYF YQIDELGKEI	RHHALDAVVV HFPQPWEFFA VHKYVTPLEFI DLEKMNVRER KAVRVEQVQK KGILPDRAVV VSLNRATGAI RPCRLKKRPP	ACSTIAMQQK QEVMIQVFGK SRAPNRKMSG EPKLYEALKA TGVVWHNHNG QKDEEDWTV DIRTHDST VR	ITRFVRYKEM PDGKPEFEEA QGHMETVKS RLEAHKDDPA IADNATIVRV MDDSEFEKFKV KGKNGIFQSV	
AJN60009.1 GI: 757015965 St1Cas9+Sp Cas9	MSDLVLGLDI QGRRLARRKK SNEELFIALK PGQIQLEERYQ QEFNPQITDE IFGILIGKCT KNQIINYVKN EAYRKMKTLE FSQKQVDELV TRLGKQKTTS GDFDNIVIE TQLQNEKLYL NKVLTRSDKN ERGGSELDDK KVITLKSCLV LESEFVYGDY NGEIRKRPLI GFSKESILPK SKKLKSVKEL FELENGRKR QKQLFVEQHK EQAENIHLF GLYETRIDL	GIGSVGVGIL HRRVRLNRLF NMVKHRGISY TYGQLRGDF FINRYLEILT FYPDEFRAAK EKAMGPAKLF TLDIEQMDRE QFRKANSSIF SSNKTKEYIDE ARENQTTQK YYLQNGRDMY RGKSDNPSE AGFIKRQLVE SDFRKDFQFY KVYDVRKMIA ETNGETGEIV RNSDKLIARK LGITIMERS LASAGELQK HYLDEIIEQI TLTNLGPAA QLGGD	NKVTGEIIHK EESGLITDFT LDDASDDGNS VEKDGGKHRL GKRKYHGP ASYTAQEFNL KYIAKLLSCD TLDKLAYVLT GKGWHNFSVK KLLTEEIYNP QKNSRERMKR VDQELDINRL EVVKMKMNYW TRQITKHVAQ KVREINNYHH KSEQEIGKAT WDKGRDFATV KDWDPKKYGG FEKNPIDFLE NELALPSKYV SEFSKRVILA FKYFDTTIDR	NSRIFPAAQA KISINLNPYQ SVGDYQIVK INVFPTSA NEKSRDYGR LNDLNNLTVP VADIKGYRID LNTEREGIQE LMELIPELY VVAKSVRQAI IEEGIKELGS SDYDVHIVP RQLLNALIT ILDSRMNTKY AHDAYLNAV AKYFFYSNIM RKVLSMPQVN FDSPTVAYS AKGYKEVKKD NFLYLASHYE DANLDKVL KRYTSTKEVL	ENNLVRRTN LRVKGLTDEL ENSKQLETK SEALRILQT YRTSGETLD TETKKLSKE KSGKAEIHTF ALEHEFADGS ETSEEQMTIL KIVNAAIKEY QILKEHPVEN QSFLKDDSID QRKFDNLTK DENDKLIREV GTALIKKYPK NFFKTEITLA IVKKEVQGT LVVAKVEKKG LIIKLPKYS KLKGSPE YNKHRDKPIR DATLIHQSI	(SEQ ID NO: 29)
Campylobac ter lari Cas9 BAK69486. 1	MRILGFDIGI RSSRRRLKRR YELRYKALTQ KNNALKLENY LEKELKLILE EEEKRACKNS DKGSITYKKF GVHSLSRQEL YINLSFKALG FAHELSPV KIEKEQKENQ KISIEHLKDE AFGKNIEKWS IATLIAKYTK TWGFDKKDRN ELTSNDYKHQ HSENKIIDKC KFYAIPYAM NDLILQKKN NAKEGSVKVE GLR	NSIGWAFVEN KARLIAIKRI NLETKDLARV QSVGEYFYKE KQKEFGYNYS YSAWEFVALT RSCINLHESI DQISTHITLI MILPLMREGK NRAISEYRKV AVNAWALKEC KALEVDHIYP KIQTALQNL EYLNFLLLSE NHLHHALDAI VKFFPEPKSF SYNSKEGLQI DFALGILPNK MQEPEFAYN SLGIQNLKVF	DELKDCGVRI LAKELKLNK ILHIAKHRGY FFQKYKNTK EDFINEILKV KIINEIKSLE SFKSLKYDKE KDNVKLKT RYDEACEIAN LNALLKKYK ENIGLKASAK YSRSFDDSF YKKKNKILDE NENANLKS IVAYSTNSII REKILSKIDE ALSCGRVRKI IVITGKDKNN DFSISTSSIC EKYIITPLGD	FTKAENPKNK DYVAADGELP MNKNEKKSND NFKIRNTKD AFFQRPLKDF KISGEIVPTQ NAENAKLIDF EKYNLSNEQI LKPKTVDEKK VHKIHLERAR NILKLKLWKE NKVLVFTKEN NFKDKQQEDF KGSKIHVQTI KAFSDFRKNQ IFVSKPPRKR GTKYVENDTI NPKQWQTIDE VEKHDNKFEN KIKADFQPRE	ESLALPRRNA KAYEGSLASV AKKGKILSAL NYNLCVSSD SHLVGACTFF TINEVLNLIL RKLVEFKKAL NNLLEIBFND DFLPAFCDSI DVGLSKKARE QKEICIYSGN QEKLNKTPFE ISRNLNDTRY SGMLTSVLRH ELLKARFYAK ARRALHKDTF VRVDIFKKQN SYEFCFSLYK LTSNQKLLFS NISLKTSKKY	(SEQ ID NO: 30)
AJN60010.1 GI: 757015966	MDKKYSIGLD LLFDSGETAE LEESFLVEED LRLIYLALAH	IGTNSVGWAV ATRLKRTARR KKHERHPFIG MIKFRGHFLI	ITDEYKVPSK RYTRRKNRIC NIVDEVAYHE EGDLNPDNSD	KFKVLGNTDR YLQEIFSNEM KYPTIYHLRK VDKLFIQLVQ	HSIKKNLIGA AKVDDSFHR KLVDSTDKAD TYNQLFEEENP	(SEQ ID NO: 31)

Table 4- Cas9 orthologs						
SpCas9+St1 Cas9	INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKS NFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASQAS FIERMTNFDDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRDNAS LGTYHDLKLI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVDELVKV MGRHKPENIV IEMARETNED DEKKAIQKI Q KANKDEKDAA MLKAANQYNG KAELPHSVFH GHKQLATKIR LWHQQGERCL YTGKTISIHD LINNSNQFEV DHILPLSITF DDSLANKVLV YATANQEKQ RTYPQALDSM DDAWSFRELK AFVRESKTLN NKKKEYLLE EDISKFDVRK KFIERNLVDY RYASRVVLNA LQEHFRAHKL DTKVSVVRGQ FTSQRLRRHWG IEKTRDITYH HAVDALIIAA SSQNLNWKQ KNTLVSYSED QLLDIETGEL ISDDEYKESV FKAPYQHFVD TLKSKEFEDS ILFSYQVDSK FNRKISDATI YATRQAKVGK DKADETYVLG KIKDIYTDG YDAFMKIYK DSKFLMYRH DPQTFEKVIE PILENYPNKQ INEKKEVPC NPFLKYKEEH GYIRKYSKKG NGPEIKSLKY YDSKLGHNHID ITPKDSNNKV VLQSVSPWRA DVYFNKTTGK YEILGLKYAD LQFEKGTGTY KISQEKYNDI KKKEGVSDS EFKFTLYKND LLLVKDTETK EQQLFRFLSR TMPKQKHVVE LKPYDKQKFE GGEALIKVLG NVANSQGCKK GLGKSNISIIY KVRTDVLGNQ HIIKNEGDKP KLDF					
SpCas9 inactive AJN60011.1 GI: 757015967	MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRKNRIL YLQEIFSNEM AKVDDSFHHR LEESFLVEED KKHHRHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSDDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKS NFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASQAS FIERMTNFDDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRDNAS LGTYHDLKLI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVDELVKV MGRHKPENIV IEMARETNED DEKKAIQKI Q KANKDEKDAA MLKAANQYNG VENTQLQNEK LYLYYLQNGR DMVVDQELDI NRLSDYDVDA IVPQSFLKDD SIDAQVLTNS DKARGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS KLVSDFRKDF QFYKVREINN YHHAAAYLN AVVGITALIKK YPKLESEFVY GDYKVYDVRK MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF ATVRKVLSP QVNIVKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPPK YGGFDSPTVA YSVLVVAKVE KGKSKLKS SV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDIIKLPK YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA PAAFKYFDTT IDRKYRSTK EVLDATLIHQ SITGLYETRI DLSQLGGD		(SEQ ID NO: 32)			
AJN60013.1 GI: 757015969	MTQSERRFSC SIGIDMGAKY TGVFYALFDR EELPTNLNSK AMTLVMPETG PRYVQAQRTA VRHRLRGQKR YTLARKLAFL VVDDMIKKQE KRLTDEEWKR GREALSGLLK RRGYSRPNAD GEDLTPLENV RADVFAAHPA FSTYFSEVRS LAEQWEEFTA NISNVEKFLG DPNIPADKEF IEFVAEGLI DKTEKKAYQS		(SEQ ID NO: 33)			

Table 4- Cas9 orthologs						
WP_005430658.1 Sutterella wadsworthensis 3_1_45B	ALSTLRANAN ERLARLLGNI PAKDQNKQHL TLLLSPEKLT NGHEPLPTLA LENCIGGQNV KKILPLLVA FGGGFNIAYN DEQKRKFANP GETVRAAQCS NGIVDVSI RIKKASRGTC SRGNQLKKNQ LKFFDLLNEH IKNLANKIRE KPDIIQPIASH IHLQAKPQEE ERCGAIEVSG AALQPLSAEE KLFQLKVELK LTSKLARIWK YQVHAINAKK TPMSEWRKV ITSPLSLPAS FWYIVVSSNK	VLTGLRQMGH SNLQLRAERW ELIKQIENSE RQYGEIWK YQLSYALQRA KTFLDCARRY ILQTDETTQ TAQYREVNKL FSLAQFYTLI RLPAETARPF EENKFEFSAS PYTGDRLAEG RYSLSDLKAN EQDCVRHALF ELQNWCKTTN SIDALCSFAV KSHFDSVAIF KQPKELLEML KRLAALLDAL GEKSFKLNGS RPVMDLAHA YRGFASAGSN AEDNLSIWIA FKVDKPAEFQ KMNESYNNVS	KPRSEYFKAI YFNAPDIMKD DIIETLCTLD SARLTSAEPT FDRSKALDPY YREADDAKVG KFLDEIWRKQ PRNAQDKELL ETEVSFGSAT DGLVRRLLVDR VADLKKNKRV GEIDHILPRS YRNEIFKTSN LDDGSEARDA NRLHFAQAAAT GSADAERDQ KEGIYAEQFL APFFNKPVG RYCTSRKSLM LTLPVKQDWL PVRREFSLPA VDWSKGILFN PGTEGRRYVR KAVGTELSSEL KS	EADLKKDSRL RGWEPDRFKK PNRTIPPYED LAPAAEILER ALRALAAGSK LWFDNADGLL IKGRETVASR TIRDRAETA TLAVHLENAW QAWETIAKRV KDKMLSEAEK LIKDARGIVF IAAITAEIED VLELLATQRR NVSDAKNLRL GFDYLDGKTV PIFTLNEKI LSAHATYRIL SLFMAANGKS RICDSPELAD IDNPSGGFRI ELQHENLTEC VETTFIQASH LGQPRSEIFI	AKINEAFGGA TLVRAFKFFH QNNRRPPLDQ STDRKSRVAV SNKLTSA ERSDLHPPMK CARIETVRKS DFIAANLGLS RMTIKDAVIN TDIQSKVDFS LETRWLKNE NAEPNLIYAS VVTKLQQT TRVNGTQIWM KLAQNQPD LGLYPQSC IGYETLNAK KKPAYEFLAK LKKREDVLK AFGKPCSADE RRTNLFGNEL GGRFITSADV WFEQSVENWA ENVGNAKHIR	
AJN60014.1 GI: 757015970 WP_011212792.1 Legionella pneumophila str. Paris	MESSQILSPI QLSQAQRRAT LNNRGYTYVD FRKILVSKVE KSDITKDNQL QTFGNELFRM TIPPYEARTN LEHTKLDRK AQGLPANLI AFSLCELSNI GRTSLKSKCK DIIQAIQSHL NYWRSQKTEI IKHIPDNSSL EEKFQRIINA LIYCSSQGNR KKYISFHLT FLGKQIMEFL RQQSFPSHAI SKEKYNKPN NKEKLFLLK HKEIVTPDDT VLGSFKHTIA DNNPNSDIPN PLYQLQTIDD AYATFDNWLT ALMIKPSDSI FESDSTPQWI	GIDLGGKFTG RHRVRNKKRN TDLDEYIKDE EKKDDKELKN DSIKKKIPSV LKNFRHLKGS TGMEKDQSL RIISPSKQDE ETQKEMETHF NPPRKQKILP EIEEARKNSG GHNDSQALIY DPEISYASRL LIPIYLEQNR SMNICPYKGA EKKEEHYLL PEQQKAARHA STLADSKQLQ DATLTMSIGL SSTPLFKDSL TYSTKNPGES TVCHFINSRL LPATKDWERL NGHNIKPKQH TPSMGIQINE LPVSTFKPEI DDPLNMPNEI QTLYVTQLKK	VCLSHLEAFA QFVKRVALQL TTINLLKELL AVKNIKNFIT CLSNLLGHLS QESLAVRNLI LNPEKLNLY KRDSYILQRY NSSLVSVLIQ LLVGAILSED NAFKIDYEEA HNPFSLSQLY PADSVRPF FEFEESFKKI SIGGQGEIDH HLSPLYLKHQ LFLDYDDEAF LEFSIKQITA KEFPQFSQEL YAERFIPVWV LQELQAKSKA YYTKKESITV FNHPNFLALK KAVRKVFSLP DRLVKQEVLM IKLEMKPHSK VCKNKLFGNE QP	ELPNHANTKY FQHILSRDLN PSESEHNFI GFEKNSVEGH NLQWKNLHRY QQLEQSQDYI PNWRNLIPGI LDLNNKKIDKF IASAYNKERE FINNKDKWAK LNHPHESNNK TILETKRDGF VLARMMQRLA KGSSSDKTLE IYPRSLSKKH FGTDNVSDIK KTITKFLMSQ EEVVDHRELL DNSWFINHLM KGETFAIGFS KWLYFPINKT KILKEPMPVL ANPAPNPKEF VIPGNAGTMM TRRYIRITQS LKPRDGKMKI	SVILIDHNNF AKEETALCHY WFLQKMQSSE LAKNPKQFDE SILEKTPPEI IDAHPFLEKD KIKKQLSFLG DAAQGIWFDN FKIFWNTHKI ALIKIIQTIP HKNCVAVTCE YEIAMAKWEQ QAIEKQNIQW FGVIFNSEVN NFIQSNVANI QKARVNGTQK SKQEPKLVKS PDEVHLNPVR EKDLFEIKPS LALEFLHHYF SVKFESSKKN NEFIRKYFLS RIRRKDNKGQ IDGINNSEGGQ LADFIKTIDE VSTGKIVTYE	(SEQ ID NO: 34)
AJN60015.1	MKKEIKDYFL EVRRLHRGAR DKTILQENTL	GLDVGTGSVG RRIERRKKRI FNDKDFADKT	WAVTDTDYKL KLLQELFSQE YHKAYPTINH	LKANRKDLWG IAKTDEGFFQ LIKAWIENKV	MRCFETAETA RMKESPFYAE KPDPRLLYLA	(SEQ ID

Table 4- Cas9 orthologs						
GI: 757015971 WP_002681 289.1 Treponema denticola ATCC 35405	CHNIIKKRGH EILKDSSLKN PDLKDAEKNS VIGDEQYLSF NNNYSYGVGV TEIETGTFLP LSHSEKIIML NFFDHIDKEK LQIIIDGKNI VIILGIDKEC EGKTILKTKI GFSEPVNIIT GLVKPLFLSP RTKTRLKILQ YTQLGKCMYC CNKNKEDKYP TAKFIARQLV KCREINDFHH YYKVFYDVVK QTIMKGLGQ LETIPLYLVK PCHITGKTND SPYEDLSFRS DTIYKKRPNS LQHIGGSKYS	FLFEGDFDSE SEKQSRLNKI ISFSKDDFDA AKVKIYEKHK CKTKSKKLI KQISKSNAEI LTFKIPYYIG TAEAFITSRT CDIKLKQKIY TSSLKSYIEL KAEYGYKCS AMRETQNNLM SVKMLWQTL DLYNNCKNDA GKPIEIGHVF LKSEIQSKQR ETRQATKVAA AHDAYLNIVV RNNITAWEEK HPLKKEGPF DIQKDQDVLK SFLLRPAVQF YIKENLWKKT ATIDILVKGK GVAKIGNKIS	NQFDTSIQAL LGLKPSDKQK LSDDLASILG TDLTKLKNVI NNSVNQEDFY PYQLRKMELE PINDNHKKFF NFCTYLVGES EDLFKKYKKI KNIFGKQVDE EQIKKILNLK ELLSSEFTFT KLVKEISHIT DAFSSEIKDL DTSNYDIDHI GFWNFLQRNN KVLEKMFPE GNVYNTKFTN KTIITVKDML NISKYGGYNK SYLTDLLGKK CCSNNEVLYF KNDEIGEKEF EKFKSLIIEN SLDNCILIIQ	FEYLRDMEV KAITNLISGN DSFELLKAK KKHFPKDYKK KFLKTILSAK KILSNAEKHF PDRCWVVKKE VLPKSSLLYS TQKQISTFIK ISTKNMLEEI FSGWGRLSRK ENIKKINSKF QAPPKKIFIE SGKIENEDNL YPQSKI KDDS FISLEKLNRL KIVYSKAETV NPWNFIKEKR KRNTPIYTRQ VSAAYYTLIE EFKILVPKIK KKIIRFSEIR YDLLQKKNLE QFEVILEILK SITGIFEKRI	DIDADSQKVK KINFADLYDN AVYNCSVLK VFGYNKNEKN SEIKEVNDIL SFLKQKDEKG KSPSGKTPW EYTVLNEINN HEGICNKTDE IRWATYDEG FLETVTSEMP EDAEKQFSYD MAKGAELEPA RLRSCKLYLY ISNRVLVCS TRATPISDDE SMFRNKF DNPKIADTYN AACKGELFN YEEKGNKIRS INSLLKINGF SQREKIGKTI IYDMLLTCHK LFSATRNVS DLLKV	NO: 35)
AJN60016.1 GI: 757015972 EFE28295.1 Filifactor alocis ATCC 35896	MTKEYYLGLD RLQRGNRRRL NDFKYPLFIE IIKTRGHFLI EILKNRSIAK GDVCKFLRVS MYDWNILVDI MFNDEKEAGS KEDLEVLTM YSLETETDKD GREDRIYPWN YMLVNLNINV DIQIDDLSGF IYGNDKEMLK SDVSTGETFD TYDSTVKEMF VKKRTKSRKA GKCMYSGDDI KKSNELLSED IARQLVETRQ VNDYHHAKDA VINGEVIWEK FNATIQNKNG HIIGVPIYIA PLRIRGENEV ENRDHITHEK IDKINVINKM VNQSVTGLYE	VGTSNSVGWAV ERKKQRIDL KDYSIDIEYK DGDLSAKQL SEKVKKLKNI KEELEIDSFS LETEEYISFA YTAYVGKLK IEECKNHTLL GYSVQKIES MEEIIDFEKS KVRGKKLPTS DKDFKTSLS RVIRANYSNQ IITAMWETDN LSPENKRAVW QLELEYAACE DINELIRGNS IQKKMHFWL STKAIADIFK YLNIVVGNVY CTYHEDTNTY NSTVSLKKG NMLEHSPSAF DTSFKRAIQL MNQLYEVL LNLLRCDNDT NRREL	TDSQYNLCF QEIFSPEICK EFPTIFHLRK RPILDTFLLS FEISDELEKE FSEGKYEDDI KVKQYEKHK NNKKYWIEKK PIQKNKDNV IFRFRIPIYV NENFITRMTN LKQKVFEDLF YLDFFKKQIFG LTEEQMKKIT NLMQILSKKF QTIQVAEEIK EDCRELIKEI KWDRDHIYPQ SKYDRLTRKG QIYSSEVVYV NKKFTSNPIQ DGGTLDRIK DVKKYGGYFS LEYCEQKGYQ KLDQKNYELV KMKKFNKKGM KADLSLIELP	KKKDMWGIRL IDPTFFIRLN HLIESEKQD LQEEQNLSVS EKKASAVIE VKNLEEKVPE NLRLRLDIIL RNPEEFYKSL IPHQVHEVEL GPLSTRHQEK KCTYLIGEDV ENKSKVTGKN EEIEKESIQN GFQYSGWGNF TFMDNVEDFN KVMGCEPKKI EDRDERDFNS SKIKDSDIDN SKYDRLTRKG KSSLVSDFRK WMKKNRDTNY IVERDNILYT ANTSYSFLIE NVRILVEKIK RNIEKFLEKY ADPSDRIEKS KNAGSFVVKK	FESANTAKDR ESRLHLEDKS IRLIYALHN LSENQKDEYE NFCFKFVGNK KVYLFEQMK KYCTKDEYNR GKLLDKIEPL KKILENAKKY GSNVWMVRKP LPKHSLLYSK LLEYLQIQDK MIEDIIKWIT SKMFLKGISG SGKVGKIDKI FIEMARGGEK MKLFLYYTQF LVLVNKTYNA DFTDEELSGF KPLNYLKSRR SLNKVFEHDV EYAYCEKGEL FEDKKGDRAR KNSLLIINGY VEKKGNYPID KPKFKLEDL NTIGKSKILL	(SEQ ID NO: 36)
AJN60017.1	MGRKPYYLSL RRQFTSRRR KNKSLSEVLS	DIGTGSVGYA KNRRIKRLGL FLGYESKKYP	CMDKGFNVLK LQELLAPLVQ TIYHLQEALL	YHDKDALGVY NPNFYQFQRQ LKDEKFDPEL	LFDGALTAQE FAWKNDNMDF IYMALYHLVK	(SEQ ID

Table 4- Cas9 orthologs						
GI: 757015973 WP_014613 259.1 Staphylococcus pseudintermedius ED99	YRGHFLFDHL EILKDNEMTK LKGANQSHTF SMAMSKVAAY TPNDQTFSSL LKVLNTTDNA IPYYVGPLVN TNKCSYLINE IKLFAVEHIF KFASKLSSYQ PELTSKQINQ LTNDVYGFQN LTSIFGEPEK IIDVANKLNN VDHIFPRSF LNKAGLISDS YPNTKVIPMK AYPNVDLDF ERLISKIKLD EVVYKGLTPY LALYLAQREN ELTVEQQLSL RVRTFFSESN GKGKDVKIAY	KIENLTNNDN NDRAKRVKNM ADNYEENLTP DKFRNELKQV CLFDQYLIRP SIPMQINLYE DHTASKFGWM DVLPKNSLLY KKYKTVSHSK DMTKIFGDIE LKKLNYSWG FIKEENQVQS IIMEFATEDQ EQLQOEKLWL KDDSIDNKVL KLHKLMKPEF AKMVSEFRKK NFKWEKVREK MNHFKINYSR QTYVVAIKSV KDEVLDQIV YNVMNNKETN QTHEDFIKAL TSISGLKTTK	MHDFVELIET EKKLEQFSIM FLTVEQSEFI KDIVYKADST KKQYSLLIKE AETILRNQQK ERKSNESSIKP QEMEVLNELN FLEIMLNSNH GKRAQIEEII RLSEKLLTHA NKIQHQDIAN QKGGKKQKSRK YLSQNGKCMY VIKKMNQTKG TAMDKEGFIQ FDIPKIRQMN WKALGEFNTK KLANIPQQFY NKKGKEKMEY YSLNKGDLly VEKLLIEYDF DELFKVVTAS PKSLFKLAES	YENLNNIKLN LLGLKFNEGK ERANKIYLSL RTQFKKIFVS LKKIIPQDSE YHAEITDEMI WNFDEVVDRS ATQIRLQTD RENFNMHGEK QWITIFEDKK YQGHSSIELL LTSPPALKKG QLWDDNIKKN SGQSIDLDAL DQVPLQFIQ RQLVETRQIS DAHHAIDAYL QKSRELFFFK NQTAVSPKTA QMDHYVDFD INNHPCYFVS IAEKVINEYH ATRSDDKIGSR RNEL	LDYEKTKVIY LFNHADNAEE TLQDILKGGK SKKSLKQYDA LYFEAENDTL EKVLSLIQFR KSATQFIRRM KNRKYRMMPO LSIFGTQDDK ILVQKLKECY RHSDEFMEI IWSTIKLVRE KLKSVDEYKY LSPNATKHYE PYERIAWKS VHVRDFLKEE NGVVYHGAQL KLEKMEVSQ ELKYESNKS YKFQNGNEKE RKEVINAKQF HYLNSKLKEK KNSMTHRAFL	NO: 37)
AJN60018.1 GI: 757015974 WP_014567 561.1 Lactobacillus johnsonii DPC 6026	MTKIKDDYIV AERLFRTH DKRKTVSSIV HHIIKYRGNF SNTAEIEKVL AVLGYKTRFD AILLTIKELF IDRKKAKELA DDSRASKEIK PFLKEINPVS QNFAWMIRKE ANSLLYQKFT NYLKENHNLV DFEKIEWST LLAQLHDHNG EDILNNAYTS PKRSQTRGSK QLGRDAYTGE SDNVPVKLFG KYKSAGFIRR DLYKSREVND DKEKAI FNKT TRDQEMFKMT IVRIDKKKGT NGKRDKKITS SMKAITNNFD EKFIKLSLED LSENAKLIYQ	GLDIGTDSG RRIKRRRWRL FPTSADKKF LYNTSVKDFK KDKQIFKRDK TIALKEISKD NEVTNLNGIVE LAYDLYVNNR KKIELD SFMP SHLKEAPYKL EGRITPWNFD VLNELNNIRI KVEIKGLADE IFEDKSIYKE QTIIEQLWDS PANKKAIRQV LQKVYKDLST PINIDEIQKY NEMAANLGMT QLVETSQIIK YHHAIDAYLS RKFSFISQLL VYPRFSDTV EYKILGIPTR FEIVKSKIPY KDSDENDALI KKDTILKVL SPTGLFKKSV	WVAMNSNNDI KLLEEFFDPY YDDYPTIYHL ASKIDVKSSI VKKIAELFAI ELSDWNFKLS DGNTLSESMI HGQLLQAKKK KORTNANGVI DELIRFRVPY QKVDRIESAN NGKRISVDLK KKFNSGLTTY KLRSIDWLNE QNNFMQIVTQ IKVVDDIVKA ELASKTIAEE DIDHILPQSF IRKMWEWKN LVSTILQSR AICGNLLYQN KNKSENSKEI KAPRNLI PKK ELVNLKKA EK KQVIQDGDKK KAYDEILD KV GLHDNAVMTK KISDL	LKLQGKTAIG MAEVDPYFFA RYKLMTEDEK EKLNELYENL KTDNKEQSKR DIDADSKFEA NKYNDHRDDL LGKIKPRSKE PYQLQQLELD KF I KRM TTKD QE IYENLFKK NRFKNLNIFD KQINALSNIR ADFKDAIAKA ASGKVPKQIA LNEAIKDKKL IKDDALDNRV IGLISKT KYN PNT E IITVKA YPNLRPFVY AKKLKRAYQF MGMSPD IYGG EDHYKSYLKE FMLGSSTYVY DKYLP LFDIN IPTIGLSTPL	SRLFEGGKSA RLKESGLSPL FDLREVLAI GLDLNVEFNI IKDISKQVAN LMGNLDENEQ KLLKEVIENH DFYKVVNKNL KIIENQSKYY STKDIQT KKN TYLFGEDVLP HTTVTVKKLE NQIDDLKYRN LQGWGRLSKK NQNLVATSV IEFTRDADEN VQDKYYLYFM LVSRAVNNGK NLLTDPDHIN KYNHYLREKF GQYKFFSSDP KYMLVSRETE YTNNSDAYMV ILTPRI LYNK NAKQTLTSTE KFREKLHSGR GFMQFPNGVI	(SEQ ID NO: 38)
Mycoplasm a	MNNSIKSKPE RRSFRGVRRL HNTVLNLKSE AKYFDKYGY	VTIGLDLGVG IRRRKYKLKR ALNAKIDPKA KGIIDSKEDN	SVGWAIVDNE FVNLIWKYNS LSWILHDYLK DNKLEELTK	TNIIHHLGSR YFGFNKEDI NRGHFYEDNR YKFSNKHWE	LFSQAKTAED LNNYQEQQKL DFNVYPTKEL EVKKVLSNQT	(SEQ ID NO: 39)

Table 4- Cas9 orthologs						
gallisepticu m str. F AJN60022.1 GI: 757015978 WP_014574 789.1	GLPEKFKEEY NIWDKTIGKC NQEFKKAYLN ENEEKLEKED LEDQYATLTK QKQKDGFLFKL KNNDKGWNFE GVKRILREAT KLIKKNNGDKI LKEIAFDDIF PYEFISSGNA SKYDIGFLAR TSFLRKNFDD YKLFKNTDGS DSNIERKARY ESAKENKNNSK INLAEQIFNE TAIIYDDKKN INSLALMIMR FVKYLLKANEI NLIETHEYKEN KNRIDEILNS	ESLFSYVRNY NIFPDEYRAP KLLDLLIKTN HKWKLKGLKL INFLQSLFVY FEQTDKDEK AIKNFDQKFI KVFNAILKQF SEGLKALGIS TKTEKFEIDH GIKWEDYEAY NLNDTRYATI SSYAKKDRDK WKKIDPKTGV SRKIENKTNI VKRQFVYRKL YTENKKIKSQ IVHRIKRLKM SIDPTAKKQY GDEYKPWRVL DDSDSKKKKK KLGLDKIVK	SEPGGSINSV KNSPIAMIFN GEKPIDARQF NTNGKIQYND LGKHLRYSNR ILAQTHSLST DITKKNNNLS SEEDYVTKVV EDEIKDILKS IIPYSISFDD CRKFKDGDSS VFRDALEDYA NIHHAVDASI VTEVTDENWK SLFNDTVYSA VNVSLNNDK NVFEKYMLDL LSSELKENKL IRVPLNTLNL TSGTLLIHKK ANRFLMTLST	SPYGIYHLDE EINELSTIRS KKLREETIAE LSSLAKFVHK VDSANLKEFS KAMLLAITRM LKQNKRYLDD IELARELSEE PTKSYKFLW SSSNKLLVLA NNHLVEDKPM ISIFSNETKT QIRVRNQVSE KKVGYEDQIK LADLFAEKED TKEFPEKFSE SNVIIRSKNQ HLGDHDFDLH DKKLMYISSF ILNDYILLDA	KEGKVQKYN YSIYLTGWFI SIGKETLKDV LKQHLKLDLFL DSNKLFERIL TNLDNDEDNQ RFINDAILSP KELENTKNYK LQQDHIDPYS ESNQAKSNQT FAKMMKTDT FKVVCINGSV LFNQLTQFAD IAKVIEKYIQ RKNLKTLDIH ILMYRANPWV FLVKSMLRNK SGTKLSYQDT NMDAYLKPKP QNLNDVIEIK KDNFDILGLS	
AJN60023.1 GI: 757015979	MRILGFDIGI RSSRRRLKRR YELRYKALTQ KNNALKLENY LEKELKLILE EEEKRACKNS DKGSITYKKF GVHSLSRQEL YINLSFKALG FAHELSPV KIEKEQKENQ KISIEHLKDE AFGKNIEKWS IATLIAKYTK TWGFDKKDRN ELTSDNYKHQ HSENKIDKC KFYAIPIYAM NDLILLQKKN NAKEGSVKVE GLR	NSIGWAFVEN KARLIAIKRI NLETKDLARV QSVGEYFYKE KQKEFGYNYS YSAWEFVALT RSCINLHESI DQISTHITLI MILPLMREGK NRAISEYRKV AVNAWALKEC KALEVDHIYP KIQTLAQNLP EYLNFLLLSE NHLHHALDAI VKFFEPFKSF SYNSKEGLQI DFALGILPNK MQEPEFAYYN SLGIQNLKVF	DELKDCGVRI LAKELKLNYK ILHIAKHRGY FFQKYKNTK EDFINEILKV KIINEIKSLE SFKSLKYDKE KDNVKLKTVL RYDEACEIAN LNALLKKYGK ENIGLKASAK YRSFDDSF YKKNKILDE NENANLKS IVAYSTNSII REKILSKIDE ALSCGRVRKI IVITGKDKN DFSISTSSIC EKYIITPLGD	FTKAENPKNK DYVAADGELP MNKNEKKSND NFIKIRNTKD AFFQRPLKDF KISGEIVPTQ NAENAKLIDF EKYNLSNEQI LKPKTVDK VHKIHLLELAR NILKLKLWKE NKVLVFTKEN GTSKYVENDT NPKQWQTIDE VEKHDNKFEN KIKADFQPRE	ESLALPRRNA KAYEGSLASV AKKGKILSAL NYYNCVLSSD SHLVGACTFF TINEVLNLIL RKLVEFKKAL NNLLEIEFND DFLPAFCDSI DVGLSKKARE QKEICISYGN QEKLNKTPFE ISRNLDTRY SGMLTSLVLRH ELLKARFYAK ARRALHKDTF VRVDIFKKQN SYEFCFSLYK LTSNQKLLFS NISLKTSSKKY	(SEQ ID NO: 30)
AJN60025.1 GI: 757015981	MSDLVLGLDI QGRRLARRKK SNEELFIALK PGQIQLERYQ QEFNPQITDE IFGILIGKCT KNQIINYVKN EAYRKMKTLE FSQKQVDELV TRLGKQKTTS GDFDNIVIE LPHSVFHGHK LPLSITFDDS	GIGSVGVGIL HRRVRLNRLF NMVKHRGISY TYGQLRGDFT FINRYLEILT FYPDEFRAAK EKAMGPAKLF TLDIEQMDRE QFRKANSSIF SSNKT KYIDE ARETNEDDEK QLATKIRLWH LANKVLVYAT	NKVTGEIIHK EESGLITDFT LDDASDDGNS VEKDGGKKHRL GKRKYHGP ASYTAQEFNL KYIAKLLSCD TLDKLAYVLT GKGWHNFSVK KLLTEEIYNP KAIQKIQKAN QQGERCLYT ANQEKQRT	NSRIFPAAQA KISINLNPYQ SVGDAQIVK INVFP TSAYR NEKSR TDYGR LNDLNNLTVP VADIKGYRID LNTREGEIQE LMMELIPELY VVAKSVRQAI KDEKDAAMLK KTISIHDLIN YQALDSMDDA	ENNLVRRRTNR LRVKGLTDEL ENSKQLETKT SEALRILQTQ YRTSGETLDN TETKKLSKEQ KSGKAEIHTF ALEHEFADGS ETSEEQMTIL KIVNAAIKEY AANQYNGKAE NSNQFEVDHI WSFRELKAFV	(SEQ ID NO: 41)

Table 4- Cas9 orthologs						
	RESKTLNKK HFRAHKIDTK LNLWKKQKNT SKEFEDSILF DIYTQDGYDA KGKEVPCNPF KDSNNKVVLQ QEKYNDIKKK KQKHVELKP TDVLGNQHII	KEYLLTEEDI VSVVRGQFTS LVSYSQDQLL SYQVDSKFNR FMKIYKKDKS LKYKEEHGYI SVSPWRADV EGVDSSEFK YDKQKFEGGE KNEGDKPKLM	SKFDVRKKFI QLRRHWGIEK DIETGELISD KISDATIYAT KFLMYRHDPQ RKYSKKGNGP FNKTTGKYEI FTLYKNDLLL ALIKVLGNVA	ERNLVDTRYA TRDTYHHHAV DEYKESVFKA RQAKVGKDKA TFEKVIEPIL EIKSLKYYS LGLKYADLQF VKDTETKEQQ NSGQCKKGLG	SRVVLNALQE DALIIAASSQ PYQHFVDTLK DETYVLGKIK ENYPNKQINE KLGNHIDITP EKGTTGYKIS LFRFLSRTMP KSNISYKVR	
WP_002664 048.1 Bergeyella zoohelcum ATCC 43767	MKHILGLDLG QAQTKNADRR DPNKIDKITI QLRGYAGGSL KKLNRRRAIIV LPNKTNWRKK TILRARYESE YRELGLEKGL FQYKQVWEQI QNKEITFSA GELWDVLGLD IVDDNAEETA LQSKILKLLN TAKEYSHDEL KPNEIRVELA ANIEKVKLWE TNKVISEKSV RKNLLATSIP YLRNHWGLTD KEVLFEQEL IVACTEPAHI ENERTEIFTQ KAGDRQIKLR SPFLSGFIAN ISTVKIYYKD LEGEIKTKKT YFEERNKAKL NIYVVQKFSK LEIDRLGNIV	TNSIGWALIE TNRGARRLNK LPISKKQEQL EPEKEDIFDE ETEEGNFEGS MENIENQLKE FEAIWNEQVK KYIINKQVVF NKLIVNTKIE IFKKLKAEPD SINRQIELWN IRISKIKFAR ENVEDPFAKA YKSYKEINLL RDLKNSAKER AQRHLSPYTG NQEKANRTAM EDPVQRQIKD KFKLLLKERY TREEFIKEYK KRLNDLNKVL IEKFRAIEMP GQLHEGTLYG HLKEYNNKKE PSKNKKKKDE SQIKRLYDII LFTLKQGDFV KQIYFIKHTI KVIKR	RNIEEKYGI RYKQRRNKLI TAFDLVSLRV EQSKDKKNKS TFLENIKVG KSKEMGREFY HYPFLENLDK YQRELKDQSH AGTNRKGEKK LREGIDFLNG ILYNEKGNEY AYSSLSLKAV AQTYLDNNQS KQGDLRNPLV ATIHKRKNKD QPIPLSDLFD EYFEVGSLEY TQYIAIRVKE EALLESEKFL ENYIRYKKNK QDWLVEHKSE WKGFPQVEQ ISQGKEAYRI EAFSAEGIMD EDLSLQKLD SFFDATNFLK YLPNENEEVI ADIKKDVEF	IGMGSRIVPM YILQKLDMLP KALTEKVGLE FIAFSKIVFL SLELLINISA ISEFLLELLK KTLIEIVSFI LISDCRYEPN YKYIDRPIPT MSPKDKLGN DLTSRDTSKV ERILPLVRAG VLSEGGVGNS EQIINEALVL QTINNKIKET KEKYDVDHII SIFTKEQFIA ELNKIVGNEN EAEYDNYKDD LIIKGWSKRI FMPNFEGSNS KLKEIIISHK PLTKFGGSKF LNNKLAQYRN EKAENEKLYV EEFRNAPDKK LDKESPLYNQ GSQNCYETVE	GAELSKFEQG SQIKLKEDFS DLGKIIYKYN GEPQEEIFKN SKSGDTITIK ENRWAKIRNN FPGEKESQKK EKAIAKSHPV ALKEWIFEEL ETKLQLQKSL LEFINKYGN KYFNNDFSQQ IATILVYDKH IRDIWKNYGI LVKNKKELSL PISRYFDDSF HVNEYFSGVK VKTTTGSITD FDSRKKEYEE DHRHHAIDAL ELLEELSLP PKDKLLLQYN ATEKNIQKIV EKGELKPHTP KTGDNYLFAV TFDKDLLFRQ YWGDLKERGK GRSIKENCFFK	(SEQ ID NO: 42)
CBK78998. 1 Coprococcus catus GD/7	MKQEYFLGLD RMFRTARRRL DAEGNCPPEL YLVLHHMMKH AAIQFVEHVL DIFNNKELDE SVLVEILGNS TEEKLNNYSA ITQEIESEIE ENAEKIQLF VIDVEASAEK GEKISVELKQ FKASLTAYHD PNLTGQLKG NDNLMQLLSR TLKVVKIEIQ ERDWITELNA	MGTGSLGWAV DRRNWRIQVL YALFVDDNYT RGHFLLSGDI KDRNLTRSTK SERPKVSFAD VSISEAKIKV YIGMTKKNGK KENFLPKQVT EFRIPYYVGP FIRMTNKCT RIYEELFCKY FKERLTDVQL ICSLSYQGWG NYGFTNEVEE VMGNAPKRVF QSDQQLRSDK	TDSTYQVMRK QEIFSEEISK DKNYHKDYPT SQIKEFKSTF KSRLIKQLNA SGYDDYIGIV YQKHQADLKT KVDLKSQCT KDNGVIPYQV LNRVDDGKDG YLVGEDVLPK RKVTRKKLER SQRAKEAIVL RLSKTFLEEI FNTLKKETDL VEMAREKQEG LFLYYIQKGR	HGKALWGTRL VDPGFFLRMK IYHLRKLME EQLIQNTQDE KSACEKAILN EAELAEQYYI LKKIVRQYMT QADFYDFLKK HDYELKKILD KFTWSVRKSD DSLVYSKFMV YLVIEGIAKK NVVLFGDDKK TVPAPGTGEV SYKTVDLYV KRSDSRKKQL CMYSGETIQL	FESASTAEER ESKYYPEDKR TTEIPDIRLV ELEWHISLDD LLSGGTVKLS IASAKAVYDW KEDYKRVFVD NVIKVIDHKE NLGTRMPFIK ARIYPWNFTE LNELNNLRLN GVEITGIDGD LLKQRLSKMY WNIMTALWQT SPAVKRQIWQ VELYRACKNE DELWDNTKYD	(SEQ ID NO: 43)

Table 4- Cas9 orthologs						
	IDHIYPQSKT QQQGFTKEK LPDTEIVYVK KFTKNAAWFI MQKNNILVTR NKAAGTYFML GLNSPEILLS AILKGVVKYV RLSAQIKTLT GSAGILVMNN	MDDSLNNRVL YVRLVRSDEL AGNVSNFRQT RNNPGRSYNL KAYEVKGGFL VKS LDKKGKE KIKIDTLFKV NRKNENKDAK EKRAKFIGLS NITACKQISV	VKKYNAIKS SADELAGFIE YELLKVREM KRMFEFDIER DQQIMKKGKG IRTIEFVPLY DGFKMWLSGR LSERDGMTEE NEDQCIVLNE INQSPTGIYE	DTYPLSLDIQ RQIVETRQST DLHHAADAYL SGEIAWKAGN QVPIKGNDR LKNQIEINHE TGNQLIFKGA KLLQLYDTFL ILHMFQCQSG KEIDLILK	KKMMSFWKML KAVATILKEA NIVVGNAYFV KGSIVTVKKV LADIEKYGGY SAIQYLAQER NQLILSHQEA DKLSNTVYSI SANLKLIGGP	
WP_002235 162.1 Neisseria meningitidis Z2491	MAAFKPNPIN VPKTDGSLAM GLIKSLPNT ADKELGALLK HTFSRKDLQA VQKMLGHCTF ERATLMDEPY KAYHAISRAL DRIQPEILEA DHYGKNTTEE IHETAREVVG KDILKLRLYE NNKVLVLGSE RILLQKFDED QITNLLRGFW NAFDGKTIDK DTPEKLRLL KRLDEGVSVL KAFAPFPYKY DVFEKGDYKY LHPNDLVEVI GVKTALSFQK	YILGLDIGIA ARRLARSVR WQLRAAALDR GVADNAHALQ ELILLFEKQK EPAEPKAAKN RKS KLTYAQA EKEGLKDKKS LLKHISFDKF KIYLPPIPAD KSFKDRKEIE QQHGKCLYSG NQNKGNQTPY GFKERNLNDT GLRKVRAEND ETGEVLHQKT AEKLSSRPEA RVPLTQLKLL DKAGNRTQQV LVPIYSWQVA TKKARMFGYF YQIDELGKEI	SVGWAMVEID LTRRRRAHRL KLTPLEWSAV TGDFRTPAEL EFGNPHVSGG TYTAERFIWL RKLLGLEDTA PLNLSPELQD VQISLKALRR EIRNPVVLRA KRQEENRKDR KEINLGRLE EYFNGKDNR RYVNRFLCQF RHHALDAVVV HFPQPWEFFA VHEYVTPLFV DLEKMNVRER KAVRVEQVQK KGILPDRADV ASCHRG TGNI RPCRLKKRPP	EDENPICLID RARRLKREG LLHLIKHRGY ALNKFESG LKEGIETLLM TKLNNLRILE FFKGLRYGKD EIGTAFSLFK IVPLMEQGKR LSQARKVING EKAAAKFREY KGYVEIDHAL EWQEFKARVE VADRMRLTGK ACSTVAMQOK QEV MIRVFGK SRAPNRKMSG EPKLYEALKA TGVWVRNHNG QKGDEEDWQL NIRIHDLDHK VR	LGVRVFERAE VLQAADF DEN LSQRKNEGET HIRNQRGDYS QGRPALS GDA QGSRPLTDT NAEASTLMEM TDEDITGR LK YDEACAEIYG VVRRYGSPAR FPNFVGEPKS PFSRTWDDSF TSRFP RSKKQ GKKRVFASNG ITRFVRYKEM PDGKPEFEEA QGHMETVKS A RLEAHKDDPA IADNATMVRV IDDSFNFKFS IGKNGILEGI	(SEQ ID NO: 44)
WP_012414 420.1 Elusimicrob ium minutum Pei191	MQKNINTKQN DGNVLLPKEI LAEQKLALPV LSLQELGYVL AKKNYRTYGE ILKKQAGKDG PASHKLNEER DYLLTGSELT MPFWARLSEA VLSSSYAPLG LPYYGAVLQE QTLNELRKL SNRNRIYEIY PNDIINNQAD AAKDSKYDYN YISKVAHKYL KESSEFNKDV INYSERNWLS SDNGELVKGT LLKFKGRESE EGKKEKNIND YDTGDSL CVD GDILEVDFDI IIKSTGGQDD	HIYIKQAQKI IMVGSRI FKA PADLDRKENS YHIAGHRGSS YLYKEFFENK FHKELTEEYI RLWETLNNAR PAQTKKLLGL QQDSFLYDWN KSAMLIILEK STQKIIAKGF NEIIDILGKK IRPQQQVIIT IEHLFPPIAES RILSNVKENI ACLFEPKNII NSNKKIRLDN KILLPPNNIV MYKIFYSERG MKNEKIKSAI ASIYQKAISL LYYDNKGKLC HSDKNSFRNN SFTINSMQQY	KEKLGDKPYR SAGAADRKLS SEGETSAKRF AIRT FENDSE EKHKREKISN EKLTKAIGYE YSDPIVDIVT KNTNFEDIIL SCPDEKLLTE IKNDLSYTEA SPQFKDKGYK PCEIGLETAR RRENPRNYIL EDNGRNNLVI PHKAWRFNQG CVKGS LTAQL RHHALDAIVI WENIDADLES YTLTTYKKLS ENNRKLF DVI SGDKYVQLSK GEIIRKIDAQ TGSAPENRVF NPRKLILSSC	IGL DLGVGSI RGQRNNHRHT LGDVLQKDIY EAQKENTENK AANNHKFSPT SEKLIPESGF GEITGYEYEQ QGRDKKAQKI KLSNEYHLTE VEEALKEGKL TPHTNKYELE ELKKS AEDRS KFELLEEQKS SHSACNADKA AFEKFIENKP RMAWGLQGLM AYASRGYGNL FESSVKTALK ALKLTD PQKK QDNLEKAKKL KEPGKFFAIS QKNPLKYKEQ IKVGTFTTEIT GFIKYRSPIL	GFAIVSMEEN RERMRYLWKV ELRVKSLDER KIAGNIKRLM RDLVIKEAEA CPYLKDEKRL FTKEQKQKLF KGYKLIKLES EEIDNAFNEI TKEKQAIKDR YGRIANPVVH KLSREQNDNE QCPCGGQIS KRSPWAAFAS MAARFKTDNS IPFAKQLITE LNKMAGKDYK NAFISVKHDH KTPKDFLETA LEEENEKSKA KPTPTTTGYG GFTLFERIYG NNNIQIWFNG KNKEG	(SEQ ID NO: 45)

Table 4- Cas9 orthologs						
WP_009105 777.1 Treponema sp. JC4	MIMKLEKWRL TKEPLAVARR SLNPYELRIK SPDEIKTQAD YPTRKMYEEE ENDKERTFKA LNSKDKVTFD LWDEIPLEEQ SGTSMCKEY YGKVLPGSTM YGKPSQIAIE FPGKSFYPNR LPFSRTLLDA NQLKNTSKKN NPSDVWTTNG NRFDRHHAI EKVKNIVVSF LDDIVDEKIK IEITSGKISR VVKPAVLENG LDIRPVYAVS VTVSPIGRVF	GLDLGTNSIG TARSQRKLIY ALDEKLEPYE MQTHLEKAIK FNLIRSKQEK MPCSQKLRL QMRKALCLAD DLIIETIITA SEKLVKRLEE EIDLSAPETN LSRDLKNNVE NDRMKYRLWS ESNLTVAHSS KFSPNAMDSF SMTKLLRDKW DAVVI GLTDR KPDHGAEGKL SKVKDYVAKH YLSPEIDYFAA KHPAAKQVC YCADWINSTN RK	WSVFSLDKDN RRKLRRKQVF LGRALFNLA ENGCRITTEF YYPQVDWDDI QDIGNLAYYE SNSFNLEENR DEDDAVYEVI IADLKYHEAV PEKHYGKISN KKAETARKQN ELGLGNKCIY CNAFKAERSP EKDSSFIAEQ EMDSILCRKF SMVQKLATKN SKETLLGKIK KGQKIEAVLS VIWEIPGEKK LLHKDDYLEF ETMLTGYWKP	SVQDLIDMGV KFLQEQGLFP RRGFKSNRND LYKNQGENGG GGSKKRVELN DFLIGNPTAV KKYDLTQEQR ESLGYKFADQ PTVHVALNQT QRAKENIAIN CGKGISGAEL FEAFGTNPSG LSDNQYIACA TEKEVALLGL SHKGNRIEIP LHGKETFVCR DFSKENGICK TFKAQYIRRN SDKGKMYFCR TPTQNWVSVN	RIFSDGRDPK KTKEECMTLK GSREEVSEKK IRFAPGRMTY KPQQRGYCIY DNQDKVLYEL KMRSKNRFGK DFIVKNTILQ TVEKYDLLPY RVVVALIKE DTISALYHTA FTKEIEIEHI YSWQEI IQRA ALRYLKCLVE KPEQIGNYKK EFPILRSDLI ENIVSLSEKN VRCVNRVQTP EVEKESKGLN IAGYATNNK VLFDKQKARL	(SEQ ID NO: 46)
WP_002460 848.1 Staphylococ cus lugdunensis M23590	MNQKFILGLD RGSRLKRRR SKDELVIAL FVCQIQLE KYIELVEMRR YAYSADLFNA IANEINVNPE DQIAEILTIY CIRLVLEEQW VVKRTFGQAI ENTRKRINEI NPNHYEVDHI LSYNQFKQHI TRYATRELTN KHHAEDALII SEMFIIPKQV VQTIKDIYAK NPLAKYHEET KKLVKLSIKP LGKAIDKNAK KEYCELNNIK KRG	IGITSVGYGL IHRLEVRKLL HIAKRRGIHK NEGQVRGEKN EYFEGPGKGS LNDLNNLVQ DIKGYRITKS QDKDSIKSKL YSSRNQMEIF NLINKIIEKY IGKYGNQNAK IPRSVSFDNS LNLSKSQDRI YLKAYFSANN ANADFLFKEN QDIKDFRNFK DNTTLKKQFD GEYLTKYSKK YRFDVYLTDK FIASFYKNDL GEPRIKKTIG	IDYETKNIID LEDYNLLDQS IDVIDSNDDV RFKTDIIEKE PYGWEGDPKA RDGLSKLEYH GKPQFTEFKL TELDILLNEE THLNIKPKKI GVPEDIIIE RLVEKIRLHD YHNKVLVKQS SKKKKEYLLE MNVKVKTING KKLKAVNSVL YSHRVDKKPN KSPEKFLMYQ NNGPIVKSLL GYKFITISYL IKLDGEIYKI KKVNSIEKLT	AGVRLFPEAN QIPQSTNPYA GNESTKEQL IIQLLNVQKN WYETLMGHCT EYHIIENVF YHDLKSVLFD NLTAANKIPK ARENNSKDKQ EQEGKCLYSL ENSKSNLTP ERDINKFEVQ SFTDYLRKVW EKPEIESKQL RQLINDTLYS HDPRTFEKLE YIGNKLGSLL DVLKKDNYYY IGVNSDTRNM TDVLGNVFTN	VENNEGRRSK IRVKG LSEAL NKNSKLLKDK FHQLDENFIN YFPDELRSVK KQKKKPTLKQ QSILENEDVL YTGTHRLSLK AMIDEFILSP KFINEMQKKN ESIPLEDLLN YQYFNSGSKS KEFINRNLVD KFKKERNHGY DIQVDSEDNY TRKKDNSTYI VIMKQYANEK DVTHQFQSST IPEQKYDKLK IELDLPDIRY TQYTKPQLLF	(SEQ ID NO: 47)
WP_011681 470.1 Streptococ cus thermophilu s LMD-9	MTKPYSIGLD LLFDSGITAE LDDSFVLPDD LRLVYLALAH SLENSKQLEE DFRKCFNLDE LLSGFLTVD FKDDTKNGYA KQRTFDNGSI YVGPLARGNS YLPKVKVLPK LYFKDKRKVT	IGTNSVGWAV GRRLKRTARR KRDSKYPIFG MIKYRGHFLI IVKDKISKLE KASLHFSKES NETEAPLSSA GYIDGKTNQE PYQIHLQEMR DFAWSIRKRN HSLLYETFN DKDIIIEYLHA	TTDNYKVPSK RYTRRRNRIL NLVEEKAYHD EGEFNSKNND YDEDLTLLG MIKRYNEHKE DFYVYLKLL AILDQAKFY EKITPWNFED YNELTKVRFI IYGYDGIELK	KMKVLGNTSK YLQEIFSTEM EFPTIYHLRK IQKNFQDFLD YIGDDYSDVF DLALLKEYIR AEFEGADYFL PFLAKNKERI VIDKESSAEA AESMRDYQFL GIEKQFNSSL	KYIKKNLLGV ATLDDAFFQR YLADSTKKAD TYNAIFESDL FLKLIVGNQA LKAKLYDAI NISLKYNEV EKIDREDFLR EKILTFRIPI FINRMTSFDL DSKQKDIVR STYHDLNII	(SEQ ID NO: 48)

Table 4- Cas9 orthologs						
	NDKEFLDDSS	NEAIIIEEIIH	TLTIFEDREM	IKQRLSKFEN	IFDKSVLKKL	
	SRRHYTGWVK	LSAKLINGIR	DEKSGNTILD	YLIDDDGISNR	NFMQLIHDDA	
	LSFKKKIQKA	QIIGDEDEKGN	IKVVKSLPG	SPAIKKGILQ	SIKIVDELVK	
	VMGGRKPESI	VVEMARENQY	TNQGKSNSQQ	RLKRLEKSLK	ELGSKILKEN	
	IPAKLSKIDN	NALQNDRLYL	YYLQNGKDMY	TGDDLDIDRL	SNYDIDHIIP	
	QAFKLDNSID	NKVLVSSASN	RGKSDDVPSL	EVVKRKRFTW	YQLLKSCLIS	
	QRKFDNLTKA	ERGGLSPEDK	AGFIQRQLVE	TRQITKHVAR	LLDEKFNKK	
	DENNRAVRTV	KIITLKSTLV	SQFRKDFELY	KVREINDFHH	AHDAYLNAV	
	ASALLKKYPK	LEPEFVYGDY	PKYNSFRERK	SATEKVYFYS	NIMNIFKFSI	
	SLADGRVIER	PLIEVNEETG	ESVWNKESDL	ATVRRVLSYP	QVNVVKKVEE	
	QNHGLDRGKP	KGLFNANLSS	KPKPNSNENL	VGAKEYLDPK	KYGGYAGISN	
	SFTVLVKGTI	EKGAKKITN	VLEFQGISIL	DRINYRKDKL	NFLLEKGYKD	
	IELIIELPKY	SLFELSDGSR	RMLASILSTN	NKRGEIHKGN	QIFLSQKFVK	
	LLYHAKRISN	TINENHRKYV	ENHKKEFEEL	FYYILEFNEN	YVGAKKNGKL	
	LNSAFQSWQN	HSIDELCSSF	IGPTGSERKG	LFELTSRGS	ADFEFLGVKI	
	PRYRDYTPSS	LLKDATLIHQ	SVTGLYETRI	DLAKLGEG		
WP_009293 010.1 Bacteroides fragilis NCTC 9343 Cas9	MKRILGLDLG	TNSIGWALVN	EAENKDERSS	IVKLGVRVNP	LTVDELTNFE	(SEQ ID NO: 49)
	KGKSITTNAD	RTLKRGMRN	LQRYKLRRRT	LTEVLKEHKL	ITEDTILSEN	
	GNRTTFETYR	LRAKAVTEEI	SLEEFARVLL	MINKKRGYKS	SRKAKGVEEG	
	TLIDGMDIAR	ELYNNNLTPG	ELCLQLLDAG	KKFLPDFYRS	DLQNELDRIW	
	EKQKEYYPEI	LTDVLKEELR	GKKRDAVWAI	CAKYFVWKEN	YTEWNKEK GK	
	TEQQEREHKL	EGIYSKRKRD	EAKRENQWR	VNGLKEKLSL	EQLVIVFQEM	
	NTQINNSSGY	LGAISDRSKE	LYFNKQTVGQ	YQMEMLDKNP	NASLRNMV FY	
	RQDYLDDEFNM	LWEKQAVYHK	ELTEELKKEI	RDIIIFYQRR	LKSQKGLIGF	
	CEFESRQIEV	DIDGKKKIKT	VGNRVISRSS	PLFQEFKIWQ	ILNNEIVTVV	
	GKKRRRRKLK	ENYSALFEEL	NDAEQLELNG	SRRLCQEEKE	LLAQELFIRD	
	KMTKSEVLKL	LFDNPQELDL	NFKTIDGNKT	GYALFQAYSK	MIEMSGHEPV	
	DFKKPVEKV	EYIKAVFDLL	NWNTDILGFN	SNEELDNQPY	YKLWHLLYSF	
	EGDNTPTGNG	RLIQKMTELY	GFEKEYATIL	ANVSFQDDYG	SLSAKAIHKI	
	LPHLKEGNRY	DVACVYAGYR	HSESSLTREE	IANKVLKDRL	MLLPKNSLHN	
	PVVEKILNQM	VNVINVIIDI	YGKPDIEIRVE	LARELKKNNAK	EREELTKSIA	
	QTTKAHEEYK	TLLQTEFGLT	NVSRTDILRY	KLYKELESCG	YKTLYSNTYI	
	SREKLSFKEF	DIEHIIPQAR	LFDDSFNSKT	LEARSVNIEK	GNKTAYDFVK	
	EKFGESGADN	SLEHYLNNIE	DLFKSGKISK	TKYNKLKMAE	QDIPDGFIER	
	DLRNTQYIAK	KALSMLENEIS	HRVVATSGSV	TDKLREDWQL	IDVMKELNWE	
	KYKALGLVEY	FEDRDGRQIG	RIKDWTKRND	HRHHAMDALT	VAFTKDVFIQ	
	YFNNKNASLD	PNANEHAIKN	KYFQNGRAIA	PMPLREFRAE	AKKHLENTLI	
	SIKAKNKVIT	GNINKTRKKG	GVNKNMQQTP	RGQLHLETIY	GSQKQYLTKE	
	EKVNASFDMR	KIGTVSKSAY	RDALLKRLYE	NDNDPKKAFA	GKNSLDKQPI	
	WLDKEQMRKV	PEKVKIVTLE	AIYTIRKEIS	PDLKVDKVID	VGVRKILIDR	
	LNEYGNDAKK	AFSNLDKNPI	WLNKEKGISI	KRVTISGISN	AQSLHVKKDK	
	DGKPILDENG	RNIPVDFVNT	GNNHHVAVYY	RPVIDKRGQL	VVDEAGNPKY	
	ELEEVVVSFF	EAVTRANLGL	PIIDKDYKTT	EGWQFLFSMK	QNEYFVFPNE	
	KTGFNPKEID	LLDVENYGLI	SPNLFRVQKF	SLKNYVFRHH	LETTIKDTSS	
	ILRGITWIDF	RSSKGLDTIV	KVRVNHIGQI	VSVGEY		
AOL40912. 1 Veillonella atypica ACS-134- V-Col7a	METQTSNQLI	TSHLKDYPKQ	DYFVGLDIGT	NSVGWAVTNT	SYELLKFHSH	(SEQ ID NO: 50)
	KMWGSRLEFE	GESAVTRRGF	RSMRRRLERR	KLRLKLLEEL	FADAMAQVDS	
	TFFIRLHESK	YHYEDKTTGH	SSKHILFIDE	DYTDQDYFTE	YPTIYHLRKD	
	LMENGTDDIR	KLFLAVHHIL	KYRGNFLEYG	ATFNSNAFTF	EDVLKQALVN	
	ITFNCFDTNS	AISSISNILM	ESGKTKSDKA	KAIERLVDTY	TVFDEVNTPD	
	KPQKEQVKED	KKTLKAFANL	VLGLSANLID	LFGSVEDIDD	DLKKLQIVGD	
	TYDEKRDELA	KVWGDEIHII	DDCKSVYDAI	ILMSIKEPGL	TISQSKVKAF	
	DKHKEDLVIL	KSLLKLDNRV	YNEMFKSDKK	GLHNYVHYIK	QGRTEETSCS	
	REDFYKYTKK	IVEGLADSKD	KEYILNEIEL	QTLPLQRIK	DNGVIPYQLH	
	LEELKVILDK	CGPKFPFLHT	VSDGFSVTEK	LKMLEFRIP	YYVGPLNTHH	
	NIDNGGFSWA	VRKQAGRVTP	WNFEEKIDRE	KSAAAFIKNL	TNKCTYLFGE	

Table 4- Cas9 orthologs						
	DVLPKSSLLY KNRIELFLKD MAEDIITDIT SKKLLKGIDG NAKLAQQQVV FVEVARTNKS LANYDDAALR LTKDDSFNLL RKYERLTRIA VKAENVSDFR FFKKNGANRT RVTRRLLEQS MTKIKNAYSI PKAKDISIKY ATYIKLLDKY YFMSKLRTPL NSKTTFDVKP	SEFMLLNELN NNYITKKHKP IFGESKKMLR CDKAGNGAPK NPHDIIDELA EKKKKDSRQK SKKLYLYYTQ VLCERTANAK PLTADDLSGF HNNNFIVKRS YNLAKMFNYD GALADATIYK IVQYTGKKGE RKLCYNQLVK DLLRKENKTL YMKMKGNKVD LGITGSRSTI	NVRIDGKALA EITGLDGEIK QTLRNKFGSQ TIIELMRNDS LSPAVKRAVW RLSDLYSAIK MGRCAYTGNI KSDIYPIDNR IARQLVETNQ LNHHHHAKDA VICTNAQDGK ASVAAKAKDG EIKEIVPLPI VNGFYIYLG KASSITTSIY ELSSSTGRSKF GVKIHNLDEF	QGVKQHLIDS NDLTSYRDMV LNDETIKKLS YNLMEILGDK QALRIVDEVA KDDVLQSGLQ IDLNQLNTDN IQTKQKPFWA SVKATTTLLR YLNIVVGNVY AWDVKTSMNT AYIGMKTYS YLINRNATDI KTNDKIYIDN NINTSTVVS IKMTLEEQSI KIINESITGL	IFKQDHHKMT RILGNNFQVS KLRYSRDWGR FLSFMECIEEE HFKKALPSRI DKEFGALKSG YDIDHIYPRS FLKHQGLISE RLYPDIDVVF HEKFTNRNRL VKKMMASNDV VFADVTKYGG ELIDYVKSIV AIELVVPDHI SNKVGIDVFD YLLEVLNLLT YSNEVTIV	
WP_013389 026.1 Ilyobacter polytropus DSM 2926	MKYSIGLDIG REKRSRRRN RAKSLEAKIS KKIMEEKGYL AKQKELGNTH GNCTFESSEK EKEVYLKKEV IGNYTIKKLN LSREDIEILL NNGIKFKRGE VEVARDLAKS KYRLYQEQEG VLVLSENGNRK KFTREDNKDW LTAQLRARWR NCKDVIPSIE REEFYNNLLSD SSKKISNQIA PEKAFAEPFY DVFRKNGKFY YPNDLIEIEF IDTQDGKIQM	IASVGWSVIN RRKKHRLDRI NKEIAQILLH TIGDMVAKDP FTDDVLEKYI RTAKNTWSGE RYEKLRLKILY LSEKLGSEIE SEEFSGTLNL LLPVVDKDLI YDDRQTIIEK RCAYSRKELS KSNLLPKEYF KSRALNDTRY LNKVRENGDL YHADEETGEV KRYLGWFNYE VSKKPLNSIK KPTNSGKRGF TVPIYVHQT GKSKSIKNDN RIGIKTVKNI	KDKERIEDMG KNILCESGLV IAKRRGFKSF KFNTHVRNKA EVFNSQRNFD RFVFLQKLNN LKEEERFGDL EDKSKLDKII SLKAIKKILP ANPVVLR AIS NKKRELENEK LSEVILDESM DRQGRDWDTF ISRFVANYLE HHALDAAVVA YFEEVKDTRF EGFIEKLRPV LKDLEKMQGR LVRGIKVEEK LKELPNRAIN KLTKEIPEV KKYQVDVLGN	VRIFQKAENP KKNEIEKIYK RKTDRNADDT GSYLFSFSRK EGPSKPSPIY FRIVGLSGKR NYSKDEKQDK EILTFNKS YLEKGLSYNE QTRKVVNAII TKKFISEEFG TDIDHIIPYS VLNVKAMKIH NALEYRDDS VTDQKAINNI PMPWSGFDLE FVSRMPNRRV DTDRKLYEAL QNVGVYVNGG GKPYKDWDLI NLSEVLGYR VYKVKREKRQ	KDGSSSLASSR NAYLKSFPWEL GKLLSGIQEN LLEDEVKRIQ SEIGQIAKMI PLTEEERDIV KTEKTKFISL TIESNLKKLE ACEKADYDYK RKYGTPHTIH IKNVKGKLLL RSMDDSYSNK PRKKSNNLKE KKRVFMIPGQ SNISRYKELK LQKRLESEN KGQAHQETIR KNRLEEYDDK QASNGSMVRI DGSFEFLYSF GMDTSTGAAT TF	(SEQ ID NO: 51)
WP_005864 263.1 Parabacteroi des sp. 20_3	MKKIVGLDLG KGNSISQTAD RYGKFLNDIE KVPYDWTIYY KLVEYYSLKV KEFIVTTDL MTVGAIYDIT RDKQLYIACL IDNCPYEENR KETDVDVTQE YRWNYVEDKP DKQEIEKALH MRMGKYWSIE ALPLWLACYL ITETLRTVRD LRIKALLTEF IILGKFNQTD	TNSIGWALIN RTSYRGIRRL CKLPWVKDET LRKKALTQKI EKVEDSGERK ADGSPKKDKE LLQKPDQKIR NELYPNNEPR YIDKESGEIK LLPTEADYVT YPCNETHAQI SFANKNNLSE NIDNGTRIRI VYDRHSEVKD IWQQVGHIDE LNPEFGIENV TLKRPTRSEI	AYINKEHLYG RERHLLRRER GSYKFIFQES SKEELAWILL GKDTWYNVHL GNIKRSFRAP GKLVRTIERK RNSISTRDFC HASIKCIAKS LFEWLNEKKE IARLGKAHIP EFIEQFKNFP NKIIDGEYDE IVKWKTPKDI IHIELGREMK RPYSPSQDL LRYKLWLEQK	IEACGSRIIP LHRILDLLGF FKEMLANFTE NFNQKRGYYQ ENGMIYRRTS KDDDWTLIK YYKNELYQIL HLFIEDIIFY HPLYQEFLRW IDQKAFFKYP KAFLSKEKEE PFKKEYGSYS NIRERVRQKA DLYLKSFKQH NPADKRARM LRIYEEGVLN YRSPYTGEMI	MDAAILGNFD LPKHYSDSLN HHPILIANNK LRGEEETPN NIPLDWEGKT KTEADIDKIK KTQSEFHEEL QRPLKSKKSL QFIVNLRIYR PFGFKKTTSN TLWHILYSIE AKAIKKLLPL INLTDITHFR SLRNPIVEQV QQMIKNENTN SILELPEDIG PLSKLFTPAY	(SEQ ID NO: 52)

Table 4- Cas9 orthologs						
	EIEHIIPQSR FDKPEVLSV ISKVVKSLLS VLPRFIRLNE DAIIIACANR VIDKPWETFT QSKGDSWAIR ILAMLELGYD KPIDTSFDDK NILILNKGKK IYETEEIDKD ILSPNDLVYV LIFALPKATA VDRLGNIQV	YFDDSLSNKV EAYEKLHVES NIVRENEQE LTESTRFTSI NIVNYLNNVS QDTLTALQKI KSMHKETVHG TKRIKNYFEE KIKESITDTG HQPIYKVRVY TKKVIRKRSY PTQEEINKGE EIYCNGENCI GSCILT	ICESEINKLK YSHNRSKMKK AISKNVIPCT NTNNTMIPSM ASKNTKITRR TVSFKQNLRV EVNLRMIKT NKDTWQDINP IQQIMLRHLE EKAKEFTVGQ STIPLNVVIE VVMPIRDRI QNEYGIGSPQ	DRSLGYEFIK LLMEDIPDQF GGITDRLKKD PLELQKGFNK DLQTLCHKD INKTTNHYQH SFNEALKKPQ SKIKVYYFTK TKDNDPTLAF KGNKRTKFVE RQKQGLSSAP YKMDSSGIT SKNQKAITGE	NHHGEKVELA IERQLNDSRY WGINDVWNKI KRIDHRHAM KTDNNGNYKW YENGKKIVSN AIVEMDLKKK ETKDRYFAVR SPDGIDEMNR AAKGTNLFFA EDENGNLPKY ANFIPASTAN MVKEICFPK	
GAP01010.1 Fructobacillus fructosus KCTC 3544	MVYDVGLDIG FRTRRRRLSR AHYYGGYLF HLVKYRGHFL TEIISGEAGY AKLFGKDISD AYDGFVLIGL EIYEQLALAT KQRTKANSVI LRFRIPIYVG NRDKAARDFI KFDSKTADL SSLSYNDLK PLDDQVKKL FMSIINNDKY VLNDIKKAMK KVLKKTLDGN IPQAFTKDDS FINRKKYERL KAVAIRSEIT RDSINVYNEF SEENTSYLK RSDTALYGGY IVNNPAIKKF DADKIASISE LDSDKRIAVI NSEMIYQSAT	TGSVGWVALD RKWRLRLDE TEEETKKFHR SSQEKITIGS GLNKSMADE KDEAKLWKLK LNGADSVSAA DEATIKNGIS PHQLHLAELQ PLVTCKDVEH ERLTGNDTYL INDLFKARKT KTFDAEYLEN SQTHYTGWR GVQAWITEQN EEPRRVYLEF KNKMSDDRYF LDNRVLVSRP TKAGKYLDGQ ADMRLLVGIK DRYTDYLNK VMMFKNILTT SNVYSAYMTL EKILISKPL NSSDEELLEA QTILRGLQID GIFETRVKIS	ENGKLARAKG LFSAEINEID SYPTIYHLRQ TYNPEDLANA ALKLFEFDNN LDDEALEEKS MVQLYDQHRE TARELVEESN KILQNGQYY AGGDADNHVW IGEKTLPQNS VLSALKDYL EDNQETLEKI LSEKLLDSKI TGSSKLTDFE AREDQTSVRS LYFQQQKDM ENARKSDSFA KTGFARQLV KHREINSFHH LRQLSSRDEV KKTEKDRGPL VRANGKNLLI GQLVNEDGNL YDILTSENVK AAYQAPVKII DLL	KNLVGVRLFD SSFFQRLKYS ELMAQPNKRF IEVYADEKGL QDKVAIKTLL QTILSQLTDE DRKLLKSLAQ LSKEVKEDTL PFLDTEFEKE ERNEGFEKSR LRYQLFTVLN KAQKGKDVTI IEIQTVEFDS IDERGQKVI KVNELTTSPA VPRYNQLKEK YTGRPINFER ETRQIIKNVA AFDALLITAA RRLKSFGFVV NKETIFSPKS KIPISIANQI IYLASNEYRH NRFPPFKKDI SKKVSDWHKL	TAQTAADRRG YVHPKDEENK DIREIYLAIH SWELNNPEQL AGLTGNQIDF EIELHAVVQ KAGLKHKRFS RRLDENEFLLP DGQDNKIEEL VTWNFDKVF ELNNVRVNGK TGLADESKFN KIASRELSKL LDKLKSTSQN NKRGIKQSFA YQSKSLSEEA LSQDYDIDHI SLWTSLLKSG SLIEGEYENS GQYMQNRYPD GTMRKGNEDW EVGNLKNINDY NAKQLWLSTT DKLSQVRDEF QQSGGIKLS	(SEQ ID NO: 53)
Bacillus smithii WP_003354.196.1	MNYKMGLDIG RIARSARRRL RVDALERKLN QSILAQYRSV QREFNNPVCT KATYTFQSF DIRKLLNLSD GKDGIRMFNE KVDKSLIDE FTGPKKKEKA ELARDLSHSF FKLWSEQQGR VLTKENREKG YEETEEKEFK TAHLRSRWDF	IASVGWAVIN RRRKHRLERI NDELARVLLH GEMIVKDSKF ERLEEKYLN VWEHINKLRL DIHFKGLLYD TDIDTFGYAL LLNLSFSKFA LLLPIPNIA DERKKIQKDQ CMYSLKPIEL NHTPVYELGL ERNLNDTRYI NKNREESDLH	LDLKRIEDLG RRLVSENVL LAKRRGFKSN AYHKRNKLD WSSQRPFASK VSPDETRALT PKSSLKQIEN TIFKDDIEDIV HLSMKAIRNI NPVVMRALTQ TENRKKNETA ERLLEPGYVE GSRWKKFEK SKFFANFIKE HAVDAVIVAC	VRIFDKAEHP TKEEMNLLFK RKSERNSKES YSNMIARDDL EDIEKKVGFC EIERNLLYKQ IRFLELDSYH AYLQNEYITK LPYMEQGEIY SRKVVNAIK IKQLIEYELT VDHILPYSRS FVLANKQFSK HLKFADGDGG ATQGMICKIT	QNGESLALPR QKKQIDVWQL SEFLKNIEEN EREIKLIFEK TFEPKEKRAP AFSKNKMTYY KIRKCIENVY NGKRVSNLAN SKACELAGYN KYGSPVSIHI KNPTGLDIVK LDDS YANKVL KKKQNLRLR QKVYTINGKI EFYKAREQNK	(SEQ ID NO: 54)

Table 4- Cas9 orthologs						
	ESAKKKEPIF PVFVSRMPKR GHFPMYQKES RTVKIIDTKN KGTLPNKAIE TNEEIIKDI VDVLGNIHKV	PQPWPHFADE SVTGAHQET DPRTYEAIRQ KVVHLDGSKT ANKPYSEWKE FAYYKTIDSA KGEKRVGLAA	LKARLSKFPQ LRRCVGIDEQ RLLEHNNDPK VAYNSNIVRT MTEEYTFQFS TGGLELISHD PTNQKKGKTV	ESIEAFALGN SGKIQTAVKT KAFQEPLYKP DVFEKDGKYY LFPNDLVRIV RNFSLRGVGS DSLQSVSD	YDRKKLESIR KLSDIKLDKD KKNGEPPGVI CVPVYTM DIM LPREKTIKTS KTLKRFEKYQ	
Mycoplasma canis PG 14 EIE39736.1 WP_004794 730.1	MEKKRKVTLG RGTRRLRRR ELKTKALSQE YKLNIFYKKY FFNVQKNILS QGGRYEHIWD KKNIQPLWKL IMISVEDIDM VLINLLDNVG FIAKNESLNI LDNNEGWNLE NVKVTQQAI KAQSKKSLV QISVNEIVNN IKQGHSGWNW KLGFLARNLN IRKNMSYDNK ISKRSEGYWV DDEVFFSRKT DIYLRLLRER TKNKINYNLY KIKFLRLVND ANNFKTSLIN SFIFPGTILL KTPKRLMFGI	FDLGIASVGW KYRNQKFYNL VCPDEIAWIL GYFKGALSQP ETFIEEFKKI KNIGKCSIFT SSVDKLNILL IKDEWAGKEP IKFEFKDRND ENLKLKLEF AIKNYDEEIK LIFNKIIQKF EERLEANNID KVEIDHIIPY DEFTKYVKRV DTRYATILFR LRLKDRSDFS IEDRYTGEIK KRKTNRQLYN EKFVINQSNP LKQYMRSLTK VKINDIRKNQ TQIAIFGDKN DKQNKEFYII KSIMNNYEQV	AIVDSETNQV VKRTEVFGLS HDYLNKRGYF TESEMKNKD FSFTRDISKG NEQRAPKYL NLFNLPISEK NVYGVGLSGL IKNLELLDN LLGAGNEFEN SQIEDNSSLM SKDFEIDKV KSVFNDKYEK SLCFDDSSAN FVNNVDSILS DQLNNYAEHH HHAYDAAIIA ELKKEDWTSI ETIYGIATKT EVIDQIIIEI SLDQFSEEFI VINKFNGKNN WDIEDFKTYN SSIQTVRDII DISPFGINKK	YKLGSRLEFDA SREAIENRFR YDEKETKEDF LKEAFFDFDS PGSDNMPSPY SALIFNFLNE KKKLTSTNIN NIEESAKENK LYLFLIYQKE HNSKTHSLSK AKQDKKYLND IELAREMTQD LIYKIFLWIS KVLVHKQSNQ KKERLKKSEN LIDNKKMFKV LFSNKTKTLY KNNVQARKIA DEDGITNYYK ESYKKNENIP NQMIANKTFV EPKAFYENIN MEKIEKYKEI EIKFLNKIEF IFE	PDTNLERRTQ ELSIKYPNII DQQTVESMPS NKEWLKEINY GIFGEFGDNG LANIRLYSTD DIVKKESIKS FKFQDLKILN SNNKDSIDL KAIDEILPKL NFLKDAILPP QENDALKGIA QDFKDPYTG EKSNSLPY LLTASYDGYD IAMNGAVTSF NLIDPSLINGI KEIEEYLIDL KEKFSILDDK SRDEANIKY LYNPTKNTTR SLGAIVFKNS YGIDKTYNFH KDEKNQDTS	(SEQ ID NO: 55)
Odoribacter laneus YIT EHP49880. 1	METTLGIDL SRNATRAKR KQKSTVRQF RRGFLSSRKG EKYRFRTERV EGSATNVRNS EEEQLKFKN TPAPLSHPEF FEKIPKHLKL FYFYDDNTLL YLKKGAYST EVIRKIKDYL NPIVQQGLNE EKQSRQIREN CPYTGKTLNI GELTPYDFYQ NEFISRQLND QSAPDITFPL LTGEVERKV DGQIVLKGR NNSKLTSSQV TPIKNAPPGV TVESCDPTLI RHHAIDAIVI	TNSIGLALVD QMRRQYFRKK PDTPAFREWL KEEGKIFTGK RARYTLRDMY KLITHLQAKY ESVLFWQRPL EEFRAYQFIN FEKFNFD DTT FEKLQKDIAL AVLLGGIRNS VHNRFGFAKN LRRTVNKLLA EKKNEAAKVK SHTLGSDNSV KDPSPEKWGA TRYISKKAVE PVSATENHRE RCKGMQEFQT EKGVFVCNQL QLFGRVREGI GGGQIILTD PIELSAPKTS ALSSQSFLQR	QEEHQILYSG LRKAKLLELL KQNPYELRKQ DRMVGIDETR IREFEIIWQR GRGHVLIEDT RSQKSLLSKC NIIYGKNEHL KVPACTTISQ QTNDLEKIKK FGKRF EYFKE DRAFQKLYHH TCREKYGPSF LAEYGLKAYR QIEHIIPYSI SSWEEIEDRA YLSAICSDVK YYVITNEQNE DVSDGKYWRR KQKLKTGLPD FRCHNYQCPA VDDKGIFHAD KGENLIEGNI LSTYNARREN	VRIFPEGINK IAYDMCPLPK AVTEDVTRPE KNLQKQTLGA QAGHLGLAHE RITVT FQLPL VFEGRNFYDP TAIQREAVFE LRKLFPHPVW IRLSESYGNV YEPEIEKAVC SQAITTQAQK KFDHIHVEMG DNIQKYLLYK SLDDSLANKT FRLLPYAKAQ AFPGQLTAE VIRLFPKQGE IKLSSSVTWS GSYWISLPVI SGADGNFWCT DDLHYELPAS WVDEHTGEVR KKRGLDSTE	DTIGLGEKEE EDVRRWKNWD LGRILYQMIQ YLYDIAPKNG QATRKKNI FL KEVLGGKIEI VHQBKI IAGP LMCTESKDFN EEKREEIWHC SLKAIRINP RILKEKNAEG ERLPETGNLR RELRSKTER EIEEKG GTVC LCDATFNREK RFIRRK PQES RHLWGLNNIL TPRTEKGELL PLFAPKPISA SQT FKEGESV LDTDTAQPAF LPKGKYYGIF FDPKKNREDQ FSPSPWPGFAQ	(SEQ ID NO: 56)

Table 4- Cas9 orthologs						
	DVRQSVVPLL	VSYKQNPRTL	CKISKTLYKD	GKKIHSCGNA	VRGQLHKETV	
	YGQRTAPGAT	EKSYHIRKDI	RELKTSKHIG	KVVDITIRQM	LLKHLQENYH	
	IDITQEFNIP	SNAFFKEGVY	RIFLPNKHGE	PVPIKKIRMK	EELGNAERLK	
	DNINQYVNP	NNHHVMIYQD	ADGNLKEEIV	SFWSVIERQN	QGQPIYQLPR	
	EGRNIVSILQ	INDTFLIGLK	EEEPEVYRND	LSTLSKHLR	VQKLSGMYT	
	FRHHLASTLN	NEREEFRIQS	LEAWKRANPV	KVQIDEIGRI	TFLNGPLC	
Akkermansi a muciniphila ATCC BAA-835 WP_012421 034.1	MSRSLTFSFD	IGYASIGWAV	IASASHDDAD	PSVCGCGTVL	FPKDDCQAFK	(SEQ ID NO: 57)
	RREYRRLRRN	IRSRRVRIER	IGRLLVQAQI	ITPEMKETSG	HPAPFYLAASE	
	ALKGHRRLAP	IELWHVLRWY	AHNRGYDNNA	SWSNSLSEEDG	GNGEDTERVK	
	HAQDLMCHKG	TATMAETICR	ELKLEEGKAD	APMEVSTPAY	KNLNTAFPR	
	IVEKEVRRIL	ELSAPLIPGL	TAEIIELIAQ	HHPLTTEQRG	VLLQHGKILA	
	RRYRGSLLFG	QLIPRFDNRI	ISRCPVTWAQ	VYEAELKKGN	SEQSARERAE	
	KLSKVPTANC	PEFYEYRMAR	ILCNIRADGE	PLSAEIRREL	MNQARQEGKL	
	TKASLEKAIS	SRLGKETETN	VSNYFTLHPD	SEEALYLNPA	VEVLQRSIG	
	QILSPSVYRI	AANRLRRGKS	VTPNYLLNLL	KSRGESGEAL	EKKIEKESKK	
	KEADYADTPL	KPKYATGRAP	YARTVLKKVV	EEILDGEDPT	RPARGEAHPD	
	GELKAHDGCL	YCLLDTDSSV	NQHOKERRLD	TMTNNHLVRH	RMLILDRLK	
	DLIQDFADGQ	KDRISRVCVE	VGKELTTFSA	MDSKKIQREL	TLRQKSHTDA	
	VNRLKRKLPG	KALSANLIRK	CRIAMDMNWT	CPFTGATYGD	HELENLELEH	
	IVPHSFRQSN	ALSSLVLTWP	GVNRMKGQRT	GYDFVEQEQE	NPVPDKPNLH	
	ICSLNNYREL	VEKLDDKKGH	EDDRRRKKKR	KALLMVRGLS	HKHQSQNHEA	
	MKEIGMTEGM	MTQSSHLMKL	ACKSIKTSLP	DAHIDMIPGA	VTAEVRKAWD	
	VFGVFKEELCP	EAADPDGSKI	LKENLRSLSH	LHHALDACVL	GLIPYIIPAH	
	HNGLLRRVLA	MRRIPKELIP	QVRPVANQRH	YVLNDDGRMM	LRDLSASLKE	
	NIREQLMEQR	VIQHVPAIDM	GALLKETMQR	VLSVDGSGED	AMVSLSKKKD	
	GKKEKNQVKA	SKLVGVFPEG	PSKLKALKAA	IEIDGNYGVA	LDPKPVVIRH	
	IKVFKRIMAL	KEQNGGKPV	ILKKGMLIHL	TSSKDKPHAG	VWRIESIQDS	
	KGGVKLDLQR	AHCAVPKNKT	HECNWREVDL	ISLLKKYQMK	RYPTS YTGTP	
	R					
Dinoroseob acter shibae DFL 12 = DSM 16493 WP_012177 079.1	MRLGLDIGTS	SIGWWLYETD	GAGSDARITG	VVDGGVRIFS	DGRDPKSGAS	(SEQ ID NO: 58)
	LAVDRRAARA	MRRRRDRYL	RRATLMKVLA	ETGLMPADPA	EAKALEALDP	
	FALRAAGLDE	PLPLPHLGRA	LFHLNQRRGF	KSNRKTDRGD	NESGKIKDAT	
	ARLDMEMMAN	GARTYGEFLH	KRRQKATDPR	HVPSVTRLS	IANRGGPDGK	
	EEAGYDFYPD	RRHLEEFHFK	LWAAQGAHHP	ELTETLRDLL	FEKIFFQRPL	
	KEPEVGLCLF	SGHHGVPPKD	PRLPKAHPLT	QRRVLYETVN	QLRVATDGRE	
	ARPLTREERD	QVIHALDNKK	PTKSLSSMVL	KLPALAKVLK	LRDGERFTLE	
	TGVRDAIACD	PLRASPAHPD	RFGPRWSILD	ADAQWEVISR	IRRQVSDAEH	
	AALVDWLTEA	HGLDRAHAEA	TAHAPLPDGY	GRLGLTATTR	ILYQLTADV	
	TYADAVKACG	WHHSDGRTGE	CFDRLPYIGE	VLERHVIPGS	YHPDDDDITR	
	FGRITNPTVH	IGLNQLRRLV	NRIIETHGKP	HQIVVELARD	LKKSEEQKRA	
	DIKIRIDTTE	AAKKRSEKLE	ELEIEDNGRN	RMLLRLWEDL	NPDDAMRRFC	
	PYTGTTRISAA	MIFDGS CDVD	HILPYSRTLD	DSFPNRTLCL	REANRQKRNC	
	TPWQAWGDT	HWHAIANL	NLPENKRWR	APDAMTRFEG	ENGFLDRALK	
	DTQYLARISR	SYLDTLFTKG	GHVWVVPGRF	TEMLRRHWGL	NSLLSDAGRG	
	AVKAKNRTDH	RHHAIDAAVI	AATDPGLLN	ISRAAGQGEA	AGQSAELIAR	
	DTPPPWEGFR	DDLVRVLDRI	IVSHRADHGR	IDHAARKQGR	DSTAGQLHQE	
	TAYSIVDDIH	VASRTDL	KPAQLLDEPG	RSGQVRDPQL	RKALRVATGG	
	KTGKDFENAL	RYFASKPGPY	QAIRRVRIIK	PLQAQARVPV	PAQDPIKAYQ	
	GGSNHLFEIW	RLPDGEIEAQ	VITSFEAHTL	EGEKRPHPA	KRLLRVHKGD	
	MVALERDGR	VVGHVQKMDI	ANGLFIVPHN	EANADTRNND	KSDPFKWIQI	
	GARPAIASGI	RRVSVDEIGR	LRDGGTRPI			
Wolinella succinogene s DSM 1740	MIERILGVDL	GISSLGWAIV	EYDKDDEAAN	RIIDCGVRLF	TAAETPKKKE	(SEQ ID NO: 59)
	SPNKARREAR	GIRRVLNRRR	VRMNMKKLF	LRAGLIQDVD	LDGEGGMFYS	
	KANRADVWEL	RHDGLYRLK	GDELARVLIH	IAKHRGYKFI	GDDEADEESG	
	KVKKAGVVL	QNFEAAGCRT	VGEWLWRERG	ANGKKRNKHG	DYEISIHRLD	
	LVEEVEAIFV	AQQEMRSTIA	TDALKAAYRE	IAFFVRPMQR	IEKMVGHCTY	

Table 4- Cas9 orthologs						
WP_011139 289.1	FPEERRAPKS AVSREKVEFR ELEFDKVEAE YYKDEGQKRR GARYDEAVLM LVRKYGAFDR FQVPLTRKNI RSADDSFANK VRTGKGKIDR NSHTKAPVQV VQKLSEYYRF VKRLISRPP FKADRVASAN SLGTNPEAMD MVLSSINNSP SQEGLRPPRK RSAKKLVKK	APTAЕКFIAI HLRKFLDLS KKAWISLRGA ELTKLPLEAE LGVPHEKESA VHFELAREIN LKKRLYIQD VLCLARANQQ LLKKNFEDENS RSGKLTSVLR KETHREKERP RARVTGQAHE QKNFYETSTI ENFFKFSIFK CEGFTCTPVS EFECDAQVKF	SKFFSTVIID NEIFKGLHYK AKLREALGNE MVERLVKIGF ILPPLNKTDI TKGEIEDIKE GRCAYTGDVI KTDRTPYEWF EMAFKDRNLN YQWGLESKDR KLAVPLANFR QTAKPYPRIK PRVDVYHKKG DDLISIQTQG MDKKHKDKCK ALDVKKYQID	NEGWEQKIIE GKPKTAKKRE FYGRFVALGK SDFLKLSLKA DILNPTVIRA SQRKNEKERK ELERLFDEGY GHDAARWNAF DTRYMARAIK ESHTHHAUDA DAVEEATRIE QVKNKKKWRL KFHLVPIYLH TPKKPAKIIM LCPEENRIAG PLGYYYEVKQ	RKTLEELLDF ATLFDPNEPT HADEATKILT IRDILPAMES FAQFRKVANA EAADWIAETS CEIDHILPRS ETRTSAPSNR TYCEQYVWFK IIIAFSTQGM NTETVKEGVE APIDEEKFES EMVLNELPNL GYFKNMHGAN RCLQGFLDYW EKRLGTIPQM	
Parasutterell a excrementih ominis YIT 11859 WP_008864 843.1	MGKTHIIGVG SKSRTAVRHR LSGYLKRRGY EAIANSPEET AILDASKYIA WRIIGNISNL DKEWSASQKQ LYLNPALSS SDVLPDSYR VLSGHQLEEF NVIVGQALGV FSEDYKFVKT EAQSKFDNLY CRLPADCVRP IEANSFNFTA RADSHGICAY GNQEKKNYI YFDLLSEKER SIFSKVRQAL KVQPVASHSI VIQNTPRNFS EEKPNRISIG SKVAGSKFTA ENLAKQKGCL KDENALRKAL FGDFVYQSOD HDRSISMSEF SISLGSSFKF PRVIFNYIVG	LDLGGTYTGT VRSYKGFDLR ARTEAETDTS KALNKELSGQ NLQSLGHKHR QERAVRWYFN IIQSLEQSGD EYGEKWKSWA LAYILQRAFD LEFANRYQYE SPAEGTDFIE ALKEGKTEKE SLAQLYNLIE FDGFIRKAID SLTDLKYIQL TGRPLDDVGE LLSNLAKNYL ACARHALFLN AAWTQETGNE DAMCIYLAAC DKTNIANSPI GKDPFSILSV EEDKAAKILE KKVEYSSKEF EAAWPSSFGT TNNLYSSFPV REVYNKDDLM SVSEVFDNKI GAASSLKEIF	FITSHPSDEA RRLLLVAEY VLESLDPSVF KEADFKKYIK SKYLSDILQD DAKFEQGGQEQ VLDVLAGLDP NKFAGAYPLL RSIALDECSI TAKAKNGLWF EIWNSKVKGR LSKKFAAVIK TERNGFSSKVS RNSWEVAKRI KEQKLKKKLE IDHII PRSLT AAVFGTSDLS SDSEARRAVI LIFDAISVPA SDPFKTKRMG FKETIYAERF LGAYLDKAPS ALHFVTVKQD KFKGSLIIPA RNLHSAKRV KNGKLDWSSP RIELAQGTSS YTENAEFTKF SEAGKERS	EHRDHSSAFT QLLQKKQTLA SSAPSFTNFF TSFPEYSAKE MKRDSRITRL LDAVKLKNVL DRTIPPYEDQ TEDLTEILKN RRTAEDFENG PENALLERAD STVRSICNAI VLKMOVSEVVP LAAHLENAWR AEEVKKSVDF DIQRNEENQE LKKSESIYNS QITNEIESTV DVLGSRRKAS ADSSEMRKRF SQLAIYEPIN LDIIVSRGEI SEKEKLTIIYR VAATVSDLIK AVEWGKVLWN FSLPVVATQS IIHPALQNRN RRYLRVEMPG LPKPRDNKH	VVNSEKLSFS PEERENLRIA NDSEPLNIQW ILANYVEGRR SEAFGSTDNL VRALKYLRSD NNRRPPEDQT TDRKSRIKIR VVIKNEKLED LHPPMKNKIL ENERKTYGPPY FIGKELRLSD MTMTDGSQC TNGTVKIPVA KRWLSKEERI EVNLI FVSAQ LQLKAAGRLG VNGTQAWFVR AEYRPEFRKP FDNLFTGSCQ FIGYPSNMFF VVKNAFELF SKKELSKDSI VFKENTAEEL GAVRIRKTA LTAYGYRFVD EKFLAWFGEN NGTIFFELVG	(SEQ ID NO: 60)
Streptococc us sanguinis SK49 WP_002933 589.1	MTKFNKNYSI IGSTTFVSAQ YRRLDESFRV ADKNSPVADI CQEVFDLEIS QLLQLLFGLK FAKLKVL RDT QNLSEQDYLD NFYDYISGKI KIIRQKGKDY	GLDIGVSSVG PAKGRVFRV LGDKSEDLQI REVYMAISHI DESKNIADIF TKFKDCFELE ILLSGMLTYT IFGRKTQNGF TGIEGAEYFL PFLLENKDKL	YAVVTEDYRV NRRRIDRRNH KQFFFFGDKEL LKYRGHFLTL KSENRRQEKV EEPDLNFSKE GATHARFSAT DVDKETKGYV NKISDGTFLR LSILTFKIPY	PAFKFKVLGN RITYLRDIFQ ETAYHKKYPT DKINPNNINM KKILPYFQQE NYDENLENFL MVERYEEHRK GYITNMVLT KLRTSDNGAI YVGPLAKGSN	TEKEKIKKNL KEIEKVDKNF IYHLRKHLAD QNSWIDFIES LLKKDKSIFK GSLEEDFSDV DLQRKFFFIK NPQKQKTIQQ PNQIHAYELE SRFAWIKRAT	(SEQ ID NO: 61)

Table 4- Cas9 orthologs						
	SSDILDDNDE VLPKRSLIYE KRVNAKSLIK ELLDSEIYQK SKLRYSGWGR SFEPKIKDIQ GYPPHNIVIE HSDEQLQSEK SFLPIDSIDN TKYQRLTTSE TKIITLKSQ KLAPELIYGH NRDLPIIKDV KTKALCLDTS DYEKFNPNPF LFESKSNLHK FILKHKEEFD TIKYFYDNFI IHQSITGLYE	DTRNGKIRPW EVMLQNELTR YLSDNHKDLN ELEEEIIVIT YSAKLLLDIR SKSTIEDDIF MARENMTTEE LYLYYLQNGK KVLTHRENNQ RTPDGVLTES ITNFRNTFHI HAHFNRRHEEN IYNSQINQVK IYGGYGPMNS QFLNDTSENG ATQFKLTKTQ NISNQLSAFS KMFSFVKSGA TRIDLSKLGE	NYQKLINMDE VKYKDKYGKA AIEIVSGVEK VFDDKKSINK DEDTGFNLLQ DEIKKLAGSP GQKKAKTRKT DMYTLDKTGS QKLNNIPDKE MKAGFIERQL AKIRELNDYH KATLRKHLYS RTMIKKGAFY ALSIIIIAER FLKKNNINRV NELFFHMKRL QKMLGNTTSL PKDINDFFDN D	TRDAFITNLI HFFDSELQRN GKSFNSTLKT YLTKEFGHLE FLRNDENRN AIKRGILNSI KLESALKNIE PAPLYLDQLD TVANMKPFWE VETRQIIKHV HAHDAYLAVV NIMRFFNNPD NQNPVGKFNK FNEKKGKIET LGFYRIPKYS LTKSNLMDLK KNLIKGYNER KCTVARMRPK	GNDIILLNEK IINGLFKNNS YNDLKTIFSE ILDEEIKINQL LTKLISDNTL KIVDELVQII NSLLENGKVP QYEVDHIIIPY KLYNAKLISQ ARILDNRFS VGQTLKVPY SKVSKDIWDC QLAANNRYPL VKEFHDIFII LMQKIDGTRM SKSAIKESQN KIKEIDIRDE PDKKLLNATL	
Actinomyces sp. oral taxon 180 str. F0310 AOL41039. 1	MLHCIAVIRV VGLRSIGFCA ARARRRGRAE IEDETEREQC VEDRTGLQFS KYMQSDLVAE LDELQALDLP EELNRVARYL YPPVDDTTVR DEEVNELISS DLSQAITRLY GVPQTVNIEH PFRRSDQVRY NLAAVCHSCN TEKAMGKRKS DRPEGCAAVQ RNHAMDALVI QDRWESWIGD KARRGSKRPR ADLNRGLKLR GPKEKVKYAL NGSAKQIGWL KITLKPRLLS QSGLIRVIRR	PPSEEPGFFE VEVDDDEDHI KQRLKKLDVL LSVAMAHIA EEVTQGEI LRQICRTQVR AAKQPRAE LNHTESSEPT VMSAEVDWLA ATAEDMLKLE GVDPGWVPTP TREGKLSASL EILDLDQDCAC SAKGGLAFGQ EVISRLKTEM VNAYSGRRLTA ALMTPAIART VEYACDRLNE PQRYVLGDAL GKRISADFP LRVCAIDLCC VLGDEIQIDP NEPLLKTSRV NALGEVRTSP	THADSCALCH RILNSVVHVH LEELGWGVSS HRGWRNSFSK TILEHDGDVT SETTFEKLVL HPAFQKFKVV WDDVARKLEV DWWDCADES LLAKKLPSGR APIEAPVGNP LEEEERERWER LYCGNEINFQ WVKRGDCPSG PYEEFDGRSM CARRAAHVDK IARVREDRREA LIDADKIPVT PADVINRVT DYFPTDSPAL IDCDDLFEVE TKFPGQSIGK GGHESDLVVA KSGLPISLNL	HGCMTYAAND DAGTGGPGET NELLD SHAPW VDTLLLEQAP IRGFVRKGGK SIFHSKEPAP ATLANMRIRE PRHRLRGSSR RGHMIDAISN VAYSCLKTLRE SVDRVLKQVA FEARREIRQK TFEVDHIIIPR VSLENAIKRV ESVAWMAIEL RVRLIRLKG QQLTRAFESW ENLRLRNSGK PGLWTALVRA AVQGGYVGLE LKPSSISMRT FLKECGPVSS ECVEKIMKKT R	KAIRYRVGID ESLRKRSOVA HIRKRLVSEY SDRMQGLKER ATKVHGVLEG SAARQREVRG QSAGERSLTS ASLETGGGLT GCGSEPDVVE VTAAILETGD RWLKFASKRW EMYKRLGISG VDASSDSRRT RSWSKDRGLL KKRIEGYFNS DGHKKNRFD KNFLGSEERM LHADQPESLK PGFDSQLGLP FHHARLYRII ADAKLKEAMG WRVSALDTPS GWVVEINALC	(SEQ ID NO: 62)
Rhodovulum sp. PH10 WP_008386 983.1	MGIRFAFDLG GQSLATMRRV TDPYHLRAKA EGSKRLAETL FYPTREMVRA GCCTFEPSE LASALLAGKS DHYGAAWHGL SIPLADGYGR WHHSDEDERGV HIGLNQLRKV EENERRTAIL AELFTSEVEI	TNSIGWAVWR PRQSRKRRDR LDES LTPHEM AATNCRTLGE EFERLWTAQS RLPRALPSVE LTFKAVRKT SFAEKDTFVG LGRTANTEIL ILDRLPYYGE VNRLIAAHGR AEHGQRDTAE DHILPVS LTL	TGPGVFGEDT FVLRRLDLLA GRVIFHLNQR FLWSRHRTGP RFAPDLLTPE ARGIYERLAH KILPHALVNF KLLDEADEER AALVEETDET ILQRHVVP GS PDQIVVELAR NKIRLRLFEE DDSLANRVLC	AASLDGSGVL ALRKAGLFPV RGFRSNRKAD RTRSPTRI RM RHEEIAGILF LRITTPVSD EEAGEKGLDG LIRRLVTENR GTVVTYAEAV GEPEEKNEAA ELKLNREQKE QARANAGIAL RREANREKRR	IFKDG RNP KD DVEEGRRLAA RQDREK GKIA EGEGAKALYA RQRDLAPPKI RGLTRPERDV ALTAKLLSKP LSEDAARRCA RRAGERTGRN RWGRLANPTV RLDREN RKNR CPYTGRAIGI QTPFQAFGAT	(SEQ ID NO: 63)

Table 4- Cas9 orthologs						
	PAWNDIVARA KIYLGKICDP DAVIGVLT STITVAVKPE EVDRVRDRAL RILKPDASVV EVNRPGRPE ANRVRLSPHN VVTLLRSNV	AKLPPNKRWR DRVYVTPGTL GMIQRIAHDA HGKGGALHED RARLGALAAP TIADRRTGVP WEVKKLGKGL DGGKLQDRHA	FDPAALERFE TGLLRARWGL ARAEDQDLDR TSYGLVPDTP FRDESGRVRD YRAVAPGENH VMRLHKGDMV DADDPFRWDL	REGGFLGRQL NSILSDSNFK VFRDVPVPFE PNAALGNLVV AKGLAQALEA HVDIVQMRDG ELSDKDGQRR ATIPLLKDRG	NETKYLSRLA NRSDBRHHAV DFRDHVRERV RKPIRSLTAG FGAENGIRRV SWRGFAASVF VKVVQQIBIS CVAVRVDPIG	
Bifidobacterium bifidum S17 WP_013362 995.1	MSRKNYVDDY TAESRRMART PDDENNADCW IYLAIHHMVK SDIEQDDENK QRAIVKAI FA QIVDSGLLDD ANWNLIKQL KANVFFKKMI NELKQIKKQ DLQKSDNSEN LGEPTLPKNS LCQRLFMKGQ CRIFGEKYVN VNTHYTGWGR ITDKQLGFSD DISKAVGKRP ADELNACSDK QAVTQNDSD SRVKFERLTR VIGLNAELTK ADGEVSDGAQ EQKRVNPRTG AATDPKGIKG QEYSLGDRP IKSAGELNNA CLYPWHRFTE VGMTGSWRRM NSKYRTNEKR	AISLDIGNAS TRRRYSRRRW YGSVLFDSNE YRGNFLVEGT LVAPANGQLA GLEGNKDLA EEREFLDRMQ RTKENEKDIN ETSDLSETDK GKYYPFLDA HWMVRKKKE LLYQEYEVNL TVTKRTAENL EHRDLMEKIV LSRKLLTTKA WIEQENLGAE SRIFLELADD DLQDDRLFLY NRVLVARAEN QNEFSEREKE EMHRYLGFSH NSYNQYL RDY EVVWSEEDKD IPFDGAKQDT SDDELRLKVL QQLWLPEEY EELADLHVAF NNKSGYTFSD LPTLSGASQP	VGWSAFTPNY RLRLLDALFD QDKRFYEKYP GQYNAIALKQ ENYGILVGWK NRLIHDIEED FEKDGKQINK ITPWNFDEM ELNNVRLSVR LRKEYGRTYE ELQTVFEDKE GECKISDDFA NGSSLMEVVD IQPSGRTISR YTQLGKDMYT ARKTDSFTYM RFVQSRSLVET KNRDINDYHH LRGYREKLSA YLRKVMNYRK SLYGGFSSAK KKKSEYAKAN CYFDDLSQK DKLPEDEKKN EDEFIFQSPS	RLVRAKGHEL QALSEIDPSF TIYHLRKALM LLGRSTSVSD IDSGQRKSVR KLFPIQRDSD IEGLLTFRVP DKDASGRKFI TGNHWNDRKR LSGLSDESKF TLLHQLRQLE PRKHSIIEIM DLRVSPKVKR KSRLQDLYRN GEELDLDRLS PQIADMRNRF AQDALCVGIA EDRKQGRAFG MLVTQKVGDD PAYAVLIESK ILLRHVPKM GSLEKDDLK VITGIVSALH GLFEKRVTVG	IGVRLFDPAD LARRKYSVWH EDDSQHDI RE RYEMSEGEQN LSAVNDLSRE NKSDYEEKCV SKCASYDAHR GESAYENMRK NGVIPHQLHQ YFVGPLVVPE ERLVGTDSYL MRLGREEKTL TSSLSTYGKM GISEADCALL RAEDRNLM EI GIIQSIRLID ANLGKEFKGI SAYDIDHIIP WQILLDNGLI MRQRYGNSAA GQFAANRGFF FIVGSMRSQD FGALYDETRY GKTRLVNVTM LIRYGGGLMV LLDSILGSVQ ADAKTANLSI ELKRKAKKEV	(SEQ ID NO: 64)
Barnesiella intestinihominis YIT 11860 WP_008863 245.1	MKNILGLDLG GVSINSQRTQ LWGLRAKAVT VKTRYDELKE CQRKFYPDIL GKKTEAGPKV FDFLNTHEKL GNYPDKKRL SIDDREQLQS FLQLGMNYAE EKILNQMVNV RENEQIAKRI DVEVEHIIPK DYVERVNTMY EILTSICYNV NHRGSVIRRE PDFNKMSLER	LSSIGWSVIR KRTQRKGYDR QRIELNELGR MRLTIGELFF TDETIRCIRD SPRTSPLFQV KGSDDLKLLG QFNLQEESSS VLRQKFGIDD ACEAAGYNHS VNALMEKYGR VEYGVPTSR SLYFDDSFAN TNNQISKTKW TATSGSVTSF QIKDWSKRFD YIQSQPHFSV	ENSEEQELVA YQLRRTLLRN VLLHLNQKRG RRLTENAFFR EIIYYQRPLK CRLWESINNI LSKTYGYRLG MVNTTETGEI DEVLERLSAI NNYTKAENEA FDEIRVELAR IQKYMWEES KVCSCRSCK QNLLTPVDKI LRHVWGWDTV HRHHAIDALT AQVREAVDRI	MGSRVVSLTA KLDTLGMLPD YKSIKSDFSG CKEQVYPRQA SCKYLVSRCE VVKDRRNEIV EQFKTGIQGN PMISLSFEQE DLVKAGFGNK RALLDRLP AI ELKQSKEERS KHCCIYCGQP EKNRRTAYDY SIDFIDRQLR LHDLNFDRYK TACTKQAYIQ LVSFRAKGRA	AELSSFTQGN DSLSYLPKLQ DKKITDYVKT YVEEFDCIMN FEKRFYLNAA FISAEQRAAL KTRVEIERAL PLYRLWHVLY SSKAIRRIPL KKNELRQPVV NTYKSINKNQ VDVGDFLRGF MKSKEKALS ESQYIARKAK KVGLTEVIEV RLNNLRAEEG VTPGKRYIRK	(SEQ ID NO: 65)

Table 4- Cas9 orthologs						
	NRKRISVQSV SINREMLDKE RTVRCFAKPA TVTFWEAVDR VLSLQQNEMF HTTETSVEDKY EIS	LIPRGALSEE KVVDKRIHRI INTLVPLKKD CRVGIPAIVT ILGMNEEDYR DGKPNLKLSM	SVYGVIVHWE LSGRLAQYND DKGNPVAWVN QPDTIWDNII YAMDQQDYAL QMGKLRVSI	KDEQGHVIQK NPKEAFAKPV PGNNHHVAIY QRNDISENVL LNKYLYRVQK KSLGLNPHK	QRAVMKYPIT YIDKECRIPI RDEDGKYKER ESLPDVKWQF LSKSDYSFRY VHISVLGEIK	
Aminomonas paucivorans DSM 12260 WP_006299 850.1	MIGEHRVGGC ARGLRRRLRR EGLDRPLSFP VGGENERLLR ARRLFQSQRT PDQIRAPRRS WKQEKVTHKT KALDKGPDPA NARALAPLSF HPKLPPLLEA QPAKVRRIE GFPCYCEEYL KGNRTPFEAF KDRNLTDTRF LHKNRKNGDL SLPYPAFRSE SRRLRGKAHM RKLYEALRER VELRDGGEVY VAHKSEEWDD LLDRHRDMP VL	LFDDHWTPNW KTQRLEDLRR EWVRVLYHIT EGGYRTAGEM LGNPHASSNL ASAETFMLLQ LRRHLEIPEE VWDTLRSRRD SGTAHLSLSA DWRNPVVFRA TEQQANEKKK NPTRLAEPGY GGDTARWDRL ITKTAATLLR HHALDAAVLA VLARLCPTRE ETLRSFKWKD LAAFSGNGKK AVADHESMVR LVDGSFDFRF RESDCGTFYV	GAFLRPNTVR LLAKEGVLSL KHRGFQSNRR LARDPKFQDH EEAFLHLVAF KTGNLRLIHR WLFTGLPYHR LLDSIADTTL LGKLLPHLEE LTQTRKVVNA QQAEREFLDI AEMDHILPYS VAWVQASHLS DRLTFHPEAP VASRSFVYRL EILLRLDQGG HPEGPRASR AFVAPFRKPC VDVYAKKNRF SLFPGDLVEI STRKDVLSMS	TFTKAENPKD SDLETLFRET NPVEDGQERS RRNRAGDYSH QNPFFASGEDI RTGEERPLTD SGDKAEKLF FYKNEDEILP GKSYTQARAD LVRRYGPPWC VGTAPGPGDL RSLDNGWHNR APKKRNLRLRE KDPVMTLNGR SSHNAAWGEL VGYDEAFRNG IPLKDLNLEK RSGEGPLVRS YLVPVYVADV EKKDGAYLGY KYQVDPLGEI	GSSLAEPFRQ PAKDPYQLRA RQEEEGKLLS TLRSRLLEE RNKAGHCSLE KEREQIHLLA VHLAIGHEIR RLESIGLSPE AGYAAPPDDR IHLETARELS LKMRLWREQG VLVHGKDNDR DFGEEAEREL LTAFLRKQWG PRGREANGF LRPVFVSRA LERMVGKDRD LRFDSGYS ARGIVKNRAI YKSHCRGDGR RLVGSEKPPF	(SEQ ID NO: 66)
Ralstonia syzygii R24 CCA84553. 1	MAEQHRWGL SLAVERRGPR PYVLQRGLD AIAAFRAGMG IAQEFDALWA NQPRVAAALP ARPKCGFDML TRWQALSLDE EDGHMRLGRS GERDALPYYG NALIQRYGKP DAGVPDNREN HILPFSKTL RAAVLPRNKR CSKDDVYVSP DAVVIGATDR ARCVVSHKPD LSNVRCDA LPIRPRGEDG NQAAYYRRFRN MSENKITLAE IGRVLDPGIK	DIGTNSIGWA QMRRRRDRYL QALTLPFEGFR NARTVGEALA SQQRFHAEVL SAQRFRLMQE RKIVFGANKE QDRLVCLLLD ALLRVLDAL ELLWRYTQDA AQIVVELARN RLRMRLFEE DSFANKVLAQ NRFAETAMQD GRLTAMLR AMLQQVATLA HGPEGGLHND QAELEPIFEQ KTLPDSAPYK DPARFRRYTA HFEGGTLKQR GD	VIALIEGRPA RRRDRFMQAL ALFHLNQRRG RRLEDGRPV ADAARDRLRA LNHLRVMTLA AYRFTIESER GENDAVLADA SGRDEQGLPL PTAKNDAERK LKAGLEEKER GQGNLGT HDANRYKGNR WLHNETDFLA WGLNRVLDGV ARAREQDAER TAYGIVAGPF DDARAREVAL AYKGSNRYC GGRPLLMRLC DADKDDPFKY	GLVATGSRI INVGLMPGDA FQSNRKTDR ARMVGQKDE ILLFQRKLLP DKRERPLSFQ RKELKGC LREHYGLTDA SYDKAVVAAG FGKIANPTVH IKKQQTANLE IYSGRQISLQ GPFEAFGANR RQLTDTAYLS MEEQGRPAVK LIGDMPTPWP EDGRYRVHR TALAERYRQ ELFINERGRW INDYIAVGTA LTKSPGALRD	SDGRNP AARKALVTEN TAKESGKVN HYELIAREW VPVGKCFLEP ERNDLAQLV AKLAKVNALG QIDTLLGLSF YPAHTADLEN IGLNQLRKLV RNERIRQKLQ RLFSNDVQVD DGYAWDDIRA RVARQYLTAI NRDDHRHHAI NFLEDVRAAV VSLFDLKPGD KVWLEELMSV DGELISTFRA AERTIFRVVK LGARRIFVDL	(SEQ ID NO: 67)
Catenibacterium mitsuokai DSM 15897	IVDYCIGLDL ASRSIRRRYN KYLGTDYKDN YLALHHIVKY EKKNLLEILEI	GTGSGWVAV KRRERIRLLR YNLFIDEDFN RGNFLYEGQK LKKPLSKKAK	DMNHRMLMKN AILQDMVLEK DYTYHHKYPT FNMDASNIED VDEVMTLIAP	GKHLWGSRLF DPTFFIRLEH IYHLRKALCE KLSDIFTQFT EKDYKSAFKE	SNAETAANRR TSFLDEEDKA STEKADPRLI SFNNIPYEDD LVTGIAGNKM	(SEQ ID NO: 68)

Table 4- Cas9 orthologs						
WP_006506 696.1	NVTKMILCEP	IKQGDSEIKL	KFSDSNYDDQ	FSEVEKDLGE	YVEFVDALHN	
	VYSWVELQTI	MGATHTDNAS	ISEAMVSRYN	KHHDDLKLLK	DCIKNNVPNK	
	YFDMFRNDSE	KSKGYNYIN	RPSKAPVDEF	YKYVKKCIEK	VDTPPEAKQIL	
	NDIELENFLL	KQNSRTNGSV	PYQMLDEMI	KIIDNQAEYY	PILKEKREQL	
	LSILTFRIPY	YFGPLNETSE	HAWIKRLEGK	ENQRILPWNV	QDIVDVDATA	
	EGFIKMRSY	CTYFPDEEVL	PKNSLIVSKY	EVYNELNKIR	VDDKLEVDV	
	KNDIYNELFM	KNKTVTEKKL	KNWLVNNQCC	SKDAEIKGFQ	KENQFSTSLT	
	PWIDFTNIFG	KIDQSNFDLI	ENIIYDLTVF	EDKKIMKRRR	KKKYALPDDK	
	VKQILKLKYK	DWSRLSKLL	DGIVADNRFG	SSVTVLVDLE	MSRLNLMELI	
	NDKDLGYAQM	IEEATSCPED	GKFTYEEVER	LAGSPALKRG	IWQSLQIVEE	
	ITKVMKCRPK	YIYIEFERSE	EAKERTESKI	KKLENVYKDL	DEQTKKEYKS	
	VLEELKGF'DN	TKKISSDSL	LYFTQLGKCM	YSGKKLDIDS	LDKYQIDHIV	
	PQSLVKDDSF	DNRVLVVPSE	NQRKLDLVLV	PFDIRDKMYR	FWKLLFDHEL	
	ISPKKFYSLI	KTEYTERDEE	RFINRQLVET	RQITKNVTQI	IEDHYSTTKV	
	AAIRANLSHE	FRVKNHIYKN	RDINDYHHAH	DAYIVALIGG	FMRDRYPNMH	
	DSKAVYSEYM	KMFRKNKNDQ	KRWKDGFIN	SMNYPYEVVG	KLIWNPDLIN	
	EIKKCFYKYD	CYCTTKLDQK	SGQLFNLTVL	SNDADHDKGV	TKAVVPVNKN	
	RSDVHKYGGF	SGLQYTIVAI	EGQKKKGKKT	ELVKKISGVP	LHLKAASINE	
	KINYIEEKEG	LSDVRIKDN	IPVNQMIEMD	GGEYLLTSPT	EYVNAARQLVL	
	NEKQCALIAD	IYNAIYKQDY	DNLDDILMIQ	LYIELTNKMK	VLYPAYRGIA	
	EKFESMNENY	VVISKEEKAN	IIKQMLIVMH	RGPQNGNIVY	DDFKISDRIG	
	RLKTKNHNLN	NIVFISQSPT	GIYTKKYKL			
Mycoplasm a synoviae 53 AOL40776. 1	MLRLYCANNL	VLNNVQNLWK	YLLLLIFDKK	IIFLFKIKVI	LIRRYMENNN	(SEQ ID NO: 69)
	KEKIVIGFDL	GVASVGWSIV	NAETKEVIDL	GVRLFSEPEK	ADYRRARRTT	
	RRLRLRRKKFK	REKFHKLILK	NAEIFGLQSR	NEILNVYKDQ	SSKYRNILKL	
	KINALKEEIK	PSELVWILRD	YLQNRGYFYK	NEKLTDEFVS	NSFPSKKLHE	
	HYEKYGFVRG	SVKLDNKLDN	KKDKAKEKDE	EEESDAKKES	EELIFSNKQW	
	INEIVKVFEN	QSYLTESFKE	EYLKLFNYVR	PFNKGPGSKN	SRTAYGVFST	
	DIDPETNKFK	DYSNIWDKTI	GKCSLFEEEE	RAPKNLPSAL	IFNLQNEICT	
	IKNEFTEFKN	WWLNAEQKSE	ILKFVFTLEF	NWKDKKYSK	KFNKNLQDKI	
	KKYLLNFALE	NFNLNEEILK	NRDLENDTVL	GLKGVKYEEK	SNATADAAL	
	FSSLKPLYVF	IKFLKEKKLD	LNLLGLLENT	EILYFLDSIY	LAISYSSDLK	
	ERNEWFKKLL	KELYPKIKNN	NLEIENVED	IFEITDQEFK	ESFSKTHSL	
	REAFNHIPL	LLSNNEGKNY	ESLKHSNEEL	KKRTEKAELK	AQQNQKYLKD	
	NFLKEALVPL	SVKTSVLQAI	KIFNQIIKNF	GKKEYSQV	IEMARELTKP	
	NLEKLLNAT	NSNIKILKEK	LDQTEKFDDF	TKKKFIDKIE	NSVVRNKLK	
	LWFEQDRKDP	YTQLDIKINE	IEDETEIDHV	IPYSKSADDS	WFKNLLVKKS	
	TNQLKKNKTV	WEYYQNESDP	EAKWNKFVAV	AKRIYLVQKS	DKESKDNSEK	
	NSIFKNKKPN	LKFKNITKKL	FDPYKDLGFL	ARNLNDTRYA	TKVFRDQLNN	
	YSKHHSKDDE	NKLFKVCMN	GSITSFLRKS	MWRKNEEQVY	RFNFWKKDRD	
	QFFHHAVDAS	IIAIFSLLT	TLYNKLRVYE	SYDVQRREDG	VYLINKETGE	
	VKKADKDYWK	DQHNFLKIRE	NAIEIKNVLN	NVDFQNVQRY	SRKANTKLNT	
	QLFNETLYGV	KEFENNIFYKL	EKVNLF SRKD	LRFKILEDLN	EESEKNKKNE	
	NGSRKRILTE	KYIVDEILQI	LENEEFKDSK	SDINALNKYM	DSLPSKFSEF	
	FSQDFINKCK	KENSLILTFD	AIKHNDPKKV	IKIKNLKFFR	EDATLKNKQA	
	VHKDSKNQIK	SFYESYKCVG	FIWLKNKNDL	EESIFVPINS	RVIHFGDKDK	
	DIFDFDSYNK	EKLLNEINLK	RPENKKFNSI	NEIEFVKFVK	PGALLNFEN	
	QQIYYISTLE	SSSLRAKIKLLN	KMDKGKAVS	MKKITNPDEY	KIIEHVNPL	
	GINLNWTKKL	ENNN				
Flavobacteri um branchiophil um FL-15	MAKILGLDLG	TNSIGWAVVE	RENIDFSLID	KGVRIFSEGV	KSEKGISSSR	(SEQ ID NO: 70)
	AAERTGYRSA	RKIKYRRKLR	KYETLKVLSL	NRMCPISIEE	VEEWKKSFGK	
	DYPLNPEFLK	WLSTDEESNV	NPYFFRDRAS	KHKVSLFELG	RAFYHIAQRR	
	GFLSNRLDQS	AEGILEEHCP	KIEAIVEDLI	SIDEISTNIT	DYFFETGILD	
	SNEKNGYAKD	LDEGDKKLVS	LYKSLLAILK	KNESDFENCK	SEIERLNKK	
	DVLGKVKGKI	KDISQAMLDG	NYKTLGQYFY	SLYSKEKIRN	QYTSREEHYL	
	SEFITICKVQ	GIDQINEEEK	INEKKFDGLA	KDLYKAIFFO	RPLKSQKGLI	

Table 4- Cas9 orthologs						
WP_014084 151.1	GKCSFEKSKS	RCAISHPDEF	EYRMWTYLNT	IKIGTQSDKK	LRFLTQDEKL	
	KLVPKFYRKN	DFNFDVLAKE	LIEKGSSFGF	YKSSKKNDF	YWFNYKPTDT	
	VAACQVAASL	KNAIGEDWKT	KSFYQQTINS	NKEQVSRTVD	YKDLWHLLTV	
	ATSDVYLYEF	AIDKLGLDEK	NAKAFSSTKL	KKDFASLSLS	AINKILPYLK	
	EGLLYSHAVF	VANIENTIVDE	NIWKDEKQRD	YIKTQISEII	ENYTLKSRF	
	EIINGLLKEY	KSENEGKRV	YYSKEAEQSF	ENDLKKKLVL	FYKSNEIENK	
	EQQETIFNEL	LPIFIQQLKD	YEFIKIQRLD	QKVLIFLKGK	NETGQIFCTE	
	EKGTAEEKEK	KIKNRLKKLY	HPSDIEKFKK	KIIKDEFGNE	KIVLGSPLTP	
	SIKNPMAMRA	LHQLRKVLNA	LILEGQIDEK	TIHIEMARE	LNDANKRKGI	
	QDYQNDNKKF	REDAIKEIKK	LYFEDCKKEV	EPTEDDILRY	QLWMEQNRSE	
	IYEEGKNISI	CDIIGSNPAY	DIEHTIPRSR	SQDNSQMNKT	LCSQRFNREV	
	KKQSMPIELN	NHLEILPRIA	HWKEEADNLT	REIBIISRSI	KAAATKEIKD	
	KKIRRRHYLT	LKRDYLQGY	DRFIWEEPKV	GFKNSQIPDT	GIITKYAQAY	
	LKSIFYKKVES	VKGGMVAEFR	KIWGIQESFI	DENGMKHYKV	KDRSKHHTHT	
	IDAITACMT	KEYDVLAHA	WTLEDQQNKK	EARSIEEASK	PWKTFKEDLL	
	KIEEEILVSH	YTPDNVKKQA	KKIVRVRGKK	QFVAEVERDV	NGKAVPKKAA	
	SGKTIYKLDG	EGKKLPRLQQ	GDTIRGSLHQ	DSIYGAIKNP	LNTDEIKYVI	
	RKDLESIKGS	DVESIVDEVV	KEKIKEAIAN	KVLLLSSNAQ	QKNKLVGTVW	
	MNEEKRIAIN	KVRIYANSVK	NPLHIKEHSL	LSKSKHVHKQ	KVYQNDENY	
	AMAIYELDGK	RDFELINIFN	LAKLIKQGGG	FYPLHKKKEI	KGKIVFVPIE	
	KRNKRDVVLK	RGQQVVFYDK	EVENPKDISE	IVDFKGRIYI	IEGLSIQIRIV	
	RPSGKVDEYG	VIMLRYFKEA	RKADDIKQDN	FKPDGVFKLG	ENKPTRKMNH	
	NQFTAFVEGI	DFKVLPSGKF	EKI			
Eubacteriu m yurii subsp. margaretiae ATCC 43715 EFM38267. 1	MENKQYYIGL	DVGTSNVGWA	VTDTSYNLLR	AKGKDMWGAR	LFEKANTAAE	(SEQ ID NO: 71)
	RRTKRTSRRR	SEREKARKAM	LKELFADEIN	RVDPSFFIRL	EESKFFLDDR	
	SENNRQRYTL	FNDATFTDKD	YYEKYKTIFH	LRSALINSDE	KFDVRLVFLA	
	ILNLFSHRGH	FLNASLKGDG	DIQGMDFVYN	DLVESCEYFE	IELPRITNID	
	NFEKILSQKG	KSRTKILEEL	SEELSISKKD	KSKYNLIKLI	SGLEASVVEL	
	YNIEDIQDEN	KKIKIGFRES	DYEESLKVK	EIIGDEYFDL	VERAKSVHDM	
	GLLSNIIGNS	KYLCEARVEA	YENHHKDLLK	IKELLKKYDK	KAYNDMFRKM	
	TDKNYSAYVG	SVNSNIAKER	RSVDKRRKIED	LYKYIEDTAL	KNIPDDNKDK	
	IEILEKIKLG	EFLKKQLTAS	NGVIPNQLQS	RELRAILKKA	ENYLPFLKEK	
	GEKNLTVSEM	IIQLFEFQIP	YYVGPLDKNP	KKDNKANSWA	KIKQGGRIPL	
	WNFEDKVDVK	GSRKEFIEKM	VRKCTYISDE	HTLPKQSLLY	EKFMLVNEIN	
	NIKIDGKIS	VEAKQKIYND	LFVKGKKVSQ	KDIKKELISL	NIMDKSVLS	
	GTDTVCNAYL	SSIGKFTGVF	KEEINKQSIV	DMIEDIIFLK	TVYGDEKRFV	
	KEEIVEKYGD	EIDKDKIKRI	LGFKFSNWGN	LSKSFLELEG	ADVGTGEVRS	
	IIQSLWETNF	NLMELLSSRF	TYMDELEKRV	KKLEKPLSEW	TIEDLDDMYL	
	SSPVKRMIWQ	SMKIVDEIQT	VIGYAPKRIF	VENTRSEGEK	VRTKSRKDRL	
	KELNGIKED	SKQWVKELDS	KDESYFRSKK	MYLYYLQKGR	CMYSGEVIEL	
	DKLMDNLYD	IDHIYPRSFV	KDSDLNLLVL	VKKEINNRRQ	NDPITPQIQ	
	SCQGFWKILH	DQGFMSNEKY	SRLTRKTQEF	SDEEKLSEFIN	RQIVETGQAT	
	KCMAQILQKS	MGEDVDVVFS	KARLVSEFRH	KFELFKSRLI	NDFHHANDAY	
	LNIVVGNSYF	VKFTRNPANF	IKDARKNPDN	PVYKYHMDRF	FERDVKSSE	
	VAWIGQSEGN	SGTIVIVKKT	MAKNSPLITK	KVEEGHGSIT	KETIVGVKEI	
	KFGRNKVEKA	DKTPKKPNLQ	AYRPIKTSDE	RLCNILRYGG	RTSISISGYC	
	LVEYVKKRKT	IRSLEAIPVY	LGRKDSLSEE	KLLNYFRYNL	NDGGKDSVSD	
	IRLCLPFIST	NSLVKIDGYL	YYLGGKNDNR	IQLYNAYQLK	MKKEEVEYIR	
	KIEKAVSMK	FDEIDREKNP	VLTEEKNIEL	YNKIQDKFEN	TVFSKRMSLV	
	KYNKKDLSFG	DFLKNKKSKE	EEIDLEKQCK	VLYNIIFNLS	NLKEVDLSDI	
	GGSKSTGKCR	CKKNITNYKE	FKLIQQSITG	LYSCEKDLMT	I	
Acidovorax ebreus	MAQHVFGLDI	GIASVGWAIL	GEQRIIDLGV	RCFDKAETAK	EGDPLNLTRR	(SEQ ID NO: 72)
	QARLLRRRLY	RRAWRLTQLS	RLKLRKGLIA	DAKLFAKAPS	YGDSAWELRR	
	QGLDRLLTPL	EWARVIYHQC	KHRGFHWTSK	AEEAKADSDA	EGGRVKQGLA	
	HTKALMQAKN	YRSAAEMVLA	EFPDARNRKR	GQYDKALSRV	LLGEELALLF	
	ATQRRRLGNPH	ASDFFEKLIL	GDGDRKSGLF	WQQKPALSGA	DLLKMLGKCT	

Table 4- Cas9 orthologs						
WP_012655 176.1	FEKGEYRAPK YQTETSKYKT LVKLPAPWHEL GAEVVQQLRQ RYDEAVAQIP FRDDADIPRN ERNKVKRAQE YSQQPLDIQR RTAFEYLTSTF KGFIDRNLD LRARWGLTKV GFPDPEETGEI QRLGSYTTED EKILLTSLKL KAFGPDNLFH VDIFTKAGKF VYPNDYVKVT GVKTALSVEK	ASFSVERHVW LKNAFIKAGL RKAFKAAGHE LALPEPAASI EYGHHSQRIE PVVLRALNQA EFRDRNDRAR VLDDHNYAQV PDGEDGERWR TRYICKFFKN RGDSDRHHAL LNPAAFDRAR LGRLRTLFSV QDFDKLLNPE KPKDKNNQPTG HLVPVYVHHR LKKEQQSGYY FNVDVLGRIY	LTRLNNLRIV WGDGVRFGGL ALWQQISTPA AVLEKISFDK PGAACHLYLP RKVVNALIRE SEFERDFGYK DHALPYRSRY TFVAWVQGNK YVEEHLQLAA DAAVVAATCH QHFPEPWTHF RAPQRRSGGA SNDHFVEPHR PVVRSIKLVR VTGLPNRAIV SGADRSTGAM LAPPETRSLG	VDGRSRPLNE AYPSQAQIDA LDGDPTLLDQ FSSLSLKALR PFYEAQRKYA YGSPIAVNIE PKAAAFEKWM DDSKNNKVLV AYRMAKRNRL RADGDTARRC GMVKALADYS AHELKARLFT VHKETIYAQP NERLYAAIRQ GKQTGIPIRG AFKDEDEWTL NLWAHDRAAS	AERQAALLLP EKTKDPEDQF IATVLSVYKD RIVPLMQSGL GKGDHIGSMQ MARDLSRPLD LYREQLGQCA LTHENQNKGN LRKNYGVDES VVVNGQLTAF RRKEISFLQE DDLAALREDM ESLKQGGVI RLEQFGGRAD GLAKNDSMLR IDESFAFLFS VGKDGILIRGI	
Porphyromonas sp. oral taxon 279 str. F0450 WP_009433 518.1	MLMSKHVLGL EAFNAGAAFT IQLPLLELWE VGDEGEKKKD GGISYRIKDQ MQRPLKSKCH KSSPLFQLCC AALKKLLKEK ETRMMTVQLT RRALITQLGM GLGYSEACAA NQMINLVNAL VAAKIRECGL EHIIPKSVLY NDLLKEKKIS IRRVASASEGG ELHITNWSKR QEAQEQQATE TRGRNIYRKK RYPLHDLKGE SVRCYAKTLS IVTFWDAVDR VDSLQQDEM HLETSVADDK	DLGVGSIGWC ASQERTARRT LRERAATAGR SNSAYLAGIR IFSRQCYIDE LVSLCEFEKQ IYEAVNNIRL ALIADQLTSK DEETGEVTER KEEDLDGGLL VGYRHSNSPT KAEYGIDEVR YPTKPRIQKY DDSYGNKTCA YSKHQRLRWL VTARLRSLWG MDHRHHAIDA TGRLSNLERW MADGREVSCV VVDPHLRELI LDKAIPMCFD ARYGIPLVIT VIGLSDEELQ NTSGRIPKFH	LIALDAQGDP MRRGFARYQL RLTLPELGRV ANDEKLQAEH YDQIMAVQRV ERVMRVQQDD TRPNGSPCDI SGLKGNSTRV EVAVVTDSYV DQLYRLDFVK PGYGNKSAKF VELARELKMS MLWKEAGRQC CRRCNKEKGN KEDIPSDFLE YGKILHTLNL LVVACTRQSY LTQRPHFSVR QRGVLVPRGE TTYNQELKSR EKGEPTAFVK HPREVMEQVL RALEAQNYRK RVQSLKAYEE	AEILGMGSRV RRYRLRRELE LCHINQKRGY KTVGQYFAEQ HYPDILTDEF GKGGWQLVER TPEERAKIVA ALASALQYPY RKPLYRLWHI EKIPLQORNE REERERMARN LYCGRSIEEE RTALEYIRAK RQLRLTQYIS DRYDSMGETE IQRLNRLSSE TVSDKVAEIL LMEASFYGKI EKGAPIPLPC SASNHHLLALY QRGDIPEQVL ISEHLYRVQK RNIRKVRVDL	VPLNNATKAI KVGMLPDAAL RHVKSDDAAAI LRQNQSESPT IRMLRDEVIF RVKFGPKVAP HLQSSASLSF QYHHLLDMEL LYSIEEREAM ICKLLPQLQQ LRQPLVEKIL NKDREERNKG QCLREGGMEV GREAEYMKRI RQAMAILQQG RVSREGEATE FGREDKKKED ISYRPGQRVV LSQGRVRIVK LDKDKKQEV RTPKGKLVES SLLPPSDWVF MSSSYVFRY LGRISLL	(SEQ ID NO: 73)
Mycoplasm a ovipneumon iae SC01 WP_010320 922.1	MHNKKNITIG RSTRNIRRK ISFFTEIYKK DLEENVADK WVNEIKKLFE DEKGVFKKY TDLKTTNKKI DFGFEKSDID INNILEFLPY DFVKSIFSNT RKWEEQSKL KYSKENIIDA NKGYKLSLE IEHIIPYSMS	FDLGSIASIGW AYRNQRFINL CAAKHSNILE YEGIEHPSIL VQEINPEFSE DNIWDKTIGK WQLSSNDRNE DQDTIEGRKI LDAICIIDLR KFNFKKIGNF GKTDKKTLYL IIIESPREKN TKPAKLLDRL YDNSQANKIL	AIIDSTTSKI ILKYKDLFEL VKVKALDSKI LYDFFKKNF KFLNLFTSVR CSFFVEENRS LLDELLKVKE IKEEPTTKLE EKSRGQDEVL SLKAIREFLP NPRIQDEII DKKTIEEIKK RFYHQDGDID TEKAENLKKG	LDWGTRTFEE KNISDIQRAN EKLDLIWILH FKSNSSIPKD DYAKGPGSEH PVNYPSEYIF KAKIISISLK VTKHLLATIY KKLTEKNIFE KMFQONKNSE SPGKNTFEQ RNKKGKGKTL LYTLDKINID KLIASEYIKR	RKTANERRAF KKDTENYEKI DYLENRGFFY LGGYSFSNLQ SASEYGIFQK NLLNQILNLS KNEIKKIILK SHSSDSNWIN VLKIDREKQL YLKWKDEEIR AVLVNLQIIK EKLFIQILNLE QLINGSQKYE NGDEFYNKYY	(SEQ ID NO: 74)

Table 4- Cas9 orthologs						
	EKAKELFINK NLNDTRYSTK FRAKPVQKNN IANLLTLADN KIIQEKYEEA IKKKLVTSTK EKYNGFSNKT QFTYHDNINN LYSLVYKVYE SDFAKPIPVN SKRFNRDTAY IKEHTDDEKL	YKKNKKLDSY LFYHALVEHF GPNENLNNNK KTDKKFLLHD KKHTAIKFSS EELKKYFENP GNAFVEYMND KSFKRIFYKNI NGKESYKSIP CKPVFVLKKG GLKPLKLSVV MCTIK	VDLDEDSAKN ENNEFFTYID PEKIEKNREN ENYKENIETG KTRTILNGGL FGKKADGKSE LALKEPTLKA RIIEYKSIPI VTSQMRNFGI SILKKKSLDI KPVAEPSTNP MCTIK	RFRFLTLDQY ENSSKHVKVI NEHHAVDAAI ELVKIPKFEV SDETLYGFKY YTVLMAQSHL EIESAKSVEK KFKILSKHDG DEFDFLDENL DDFKETKETE IFKEYIPIHL	DEFQVEFLAR STIKGHVTKY VAIIGNKNPQ DKLAKVEDLK DEKEDKYFKI SEFNKLKEIF LLYYNFKPSD GKSFKDTLFS YNKEKLDIYK EGNYYFISTI DELGNEYFPVK	
Wolinella succinogenes WP_011139 431.1	MLVSPISVDL SKRNNLRNKL DEKKKDALES FEAVFASATH ENQGNLPRAK LKELRRYFND KDLKSKEIME NWRDIAKRLN KKPNELLVKT LEDNSNMLKR VGNKKLSAYC DDVILNEWSQ KRCTAENRFR DKLGYELAKI VINSKKDRDG EIDHILPRSH SAQWIEEQVA TDQSARVNGT ARQNPLLAKE LDSWNPASNE GERYKPIVVQ DLISLKENQY VENCRYYTKK KDRFVIDKSS KSIRIKARKT ANILLEANSK ADFDTHGIRL TTISFKESGP SEHIQKLLK	GGKNTGFFSF VKRLFLLILQ DTLKEFLSEK SPNKKLELKD YFEELGEYIA KEMEKGDIWI LLTTTEPVMT HGKFNDDLAD LGQSDANRLY CNHNPPHKKK KNIEELRKTH KIANFFDIDD SETAFYNDET KAKELEGMEA KKLAKAKENA TLKIYGTVMN NIKGYKTFSV QKYLAKKIQE EKQAPSSHAI PLTKGLSTNQ EGGYYIGYPA IKIFSINKQT VDVKFAPKYI LNEYYPDPKN FSLLAESGVQ TSLSLVPKSGI IVENTSVKVG VQDYLVKFLK	TDSLDSNSQSG EHHGLSIDVL LQSIDRSDSV ELKSEYGENA TNEKVKSFDD PNKLHKITER IPPYDDMNRR STVKGYSDDS GFAQNYIELI QIHNLVAGIL GNTFKIDIEE KHRQRFNNLF GEFHKKATAT KEIKVPIILE EDRLKDKDKR PEGNLIYVHQ LSAEQQKAFR KLTKMLPNKH DAVMAFVARY KIEKMIKSGD TVKKGYELKN ISELSNRYFN HETKYPFYDD EYKLDVDTQP GKVFRARSKI SIEKQLDKKL NFPISAIDKS KIVKIQLRTD	TVIYDESFVL PDEIRGLFNK EDFLNQIASN KELLAGLRVT SNSLKLDTMT FVRSWHPKND GAVKQCOTLRL TLLHRLLDTS RQKVRAGIWW GVKLDEAKFA LRKKDPAELS SMAQLHTVID CQRLPADTQR QNAFEYEESL IKAFSSGICP KCNQAKADSI YALFLQNDNE LSFEFILADA QKVFKDGTTP YGQKNMREVF CKVVTSKNDI MNYKNLVERD WRRFDEAWRY IWDDFCRWYF PTGYAYQALP DVIKKT DVRG AKRMIFRALF GSISNIVVRK	SQVGRRSKRH RGYTYAGFEL AESFKDYKKG KEILDEFDKQ KLIGNISNYQ ADRQRAELM NEEYLDKHLR KEIDIYELRG PVKNKDDSLN EFEKELWSAK KEEKAKLRLT TPRSGFSSTC PFSGKIERYI RKSKTGSNDR YCGDTIGDDG YKLSDIKAGV AYKKVVDWLR TEVSELRRQY NADEVAKLAM GKSIFGENAI AKLEKIKNKQ KEIVGLLEFI LQENQNTSS LDTRYKTANDK MDNNVIAGDY LAIDNNSFFN EKEKGKRRKK NAADFTLSFR	(SEQ ID NO: 75)
Streptococcus mutans UA159 WP_002263 549.1	MKKPYSIGLD LLFDSGNTAE LEDNFLVTE LRLVYLALAH LQEQNVQVEE DFKKHFELEE LLSGILTVTD FSDVSKDGYA KQRTFDNGSI YVGPLARGKS YLPNQKVLPK VFKVYRKVTK KILDKDFLDN KLERRHYTGW DALSFKEEIA	IGTNSVGWAV DRRLKRTARR KRGERHPIFG IIKFRGHFLI ILTDKISKSA KAPLQFSKDT VGTKAPLSAS GYIDGKTNQE PHQIHLQEMR DFAWLSRKSA HSLLYEKFTV DKLMDFLEKE SKNEKILEDI GRLSAELIHG KAQVIGETDN	VTDDYKVPK RYTRRRNRIL NLEEEVKYHE EGKFDTRNND KKDRVLKLF YEEELEVLLA MIQRYNEHQ AFYKYLKGLL AIIRRQAEFY DKITPWNFDE YNELTKVKYK FDEFRIVDLT VLTLTFLFEDR IRNKESRKT LNQVVSIDIAG	KMKVLGNTDK YLQEIFSEEM NFPTIYHLRQ VQRLFQEFLLA NEKSNGRFAE QIGDNYAELF DLAQLKQFIR NKIEGSGYFL PFLADNQDRI IVDKESSAEA TEQGKTAFFD GLDKENKVFN EMIRKRLNENY LDYLIDDGNS SPAIKKGILQ	SHIEKNLLGA GKVVDSFFHR YLADNPEKVD VYDNTFENSS FLKLIVGNQA LSAKKLYDSI QKLSDKYNEV DKIEREDFLR EKLLTFRIPY FINRMTNYDL ANMKQEIFDG ASYGTYHDL SDLLTKEQVK NRNFMQLIND SLKIVDELVK	(SEQ ID NO: 76)

Table 4- Cas9 orthologs						
	IMGHQPENIV VENSQIQNDR SIDNRVLTSS TKAERGGGLTD RQVKIVTLKS YPQLEPEFVY IIWKKDEHIS RKTKKFYWDT MEKMTFERDP ELQKGNEIVL SNFSKKYTLA FKFFDKNIDR	VEMARENQFT LFLYYLQNGR KENRGKSDDV DDKAGFIKRQ NLVSNFRKEF GDYPHFHGHK NIKKVLSYPQ KKYGGFDSPI VAFLEKGYR PNHLGTLTYH EGNLEKIKEL KRYTSTTEIL	NQGRNSQQR DMYTGEELDI PSKDVVRKMK LVETRQITKH ELYKVREIND ENKATAKKFF VNIVKKVEEQ VAYSILVIAD NVQEEENIIL AKNIHKVDEP YAQNNGEDLK NATLIHQSI	LKGLTDSIKE DYLSQYDIDH SYWSKLLSAK VARILDERFN YHHAHDAYLN YSNIMNFFKK TGGFSKESIL IEKGSKKKLK PKYSLFKLEN KHLDYVDKHK ELASSFINLL GLYETRIDLN	FGSQILKEHP IIPQAFIKDN LITQRKFDNL TETDENNKKI AVIGKALLGV DDVRTDKNGE PKGNSDKLIP TVKALVGVTI GRKRLLASAR DEFKELLDVV TFTAIGAPAT KLGGD	
Prevotella timonensis CRIS 5C- B1 WP_008122 718.1	MNKRIILGLD SSKAAERSGY QKQYPKSDEL LYHLTQRRGF YKLYDAQGNK FQLPLKSQRH LELRPLTYEE AYKFNYRMTQ EMVDDVWNVL LKAIKFLPF LVFNYQPKHR QRNEFGILQL VEYARELND TDVLKFQLWE QMNLTLCNDR QRTFAGMDKA TGIGLISRYA EKKCRDNHSH KPWATFTEDV GSLHLDITYG IADKNFKQAI YKQYHVMNE FSSIVPTHST VVGIEKDGR VDGIDFTIDI	GTNSLGWAVV KAIRKQYFRR MLWQRTSDEE LSNRLDTSAD VRIRQRYTDR GVGRCTFERG REKIEPLFFR GVPGCPTIAQ YSFSSVEKLE LRKGMYYTHA EVQGTIEMLI GSPRTNAIRN NKRRAIADRO EQNHHCCLYT FNREVKKAKL VKDIRIQKRH GLYLKSLFHQ HCMDAITIAC LNIYKNLLVV AIERDGEIRY AEPIYMNEEK NNYLLAIYEG KYGLPLKTKL KFKYHQEARK LGKVTILKE	DWDEHAQSYE RLRKIQVLKV GKNPPYDRHR NKEDGVVKS NKHYQHEFDA KPRCADSHPD KSKPNFDFED LKSIFGDDWK EFAHKKLQLD SFFANIPTIV KDFLANNFEL PMAMRSLHIL KEQDKQHKKY EQIGITDFIG PTELANHEEI KLQMEIDYWR ADSRNKSNNV IGKREYDLMA HDTNPNMPKH VVRRLSSFT GILIKKVRCF LVKNKVREF LMGQLVLMFE EGLPIFSTPY	LIKYGDVIFQ LVKYHLCPYL CLHEKLDLTV ISQLSTEMEE ICEKQELSSE YEEFRMLCFV IAKALAGKKK TGIAETITLI EESAEEKFAKI GKEIWNKEQN PAGATDKLYH RRVVNQLLKE GDEIRKLYKE SNPKFDIEHT LTRIEPWKNK VKYERFTMT VVKGVATAEF EYRMEETFK TKKYVQTSIG KPEELENIVD AKSVKQPINI EIVSYIEAAK ENPDEIQVDN KNNDYAPIF	EGVKIEKGIE SDDDLRQWHL EADRYTLGRA AGCEYLGDF LIEDLQRAIF NNIQVKGPHD YAWIHDKEER QKNGSKSLQ KLSHSFAALS RKYIMENVGE PSMIETYPNA SIIDENTEVH ETGKDIEPTQ IPQSVGGDST YEQLVKERDK VPEGFRRQ RKMWGLQSEY QGRGSKPKFS KVLAQGDAR ETVKRTIKEA RQHRDLKKE YYKRSQDRNI TKDLVKRLYK RQSINNINIL	(SEQ ID NO: 77)
Clostridium cellulolyticu m H10 ACL77411. 1	MKYTLGLDVG RIARGMRRRI WELRYNALS KNNSEMLRTK YIFSIQRKLG ELRAPTSCYT EIKYSEIRKL LPTDIWGLKH LPTFNKFKHL LTVEPIIENV RQDRDNLKKE AYSGKPIPV YTPYEVWGST LNDTRYISRF REESDLHHA QDLMETLAGV LEKLETMVKN TNGAIRSIR YVAHMIKKEL	IASVGWAVID SRRSQRLRLV ILKPYELVQV NYRTIGEMIF SPFVTEKLEH SEYFGLLQSI LDIEPEILFK SNKESLDNLF SLVAMKRIIP TNPVVIRALT HENNRKAREK DLLNDSLTIQ EKWEDFEARI LKNYIESYLQ DAAVIACADR FISRAPRRKI TKGGISDKAV VETPSYTGVE PSKAIVPLKP	KDNKKIIDLG KKLFVQYEII LTHITKRRGF METPENSNNR DFLNIWEFQR NNLVLVEDNN AHNLTHKNPS YCLTVYKNDN FMEKGYKYS QARKVINAI ISDLIRQNGR DHIYPYSRSM YSMHLPSKE FSNDSKSCV KIIKEITNYY TGPAHDETIR YNVLKNRLIE RNEGKGISDN ESQWELIDST	VRCFDKAEES KDSSEFNRI KSNRKEDLST NKVDEYIHT PFASGDSILS TLTLNNDQRA GNNESSKKFY EIKDYLOANN ACNMAELDFT QKYGLPYMVN VASGLDILKW DDSYMKNVLV KRLNLRNFIT VCVNGQCTAQ NERENHNYKV SPKHFNKGLT HNNKPLKAF SLMVRVDVFK HEFLFSLYQN	KTGESLATAR DTSRDGWKDP TKEGVVITS AREDLLEIK KVGKCTLLKE KIIIEYAHFN MKSYYHLKST LDYLIBEYIAK GSSKLEKCNK IELAREAGMT RLWEDQGGRC LTDENQNKRS KDLSFISRN LRSRWGLNKN KYPLPWSHFR SVKIPLTTVT EKIYKPLKNG KKDKYYLVP DYLVIKTKKG	(SEQ ID NO: 78)

Table 4- Cas9 orthologs						
	ITEGYYSRCH KSIVKGEPRR	RGTGSLSLMP GMEKYNSEFKS	HFANNKNVKI N	DIGVRTAISI	EKYNVDILGN	
Francisella tularensis subsp. novicida U112 WP_003038 941.1	MNFKILPIAI LMNNRTARRH RRGFSFITDG ESKISEIYNK YLANYESLK LDTLTDDLD IKSEMASGGR KNLVNLIGNL RVGVKQDQDK DNNNRKPPKC LKVLSKSKDQ LLLNEIYFQA IFEQGTFLHL NQLLTYCNHK EHIRGFKKAC GSEDKKGNK QIAFAERKGN PAIPTRIVDG AFEFEPALAD SGANLTDGDF RIFCLRDLD NLTPQEQKAF NNIYLRAKKE AKGDKPQASY GEVFTKDIFS PILIHKELNE SIDIQISTLE YKKYSKEMEF EWQRLYREWQ KFLVKRKTWD TSKNIFWLPK IDNNSRPKVR EFESSGFNKT	DLGVKNTGVF QRRGIDRKQL YSPEYLNIVP LMQKILEFKL TQKFSYTDKQ IWNFNFEKFD HRSQYFQEIT SNLELKPLRK DGAKEYSYKDL QSLILNPKFL PYFVEYKSSN KKLKQKASSE VCKYYKQRQR PRQKRYQLLN EDSLKIQKDN HGLAYELGVL ANTCAVCSAD AVKKMATILA VKGKSLKDRR DGAKEELDHI NYKLLQFETT RHAFLLADEN NLNTDKISFD SHLIDAMLAF QIKITDNEFS VRKGYTWKNS ELRNILTTNN LRSLAYRSE NTTIKDDYEF NNFIYQILND NIELQKVDNK VKLDYVIDDD IKEMLGMKLA	SAFYQKGTSL VKRLFKLIWT EQVKAILMDI MKLCTDIKDD GNLKELSYHH FDKNEEKLQN NVLDENNHQE YFNDKIHAKA CNELKQKVTK DNQYPNWQQY QQIASGQRDY LEKLESSKKL ARDSRLYIMP DLAGVLQVSP RGLLNHKINI LFGEPNEASK NAHRMQQIKI KNIVDDNWQN KKALERISPE IPRSHKKYGT DDLEIEKKIA PIKQAVIRAI YFGIPTIGNG CIAADEHRND DKKLVRKKAI EIKIFKGKK IAATAEYYYY VKIKSIDDVK LKSEFFNVKSI SDSRADGTFP NIFAIDTSKW SKINYFMNHS GIYNETSNN	ERLDNKNKGV EQLNLEWDKD FDDYNGEDDL KVSTKTLKEI HDKYNIQEFL QEDKDHIQAH GYLKNFCENL DHWDEQKFTE AGLVDFLLEL LQELKKLQSI KDLARILQF DEVIANSQLS EYRYDKKLHK NFLKDKIGSD ARNTKGKCEK PEFDRKIKKF TEPVEDNKDK IKQVLSAKHQ NIFKDKNNRI LNDEANLICV DTIWDANKKD NNRNRTFVNG RGIAEIRQLY GSIGLEIDKN EGFNTHRQMT YDIQQLNNLV NLKTQKLHEY QVLDKDSNFI TKLHKKVRKD FIPAFDISKN FEVETPSDLR LLKSRYPDKV	YELSKDSYTL TQQAISFLFN DSYLKLATEQ TSYEFELLAD KRHATINDRI LHHFVFAVVK HNKKYSNLSV TYCHWILGEW DPCRTIPPYL QNYLDSFETD IFDRVKASDE QILKSQHTNG YNNTRGFDDD DDLFISKWL EIFNLICKIE NSIYSFAQIQ IILSAKAQRL LHIPITESN KEFAKGISAY TRGDNKNKGN FKFGNYRSFI TQRYFAEVLA EKVDSDIQAY YSLYPLDKNT RDGIYAENYL YCIENYNTALG IGKITLPFKK FSLPISTNEG EIVEAIIIDSF DIGIATIQYK LEILKQSTII	(SEQ ID NO: 79)
Azospirillum sp. B510 AOL40891. 1	MARPAFRAPR GTSRDHTDKF GSRI FSDGRD MPADPAARKR RTATKPDEEA AGKGKEAAYP HQRPLKPPPV RPLTPAERDR SLES DKRPEL AIAALTARWA PDGRVRPIRL TGNPADPEEK DLKRS AEDRR QGPVENRRCP EANRIKRNRS EGGLRARHLN LAEADGPPPE LAAASAEREA HGIGGALHKE LIDRLAIRRR EHGNGPSGPR	REHVGWTPD AEWECAVQPY PQDKASLAVA LEVAVDPYLA GKVKEAVERL FYPARRMLEA GRCTLYPDDG IVAFVQGRPP LGDETGARIA LDEATAAKLA DEAVKLLRGG RVGRVANPTV REDKRQADNQ YSGETISMRM PWEAFGHDSE DTRHLSRLAV VPAETLDPSP AAREDNIRRV TAYGPVDPPE DANDPATALA SGGPFHKKLL	PHRISKPFPI RLSFDLGTNS RRLARQMRRR RERATRERLE EAAIAAAGAP EFDTLWAEQA RAPRALPSAQ KAGRKPGKVQ PAFGPGWTAL GATLPDFHGR KDHSDFSREG HIALNQLRHL KRNEERKRLI LLSEQVDIDH RWAGILARAE EYLRVCVCPKV AEKNRADHRH LEGFKEEPWD EGFNLVVRKP KAAEDLAAQP AGEVHHVDVA	LVSWHLLSRV IGWGLLNLDR RDRYLTRRTR PFEIGRALFH TLGAWFAWRK RHHPDLLTAE RLRLFQELAS KSVPFKEKLRG PLEEQDALVE YGRRAVAELL ALLDALPYYG VNAILARHGR LSLGERPTPR ILPFSVSLDD ALPKNKRWRF RVSPGRLTAL HALDAVVIGC GFRAELERRA IDGLSKDEIN ASRGIRRVRV LRADGRRWVG	VIDSSSGCFP QGKPREIRAL LMGALVRFG LNQRRGYKPV TRGETLRARL AREILRHRIF LRVIHLDLSE LLELPPGTGF LLLTEAEPPER PVLERETRGD AVLERHVAFG PEEIVIELAR NLLKLRLWEE SAANKVVCLR APDALEKLEG LRRRWGIDAI IDRSMVQRVQ RTIVVSHRPE SVRDPRLRRA LKKESNPIRV HWTLFEAHG	(SEQ ID NO: 80)

Table 4- Cas9 orthologs						
	GRGADGAAAP SVVVVEPHQV GIGGDAGRTA	PRLGDGERFL KTDRSKHVKI MEPAEDIS	MRLHKGDCCLK SCDQLRARGA	LEHKGRVRVM RRVTVDPGLGR	QVVKLEPSSN VRVHAPGARV	
Peptoniphil us duerdenii ATCC BAA-1640 WP_008901 059.1	MKNLKEYYIG ERRTQSGSRR KDEKLKSKYT ACHYLLKNRG LIEIITDTTL EFEETT VKSV LKDADKSKDG KSINDNYVAY SENEDLKID KYYDFL NESD TPWNFKDIVD LNNLKLNEFL LSGTDGDFKQ YLKKKIKSAY KGSIIYFMRE SPSVKRMLWQ LLEFYKFGKK CMYSLEPIDL NQYPIPEKVL ERQIVETRQA NDLHHMHDAY VRGSYTAWIA NLMRKGKGQI ETIPIMVYNQ LYNIKGKTGD FDNIKEEELI KIDVLNQIIL ITGLFENEVD	LDIGTASVGW RLNRRKERIN LFNDKDFKDR HFIFEGQKFD NKTNKKKELK DFSTTAFDDK NKYISKAFVK TKSNITSNKR EMLTDIEFKT EYGTVKDKVE LDSSREEFID ITEEMKKAIF GLNSYIDFKN KNDFTDDEIK YNLNLMELEMS SLRVVDEIKR AFINEIGEER ADLKSNNIYD NKNAYGFWKI TKETANLLKN LNIVVGNVYN DDSEGNVCAA PQKENTNKS EKYGNTEAVD SLSIAGSVQL KLYKTLSDKL LFQSYNNGCN LLNL	AVTDESYNIP LLQDLFATEI DYHKKYPTIH TKNSFDKSIN NIVGDTKFLK YSEYEEALGD KFNKHGKDLK TKASKFTKQE FIPKLKSSDN SIMEFRIPYY RLIGRCTYLK EELFKTKKKV KIAALNYKDW GHYTFTEEVE IIGKDPKKIF YNYLLNEINS QDHIYPKSKI LFDKGLIGQK ICQDSEIVYS TKFTKNPLNF TIKKVKRELE IEKYGGYNKA KYLKDNLELQ IVNKEEQKLI NNGIYSNKR LKSIGLSAKT	KFNGKKMWGV SKVDPNFFLR HLIMDLIEDE DLKIHRLDEY AISAIMIGSS TISLLNILKS TLKRIKKYL IEMARAKEAK EESKFRWDN YDDSLNRLV KYTRLTRRTP KAENASRFRQ IKDKDNVRSY GKNYRFRTRMS SSAYFALIES DPKILKDKIK KKMDKFLVKK NQAKNISEAL GVVFIPKKLN	RLFDDAKTAE LDNSDLYRED GKKDIRLLYL NIDLEFNED QKLVDFEDG IYDSSILENL PSEYANIFRN DTIKETKLS ELKKILDNQS AWIKRENTKI IYNEFMVLNE KEFNLTGDIL LIVLYEDDKT IEGVDKTTGE CYEMVDELYL NSRKESRKNK LYLYYTQLGR VKKNLNHEKG FEERELAEFI EFDIICKRTV NLENMFYDV YIGTGGLYDQ DGKAGRERTL INSLIKLDGF KDNKDIKVT DKFKEISIEE YKECKLINQS	(SEQ ID NO: 81)
Lactobacillu s coryniformi s subsp. torquens KCTC 3535 WP_010014 406.1	MGRIGLDVVG AEPRRIKRGL SVYELRVAGL MGQLNSIEE NRQMLVDEIK EGPGAGPYAG QNTTGDWHT DARFNLVNYG LDHIGTALT MQNIIPYLEM TIKIVEQIIR AAELTELGIP IIPYSISWDD NIIRNSRKRQ NPELEKKQRV KFIQQVTKY FPLPWPEFLD HKISGPAHLD ENSNNGDKVV AKKVSPLPVL KAIARGKSLW VPIKNFYGYF GRYHKVHEKK	ITSTGYAVLK RRRTRRTKFR DRRLTNSLEF IRKAIAEKGY QILDQKREL DQIKKMVGKD LNGLDRAII SLQTQKEIET YSSDKRRRRY GQVYSEATTN RYGKPDGINI VNGQNIIRYK SYTNKVLTS KLLKQKLSDE LPLNGEVT QHGEVKNNQA ELLARISDN TIRSAKLFDE YSAIRQALEA KNKAAADNGS YQITESDQFC DGADIATASI RQLFVKRDE	TDKNGLPYKI KQRTQQLFIH RVLYFFIGHR RTVGELYLKD NEKLTDEFYA IFEPTEDRAA DAVFAKAEKP VEKKTRFVDF FAEELNLP TGDFRKKQI ELARELGRNF LHKEQNGVDP KCNREKGNRI ELKDWKQRNI IRSRWGFLKV LWHD AEIKDA VEMMKSRSWN KGIVLSRVSI HNGSGELAFP MVRIDVFNTG FSLYPGDMVH LVRAHDSSYT	LTLDSEVIYPR SGLLSKPEIE GFKSNRKAEL PKYNDHKNRK TYLLGDENRA KATYTFQYFN TKTYKPTDFG KAYHDLVKVL LIEKLLPLNF SKDTIREEIT KERGDIQKRQ YTGDQIPFER PMVYLANNEQ NDTRFITRVL REDGDLHHAI EYAAEAQRMD TYTPIEIAKL TKLKINKKGQ DGYLEYVDHG KKEFVFPYIY IESKTGIKPK ARSIGIAGLL	AENPQTGASL QILATPQAKY NPENEADKKQ GYIDGYLSTP GIFQAQRDFD LLQKMTSLNY ELRKLLKLPD PEEMWQSRQL SKFGHLSIKS NPVVRRAVTK DKNRQTNDKI AFSEGEYVDH RLNALTNIAD YNYFRQAIEF DATVIAAITP ADLFNKIFNG KPVFVVRLAN VATGDGIYDP TKKLVRKVRV KDTVEQVLPN YSNKENNTSV KFEKYQVDYF	(SEQ ID NO: 82)

Table 4- Cas9 orthologs						
Ignavibacterium album JCM 16511 WP_014561 873.1	MEFKKVLGLD	IGTNSIGCAL	LSLPKSIQDY	GKGGRLLEWLT	SRVIPLDADY	(SEQ ID NO: 83)
	MKAFIDGKNG	LPQVITPAGK	RRQKRGSRRL	KHRYKLRRSR	LIRVFKTLLNW	
	LPEDFPLDNP	KRIKETISTE	GKFSFRISDY	VPISDESYRE	FYREFGYPEN	
	EIEQVIEEIN	FRRKTGKGNK	NPMIKLLPED	WVYYLRLKKA	LIKPTTKEEL	
	IRIIYLFNQ	RGFKSSRKDL	TETAILDYDE	FAKRLAEKEK	YSAENYETKF	
	VSITKVKEVV	ELKTDGRKKG	KRFKVILEDS	RIEPEIERK	EKPDWEGKEY	
	TFLVTQKLEK	GKFKQNKPD	PKEEDWALCT	TALDNRMGSK	HPGEFFFDL	
	LKAFKEKRGY	KIRQYPVNRW	RYKKELEFIW	TKQCQLNP	NNLNINKBIL	
	RKLATVLYPS	QSKFFGPKIK	EFENSVDLHI	ISEDIIYYQR	DLKSQKSLIS	
	ECRYEKRKGI	DGEIYGLKCI	PKSSPLYQEF	RIWQDIHNIK	VIRKESEVNG	
	KKKINIDETQ	LYINENIKEK	LFELFNSKDS	LSEKDILELI	SLNIINSQIK	
	ISKKEEETH	RINLFANRKE	LKGNETKSRY	RKVFKKLGFD	GEYILNHPSK	
	LNRLWHSYD	NDYADKEKTE	KSILSSLGWK	NRNGKWEKSK	NYDVFNLPLE	
	VAKAIANLPP	LKKEYGSYSA	LAIRKMLVVM	RDGKYWQHPD	QIAKDQENTS	
	LMLFDKNLIQ	LTNNQRKVLN	KYLLTLAEVQ	KRSTLIKQKL	NEIEHNPKYL	
	ELVSDQDLEK	QVLKSFLEKK	NESDYLGKGL	TYQAGYLIYG	KHSEKDVPIV	
	NSPDELGEYI	RKKLPNNSLR	NPIVEQVIRE	TIFIVRDVWK	SFGIIDEIHI	
	ELGRELKNNS	EERKKTSESQ	EKNFQEKERA	RKLLKELLNS	SNFEHYDENG	
	NKIFSSFTVN	PNPDSPLDIE	KFRIWKNQSG	LTDEELNKKL	KDEKIPTEIE	
	VKKYILWLTQ	KCRSPYTGKI	IPLSKLFDSN	VYEIEHIIPR	SKMKNDSTNN	
	LVICELGVNK	AKGDRLAANF	ISESNGKCKF	GEVEYTLLKY	GDYLYQCKDT	
	FKYQKAKYKN	LLATEPPEDF	IERQINDTRY	IGRKLAEELLT	PVVKDSKNII	
	FTIGSITSEL	KITWGLNGVW	KDILRPRFKR	LESINKKLI	FQDEDDPNKY	
	HFDLSINPQL	DKEGLKRLDH	RHHALDATII	AATTREHVRY	LNSLNAADND	
	EEKREYFLSL	CNHKIRDFKL	PWENFTSEVK	SKLLSCVVS	YKESKPILSDP	
	FNKYLKWEYK	NGKWQKVFAI	QIKNDRWKAV	RRSMFKEPIG	TVWIKKIKEV	
	SLKEAIIQIA	IWEEVKNDPV	RKKKEKYIYD	DYQKVIKAI	VQELGLSSSM	
	RKQDDEKLNK	FINEAKVSAG	VNKNLNTTNK	TIYNLEGRFY	EKIKVAEYVL	
	YKAKRMPLNK	KEYIEKLSLQ	KMFNDLPNFI	LEKSILDNYP	EILKELESND	
	KYIIEPHKKN	NPVNRLLLEH	ILEYHNNPKE	AFSTEGLEKL	NKKAINKIGK	
	PIKYITRLDG	DINEEEIFRG	AVFETDKGSN	VYFVMYENNQ	TKDREFLKPN	
	PSISVLKAIE	HKNKIDFFAP	NRLGFSRIIL	SPGDLVYVPT	NDQYVLIKDN	
	SSNETIINWD	DNEFISNRIY	QVKKFTGNSC	YFLKNDIASL	ILSYSASNGV	
	GEFGSQNISE	YSVDDPPIRI	KDVCIKIRVD	RLGNVRPL		
uncultured delta proteobacterium HF0070_07 E19 ADI19058.1	MSSKAIDSLE	QLDLFKPQEY	TLGLDLGIKS	IGWAILSGER	IANAGVYLFE	(SEQ ID NO: 84)
	TAEELNSTGN	KLISKAAERG	RKRRIRRMMLD	RKARRGRHIR	YLLEREGLPT	
	DELEEVVHQ	SNRTLWDVRA	EAVERKLTKQ	ELAAVLFHLV	RHRGYFPNTK	
	KLPPDDES	ADDEQGKINR	ATSRLREELK	ASDCKTIGQF	LAQNDRQRN	
	REGDYSNLMA	RKLVFEEALQ	ILAFQRKQGH	ELSKDFEKT	LDVLMGQSG	
	RSPKLGNC	IPSELRAPSS	APSTEFKFL	QNLGNLQISN	AYREEWSIDA	
	PRRAQIIDAC	SQRSTSSYWQ	IRDFQIPDE	YRFNLVNYER	RDPDVLQEQ	
	LQQQERKTLA	NFRNWKQLEK	IIGTGHPQIT	LDEAARLITL	IKDDEKLSQ	
	LADLLPEASD	KAITQLCELD	FTTAAKISLE	AMYRILPHMN	QGMGFFDACQ	
	QESLPEIGVP	PAGDRVPPFD	EMYNPVVNRV	LSQSRKLINA	VIDEYGMPAK	
	IRVELARDLG	KGRELREIRK	LDQLDKSKQN	DQRAEDFRAE	FQQAPRGDQS	
	LRYRLWKEQN	CTCPYSGRMI	PVNSVLSED	QIDHILPISQ	SFDNSLSNKKV	
	LCFTEENAQK	SNRTPFEYLD	AADFQRLEAI	SGNWPEAKRN	KLLHKSFGKV	
	AEEWKSRA	DLTRYLTSALA	DHLRHLPDS	KIQTVNGRIT	GYLRKQWGLE	
	KDRDKHTHHA	VDAIVVACTT	PAIVQQVTLY	HQDIRRYKKL	GEKRPTWPPE	
	TFRQDVLQVE	EEIFITRQPK	KVSGGIQTKD	TLRKHRSPKD	RQRVALTKVK	
	LADLERLVEK	DASNRNLYEH	LKQCLEESGD	QPTKAFKAPF	YMPSGPEAKQ	
	RPILSKVTLL	REKPEPPKQL	TELSGGRRYD	SMAQGRLDIY	RYKPGGKRKD	
	EYRVVLQRM	DLMRGEENVH	VFQKGVYDQ	GPEIEQNYTF	LFSLYFDDL	
	EFQRSADSEV	IRGYRTFNI	ANGQLKISTY	LEGRQDFDFF	GANRLAHFAK	
	VQVNLGKVI	K				

Table 4- Cas9 orthologs						
Ruminococcus albus 8 WP_002846 926.1	MGNYYLGLDV QQCRRSRGLR QLRNLKGLSEK EEHDSIVLIS YLITRHLMLVK NERFRRTGY RGESVFDTSF KCFKYIKVVK RREKLKALNI KYQAKFNKEI NAIIDKYGY ECIKCDEVHA HIVPYSLILD MFKSKKCSKK NLRFDRSYES AELKKLTYLD EQSKKACIDN NIYEQFWKDD YRGITIGEEA KTIFEQADYK DDNNFSLND FNYPDDYYTH RFISDNKPCA LLKEKN	GIGSIGWAVI RLYRRKSHRK LTPEEIAACL NLMNEGGYCT EVDLILENQS LDSIGKCQFF ANDLINSALK PLFEKYGYDW GLDDGLINEL PDIDENAKPQ AAVNIETADE RHLIEKYKLW NTINNKAALVY KYQYLMLPDL SDEDDLKIRD HAADAIITAN LYKLFRMDRR KDKKSCEELY IRVKEIDGKL DVG DYLKKTN SSYYCIELYK VMYIFPGDYL KDKRVSITKK	NIEKKRIEDF LRLKNYLSII IHICNNRGYK PAEMICNCRE KDYGILDDKT KDQERGSRFT NGSMDKRELK DKLIENYTD TKLKLSTAN KLPPFKNEDD LNKTFEDRAI EAQEGKCLYS AEENQKKGQR NDQELLGGWR HYRVFPVKSR CRPEYVVLG TAEKLLSGHG RENVASLYKG IKLKRKSISE QHFFTTSSGK DSKGDNNLQG RIKSTSKKSG DIVIKLAVDL	NVRIFKSGEI GLTTSEKIDY DFYEVNVEDI FDEPNSVYRK IAKIKDIIFA VIADIYAFVN AIAKSYHIDI DNNVLNRIGI VSYKYMQGS CEFFKNPVVF DTKRNNDNQK GETITKEDML TPLMYMNEAQ SRNLNDTRYI FTSMFRRWWL EKLKLNKMYH RLTPIIPNLS DPKFASSLSM ITAESINSIY RVNKVTVIEK IAMS DIVHDR EQLKFEGYFI MGKVQGENNG	QEKNNRSRAS YYETADNNVI EDPDERNEYK FHNSAASKNH QRDFEIGPGK VLSQYTYTNN SDKNSDTSLT VLSQAQTPKR EAFCEGDLYG RSINETRKL ENDRIVKEII RDKDKLFEVD AADYRVRVNT CKYLVNLYRK NEKTWGRYDK QAGKRITPEY EEVDKRLWDK PVISLKPDKH TDDKILIDSL VPSRWLRKEI KTKKLYLKPD SVKNVNENSF KGISCGEPLS	(SEQ ID NO: 85)
Lactobacillus farciminis KCTC 3681 WP_010018 949.1	MTKKEQPYN KETRLNRSTR DPDRKRDKYN ALHNILKYRG DWNHISDILI NTIFKTSITK ITLNEILNGE YDDYINKPKY SENGVVPYQL LQSAEKNPFA GEPVLPKNSL LFKKYKKVTV FGSSF MEDNK SNMRYKGWGR MSNDKYDFKN EIVKFMGHAP QLREYLVPNK SDYQVDHILP WTYMLNNNMI DNMYKNQGT HAQDAYLASF SRKNGFIISP EIKTGQFYDQ IDNNKIKPVA IIYSKKVGLL YDKNILVEIL EELITLLHAN TRIHIE	GLDIGTSSVG RRYRRRNRI LFIDGPYTDK NFTYEHQKFN GRGNATQKSS DEEKLNFSGK SYFSMAKVNQ GTEKELYTSK NKIEMEKIID WMERKSNHGA IYQRYEVLNE KKLTKWLIAQ NYDQIEELIE LSKKILMDIT YIENHNLNKN KHIFIEVTRE KIQEELKKHK RTYIPDDSLE GLKKFKNLTR CIQARANLST LGTYRLRRFP LVNGTTQYDR TIYSPKNPKY IPIRLINDLK SLNSDREAN QELITKMKKF STSAHLIFNN	WAVTNDNYDL NWLNEIFSEE EYREFPTIF ISNLNNNLSK NILKDFTLTK DIESKLDDLD YENHAIDLCK KFLKVALPTN NQSQYYPFLK RPWNFDEIVD LNNIRITENL GYYKNPILIG WLTIFEDKQI TETNTPQLLQ EDQNISDLVN TKKSEITTSR NDLSSERIML NKALVLAKEN RVITDKDKLG AFRKALSGQD TNEMLLMNGE NTGEIINWVG KKLIAQKKDM DKKTLQNWLE RQQLILPPEH YPFYKGEREF IEKKA FGRKT	LNKKKNLWG LAKTDP SFLI HLRKELILNK ELIELNQQLI ETKLLKEVI SILDDDQFTV LRDMWHTTKN LAKEAE EKIS ENKEKLLSIL REKSSNKFIR KTNPIGSRLT LSQKDEFNST LNEKLHSSKY LSNYSILDLM DIHVSPALKR EKRIKRLQSK YFLQNGKS LY QRKADDLLLN FIHRQLVQTS DTYHFKHPEL YNKFYGVQKE FRDKILKIFN DPNIYGGFSG ENVKHKKSIQ SALLRLLQIP LIANIENFNQ HGLTLNNTDF	VRLFEEAQT RLQNSWVSKK DKADIRLIYL KYDISFPDDC NLILGNVAHL LDAANRIYST EEAVEQSRQA KGTYLKPRN SFRIPYYVGP RMTVTDSYLV VETKQRIYNE LT TYLDMKKI SYTPDQIKKI WATNNNFISI GITQSIKIVQ LLNKANDFKP SEESLNINKL SNVIDRNLER QMVKG VANIL VKNRNVNDFH LYSKKKKL PD YHQC NVTRKT DNKSSITIVK IIKNNVPIGQ DEDLDQILAF ATTSEKVNLS IYQSVTGLYE	(SEQ ID NO: 86)
Eubacterium dolichum DSM 3991	MMEVFMGR LV RRTRKGRRL ERLNGEELAT KSGKYVCEIQ	LGLDIGITSV KRRRVTRRED ALLHLCKHRG KERLRTNGHI	GFGIIDLDES MLHLLKQAGI SSVETIEDDE RGHENNFKTR	EIVDYG VRLF ISTSFHPLNN AKAKEAGETK AYVDEAFQIL	KEGTAAENET PYDVRVKGLN KVLSMNDQLL SHQDLSNELK	(SEQ ID NO: 87)

Table 4- Cas9 orthologs						
WP_004800 457.1	SAIITIISRK FPNEPRAPKL ITPKQLAKEV LEDHVFYDEI YHLSLFSKALR TAILSPVAKR ERQKFFEMRN KLLIDDPNAY KGRFTKGSIA KEFINRNLVD NLKKDRDEDY SSGDDSMYYD IDGKEKVVKK VVNYFFEEMS GPAITQMKYF YKFVTIRYKD LIGITKAEGE KPIHCYCKMR	RMYDGPGGP AYSAEFLNLL GVSLEQIRGF AEILTKTKDI LINEEMLKTE AQRETFKVVN KQVADIIGDD EVDHIIPIISI QYKAYCLKLK TSYASRVVLN GHHAIDALII DRYFAFIASL YKDIYDPKFT KSEKYFTKDK DGVLGNHIDI VRWSEKKKKY ALIYPDETHW LMPTISKKIV	LSPTPYGRYT NDLNNLSIEG RIDTKGSPLL EGRKKQISEL LNQMQSITLF RLREIYGEFD RKINAKLREK SLDDSIITNKV EKNIKTNGY TLTTYFKQNE ASMPKMRLLS KAIKVRKFSH ALAEDILNNA KGRIKISGMN SAHYQVRDKK VIDQQDYAMK NFNFFFHAGE RIDKYATDVV	YFGQKEPIDL EKLTSQKAM SELTGYKMIR SSDLNEESVH GLKQNNELSV SIVVEMAREK LVLYQEODGK LVTHRENQEK RKKVEQYLLN IPTKVFTVKG TIFSRYKIED KIDTKPNRSV YQEKYLMALH PLSLYRDEHG VVLQQISPYR KAEKKIDDTY TPEILKFTAT GNLYKVKKNT	IEKMRGKCSL ILKIVHEKGK EVLEKSNDHEH QLAGLTKFTA KGMKNIQADD NSEEQRKAIR TAYSLEPIDL GNLTPIISAFV ENDIYKYDIQ SLTNAFRRKI IYDESTGEVF ADETIYSTRV DPQTFDQIVK MLKKYSKKGD TDFYYSKENG EFQFSMHRDE NNDKSNKIEV LKFEFD	
Nitratifracto r salsuginis DSM 16511 ADV46720. 1	MKKILGVDLG SIRSTQKSMR NRWQLRAVDA LNDPEKESEK RSYRNHDDYE ACITDQEMPT GYEVTQEDRE DERIVKGKKE QRSKTPQEAL ALPLFLEGYD INNHAVKSLA MREREKALDK QLLDGSADIE DYRERIGMLS VAQVLKRYYP HAEDALILST ADYIDEAFRR SSRHEVPTLR IGNTNDESYR SPIEVTGKLL FRMRDVNNAH SKYIALFNPR RSYEVFGTLP	ITSFGYAILQ RLIEKRKKRI WKRPLSPQEL KADERRQVYN KMIRREDIEE IDESLFGKCT KVIEWVEKKI PRTFVPFFFL DRLRALMAGK EKEVQRILGF SWALGLIADL IIGKYKKEFP HIVPQSLGGL EKGLIDWKKR FPDPELRKNG LTRGWQNR LH FMSKGEESLF KNILEAFDSL AVEERATQIA RKMRFVYDKL ELQKERSGIL FPANPKAQPS EGSIEWFKEE	ETGKDLRYCL RCVAQTMERY FAIFAHMAKH ALRHLEELRK EIEKVLLRQA FYKDELAAPA AQGKNLKKIT ADIAKFKELF GIDTDDRELL DDREDYSRYP SWRYGPFDEI SIDKRLARKI STDYNTIVTL KNLLAQSLDE IGVRMIPGKV RMLRDNYGKS YRDMFDTIRS NVIKDRHKLT QILTRYQLMD NAMQIDRGLV CYLNEMLFIF KFTSDSKIKQ SGYGRVEDDP	DNSVVMRNNP GILDYSETMK RGYKSIATED KYGGETIAQT ELGALGLPEE YSYLYDLYRL HKDLRKILGL ASIQKHPDAL ELFKNKRSRG KSLRHLHLRE ILETTRDALP QLWERQKGLD KSVNAAGKGNR TSKTRSLGI EAELEKELWK ISYWVDKKPL TEEFMKRYDK AQNDKEIDEK ETDNMLGIH NKKGLIHYGC VGIGSATGII HH	YDEKSGESSQ INDPKNNPIK LIYELELELG IHRAVEAGDL QVSELIDELK YKKLADLNID APEQKIFGVE QIFRELAEL RELSHRYILE GNLFKEENP EKIRKEIDKA LYSGKVINLS LPGDWLAGNP IRATSYLEAL KSKSRETNFH YMPHIEGLTL SASSHKETVY EIRQKLWLHR FQQALKELIT ISKGPNEKLI LRSYLEKGQG KAHLDLDDGHV	(SEQ ID NO: 88)
Rhodospirill um rubrum ATCC 11170 WP_011388 212.1	MRPIEPWILG KDHTSRQAER LALRREAVSR PAKATAKATA RGQPVRMRYN LVLDQRPLRS ADEPRPLTDE AKQAARGTAG ALIGDPRGPT GIGYDEAVTL EDDGAAAEAY ELKAGADERK RRQNNLCPYT NKTKGDKTPW ARKSAEDEDQ	LDIGTDSL GW GAFRRARRQT PLAPDALWAA PAKEADDEAG QSDRDGVVAP KGAGPCAFLP EHAKALALLS NLTEAILAPL RVTEDETAEA ALGLHHSHRP YGRIGNISVH RMIAEQAERE STPIGHADLL EAFHDKPGWI GFLPRQLTDT	AVFSCEEKGP RTWPWRRDRL LLHLAHRGF FWEGAEAAALR TRALIAEELA GEDRALRALP TARFVEWPAL IPGWSGWDL VADAIQIVLP RQERLARLPY IALNETRKIV RENAEIDVEL GDAYDIDHVI AQRDDFLARL GYIARVALRY	PTAKELLGGG IALFQAAGLT RSNRIDKREK QRMAASGAPT EIVARQSSAY TVQDFIIRQT RRALGLKRGV RKDRVFSDLW TGRASLSAKA YAAALPDVGL NALLHRHGPI RKSDRWMANA PLARGGRDSL DPQTAKALAW LSIVTNEPNA	VRLFDSGRDA PPAAETRQIA AAAKALAKAK VGALLADDLD PGLDWPVAVTR LANLRLPSTS KFTAETERNG AARQDRSALL ARAIQAMAP DGDVPGPPPA LRLVMVETTR RERRQVRVLA DNMVLCSQSDA RFADDAGERV VVATNGRLTG	(SEQ ID NO: 89)

Table 4- Cas9 orthologs						
	LLRLAWDITP VQARLAALGR VTTRGLINQI SIRPAHRDGG KIAEIIDGFA GMTANTVIAG RPRWTHRVL VKKMAVDGRL RKEKIRTTSC	GPAPRDLLPT SRVADAGLAD NQASGAGRIL SLHAATVFGV SPRMAKRFGA RSDGDGEDAG RFDRTQPAPP ALWPARLATG TALGRLRLSK	PRDALRDDTA ALGLTLASLG DLRKWPRTNF RNRPDARVLV LLARYQAAHP LITPFRANPK PPPENARLVM KATALYAQLS KAT	ARRFLDGLTP GGGKNRADHR EPPYPTFRAE QRKPVEKLFL EVPPALAALA AAVRTMGNAV RLRRGDLVYW CPNINLNGDQ	PPLAKAVEGA HHFIDAAMIA VMKQWDHIHP DANAKPLPAD VARDPAFGPR YEVWEIQVKG PLESGDRLFL GYCVQSAEGI	
Finegoldia magna ATCC 29328 WP_012290 141.1	MKSEKKYYIG NTRIFRSGRR KKVSGKYSLF NQMMKRRGHF EILADVNEKR IISDEDILEK YDFLILQSI VFKEDNGNGY DEDVKYLLGK FNNKKEEFTK RPWNFEKIVD LNNLRINGKP DNSEKNLEIA KYIEETYPDL INFLRNSSDN SVKKMIWQVI LYKAIKKDER IDLKLFDPST PSSIRNNPKI VETRQSTKAI HAHDAFLNIV KGKVIWTPFK GKIYLPKLD HLLSKFYEDK KSNDSGSI DKIYQQFKAG KTSNNAVDFT	LDVGTNSVGW RNDRKGMRQ NDKNFSDKQY LIDGQISHVT TDDKNNLKE IKEDNKEDFV KKGSTLSDAQ VSYIGYYLNK IEQENFLLKQ IEKIRKTFTF LHKSEEEFIK LDTDVKKLKI SNMKSIDFN SSSQIQKIIN LMQIIGSQNY RVTEEITKVM DSQYEKLLTG HYDKDHIIIP YNYWKYLMEK KELFEKFKYQ AGDVWNREFT GRKLIVDTLN DRLQDVSKY NTVLDYAINV PNEQMYWRVD KYKNRRTTDT VIGLGTECGK	AVTDEFYNIL ILREIFEDEI FEKFPTIFHL DDKPLKEQLI IKGQDFNKQE LTGDSYEENL VERYDEHKKD NKKITAKKKI ISSINSVIPH RIPYYVGPLN RMLNQCTYLP EELFKKKTKV NILEDKFDVE LKYKDWGRSL SFNEYIDKLR GYDPDKIFIE LNKLDSDSLR SMKKDDSIIN EFISKEKYNR SKIIPVKASL SNPINYVKEN KPSVLISNES GYKAINGAFF LQLQDPKIII EISNLKKIEN IIEKYEIIDL PRITNLPDNT	RAK GKDLWGV KKVDKDFYDR RKYLMEEHGK LLINDLLKIE GNILNSIFES QYFEEVLQEN LEILKKVIKK SNIEFTKYVK QIHLFELDKI DYHKNNNGNA EETVLPKSSI TLKSIRDYMV MVEDLIEKIT RKLLDGIGKT KKYIPQEISY MAKSEEEKKT SRKLYLYYTQ NLVLVNKNAN LIRNTPLTNE ASDLRKDMNT REGDKVKYSL HVKKGELFNA FLVEHTKSKK DKINYRTEII KYKKDAILTE DTLDNKQLYQ YLVYKSITGI	RLF EKADTAA LDESKFWAED VDIRYYFLAI LEEELMDSIF IVTGKAKIKN ITLFTNLKST YDEDGKLFQK GILEKQCDC LENLAKNYP WIFRNKGK LYSEYMLNE RNNFADKEDF IHTGNKKLLK KKETEKTDTV EVVENLYVSP TISRKNKLLD MGRDMYTGEK QTTKGNLYPV ELGGFINRQL LKSREVNLDH SKDFTRP TIAGKKDYKK RIRSIELFPL IDNFSYLIST EDRKIMESYI LLVAFISLSY YEKRIRIK	(SEQ ID NO: 90)
Eubacteriu m rectale ATCC 33656 WP_012742 555.1	MNYTEKEKLF DNQLRRDMRG AITEEITLDE ELETHYPCEI QRVFEIQKKY KLDANGNYIT TINGRKLEEN GKEIFHKFEV EAFQKSWIDL PKEQMTLLTE ILNAVLKKYK AVTYGIKLS LFEIDHII EQYKATVMNL YASRLVLNTI HAVDAMLIGY YLYGKKWANI YKINKLDIRT DYSDAANPFV KYGFEKGSKK	MKYILALDIG AKRNNRRLKT LYLILYSYLK QQRERLNTIGK HPELTDEFCD EDNIFEKLI EKHEIVERIK YNKMRKALLE SDDVKQCLIN MGVTKGTEEE ALDTIVIEEMP SDFSSQKQLS SISFDDARSN SKKKEYAISR QNFFMANEAD SELGYEAYHK RNEVVKAEKN KEGIKVF QYEKETGDII VILES VILESLVPYR	IASVGWAILD RINDFIKLWE HRGISYLEDA YRGQSQIINE GYMLIFNRKR KCSVPDELR SSNTINMRKI IGIDISNYSR MRKTNGALFN FAGLKYIPVD RDRNSEEQKK LKLKLWNEQD KVLVYRSENQ KKIQNLLYSE TKVKVIKGSY LQGEFIDFET VKYWHYVMRK AFSKKDSRE RKYSKKHNGP MDVYYKEENH	KESETVIEAG NNNLSIPQFK LDDTVSGSSA NGEVLDSLNV KYYEGPGNEK AAAASYTAQE ISDCMGENID EELDEIGYIM KWQSFSLKIM VVEDIFNPV RINDSQKLNE GICLYSGKTI KKGNQTPYYY DITKMDVLKG THQMRCNLKL GEILRKDMWD SNRGLCNQTI RLLVYLNDRR RIDKLKYKDG SYLLVGVKQS	SNIFPEASAA STEIVGLKVR YANGLKLN FTIGAYRKEI SRTDYGRFTT YNVLNDLNNL DFAGARIDKS TINTDKEAMM NELIPEMYAQ VRRSVRISFK KEMEYIEKKL DPNDIINNPF LTHSHSEWSF FINRNINDTS DKNRDESYSH ENMSDEVYAD RGTREYDGKQ TFDDLCKIYE EVGACIDISH DIKFEKGRNV	(SEQ ID NO: 91)

Table 4- Cas9 orthologs						
	IDEEAYARIL YTERLVSRTM NKYSCSEEF	VNEKMIQPGQ PKQRNYIETK TSFC	SRADLENLGF PIDKAKFEKQ NLVGLGKTKF	KFKLSFYKND NLVGLGKTKF IKKYRYDILG	IIIEYEKDGKI IKKYRYDILG IKKYRYDILG	
Corynebacte rium diphtheriae C7 (beta) AEX66236.1 WP_014318431.1	MKYHVIGIDVG VTRLASSGIA WKVRAELAAS PSDAFKAIRES LQQSDHAREI QPGKNRALK NLAPKKEPEW VNSRIAPLVD DDDVFHAKLDS PSWTPPAPRI GFTVTEKRARE QRQNCQCA GNTPFALWAK FQRATMDDEE ARRASGISGK LKQSQAHREQ VVVMSNVRLR TREPDPFDPK ELGSSFHAR SMRQAEKKLR TIRRWVVDGF KLFSEGNVT	TFSVGLAAIE RRTRRLYRRK YIADEKERGE EIKRASGQPV QEICRMQEIG SDAFQRYRIA VTIAEILGID WWKTASALEQ LHLPVGRRA GEPVGNPAVD MDGDMRRRA GSPITFSNSE NTSIEGVSVK DARSMESVAW LEFLDGVGKS APQWREFTGK LGNGSAHEET GLPANPERHI VYKITSGKKP DALATGNAEY FGDTRLRLRP VRRDSLGRVR	VDDAGMPIKT RRRLQQLDKF KLSVALRHIA PETATVGQMV QELYRKIIDV ALIGNLRVRV RGQLIGTATM HAMVKALSNA SEDTLVRLTR RVLKTVSRWL RNAKLFQEMQ MDHIVPRAGQ EAVERTRHWW MANELRSRVA RLDRRHHAID DAEHRAAWRV IGKLSKVKL RVNGTHVYAG AFAMLRVYTI LGWLVVDEL LQMSKEGIKK LESTAHLPT	LSLVSHIHDS IQRQGWVPIE RHRGWRNPYA TLCELGTLLK VFAAESPKGS DGEKRILSVE TDDGERAGAR EVDDFDSPEG RMLADGVLDY ESATKTWGAP EKLNVQGKPS GSTNTRENLV TDTGMRSTDF QHFASHGTTV AAVIAFTSDY WCQKMEKLSA SQLSVSDIDK DNIGLFPVSA DLLPYRNQDL VVDTSKIATD ESAPELSKI WKVQ	GLDPDKIKSA LEDYSDPLYP KVSSLYLPDE RGEVGVSAR ASSRVGKDPL EKNLVFDHLV PPTHDTNRSI AKVQAFADL TARLQEFGIE ERVIEHVRE RADLWRYQSV AVCHRCNQSK KKFTKAVVER RVYRGSITAE VAETLAVRSN LLTEDLRDDR ASSEALWCAL GSIALRGGYA FSVELKQPMT QVKAVEAELG DRPGWLPVAVN	(SEQ ID NO: 92)
Roseburia inulinivoran s DSM 16841 WP_007889305.1	MNAEHGKEGL VRIFDVAENP EMDSFMERY RKAETKEKEG KGYVLTNRN VEIMTSQRSF YTAELFVALQ RKKNLNDPSL LKDRTEEMPV LDLSPAKYQR LLKGEEINAI MSKNFQDRTN QGGYCLYSGK QKGNRTPY FKERNLNDTK KRWGLMQKDR FPDEETGEIL FLLTHADVQR SAHDATIRSA ALVRQLLLHG GGTGIAANGE EWRVMDANF DIATASIAGF FH	LIMEENFYQY KNGDALAAPR KGNLPDVYQL GAVLKATTEN PDDYKHTILR DMGPGQLPDG KINHTKLIDE KFNSLNYS GEKADLFDRI VSLKAMRKMQ VNDITNPVVK LEKEMKKRQ KIPLEELFDG FGADEKRWED YITRVVYNMI STDRHHAMDA NRDNFTREQW ELDYPGWYMG RDYETRGVVI NDGKKAF MVRIDVFREN VFSLYSRDLI ANDSNFKFRG	IGLDIGITSV RDARTTRRL RYEGLDRKLK QKIMQEKGYR SMLVEEVHAI KPSPYAMEGF FGTGRFFSEE KEGETEEERV GEILTAYKND PYLEDGLIYD RSVSQTIKVI ENERAKQII GYDIDHILPY YEASVRLV RQNLLELEPFN VVIACCTDGM DEKFGVKVPL EESPIEEGR TKVPLTDLKL FHKPKADGTE GKYFVVPVYT HVKSCKDIKT LGIQSLEIFE	GWAVLQNN RRRRHRLERI DEELAQVLIH TVGEMLYLDE FAAQRAHGNQ GDRVGKCTFE ERKTIIGLLL LDTEKAKFAS DSRSSRLKEL KACEAAGYDF NAIIQKYGSP ELGKQNP DYKKQKLLK HPEKKKQVWA IHKISRYMQG PWNSFRDELD YINYIRPLFV NKDNEIEGYY GPVVRKV ADVVRKVLPN NLVNGGLLQ KCQVDILGNI	DEPVRTDLG KFLQLQENGLI IAKHRGFRST AFHTECLWNE KATEGLEEAY KDEYRAPKAT SSKELKYGTI MFWTYEYSKC GLSGEEIDGL RALNDGNKKH QAVNIELARE DILKYRLWND KVLVTAQENR KNFTTEERKE VNGAVTSYLR RELAYS IRLLNEDPKN SRMPNHKVTG DKDSRLLYQ KKQTS RAATHKPY KEIFAYYTGA SVVRHENRQE	(SEQ ID NO: 93)
Alicyclophil us denitrificans K601	MRSRLRYRLAL SSLAVTRRAA NPYALRAKGL AIGQLRQOMA KSYDLYIDRA PVKPGRCTLL QRDAVVTALE	DLGSTSLGWA RAMRRRRDRL QEALLPGEFA EQGSRTVGEY MIEQEFDALW PEEERAPLAL TKAKLSFEQI	LFRLDACNRP LKRKTRMQAK RALFHINQRR LWTRLQGGQ AAQAAFNPTL PSTQRFRIHQ RKLLKLSGSV	TAVIKAGVRI LVEHGFFPAD GFKSNRKT VRARYREKPY FHEAARADLK EVNHLRL QFNLEDAKRT	FSDGRNPKDG AGKRKALEQL KDNDSGVLKK TTEEGKKRID DTLLHQRP NLREVALTLA ELKGNATSAA	(SEQ ID NO: 94)

Table 4- Cas9 orthologs						
WP_013517 127.1	LARKELFGAA QAIVDVSLPE FEYDASEVED RYGKIANPTV EAQRRQADNQ DRRCPSGVQ KRNRTPWDAR KDFLARALND GLDGEKNRND MPMPWPTYRD RKADGSAGRE TVNDKGKWEQ SVRLEVDGAE KYAGSLQKAK	WSGFDEALQD GYGNLSRKAL LVHPETGEIR HIGLNQVRMV RRNARIRRSI ISAAMLLSDE AEFEAQGWSY TRYLSRVAAE HRHHAVDACV HVERAVRHIW ISNLIRIHEA EVISTFRAYG RTMRIVKISG TRRVTISPIGEV	EIVWQLVTEE ARIVPALRAA SVFKQLPYYG VNALIRRYGR AEVLGIGEER VEVEHILPFS EDILQRAERM YLRLVCPGTR IGVTDQGLMQ VSHRPDHGFE TQPLRHGVSA VVRAGGMGRL SNGQIFMAPI HEANVDARNT	GEGALIAWLQ VITYDKAVQA KALQRHVAFG PTEVVIELAR VRGSDIQKWI KTLDDSLNNR PLRKRYRFAP VIPGQLTALL RFATASAQAR GAMMEETSYG DGQPLAYKGY RNPHEGQNGR DKQDAFTYTS	THTGVDEARA AGFDHHSQLG SGKPEDPDEK DLKQSRQKQV CWEELSFDAA TVAMRQANRI DGYERWLGGD RGKFGINDVL GDGLTRLVDG IRKDGSIKQR VGGSNYCIEI KLIMRLVIGD DKQDAFTYTS	
Sphaerochaeta globosa str. Buddy WP_013607 849.1	MSKKVSRRYE SDLVFVSSRI LSYSEQDVPD NHRGSSSVRT TNGLIGYRNS DRIVSAILFE IKMPMQEGAA SIIVLQKKEK KRLSKYLMKH TYTEALERGM KEKRDSEGGF PQEITVELAR IERFRLLEDQ VVAHRQCNDI EEYAKYLQSK TSILRKAWNLI YCSRSVLQMI ILEFMDLHAF DIGVKIGDYE KESLVQAAAV NKRTKTFVVK NPSYKPAYMK KPGRTFVII QLSSAGLVRY	EQAQEICQRL FIPSTGAAER PARLRFEDAV FLDEEKSSDD ESVKSQGVVP NEKIVPEAGC KRYQSASFSD AIQKIAGFRF LLLTENEVD ETGEFQELSV KPNTNSDEEK ELKVGAEKRE AFVCPYCLEH KGKRTPYAAF GFVSRFESDN QGIDDLLGSR NTMSEQGKRA VSMKTDNDAN EVASAIRGRI LEESNRKLI EPSNEVKGFA QGYSLYVRLY TFTEMGSGYQ VSPLLVDKIE	GSRPYSIGLD RQKRQQRNSL VRANPYELRL KKLEEQQAMT PTRDIISNEI CPYFPDEKKL EQRHILFHIA RRLEEKSFWK ALKTVSLIGD WEQQSLLPYY YGRIANPVVH DIIKQQTQKE ISVADIAAGR SNTSAWGPIM SYIAKAAKEY HWSKDADTSP VEIEAMIPIP GALLKDTVYS TDKQPKWYPM SGKKPIQLSE FDTGSNLCLD QGDVCELRAS IYFSNLAKSK KDEVALCGE	LGVGSIGVAV RHRANRLKFL KGLNEQLTSL EQLAKEKGIS DVLLOTQKQF PRCHFLNEER RSGTDITPKL RLSEEQKDDF YGPIGKTATQ GQILTGSTQA QTLNELRKLM KEAVLAYSKY ADVDFHIFPRD HYLDETGMW LRCLFNPNNV TMRKNRDDNR GYASEPNLSF ILGADTQGED EMKDKEQLQ KTISKKALEL FYHDAQGKLC DLTEAESNLA KGQDTSFTLT	AAYDPIKKQP WKLLAERNLM ELGYALYHIA TFIEVLTAFN YQEILSDEYC RLWEAINNAR VQKEFPALKT FSAWTNTPDD LLMKHLEDGL LMGKYWHSFAF NELITILGAK CEPNNLDKRY DTADNSYGNK RKRKFETNE TAVGSLKGME HHGLDAIVAL EAQRELFRKK LVFVVKKKIK SKNEAALQKY VGGYYYLISN GEIIRKIQAM KTTHVRLPNA TIKNYDVRKV	(SEQ ID NO: 95)
Fusobacterium nucleatum subsp. vincentii ATCC 49256 WP_005888 649.1	MKKQKFSDYY AAERRVQRNS EDKNSKEKFT ALHSIFKSRG EKLEKICDS TDEYKKEEVE LNNILSDSNY NNYPAYIGLN FNEILNKIEL EENGVSTKDK NFEQKVDIEK VQVNDEFLEK IKDSFNSNYV KKIKNEYGDI SVMEALRRTN SLKRAILQTL LKKLYDSCGN YTGREIDLDR	LGFDIGTNSV RRRLKRRKWR LFNDDNYKDY HFLFEGQNLK GKGLKDKEKE KEKISFREQI ISEAKVKLYE KEKDKKEVVE KTILPKQRIS LLKTFKFRIP SAEEFIKMT ENKRKIIDEL SYIKFKDIFG LNKDEIKKIN YNLMELLSSK KIYEEIKKIT DIANFSIDIK LLQNNDTYDI	GWCVTDLDDYN LNLLEEIFSD DFYKQYPTIF EIKNFETLYN FKGIFNSDKQ YEDDKPIIYYS EHKKDLKNLK KSRLKIDDLI DNGTLPIQIH YYVGPLNSYH NKCTYLNSED FKENKKVSEK EKLNLDIYKE SFKFNTWGRL FTLQESIDNE GRVPKKVFIE EMKNSLSSYD DHIYPRSKVI	VLRFNKKDMW EIMKIDSNFF HLRDELIKNP NLISFLEDNG LVAIFKLSVG ILGEKIELLD YIIRKYNKEN KVIKGYLPKP EVELEKILEN KDKGGNSWIV VIPKDSFLYS KFKEYLLVNQ ISEKSILWKC SEKLLTGIEF NKEMNEVSYS MARGGDESMK NNSLRQKKLY KDDSFNLDVL	GSRLFDEAKT RRLKESSLWL EKKDIRLIYL INKSIDKDNI SSVSLNDLFD IAKSFYDFMV YDKLFDKNE ERIEEKDKTI QSKYYDFLNY RKEEGKILPW EYIILNELNK IANRTVELKG LYGDDKIFIE INLETGECYS DLIEESYVSP NKKIPARQEQ LYYLQFGKCM VLKNENAEKS	(SEQ ID NO: 96)

Table 4- Cas9 orthologs						
	NEYPVKKEIQ VNVRQTKEV HAKDAYLNIV ENSLEIVKKN LYDGKDNKLN KNEKNLIKYL LKNKKQLYLE TIDAVKERYN LKLWEKSLIL SSLGNNKELL	EKMKSFWRFL GKILQQIEPE AGNVYNTKFT MEKNTVNITR EKYGYTSLK VSQKKLLNPK KKYEQILKNA IEFNEMYDKF REFLKIFNKN EESVTGLFVK	KEKNFISDEK IKIVYSKAEI EKPYRYLQEI FIKEEKGELE AAFYIYVEHE IIKKIYKEQT LKFVEDNQGE LEKLSSKDYK TYGKYEIKDS KIKL	YKRLTGKDDF ASSFREMDF KENYDVKKIY NLNPIKKGET KKNKVKTFE LIIDSYPYTF TEENYKFIYL NYINNKLVTN QTKEKLFSPF EDTGRIRLGQ	ELRGFMARQL IKVRELNDTH NYDIKNAWDK SNEIISIKPK RITRIDSTLI TGVDNKKVE KKRNNNEKNE FLNSKEKFKK EDTGRIRLGQ	
Pasteurella multocida subsp. multocida str. Pm70 WP_010907 033.1	MQTTNLSYIL TGESLALSRR QAWELRVAGL LSGVAQNHQL LAELNLLFAQ THEKNEFKAA PYEKSKLTYA ALENQGLKDT NALLVSLNFD KTSQLLPAIP LGKSFKERRE YEQQHGKCLY SENQNKGNQT DNKFIDRNLN RWGLIKAREN YEMVDQESGE NHQFVQPLFV LLENMVNKER EQVQKSGVLV NKAIVAHKNE ATGNISLKEH QRQPVR	GLDLGIASVG LARSTRRLIR ERRLSAIEWG LQSDDYRTPA QHGFNPHCK KHTYSAERFV QVRKLLGLSE WQDLAKKPD KFIELSLKSL AQEIRNPVVL IQKQQEDNRT SGKEINIHRL PYEWLQKIN DTRYIARFLS NNRHHALDAI IISPHFPEPW SRAPTRKMSG EPALYAGLKA RENNGVADNA DEWEEMDEGA DGEISGKGDG	WAVVEINENE RRAHRLLLAK AVLLHLIKHR ELALKKFAKE EHIOQYMTL WLTKLNNLRI QAIFKHLRYS LDEIGTAFSL RKILPLMEQG RTLSQARKVI KRESAVQKFK NEKGYVEIDH SERWKNFVAL NYIQENLLLV VVACATPSMQ AYFRQEVNIR QGHMETIKSA RLAEFNQDPA SIVRTDVFIK KFKFSLFPND VYRVGVKLAL	DPIGLIDVGV RFLKREGILS GYLSKRKNES EGHIRNORGA LMWQKPALSG LEDGAERALN KENAESATFM YKTDEDIQQY KRYDQACREI NAIRQYQSGP ELFSDFSSEP ALPFSRTWDD VLGSQCSAAK GKNKKNVFTP QKITRFIRFK VFDNHPDTVL KRLAEGISVL KAFATPFYKQ NNKFFLVPIY LVELKTKEY SFEKYQVDEL	RIFERAEPVK TIDLEKGLPN QTNNKELGAL YTHTFNRLDL EAILKMLGKC EEERQLLINH ELKAWHAIRK LTNKVPNSVI YGHYGEANQ ARVHIETGRE KSKDILKFR SFNNKVLVLA KQRLLTQVID NGQITALLRS EVHPYKIENR KEMPLPDRQA RIPLTQLKPN GGQVQKAI TWQVAKGILP FFGYIIGLDR GKNRQICRPQ	(SEQ ID NO: 97)
Alcanivorax pacificus W11-5 WP_008738 269.1	MRYRVGLDLG PKKAARRLAR TLRAMAVNER RLEEEMVALA LFRQYLGLPA VKEEIRNAIF ADLRWGMGRR GKGLNMDRAA ADLGSPEQLD MGFEGGRSSY ATPAQGGQRS EMAREMKGGV LWIEQGHQCP HRECNDKGN EGEALTDESI RRRWGLDVI RTDRRPDKRI VKVELPMPIL ERSGEKVDWL FEKRVSGGMT HYKMLLNDGF KGDTVKNIKT GGRNIHLRL	TASVGAAVFS QQRQIDRRA IELGQLRAVL SVQNKDSVT IKGRPDCLPN FQRPLKSPAD AEMLNDHQKA LGGRDDLSGN TDDWSCRFGM SIKALKALTE KLEPPPLTGN TRNDIEKQN YCESNISLEQ RTPYQAFGHD DEFADRQLHE PQVRFESGMP DHRHHLVDI TIRDIALEAV VSRKSLTDLA PQQALREPIE AYMEVPCKEG GRVYTIKQIL CAE	MDEQGNPMEL SRLRRIAIVS LRMGKKRGY GEYLAARVEH MYALRHQIEH KVGRCSLQTN VIRELLNQK TTLAAWRKL KNGRPRNFSD WMIAPHWRET EVVDVALRQV KRFASERKKA ALSGAYTNFE DRRWRIEQR SSWLAKVTQ VVDEEGAEIT VTALTSRSLY RSVRISHKPD PEKKSIDVDK FQGNILRKVR ILYGVNPLVR GDGGGKLILT	IWHYERLFSE RRLGIAPGRN GTFKAVRKVG GLPSKLKVAA EFERIWATQS LPRAPRAQIA ELSFRIYKE LEDRWQELDE EFVAFMNELR PETHRVDEEA RHTINMMIDD AQSIENGKT HILPRTLTQI ANALPKKSSR WLSSLGSDVY PEEFKFKRLQ QQYAKAWKVA RYPDGRFFEA VRANISRIVG CFYSKADDCV PSEAVGIKRA PVTETKPADL	PLVPDMGQLK DSGVHGNVDP EAGEVASGAS NNEYAPEYA QFHDVMDH AQNFRIEKM LERAGCPGPE VTQIQVINFL MTDGFDRLSK AIRECYPESL LGSVPAQIVV PTPARILRYQ GRKRSELVLA KTRLLLLKDF VSRGSLTAE WEGHRVTREM DEKQRHGRVD TAYGIAQRLD EAIRLHISNI RIEHSSRRGH PESGDFIRFY LSAKWGRLKV	(SEQ ID NO: 98)

Table 4- Cas9 orthologs						
Mycoplasma mobile 163K AAT27519.1	MYFYKNKENK	LNKKVVLGLD	LGIASVWGCL	TDISQKEDNK	FPIILHGVRL	(SEQ ID NO: 99)
	FETVDDSDDK	LLNETRRKKR	GQRRNRRLF	TRKRDFIKYL	IDNNIELEF	
	DKNPKILVRN	FIEKYINPFS	KNLELKYKSV	TNLPIGFHNH	RKAAINEKYK	
	LDKSELIVLL	YFYLSLRGAF	FDNPEDTKSK	EMNKNEIEIF	DKNESIKNAE	
	FPIDKIEFY	KISGKIRSTI	NLKFQHDYL	KEIKQVFQKQ	NIDFMNIEKF	
	AMEEKSFFSR	IRNYSEGPGN	EKSFSKYGLY	ANENGNEPELI	INEKGQKIYT	
	KIFKTLWESK	IGKCSYDKKL	YRAPKNSFSA	KVFDITNKLT	DWKHKNEYIS	
	ERLKRKILLS	RFLNKDSKSA	VEKILKEENI	KFENLSEIAY	NKDDNKINLP	
	IINAYHSLTT	IFKKHLINFE	NYLISNENDL	SKLMSFYKQQ	SEKLFVFNK	
	GSYEINQNNN	VLHIFDAISN	ILNKFSTIQD	RIRILEGYFE	FSNLKKDVKS	
	SEIYSEIAKL	REFSGTSSLS	FGAYYKFIPN	LISEGSKNYS	TISYEEKALQ	
	NQKNNFSSHN	LFEKTWVEDL	IASPTVKRSL	RQTMNLLKEI	FKYSEKNLE	
	IEKIVVEVTR	SSNNKHERKK	IEGINKYRKE	KYEELKKVYD	LPNENTTLLK	
	KLWLLRQQQ	YDAYSLRKIE	ANDVINKPWN	YDIDHIVPRS	ISFDDSFNL	
	VIVNKLDAK	KSNDSLAKQF	IEKIYGIEKL	KEAKENWGNW	YLRNANGKAF	
	NDKGKFIKLY	TIDNLDEFDN	SDFINRNLSD	TSYITNALVN	HLTFSSNSKY	
	YSVSVVNGKQ	TSNLRNQIAF	VGIKNNKETE	REWKRPGEFK	SINSNDFLIR	
	EEGKNDVKDD	VLIKDRSFNG	HHAEDAYFIT	IISQYFRSFK	RIERLVNRYR	
	KETRELDLLE	KNNIKFKEKA	SFDNLLINA	LDELNEKLNQ	MRFSRMVITK	
	KNTQLFNETL	YSGKYDKGKN	TIKKVEKLNH	LDNRDQIKK	IEEFFDEDKL	
	KENELTKLHI	FNHDKNLYET	LKIIWNEVKI	EIKNKNLNEK	NYFKYFVNKK	
	LQEGKISFNE	WVPILDNDFK	IIRKIRYIKF	SSEKETDEI	IFSQSNFLKI	
	DQRQNFSEHN	TLYWVQIWWY	KNQKDQYCFI	SIDARNSKFE	KDEIKINYEK	
	LKTQKEKLQI	INEEPILKIN	KGDLFENEK	ELFYIVGRDE	KPQKLEIKYI	
	LGGKIKDQKQ	IQKPVKKYFP	NWKKVNLTYM	GEIFKK		
gamma proteobacterium HTCC5015 WP_008284.239.1	MTKNYISPIA	IDLGAKFTGV	ALYQYLEGAD	CTQEVAKGLL	VDDRGNVTWS	(SEQ ID NO: 100)
	QEGRRGKRHQ	VRGYKRRKMA	KRLWLILDS	EYGIKREEVT	EPLLKFINGL	
	LNRRGTYTIS	EEVDEESMN	SPLPFSEMP	DYFNSSAPL	EQLAKLLSDK	
	NKLVRFAEG	KIPSNKNEFK	KLLDTALDGK	YKDEKKELSE	AWGNILIASE	
	NVLKSTVDGH	KSRSEYLANI	KEDIKSNEEL	EKQISSKEID	GFYNLVGHLS	
	NFQLRLLRKY	FNDPNMSGVS	YWDEKRLEKY	FYQWVQGWHT	KGGTDEAEKK	
	NIILKTGAP	LLKTLKSLA	DLTIPPYEDQ	NNRPPKQCS	VLLSDEKLT	
	HYPKWKEWVG	QLVKQNDNAY	LNENVTLANA	LHRIVERSRS	IDPYQLRLLI	
	SITDAEKRND	LAGYKRLKLS	LGSEVDEFL	LVKNIVDETK	EAREGLWFET	
	ENKLFKCGK	TPPRKEKLKS	TLLSAVLGKN	LSDDQSSFI	EEFWKSGTPK	
	IERNVRGWC	RLASQVQKTY	GVYLKEYGLQ	QLHKEAGKK	LDDKPLALLY	
	KNSGLIASKI	GEALNIEPDE	VSRFASPHSL	AQIFNIEGD	VAGFNKTCRA	
	CTYENIWRMQ	EEKVESLITN	QLLSEIHGER	KVPLKSAMCT	RLSADSTRPF	
	DGQMASIIEH	IARKIAQHKI	AQINDVPKEF	SIDIPIIIES	NQFSFTAEL	
	EIKRGRGSAK	AKKAKELGEK	SKAGWVSKTE	RIKTSSEGC	PYTGAPLGGS	
	GEIDHIIPRS	LTGRTKKTVE	NSEANLIYCS	SKGNHDKGNR	VYVIEQLNDK	
	YLKKQFSTSD	VNLIKKKIKT	TIQRFTEGGE	KLRSFSSELSR	EDQKAFRHAL	
	FVPELKSEVT	SLLAVKNITR	VNGTQAWLAK	KIASLLAEHL	DKQGRDYTLS	
	AHQIDPWSVS	KQRKMLASAE	PIWAKKDPQP	AASHVVDVAV	TFLEALEQPH	
	TASRLKTISS	TSFEKTGWRS	ALIPDLIKVD	ALDRRPKYRR	YNIGSTSLFK	
	DGIYAERFLP	ILIDENGLMA	GYDIDNSLKA	KGADVVFESL	SPFLLFKGEE	
	VGAQSLSDWQ	ERIDGRYLYM	SIDKVKAIFY	LQEKVGEKDI	AAELLNSIHF	
	TQRKTELRAK	FSDDSGKKMK	TLDAIRKSLK	LTVTVNEIGK	RKEKCGFSGT	
	IGIPAKSAWE	NLLDEPLET	YWGTMPPQE	IWEKVYRKHF	PRNIPNQAHR	
	KVRKDFSLPV	VDSVSGGFRV	KRKTPNGYNY	QLLAIDGYSA	VGFKKEGDNV	
	DFKSPALVPQ	IAESKSVTP	SSELVHLDKN	EIVYFDEWRK	IDISDSDLKQ	
	FVSSLELAPG	SQNRFYIRFT	VDEDQFERHF	KSALRVNGIQ	DLDTVNKTFD	
	WNREIPSLLI	PPRSNLFLE	TGQKITFEYI	ANGANADEVK	AYSLRRA	

Table 4- Cas9 orthologs						
Planococcus antarcticus DSM 14505 ANU10858.1	MKNYITIGLDI	GVASVGWVCI	DENYKILNYN	NRHAFGVHEF	ESAESAAGR	(SEQ ID NO: 101)
	LKRGMRRRYN	RRKKRLQLLQ	SLFDSYITDS	GFFSKTDSQH	FWKNNNEFEN	
	RSLTEVLSSL	RISSRKYPTI	YHLRSDLIES	NKKMDLRLVY	LALHNLVKYR	
	GHFLQEGNWS	EAASAEGMDD	QLLELVTRYA	ELENLSPLDL	SESQWKAET	
	LLLNRNLTKT	DQSKELTAMF	GKEYEPFCKL	VAGLGVSLHQ	LFPSSSEQALA	
	YKETKTKVQL	SNENVEEVME	LLLEESALL	EAVQPFYQQV	VLYELLKGET	
	YVAKAKVSAF	KQYQKDMASL	KNLLDKTFGE	KVYRSYFISD	KNSQREYQKS	
	HKVEVLCKLD	QFNKEAKFAE	TFYKDLKKLL	EDKSKTSIGT	TEKDEMLRII	
	KAIDSNQFLQ	KQKGIQNAAI	PHQNSLYEAE	KILRNQQAHY	PFITTEWIEK	
	VKQILAFRIP	YYIGPLVKDT	TQSPFSWVER	KGDAPITPWN	FDEQIDKAAS	
	AEAFISMRK	TCTYLKGQEV	LPKSSLTYER	FEVLNELNGI	QLRRTGAESD	
	FRHRLSYEMK	CWIIDNVFKQ	YKTVSTKRLL	QELKKSPYAD	ELYDEHTGEI	
	KEVFGTQKEN	AFATSLSGYI	SMKSILGAVV	DDNPAMTEEL	IYWIAVFEDR	
	EILHLKIQEK	YPSITDVQRQ	KLALVKLPGW	GRFSRLIDG	LPLDEQGQSV	
	LDHMEQYSSV	FMEVLKNKGF	GLEKKIQKMN	QHQVDGTTKI	RYEDIEELAG	
	SPALKRGIWR	SVKIVEELVS	IFGEPANIVL	EVAREDEGEK	RTKSRKDQWE	
	ELTKTTLKND	PDLKSFIGEI	KSQGDQRFNE	QRFWLYVTQQ	GKCLYTCKAL	
	DIQNLSMYEV	DHILPQNFBK	DDSLDNLALV	MPEANQRKNQ	VGQNKMPLEI	
	IEANQQYAMR	TLWERLHELK	LISGKLGRL	KKPSFDEVDK	DKFIARQLVE	
	TRQIIKHVRD	LLDERFSKSD	IHLVKAGIVS	KFRRFSEIPK	IRDYNNKHHA	
	MDALFAAALI	QSILGKYGKN	FLAFDLSKKD	RQKQWRSVKG	SNKEFFLFKN	
	FGNLRQLSPV	TGEEVSGVEY	MKHVYFELPW	QTTKMTQTGD	GMFYKESIFS	
	PKVKQAKYVS	PKTEKFVHDE	VKNHSICLVE	FTFMKKEKEV	QETKFIDLKV	
	IEHHQFLKEP	ESQLAKFLAE	KETNSPIIHA	RIIRTIPKYQ	KIWIEHFPYY	
	FISTRELHNA	RQFEISYELM	EKVKQLSERS	SVEELKIVFG	LLIDQMNDNY	
	PIYTKSSIQD	RVQKFVDLQL	YDFKSFEIGF	EELKKAVAAN	AQRSDTFGSR	
	ISKPKPEEV	AIGYESITGL	KYRKPRSVVG	TKR		
Prevotella sp. C561 WP_009013303.1	MTQKVLGLDL	GTNSIGSAVR	NLDLSDDLQW	QLEFFSSDIF	RSSVNKESNG	(SEQ ID NO: 102)
	REYSLAAQRS	AHRRSRGLNE	VRRRLWATL	NLLIKHGFCP	MSSESLMRWC	
	TYDKRKGLFR	EYPIDDKDFN	AWILLDFNGD	GRPDYSSPYQ	LRRELVTQRF	
	DFEQPIERYK	LGRALYHIAQ	HRGFKSSKGE	TLSQQETNSK	PSSTDEIPDV	
	AGAMKASEEK	LSKGLSTYMK	EHNLLTVGAA	FAQLEDEGVR	VRNNNDYRAI	
	RSQFQHEIET	IFKFQQGLSV	ESELYERLIS	EKKNVGTIFY	KRPLRSQRGN	
	VGKCTLERSK	PRCAIGHPLF	EKFRAWTLIN	NIKVRMSVDT	LDEQLPMKLR	
	LDLYNECFLA	FVRTEFKFED	IRKYLEKRLG	IHFSYNDKTI	NYKDESTVAG	
	CPITARFRKM	LGEWESFRV	EGQKERQAHS	KNNISFHRVS	YSIEDIWHFC	
	YDAEPEAVL	AFAQETLRLE	RKKAEEVLRI	WSAMPQGYAM	LSQKAIRNIN	
	KILMLGLKYS	DAVILAKVPE	LVDVSDEELL	SIADYLLVE	AQVNYDKRIN	
	SIVNGLIAKY	KSVSEERYFA	DHNYEYLLDE	SDEKDIIRQI	ENSLGARRWS	
	LMDANEQTDI	LQKVRDRYQD	FFRSHERKFV	ESPKLGESFE	NYLTKKFPMV	
	EREQWKKLYH	PSQITIYRPV	SVGKDRSVLR	LGNPDIGAIK	NPTVLRVLNT	
	LRRRVNQLLD	DGVISPDETR	VVETARELN	DANRKWALDT	YNRIRHDENE	
	KIKKILEEFY	PKRDGISTDD	IDKARYVIDQ	REVVDYFTGSK	TYNKDIKKYK	
	FWLEQGGQCM	YTGRTINLSN	LFDPNADFIE	HTIPESLSFD	SSDMNLTLCD	
	AHYNRFIKKN	HIPTDMPNYD	KAITIDGKEY	PAITSQLQRW	VERVERLNRN	
	VEYWKQARR	AQNKDRKQD	MREMHLWKME	LEYWKKKLER	FTVTEVTDGF	
	KNSQLVDTRV	ITRHAVLYLK	SIFPHVDVQR	GDVTAKFRKI	LGIQSVDEKK	
	DRSLHSHHAI	DATTLTIIPV	SAKRDRMLEL	FAKIEEINKM	LSFSGSEDRT	
	GLIQLEGLK	NKLQMEVKVC	RIGHNVSEIG	TFINDNIIVN	HHIKNQALTP	
	VRRRLRKKGY	IVGGVDNPRW	QTGDALRGEI	HKASYGGAIT	QFAKDEGKV	
	LMKEGRPQVN	PTIKFVIRRE	LKYKSAADS	GFASWDDLK	AIVDKELFAL	
	MKGQFPAETS	FKDACEQGIY	MIKKGKNGMP	DIKLHHIRHV	RCEAPQSGLK	
	IKEQTYKSEK	EYKRYFYAAV	GDLYAMCCYT	NGKIREFRIY	SLYDVSCHRK	
	SDIEDIPEFI	TDKKGNRLML	DYKLRTGDMI	LLYKDNPAEL	YDLNVLNLSR	
	RLYKINRFES	QSNLVLMTHH	LSTSKERGRS	LGKTVDYQNL	PESIRSSVKS	
	LNFLIMGENR	DFVIKNGKII	FNHR			

Table 4- Cas9 orthologs						
Alicyclobacillus hesperidum URH17-3-68 WP_006446566.1	MAYRLGLDIG	ITSVGWAVVA	LEKDESGLKP	VRIQDLGVRI	FDKAEDSKTG	(SEQ ID NO: 103)
	ASLALPRREA	RSARRRTRRR	RHRLWRVKRL	LEQHGILSME	QIEALYAQRT	
	SSPDVYALRV	AGLDRCILAE	EIARVLIHIA	HRRGFQSNRK	SEIKDSDAGK	
	LLKAVQENEN	LMQSKGYRTV	AEMLVSEATK	TDAEGKLVHG	KKHGVSNNVR	
	NKAGEYRHTV	SRQAIIVDEV	KIFAAQRALG	NDVMSEELED	SYLKILCSQR	
	NFDDGPGGDS	PYGHGSVSPD	GVRQSIYERM	VGSCFTFETGE	KRAPRSSYSF	
	ERFQLLTQV	NLRIYRQQED	GGRYPCELTQ	TERARVIDCA	YEQTKITYGK	
	LRKLLDMKDT	ESFAGLTYGL	NRSRNKTEDT	VFVEMKFYHE	VRKALQIRAGV	
	FIQDLISIETL	DQIGWILSVW	KSDDNRRKKL	STLGLSDNVI	EELLPLNGSK	
	FGHLSLKAIR	KILPFLEDGY	SYDVACELAG	YQFQGKTEYV	KQRLPLPLGE	
	GEVTNPVVR	ALSQAIRVVN	AVIRKHGSPE	SIHIELAREL	SKNLDERRKI	
	EKAQKENQKN	NEQIKDEIRE	ILGSAHVTGR	DIVKYKLFQK	QQEFCMYSGE	
	KLDVTRLFEP	GAEVDHIIP	YGISFDDSYD	NKVLVKTEQN	RQKGNRTPLE	
	YLRDKPEQKA	KFIALVESIP	LSQKKKNHLL	MDKRAIDLEQ	EGFRENLSD	
	TRYITRALMN	HIQAWLLFDE	TASTRSKRVV	CVNGAVTAYM	RARWGLTKDR	
	DAGDKHHAAD	AVVVACIGDS	LIQRVTKYDK	FKRNALADRN	RYVQQVSKSE	
	GITQYVDKET	GEVFTWESFD	ERKFLPNEPL	EPWPFPRDEL	LARLSDDPK	
	NIRAIGLLTY	SETEQIDPIF	VSRMPTRKVT	GAHKETIRS	PRIVKVDDNK	
	GTEIQVVSX	VALTELKLT	DGEIKDYFRP	EDDPRLYNL	RERLVQGGD	
	AKAAFKEPVY	KISKDGSVRT	PVRKVKIQEK	LTGVPVHGG	RGIAENGGMV	
	RIDVFAKGGK	YYFVPIYVAD	VLKRELPNRL	ATAHKPYSEW	RVVDDSYQFK	
	FSLYPNDAMV	IKPSREVDIT	YKDRKEPVC	RIMYFVSANI	ASASISLRTH	
	DNSGELEGLG	IQGLEVFEEKY	VVGPLGDTHP	VYKERRMPFR	VERKMN	
Lactobacillus rhamnosus GG WP_014569977.1	MTKLNQPYGI	GLDIGSNSIG	FAVVDANSHL	LRLKGETAIG	ARLFREGQSA	(SEQ ID NO: 104)
	ADRRGSRTR	RRLSRTRWRL	SFLRDFPAPH	ITKIDPDPFL	RQKYSEISPK	
	DKDRFKYEKR	LFNDRTDAEF	YEDYPSMYHL	RLHLMTHTHK	ADPREIFLAI	
	HHILKSRGHF	LTPGAAKDFN	TDKVDLEDIF	PALTEAYAQV	YPDLELTFDL	
	AKADDFKAKL	LDEQATPST	QKALVNLLLS	SDGEKEIVKK	RKQVLTEFAK	
	AITGLKTKFN	LALGTEVDEA	DASNWQFSMG	QLDDKWSNIE	TSMTDQGTET	
	FEQIQELYRA	RLLNGIVPAG	MSLSQAKVAD	YGQHKEDLEL	FKTYLKKLND	
	HELAKTIRGL	YDRYINGDDA	KPFLREDFVK	ALTKEVTAHP	NEVSEQLLNR	
	MGQANFMLKQ	RTKANGAIP	QLQQRELDQI	IANQSKYYDW	LAAPNPVEAH	
	RWKMPYQLDE	LLNFHIPYYV	GPLITPKQQA	ESGENVFAMW	VRKDPSPGNIT	
	PYNFDEKVDR	EASANTFIQR	MKTDTTYLIG	EDVLPKQSL	YQKYEVNLNL	
	NNVRINNECL	GTQKQRLIR	EVFERHSSVT	IKQVADNLVA	HGDFARRPEI	
	RGLADEKRFL	SSLSTYHQLK	EILHEAIDDP	TKLLDIENII	TWSTVFEDHT	
	IFETKLAEIE	WLDPKKINEL	SGIRYRGWGQ	FSRKLLDGLK	LGNHGTIVQE	
	LMLSNHNLMO	ILADETLKET	MTELNQDKLK	TDDIEDVIND	AYTSPSNKKA	
	LRQVLRVVED	IKHAANGQDP	SWLFIETADG	TGTAGKRTQS	RQKQIQTVYA	
	NAAQELIDSA	VRGELEDKIA	DKASFTDRLV	LYFMQGGGRI	YTGAFLNIDQ	
	LSHYDIDHIL	PQSLIKDDSL	DNRVLVNATI	NREKNNVFAS	TLFAGKMKAT	
	WRKWHEAGLI	SGRKLRNLML	RPDEIDKFAK	GFVARQLVET	RQIIKLTEQI	
	AAAQYPNTKI	IAVKAGLSHQ	LREELDFPKN	RDVNHYHHAF	DAFLAARIGT	
	YLLKRYPKLA	PFFTYGEFAK	VDVKKFREFN	FIGALTHAKK	NIIAKDTGEI	
	VWDKERDIRE	LDRIYNFKRM	LITHEVYFET	ADLFKQTIYA	AKDSKERGGG	
	KQLIPKKQGY	PTQVYGGYTQ	ESGSYNALVR	VAEADTTAYQ	VIKISAQNAS	
	KIASANLKS	EKGKQLLNEI	VVKQLAKRRK	NWKPSANSFK	IVIPRFMGMT	
	LFQNAKYGLF	MVNSDTYYRN	YQELWLSREN	QKLLKKLFSI	KYEKTQMNHD	
	ALQVYKAIID	QVEKFFKLYD	INQFRAKLSD	AIERFEKLPI	NTDGNKIGKT	
	ETLRQILIGL	QANGTRSNVK	NLGIKTDLGL	LQVGSIGIKLD	KDTQIVYQSP	
	SGLFKRRIPL	ADL				
Enterococcus faecalis TX0012	MYSIGLGLGI	SSVGWSVIDE	RTGNVIDLGV	RLFSAKNSEK	NLERRTNRGG	(SEQ ID NO: 105)
	RRLIRRKTNR	LKDAKKILAA	VGFYEDKSLK	NSCPYQLRVK	GLTEPLSRGE	
	IYKVTLHILK	KRGISYLDEV	DTEAAKESQD	YKEQVRKNAQ	LLTKYTPGQI	
	QLQRLKENNR	VKTGINAQGN	YQLNVFKVSA	YANELATILK	TQQAFFYPNEL	
	TDDWIALFVQ	PGIAEEAGLI	YRKRPPYHGP	GNEANNSPYG	RWSDFQKTGE	

Table 4- Cas9 orthologs						
WP_002408 901.1 EFT93846.1	PATNIFDKLI	GKDFQGELRA	SGLSLSAQQY	NLLNDLTNLK	IDGEVPLSSE	
	QKEYILTELM	TKEFTRFGVN	DVVKLLGVKK	ERLSGWRLDK	KGKPEIHTLK	
	GYRNWRKIFA	EAGIDLATLP	TETIDCLAKV	LTLNTEREGI	ENTLAFELPE	
	LSSESVKLLVL	DRYKELSQSI	STQSWHRFSL	KTLHLLIPEL	MNATSEQNTL	
	LEQFQLKSDV	RKRYSEYKKL	PTKQVLAIEY	NPTVNKTVSQ	AFKVIDALLV	
	KYGKEQIRYI	TIEMPRDDNE	EDEKKRIKEL	HAKNSQRKND	SQSYFMQKSG	
	WSQEKFQTTI	QKNRRFLAKL	LYYYEQDGIC	AYTGLPISPE	LLVSDSTEID	
	HIIPISISLD	DSINNKVVLV	SKANQVKGQQ	TPYDAWMDGS	FKKINGKFSN	
	WDDYQKWVES	RHFSHKKENN	LLETRNIFDS	EQVEKFLARN	LNDTRYASRL	
	VLNTLQSFFT	NQETKVRVVN	GSFTHTLRKK	WGADLDKTRE	THHHHAVDAT	
	LCAVTSFVKV	SRHYHAVKEE	TGEKVMREID	FETGEIVNEM	SYWEFKKSKK	
	YERKTYQVKW	PNFREQLKPV	NLHPRIKFSS	QVDRKANRKL	SDATIYSVRE	
	KTEVKTLSKG	KQKITTDDEY	IGKIKDIYTL	DGWEAFKKKQ	DKLLMKDLDE	
	KTYERLLSIA	ETTPDFQEVE	EKNGKVKRVK	RSPFAVYCEE	NDIPAIQKYA	
	KKNNGPLIRS	LKYYDGKLNK	HINITKDSQG	RPVEKTKNGR	KVTLSQLKPY	
	RYDIYQDLET	KAYYTVQLYY	SDLRFVEGKY	GITEKEYMKK	VAEQTKGQVV	
	RFCFSLQKND	GLEIEWKDSQ	RYDVRFYNFQ	SANSINFKGL	EQEMMPAENQ	
	FKQKPYNNGA	INLNIACYGK	EGKKLRKFNT	DILGKKHYLF	YEKEPKNIK	
Candidatus Puniceispiril lum marinum IMCC1322 WP_013047 413.1	MRRLLGLDGT	NSIGWCLDL	GDDGEPVSIF	RTGARIFSDG	RDPKSLGSLK	(SEQ ID NO: 106)
	ATRRREARLTR	RRDRFIQQR	KNLINALVKY	GLMPADEIQR	QALAYKDPYP	
	IRKKALDEAI	DPYEMGRAIF	HINQRRGFKS	NRKSADNEAG	VVKQSIADLE	
	MKLGEAGART	IGEFLADRQA	TNDTVRARRL	SGTNALYEFY	PDRYMLEQEF	
	DTLWAKQAAF	NPSLYIEAAR	ERLKEIVFFQ	RKLKPQEVGR	CIFLSDEDRI	
	SKALPSFQRF	RIYQELS NLA	WIDHDGVAHR	ITASLALRDH	LFDELEHKKK	
	LTFKAMRAIL	RKQGVVDYPV	GFNLESDNRD	HLIGNLTSCI	MRDAKKMIGS	
	AWDRLDDEEQ	DSFILMLQDD	QKGDDEVRSI	LTQQYGLSDD	VAEDCLDVRL	
	PDGHGSLSKK	AIDRILPVL	DQGLIYYDAV	KEAGLGEANL	YDPYAALSDK	
	LDYYGKALAG	HVMGASGKFE	DSDEKRYGTI	SNPTVHIALN	QVRVAVNELI	
	RLHGKPDDEV	IEIGRDLPMG	ADGKRELERF	QKEGRAKNER	ARDELKKLGH	
	IDSRESRQKF	QLWEQLAKEP	VDRCCPFTGK	MMSISDLFSD	KVEIEHLLPF	
	SLTLDDSMAN	KTVCFRQANR	DKGNRAPFDA	FGNSPAGYDW	QEILGRSQNL	
	PYAKRWRFLL	DAMKRFEADG	GFLERQLNDT	RYISRYTTEY	ISTIIPKNKI	
	WVVTGRLLTS	LRGFWGLNSI	LRGHNTDDGT	PAKKS RDDHR	HHAIDAIVVG	
	MTSRGLLQKV	SKAARRSEDL	DLTRLFEGRI	DPWDGFRDEV	KKHIDAIIVS	
	HRPRKKSQGA	LHNDTAYGIV	EHAENGASTV	VHRVPITSLG	KQSDIEKVRD	
	PLIKSALLNE	TAGLSGKSFE	NAVQKWCADN	SIKSLRIVET	VSIIPITDKE	
	GVAYKGYKGD	GNAYMDIYQD	PTSSKWKGEI	VSRFDANQKG	FIPSWQSQFP	
	TARLIMRLRI	NDLLKLQDGE	IEEIYRVQRL	SGSKILMAPH	TEANVDARDR	
	DKNDTFKLTS	KSPGKLQSAS	ARKVHISPTG	LIREG		
Oenococcus kitaharae DSM 17330 EHN59352. 1	MARDYSVGLD	IGTSSVGWAA	IDNKYHLIRA	KSKNLIGVRL	FDSAVTAEKR	(SEQ ID NO: 107)
	RGYRTTRRRL	SRRHWRLLRL	NDIFAGPLTD	FGDENFLARL	KYSWVHPQDQ	
	SNQAHFAAGL	LFDSKEQDKD	FYRKYPTIYH	LRLALMNDQ	KHDLREVYLA	
	IHHLVKYRGH	FLIEGDVKAD	SAFDVHTFAD	AIQRYAESNN	SDENLLGKID	
	EKKLSAALTD	KHGSKSQRAE	TAETAFFDILD	LQSKKQIQAI	LKS VVG NQAN	
	LMAIFGLDSS	AISKDEQKNY	KFSFDDADID	EKIADSEALL	SDTEFEFLCD	
	LKAAFDGLTL	KMLLGDDKTV	SAAMVRRFNE	HQKDWEYIKS	HIRNAKNAGN	
	GLYEKSKKFD	GINAAYLALQ	SDNEDDRKKA	KKIFQDEISS	ADIPDDVKAD	
	FLKKIDDDQF	LPIQRTKNNG	TIPHQLHRNE	LEQIIEKQGI	YYPFLKDTYQ	
	ENSHELNKIT	ALINFRVPYY	VGPLVEEEQK	IADDGKNIPD	PTNHWMVRKS	
	NDTITPWNLS	QVVDLDKSGR	RFIERLTGTD	TYLIGEPTLP	KNSLLYQKFD	
	VLQELNNIRV	SGRRLLDIRAK	QDAFEHLFKV	QKTVSATNLK	DFLVQAGYIS	
	EDTQIEGLAD	VNGKNFNNAL	TTYNYLVSVL	GREFVENPSN	EELLEIEITEL	
	QTVFEDKKVL	RRQLDQLDGL	SDHNREKLSR	KHYTGWGRIS	KKLLTTKIVQ	
	NADKIDNQTF	DVPRMNQSI	DTLYNTKMNL	MEIINNAEDD	FGVRAWIDKQ	
	NTTDGDEQDV	YSLIDELAGP	KEIKRGIVQS	FRILDDITKA	VGYPAPRVYL	
	EFARKTQESH	LTNSRKNQLS	TLLKNAGLSE	LVTQVSQYDA	AALQNDRLYL	

Table 4- Cas9 orthologs						
	YFLQQGKDMY RRKSDSSNYL RFIARSLVET NRDINDYHHA HAARLNAQSQ DYLHKVMNFR PVDLYGGFSS GKYPKAEMVL NTLLKTKEKN ALVKLQAHRD RLFSPSHFTL	SGEKLNLDNL PALIDKMRPF RQIIKNVASL FDALLFSTVG RVNPFQFVVG KITVTRRLKD AKPAYAALIK SHIPYQQLVK LVDILHNRLD DFNALNYEDK ADTDEFIFQS	SNYDIDHIIP WSVLSKQGLL IDSHFGGETK QYTENSGLMK SMRNAAPGKL QKGQLYDESR FKNKFRLNVV KDGALVTISS LPEMTIESAF QOTLERILDA VTGLFSTQKT	QAYTKDNSLD SKHKFANLTR AVAIRSSLTA KGQLSDSAGN NPETGEITPE YPSVLHDAKS LRQWTYSKDN ATELHNFEQL YKAFDSILSF LHASPASSDL VAQLYQETK	NRVLVSNITN TRDFDDMEKE DMRRYVDIPK QYNRYIKEWI ENADWSIADL KASINFDKHK SEDIYILEQIR WLPLADYKLI AFNRYALHQN KKINLSSSGFG	
Helicobacter mustelae 12198 WP_013022 389.1	MIRTLGIDIG ESLALPRTLA LFQTSKDFLS NDSGKIKKAI VGRSELREEL YQRPKLGFGD LTNRHGLCFS GLDYSKGNPE KELESQQLSK VEILQERGIF VVNALLEKYG LEVGLPNTY PLSRSLDDSQ NFSPPSKKRKL LPDGKKEHIR MIQKYTTYLK QNIIPSHHVS NQGIVDNGAM GIIKWLEMD ELEHLSKYAL LGEVLEHKPR	IASIGWAVIE RSARRRNARK PWELRERALLY AENSKRIKEE KTIFQIQQEL KIGKCSHIKK QEDMCVYLK KAVFLSLPST EQIQTIKDAK SKPQPKNRQL GFHYFHIELT NNRLKCKLWK SNKVLCLTSS TQKNFKERNE IISGSMTSTM DKETHRLKSH HKVTGELHQE VRVDIFKSKD ENYEFCSLSF KNEDEEKMFT NRQNIALKTT	GEYTDKGLN KGRIQQVKHY RVLDKEELAR QCKTIGEMMY KSPWVNEELI GENSPYRACK ILQEAQKNEK FKLNKITQDR LNFSKHINLS LPPLSELAKE RDVCKAKSAR QQEYCYLYSG NQEKSNTKPY EDFLARNLVD RSFWGVQEK QKAQILREGD TVRTKEFYQ KGKFYAVPIY KNDCIKIQTK DTDKEKNKTM PKHV	KEIVASGVRV LSKALGLDLE VILHIAKRRG KLYFQKSLNV YKLLGNPDAQ HAPSAEEFVA GLTYSKLKLL KTQDKIANIL LEALYHLLPL ESYFDIPNPV MQLEKINKKN EKITIDHLKD EWLGSDEKKW TG YIGRVTK RDHHLHHAQD HKLSLRWPMS AFGGEEGVKK TYDFAIGKLP EMQEAVALAIY TRESGCIQGL KVFQKVKLSV	FTKAENPKNK CFVQGEKLAT YDDITYGVED RNKKESYNRC SKQEREGLIF LTKSINFLKN LDLPSDFEFL GANKDWEAIL MREGKRYDEG LRRALSEFRK KSEDAASQL QRALQIDHAF DMYVGRVYSS YIKHSLSFLP AII IACIEPS NFKDKIQESI ALKFKGIREI NKAIVQGKKN KSTNSAKATI KVFQKVKLSV	(SEQ ID NO: 108)
Bradyrhizobium sp. BTai1 WP_012044 026.1	MKRTSLRAYR QSKQSNAAGR ECLDPYGLRA IKAAAGRLQT YPTRAMVDAE CSLDPATSDQ EDQDKVARAL ILSARRHFGP AQRALSALLP GKLSPTGYLP RVVNELIRWH LLRKLGLPAS HLIPFSISWD ERAARLPRGK THPHRIHVLP AALTDQALLR HGLQARLHED DLVRAHVAGE LVPIRDKAGR DAPAAQPIMR FQTRHDDPED DATRIGWTRP	LGVDLGANSI RLARSARRRR KALDEVPLPH SMQACGARTL FEAIWAAQAP DVDGFRCAWS LQTDQLSFDE AWHDSRLDRQ DGYCRLGLRA YYGQWLQNDV GPPAEITVEL THNLLKLRLW DSAANKVVCN RWRFGPGARA GRLTALLRAT RMANAHDDTR TAYGTVEHPE RQQGKTLKAA IYKSYNAGEN VFKGDMRLID SFRWLFASF KKWP	GWFFVWLDDH DRYLQRRGKL HVGRALFHLN GEFLNRRHQL HHPTMTAEAH HPLAQRFRIW IRGLLGLPSD IDIVALLES IKRVLPLMEA VGSDDERDTN TRDLKLSPRR DEQGDVASEC RYANREKGNR QFEELGDFQA WELNDLLPGS RKIEVLLPWP TEDGANLVYR VLSFAQRRDI AFVDILQAES HAGAEKFVKI KLREWNALV RIDTLGQPPW	GQPEGLGPGG MGLLVKHGLM QRRGLFANRA RATVRARSPV DTIREAIFSQ QDVRNLAVVE ARFNLESDDR LDEAAIIASL GRTYAEAAASA ERRWGRLPNP LAELEREQAE PYTGEAIGLE TPFEAFGHRQ RLLNETSWLA DDRAAKSRKD TFRIDLETRL KTFVDISEKE AGHPNGIRHV GRWIARATTV VRLSPSNNLL RIDTLGQPPW	VRIFPDGRNP PADEPARKRL IEQGDKDASA GGDVQARYEF RAMKRPSIGK TGPTSSRLGK DHLKGDATGA GTTHSLDEAA AGYDHALLPG TVHIGIGQLR NQRKNDKRTS RLVSDDVDID GRPYDWADIA RVAKQYLAAY HRHHAI DALV KAMLVSHKPD IDRIRDRRLR RLTKSIKPDY FQANQANESH YLVEHHQAGV RKRGLTGESE	(SEQ ID NO: 109)

Table 4- Cas9 orthologs						
Acidaminococcus sp. D21 WP_009016219.1	MGKMYYLGLD RSFRTSRRRL ETDKHIFFND LVAHRGHFLN ATLLSRNSVN NKLFPQESND HALKDGRTIS TKNYVAYSYS EMIQRLLTNS KDAISNQDKL VDLDKSEAEF YPISVDVKQQ SNYNTYHHFK TADDVKHISR LMQLLSECYT LAVVNDIRKA FQQEVDLFLEK SFYNIDHIYP WDAYYNHGLI LKRKFDPTEI VYHERFNRRW KQNKNTIHFT TAAYYLLVRF EILQKDKQKV AMVPLIVPHK LFLQKLQHNP CRSSGCDLKS SQNLLEYL	IGTNSVGAYV DRRQQRVKLV PTYTDKEYYS EVDKDNIGDV DKYKALKSLI ASFTLNDKED ESKVKLYEQH VKEVKGTLPK FMPKQVSGEN LSIMTFRIPY VFGLEFEKRR SLMKEKRLTR LKRKDFGRSL FSDEITKLQE CGTAPKRIFI ILENKSDGQL QSMVKDDSLD SLKKYQRLTR VYSKAGLSSD FMVNQPYSVK RFSFDRKEGL TLEDKKTQHK INIMFPMGTR IECYIKAMES YNKFFSTQFD INGSAQAARI SQNLEYL	TDPSYHLLKF QEIFAPVISP DYPTIHHLIV LSFDAFYPEF FGSQKPEDNF AIEEILGTLT HHDLTQLKYF NKATQEEFCK RVIPYQLYYY FVGPLRKDNS VTVKDIQNLL DDVERIVERM KMFLTGLKGV AYYAKAQLSL EMARDGESKK QSDALYLYFA NKVLVQSEIN STPFTDDEKW FRHEFGLVKS TKTLFTHSIK FDIQPLKAST LMMIPVEGLY HIKLNSMISI FARKFKENNK VLTNGRSTFT MISADLTGLS	KGPEMWGAHV IDPRFFIRLH DLMESSEKHD LAFLSDNGVS DANISEDGLI PDECEWIAHI VKTYLAKEYD YVLGKVKNIE ELKTILNKAA EHAWLERKAG SLIYEKFMIL TYSDDTKRVR HKETGERASI NDFLDSMYIS KRSVTRREQI QLGRDMYTG GEKSSRYPLD DFINRQLVET RNINDLHHAK NGNFVAWNGE GLVPRKAGLD KARIDHDKEF DGFYLSIGGK LRIVEKFDKI KLSPEEQVQT KKYSDIRLVE	FAAGNQSAER ESALWRDDVA PRLVYLAVAW PWVCESKALQ QLLAGKKVKV RRLFDWAIMK DIFRNVDSST CSEADKVDFF SYLPFLTQCG KIYPWNFNCK NEINNIRIDG KLTGIDTTIH LWLNNNYCTL LDFMWNTNDN NAVKRPIYRT KNLYRSIRKD PIKLEHIKQD AAIRNKMKPL RQSTKALAIL DAFLAIVTGN EDLGRIVKML VVKYGGYDKS LTDYAQTTIS SSKKGKSVLCH TVEDNLNLYE LLNLSIFPKT QSASGLFVSK	(SEQ ID NO: 110)
Methylosinus trichosporium OB3b WP_003611034.1	MRVLGLDAGI GAKLKSEQRR GLDPWRIRAE SKMKRAVNET REGDYSRSL LQDSEKLVGP TRDELALAAA RSGKAAEGTA ISEGLAQAGC GMTYDKACCA PTARKALIES KRNRRQDKLR ISPSQLVATD WFKAEKTDTE DTRWACRLLA WMKKSTKGDR LVKNVTPPWP EQRVYERRKV PKGDPFVKVR KYEYLVPIY TYLQVISSKG KKFNVDRFGR	ASLGWALIEI TFRGQRRVVR ALDRLLGPVE REKLARFGSA RDDLAAEMRA CPFEVDEKRA DFGAAAKVSF RLRSVIVDAL NAPLVDALTA ADYDHTASRE IKQVKAIVER GLFEKEVGRP DAVQVDHILP WDAFIVRVEA EALKQLYPKG IPDDRHHALD GFREQAVEAV AELKLADLDR LVTKSKVNIA PHDIATMKTP EIFEGYYRGM KHEVERELRT	EESNRGELSQ RRRQRMNEVR LAVALGHIAI AKMLVEDESF LFTAQARFQS PKRGYSFELF TALRKKLKLP GELAWGALLC AASDGRFDPF RGAFDVGGHG YGVDPRIHVE PQDGARGKEE WSRFADDSYA LADMKGFKKR EKDKDGKERR AIVIAATTES EKVFVARAER VKDAERNARL LDTGNPKRPG PIRAVQAYKP NRSVGAIQLS WRGETWRGKA	GTIIGAGTWM RILHSHGLLP HRGFKSNSKG VLRQTPTKNG AIATADLQTA RFLSRLNHVT ETTTFVGVKA SPEKLDKIAE TGAGHISKA REALKRILQE LARDVGKSIE LLRFELWSEQ NKTLCMAKAN NYKLRNAEEA RVFSRPGALT LLQRATREVQ RRARGKAHDA IEKLNRNIEA TVDRGEMARV EDEWPEMDSS AHSNSSDVVQ YI	FDAPEEKTQA SSDRDALKQP AKTNDPADDT ASEIVRRFRN FTKAAFFQRP LRDGKQERTL DEESKLDVVA VISFRSDIGR ARNILSGLRQ ERISRELVGS EREETRGIE MGRCLYTDY QDKKGRTPEY AAKFRNRNLN DRLRRAWGLQ EIEDKGLHYD TIRHIAVREG GSPKDDPPLS DVFRKASKKG YEFWCWSLVP GIGARTLTET	(SEQ ID NO: 111)
Actinomyces coelicolor DSM 15436	MDNKNYRIGI KTNTTRLAMS YTPWLVRAL EQPPSDQYLA KTDTRPENKK ADSPKGASGE	DVGLNSIGFC GVARTRRLF AQTPIRDEND LKERVEAKTL PGFLGGKLMQ LVGYDVLPGQ	AVEVDQHDTP RKRKRRLAAL LHEKLAIABR LQMPGATPA SDNANELRKI HGKRAEAKAH	LGFLNLSVYR DRFIEAQGT HIARHRGWRS EMVVALDLSV AKIQGLDDAL PAFQRYRIAS	HDAGIDPNGK LPDHADYKDP PWVPSRLHV DVNLRPKNRE LRELIELVFA IVSNLRIRHL	(SEQ ID NO: 112)

Table 4- Cas9 orthologs						
WP_006546 479.1	GSGADERLDV QTADGERASA AEKLTEGTAA KRMIENGEDL LMAAAEAWGA ADHINATSGV LGSTNTRDNL IADGFASSKE QYYFDEKHTG AMDAATVAML FESWSENMRV KVGDAWSLTE GPLDKVGIFG APDLLRYRNE ELLLDLSSET GEDVSEGSKS VSWQG	ETQKRVEFYL KPPVDVTNVA EVEVAEFLQN FEARVNEFGV PLSVNIEHVR RGSVDVTRYLA VATCERCNKS HRELQKGVKD TKVRVFRGSL RNSVAKTLVL LVEKFNLALY IDRASTPALW KAAASLLVRG DLFNVELPPQ SGQIAELQQD IIAGQGWRPA	INAKPTADIT FATCKIKPLK LSDEDNEKLD SEDWRPPAEP EGFISKRQAV IQRQNGECLY KSNKPPFAVWA RLKRKVSDPE TSAARKASGF RGNIRASERA NDEVSIFSSL CALTRQPDFT GSVDIGSAIH SVSMRYAEPK FPGTTHWTVA VNKVFGSAMP	WSDVAEEIGV EWWLNADYEA SFSLPIGRAA IGARVGNPAV EIDRENQKRY CGTAITFVNS AECGIPGVSV IDNRSMESVA ESRVNFIGGN IGAAETWKSF RLQLGNGKAH WKDGLPANED HARIYRIAGK VREAIREGKA GFFSPSRLRL EVIRRDGLGR	ERNLLMGAT RCVMVSALSH YSVDSLRLT DRVLKAVNRY QRNQAVRSQI EMDHIVPRAG AEALKRVDFW WMARELAHRV GKTRLDRRHH RGENVADRQI DDTITKLQMH RTIIVNGTHY KPTYGMVRVF EYLGWLTVGD RPVYLAQEG KRRFSYSGLP	
Caenispirillum salinarum AK4 WP_009541 330.1	MPVLSPLSPN TLFPSAWSNE RSPEDLKDLT IRLAELQEVD RHGTQPTCGE FDAIRAIQAP FQGRITITREA RAMPELGLSH DNSPVCFAQE ARGADSAALA WAAEAILDAW RTMAAARTPN EERNPFLRTW LAREAKHGI KERAALRLRL DGPDLNLVLAH KGKGHKARTP TRLARLYLAA NGSVTHMLRQ GADAPDDRPV QDPETGRVHW ATHYGRREIT TILKEICTEI LADLDALADA TRALDPHWG TDGNAVSQLW GVTGNNAPPR GKAMRQTLTD TQGIFGMPVD	AAQGRRRWSL NGTYVAHGAA RTPPKADPRA PPESPADDDA IMAGRLRETP RHPDLPWDSL IDRGLTVDP GELTAPERDT RDTSGGGITV RLLAEGAHGV ANAPTEGFYD RAAQQRGPLP TGNAATDHIL RRNEIAKENR AQRQEFFCPY KDCNNAGKK REDKDFMDRV AVMPEDPAEI RLLQRDKNRD DLMPYSDAYH RLTRAGRGLK VDGRTDTVVT ADRHDRVDP DRTPEQEARR PRGLRHLIMR RLTSVVTDDG PLRQDIDRLT AGITAEAGWR AAGSAPRTPE	ALDIGEGSIG DRAVRGQQQR IFFLRADAAR APAATEDEDG AGAQPVTAR RRLVLDQAPI IQALRIRETV LVRALMHDPD NPTDPLMARW PPVAAAAPPA VTRGLFGFAP AYESVIPSQ NQVRKTANEV ENEGRRKKES CAERPCLRAT TPHEHAGDLL GWRFEEDARA GAPPVETPPS YQTHHAEDAC QRRRARALGR RRIDDLTRNC QRMNARDLVA EGTHARRWIS SALRQSPYLG EARAPSLVRI RRIPLPKPIE PLWRDHGTAP LDSEGAVCDL DCEKFEEQYG	WAVAEVDAEG HDSRRRLLAG RPLDGPFLFR TRRAAADERA DGLRVGGGVA AVPPATPCLF GNLRLHERIT GLAAKDGRIP IDGWVDLPLK ATAAILEDSDI GEIVLEDLRR ITSLRRAHKG ITKYGNRRGW AALDTFCQDN LLDNAKIVPA RAISAKKADG RANKTDAFGR KRIEISNLEY GGYLGTAUGE EVAKGDTVKK IKPWKAKGIP	RVLQLTGTGV LARLCAPVLE VLHHMAAHRG FRRLMAEHMH VPTRALIEQE LEELRRRGET EPDGRQRYVP YTRLRKLIGY ARSLYVRDVG MQPGRYSVCP ARGALLAHL RAADWSAADP DPLPSRITVE TVSWQAGGLP DHVIERRMGG GWRKENADRL RRMLHDTARA YRTISRQPV QAFNTEAAQH ADIVMPESDR ETGTPGALHN ARLDAAAPGD AEAVARDIAE RARAREQEIL PVPDAAVWVK ARLNGLDEGA LEDKARSALR DGKTYKVGVI LA	(SEQ ID NO: 113)
Coriobacterium glomerans PW2 WP_013709 575.1	MKLRGIEDDY TAAVARMPRG RDDRAEEHAD LIYLALHNIV ECSIADNGSI GAVIGLKTEF CEVYSAYVLQ YRMFFSGATY VEKLFKATDA AIVENQGRFY	SIGLDMGTSS QRRRYVRRRW YRWPLFNDCK KHRGNFLREG LAMLTHPDLS KNIFGDFPCE GLLSYAPGQT PGTGIYDAAQ RADERYRMMM PFLKRDADKL	VGWAVTDERG RLDLLQKLFE FTERDYYQRF QSLSAKSARP PSDRKKIAP DSSIYLSNDE ISANMVEKYR ARGYTKYNLG DRFDKQQFLR VSLVSFRIPY	TLAHFKRKPT QQMEQADPDF PTIYHVRSWL DEALNHLRET LFDVKSDDAA AVDAVRSACP RYGEDLALLK PKKSEYKPSE RLKTSNNGSI YVGPLSTRNA	WGSRLFREAQ FIRLRQSRL METDEQADIR LRVWSSERGF ADKKLGIALA DDCAELFDRL KLVKIYAPDQ SMQYDDFRKA YHQLHLEELK RTDQHGNGRF	(SEQ ID NO: 114)

Table 4- Cas9 orthologs

AWSERKPGMQ	DEPIFPWNWE	SIIDRSKSAE	KFILRMTGMC	TYLQQEPVLP
KSSLLYEFC	VLNELNGAHW	SIDGDDEHRF	DAADREGIIE	ELFRRKRTVS
YGDVAGWMER	ERNQIGAHVC	GGQGEKGFES	KLGSYIFFCK	DVFKVERLEQ
SDYPMIERII	LWNTLFEDRK	ILSQRLKEEY	GSRLSAEQIK	TICKKREFTGW
GRLSEKFLTG	ITVQVDEDSV	SIMDVLREGC	PVSGKRGRAM	VMMEILRDEE
LGFQKKVDDF	NRAFFAENAQ	ALGVNELPGS	PAVRRSLNQS	IRIVDEIASI
AGKAPANIFI	EVTREDEPKK	KGRRTKRRYN	DLKDALEAFK	KEDPELWREL
CETAPNDMDE	RLSLYFMQRG	KCLYSGRAID	IHQLSNAGIY	EVDHIIPRTY
VKDDSLENKA	LVYREENQRK	TDMLLIDPEI	RRRMSGYWRM	LHEAKLIGDK
KFRNLLRSRI	DDKALKGFIA	RQLVETGQMV	KLVRSLLEAR	YPETNIISVK
ASISHDLRTA	AELVKCREAN	DFHHAHDAFL	ACRVGLFIQK	RHPCVYENPI
GLSQVVRNYV	RQQADIFKRC	RTIPGSSGFI	VNSFMTSGFD	KETGEIFKDD
WDAEAEVEGI	RRSLNFRQCF	ISRMPFEDHG	VFWDATIYSP	RAKKTAAALPL
KQGLNPSRYG	SFSREQFAYF	FIYKARNPRK	EQTLEFEFAQV	PVRLSAQIRQ
DENALERYAR	ELAKDQGLEF	IRIERSKILK	NQLIEIDGDR	LCITGKEEVR
NACELAF AQD	EMRVIRMLVS	EKPVSRCEVI	SLFNRIILLHG	DQASRRLSKQ
LKLALLSEAF	SEASDNVQRN	VVLGLIAIFN	GSTNMVNLS	IGGSKFAGNV
RIKYKKELAS	PKVNVHLIDQ	SVTGMFERRT	KIGL	

[0300] In some embodiments, prime editors utilized herein comprise CRISPR-Cas system enzymes other than type II enzymes. In certain embodiments, prime editors comprise type V or type VI CRISPR-Cas system enzymes. It will be appreciated that certain CRISPR enzymes exhibit promiscuous ssDNA cleavage activity and appropriate precautions should be considered. In certain embodiments, prime editors comprise a nickase or a dead CRISPR with nuclease function comprised in a different component.

[0301] In various embodiments, the nucleic acid programmable DNA binding proteins utilized herein include, without limitation, Cas9 (e.g., dCas9 and nCas9), Cas12a (Cpf1), Cas12b1 (C2c1), Cas12b2, Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), C2c4, C2c5, C2c8, C2c9, C2c10, Cas13a (C2c2), Cas13b (C2c6), Cas13c (C2c7), Cas13d, and Argonaute. Cas-equivalents further include those described in Makarova et al., “C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector,” *Science* 2016; 353(6299) and Makarova et al., “Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?,” *The CRISPR Journal*, Vol.1. No.5, 2018, the contents of which are incorporated herein by reference. One example of a nucleic acid programmable DNA-binding protein that has different PAM specificity than Cas9 is Clustered Regularly Interspaced Short Palindromic Repeats from *Prevotella* and *Francisella* 1 (i.e., Cas12a (Cpf1)). Similar to Cas9, Cas12a (Cpf1) is also a Class 2 CRISPR effector, but it is a member of type V subgroup of enzymes, rather than the type II subgroup. It has been shown that Cas12a (Cpf1) mediates robust DNA interference with features distinct from Cas9. Cas12a (Cpf1) is a single RNA-

guided endonuclease lacking tracrRNA, and it utilizes a T-rich protospacer-adjacent motif (TTN, TTTN, or YTN). Moreover, Cpf1 cleaves DNA via a staggered DNA double-stranded break. Out of 16 Cpf1-family proteins, two enzymes from *Acidaminococcus* and *Lachnospiraceae* are shown to have efficient genome-editing activity in human cells. Cpf1 proteins are known in the art and have been described previously, for example Yamano et al., “Crystal structure of Cpf1 in complex with guide RNA and target DNA.” *Cell* (165) 2016, p.949-962; the entire contents of which is hereby incorporated by reference.

6.3. Type V CRISPR proteins

[0302] In some embodiments, prime editors used herein comprise the type V CRISPR family includes *Francisella novicida* U112 Cpf1 (FnCpf1) also known as FnCas12a. FnCpf1 adopts a bilobed architecture with the two lobes connected by the wedge (WED) domain. The N-terminal REC lobe consists of two α -helical domains (REC1 and REC2) that have been shown to coordinate the crRNA-target DNA heteroduplex. The C-terminal NUC lobe consists of the C-terminal RuvC and Nuc domains involved in target cleavage, the arginine-rich bridge helix (BH), and the PAM-interacting (PI) domain. The repeat-derived segment of the crRNA forms a pseudoknot stabilized by intra-molecular base-pairing and hydrogen-bonding interactions. The pseudoknot is coordinated by residues from the WED, RuvC, and REC2 domains, as well as by two hydrated magnesium cations. Notably, nucleotides 1–5 of the crRNA are ordered in the central cavity of FnCas12a and adopt an A-form-like helical conformation. Conformational ordering of the seed sequence is facilitated by multiple interactions between the ribose and phosphate moieties of the crRNA backbone and FnCpf1 residues in the WED and REC1 domains. These include residues Thr16, Lys595, His804, and His881 from the WED domain and residues Tyr47, Lys51, Phe182, and Arg186 from the REC1 domain. The structure of the FnCas12a-crRNA complex further reveals that the bases of the seed sequence are solvent exposed and poised for hybridization with target DNA. Structural aspects of FnCpf1 are described by Swarts et al., Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a, *Molecular Cell* 66, 221-233, April 20, 2017.

[0303] *Pre-crRNA processing*: Essential residues for crRNA processing include His843, Lys852, and Lys869. Structural observations are consistent with an acid-base catalytic mechanism in which Lys869 acts as the general base catalyst to deprotonate the attacking 2'-hydroxyl group of U(-19), while His843 acts as a general acid to protonate the 5'-oxygen

leaving group of A(-18). In turn, the side chain of Lys852 is involved in charge stabilization of the transition state. Collectively, these interactions facilitate the intra-molecular attack of the 2'-hydroxyl group of U(-19) on the scissile phosphate and promote the formation of the 2',3'-cyclic phosphate product.

[0304] *R-loop formation:* The crRNA-target DNA strand heteroduplex is enclosed in the central cavity formed by the REC and NUC lobes and interacts extensively with the REC1 and REC2 domains. The PAM-containing DNA duplex comprises target strand nucleotides dT0–dT8 and non-target strand nucleotides dA(8)*–dA0* and is contacted by the PI, WED, and REC1 domains. The 5'-TTN-3' PAM is recognized in FnCas12a by a mechanism combining the shape-specific recognition of a narrowed minor groove, with base-specific recognition of the PAM bases by two invariant residues, Lys671 and Lys613. Directly downstream of the PAM, the duplex of the target DNA is disrupted by the side chain of residue Lys667, which is inserted between the DNA strands and forms a cation- π stacking interaction with the dA0–dT0* base pair. The phosphate group linking target strand residues dT(-1) and dT0 is coordinated by hydrogen-bonding interactions with the side chain of Lys823 and the backbone amide of Gly826. Target strand residue dT(-1) bends away from residue T0, allowing the target strand to interact with the seed sequence of the crRNA. The non-target strand nucleotides dT1*–dT5* interact with the Arg692-Ser702 loop in FnCas12a through hydrogen-bonding and ionic interactions between backbone phosphate groups and side chains of Arg692, Asn700, Ser702, and Gln704, as well as main-chain amide groups of Lys699, Asn700, and Ser702. Alanine substitution of Q704 or replacement of residues Thr698–Ser702 in FnCas12a with the sequence Ala-Gly3 (SEQ ID NO: 115) substantially reduced DNA cleavage activity, suggesting that these residues contribute to R-loop formation by stabilizing the displaced conformation of the nontarget DNA strand.

[0305] In the FnCas12a R-loop complex, the crRNA-target strand heteroduplex is terminated by a stacking interaction with a conserved aromatic residue (Tyr410). This prevents base pairing between the crRNA and the target strand beyond nucleotides U20 and dA(-20), respectively. Beyond this point, the target DNA strand nucleotides re-engage the non-target DNA strand, forming a PAM-distal DNA duplex comprising nucleotides dC(-21)–dA(-27) and dG21*–dT27*, respectively. The duplex is confined between the REC2 and Nuc domains at the end of the central channel formed by the REC and NUC lobes.

[0306] *Target DNA cleavage:* FnCpf1 can independently accommodate both the target and non-target DNA strands in the catalytic pocket of the RuvC domain. The RuvC active site contains three catalytic residues (D917, E1006, and D1255). Structural observations suggest that both the target and non-target DNA strands are cleaved by the same catalytic mechanism in a single active site in Cpf1/Cas12a enzymes.

[0307] Another type V CRISPR is AsCpf1 from *Acidaminococcus sp BV3L6* (Yamano et al., Crystal structure of Cpf1 in complex with guide RNA and target DNA, Cell 165, 949-962, May 5, 2016)

[0308] In certain embodiments, the nuclease comprises a Cas12f effector. Small CRISPR-associated effector proteins belonging to the type V-F subtype have been identified through the mining of sequence databases and members classified into Cas12f1 (Cas14a and type V-U3), Cas12f2 (Cas14b) and Cas12f3 (Cas14c, type V-U2 and U4). (See, e.g., Karvelis et al., PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Research*, 21 May 2020, 48(9), 5016-23 doi.org/10.1093/nar/gkaa208). Xu et al. described development of a 529 amino acid Cas12f-based system for mammalian genome engineering through multiple rounds of iterative protein engineering and screening. (Xu, X. et al., Engineered Miniature CRISPR-Cas System for Mammalian Genome Regulation and Editing. *Molecular Cell*, October 21, 2021, 81(20): 4333-45, doi.org/10.1016/j.molcel.2021.08.008).

[0309] Exemplary CRISPR-Cas proteins and enzymes used in the prime editors herein include the following without limitation.

Table 5 - Cas12a orthologs						
KKP36646_(modified) hypothetical protein UR27_C0015G0 004 [Candidatus Peregrinibacteria bacterium GW2011_GWA 2_33_10]	MSNFFKNFTN	LYELSKTLRF	ELKPVGDTLT	NMKDHLEYDE	KLQTFLLKDN	(SEQ ID NO: 116)
	IDDAYQALKP	QFDEIHEEFI	TDSLESKKAK	EIDFSEYLDL	FQEKKEKELNDS	
	EKKLRNKIGE	TFNKAGEKWK	KEKYPQYEWK	KGSKIANGAD	ILSCQDMLQF	
	IKYKNPEDEK	IKNYIDDTLK	GFFTYFGGFN	QNRANYYETK	KEASTAVATR	
	IVHENLPKFC	DNVIQFKHII	KRKKDGTVEK	TERKTEYLNA	YQYLKNNNNKI	
	TQIKDAETEK	MIESTPIAEK	IFDVYFSSC	LSQKQIEEYN	RIIGHYNLLI	
	NLYNQAKRSE	GKHLSENEKK	YKDLPKFKTL	YKQIGCGKKK	DLFYTIKCDT	
	EEEANKSRNE	GKESHVSVEEI	INKAQEAINK	YFKSNNDNEN	INTVPDFINY	
	ILTKENYEGV	YWSKAAMNTI	SDKYFANYHD	LQDRLKEAKV	FQKADKKSED	
	DIKIPEAIEL	SGLFGVLDL	ADWQTTLFKS	SILSNEDKLK	IITDSQTPSE	
	ALLKMFNDI	EKNMESFLKE	TNDIITLKKY	KGKKEGTEKI	KQWFDYTLAI	
	NRMLKYFLVK	ENKIKGNSLD	TNISEALKTL	IYSDDAEWFK	WYDALRNYLT	
	QKPQDEAKEN	KLKLNFDNPS	LAGGWDVNKE	CSNFCVILKD	KNEKKYLAIM	
	KKGENTLFQK	EWTEGRGNL	TKKSNPLFEI	NNCEILSKME	YDFWADVSKM	
	IPKCSTQLKA	VVNHFKQSDN	EFIFPIGYKV	TSGEKFREEC	KISKQDFELN	
	NKVFKNELNS	VTAMRYDLSS	TQEKQYIKAF	QKEYWELLFK	QEKRDTKLTN	

Table 5 - Cas12a orthologs						
	NEIFNEWINF YSFKESSEYN IKNQDSNDGK DKLGIVKGGK TKFYNFNSLH RSIKKEKYFY IKEAKNGYVS QKFELALAKK IMLYTRANYT ETFTDIGFDG GEREKDKYEW GTGESLRFAI TKGEKISMP LHLNNKTEW	CNKKYSELLS SLDEFYRDVD SIGHKNNLHT TKNGTEIIKN FLGIDRGEKH NKQEDKWEAK LVIRKIADLA LNFLVDKNAK SQTDPATGWR KDYFFEYDKG KIDKIDIVKI NLIQQIRNTG SGDANGAFNI KQLNIFSSRK	WERKYKDAL ICSYKLNINT IYWNAIFENF YRFSKEKFI LAYYSLVNKN EVDCWNYNDL VNNERPFI RDEIGSP KTIYKAGPE EFVDEKTGEI LDDLFVNFDK NNERDND ARKGIIMNAH AMAKRKK	NWINFCKYFL TINKSILDR DNRPKLNGEA HVPITLNFCS GEIVDQGT LDAMASNRDM LEDLNTGF LQLTPPVNNY ETTYKKDGKI KPKKWRLYS NISLLKQLKE SPVRDENGKH ILANSDSKDL	SKYPKTTLFN VEEGKLYLFE EIFYRKAI NNEYVNDIVN LPFTDKDGNQ ARKNWQRI SRQKIDKSVY GDIENKKQAG KNKSVKDQII ENGKSLDRFR GVELTRNNEH FDSREYWDKE SLFVSDEEWD	
KKR91555_(modified) hypothetical protein UU43_C0004G0003 [Parcubacteria (Falkowbacteria) bacterium GW2011_GWA 2_41_14]	MLFFMSTDIT DEDFSVLRKY NLENYDRLLG VFESNGSGIL DGKEETKSFF NKKLYEKYKN LGGESNKQER YPTFTCLQKQ EFLNHEEND VSFDKIKTHL TVTLGGKKES LAYSQAVLNI FFYFDKFGNY RNQYYLGIGK DRFEGIAFPK KARRQKEKDW DRSDFVRHIQ KSEDKNHTA PKTDDVKLKI LNFGKHVNL ENGKCGSLNR MTRIKDLKKA RERQVYQGFE GTLVYVPASY DQKNDRFYFE FVDGTGDSIT NLMQIRNYD YNIARKGFMQ EEN	NKPREKGVFD GVIEEDKRIA ETEKNNKNEN SSKNCLDELT AMDVGYLDDF LDIDFSEIER TANIHGLNQI ILSQVFRKEI IDLAMVYLPK ENNKLTYKIF ITSLDEKKNR TKRAEIFWLN LKRRSRNTTK KDGEIFHKKL KTKTEEAFRQ DSKFSKEKMS RQAYWIDPRK NLFTQYLLLE YQKKGKDVY DFNKLNVNTE VGEQIITLED YLGNVVSWIS KALVNKLGYL TSKICPHPKC YQWEQEHKSD NKLKSLKGGK KEKSGSEADY LCRIRENPQE	NFTNLYEFSK ESIKIAKFYL ISEDKKKEIR KRFTQRQEV HKNRQKLYSV NFNLTLEKVF INLYIQKKQS IESDRDLIR SKINSFVYKV FKTLIKENHD LKEKLGWFEG EKQDAKVGED EIKLHFGNDD GNSVEAVKEA VCKKRADEFL KLIEYYITCL VSKDYVDKKV FSCENIKNIK ADKRDGNKEK FAKVPVKILG DKKVKEPVDY KEMLSGIKEG VDKKYDNYRN GWRERLYMKN GKKKKYSVD GIELDNINQQ IQCPSCFLDS PMKLITNREW	TLTFGLIPLK NILHRELIGK KNFKKELSLA LRRENKDIGV KGKKNLGR DFDNYNERLT EQKAEQKETG ELKFFVEESK FKEPQDFLSV FESFLILLQ KVRENEKMKD NKDMIFYKFF LLEGWDMNKE AEGEMFLFKV GGDTYEFKIL GKRDDWKRFN AEGEMFLFKV SKDLIESIFE EVIQHRRFAK MDRGENNLIY FQLLDVREGQ VVTIGVLEDL VYQFAPIVDS SASKEKIVGL KVFSNVSRMR IVNQKQELGV RKPEMNGKLS DEAVREWDIY	WDDNKKMIVE VLGSLKFEKK QDILLKKVGE EYDPVAYREK ILDNFEIFCK QEGLEDEYAKI KKKIKFNKKD EKVDKARGII FQDGASNLD EIDLLIDGGE EEEGEFCSTV DEFADDGFAP PEYWSFILRD EKIDYKQLNI LAIKKEYDDF LNRQPKYEYE HNKDFYDFER LDGKAEIRFR DALTLHLKIR YCFLEDEHGEI RDWEQKNWQK NSNFKRTRFF VEEMEKNKQI LKSDBGIKISY WDVEQKKSID EFFQSIIIFYF AITNGDANGA SAAQKIPVLS	(SEQ ID NO: 117)
KDN25524_(modified) hypothetical protein MBO_03467 [Moraxella bovoculi 237] > WP_052585281.1 type V CRISPR-associated protein Cpf1 [Moraxella bovoculi]	MLFQDFTHLY VILDDYHRDF AVLRKEIVKP SPKLAHLAHF DNLQILTTIK AYNTLLGGIS SDGMSVSFLP GIYVEHKNLN AKAKLTKEKD LAGVDNPIQK KELLDNALNV VRDYLKQKPF LLDKAHKKVF PSKELVEIKD FKKDKKGREV VDQYNIYKKI	PLSKTVRFEL IADMMGEVKL IGNGGKYKAG EKFSTYFTGF QKHSALYDQI GEAGSPKIQG SKFADDSEMC ELSKQAFGDF KFIKGVHSLA IHNHSTIKG AHFAKLLTTK STEKYKLNFG DNAPNTGKSI KGRQSRDDE PISEKDLFDK DSNDNRKKEN	KPIDRTLEHI TKLAEFYDVY YDRLFGAKLF HDNRKNMYS INELTASGLD INELINSHHN QAVNEFYRHY ALLGRVLDGY SLEQAIEHYT FLERERPAGE TTLDNQDGNF NPTLLNGWDL YQKMIYKYLE LKLRYFILEC INGIFSSPKP FYNNHPKFKK	HAKNFLSQDE LKFRKNPKDD KDGKELGDLA EDKHTAIAYR VSLASHLDGY QHCHKSERIA ADVFAKVQSL YVDVNPPEFN ARHDDSVQA RALPKIKSGK YGEFGVLYDE NKEKDNFVGI VRKQFPKVVFF LKIHPKYDKK LEMEDFFIGE DLVRYYYESM	TMADMHQKVK ELQKQLKDLQ KFVIAQEGES LIHENLPRFI HKLLTQEGIT KLRLPHKQIL FDGFDDHQKD ERFAKAKTDN GKLGQYFKHG NPMTQLRQL LAKIPTLYNK LQDKGCYYLA SKEAIAINYH FEGAIGDIQL FKRYNPSQDL CKHEEWEESE	(SEQ ID NO: 118)

Table 5 - Cas12a orthologs						
	EFSKKLQDIG YLFQIYNKDF KASLDMNETT PITMNFVQVQ KGEILEQCSL IKELKSGYLS FENALIKKLN VPAWNTSKID HIDYAKFTDK KSLFARHHIN EDFILSPVAN SDDLNVKLA	CYVDVNELEFT SPKAHGKPNL IHRAGEVLEN MTIKEFNKKV NDITTASANG HVVHQISQLM HLVLKDKADD PETGFVDLLK AKNSRQIWTI EKQPNLVMDI DEGVFFNSAL IDNQTWLNEA	EIETRRLNYK HTLYFKALFS KNPDNPKKRQ NQSIIQQYDEV TQMTTPYHKI LKYNAIVVLE EIGSYKNALQ PRYENIAQSQ CSHGDKRYVY CQNNDKFEHK ADDTQPQNAD QNR	ISFCNINADY EDNLADPIYK FVYDIIKDKR NVIIGIDRGER LDKREIERLN DLNFGFKRGR LTNNFTDLKS AFFGKFDDIC DKTANQNKA SLMYLLKTL ANGAYHIALK	IDELVEQGQL LNGEAQIFYR YTQDKFMLHV HLLYLTVINS ARVGWGEIET FKVEKQIYQN IGKQTGFLFY YNADKDYFEF AKGINVNDEL ALRYSNASSD GLWLLNELKN	
KKT48220_(modified) hypothetical protein UW39_C0001G 0044 [Parcubacteria bacterium GW2011_GWC2 _44_17]	MENIFDQFIG AKFYFDSLHQ LRKRDKNVAVG YQEREVDGKK NQPSLFVEEK ATRIADNFII IEHIKNENSY GEVEKEKQLI EFESYEGYI KKFISIAEIK YLFRIEREN GLGIFQMMKY NFTTKPFDE HKNHRKLFQS FKPSQEI LRI QFQDTGKYQN NKDIYNTKNG REKIT TQKNQ NKIINQRISS NGVNYRDKLI YSAIVVLEDL PGGVLNGYQL SNSASLDKIK WAIFASAPRI EIIKKEKAGD GEVIAVDEGV KGMILKKIR DH	KYSLSKTLRF KFIDAALASD IDRLQKEIND ITFSKADLQ ENPGQKRYIF FHQNTKVFRD NKIIGRINKK EKTREKTEED LKNKAINTIS NAVEELDGD GEKLLGYDSC FSLDDKDRKN DKIKLNFENG MGNAGKDNAN KKEKTFKRES IKEFTDDVQK SKNLHTLYFE CILDKGDRAY SDNEMRVNVI EREKERLKNR NMRFKQIRGG SAPFVSFEKM EAVKKFDAIG RRQKGEDGYW LQGEKELDGR DFIASPVKPF EHAKKDPEFK	ELKPVGKTED KTSELSFQNF AEDIIQKEKE KGADFLTAAG DSFDKFAGYL KYKNNHTDLG IKEYRDQKAK VLIERFKEFI RRWFVSDRDF FKAVFYDKKI LKI AKQLGIF TPGQLSTNFY ALLKGWDENK RYQKMIYKQI KNFSLRDLHA YGYKISFRDI HILSAENLND KYRRYTEKKI GIDRGEKNLL QSWKPVVKIK IERSVYQQFE RKQTGILFYT WDGKEQSYFF KYDRVKVNEE LRNFWHSFIY FTTPNPYIPS KLPNLFISNA	FLKINKVFEEK ADVLEKQNKI KIYKDVRTL ILKVLKYEF TKFQQTKKNL FDEENIFEIE DTKLTKSDFP ENNEERFTAA ELKLPQQKSK IAQGGSKLEQ PQEKEAREKA AEYDGYKDF EYDFMGVILK ADASKDVPR LIEYYRNCIP DDEYINQALN PVFKLSGMAE MFHMSLVNLT YYSVVKQNGE DLKKGYISHV KALIDKLGYL QA EYTSKTD KYNPNLADE FEKLLKVWNF LFNLVLELRN NLCWLAVENA EWDEAARDWG	DQTIDDSYNQ ILDKKREMG DNEAESWKTY EEKEKEFQAK YAADGTSTAV RYKNCLLQRE FFKNLQKQIL KKLMNAFCNG NKSEKNPKV FLVIWKYEF TAVIKNYADA EFIKYYNEFR KEGLYLGIM LLTSKKAMEK QYSNWSFYDF EGKMYLFEV IFQRQPSVNE GKGEIKQVQF IEEQASLNEI IHKICQLIEK VFKDNRDLRA ITGFRKNVYI KYKNSTVSKE VNPKATDIKQ SFSLQIKIKA DANGAYNIAR KYAGTTALNL	(SEQ ID NO: 119)
WP_031492824_ (modified) hypothetical protein [Succinivibrio dextrinosolvens]	MSSLTKFTNK KIIVDNFLRD VSKFETFDLF FKLVLP SYLK AYRIVHDNFP NYFTVGAYDY SELKSKLKNR EQCKDNNVIF LRNDIEDSAN TKFSDLLSCT YNTLLIFNCK DKFKLNFNSP TQAIADNTDN SDKEKFASPL NEWIAFCKEF CPIKTSFIEN LNNPTIMLNG RNEIYQYENK PLTINYKEGD	YSKQLTIKNE FINKALNNTQ SSYSIKKDEK TTNQDKLKII KFLDNIRCFN FLSQNGIDFY HAFKMAVLFK NLLNLIK NIA SKQGNKELAK LHKVASEKLV SFNKNGNFYV QLGEGFSKSK CIFKMNYFLL VIKKSTFLLA LKT YKAATIF LIDNGDLYLF GAELFYRKES FIDTLSDEAK TKQFNNEVLS	LIPVGKTLEN IGNWRELADA IIDDNDVEE SSFDNFSTYF VWQTECPQLI NNIIGGLPAF QILSDREKSF FLSDDEL DGI KIKTNKGDVE KVNEGDPKH DYDRCINELS ENDCLTLLFK KDAKKFIPKC TAHVKGKKGN DITTLKKAEE RINNKFSSSK IEQKNRITHK KVLPNVIKKE FLRGNP DINI	IKENGLIDGD LNKEDEDNIE EELDLGK KTS RGFFENRKN VKADNYL KSK AGHEKIQGLN VIDEFESDAQ FIEGKYLSSV KAISKYEFSL LKNNEEKQKI SVVYLYNKTR KDDNYVVGII SIQLKEVKAH IKKFQKEYSK YADIVEFYKD STGTKNLHTL AGSILVNKVC ATHDITKDKR IGIDRGERNL	EQLNENYQKA KLQDKIRGII SFKYIFKKNL FTKKPISTSI NVIADKSLA EFINQECQKD VIDAVKNFYA SQKLYSDWSK SELNSIVHDN KEPLDALLEI NYCTKKPYNT RKGAKINFDD FKKSEDDYIL ENPTEYRNSL VDNL CYKLEF YLQAI FDERN KDGTS LDDKI FTSDKFFFHC IYVTVINQKG	(SEQ ID NO: 120)

Table 5 - Cas12a orthologs						
	EILDSVSFNT LKEGYLSAIV KMLINKLNYF PAAATSKIDP SFDLDSLSKK FEMKKLLNEY VTNGKEDVLI ERNNLVREEK	VTNKSSKIEQ HEICLLMIKH VSKKESDWNK TTGFANVLNL GFSSFVKFSK KVSFDLENNL SPVKNAKGEF DTKKIMAI SN	TVDYEEKLAV NAIVVLENLN PSGLLNLQQL SKVRNVDAIK SKWNVYTFGE IPNLTSANLK FVSGTHNKTL VDWFEYVQKR	REKERIEAKR AGFKRIRGGL SDQFESFEKL SFFSNFNEIS RIIKPKNKQG DTFWKELFFI PQDCDANGAY RGVL	SWDSISKIAT SEKSVYQKFE GIQSGFIFYV YSKKEALFKF YREDKRINLT FKTTLQLRNS HIALKGLMIL	
KKT50231_(modified) hypothetical protein UW40_C0007G 0006 [Parcubacteria bacterium GW2011_GWF2 _44_17]	MKPVGKTEDE TSELSFQNF EDIIQKEKEK GADFLTAAGI SFDKFAGYLT YKNNHTDLGF KEYRDQKAKD LIERFKEFIE RWFVSDRDFE KAVFYDKKII KIAKQLGIFP PGQLSTNFYA LLKGWDENKE YQKMIYKQIA NFSLRDLHAL GYKISFRDID ILSAENLNDP YRRYTEKKIM IDRGEKNLLY SWKPVVKIKD ERSVYQQFEK KQTGILFYTQ DGKEQSYFFK YDRVKVNEEF RNFVHSHFIYL TTPNPYIPSN LPNLFISNAE	LKINKVFEKD DVLEKQNKII IYKDVRTLFD LKVLYEFPE KFQQTCKNLY DEENIFEIER TKLTKSDFPF NNEERFTAAC LKL PQQKSKN AQGGSKLEQF QEKEAREKAT EYDGYKDFE YDFMGVILKK DASKDVPRLL IEYYRNCIPQ DEYINQALNE VFKLSGMAEI FHMSLVNLNTG YSVVKQNGEI LKKGYISHVI ALIDKLGILV AEYTSKTDPI YNPYNLADEK EKLKLVWNFV FNLVLELRNS LCWLAVENAD WDEAARDWKG	QTIDDSYNQA LDKKREMGAL NEAESWKTY EKEKEFQAKN AADGTSTAVA YKNCLLQREI FKNLDKQILG KLMNAFCNGE KSEKNEPKVK LVIWKYEFY AVIKNYADAG FIKYYNEFRN EGRLYLGIMH LTSKKAMEKF YSNWSFYDFQ GKMYLFEVVN FQRQPSVNER KGEIKQVQFN IEQASLNEIN HKICQLIEKY FKDNRDLRAP TGFRKNVYIS YKNSTVSKEW NPKATDIKQE FSLQIKIKAG ANGAYNIARK YAGTTALNLD	KFYFDSLHQA RKRDKNAVGI QEREVDGKKI QPSLFVEEKE TRIADNFIIF EHIKNENSYN EVEKEKQLIE FESEYEGIYL KFISIAEIKN LFRDIERENG LGIFQMMKYF FITKKPFDED KNHRKLFQSM KPSQEILRIK FQDTGKYQNI KDIYNTKNKS EKITTQKNQC KIINQRISSE GVNYRDKLIE SAIVVLEDLN GGVLNGYQLS NSASLDKIKE AIFASAPRIR IIEKEKAGDL EVIADDEGVD GVMILKKIRE H	FIDAALASDK DRLQKEINDA TFSKADLKQK NPGQKRYIFD HQNTKVFRDK KIIGRINKKI KTREKTEEDV KNKAINITISR AVEELDGDIF EKL LGYDSC SLDDKDRKNT KIKLNFENGA GNAKGDNANR KEKTFKRESK KEFTDDVQKY KNLHTLYFEH ILDKGDRAYK DNEMRVNVIG REKERLKNRQ MRFQIRGGI APFVSFEKMR AVKKFDAIGW RQKGEDGYWK QGEKELDGR FIASPVKPPF HAKKDPFKK	(SEQ ID NO: 121)
WP_004356401_(modified) hypothetical protein [Prevotella disiens]	MKVMENYQEF VRADNVSYVK DCKSDEEEVK VFKADENVQH IKNIYIFEKL KGIDNYNAVI LSDREALSWL YNLNGIFIRN RAKIKKETKQ NSGNMPRKVE DIFKKDENSEK EKFEELTLLY GTQYGGYLFR ANTIYGSAYE NISDDDKVTP PLKNKAFLD DKDKPLCLFQ LDLGSGAIFY RYMENKFLFH NLLYYSVIDM EAVEGIDKDL AVYQQFEKSL FLFYVPAWNT FEFETNYDKF	TNLFQLNKTL KEIDKHKHIF KTALRNKCTS FSEFTSYFSG KEQFDAKTL GKIVKEDKQE PDMFKNDSEV NEALSSLSQN GRKSFEKYEE DYFSLMRKGD LIKELLDATK NKVRNRLTQK KKNEIGEYDY GENSYKEDKK SSLLEKIKKV LINKDYQIFT ISNKDLSFAK RAKSLDGNKP LSIVQNYKAA KGNIVEQDSL KGYLSQAVHQ SKIDPVTGFT KIRMKSAQTR	RFELKPIGKT IEETLSSFSI IQRAMREAIS FETNRENFYS EIFENYKLYV IQGLNEHINL IKALKGFYIE VYRNFSIDEA YIDKKVKAID FGSNDLIENI QFQHFIPKLL PYSKDKIRLC FLGISSKAQL RLNKVIAIYI SIDSYNGILS EVQAVIDEIC TFSANLRKKR THPANEAIKC NDSAQLNSSA NII RNNDLET IAQLMLKYNA KRPYNELGGI DLLRPKAMTI WTICTFGNRI	CELLEEGKIF SNDLLKQYFD QAFKSPQKK DEEKSTSIAY AGSSLEDEVFS YNQKHKDRRL DGFENNVLT SLSIQEINEL KTKLSAAEKL GTGEEADRDL FNKPKLMTGW FRKNEAVIGD EQIKQTNIKK FKSFQSVNKE KQKTFIYFPI GAENLHTMLF RNVANKDKVS TEYIRKADDL DYHDLDDKRE IIALEDLGQM LKAYQLASSI KEAQDFFGAF KRKKDKNYWN	ASGSFLEKDK CYNELKAFKK LLAIKNLIEN RLVHDNLPIF LEYFNNTLTQ PFFISLKKQI LATLLSSLDK NNYELIANAL VENYVSEFNS LGTKYQETAK VFYGDPLPLY VDSKTEKSDN YERLDYYQPK SIIESISKYP VIDNLLKTIS SNVELEKEMG KALMEGNQDN LFTYDIYKNR HIIGIDRGER KERKANRQNW FVTRGQKIEK TKNNSDKQNG DNISYNDKGY YEEVELTEEF	(SEQ ID NO: 122)

Table 5 - Cas12a orthologs						
	KKLFKDSNID DYIISP VANA KNDKKLNLSI	YENCNLKEEI EGQFFDSRNG SSTEWLDFVR	QNKDNRKFFD DKKLPLDADA EKPYLK	DLIKLLQLTL NGAYNIARKG	QMRNSDDKGN LWNIRQIKQT	
CCB70584 (modified) Protein of unknown function [Flavobacterium branchiophilum FL-15]	MTNKFNTQYS TIDKFHKYFI NLRKEIVKSF KTFTTYFTGF QVESLQVNFR IISGFTKNDI FLPDAFTDGK SFDTQKIYLYK KANekkREIL SNCIADYFKN LKQDQKLIDN EALSLLTPLY TILKKDGNFY FFSNKNIAFY EDWKYFDFQF KLFLFQIYSK FRKASIKPKN KISEKELTQD FKATGGSYIN ESLNTITDSK VHKIATLMLE VLKDKQPQEL TGFVNYFYTK EGTQQAWTIC GDGNCIKSQI DNGTFYNsRD VDLSISNRDW	LSKTLRFELI DLALsNAKLT SDGDAKSIFA HQNRKNMYSV ELMGEFGDEG KYGKLNeyIN QVLKAIFDFY NDTHLTTISQ DKAKAVFTKQ HFVAKKENET LKFFFLDAILE NMVRNYVTQK LAIMDKKHnk NPSKELLENY SETKSYQDLS DFSPPFSKGP IILHKKKIki DLRYIDNFSI QTVLEYLQNN ISTPYHKLLD ENAIvVMEDL GGLYNALQLT YENVDKAKAF TYGERIETKR ASKDDKAFFE YEKLENPTLP LQFVQKNK	PQGKTLEFTQ HLETYLELYN ILDKKELITV EPNSTAIAYR LIFVNELEEM NYNQTKDKKD KINLLSYTIE QVFGDFSvFS DYFSIAFLQE DKTFDFIANI LLHFIKPLHL PYSTEKIKLN AFQKFPEGKE KKETHKKGDT GFYREVEHQG NMHTLYWKAL AKKHFI DKKT FNEKNKTIDI PEVKIIGLDR NKENERDLAR NFGFKRGRFK NKFESFQKMG FEKFEAIRFN QKDQNNKFVS TLlyWFKMTL KDADANGAYH	EKGLLSQDKQ KSAETKKEQK ELEKWfENNE LIHENLPKFL FQINYYNDVL RLPKLKQLYK GQEEsQNLLL TALNYWYETK VLSEYILTLd TAKYQCIQGI KSEsITEKDT FENAQLLNGW NYEKMVYKLL FNLEHCHTLI YKINFKNIDS FEEQNLQNVI KTSEIVPVQT IKDKRFTVDK GERHLVYLTL KNWGTVENIK VEKQIYQKLE KQSGFLFYVP AEKKYFEFEV TPINLTEKIE QMRNSETRTD IAKKGLMLLN	RAESYQEMKK FKDDLKKVQD QKDIYFDEKF ENAKAFEKIK SQNGITITyNS QILSDRISLS LIRQTIENTLS VNPkFETEYS HTSDIVKKHS LENADQYEDE AFYDVfENYY DANKEGDYLT PGVNMKLPKV DFFKDSL NKH EYIDGLVNEG YKLNQQAEIF IKNLNMYyQG FQFHVpITMN IDQQGNILKQ ELKEGYISQV KMLIDKLNyL AWNtSKIDPT KKYsDFNPKA DFLGKNQIVY IDYLLSPVMN KIDQADLTkK	(SEQ ID NO: 123)
WP_005398606_ (modified) hypothetical protein [Helcococcus kunzii]	MFEKLSNIVS ISDDYRQYI KVLFNILSGE DSYSEEDLEG GFYYDNiKIF YVITQDGIDV LGVLTKLRKQ SLFEQIERLY TEGSKEEKKK LLEKFkMNNI DDLDTDETFY TLADGWSESK IYNQIPEFSK IKKEKKYKKD YDFSTLKDTS FQIYNKDFSP LEKKPIIHKE IALFNKDVLK VQGMiKDGEI NSTVKVDYQN IIKYEAIvIM ERKNLEPSGI TNLFRLKSIN NNIKSKDWKI IIPLiDNLee YYNSNEIDID NEDWINFIIS	ISKTIRFKLI KEKLSDLNLD LDKSGEKNSK LNLYSKFTTR NKNIEYLENK YQAIRGGFTK ILEYSESTSF NALKSikKEE YKEYLELDET ERSEFEAPIY PLLNEVISDY ISDNRSiILR SIPNYPFTKK FYKDNNTNKN LYVNLNDFYA ESKGKKNLHT GSILLNKIDK YKEARFDiIK KHIIIGIDRGE KLRTREEDRD ENLNQGFKRG LNPIQACYPV SSKYEEFIKK STRGERISYN KAKRKILKAL KTNLPNNGDA	PVGKTLENIE WQKLYDAHEL DLIKNNKALY LKNFWETRKN IPNLENELKE ENGEKVQGIN LIDQIEDDND VYIDARNTQK SKAKRYLNIR GSPiKLEAIK YIIPLYNLTR KGGYYYLGIL VKEHFkNNVS YLYTIYKWIE DVNSCAYRVL LYWLSMFSEE EGNTIPENIY DRRYSESQFF RHLLYYSVID RARKNWTNIN RFKVERQVYQ FKNIYFDNEE SKKKEYFYVQ FDTFKYSVQL NGAFNIARKG	KLgKLEKDFE LDSSKKESQK GKLfKKQFiL VFTDKDIVTA ADILDDNRSV EILNLTQQQL LVDRINKFNV FSQMLFGQWD EIEELVNlVE EYLEKHLEeY NYLTRKHSDK IDNKLLINKK DFQLIDGYVS FCKQFLYKYK FNKIDENTID NLRTRKLKLN HECYRYLNKK FHVPItFNWD LEGNIvEQGS KIKELKDGYL KFELALMNKL EDFKFiFNyK PTEFLINKLK RNYDFENDYI LLLKDRIvNS	RSDFYPILKN NLEMIQAQYR EVLpDFVNNN IPFRAVNENF KDYFTPNGFN RRKPETKNVK SFFESTEVSP VIRRGYTVKI GFEEVDVfSV HKWKLlLiGN DKIKVNFDfP NKSKKIYEIL PLIITKEiYD GPNKESYKEM NAVEDGKLLL GQAEiFYRKK IGREDLSDEA IKTNKNVNQi LNTLEQNRFd SHVVHKLsRL SALSfKEKYD TSVIDPVTGF DFAKANLVIL ELNIDYENID ISPTADDNGN NESKVDLKIK	(SEQ ID NO: 124)

Table 5 - Cas12a orthologs						
WP_021736722_ (modified) CRISPR-associated protein Cpf1, subtype PREFRAN [Acidaminococcus sp. BV3L6]	MTQFEGFTNL KPIIDRIYKT TYRNAIHDF TTEHENALLR FKENCHIFTR TQTQIDLYNQ RFIPLFKQIL ALFNELENSID ITKSAKEKVQ DQPLPTTLKK TGKLEMEPS NNGAILFVKN AAKMIPKCST EPKKEQTAYA SSQYKDLGEY AKGHHGKPNL RLGEKMLNKK TKEVSHEIIK ETPIIGIDRG RVAARQAWSV SKRTGIAEKA SFAKMGTSQG FDFLHYDVKT GTPPIAGKRI PKLLENDDSH SRFQNPPEWPM YIQELRN	YQVSKTLRFE YADQCLQLVQ IGRTDNLDTA SFDKFTTYFS LITAVPSLRE LLGGISREAG SDRNTLSFIL LTHIFISHKK RSLKHEDINL QEEKEILKSQ LSFYNKARNY GLYYLGIMPK QLKAVTAHFQ KKTGDQKGYR YAEINPLLYH HTLYWTGLFS LKDQKTPIPD DRRFTSDKFF ERNLIYITVI VGTIKDLKQG VYQQFEKMLI FLFYVPAPYT GDFILHFKMN VPVIENHRFT AIDTMVALIR DADANGAYHI	LIPQGKTLKH LDWENLSAAI INKRHAIEYK GFYENRKNVF HFENVKKAIG TEKIKGLNEV EEFKSDDEVI LETISSALCD QEIISAAGKE LDSLLGLYHL ATKKPYSVKE QKGRYKALSF THTTPILLSN EALCKWIDFT ISFQRIAEKE PENLAKTSIK TLYQELYDYV FHVPIITLNYQ DSTGKILEQR YLSQVIHEIV DKLNCVLKLD SKIDPLTGFV RNLSTFQRLP GRYRDLYPAN SVLQMRNSNA ALKGQLLLNH	IQEQGFIEED DSYRKEKTEE GLFKAELFNG SAEDISTAIP IFVSTSIIEV LNLAIQKNDE QSFCYKKTLL HWDTLRNALY LSEAFKQKTS LDWFAVDESN FKLNFQMPTL EPTEKTSEGF NFIEPLEITK RDFLSKYTKT IMDAVETGKL LNGQAEFYR NHRLSHDLSD AANSPSKFNQ SLNTIQQFDY DLMIHYQAVV YPAEKVGGVL DPPFVWTKIKN GFMPAWDIVF ELIALLEEKG ATGEDYINSP LKESKDLKLQ	KARNDHYKEL TRNALIEEQQA KVLKQLGTVT HRIVQDNFPPK FSFPFYNQLL TAHIIASLPH RNNVLETAE ERRISELTGK EILSHAHAAAL EVDPEFSARL ASGWDVNKEK DKMYDYDFPD EIYDLNNPEK TSIDLSSLRP YLFQIYNKDF PKSRMKRMAH EARALLPNVI RVNAYLKEHP QKKLDNREKE VLENLNFQFK NPYQLTDQFT HESRKHFLG EKNETQFQDAK IVFRDGSNIL VRDLNGVCDF NGISNQDWLA	(SEQ ID NO: 125)
WP_004339290_ (modified) hypothetical protein [Francisella tularensis]	MSIYQEFVVK KQIIDKYHQF AKDTIKKQIS ELFKANSDIT YRIVDDNLPK SEVNQRVFSL NEYINLYSQQ TMQSFYEQIA DLSQQVFDDY LSLETIKLAL QISIKYQNGQ KANILDKDEH ENSTLASGWD GEGYKKIVYK KHLPENIYRI SNEYSDFNDF SAYSKGRPNL PAKETIANKN NKFNDINLL IGNDRMKTNY KLVEYNAIV EFDKTGGVLR QLYPKYESVS IASFGSRLIN AICGESDKKF RQAPKNMPQD FVQNRNN	YLSKTLRFE FIEEILSSVC KYINDSEKFK DIDEALEIHK FLENKAKYES DEVFEIANFN INDKTLKKYK AFKTVEEKSI SVIGTAVLEY EEFNKHRDID KKDLLQASAE FYLVFEECYF KNKESANTAI QIADASKDIQ KETKSYLKNE TNHIGSQGYK HTLYWKALFD KDNPKKESVF LKEKANDVHI HDKLAAIEKD VFEDLNFQFK AYQLTAPFET KSQEFFSKFD FRNSDKNHNW FAKLTSVLNT ADANGAYHIG	LIPQGKTLEN ISEDLLQNY NLFNQNLIDA SFKGWTTFYK LKDKAPEAIN NYLNQSGITK MSVLFKQILS KETLSLLFDD ITQQVAPKNL KQCRFEEILS EDVKAIKDLL ELANIVPLYN LFIKDDKYLL NLMIIDGKTV ARFSRKDLYD LTFENISQDY ERNLQDVVYK EYDLIKDKRF LSIDRGERHL RDSARKDWKK RGRFKVEKQV FKKMGKQTGI KICYNLDKGY DTREVYPTKE ILQMRNSKTG LKGLMLLDRI	IKARGLILDD DVYFKLKKSD KKGQESDLIL GFHENRKNVY YEQIKKDLAE FNTIIGGKFV DTESKSFVID LKAQKLDLSK DNPSKKEQDL NFAAIPMIFD DQTNNLLHRL KIRNYITQKP GIMDKKHNNKI CKKGRKDRNG FIDYYKDRLD INSLVNEGKL LNGEAELEFYR TEDKFFFHCP AYYTLVDGKG INNIEKMEKEG YQKLEKMLIE IYYVPAGFTS FEFSFDYKNF LEKLLKDYSI TELDYLISPV KNNQEGKKLN	EKRAKDYKKA DDNLQKDFKS WLKQSKDNGI SSNDIPTSII ELTFDIDYKT NGENTKRKGI KLEDSDSVVT IYFKNDKSLT IAKKEKAKY EIAQNKDNLA KIFHISQSED YSDEKFKLNF FSDKAIEENK VNRQLLSLKR YYDFEFELKP YLFQIYSKDF KQSIPKKITH ITINFKSSGA NIIKQDNFNI YLSQVVHEIA KLNLYLVFKDN KICPVTGFVN GDKAAAGKWT EYGHGECIKA ADVNGNFFDS LVIKNEEYFE	(SEQ ID NO: 126)
WP_022501477 type V CRISPR-associated protein Cpf1	MNKAADNYTG YRAQQSLELK NRDKLDLVEK DKEKAETLA FYDNLKTFMR	GNYDEFIALS KIADEYYRNY SKRGEIAKML LFBGFTTYFK IQEKAGDEIA	KVQKTLRNEL ITHKLNLDINN SADDNFKSMF GYFKTRKNMF LIEEELTEKL	KPTPFPTAEHI LDFYNLFDAL EAKLITKLLP SGEGGASSIC DGWRLEHIFS	KQRGIISEDE EEKYKNDKDK DYVERNVTGE HRIVNVNASI RDYYNEVLAQ	(SEQ ID NO: 127)

Table 5 - Cas12a orthologs						
[Eubacterium sp. CAG:76]	KGIDYYNQIC PPMYQNDEEV YYTNVSSYVY VKELDSIVAE SLYTIDYNSD AFYAELEDIC SRSKEFDNNA VYNLLPGPNK NYCHDLIDYY SWAYISEEDI DNLKNICIEL SDEDYMKVYN DKYFLYLPIT VIDNKGNIIE SGYLSGVISK ISKLYLVFK TSKIDPATGF YEKIGHRELF RSEVLNLTEN LKYTVQLRNS PKDADANGAY NRRFE	GDINKHMNLY YASFNEFISR GGWGVIDSAI YEPDYFNAPY DSLIENKESA CELENVVTLY IILLRNNKYY MLPKVFIKSD KACINKHPEW NKLDEEGKIY NGEAELFYRK YNNNNYVIDT INYGIEDENV QKSFNLVNNY IARMVIDYNA ERKADENGGI INIFDFKKYS AFSFDYNNFK LKLMEQYNI KSEAEDENYD CIALKGLYEI	CQONKFKANI LEEVKLTDR ERYLYNTIAG IDDDDNAVKA LRIKKQLDDI TAIFNVNNKP TGKRDYNPSS KNYGFKFKET LFEIYNKDLS SSMKSNI THK ENDKNLIDII NSKIIEYIAK DYKNKLKNME IIVMEDLNKG LRGYQLTYIP GSGINAKVKD TYNVSSPVNE EYKDGHDIRE RLVSPILNSS NKIKQNSWSD	FKMMKIQKQI INILQNINIY KGQSKVKKIE FGGQGVLYGF MSLYHWLQTF PYSTQKFKLN DKQIIKGSEE YILEGYEKNR NQYNDIGQFY AHSTGRDNLH KDTILVNKTY EKIGHRKS QDNMNVIGID KTRDNARKNW FKRGRFKVER KSIKNVGKQC KKEFLMSMNS WTAYTYGERI DISHMDETRN NGFYDSSDYM KKFKENELYI	MGISEKAFEI NTAKIYINAR NAKKDNKFMS NKMSSELLADV IIDEVVEKDN FASPTLAAGW QRLSTDYKKM HIKSSGNFDI KDVEKQGYSI TMYLKNIFSE INETGVRVSL DIVKDKRYTE RGERNLIYIS QEIGKIKDVK QVYQKFENML GCIFYVPAAY IRYINECSEE KKLYKDGRWL ADFTCSLFEE ENENNTTHTM NVTWELDYIQ	
WP_014550095 type V CRISPR- associated protein Cpf1 [Francisella tularensis]	MSIYQEFVNK KQIIDKYHQF AKDTIKKQIS ELFKANS YRIVDDNLPK SEVNQRVFSL NEYINLYSQQ TMQSFYEQIA DLSQVVFDDY LSLETIKLAL QISIKYQNG KANILDKDEH ENSTLANGWD GEGYKKIVYK GNPQKGYEKF DEFYREVENQ PNLHTLYWKA NKNKDNPKKE NLLKEKAND TNYHDKLAAI AIVVFEDLNF VLRAYQLTAP SVSKSQEFFS LINFNRSDKN KKFFAKLTSI PQDADANGAY	YSLSKTLRFE FIEEILSSVC EYIKDSEKFK DIDEALEIIK FLENKAKYES DEVFEIANFN INDKTLKKYK AFKTVEEKSI SVIGTAVLEY EEFNKHRDID KKDLLQASAE FYLVFEECYF KNKEPDNTAI LLPGANKMLP EFNIEDCRKF GYKLTFFENIS LFDERNLQDV SFFEYDLIKD VHILSIDRGE EKDRDSARKD GFKRGRFKVE FETFKKMGKQ KFDKICYNLD HNWDTREVP LNTILQMRNS HIGLKLMLL	LIPQGKTLEN ISEDLLQNY NLFNQNLIDA SFKGWTYFYK LKDCAPEAIN NYLNQSGITK MSVLFKQILS KETLSLLFDD ITQQVAPKNL QOCRFEELI DDVKAIKDLL ELANIVPLYN LFIKDDKYLL KVFFSAKSIK IDFYKESISK ESYIDSVVNQ VYKLNGEAE KRFTEDKFFF RHLAYYTLVD WKKINNIKEM KQVYQKLEKM TGIIYYVPAG KGYFEFSFDY TKELEKLLKD KTGTELDYL DRIKNNQEGK	IKARGLILDD DVYFKLKKSD KKGQESDLIL GFHENRKNVY YEQIKKDLAE FNTIIGGKFV DTESKSFVID LKAQKLDLSK DNPSKKEQDL NFAAIPMIFD DQTNLLHRL KIRNYITQKP GVMNKKNNKI FYNPSEDILR HPEWKDFGFR GKLYLFQIYN FYRKKSIPKK HCPITINFKS GKGNIIKQDT KEGYLSQVVH LIEKLNLYLVF FTSKICPVTG KNFGDKAAKG YSIEYGHGEC SPVADVNGNF KLNLVIKNEE	EKRAKDYKKA DDNLQKDFKS WLKQSKDNGI SSNDIPTSII ELTFDIDYKT NGENTKRKGI KLEDDSDVVT IYFKNKSLT IAKTEKAKY EIAQNKDNL KIFHISQSED YSDEKFKLNF FDDKAIKENK IRNHSTHTKN FSDTQRYNSI KDFSAYSKGR ITHPAKEAIA SGANKFNDEI FNIIGNDRMK EIAKLVIEHN KDNEFDKTGG FVNQLYPKYE KWTIASFGSR IKAACIGESD FDSRQAPKNM YFEFVQNRNN	(SEQ ID NO: 128)
WP_003034647 type V CRISPR- associated protein Cpf1 [Francisella tularensis]	MSIYQEFVNK KQIIDKYHQF AKDTIKKQIS ELFKANS YRIVDDNLPK SEVNQRVFSL NEYINLYSQQ TMQSFYEQIA DLSQVVFDDY LSLETIKLAL QISIKYQNG KANILDKDEH ENSTLANGWD GEGYKKIVYK GNPQKGYEKF DEFYREVENQ PNLHTLYWKA NKNKDNPKKE NLLKEKAND TNYHDKLAAI AIVVFEDLNF VLRAYQLTAP SVSKSQEFFS LINFNRSDKN KKFFAKLTSI PQDADANGAY	YSLSKTLRFE FIEEILSSVC EYIKDSEKFK DIDEALEIIK FLENKAKYES DEVFEIANFN INDKTLKKYK AFKTVEEKSI SVIGTAVLEY EEFNKHRDID KKDLLQASAE FYLVFEECYF KNKEPDNTAI LLPGANKMLP EFNIEDCRKF GYKLTFFENIS LFDERNLQDV SFFEYDLIKD VHILSIDRGE EKDRDSARKD GFKRGRFKVE FETFKKMGKQ KFDKICYNLD HNWDTREVP LNTILQMRNS HIGLKLMLL	LIPQGKTLEN ISEDLLQNY NLFNQNLIDA SFKGWTYFYK LKDCAPEAIN NYLNQSGITK MSVLFKQILS KETLSLLFDD ITQQVAPKNL QOCRFEELI DDVKAIKDLL ELANIVPLYN LFIKDDKYLL KVFFSAKSIK IDFYKESISK ESYIDSVVNQ VYKLNGEAE KRFTEDKFFF RHLAYYTLVD WKKINNIKEM KQVYQKLEKM TGIIYYVPAG KGYFEFSFDY TKELEKLLKD KTGTELDYL DRIKNNQEGK	IKARGLILDD DVYFKLKKSD KKGQESDLIL GFHENRKNVY YEQIKKDLAE FNTIIGGKFV DTESKSFVID LKAQKLDLSK DNPSKKEQDL NFAAIPMIFD DQTNLLHRL KIRNYITQKP GVMNKKNNKI FYNPSEDILR HPEWKDFGFR GKLYLFQIYN FYRKKSIPKK HCPITINFKS GKGNIIKQDT KEGYLSQVVH LIEKLNLYLVF FTSKICPVTG KNFGDKAAKG YSIEYGHGEC SPVADVNGNF KLNLVIKNEE	EKRAKDYKKA DDNLQKDFKS WLKQSKDNGI SSNDIPTSII ELTFDIDYKT NGENTKRKGI KLEDDSDVVT IYFKNKSLT IAKTEKAKY EIAQNKDNL KIFHISQSED YSDEKFKLNF FDDKAIKENK IRNHSTHTKN FSDTQRYNSI KDFSAYSKGR ITHPAKEAIA SGANKFNDEI FNIIGNDRMK EIAKLVIEHN KDNEFDKTGG FVNQLYPKYE KWTIASFGSR IKAACIGESD FDSRQAPKNM YFEFVQNRNN	(SEQ ID NO: 129)

Table 5 - Cas12a orthologs						
	KANILDKDEH ENSTLANGWD GEGYKKIVYK GNPQKGYEKF DEFYREVENQ PNLHTLYWKA NKNKDNPKKE NLLLKEKAND TNYHDKLAAI AIVVFEDLNF VLRAYQLTAP SVSKSQEFFS LINFRNSDKN KKFFAKLTSV PQDADANGAY	FYLVFEECYF KNKEPDNTAI LLPGANKMLP EFNIEDCRKF GYKLTFFENIS LFDERNLQDV SVFEYDLIKD VHILSIDRGE EKDRDSARKD GFKRGRFKVE FETFKKMGKQ KFDKICYNLD HNWDTREVYP LNTILQMRNS HIGLKGLMLL	ELANIVPLYN LFIKDDKYLL KVFFSAKSIK IDFYKESISK ESYIDSVVNQ VYKLNGEAE KRFTEDKFFF RHLAYYTLVD WKKINNIKEM QOVYQKLEKM TGIIYYVPAG KGYFEFSFDY TKELEKLLKD KTGTELDYL DRIKNNQEGK	KIRNYITQKP GVMNKKNNKI FYNPSEDILR HPEWKDFGFR GKLYLFQIYN FYRKQSIKK HCPITINFKS GKGNIIKQDT KEGYLSQVVH LIEKLNLYLVF FTSKICPVTG KNFGDKAAKG YSIEYGHGEC SPVADVNGNF KLNLVIKNEE	YSDEKFKLNF FDDKAIKENK IRNHSTHTKN FSDTQRYNSI KDFSAYSKGR ITHPAKEAIA SGANKFNDEI FNIIGNDRMK EIAKLVIENH KDNEFDKTGG FVNQLYPKYE KWTIASFGSR IKAAICGESD FDSRQAPKNN YFEFVQNRNN	
WP_003040289.1 type V CRISPR-associated protein Cpf1 [Francisella tularensis subsp. novicida U112]	MSIYQEFVNK KQIIDKYHQF AKDTIKKQIS ELFKANSDIT YRIVDDNLPK SEVNQRVFSL NEYINLYSQQ TMQSFEQIA DLSQQVFDDY LSLETIKLAL QISIKYQNGQ KANILDKDEH ENSTLANGWD GEGYKKIVYK GSPQKGYEKF DEFYREVENQ PNLHTLYWKA NKNKDNPKKE NLLLKEKAND TNYHDKLAAI AIVVFEDLNF VLRAYQLTAP SVSKSQEFFS LINFRNSDKN KKFFAKLTSV PQDADANGAY	YSLSKTLRFE FIEEILSSVC EYIKDSEKFK DIDEALEIHK FLENKAKYES DEVFEIANFN INDKTLKKYK AFKTVEEKSI SVIGTAVLEY EEFNKHRDID KKDLLQASAE FYLVSFECCYF KNKEPDNTAI LLPGANKMLP EFNIEDCRKF GYKLTFFENIS LFDERNLQDV SVFEYDLIKD VHILSIDRGE EKDRDSARKD GFKRGRFKVE FETFKKMGKQ KFDKICYNLD HNWDTREVYP LNTILQMRNS HIGLKGLMLL	LIPQGKTLEN ISEDLLQNY NLFNQNLIDA SFKGWTYTFK LKDKAPEAIN NYLNQSGITK MSVLFKQILS KETLSLLFDD ITQOIAPKNL KQCRFEEILA DDVKAIKDLL ELANIVPLYN LFIKDDKYLL KVFFSAKSIK IDFYKQSIK ESYIDSVVNQ VYKLNGEAE KRFTEDKFFF RHLAYYTLVD WKKINNIKEM QOVYQKLEKM TGIIYYVPAG KGYFEFSFDY TKELEKLLKD KTGTELDYL GRIKNNQEGK	IKARGLILDD DVYFKLKKSD KKGQESDLIL GFHENRKNVY YEQIKKDLAE FNTIIGGKFV DTESKSFVID LKAQKLDLSK DNPSKKEQEL NFAAIPMIFD DQTNLLHLKL KIRNYITQKP GVMNKKNNKI FYNPSEDILR HPEWKDFGFR GKLYLFQIYN FYRKQSIKK HCPITINFKS GKGNIIKQDT KEGYLSQVVH LIEKLNLYLVF FTSKICPVTG KNFGDKAAKG YSIEYGHGEC SPVADVNGNF KLNLVIKNEE	EKRAKDYKKA DDNLQKDFKS WLKQSKDNGI SSNDIPTSII ELTFDIDYKT NGENTKRKGI KLEDDSDVVT IYFKNDKSLT IAKKTEKAKY EIAQNKDNLA KIFHISQSED YSDEKFKLNF FDDKAIKENK IRNHSTHTKN FSDTQRYNSI KDFSAYSKGR ITHPAKEAIA SGANKFNDEI FNIIGNDRMK EIAKLVIENH KDNEFDKTGG FVNQLYPKYE KWTIASFGSR IKAAICGESD FDSRQAPKNN YFEFVQNRNN	(SEQ ID NO: 130)
KKQ38174 hypothetical protein US54_C0016G0015 [Candidatus Roizmanbacteria bacterium GW2011_GWA 2_37_7]	MKSFDSTNL KPYFDRHRE LQRLRTEIGK GEEKDTFIEV DGTSTAIATR SIDFYNKCLL IPFFKLLDKQ ADFEVNNISKY KYEKKDNSYK EQFLAIFLYE KDSKVTIKDF QFYDNAYEDI DNSAVILQKG FPDQAKMFPK IDFYKDCLTK ISDQYIHEKN FSNDNVVQNF NKIFFHVPLT	YSLSKTLKFE FIEEALTGVE IFNLKAEDWV EEDDKTGSKS IIDQNLKRFI QDGDIDYNNKI ILSEKILFLD DLAQIYISQE FPDFIALSQM FNSLFSKIN ADSVLTIIYQM VQVYNKLRNY GKYLLGLITK VCFSAGLEF YEGWACYTFR EKGELHLFEI PIKLNQAEI LNRTKNDYSR	MRPVGNTQKM LIGLDENFRT KNKYPILGLK INQISIFDSW DNLSIVESVR IGGETLKNGE EIKNDTELIE AFNTISNKWT KSALLSISLE TKDGETKQVG AKYFAVEKKR LTKKPYSEEK GHNKIFDDR FRPSEELRI HLKPTEEYQN HNKDNWLDKA FYRPKTEKDK FNAQINNFLA	LDNAGVFEKD LVDWQKDKKN NKNTDILFEE KGFTGYFKKF QKVDLAETEK KLIGLNELIN ALSQFAKTAE SETETFAKYL GHFWKEKYYK YYLFAKDLHN AWLAEYELDS WKLNFENSTL QEKFIVGIEG YNNAEFFKKE NIGEFFRDVA RDGKSKTTQK LESKKDKKGN NNKDINIIGV	KLIQKKYKGT NVAMKAYENS AVFGILKARY FETRKNFYKN SFSISLSQFF QYRQNNKDQK EKTIVKKLF FEAMKSGKLA ISKFQEKTNW LILSEQIDIP FYTQPDGTGL ANGWDKNKES GKYEKIVYKF TYSIDSMQKL EDGYRIDFQG NLHTLYFRESL KVIDHKRYSE DRGEKHLVYY	(SEQ ID NO: 131)

Table 5 - Cas12a orthologs						
	SVITQASDIL KKGYSQVVR LIDKLSFLVD SYTSKSDPVT DFQQAKEYPN TKYGIDITKD KNGKDDFILS QYSEKNENCE	ESGSLNELNG KLADLAIKHN KGEKNPEQAG GWRPHLYLKY KTVWKVCSNV LLTQISTIDE PVEPFFDSRK KMKWGDLYVS	VNYAEKLGKK AIIILEDLNM HLLKAYQLSA FSAKKAKDDI ERFRWDKNLN KQNTSFFRDF DNGNKLPENG NIDWDNFVTQ	AENREQARRD RFKQVRGGIE PFETFQKMGK AKFTKIEFVN QNKGGYTHYT IFYFNLICQI DDNGAYNIAR ANARH	WQDVQGIKDL KSIYQQLEKA QTGIIIFYTQA DRFELTYDIK NITENIQELF RNTDDSEIAK KGIVILNKIS	
WP_022097749 type V CRISPR- associated protein Cpf1 [Eubacterium eligans CAG:72]	MNGNRSIVYR KSTELKNIMD LEKEHNKMRE AGKLETALF ANMTSYKKIS IDFYNEICGV MFEDDMSVYN ETLSCFMSGN DVNDLVEKYI IESEEKADEI ENIVPLYNRV IRDNKYLLAI KVFLSKKGIE EKHAEWRKYE EEGKIYLFQI ELFYRKASVK MYNGYIKESD NYKV TARNNV QKSYNILNNY IAMLMEVEYNA GKSVDEPGGL ISAFNFKSIS GKTQWTVYTN HDVRIDMEKM IISPVINDEG IKSEWTEGDF	EFVGVTPEVAK DYYREYIDKS QICTHLQSDS NGFSTYFTDF EKALDEIEVI VNAHMNLYCQ AVNAFIDETE WNLITGCVEN DEKERNEFKN KKRLDMYMMN RNYVTQKPYT FNAKNKPKDK TFKPSDYIIS FKFSATDSYN YNKDFAENST NPVKHKKDSV LSEAAKEYLD NDMAVKYIAQ DYKKKLVEKE IIAMEDLNYG LKGQQLTYVP TNASRKQFFM GERLQSEFNN YEDKNSEFFA EFFDSDNKYE DRNCLKLPHA	TLRNELRPVG LSGLTDLDFT DYKNMFNAKL FEKRKNVFTK EKNNQDKMGD QTKNNYNLFK KGNIIIGKLD FYDENIHAKG SNAKQYIREI YHVVKA FIVD SKKIKLNFQS IIQGNSDKKN GYNAHKHIKT DISEFYREVE GKENLHTMYF LVNKT YKNQL KVEVRTAQKD NDDIHVIGID KTREYARKNW FKRGRFKVER DNIKNLGKQC QFDEIRYCAE ARRTGKTKSI QLLSLYKLTV SDDKECKMPK EWLDFIQNKR	HTQEHI IQNG LLFELMNSVQ FKEILPDFIK EAVSTSIAYR WELNQIFNPD MRKLHKQILA IVNKYDELDE KSKEEKVKKA SNIITDTETA EVLDRDEMFI PTLANGWSQS DNDYKKMVYN SENFDISFCR MQGYRIDWTY KNIFSEENLK DNGDVVRIP IVKDYRYTVD RGERNLIYIS KSIGNIKELK QVYQKFESML GVIFYVPAAF KDMFSFGFDY NLTETIKLLL QMRNSYTEAE DADANGAYCI YE	LIQEDELQRQ SSLSKDNKKA NYNQYDVKDK IVHENS LIFL FYNMVL IQSG YTSTSFEVPK KRIYISKDFY VKEDKYKSIN HLEYDEHISL SDIDDIYNIL KEFDNNAIIL LLPGANKMLP DLIDYFKNSI ISEADINKLD NIVIKLNGQA PDDIYNEIYK KYFIHTPITI VIDSHGNIVK EGYISGVVHE INKLNYFASK TSKIDPSTGF NNFDTYNITM EDNEINYADG EQEKGISYDK ALKGLYEVLK	(SEQ ID NO: 132)
WP_021739647 hypothetical protein [Eubacterium ramulus]	MIKKTIDTVL AAYVNEVVQH FHF GIGRDI SWDEFDTWMR WGIQFRHNV LIANA EACTM SEEKFTLCMT AIVFDKEGHM RYRLAPFAYS AGYEMKNCDQ LGIEAVIA	NVRPIFVGIQ ETEGVHIMDP YLEFAERYKK VLRVRKCLKN HELLDQTHPV EPDMVKKTIQ HSLNDENGIS IPFHKFNDNT GFGATIRYDF GVFFNVADKV	HLYFYE GPCR IYVERDDWFR PVGAAPGLCC TRVLLAVRWD DPTTNPSTPG AAYTVQKLLD SACEYDINSV IEDIADKTNL AQDIGQVITM DFYHKQYQFG	FGEGDELMPE SPEAMY EKMA DGIGNTAAVK SNRSYSSYDN RKALNINDED AYDCNAFTAP IGKVIMTNLS YMTFHSTPNR IRISPDATKI NHTVLAYGDY	YDAMMNQEMN EDIDKVD FYL NRGLEAYAFM FINQSDVTNK MKEIEKITDE CPDLCSTRRF GKAPYMGNTN NLKGLKAEKE FIAKGTISGG VEELKMLAEA	(SEQ ID NO: 133)
gi 800943167 WP_045971446. 1 type V CRISPR- associated protein Cpf1 [Flavobacterium sp. 316]	MKNFSNLYQV IDEFHKYFID LRKEIVSSFT FTTYFTGFHQ AELQPKIEKL RTATEGKQKI PDAFTEGQV DVNKIYLNKD NEAKQEKLDK CIADYFTTHF QDQKLIDNIK LNVVTPLYNM LKRDN YFLA SNKNIAFFNP	SKTVRFELKP LALNNKKLSY EGEAKAIFSV NRKNMYSAEA YKEFEAYLNV QGLNEIINLY LKAVFEFYKI TSLTTIAQQL EKIKFVKQDY IAKKENEADK FFLDAILEVV VRNYVTQKPY IMDKKHNTF SKEILDNYKN	IGNTLENIKN LNEYIALYTQ LDKKELITIE NSTAIAYRLI NSISELFEID NQTKPKNERL NLLSYKQDGV FGDFS VFSAA FSIAFLQEVV TFNF IANIK HFIKPLHLKS STEKIKLNFE QOFTEDDENY NTHKKGATFN	KSLLKNDSIR SAEAKKEDKF LEKWKNNENL HENLPKFIEN YFNEVL TQKG PKLKQLYKQI EESQN LLELI LQYRYETVVN ADYVKTLDEN KYQC IQGILE ESITEKDNAF NAQLLNGWDA EKIVYKLLPG LKDCHALIDF	AESYQKMKKT KADFKKVQDN AVYLDESFKS SKAFEKSSQI ITVYNNIIGG LSDRISLSFL QQVVKNLGNQ PKYTA EYQKA LDWKQKYTPS QADDYEDELK YDVFENYEEA NKEKDYLTTI VNKMLPKVFF FKDSL NKHED	(SEQ ID NO: 134)

Table 5 - Cas12a orthologs						
	WKYFDFQFSE YLFQIYNKDF KASIKKKNI VPITMNFKAT GTILLQESLN GYISQVVHKE DKLNYLVLDK SKIDPTTGFV DFNPKAADTR KVGDLDSKEL PVANEKGI FY LKKVKLAISN	TKTYQDLSGF SPYAKGKPNM THKAHQPIAA GNSYINQDVL VIQDEKTHTP TKMMIEHNAI KQPHELGGGLY NYFYTKYENV QEWITICTHGE KTQLIAQNEK DSRKATASLP REWLQYVQQV	YKEVEHQGYK HTLYWKALFE KNPLTP TAKN AYLKDNP EVN YHTLLDNKEI VVMEDLNFGF NALQLTNKFE EKAKTFFSKF RIETKRQKEQ SFFEELFHLL IDADANGAYH QKK	INFKKVSVSQ TQNLENVIYK TFAYDLIKDK IIGLDRGERH ARDKARKNWG KRGRFKVEKQ SFQKMGKQSG DSILYNKTKG NNNFVSTTIQ KLTLLQMRNSE IACKGLWIME QINKTNSDD	IDTLIEEGKM LNGQAEIFFR RYTVDFKQFH LVYLTLDQK SIESIKELKE IYQKLEKMLI FLFYVPAWNT YFEFVVKNY LTEQFVNFFE SHEIDYLLIS QINKTNSDD	
WP_044110123.1 type V CRISPR-associated protein Cpf1 [Prevotella brevis]	MKQFTNLYQL IDDYHKDYIE ASALKGDDRY ENRKNMYSDE YKNFEEYLN GINEFVNLN LNAVKEYCTS GEWSIIPNAI CISELNGIDI REKNLSQDKL QLWEVIDSIT NASIILRKNG VPKCTTQMK VKKFQIGYLN IGCYNDLNSF STYSKGT HPANKPIDNK NGDINMQVRE IVNNYKGT HKITQLMIKY DKNKDANEIG GFVNLDTTRY TRTQWTLCTY NLKEAILLQA FFDSRKVKEG EWLQFAQKKQ	SKTLRFELKP NVLNNFKLNG KTIDKKELIR AKSTAIAYRL NSIDEIFSLD QQLANKKDRL YFDNVEESVK EQRLRSDNPK ADYYATLGAI TVMQLKDLDL PLYNKVRNYC MYYLGIMKKE VKEHFSNSND NTGDSFGYNH YGAVNLLLYN HTLYWKMLFD NVNPNPKQSN YIKTTDDLHF KTDYHTLLSE NAIVLLEDLN GLLHAYQLTN ENVAKAQAFF GTRIETFRNA NKNFYTDILH LPVNADANGA YLKD	IGKTLEHINA EYLQAYFDLY QDMKTFLKDD IHDNLPKFID YNNIVLTQTQ PKLKPLFKQI VLLTGISDYD KTNEKEEKYS NTESKQEPSI DFKALQHFIK TRKPFSTEKI YRNILTKPMP DYTLFEKDKF AVEIWKSFCL LTYRKVSVDY ESNLNDVVYK FEYDLIKDKR IGIDRGERHL RDKKRKEERS MGFKRGRQKV DPKLPNKNSK KKFDSIRYNK EKNSNWDSRE LMKLTLLQMRN YNIARKGLWL	NGFIDNDAHR SQDTKDKQFK TDKALLDEFY NIAVFKKIAN IEVYNSIIGG LSDRVQLSWL LSKIYITNDL DRISKLLKLP PTSIQVHYNA PLLGNDEAE KVNFEQAQLL SDGDCYDKVV IAPVVITKEI KFLKAYKSTS IHQLVDEDKM LNGQAEVFYR YTVDKFMFHV LYICVINGKG SWQTIEGIKE ESSVYQQFEK QSGFLFYVPA EYDRFEFKFD IDLTTEWCTL SVTGTDIDYM AQQIKNANDL	AESYKKVKKL DIQDKLRKSI EFTTYFTGYH TSVADNFSTI RTLEDDTKIQ QEEFNTGADV ALTDVSQRMF KSYSLGYINE LKPILDTDYP KDEKFGELM DGWDENKES YKFFKDITTM FDLNNVLYNG IYDFSSIEKN YLFMIYNKDF KKSITYQHPT PITLNFKGMG EIVEQYSLNE LKSGLYSQVI ALIDKLNLYLV WNTSKIDPVT YSNFTAKAED FTQHNIPLNA VSPVANECGE SDVKLAITNK	(SEQ ID NO: 135)
WP_036388671.1 type V CRISPR-associated protein Cpf1 [Moraxella caprae]	MLFQDFTHLY AILEDYHRDF AVLRKEIVKP SPKLAHLAHF DNLQILATIK AYNTLLGGIS SDGMGVSLP GIYVEHKNLN AKAKLTKEKD LAGVDNPIQK KELLDNALNV VRDYLSQLPF LLDKAHKKVF PSAELLDKYA PTSSYQDLSD FSPKAHGKPN TIHRAGEVLE GMTIKEFNKK LNDITTASAN SHVVHQISQL NHLVLKDEAD	PLSKTMR FEL IADMMGEVKL IGNGGKYKAG EKFSTYFTGF QKHSALYDQI GEAGSRKIQG SKFADDSEMC ELSKQAFGDF KFIKGVHSLA IHNNHSTIKG AHFAKLLTTK STEKYKLNFG DNAPNTGKNV QGTHKKGNF FYREVEPQGY LHTLYFKALF NKNPDNPKKR VNQSIQQYDE GTQMTTPYHK MLKYNAIVVL DEIGSYKNAL	KPIGKTLEHI TKLAEFYDVY YDRLFGAKLF HDNRKNMYS INELTASGLD INELINSHHN QAVNEFYRHY ALLGRVLDGY SLEQAIEHYT FLERERPAGE TTLDNQDGNF NPTLLNGWDL YQKMIYKLLP NLKDCHALID QVKFVDINAD SKDNLNPIY QFVYDIKDK VNVIGIDRGE ILDKREIERL EDLNFGFKRG QLTNNFTDLK	HAKNFLSQDE LKFRKNPKDD KDGKELGDLA EDKHTAITYR VSLASHLDGY QHCHKSERIA ADVFAKVQSL YVDVVPNEFN ARHDDSVQA RALPKIKSGK YGEFGALYDE NKEKDNFGII GPNKMLPKVF FFKAGINKHP YINELVEQGG KLNGEAQIFY RYTQDKFMLH RHLLYLTVIN NARVGWGEIE RKFVEKQIYQ SIGKQTGFLF	TMADMYQKVK GLQKQLKDLQ KFVIAQEGES LIHENLPRFI HKLTLQEGIT KLRPLHKQIL FDGFDDHQKD ERFAKATDN GKLQGYFKHG NPMTQLRQL LAKIPTLYNK LQKDGCCYYLA FAKSNLDYYN EQHFGFKFS LYLFQIYNKD RKASLDMNET VPITMNFQVQ SKGEILEQRS TIKELKSGYL NFENALIKKL YVPAWNTSKI	(SEQ ID NO: 136)

Table 5 - Cas12a orthologs						
	DPETGFVDLL KAKNSRQIWK NDKQPNLVMD NDEGMFFNSA AIDNQTWLNF	KPRYENIAQS ICSHGDKRYV ICQNNDKEFH LADDTPQNA AQNR	QAFFGKFDDKI YDKTANQNG KSLIYLLKTL DANGAYHIAL	CYNADKDYFE ATKGINVNDE LALRYSNASS KGLWVLEQIK	FHIDYAKFTD LKSLEFARHHI DEDFILSPVA NSDDLNVKVL	
WP_020988726.1 type V CRISPR-associated protein Cpf1 [Leptospira inadaai]	MEDYSGFVNI KKIIDKYHRA KNIPDKERLE KFCETDEERK LPKFLDNLKI VLNQKGIDAY ILFKQILGDR KKFLSSFNRY DPKKKIKSPL EAYFAEFKSK VGKIKDFLDS LYNKVRNYLT YYLGVMKDEN NLSIYNPSKS FDFKFSKTSS QIYNKDFSIF INYDEKKKRE LNFNLKVNEF QSGKGRPEIN SKLMVENNAI NKPTPEGGVL DFLHPAYENI RVWVICCTNV PEILRKNDAY FFNSLEASDD NKDWLEFVWE	YSIQKTLRFE YIEEVFDSVL ALSEKLRKML QVSNFKSFTT IESIQRRFKD NTILGGKSEE ETKSFIEPAF ELDGIYLAND KYEKEKEKWL DDAKKQFDLL IKSLQFFLKP GKIYSKEKFK NTILSDIPKV ILKIREAKSF YENISEFYRE SKGKPNLHTI GHPPELFELK LKRNDKINII YKEKLQEKEL VLEDLNI GF KAYQLTDEFQ EKAKQWINKF ERYFTSKTAN FFKSLLFYIK EPKADADANGA RNR	LKPVGKTLLEH HQKKKKDKTR VGAFKGEFSE YFTGFHSNRQ FPWSDLKKNL SGEKIQGLNE PDDQSVLNSI NSLASISTFL KQKYTISFL ERIEEAYAIV LLSAEIFDEK LNFENSTLLK KPNELFYEKM KEGKNFKLKD VERQGYNLDF YFRSLFSKEN KYPILKDKRY GIDRGERNLL ERDKARKSWG KRGRQKVERQ SFEKLSKQTG DSIRFNSKMD SSIQYNSIQI TTLSLRQNNG YHIALKGLMN	IEKKGFLKGD FSTQFIKEIK EVAEKYKNLF NIYSDEKKST KKIDKNIKLT YINLYRQKNN TEFAKYLKLD FDDWSFIKKS NDAIESYSKS EPLLGAEYPR DLGFYNQLEG GW DENREVAN VYKLIPTPHM CHKFIDFYKE KKVSKFYIDS LKDVCCLKNG SEDKFQFHLF YLVMINQKGE TVENIKELKE VYQKF EKMLI FLFYVPSWNT WFEEFADTRK TEKLKELFVD KKGEEEEKDFI LLVLNETKEE	KIRAEDYKAV EFSELYYKTE SKELIRNEIE AIGYRIIHQN EYFSIDGFEV IDRKNLPNVK KKKKSIIAEL VSFKYDESVG QDEKRVKIRL DRNLKADKKE YYEEDISIGH LCVIFREDQK QLPRIIFSSD SISK NEDWSR LVEDGKLYLF EAEMFFRKKS ISLNFKSKER ILKQTL LDSM GYLSIVIHQI DKLNLVLFKE SKIDPRTGFI FSENMLGKN IPFNGQDLK LSPVVD SKGR NLSRPKW KIK	(SEQ ID NO: 137)
WP_023936172.1 type V CRISPR-associated protein Cpf1 [Porphyromonas crevioricanis]	MPWIDLKDFE RVKKIIDTYH DKALDKIRAV FVLSTEAESL IAYRLIHENL EDIFSLNYYI GREDRLPLFR ILGRTQQ LMT DHEQAPKRIT LSTLGQKEGP IKNLLDNISD KVRNYLTRKP AIMNNRHKRS IEIYEPSPKL GFKFSDTATY IYNKDFSPCS KNDHPTHPAG NFKCSAGSKV QISLNTINDI IAELMVAYKA KKRPEDIGGL VNLFHAQYEN SMWILCTHGS KTAIVDEKQK DTREGNKS LP LQFVQERSYE	NLYPVSKTLR KV FIDSSLEN LRGLIVGAFT PFSVEEATRS PKFIDN ILVF HVLSQAGIEK PLYKQILSDR SISEYDLSRI AKYERDRIKA HGLSNLVENV LQRFLKPLWG YSTRKVKLNF FENKVLPEYK LEQYGHGTHK ENVSSFYREV KGT PNLHTLY KPIKKKSRQK NDMVNAHIRE DYHDLLESRD VVALEDLNMG L RAYQFTAPF VDKAKSFFQK RIKNFRNSQK DFFVDLLKLF KDADANGAYN KD	FELKPVGKTL MAKMGIE NEI GVCGRRENTV LKEFDSFTSY QKIKEPIAKE YNALIGKI VT EQLSYLPESF YVRNDSQLTD LKGEESISLA FASYHEAEQL MGDEPDKDER GNSQLLSGWD EGEPYFEKMD KGD TFSMDDL EDQGYKLSFR WRMLFDERNL KGEESLFEYD AKDMH VIGID KDRQQERRNW FKRGRQKVES KSFKE MGKQN FDSISYNPKK NGQWDSE EFA KLT VQMRNSW IALKGLWALR	ENIEKAGILK KAMLQSFCEL QNEKYESL FK FAGFYENRKN LEHIRAD FSA EGDGEMKGLN EKDEELLRAL ISKKMLGDWN NLNSCIAFLD LSFPYPEENN FYGEYNYIRG RNKEKDN SCV YKFLPD PNKM HELIDFFKHS KVSESYVYSL ADV IYKLDGK LVKDRRYTMD RGERNLLYIC QTIEG IKELK SVYQQFEKQL GFLFYI PAWN DWFEFAFDYK LTEAFKSLFV KEKDL DY LIS QIRQTSEGGK	EDEHRAESYR YKKDHRTEGE EKLKEILPD IYSTKPQSTA GGYIKKDERL EHINLYNQQR KEYFDHIAED AIYMARERAY NVRDCRVDTY LIQDKDNVVL ALDQVIPLYN ILRKQNFYFL LPKVFLSKKG IEAHEDWKQF IDQGKLYLFQ AEIFFREKSL KFQFHV PITM VIDSRGTILD QGYLSQAVHR IDKLNLYLVDK TSNIDPTTGF NFTKKAEGSR RYEIDYTADL PVAGADGRFF LKLAI SNKEW	(SEQ ID NO: 138)

Table 5 - Cas12a orthologs						
WP_009217842.1 type V CRISPR-associated protein Cpf1 [Bacteroidetes oral taxon 274]	MRKFNEFVGL KKIIDKYHKC KTFKTIQNNL EKFKNFTTYF KVVSSPLAEK IVGGRTEEDN SVSWLPPKFD IYISNDSQLS DKKLKTIDSI SFKKNIEGAY FIKPLLKGKD KKIKLNFENS REDLPHDGEC RGTHKKGAGF FYREVEQQGY MHTLYWKMLF KNKDNQKKES RHRSATDLH YRTDYHELLD YNAVIVLEDL GGLLRAYQLT YENIEKAKVF TQGKRIQICR QNKKEFFFEEL ELKENAVLPM LKFAQTTPYL	YPVSKTLRFE LIDEALSGFT RKQIVNKLTD TEFHKNRKNN INALYEDFKE KIQIKGLNQY SDKNLLIKIK YISQKMFGRW SIGDVDECLA ESVKELLNNA EADKDGVFYF TLMGWDNLK YDKMEYKLLP DLGDCRALID KMSFRKVSVD DEENLKDVMY KFEYDLIKDR IIGIDRGERH TREGERTEAR NFGFMRSRQK GEFESFKTLG FDKFKSIRYN NHQRNNEWEG LRLRLTLQML NADANGAYNI ED	LKPIGKTLLEH FDTEADGRSN SEKYKRIDKK YSKEEKSTAI YLNVEEISRV INEYNQQQTD ECYDALSEKE DIISKAIRET QLGETYVKRV DNITDNNLMQ EFTSLWTKLD EPDNTTVIFC GANKMLPKVF FFKKSIERHD YIKSLVEEGK KLNGEAEVVF RYTVDKFLFH LLYLTVIDSR RNWQTIQNI VEKQVYQKFE KQSGILFYVP SDKDWFEFVV QEIDLTKAFK RNSMPSSDID ARKGLLAIK	IQRNKLEHD NSLSEYYLYY ELITTDLPDF AFRLINENLP FRLDYYDELL RSNRLPKLKP KVFDKLESIL CAKRNPKQSR EDYFVAMGES DKGNVEKIKT QVTPLYNMVR KDGLYYLGIM FSETGIQRFL DWKKFDFKFS LYLFQIYNKD RKSSITVQSP VPITMNFKSV GNIKEQFSLN ELKEGYLSQV KMLIDKLNLY AWNTSKIDPV DDYTRFSPKA EHFEAYGVDI YLISPVANDT MKQEENDSAK	AVRADDYVKV NLKKRNEQEQ LTNESEKELV KFVDNIAAFE TQKQIDLYNA LYKQILSDRE KSLSTYDLSK ESLEKFAERI EIDDEQTDTT LLDAIKDLQR NYLTSKPYST GKKYNRVFDV PSEELLGKYE DTSTYQDISE FSAHSGKTPN THPANSPKIN GGSNINQLVK EIVNEYNGNT IHKTSELAIK VDKKKPAVET TGFVNLFDT EGTRRDWTIC SKDLREQINT GCFFDSRKQA ISLAISNKEW	(SEQ ID NO: 139)
WP_036890108.1 type V CRISPR-associated protein Cpf1 [Porphyromonas crevioricanis]	MDSLKDFTNL KKIIDTYHKV ALDKIRAVLR LSTEAESLPF YRLIHENLPK IFSLNYYIHV EDRLPLFRPL GRTQQLMTSI EQAPKRITAK TLGQKEGPHG NLLDNISDLQ RNYLTRKPYS MNNRHKRSFE IYKPSPKLLE KFSDTATYEN NKDFSPPCSK DHPHTPAGKP KCSAGSKVND SLNTINDIDY ELMVAYKAVV RPEDIGGLLR LFHVQYENV WILCTHGSRI AIVDEKQKDF REGNKSLLPK FVQERSYEKD	YPVSKTLRFE FIDSSLENMA GLIVGAFTGV SVEEATRSK FIDNILVFQK LSQAGIEKYN YKQILSDREQ SEYDLSRIYV YERDRIKALK LSNLVENVFA RFLKPLWGMG TRKVKLNFNG NKMLPEYKEG QYGHGTHKKG VSSFYREVED TPNLHTLYWR IKKKSRRQK MVNAHIREAK HDLLESRD ALEDLNMGFK AYQFTAPFKS KAKSFFQKFD KNFRNSQKNG FVDLLKLFKL ADANGAYNIA	LKPVGKTLEN KMGIEINEIKA CGRRENTVQN EFDSFTSYFA IKEPIAKELE ALIGKIVTEG LSYLPESFEK RNSQLTDIS GEESISLANL SYHEAEQLLS DEPDKDERFY SLLSGWDRN EPYFEKMDYK DTFSMDLHE QGYKLSFRKV MLFDERNLAD EESLFEYDLV DMHVIGIDRG RQOEHRNWQT RGRQKVESV FKEMGKQNGF SISYNPKKDW QWDSEEFALT TVQMRNSWKE LKGLWALRQI	IEKAGILKED MLQSFCELYK EKYESLFKEK GFYENRKNIY HIRADFSAGG DGEMKGLNEH DEELLRALKE KKMLGDWNAI NSCIAFLDNV FPYPEENNL GEYNYIRGAL KEKDNSCVIL FLPDPNKMLP LIDFFKHSIE SESYVYSLID VIYKLDGKAE KDRRYTMDKF ERNLLYICVI IEGIKELKQG YQQFEKQLID LFYIPAWNTS FEFAFDYKNF EAFKSLFVRY KDLDYLISP RQTSEGGK	EHRASRYRRV KDHRTGEDK LIKEILPDFV STKPQSTAI YIKKDERLED INLYNQQRGR FYDHAIEDIL YMARERAYDH RDCRVDTYLS QDKDNVVLK DQVIPLYNKV RKGQNFYLA KVFLSKKKGIE AHEDWKQFGF QGKLYLFQIY IFFREKSLKN QFHVPIITMNF DSRGTILDQI YLSQAVHRIA KLNLYLVDKK NIDPTTGFVN TKKAEGSRSM EIDYTADLKT AGADGRFFDT LAISNKEWLQ	(SEQ ID NO: 140)
WP_036887416.1 type V CRISPR-associated protein Cpf1	MDSLKDFTNL KKIIDTYHKV ALDKIRAVLR LSTEAESLPF YRLIHENLPK IFSLNYYIHV EDRLPLFRPL	YPVSKTLRFE FIDSSLENMA GLIVGAFTGV SVEEATRSK FIDNILVFQK LSQAGIEKYN YKQILSDREQ	LKPVGKTLEN KMGIEINEIKA CGRRENTVQN EFDSFTSYFA IKEPIAKELE ALIGKIVTEG LSYLPESFEK	IEKAGILKED MLQSFCELYK EKYESLFKEK GFYENRKNIY HIRADFSAGG DGEMKGLNEH DEELLRALKE	EHRASRYRRV KDHRTGEDK LIKEILPDFV STKPQSTAI YIKKDERLED INLYNQQRGR FYDHAIEDIL	(SEQ ID NO: 141)

Table 5 - Cas12a orthologs						
[Porphyromonas crevioricanis]	GRTQQLMTSI	SEYDLSRIYV	RNDSQLTDIS	KKMLGDWNAI	YMARERAYDH	
	EQAPKRITAK	YERDRIKALK	GEESISLANL	NSCIAFLDNV	RDCRVDTYLS	
	TLGQKEGPHG	LSNLVENVFA	SYHEAEQLLS	FPYPEENNL	QDKDNVVLK	
	NLLDNISDLQ	RFLKPLWGMG	DEPDKDERFY	GEYNYIRGAL	DQVIPLYNKV	
	RNYLTRKPYS	TRKVKLNFGN	SQLLSGWDRN	KEKDNSCVIL	RKGQNFYLAI	
	MNNRHKRSFE	NKVLPEYKEG	EPYFEKMDYK	FLPDPNKMLP	KVFLSKKGIE	
	IYKPSPKLLE	QYGHGTHKKG	DTFSMDDLHE	LIDFFKHSIE	AHEDWKQFGF	
	KFSDTATYEN	VSSFYREVED	QGYKLSFRKV	SESYVYSLID	QGKLYLFIQIY	
	NKDFSPCSKG	TPNLHTLYWR	MLFDERNLAD	VIYKLDGKAE	IFFREKSLKN	
	DHPHTPAGKP	IKKKSQRQKG	EESLFEYDLV	KDRHYTMDKF	QFHVPIITMNF	
	KCSAGSKVND	MVNAHIREAK	DMHVIIGIDRG	ERNLLYICVI	DSRGITILDQI	
	SLNTINDIDY	HDLLESRDKD	RQQERRNWQT	IEGIKELKQG	YLSQAVHRIA	
	ELMVAYKAVV	ALEDLNMGFK	RGRQKVESSV	YQQFEKQLID	KLNYLVDKKK	
	RPEDIGGLLR	AYQFTAPFKS	FKEMGKQNGF	LFYIPAWNTS	NIDPTTGfVN	
	LFHAQYENV	KAKSFFQKFD	SISYNPKKDW	FEFAFDYKNF	TKKAEGSRSM	
	WILCTHGSRI	KNFRNSQKNG	QWDSEEFALT	EAFKSLFVRY	EIDYTADLKT	
	AIVDEKQKDF	FVDLLKLFKL	TVQMRNSWKE	KDLDYLISPV	AGADGRFFDT	
	REGNKS LPKD	ADANGAYNIA	LKGLWALRQI	RQTSEGGKLG	LAISNKEWLQ	
	FVQERSYEKD					
WP_023941260.1 type V CRISPR-associated protein Cpf1 [Porphyromonas crevioricanis]	MSLSKDFTNL	YPVSKTLRFE	LKPVGKTLEN	IEKAGILKED	EHRAESYRRV	(SEQ ID NO: 142)
	KKIIDTYHKV	FIDSSLENMA	KMGIE NEIKA	MLQSFCELYK	KDHRTEGEDK	
	ALDKIRAVLR	GLIVGAFTGV	CGRRENTVQN	EKYESLFKEK	LIKEILPDFV	
	LSTEAESLPF	SVEEATRS LK	EFDSFTSYFA	GFYENRKNIY	STKPQSTAI A	
	YRLIHENLPK	FIDNILVFQK	IKEPIAKELE	HIRADFSAGG	YIKKDERLED	
	IFSLNYYIHV	LSQAGIEKYN	ALIGKIVTEG	DGEMKGLNEH	INLYNQQRGR	
	EDRLPLFRPL	YKQILSDREQ	LSYLPESFEK	DEELLRALKE	FYDHIAEDIL	
	GRTQQLMTSI	SEYDLSRIYV	RNDSQLTDIS	KKMLGDWNAI	YMARERAYDH	
	EQAPKRITAK	YERDRIKALK	GEESISLANL	NSCIAFLDNV	RDCRVDTYLS	
	TLGQKEGPHG	LSNLVENVFA	SYHEAEQLLS	FPYPEENNL	QDKDNVVLK	
	NLLDNISDLQ	RFLKPLWGMG	DEPDKDERFY	GEYNYIRGAL	DQVIPLYNKV	
	RNYLTRKPYS	TRKVKLNFGN	SQLLSGWDRN	KEKDNSCVIL	RKGQNFYLAI	
	MNNRHKRSFE	NKVLPEYKEG	EPYFEKMDYK	FLPDPNKMLP	KVFLSKKGIE	
	IYKPSPKLLE	QYGHGTHKKG	DTFSMDDLHE	LIDFFKHSIE	AHEDWKQFGF	
	KFSDTATYEN	VSSFYREVED	QGYKLSFRKV	SESYVYSLID	QGKLYLFIQIY	
	NKDFSPCSKG	TPNLHTLYWR	MLFDERNLAD	VIYKLDGKAE	IFFREKSLKN	
	DHPHTPAGKP	IKKKSQRQKG	EESLFEYDLV	KDRRYTMDKF	QFHVPIITMNF	
	KCSAGSKVND	MVNAHIREAK	DMHVIIGIDRG	ERNLLYICVI	DSRGITILDQI	
	SLNTINDIDY	HDLLESRDKD	RQQERRNWQT	IEGIKELKQG	YLSQAVHRIA	
	ELMVAYKAVV	ALEDLNMGFK	RGRQKVESSV	YQQFEKQLID	KLNYLVDKKK	
	RPEDIGGLLR	AYQFTAPFKS	FKEMGKQNGF	LFYIPAWNTS	NIDPTTGfVN	
	LFHAQYENV	KAKSFFQKFD	SISYNPKKDW	FEFAFDYKNF	TKKAEGSRSM	
	WILCTHGSRI	KNFRNSQKNG	QWDSEEFALT	EAFKSLFVRY	EIDYTADLKT	
	AIVDEKQKDF	FVDLLKLFKL	TVQMRNSWKE	KDLDYLISPV	AGADGRFFDT	
	REGNKS LPKD	ADANGAYNIA	LKGLWALRQI	RQTSEGGKLG	LAISNKEWLQ	
	FVQERSYEKD					
WP_037975888.1 type V CRISPR-associated protein Cpf1 [Synergistes jonesii]	MANS LKDFTN	IYQLSKTLRF	ELKPIGKTEE	HINRKLIMH	DEKRGEDYKS	(SEQ ID NO: 143)
	VTKLIDDYHR	KFIHETL DPA	HFDWNPLAEA	LIQSGSKNNK	ALPAEQKEMR	
	EKIISMFTSQ	AVYKKLFKKE	LFSELLPEMI	KSELVSDLEK	QAQLDAVKSF	
	DKFSTYFTGF	HENRKNIYSK	KDTSTSI AFR	IVHQNF PKFL	ANVRAYTLIK	
	ERAPEVIDKA	QKELSGILGG	KTLDDIFSIE	SFNNVLTQDK	IDYNNQIIGG	
	VSGKAGDKKL	RGVNEFSNLY	RQQHPEVASL	RIKMVPLYKQ	ILSDRTTL SF	
	VPEALKDDEQ	AINAVDGLRS	ELERN DIFNR	IKRLFGKNNL	YSLDKIWIKN	
	SSISAFS NEL	FKNWSFIEDA	LKEFKENEFN	GARSAGKKA E	KWLKSKYFSF	
	ADIDAAV KSY	SEQVSADISS	APSASYFAKF	TNLIETAAEN	GRKFSYFAAE	
	SKAFRGDDGK	TEI I KAYLDS	LNDILHCLKP	FETEDISDID	TEFYSAFAEI	
	YDSVKDVI PV	YNAVRNYTTQ	KPFSTEKF KL	NFENPALAKG	WDKNKEQNNT	
	AIILMKDGKY	YLGVIDKNNK	LRADDLADDG	SAYGYMKMNY	KFIPTPHMEL	
	PKVFLPKRAP	KRYNPSREIL	LIKENKTFIK	DKNFNRDCH	KLIDFKDSI	
	NKHKDWRTFG	FDFS DTD SYE	DISDFYMEVQ	DQGYKLTFT R	LSAEKIDKWV	

Table 5 - Cas12a orthologs						
	EEGRFLFLFQI ELFFRRKSID ASLNAGARGL QGNINSDVNL IGGMDYHAKL DNNAVIVMED PCGILKGLQL NDVSSKEKQK WTNGKRIVRE SRDKSLASEI AALPIDADAN FMQTRKFDF	YNKDFADGAQ KPAVHAKGSM IESGLVWRITE FLRNNKDVNI NQKEKERDTA LNIGFKRGRF TEKFESFTKL DFIGKLDISR KDKDGKFRMN FYVFKNTLQM GAYHIALKGS	GSPNLHTLYW KVNRRDIDGN VKHELVDKDR IGIDRGERNL RKSWKTI GTI KVEKQVYQKF GKQCGIIFYI F'DAKRDMFTF DRLLTEDMKN RNSKRDGTED LVLDIDAIDEKL	KAIFSEENLK PIDEGTYVEI YTIDKYFFHV VYVSLIDRDG KELKEGYLSQ EKMLIDKLN PAGYTSKIDP EFDYDKFRTY ILNKYALAYK FIISPVLNAK KEDGRIDYKD	DVVLKLNGEA CGYANGKRDM PFTINFKAQG HIKLOKDFNI VVHEIVRLAV LVFKDAGYDA TTGFVNLFN QTSYRKKWAV AGEDILPDVI GRFFDSRKTD MAVSNPKWFE	
WP_081839471.1 type V CRISPR-associated protein Cpf1 [Synergistes jonesii]	MENMANSLKD YKSVTKLIDD EMREKIISMF KSFDKFSTYF LIKERAPEVI IGGVSGKAGD LSFVPEALKD IKNSSISAFS FSFADIDAAV AAESKAFRGD AEIYDSVKDV NNTAILMKD MELPKVFLPK DSINKHKDWR KWVEEGRFLF GEAELFFRRK RDMASLNAGA AQGGQGNINS FNIIGGMDYH LAVDNNNAVIV YDAPCGILKG FNINDVSSKE WAVWTNGKRI DVISRDKSLA KTDAAALPIDA WFEFMQTRKF	FTNIYQLSKT YHRKFIHETL TSQAVYKKLF TGPHENRKN DKAQKELSGI KKLRGVNEFS DEQAINAVDG NELFKNWSFI KSYSEQVSAD DGKTEIIKAY IPVYNAVRNY GKYLLGVIDK RAPKRYNPSR TFGFDFSDTD FQIYNKDFAD SIDKPAVHAK RGLIESGLVR VNLFLRNNKD AKLNQKEKER MEDLNIGFKR LQLTEKFESF KQKDFIGKLD VREKDKDGKF SEIFYVFKNT DANGAYHIAL DF	LRFELKPIGK DPAHFDWNPL KKELFSELLP YSKKDTSTSI LGGKTLDDIF NLYRQQHPEV LRSELERNDI EDALKEFKEN ISSAPSASYF LDLNDILHC TTQKPFSTEK NNKLRADDL EILLIKENKT SYEDISDFYM GAQGSPNLHT GSMKVNRRDI ITEVKHELVK VNIIGIDRGE DTARKSWKTI GRFKVEKQVY TKLGKQCGII SIRFDAKRDM RMNDRLLTED LQMRNSKRDT KGSVLDAID	TEEHINRKLI AEALIQSGSK EMIKSELVSD AFRIVHQNFP SIESFNNVLT ASLRIKMVPL FNRIKRLFGK AKFTNLIETA LKPFEFEDIS FKLNFENPAL DDGSAYGYMK FIKDKNFNRT EVQDQGYKLT LYWKAIFSEE DGNPIDEGTY DKRYTIDKYF RNLVYVSLID GTIKELKEGY QKFEKMLIDK FYIPAGYTSK FTFEFDYDKF MKNILNKYAL GEDFIISPVL EKLKEDGRID	IMHDEKRGED NNKALPAEQK LEKQAQLDAV KFLANVRAYT QDKIDYNNQI YKQILSDRTT NNLYSLDKIW KAEEKWLKSKY AENGRKFSYF DIDTEFYSAF AKGWDKNKEQ MNYKFIPTPH DCHKLIDFFK FTRLSAEKID NLKDVVLKLN VEICGYANGK FHVPFITNFK RDGHIKLQKD LSQVVHEIVR LNYLVFKDAG IDPTTGFFNL RTYQTSYRKK AYKAGEDILP NAKGRFFDSR YKDMAVSNPK	(SEQ ID NO: 144)
WP_006283774.1 type V CRISPR-associated protein Cpf1 [Prevotella bryantii B14]	MQINNLKIIY HRADSYKKVK KRIEKTEKDK SDEERTLIKE VDNINAFSKI QKQIEVYNAI LSDRIAISWL DTYNLKGIFI EDYNDRLKKL QTINLFAQVR QRFIKPLLKG SQEKIKLNFE DKDKLDNSGD KKGTHKKGAN DFYREVEQQG NMHTLYWNSL NKNKCNEKKE IDYLRTEDDT IYRTNYHDL KYHAVVVLED AGGLLHAYQL	MKFTDFTGLY KIIDEYHKAF FAKIQDNLRK FKDFTTYFKG ILIPELREKL IGGKSTNDKK PDNFKDDQEA RNDLQLTDIS YTSQESFSIQ NAYTSVQAIL GDESDKDERF NSTLLGGWDL CYEKMVYKLL FNLADCHNLI YSISFCDVSV FSKENLNNII SIFDYDLVKD HIIGIDRGER TREQNREKA LNMGMFRGRQ TSKFESFQKL	SLSKTLRFEL IEKSLSNFEL QIADHLKGDE FYENRENMY NQIYQDFEY IQGLNEYINL LDSIDTCYKN QKMYASWNV YLNDCLRAYG TTPYPENANL YGDFTPLWET NKEHDNTAI PGANKMLPKV DFFKSSISKH EYINKMVEKG YKLNQAEIF KRYTVDKFQF HLLYLVVIDS RESWQTIENI KVEKQVYQKF GKQSGFLFYI	KPIGKTLENI KYQSEDKLDS SYKTIFSKDL AEDKSTAISH LNVESIDEIF YNQKHKDKCL LLNDGNVLGE QDAVILDLKK KTENIQDYFA AQDKETVALI LNQITPLYNM LRKNGLYYLA FFSKSRIDEF EDWSKFNFHF DLYLFQIYNK FRKKSINLYKR HVPITMNFKS HGKIVEQFTL KELKEGYISQ EEMLINKLNY PAWNTSKIDP	KKAGLLEQDQ LEEYLMYYSM IRKNLPDFVK RIIHENLPKF HLDYFSMVM PKLKLFLFKQI GNLKLLENI QVSRKKKESA KLGAVNNEHE KNLLDSLKRL VRNYMTRKPY IMKKSANKIF KPSENIENY SDTSSYEDLS DFSEFSKGTP PTHPAHQAIK TGNTNINQQV NEIVNEYGGN VIHKITDLMQ LVNKKADQNS VTGFVNLFD	(SEQ ID NO: 145)

Table 5 - Cas12a orthologs						
	RYESIDKAKA TYGSRIRTFR ETEKSFEDL NSLPANADAN EKPYLND	FFGKFDSIRY NQAKNSQWDN LHLLKLTLMQ GAYNIARKGL	NADKDWFEFA EEIDLTAKAY RNSITGTTTD MLIQQIKDST	FDYNNFTTKA AFFAKHGINI YLISPVHDSK SSNRKFSPFI	EGTRTNWTIC YDNIKEAIAM GNFYDSRICD TNKDWLIFAQ	
WP_024988992 type V CRISPR- associated protein Cpf1 [Prevotella albensis]	MNIKNFTGLY GFIDEYHKQF FTEIKDNLRS QFEDFTTYFT IANSSVSEHF IGKKVLEDGT SWIAEEFDS INITNDLSLT NKIFKTQKSF IFAQIENAYE PLGSGEENE KLNFDNAQLL PSDGECYDKI KKGKNFNLT VEQQGYRLTS YWKMLFDKRN LNKKHTSTFK RQNGITHIIG NYHNLLEKRE IVVLEDLNGG LNAYQLTNKF IKETKVFWSK RIQTFRNPEK KFFQELHNL ENADANGAYN LKD	PLSKTLRFEL IKDRLWDFKL SITEQLTKSG GFYENRKNMY SDIYESWKEY EIKGINEYVN KKMLSAITES EISQNLFGRY SIAFLNKLPO DAQVFLQIKD KDELFGSFL GGWDVNKEHD DYKLLPGANK CHRLINFFKT HPVSASYIHS LSDVVYKLN YDIKDRRYT IDRGERHLLY KERTEARHSW FMRGRQKVEK ESFKKIGKQS FDIIRYNKEK NAQWDNKEIN HLLTQMRNSV IARKGLMLIR	KPIGKTENI PLESEGEKNS SAYDRIFKKE SSEEKSTAI LNVNSIEEIF LYNQQQKDKS YNHLHNVLMG DVFTNGIKNK PEMEDGKPRN TDNKLSONKT AIWDELDTIT CAGILLRKND MLPKVFFSKS SIEKHEDWSK LVKEGKLYLF QAEVFYRKSS VDKFQFHVPI LSLIDLKGN SSIESIKELK QVYQKFEKKL GFLFYIPAWN NWFEFVFDYN LTESFKALFE TGTDIDYLLS QIKQADPQKK	EKNIGILTKE LEEYQELYEL FIREDLVNFL YRLIHQNLPK QLDYFSETLT KRLPFVLPLY NENESLRNLL LRVLTPRKKK IEDYFITQGA AVEKIKTLLD PLYNKVRNWL SYLLGIINKK RIKEFEPSEA FGFKFSDTET QIWNKDFSQF IEHQNRRIHP TINFKATGQN IKQMTLNEII DGYMSQVIHK IDKLNVLVDK TSKIDPITGF TFTTKAEGTR KYKIDITSNL PVADEGDNFY FKFETITNKD	QRAKDYLVK TKRNDQAEAD EDEKDKNIVK FMDNMRSFAK QPHIEVYNYI KQILSDREKL LNIKDYNLEK ETDENFEDRI INTKSIQKED ALKELQHF TRKPYSTEDI TNHIFDTDIT IINCYKKGTH YEDISGFYRE SGKTPNLHTL AQHPITNKNE NINPIVQEV NEYKGVTYKT ITDMMVKYNA KLDANEVGGV VNLNTRYES TKWTLCTHGT KESIMQETEK DSRINGKNFP WLKFAQDKPY	(SEQ ID NO: 146)
WP_039658684. 1 type V CRISPR- associated protein Cpf1 [Smithella sp. SC_K08D17]	MQTLFENFTN VKNIIDEYHK NLRKQIANAF FTGFHQNRAN LSPFNQTLKD EEGKTKIKGL EAFKNDTEIL KMYFRSGASL KEKWLKQDFN IQKVNEGYIA PLSLKDTDKE KLNFNSTLL ADKKDFCYEK HKKGDNFNLN EVEHQGYKIS LYWKMLFDEN PDNPKATSTF LKANPDINII NLLDKKEGDR MEDLNFQFKR FQLANKFESF AKDFFEKFDS AWNRLNNNR FKALMKNLSI PKNADANGAY	QYPVSKTLRF DFIEKSLNGL RNNEKFKTLF MYVADEKRTA MKDVIKGTTL NEYINTDFNQ EAIEKFYVNE TDVSRKVFG VSLIQTAIDE VKDLLNTPCP KDETFFSLFT GGWDLNKETD MVYKLLPGAN HCHKLIDFFK FQSVADSFID NLKDVVYKLN NYDIVKDKRY GIDRGERHLL ATARQEWGVI GRQKVEKQVY QKMGKQNGFI IRLNSKADYF GSQEKYDITA TSLSRHNNGE HIALKGLWCL	ELIPQGKTKD KLDGLEKYKT AKELIKNDLM IASRLIHENL EEIFSLDYFN KQTDKKKRQP LLHFSNEGKS WSIINRALDN YDNETVKGKN ENEKLGSKND PLYDHLTQTI NTAII LRKDN KMLPKVFFSQ DSINKHEDWK DLVNEGKLYL GEAEVFYRK TIDKFQFHIP YYALINQKGK ETIKELKEGY QKFEKMLIDK FYVPAWNSTK EFAFDKFNFT ELKSLFDGKV KGDNEQDYIL EQISKTDLLK	FIEQKGLLKK LYLKQEKDDK SFACEEDKKN PKFIDNIKIF KTLTQSGIDI KFKQLYKQIL TNVLDKAIKNA YYATTYPIKP SGKVIADYFA QVKQIKAFMD ALYNKVRNYL LYYLGIMDKR SRIQEFTPSA NFDREFSATS FQIYNKDFSP ITMNFKAEGI ILKQDTLNVI LSQVIHKLTD LNYLVDKNKK TDPATGFIDF EKADGGRTKW DYKSGKDLKQ SPVADSKGRF KVKLAISNKE	DEDRAEKYKK DKKAFDKEKE VKEFEAFTTY EKMKEAPEL YNSVIGGRTP SDRQSLSFIA VSNLESFNLT REKSEKYEER KFCDDKETDL SIMDIMHFVR TQKPYSTEDI HNRIFRNVPK KLENNYANET TYADLSGFYH FSKGGPNLHT KANESIINKN FNMNQRVNQF ANEKQKVDYH LMIENNAIIV ANELGGLLNA LKPRYENLNQ TVCTTNEDRY QIASQESADF FDSRKADDDM WLEFVQTLKG	(SEQ ID NO: 147)
WP_037385181	MQTLFENFTN VKNIIDEYHK NLRKQIANAF	QYPVSKTLRF DFIEKSLNGL RNNEKFKTLF	ELIPQGKTKD KLDGLEKYKT AKELIKNDLM	FIEQKGLLKK LYLKQEKDDK SFACEEDKKN	DEDRAEKYKK DKKAFDKEKE VKEFEAFTTY	(SEQ ID

Table 5 - Cas12a orthologs						
type V CRISPR-associated protein Cpf1 [Smithella sp. SCADC]	FTGFHQNRAN LSPFNQTLKD EEGKTKIKGL EAFKNDTEIL KIYFRSGTSL KEKWLKQDFN IQKVNEGYIA PLSLKDTDKE KLNFNSTLL ADKKDSCYEK HKKGDNFNLN EVEHQGYKIS LYWKMLFDEN PDNPKATSTF LKANPDINII NLLDKKEGDR MEDLNFGFKR FQLANKFESF AKDFFEKFDS AWNRRALNNR FRTLMKYLSV PKNADANGAY	MYVADEKRTA MKDVIKGTTL NEYINTDFNQ EAIEKFYVNE TDVSRKVFGE VSLIQTAIDE VKDLLNTPYP KDETFYSLFT GGWDLNKETD MVYKLLPGAN HCHQLIDFFK FQSIADSFID NLKDVVYKLN NYDIVKDKRY GIDRGERHLL ATARQEWGVI GRQKVEKQVY QKMGKQNGFI IRLNSKADYF GSQEKYDITA TSLSRHNNGE HIALKGWLWCL	IASRLIHENL EEIFSLDYFN KQTDKKRQF LLHFSNEGKS WSIINRALDN YDNETVKGKN ENEKLGSKND PLYDHLTQTI NTAIILRKEN KMLPKVFFSQ DSINKHEDWK DLVNEGKLYL GEAEVFYRKK TIDKFQFHVP YYTLINQKGG ETIKELKEGY QKFEKMLIDK FYVPAWNTSK EFAFDFKNET ELKSLFDGKV KGETEQDYIL EQISKTDLLK	PKFIDNIKIF KTLTQSGIDI KFKQLYKQIL TNVLDAIKNA YYATTYPIKP SGKVIVDYFA QVKQIKAFMD ALYNKVRNYL LYYLGIMDKR SRIQEFTPSA NFDFFRSATS FQIYNKDFSP SIAEKNTTIIH ITMNFKAEGI ILKQDTLNVI LSQVIHKLTD LNYLVDKNKK TDPATGFIDF GKADGGRTKW DYKSGKDLQK SPVADSMGKF KVKLAISNKE	EKMKEAPEL YNSVIGGRTP SDRQSLSFIA VSNLESFNLT REKSEKYEER KFCDDKETDL SIMDIMHFVR TQKPYSTEKI HNRIFRNVPK KLENYENET TYADLSGFYH FSKGGPNLHT KANESIINKN FNMNQVRNQF ANEKQKVDYH LMIENNAIIV ANELGGLLNA LKPRYENLKQ TVCTTNEEDRY QIASQELADF FDSRKAGDDM WLEFMQTLKG	NO: 148)
WP_039871282.1 type V CRISPR-associated protein Cpf1 [Prevotella bryantii B14]	MKFTDFTGLY KIIDEYHKAF FAKIQDNLRK FKDFTTYFKG ILIPELREKL IGGKSTNDKK PDNFKDDQEA RNDLQLTDIS YTSQESFSIQ NAYTSVQAIL GDESDDKDERF NSTLLGGWDL CYEKMVYKLL FNLADCHNLI YSISFCDVSV FSKENLNNII SIFDYDLVKD HIIIGDRGER DTREQNREKA LNMGMFRGRQ TSKFESFQKL FFGKFDSIRY NQAKNSQWDN LHLLKLTLMQ GAYNIARKGL	SLSKTLRFEL IEKSLSNFEL QIADHLKGDE FYENRENMYN NQIYQDFEY IQGLNEYINL LDSIDTCYKN QKMYASWNV YLNDCLRAYG TTPYPENANL YGDFTPLWET NKEHDNTAI PGANKMLPKV DFKSSISKH EYINKMVEKG YKLNQAEIF KRYTVDKFQF HLLYLVVIDS RESWQTIENI KVEKQVYQKF GKQSGFLFYI NADKDWFEFA EEIDLTKAYK RNSITGTTTD MLIQQIKDST	KPIGKTLENI KYQSEDKLDS SYKTIFSKDL AEDKSTAISH LNVESIDEIF YNQKHKDCCL LLNDGNVLGE QDAVILDLKK KTENIQDYFA AQDKETVALI LNQITPLYNM LRKNGLYYLA FFSKSRIDEF EDWSKFNHFH DLYLFQIYNK FRKKSILNYK HVPITMNFKS HGKIVEQFTL KELKEGYISQ EEMLINKLNY PAWNTSKIDP FDYNNFTTKA AFFAKHGINI YLISPVHDSK SSNRKFESPI	KKAGLLEQDQ LEEYLMYYSM IRKNLPDFVK RIIHENLPKF HLDYFSMVMT PKLKLFLKQI GNLKLLENI QVSRKKKESA KLGAVNNEHE KNLLDSLKRL VRNYMTRKPY IMKKSANKIF KPSENIENY SDTSSYEDLS DFSEFSKGTP PTHAPHAQAIK TGNTNINQQV NEIVNEYGGN VIHKITDLMQ LVNKKADQNS VTGFVNLFD EGTRTNWTIC YDNIKEAIAM GNFYDSRICD TNKDWLIFAQ	HRADSYYKKVK KRIEKTEKDK VDNINAFSKI QKQIEVYNAI LSDRIAISWL DTYNLKGIFI EDYNDRLKKL QTINLFAQVR QRFIKPLLGK SKEIKLNF DKDKLDNSGD KKGTHKKGAN DFYREVEQQG NMHTLYWNSL NKNKCNEKKE IDYLRTEDDT IYRTNYHDL KYHAVVLED AGGLLHAYQL RYESIDKAKA TYGSRIRTFR ETEKSFEDDL NSLPANADAN EKPYLND	(SEQ ID NO: 149)
EKE28449.1 hypothetical protein ACD_3C00058G0015 [uncultured bacterium (gcode 4)]	MFKGDAFTGL KAYFDLIHKE EKVMKSLKKE PDEKELFDKF KNIEDFKKLN WWYSLENWEK KKVYIDEIRD IYLTRQSINT TLENIFKDCY LEWELDKIQN LDTDNFFYND LTWWDKDKEN	YEVQKTLRFE FVRQSLENAN FVSFFDAKWN DKFFTYSNF SSFREFVNDN VQWLNEKINN DKNLIDLIDN ISSKYFSSWD VKGINTESQN EDIKKNEKQV FNEYLEGFEI ERLGIILRRE	LVPIGLTQSY LCQLDDFYEK EWKQKFSFLK KESRKNFYAD FSEEDKQIFE FKQNQNKSNS SKRRNQIKID YIRWYFWTGE NIVFETQGIY EVIKNYFDSV WKDYNLVRNY WKYYLWILKK	LENDWVIQKD YIELHNSLET KWWIDVLNEK DGRAWAIAIR IDFYNNCLLQ KDLKFPRMKL NANDIINDFI LQEFVSFYDL ENFLNIFKFE MSVYKMTKYF ITKKQVNTDK WNTLNFQDYL	KEVEENYGKI RKDKNLAKQF EVLDLMAEFY AIDENLITFI PWIDKYNKIV LYKQILGDKE NNNAKFELDK KETFWKIEYE FNQNISQISL SLEKWKRRVE IKLNFNDSQF QKEWEIFYEK	(SEQ ID NO: 150)

Table 5 - Cas12a orthologs						
	MNYKQLNNVY QIKIEFDIFQ KNYDDLSIFY ESKKEWSTEN KDKNGQEILD IKIWIWDRWE SRKDSRISWW KRWRQKIEKQ DYQDIWNYKQ NSIAIKYNDW LTKVIDVNKK LRNSDSKDNK RIRKNQEKPD	RQLPRLLFPL ESKEKWEKFD SDVEKKMYNL IHTKYFKLLF HRRFSQDKIM KHLAYYCVID EENIKELKN IYQKLELALA SWIMFYVRAN KFSFSYEIDD LLELFRDFNL DYISCPSCYY LYVSDIDWDN	TKKLNELKWD IDKLRKLEIY NFTKIDKSLI NEKNLQNLVV FHSITLNAN KSWKIFNNEI GYISQVINKL KKLNYLTQKD YTSVTCPCNW KSWKQKQSLN SLDINKQIQE HSNNWLQWFE FVQSNQFPNT	ELKKYLSKYI YKKWVLALYS DGKVKSWELY KLSWWADIFF CWDKYWFNQY WTLNELNWNV TELIVKYNAI KKDDEILWNL LRKNLYISNS KKKFIVYSOI KDLDSVFFKS FNWDANWAYN IIPQNIQIEKQ	QNFWYNEEIA DLYDLEFIKY LFQIYNKDFS RDKTENLKFK VNEYMNKERD YLEKLEKIES IVFEDLNIWF KALQLVPKVN ATKENQKKSL ERFVYSPLEK LTHLENLILQ IARKGIILLD VPLNIKI	
WP_018359861.1 type V CRISPR-associated protein Cpf1 [Porphyromonas macacae]	MKTQHFFEDF EKLKVIDEY DTLAKAFSED HENRKNLYTS MMENLRNVWP KHQGINWIN VFCVLRQFRK ISKNI FDRWN AELDDLALHY DEVLI AFRDL SGTGAEIRRD HFDNPSLLSG AEEMFYEKME TGQKGFNKKD REQAYKVMV WKALFSEQNQ KGETSLFNVD NDDLIQIIGID ILGDREQERL ENLNL S FMKG YQLTNPLFSF VGDARKFFDR IAKSKKSGKW FYVRLIYLFN ADANGAYNVA	TSLSLSLTKTI HEDFIANILS ERYKSIFKKE NEITASIPYR SFVKTPDDLC IYRQRNKEMR LEWNTVSSKE YISDAIRRKT SEESLPAGFS QVILDKDFTE SSFYALYTDR WCKNKELNNL YKQIAEPMML LYRLIDFYKE NVPASYIDEA SRVYKLCGGG LVKDKRFTED RGERNLLYIS RRRQEWKSIE RKKVEKSVYE EELHRYPQSG FNAIRYDGKG MVERIENLSL LMMQIRNSDG RKGLMVVQRI	RFELKPIGKT SFSFSEEILQ LVKKDIPVWC IVHVNLPKFI NLKTYNHLMV LPGLVFLHKQ DDAASLKDLF EVLMPRKKE LLSYFTSLGG KKLGKDEEAV MDKLKGLLKM SVIFRQNGYY LPKVFFPKKT ALTVHEWKLF VENGKLYLFQ ELFYRKASLH KFFFHVPI SI RIDTRGNLLE SIKDLKGYM KFERMLVDKL ILFFVDPWNT NILFDLDSR CFLELFEQFN EEDYILSPAL KRGDHESIHR	LENIKKNGLI SYIQNLSESE PAYKSLCKKF QNIEALCELQ QSSISEYNRF ILAKVDSSSF CGLSGYDPEA VERYAEKISK QKYLVSDEGEV SVIKKALDSA YDKVRNYLTK YLGIMTPKGK KPAFAPDQSV NFSFSPTEQY IYNKDFSPYS MQDTTVHPKG NYKNKKITNV QFSLNVIESD SQVVHKICNM NYLVVDKKNL SLTDPSTGFV FDVRVETQRK IGYRVEKDLK NEKNLQFDSR IGRAQWLRYV	RRDEQRLDDY ARAKIEKTM DNFTTSLVPF KKMGADLYLE VGGYSTEDGT ISDTLENDQ IYVSDAHLAT QIKKRQSYSL ILYEEGSIW LRLRKFDDL KPYSEKFKL NLFKTLPKLG VDIYNKKTFF RNIGEFFDEV KGIPNLHTLY ISIHKKNLNK NQMV RDYIAQ KGDRLTDYQK VVEHKAIVVL SNEPGGLYAA NLLGRINVTN LWTLTTFGSR KAILSQDRKE LIEAKDLPVD QEGIVE	(SEQ ID NO: 151)
WP_013282991 type V CRISPR-associated protein Cpf1 [Butyrivibrio proteoclasticus]	MLLYENYTKR KPVIDEGIKD KTLIKNLPEG TKGYLALFSK SFAIEYESIC EKNIKIKSEH EAFEQTKEQW RIRLLLEEYE KMTGDDRLFI LVSFYDELTK CRNYVTKSPA YYYYVLTDPV FESPEHKNDK KFKDDCRYLI LCTMEWIPVS DSNMEHTSML IREATYKMKN FFLSLSYTKN VDEKNNVLEE EAYVELAVTQ CARMSDLSFN	NQITKSLRLE IARDTLKNCE IGKISEINSA FLTSTRITTLE SVDYYNQILS LEEKPFRIK ENIKKINGFY KNTLEVLPRR LYRDELARNY FRSALRIINS DDARIMASCL KPVLDLTKDE NCVIEEYVSK DVKYEFIAVY FERINDMVDK LNSRAMIQYR NFDISSEDFI ADISARTLDY KSLNEIDGVN ISKLATKYNA TIKEGEAGSI	LRPQKTLRN LSFEKLYEHF KYLNGVLYDF QSMPKRVIEN QEDIDSYNRL QLHKQILEER AKDPGDITLF TKSEKARYDK MRIKEAYGTF GNDEKADPIF GTPARLRTHW TDAQIFVQRK PLTIDRRAYD TRYSCFNMSG KEGLLFLVRS AASLPRRVTH MAKAYLAEHD INDKVEEDTQ YRELLKERTK VVVVESMSST SNPIQVSNNN	IKELNLEQD LSGDKKAYAK IDKTHKDSEE FEIYAANIPK ISGIMDEGGA EKAFTIDHID IVVGPNQTHV FVNAVPPKVA ERDILKSRRG YNTFDGIFEK WNGEENFAIN GAKSFLGLPK IFKNGTFKKT LKRADYNDI MFLYNRPKRP KKGSI LVALR VAIKKANEDI DSRMAVIVTR IKYHDKTRLW FKDKFSFLDE GNSYQDGVYI	KAIYALLERL ESERLKKIEV QKNILSDILE MQDALERGAV KEKGINQTIS SDEEVVQVTK LSQLIYGEHD KESHTFDGLQ IKGNRDVQES ANRTYKAENL DVAMIRRGDE ALFKCILEPY NIGIDGLTEE GEFFSDVDTR YERTFIQLFS DSNGEHI PMH IRNRRYTEDK NLKDLTYVAV QYDVSSKGLK QIFKAFEARL FLNNAYTRTL	(SEQ ID NO: 152)

Table 5 - Cas12a orthologs						
	CPDTGFVDVF RLDRKEWTV NMAEKFLDKT DVSKNTTYNF	DKTRLITMQS KIAGDGSYFD AISGKSVELI GKKFMFKQEY	KRQFFAKMKD KDKGEYVYVN YKWFANALYG RGDGDFLDAF	IRIDDGEMLF DIVREQIIPA IITKKDGEKI LNYMQAQDIA	TFNLEEYPTK LLEDKAVFDG YRSPITGTEI V	
WP_048112740.1 type V CRISPR-associated protein Cpf1 [Candidatus Methanoplasma termitum]	MNNYDEFTKL KEAIDEYHKK RMRQEIVSEF FSGYFIGLHE YPEWIIKAES SCEKMMGLND EDSQLLPSIG NRVSNVIFGE DVLEAIKRTG HIIKTLLDSV NKVRNYLTKN LGIINPKRKK KEYKPSKEII FYFSPTESYG YNKDFVKAAT KEIVHREGEI VRYFKAHYDI EKAHIIIGIDR EMKDARQSWN KRGREFKVEKQ SFAKLGKQTG FESISYSASKD RNELFDPDSKE AIQMRVYDGK GELTMRAIAE	YPIQKTIRFE FIDEHLTNMS KKDDRFDKDLF NRKNMYSDDG ALVAHNIMKD ALNLAHQSEK GFFAQIENDK WGTLGGLMRE NNDAFNEYIS QQFLHFFNLF NLNTKKIKLN NIKFEQSGSN EGYEADKHIR DISEFYLDVE GKKDMHTIYW LVNRTYNGRT TKDRRYLNDK GERNLLYYSI AIGKIKDLKE IYQKFENMLI ILFYVPAAYT GGIFAFADFY IKEALTSSGI EDYIISPIKN KFDPDSEKMA	LKPQGRTEH LDWNSLKQIS SKKLFSSELLK EITAI SNRIV EVFSLEYFNK SSKGRIHMTF DGNIFDRALE YKADSINDIN KMRTAREKID KARQDIPLDG FKNPTLANGW GPFYRKMYVK GDKFDLDFCH KQGYRMHFEN NAAFSPENLQ PVPDKIHKKL IYFHVPLTLN IDRSGKIIDQ GYLSKAVHEI DKMNYLVFKD SKIDPTTGFV RKFGTSKTDH KYDGGQNILP SKGEFFRTDP KLELKHKDW	LETFNFFFEED EKYYKSREEK EEIYKKGNHQ NENFPKFLDN VLNQEGIQRY LKFQILSEKE LISSYAEYDT LERTCKKVDK AARKEMKFIS AFYAEFDEVH DQNKVYDYAS QIPGPNKNLP KLIDFFKESI ISAETIDEYV DVVVKLNGEA TDYHNGRTKD FKANGKKNLN QSLNVIDGFD TKMAIQYNAI APDESPGGVL NLFNTSSKTN KNVWTAYTNG DILRSNNNGL KRRELPIAD EFMQTRGD	RDRAEKYKIL DKKVFLSEQK EIDALKSFDK LQKYQEARKK NLALGGYVTK SFSYIPDVFT ERIYIRQADI WLDSKEFALS EKISGDEESI SKLFAIVPLY LIFLDRGNYY RVFLTSTKGG EKHKDWSKFN EKGDLFLFQI ELFYRDKSDI LGEAKEYLKD KMOVIEKFLSD YREKLNRQEI VMEELNYGF NAYQLTNPLE AQERKEFLQK ERMRYIKEKK IYTMYSFFIA ANGAYNIALR	(SEQ ID NO: 153)
WP_027407524.1 type V CRISPR-associated protein Cpf1 [Anacrovibrio sp. RM50]	MVAFIDEFVG GVLKELLDDY LRKDLVKAFG NWLNNNAKYS IAFRVIDQNM SQSKIDEYNY KQILSDRDNS DNYGNIYVKT SEYCKVLDAD MLQAIRKYKL RNFATKKPYS ADSKSKNIFD NEKIFGHLIS EYFKFKFKNT LFQLYTKDFS EAEMFMRNPS KFYFHCPIITL VINQTDGILE QAWETIENIK VEKQVYQKFE KQGTGCIYYVI YNAKADYFEF TDRKLKLFKA VSGTENENDF TIRGIEDGKL	QYPVSKTLRF YRAYIEDALT NLKDYLIGSD AEDREKYIKA VTYFGNIRIY QCIGRPIDDA FMSEVINRNE QPLTELSQAL DAAKIQDKFN FSMYNGRKKM DEKMKLSFNM FKKNPHLLDK KRILEIREKK AEYDNANKFY DKKKKKGTDN IEFQVTHEHN NFRADKPIKY QGS LNKISGS ELKAGYLSQV KAMIDKLNLY PSYTSHIDPK AFDYRSFGVD HGIDYVGGEN ILSPVEYAPG HNYGKGGENA	EARPVPETKK SFTLDKALLE QLKDLVKLKA IESFEGFVTY EKIKAKYPEL DFKGVNSLIN EAIECAKNGY FGDWSILRNA LKDYFIQKNA DVPENGIDFS PTMLAGWDYN YSSKDIYYKV LYTAAAGDRK EDIDKQTYSL LHTMYWHGVF KPIANKNPLN NEKINRFVEN YTNDKGEKVN VYKLTQLMLQ VFKDRGYEMN TG FVNLLNAK MARNEWVCT LVSHITEVAD KFFDSREATS AWFKFMQONQE	WLES DQCSVL NAYDLYCNDR KVDAPAGKGK LTNYKQAREN YSALKGFEKF EYRQKNGIKA KVS YALFNEL LDNGKYDKDI LDATLPDLDK NEFNAIYDKL NETANGCFLF KYKQVSGSAK AVA EWIDFMK EKVEIPT EYI SDENLKAVTE TKKESVFNYD NPDVCIIGID KETDYHDLLD YNAVIVLENL GSYAKGLQLT LRYENITKAQ CGDLRWEYSA KHFLSTLLFY TEPMNADANG YKNNG	FNDQKRNEY TNAFSSCCEK KKIEVD SRLI MFSSSEDKSTA FSP TAYSEIL REL PVM SMLY LQLYKKIFTE INLAELEKYF ITQYKPHLDA SEFSILYDRI IKDGKYLGLV MLPKVVFAGS SAIAIHPEWN DEMVSQHKLY GTQPIIKLNG LIKDKRYTER RGERHLLYYT RKEKGKHVAQ NVGFKRGRTK DKFESFDKIG DTIRKFD SIS KTRETKAYSV LRLVLKMRYT AYHIALKGLM	(SEQ ID NO: 154)
WP_044910712.1 type V CRISPR-associated	MDYNGNQFER VTPIVDDCIR LLAKVL TENL TVLMQRFLTT KFDPDAEQYA	RAPLTKTITL KIADNALCHF PDGLRKVNDI RITALTVWLP SLEFYGQCLS	RLKPIGETRE GTEYDFSCLG NSAAFIQDTL DRV FENFNI F QKDIDSYNLI	TIREQKLLEQ NAISKND SKA TSFVQDDADK IENAEKMRI L ISGIYADDEV	DAAFRKLVET IKKETEKVEK RVLIQELKGG LDSPLNEKIM KNPGINEIVK	(SEQ ID NO: 155)

Table 5 - Cas12a orthologs						
protein Cpf1 [Lachnospiraceae bacterium MC2017]	EYNQQIRGDK ILKDTEMLPS DKESKRISL VMAAYIAAVE WTVFRNLIRI KIGSDLKPEI KGAKPVELED FVFLSGMKAP AYKEFLENRM QLDDLKSGN SMRTLLNSRP LYRFFNNLLI FSVKKNADCL VLKQRLSLNII LNYALKEIAE LSDLHFRGIK DTGFIFAIND TDNRVWNCCC KLNERENVPG FIQAMNLRNK	DESPLPKLKK KIIIEAMKEAD METLSPKERK ESCAEIMRKE LRRKSEAEID ATYGSALRPN MDGDIECLQM VTITRETYEA IYADLNFGFK GLLFEIWSER MLVYRPKESS KEKLGEKARK DTTKDLNAEV NDGARDVDWK AVIEYNAILI DGEPCSFNTN HNIRTKKAKL NHPITNYDRE KVVDIAISLV CDYYRIGSKE	LHKQILMPVE AGDIAVYGSR ESKKRLEGLE KDLRTLSSKE SDFYDVLDD SRWWSPEGEK RKIPNPTIFL YRYKLYTVGK DLEEKDSSE LESYYKYGNE KPMVVHRDGS ILDNKKVKVK MEQYSESNRL ERFRQVTKDR IEKMSNAFKD LQLCQNDSENK NFLSKFDQLK TKKVEFIEEP LNYLRGTVSG RGEWTDFAVQ	KAFFVRVLSN LHELHSHVIYG EHIRKSTYTF DVKIRGNRHN VEVLSLTYKG NVKFHTIVRR PKLVFKDPEA LRDGEVSEEE FIKQVETHNT KVLRGYEGVL RVVDRFDKDG VLESERVKWS ILIRNTDIL KYSFLDDVTF ILQDGVIFMV VSSEGCLIMK VEELSRVLEE VAGQRAVYYS LIN	DSDARSILEK DHGKLSQIIY DELNRYAEKN TLIVKNYFNA ENLCRSYITK DGRLYYFILP FFRDNPEADE YKRALLQVLT FMCWAKVSSS LSILKDENLV PNSMTRSLDP KGYDEQFAVT YYLVLDDKNGK RTSNDLKEVY KGFETKLLAK PNSMTRSLDP YSGDSLPTHN NGIETDTELN PVTGKKYDIS	
WP_081834226 type V CRISPR- associated protein Cpf1 [Lachnospiraceae bacterium MC2017].	MTMDYGNQGF ETVPIVDDC EKLLAKVLTE GKTIVLMQRF IMKFDPAEQ VKEYNQQIRG EKILKDTEML IYDKESKRIS KNVMAAYIAA NAWTVFRNLI TKKIGSDLKP LPKGAKPVEL DEFVFLSGMK LTAYKEFLEN SSQLDDLKVS LVSMRTLLNS DELYRFFNNL VTFVSKKNAD GKVLKQRLSL VYLYALKEI AKLSDLHFRG DPDTGFIFAI HNTDNRVWNC LNKLNERENV ISFIQAMNLRN	ERRAPLTKTI IRKIADNALC NLPDGLRKVN TTRITALTVW YASLEFYGQC DKDESPLPKL PSKIIIEAMKE ELMETLSPKE VEESCAEIMR RILRRKSEAE ETATYGSALR EDMDGDIECL APVTITRETY RMIYADLNFG GNGLLFEIWS RPMLVYRPKE LIKEKLGEKA CLDTTKDLNA IINDGARDVD AEAVIEYNAI IKDGEPCSF NDHNIRTKKA CCNHPITNYD PGKVVDIAIS RKCDYYRIGS	TLRLKPIGET HFGTEYDFSC DINSAAFIQD LPDRVFENFN LSQKDIDSYN KKLHKQILMP ADAGDIAVYG RKESKKRLEG KEKDLRTLSS IDSDFYDVLDD PNSRWWSPE QMRKI PNPTI EAYRYKLYTV FKDLEEKDS ERLESYYKYG SSKPMVVHRD RKILDNKKVK EVMEQYSESN WKERFRQVTK LIIEKMSNAF NPLQLCQND KLNFLSKFDQ RETCKVEFIE LVLNLYRGTV KERGEWTDFAV	RETIREQKLL LGNAISKND TLTSFVQDDA IFIENAEKMR LIISGIYADD VEKAFFVRVL SRLHELHSHVI LEEHIRKSTY KEDVKIRGNR DSVEVLSLTY KFNVKFHTIV FLPKLVFKDP GKL RDGEVSE SEFIKQVETH NEKVLRGYEG GSRVVD RFDK VKVLESERVK RLILIRNTTD DRNEGYNEWD KDKYSFLDDV NKILQDGVIF LKVSEGCCLI EPVEELSRVL SGVAGQRAVY AQLIN	EQDAAFRKLV KAIKKETEKV DKRVLIIQELK ILLDSPLNEK EVKNPGINEI SNDSDARSIL YGDHGKLSQI TFDELNRYAE HNTLIVKNYF KGENLCRSYI RRDGRLYYFI EAFFRDNPEA EYKRALQV NTFMCWAKVS VLLSILKDEN DGKYPPEVH WSKFYDEQFA ILYYLVLDDK YSRTSNDLKE TFKGFETKLL MVPNSMTRSL MKYSGDSLPT EENG IETDTE YSPVTGKKYD	(SEQ ID NO: 156)
WP_027216152. 1 type V CRISPR- associated protein Cpf1 [Butyrivibrio fibrisolvens]	MYYESLTKLY AIMDDYHKRL RKAIAKRLRE SYNDVRLNLY LSSEAQDGLF FRKVPKMKVL KVQLLFTALR NNRRKDEKYY DDIRFKGKNF INGSGQELEK LNFNKPTFLD DKSDCYKKVI LEELKNITYLP YSFYFKSDDD	PIKKTIRNEL INEALSGFAL HENFGKIGNK SDKEKSSTVA LVNTFNNVLI YKQILSDREE ESKGVNVYVK DKRQKELKKN YEAVLCGHDR NLVVYGEQEA GWDYGNEEAY YKQISSPEKS KKINDIRKRR YASFKEFQED	VPIGKTLENI IDLDKAANLY DIIPLLQKLS YRLINENLPR QDGINTYNED SFIDEFESDT NDIAKTSFSN KSYTIEKIIT SKKLSKNKGA VLSELSGIDS LGFFMIKEGN IQNLIMVIDGK SYLNGDTFSK VGRQAYQISY	KKNNILEADE LSRSKSADDI ENEDDYNALE FLDNIRAYDA IGKLNVAINL ELLDSLESHY VVFGSWSRID LSTEDVDVIG VEAIKGYLDS LYNMTNRNYLT YFLAVMDANW TVKKNGRKEK KDLTEFIGYY VDVPVSFVDD	DRKIAYIRVK ESFSRFQDKL SFKNFYTYFE VQKAGITSEE YNQKNASVQG ANLAKYFGSN ELINGEYDDN KYIEKLES DI VKDFERDLKL KKPFSTEKIK NKEFRNIPSV EGIHSGENLI KQRVIEYYNG LINSGLKLYLF	(SEQ ID NO: 157)

Table 5 - Cas12a orthologs						
	RVYNKDFSEY IKKEELIVHR SIIMNFGAEE KILEQRSLNC AGYLSQVVHI IDKLNLYLMD PAYLTSKIDP GYFEFAFDYK LTNSLKELFD SSCDGSRDYI LNEIRNSEEG	SKGRNLNLHTL AGEEIKNKNP NVSFNDIVNG ITDSNLDIET VAELVLKYNA KSREQLSPEK MTGFANLFYV KFTDRAYGRV SYKVTYESEV ISPVKNSKEG SKINLAMSNA	YFKMLFDERN KRAAQKPTRR VLRNEDKVVV DYHRLLEKE IICLEDLNFG ISGALNALQL KYENVDKAKE SEWTVCTVGE DLKDAILAID FFCDNNDDT QWLEYAQDNT	LKNVVYKLN LDYDIVKDRR IGIDRGERNL SDRKIARRDW FKRGRQKVEK TPDFKSFKVL FFSKFDSIKY RIIKFKNKEK DPAFYRDLTR TPNDADANGA I	QAEVFYRPSS YSQDKFMLHT LYVVVIDPEG TTIENIKELK QVYQKFEKML GKQTGIIYYV NKDGKNWNTK NNSYDDKVID RLQQTLOMRN FNIARKGLWV	
WP_016301126.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium COE1]	MHENNGKIAD DYKSVKKIID IQMSLRKQIV EFTTYFEGFH TDIRSDLTEL FSTDNTKIK FDSDEVLLEA DTTISAISND IKMYSIEELN KGLINNISLC TELLRIWEEL KEKDNLGIIL KVSANLPRIF GIKQYEDWGQ LVNEGKLYLF NAEIFYRKAS CDKYQFHVPI LCMIDMKGNI WQTIHTIKEL RQVYQKFEKM SGFLYYIPAW REYFEFLFDY DLTFELKKLF SDEREDKLIS IKNTDVEQLD	NFIGIYPVSK AYHKYFIDEA KRFSEHPQYK QNRKNMYSDE FNNLKTMDI GLNEYINLYN VDMFYNRLLQ LFDDWSYISK FFVKKYSCNE QDRQAISELK EPITLLYNKV LKDGOYYLGI LKDKYNPSEE FDFKFSDES QIYNKDFSPY INQKDVVHK TMNFKALGEN VKQISLNEII KEGYLSQVIH LTDKLNLYLD NTSKLDPTTG SAFTSKAEGS NDYSISLLDG PVLNKYGAF KVKLTISNKE	TLRFELKPVG LNGIQLDGLK YLFKKELIKN EKSTAIAYRV TIVEEYFAID QKNKAKLPKL FVIENEGQIT AVRENYDSEN CHIEGYFERR DFLDSIKEVQ RNYVTKKPYT MNRNKNKIAD MLEKYEKGTH YDDISAFYKE SKGTKNLHTL ADLPIKNKDP HFNKRVNRLI SYDKNKLEHK VITDLMVEYN KSKGMDDEDG FVNLFYTKYE RLKWTVCSSG DLREQMGKID ETGKNERMPL WLQYAEHIL	KTQEYIEKHG NYYELYEKRR VLPEFTKDNA VHQNLPKYID GFNKVVNQKG KPLFKQILSD ISKLLTNFSA VDKNKRAAAY ILEILDKMRY WLLKPLMIGQ LEKVKLNFKYK DAPLAKTDNV LKGENFCIDD VEHQGYKITF YWEMLFSSQN QNSKKESMFD HDAENMHIIG RNYHQLLKTR AIVVLEDLNF LLHAYQLTDE SVEKSKEFIN ERVETYNPK KADFYKKFMK DADANGAYNI	ILDEDLKRAG DNNEEKEFQK EEQTLVKSFO NMRIFSMILN IDVYNTILGA RDKISFIEPQ YDLNKIYVKN EEKKEKALS AYESCKILHD EQADKEEAFY STLLDGWDDN YRKMEYKLLT CRELIDFFKK RDIETIYIDS LQNIYVKLNG YDIIKDKRFT IDRGERNLIY EDENKSARQS GFKQGRQKFE FKSFKQLGKQ NFTSILYNQE KNNEWDTQKI LFBALIVQMRN ARKGLWII EK	(SEQ ID NO: 158)
WP_035635841.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium ND2006]	MSKLEKFTNC KKLLDRYYLS LRKEIAKAFK FTGFFDNREN EVQEIKEKIL KIKGLNEYIN LEVFRNTLNK IFGEWNVIRD QEYADADLSV AVVAIMKDLL DHIYDAIRNY SKYYLAIMDK KWMAYYNPSE NAYDFNFSET MFQIYNKDFS LKKEELVVHP AINKCPKNIF VEQYSLNEII AGYISQVVHK IDKLNMYMDK TSKIDPSTGF NFSRTDADYI KYGINYQOGD	YLSKTLRFK FINDVLHSIK GNEGYSKLFK MFSEEAKSTS NSDYDVEDFF LYNQKTKQKL NSEIFSSIKK KWNAEYDDIH VEKLKEIIIQ DSVKSFENYI VTQKPYSKDK KYAKCLQKID DIQKIYKNGT EKYKDIAGFY DKSHGTPNLH ANSPIANKNP KINTEVRVLL NNFNGIRIKT ICELVEKYDA KSNPCATGGA VNLLKTKYTS KKWKLYSYGN IRALLCEQSD	AIPVGKTQEN LKNLNNYISL KDIIETILPE IAFRICINENL EGEFFNFVLT PKFKPLYKQV LEKLFKNFDE LKKKAVVTEK KVDEIYKVYK KAFFGEGKET FKLYFQNPQF KDDVNGNYEK FKKGDMENLN DNPKKTTLT KHDDNPYVIG DYHSLLDKKE VIALEDLNSG LKGQITNKF IADSKKFIS RIRIFRNPCK KAFYSSFMA	IDNKRLLED FRKKTRTEKE FLDDKDEIAL TRYISNMIDF QEGIDVYNAI LSDRESLSFY YSSAGIFVK YEDDRRSFK SSEKLFDAF NRDESFGDF MGGWDKDKET INYKLLPGPN DCHKLIDFFK NNHGQIRLSG YDVYDKRFS IDRGERNLLY KERFEARQNW FKNSRVKVEK ESFKSMSTQN FDRIMVPEE NNVFDWEEVC MSLMLQMRNS	EKRAEDYKGV NKELENLEIN VNSFNGFTTA EKVDAIFDKH IGGFVTESGE GEGYTSDEEV GPAISTISKD KIGSFSLEQL VLEKSLKKND VLAYDILLKV DYRATILRYG KMLPKVFFSK DSISRYPKWS GAEFLMRRAS EDQYELHIP IVVVDGKGN TSIENIKELK QVYQKFEKML GFIFYIPAWL DLFEFALDYK LTSAYKELFN ITGRTDVDFL	(SEQ ID NO: 159)

Table 5 - Cas12a orthologs						
	ISPVKNSDGI AEDEKLDKVK	FYDSRNYEAO IAISNKEWLE	ENAILPKNAD YAQTSVKH	ANGAYNIARK	VLWAIQGFKK	
WP_051666128.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium ND2006]	MLKNVGIDRL NKRLLEVEDEK KKTRTEKENK DDKDEIALVN YISNMDIFEK GIDVYNAIIG DRESLSFYGE SAGIFVKNGP DDRRKSFKKI EKLFDADFVL DESFYGDVFL GWDKDKETDY YKLLPGPNKM HKLIDFFKDS ESASKKEVDK HGQIRLSGGA VYKDKRFSED RGERNLLYIV RFEARQNWTS NSRVKVEKQV FKSMSTQNGF RIMYVPEEDL VFDWEEVCLT LMLQMRNSIT GAYNIARKVL	DVEKGRKNMS RAEDYKGVKK ELENLEINLR SFNGFTTAFT VDAIFDKHEV GFVTESGEKI GYTSDEEVLE AISTISKDIF GSFSLEQLQE EKSLKKNDV AYDILLKVDH RATILRYGSK LPKVFFSFKW ISRYPKWSNA LVEEGKLYMF ELFMRRASLK QYELHIPIAI VVDGKGNIVE IENIKELKAG YQKFEKMLID IFYIPAWLTS FEFALDYKNF SAYKELFNKY GRTDVDFLIS WAIGQFKAKE	KLEKFTNCYS LLDRYYLSFI KEIAKAFKGN GFFDNRENMF QEIKEKILNS KGLNEYINLY VFRNTLNKNS GEWNVIRDKW YADADLSVVE VAIMKDLLDS IYDAIRNYVT YYLAIMDKKY MAYYNPSEDI YDFNFSETEK QIYNKDFSDK KEELVHPAN NKCPKNIFKI YISQVVKIC KLNVMVDKKS KIDPSTGFVN SRTDADYIKK GINYQQGDIR PVKNSDGIFY DEKLDKVKIA	LSKTLRFKAI NDVLHSIKLK EGYKSLFKKD SEEAKSTSIA DYDVEDFFEG NQKTKQKLPK EIFSSIKKLE NAEYDDIHLK KLKEII IQKV VKSFENYIKA QKPYSKDKFK AKCLQKIDKD QKIYKNGTFK YKDIAGFYRE SHGTPNLHTM SPIANKNPDN NTEVRVLLKH ELVEKYDAVI NPCATGGALK LLKTKYTSIA WKLYSYGNRI ALLCEQSDKA DSRNYEAOEN ISNKEWLEYA	PVGKTQENID NLNNYISLFR IETILPEFL FRCINENLTR EFFNFVLTQE FKPLYKQVLS KLFKNFDEYS KKAVVTEKYE DEIYKVGSS FFGEGKETNR LYFQNPQFMG DVNGNYEKIN KGD MFNLNDC VEEQGYKVSF YFKLLDFDENN PKKTTTLSYD DDNPYVIGID HSLLDKKEKE ALEDLNSGFK GYQITNKFES DSKKFISSFD RIFRNPCKNN FYSSFMALMS AILPKNADAN QTSVKH	(SEQ ID NO: 160)
WP_015504779.1 type V CRISPR-associated protein Cpf1 [Candidatus Methanomethylphilus alvus]	MDAKEFTGQY ELLDDNHRAF KRRKEISAYL SVYTYGYHES DIISEVSGNL QAFNVVLNLR DCICDYVSKI DWDAIETALM NRAEDICRVI FHELEIFSVG TDKIKVNLKF RTSDEDESSF KGMHKSASF FNEDVAGAGY MHTMYWEGLF PRNDVNGRRI KDRRFTVDM ERNLIYVTMV VEGIRKMKEG YQKFESMLIN IFYIPAAFTS KFAFTFDYLD DALQKAGISV SPVKNASGEF AESIRLPLIT	PLSKTLRFEL LNRVLPQIDM QDADGYKGLF RENIYSDEDM GVDDIGKYFD HQKDPGFKEI EKSETVERAL HSSSENDKK SETAPFINDL DEFPKCAAFY PTLADGWDLN EKMAYKLLPS DLGFCHELID YMSLRKIPCS SPQNLESPVF PDSIYRELTR MFHVPIAMNF DRKGNILYQD YLSLAVSKLA KLGVMVLKDK KIDPTTGAFD YNVKSECGRT EGDLRDRIAE FCSKNAGKSL NKAWLTFMQS	RPIGRTWDLN DWHPIAEAFK AKPALDEAMK VSVAYRITED VSNYNNFLSQ QFKQLYKQIL KLVRNISSFD SVYDSAEAFK RAVDLDSLND SELEEVSEQL KERDNKAIL PVKMLPKIFV YYKRCIAEYP EVYRLLDEKS KLSGGAELEFF YFNRGDCRIS KAISKPNLNK SLNINLGYDY DMIIENNAII SIDQSGGALH LFALSNVKNV LWTVYTVGER SDGDTLKSIF PQSDANGAY GMKTWKN	EASGYLAEDR KVHKNPGNKE IAKENGNESD NFPRFVSNAL AGIDDYNHII SVRTSKSYIP LRGIFVNKKK LDDIFSSVKK DGEEAAVSKI IEIIPLFNKA RKDGKYLLAI KSKAAKEKYG GWDVDFKFR IYLFQIYNKD RKSSIPNDK DEAKSYLDKV KVIDGIIDDQ RKALDVREYD VMEDLNHGFK GYQLANHVTT ASMREFFSKM FTYSRVNREY YAFKYALDMR NIALKGILQL	HRAECYPRAK LAQDYNLQLS IEVLEAFNGF IFDKLNESH GGHTTEDGLI KQFDNSKEMV LRILSNKLIG FSDASAEDIG RESLEPYMDL RSFCTRKRYS LDMKKDLSSI LTDRLMECYD ETSDYGSMEKE YSENAHGKNK TVHPKGSVLV KTKKADHDIV DLKIIIGIDRG NKEARNWTK AGRSKIEKQV LASVGKQCGV KSVIYDKAEG VRKVPTDIIY VENREEDYIQ RMLSEQYDPN	(SEQ ID NO: 161)
WP_044910713.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae]	MGLYDGFVNR KRLIDEYHKD RGEIASFLTG FTGFFTNRN IAEFESNFKA KIQGLNELIN VINAVKAVTD	YSVSKTLRFE YISRCLKNVN NDEYKEQKSR MYSAEKKSTA ITDEHEVSDL LHNQKHGPKK TFSETLLGSL	LIPQGR TREY ISCLEEYHL DIIINERIN IAHRIIDVNL LDITKYSRFI VPLLKVLYKQ KIIINNIGHY	IETNGILSDD YNSSNRDKRH FASTDEELAA PKYVDNIKAF RNEDIIINYNT ILGDSQTHSF DLDRIYIKAG	EERAKDYKTI EELDALSQDM VKRFRKFTSY NTAIEAGVFD LLGGISMKDE VDDQFEDDQ QDITTL SKRA	(SEQ ID NO: 162)

Table 5 - Cas12a orthologs						
e bacterium MC2017]	LNDWHIITEC VTTYTEQACF LITDYESVRI KLLPVIKVYN ILRKDGQYYL FFSKSNIAFF EEWRKFNFKF KLYLFKLHNK FRKASIPENE FQGHVPLKMN INEEGRIIEQ KDLKTGYLSQ EKMLIDKLNY PAWNTSKLDP LHYSNFTERA DLFEKNGIDV DYMISPAWND TGNNVKMAMS	LESEYDDKFP LENYLGSFGG LKRLLDSEME RVRNYLTRKP GIMTPSNKKI NPSDEILRIK SDTDSYEDIS DFSEHSGKGP RVIHKKNVPI FKADEKKRIN RSLNIIDSGE AIHTITKWMK YVDKDEEFDR TTGFVDLLYP IGCRDEWILC KQENLIDSIC RGEFFDTRKG NREWMHFAQE	KNKKSPTYEE DTDNCLTDF VIHFLKPLL FSTEKIKLNF FSEAPKPED KQESFKKGKS RFYKEVENQA NLHTVYWNAL TRKVAELNGK DDVIEAIRSN GHTQNYRDL EYNAIIVLED MGGTHRALQL KYKSVDATKD SYGNRIVNFR EMKDKPFFKS LQQLPLDADA SRLA	MNRNVKSKF TNSLMEVEHL NGNESDKDLV NSPTLLCGWS CYEKMVLRYI FNRDDCHKFI FMSFTKIPT FSEYNLQNTV KEVSVFPYDI KGIHVIDR DSREKDREKA LNDRFTNGRK TEKFESFQKL FIKKFDFIRF NAAKNNSWDY LIANIKLILQ NGAYNIAKKG	SFSIGRLNSL LNSEYPVTNR FYGEFEAEYE QSKEKEYMGV PHPYQMLPKV DFYKDSINRH VYIDSLVDEG YQLNGSAEIF IKNRRYTVDK GERNLLYLSL RENWQEIQEI KVEKQVYQKF GRQTGFIFYV NSEKNYFEFG KEIDITKQLL IRNSASGTDI LWIVDQIRNT	
KKQ36153.1 hypothetical protein US52_C0007G0 008 [candidate division WS6 bacterium GW2011_GWA 2_37_6]	MKNVFGGFTN EMKPILDEIH QVKLLQKEIT TEQAVLDILK ANRIVNENLS DKYNIIGGY KGFGIFEIKE EIKRDSNSQA GEKKIPEQIS TFLNIWQNEI LAIERMAKYH PYNKSKMKLN RTDKINNEIY SVPSLQKFIR INIAENKLE CKSPIIGLNG SITINYGPKP ILKQGSLNKI LSQAIHEIYQ KLNHLILKDR NYTSTDPVT WGFYRVQKET DKSADIYSQ YPFFDSRKT TEWDAWIONF	LYSLTKTLRF RRFINDALAQ IKTLGLRREV IEFPNKAKYI RYIDNIIAFE IVDKDKKIQQ GKEWEQLGDL RYSLDKIYFN IEDIKDSLIS EESFNKLEEK LPKDSNLSRE FENGILLGGW RTGSSKWERL ENYLKKYPQL AEDKGDLYLF GAEIFFREGQ NIKFRDIINQ RVGDKEVDNF LMIKYNAIIV NTNEIGGVVK GWRKHLYISN GKKWELFATK CNVLDLFRWKS GKTEPINGDA NKNV	ELKPTSSTQK KIFISASLDN VSGFITVSKK DKFRGFWTYF EILQKIPNLK INEKVNLYTQ FKLQRTKINS KASVNTISNS IPKENLEELF EKDFKKNAAI EDFYWIIDNL SDGQERNKAG ILSNLKFQTL TEVSNTKFLS EITNKDFSGK KDKLHTKLDK LITSMNVNII KKLTERANEM LEDLNTFKA AYQLTPTIGG FSNNSVIKSF ELERFKYNQT LVYLLWNLNQ NGALNIARKG	LMKRNNVIQT FLKVVKYKQV WKDKYVGLGI SGFNENRKNY KFKQDLIDTS QTKKKLPKLK NGREKGLFDS WFTNWNKFAE KLTNREKHDR KFQKNNLVQK SEQREIYKYY VILRNGNKYY AGKGFLGKHG KKDFDAAIKE KSGKDNHTI KGKKVFDAGR GIDRGEKHL KKARQSWEQI KRLSKVEKSV GDVSKFEKAK FDPTNRDTGI TKLCEKINLY IRNVDKNAEG LMLVERIKNN	DEEIDKLYHD ESAKNIKQON KLKGDGYKVL YSEEDKATSI YNYLNLQAGI FLFKQIGSER LRTMYREFFD LLNIKEDKKN TRFFGSNAWV NYIKEVCDRM NAFRNYISKK LGVLINRGIF VSYGNMNPKE ALKECFTMNF YWKYLFSES YSEDKLFFHV YYSVIDSNGI GNIKNFKEGY YKKFELKLAR QWGMFFYVRA EIFYSGKYRS DKFEELFKGI NKNDFIQSPV PEKYEQLIRD	(SEQ ID NO: 163)
WP_044919442. 1 type V CRISPR- associated protein Cpf1 [Lachnospiracea e bacterium MA2020]	MYEESLTKQY GILDEYHKQL RKEVVEKLKA TSYKNVREN GLGEEQDSL GFRKIPKMKM SKVSAAFFDAL ANENKKKDDK DIENIIINNE KLINSSGQEL VKLNFNRSTL VAKTEKVFKK HKKGNMFSLE EVEKQGYKLT LYFMMLFDQR PNRARTKETS	PVSKTIRNEL INEALDNCTL HENFTKIGKK YSDKEKSSTV FIVETFNKTL LYKQILSDRE RESKGNVYV YFEKRQKELK TFLRIVINEH EKDLIVYSAH LNGWDRNKET VDYKLLPVPN DCHNLIDFFK YTDIDETYIN NIDDDVYKLN TFSYDIVKDK	IPIGKTLTDLN PSLKIAAEIY DILDLEKLP AYRLINENFP TQDGIDTYNS ESFIDEFQSD KNDLAKTAMS KNKSYSLEHL DRSRKLAKNR EELLVELKQV DNLGVLLKLD QMLPKVFFAK ESISKHEDWS DLIERNELYL GEAEVYRPA RYSKDKFTLH	RQNNILES LKNQKEVSDR SISEDDYNAL KFLDNVKS QVGKINSIN EVLIDNVESY NIVFENWRTF CNLSEDS KAVKAIKDFL DSLYNMTRNY GKYYLGIMNT SNIDFYNPSS KFGFKFSDTA FQIYNKDFSM SISEDELIH IPITMNF	KRKQNYEHVK EDFNKTQDLL ESFRNFYTYF FVKTAGILAD LYNQKNQKAN GSVLIESLKS DDLNLQEYDL IENYIHQISD DSIKVLEREL LTKKPFSTEK SANKAFVNPP EIYSNYKKGT SYNDISEFYR YSKGLNLHT KAGEEIKNNK EVKREND	(SEQ ID NO: 164)

Table 5 - Cas12a orthologs						
	SAIRIDENVN TDYHALLDER AIICLEDLNF ELGGALNALQ MKCENVEKSK CTNGERIICKY IISNEAEAFY LSDGSVPKDA AQTHLL	VIGIDRGERN EGGRDKARKD GFKRGRQKVE LTSKFKSFE RFFDGFDFIR RNPDKNNMFD RRLYRLQQT DANGAYNIAR	LLYVVVIDSK WNTVENIRDL KQVYQKFEKM LGKQSGVIYY FNALENVFEF EKVVVVTDDEM LQMRNSTSDG KGLWVLEQIR	GNILEQISLN KAGYLSQVNN LIDKLNLYVI VPAYLTSKID GFDYRSFTQR KNLFEQYKIP TRDYIISPVK QKSEGEKINL	SIINKEYDIE VVAKLVLKYN DKSREQTSFK PTTGFANLFY ACGINSKWT YEDGRNVKDM NKREAYFNSE AMTNAEWLEY	
WP_035798880.1 type V CRISPR-associated protein Cpf1 [Butyrivibrio sp. NC3005]	MYYNLTCKY GIMDEYHKQL RREVTGRLKE TSYNKVREN CIEEEQDAL EFKKIPKMKV EKINIFFDAL ANENKKDDK DIEKICIYNG KLINGSGQEL VKLNENKSTL AAKTENVFKK HKKGPSFSID EVEKQGYKLT LYFMMLFDQR PNRAKVKETS NALRTDDNVN TNYHALLDER AIICLEDLNF KMGGALNALQ IKYENIEKAK YTNGERIICKY IISKAEADFY FSEGTMPKDA AQLHLL	PVSKTIRNEL INEALDNYML HENYTKIGKK YSDEEKSSTV FMVETFNMTL LYKQILSDRE RESEGKNVYV YFEKRQKELK EFEKIVVNEH EKNLVVYVGQ LNGWDKNKET VDYKLLPGSN DCHNLIDFFK FTDIDESYIN NLDNVVYKLN TFSYDIVKDK VIGIDRGERN EDDRNKARKD GFKRGRQKVE LTSKFKSFAE QFFDGFDFIR PNPEKNNLFD KRLFRLLHQT DANGAYNIAR	IPIGKTLENI PSLNQAAEY DILDLEKLP AYRLINENLP TQEGIDMYNY EVFIGEFKDD KNDLSKTMS KNKSYTLEQM DSSRKLKSKNI EEALEQLRPV DNLGILFFKD KMLPKVFFAK ESIKKHEDWS DLIEKNELYL GEAEVYRPA RYSKYKFTLH LLYVVVINSE WNTIENIKEL KQVYQKFEKM LGKQSGIYY FNKKDDMFEF EKVINVTDEI LQMRNSTSDG KGLWVLEQIR	RKNNILES DV LKKHVDVEDR SISEEDYNAL KFLDNIKSYA QIGKVN SAIN ETLLSSIGAY NIVFGSWSAF SNLSKEDISP KAVKVIKDYL DSLNLTRNY GKYLLGIMNT SNIGYNNPST KFGFEFSDTA FQIYNKDFSE SIAENELVIH IPITMNFVD GKILEQISLN KTGYLSQVNN LIEKLNLYVI VPAYLTSKID SFDYKSFTQK KGLFKQYRIP TRDYIISPVK QKDEGEKVNL	KRKQDYEHVK EEFKKTQDLL ESFRNFYTYF FVKAAGVLAD LYNQKNHKVE GNVLMTYLKS DELLNQEYDL IENYIERISL DSIKELEHDI LTKKPFSTEK TANKAFVNP ELYSNYKKG DYRDISEFYR YSKGLNLHT KAGEGIKNKN EVRRENDVIN SIINKEYDIE VVAKLVLKYN DKSREQVSPE PTTGFVNLFY ACGIRSKWIV YENGEDIKEI NDRGEFFCSE SMTNAEWLKY	(SEQ ID NO: 165)
WP_027109509.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium NC2008]	MENYDSLTR VKELMDEFHK YMRQIADAL YFTDYKNRK NDNYDEETLD EENFVRVPMK LPINIDRYIK YDAKYTGKVN DIKEICEAYK KLLNGTGLES KLNFNPNQLL SKNENAYKKV KKGNSFNLED VEAQNYKITY YLKMLFDERN NSKKKESIFS KSKSNIRVIG LLDNKEKERL EDLNIGFKNG LNAYQLTAKF KQGLQLFFRK RIAKYRSDDA KSAVLKEIYY SKELPCDADA NNIPSL	QYPVTKTIRQ SIEKSLVGI KNSRDYGYVD NIYSDEEKHS KTFMIESFNE KVLFKQILSD TLMDSNNGTG DKYEEKREKA EMTKNCFEDH RNEFFYGEQS GGWDVNKERD NCKLLPGPNK CHKLIDFYKE TNVSCDFIES LKDLCIKMNG YDIVKDKRYT IDRGERNLLY SSRRDWKTIE RKKVEKQVYQ ESFEKIGKQT FKKINFDKVA NGKWISRTVH GFKLTQLLRN NGAYNIARKG	ELKPVGKTLE KLDGLSEFEK NKDLIEKILP TVGYRCINEN CLTQSGVEAY RTSLFDGLII IYVKNDSSLT YKSNDSEFELN DKTKKLQKNI TVLEEITKVD CYGVILIKDN MFPKVFFAKS SIKKNDWKN LVDEGLYLF EAEVYRPA VDKFFIHLPI VVVCDSDGSI NIKDLKAGYM NFEKALIDKL GCIFYVPAYL SNFEFVFDYN PTDIIKEALN STLANEEQE LWALEQIRNS	NIKNAEIEA ERVKDENSELN LLIFMSNIEV NSVVASIKTA ESDDELDDKL TLSNYLTD SW YIQNLLGINV KAVASIKSYL ELYNITRNYL NYYLGIMDKS NIDYYDP THE FNFNFSDTKD QIYNKDFSEY ILDEDKV VHK TLNYKEQNV S LYQRSINEIV SQVVNEIYNL NYLCIDKTRE TSQIDPTTGF DFTNKAEGTK REKINYNDGH DYIISPVKNS ENVSKVKLAI	DKQKKEAYVK DKNRISELYF ALSCFKGFTT YQIYKKANIK TNLYIQKNNK CSFSAEVDKF SSIRNAFNEN IDKYIERINF DSLKNIERDI TKKPFSTEKM ANKSFLNIKE IKKLYDKGTF YEDISGFFRE ATGNLNLHTL ANQKITNKNT RFNDYIREIL SGSHKT DYHK ILKYNAIVVL QLSPSSPGGV VNLFYQKDT S TNWTISTQGT DLIDEIVSIE SGNYFDSRIT SNKEWF EYTO	(SEQ ID NO: 166)

Table 5 - Cas12a orthologs						
WP_049895985.1 type V CRISPR-associated protein Cpf1 [Oribacterium sp. NK2B42] WP_029202018	METEILKYDF RILEADIKRK QNDERDIDAF DRILIESFDN IKAYRTVRNA NSAINLYNQK SVHNYVSYLI NWSTIDDLIN CRDRTILEKY AIKNMLDSIK NLTRNYLTCK LGIMNTSSNK YYAPSEELLS NFSDTNSYND NKDFSPYSKG DEQIIHKSQ VNFGTNESGK EQISLNTIVD LSQVNNIIAK MNYLVLDKSR LTSKIDPTTG KSFTERASGC DKYLQNYIDE IISPVKNYRE GTKINLAMSN	FEREGKYYMY SDYEHVKKLM LKIQDKLRKE FYTYFTSYNK GVISGDMSIV MHGAGSFKKL DYLNSDKVES HEYDSAPENV IRRLVDDIEK DFEHDVMLIN PFSTEKIKLN AFVNPPKAIS KYRKGTHKKG ISDFYREVEK KLNHLHTLYFK NIKNNQKRS FNELVNNAIR KEYDIETDYH LVLKYDAIIC KQESPQKPGG FANLFYIKYE KSKWIACTNG DDLKDQILQI EFFCSEFSDD SEWLEYAQCN	DGLTKQYALS DMYHKKIINE IVEQLKGHTD VRSNLYSAED EQDELFEVDT PKMKELYKQL FFDTLRKSDG KKTKDDKYFE IYTSNNVFS GSGQEIKKNL FNRPTFLDGW NDIYKKVDYK DSFSIDDCRN QGYKLSFTDI MLFDQRNLDN NCKKTSTFDY ADKDVNVIGI QLLDEKEGSR LEDLNFQFKR ALNALQLTSA SVDKARDFFS ERIVKYRNSD DSADFYKNLI TFPRDADANG LL	KTIRNELVPI ALDNFKLSVL YSKVGNKDFL KSSTVAYRLI FNHTLTQYGI LTEREEEFIE KEVFIKNDVS KRQKDLKKNK IVLSKHDRSK NVYSEQEALA DKNKEEANLG LLPGPNKMLP LIDFFKSSIN DACYIKDLVD VVYKLNGEAE DKARKDWNTI GRQKVEKQVY FKSFKELGKQ KFDFIRYNQM KNNSFDDKT KLFQLTQLQMR AYNIARKGLW	GKTLDNIKKN EDAADIYFNK GLLKAASTE NENLPKFFDN DTYNHMIGQL EYTDDEVLT KTTMSNILFD SYSLSKIAAL KLSKNTNAVQ GILRQVDHIY ILLIKDNRY KVFFATKNIA KNTDWSTFGF NNELYLFQIY VFYRPASLIE DKFMLHLPI VVVDPCGKII ENIKELKEGY QKFEKMLIDK TGIYYVPAY DNYFEFGFDY ILTDEYRSLF NSSSDGKR VIKQIRETKS	(SEQ ID NO: 167)
WP_028248456.1 type V CRISPR-associated protein Cpf1 [Pseudobutyrio ruminis]	MYYNLTMY KLMDNYHKQL RKEIVSLLKN SYNEVRKNLY LSEEDQDCLF FRKVPQMKCL IITDFKIALV NSKKKKDEKY ITATAEAKND LINGSGQVEE KLNLVVYAEQ KLNFRATLL ATSNVYHKV KKGDNFSLD VEKQGYKITY YLQMLFDQRN KRAVDKPTSK ALRNDEKVRV DYHKLLDEKE IICLEDLNF FGGALNALQL KYENVEKAKE TYGERIIKFR MSVDEADFFR CDASKPKDAD QRNQI	PISKTLRNL INEALQGVHL HENFPKIGNK SDEEKSSTVA IITETFERTLT YKQILSDREE ESDGLSVYIK YEKRQKELKK FDEIVLRKHD KNLAEIKNVD NLGILFEKDG MLPKVFFAQ SIEKHPDWSS LVERDELYLF EAEVFYRPAS YKDKFMLHI LYVVVVDTDG GDRDRARKNW FKRGRQKVEK TSKFTSFKDM FFSRFDSISY NTEKNNSFDD SLTRLFQQTM ANGAFNIARK	IPVGKTLENI SDLSDAYDLY EIIKLLQSLY YRLINENLPK QDGIDNYNAA AFIDEFSDE NDVSKTSFSN NKSVDLETII KNKSLRKNTN NILAEIKNVD NLGILFEKDG MLPKVFFAQ SIEKHPDWSS LVERDELYLF EAEVFYRPAS YKDKFMLHI LYVVVVDTDG GDRDRARKNW FKRGRQKVEK TSKFTSFKDM FFSRFDSISY NTEKNNSFDD SLTRLFQQTM GLWVLEQIRN	RKNGILEADI FNLSKEKNSV DNDTDYKALD FLDNIAKAYAI IGKLNNTAINL DLITNIESFA IVFGSWNAID GLFDDNSDVI AVEAIIKSYLD SLYNSMRNYL MYYLGIMNTK NLDYYPSEE FGFEFSETCT QIYNKDFSPY INDEEVIHK PVTMNFVDE TILEQISLNS EGYLSQVVNV IDKLNLYLVID PAYLTSKIDP FDYKKFTDRA ELFSIYGISY RDYIISPIMN TPSGDKLNLA	QRKADYEHVK DAFSKCQDKL SFSNFYTYFS AKKAGVRAEG FNQONKKQEG ENMNVLNSE EKLSDYDLA GKYIEKLESD TVKDFERDIK TQKPFSTEFK ANKIFVNIPK LLAKYKAGTH YEDLSGFYRE SKGNLNLHTI AGEEIKNNKS TRRFNDVVND IINNEYSIET IAKLVLKYNA KSRKQDKPEE TTGFANLFYV CGARSQWTV EDGAELKNKI DRGEFFNSEA MSNAEWLEYA	(SEQ ID NO: 168)
WP_028830240 type V CRISPR-associated protein Cpf1 [Proteocatella sp. phenisci]	MENFKNLYPI IDEKFKETIE DEIFNNYLKP ESSGSMAYRI FITQSGITHY KMILSDRVSN KYYQLGNLPG SINYISELLT EKTNLIKDLL	NKTLRFELRP ERLKYTEFSE DKNIDALFKN IDENLTYYLN NEIIGGISKS SFVLDTIEND ISYSSIVNAI DTDVSSNIKM DILKSIQRFY	YGKTLENFKK CDLGNMTSKD DPSNPVISTF NIEKIKKLPE ENVKIQGINE TELIEMISDL CSDYDNNFGD RYKELEQNYQ DLFDIVDEDK	SGLLEKDAFK KKITDKAATN KGFTTYFVNF ELKSQLEGID GINLYCQKNK INKTEISQDV GKRKKSSEND VCKENFNATN NPSAEFYTWL	ANSRRSMQAI LKKQVILSFD FEIRKHIFKG QIDKLNNYNE VKLPRLTPLY IMSDIQNIFI RKKHLETNVY WNKNIKQKS SKNAEKLDFFE	(SEQ ID NO: 169)

Table 5 - Cas12a orthologs						
	FNSVYNKSRN FNNDRGDYDY KQFLSKIWKA NHDEKYQDVF DGKLYVFQIW IFYRPASLNY INYKSEKLNS VEQKHLDEII KEGYISHVIN LIKKNFYIID IDPVTGFVNL NNRYSISKTK GTLKSQDVET RENIKNLPAD NQQE	YLTRKQYSDK FLGIWNKSTP KHPTTPEFDK GFNLNRTEDY SKDFSIDSKG CEDIIKKGHH KSLNNRTNEN NTDTKGVEHK EIQKLQEKYN KKDPETYIHG LYADDLKYKN WTLTSYGTRI YKKFMSLFKL ADANGAYNIA	KIKLNFDSPT ANEKIIILED KYKEGRHKKG NSYTEFLEDV TKNLNTIYFE HAELKDKFDY LGQFTHIIGI YQLTNPITTL QEQAQSFQK QTFRNPQKNN MLQLRNSVTG RKGIMAIENI	LAKGWDANKE NGLFEKMQYK PDFEKEFLHE ERCNYNLSFN SLFSEENMIE PIIKDKRYSQ DRGERHLYL IDNIIYFENGE KWDSAEYDLT TDIDYMISPV MNGISDPLKI	IDNSTIIMRK LYPDPSKMLP LIDCFKHGLV KIADTSNLIN KMFKLSGEAE DKFFFHVPMV TVVDVSTGEI WEAETIKEL KQVYQKFETA LYIPAWNTSK FKFDIDFSKW EEFKLILNID TDKTGTHFDS SNEDYLKYIQ	
WP_084502895.1 type V CRISPR-associated protein Cpf1 [Proteocatella sphenisci]	MIILYISTSN LEKDAFKANS TDKAATNLKK TTYFVNFFEI SQLEGIDQID LYCQKNKVKL TEISQDVIMS KKSSENDK ENFNATNWMN AEFYTWLSKN GWDANKEIDN FEKMQYKLYP EKEFLHELID NYNLSFNKIA SEENMIEKMF KDKRYSQDKF ERHLYLTVV RDNERKSWEA NSRIKVEKQV GNQSGIVLYI IYFENGFEKF SAEYDLTEEF DYMISPVTDK ISDPLKISNE	MNMEGVFMEN RRSMQAIIDE QVILSFDDEI RKHIFKGESS KLNNYNEFIT PRLTPLYKMI DIQNIFIYK HLETNVYSIN IKNIKQSEKT AEKLDFEFNS STIIMRKFN DPSKMLPKQF CFKHGLVNHD DTSNLINDGK KLSGEAEIFY FFHVPMVINY DVSTGEIVEQ IETIKELKEG YQKFETALIK PAWNTSKIDP DIDFSKWNNR KLILNIDGTL TGTHFDSREN DYLKYIQNQQ	FKNLYPINKT KFKETIEERL FNNYLKPDKN GSMAYRIIDE QSGITHYNEI LSDRVSNSFV QLGNLPGISY YISELLTDTD NLIKDLLDIL VYNKSRNYLT DRGDYDYFLG LSKIWKAKHP EKYQDVGFEN LYVFIWQSKD RPASLNYCED KSEKLSKSL KHLDEIINTD YISHVINEIQ KFNYIIDKKD VTGTFVNLLYA YSISKTKWTL KSQDVETYYK IKNLPADADA E	LRFELRPYGK KYTEFSECDL IDALFKNDPS NLTTYLNIE IGGISKSENV LDTIENDTEL SSIVNAICSD VSSNIKMYK KSIQRFYDLF RKQYSDKKIK IWNKSTPANE TTPEFDKQYK LRNTEDYNSY FSIDSKGTKN IIEKKGHHAE NNRTNENLQ TKGVEHKTHY KLQEKYNALI PETYIHGYQL DDLKYKNQEQ TSYGTRIQT FMSLFKLMLQ NGAYNIARKG	TLENFKKSGL GNMTSKDKKI NPVISTFKGF KIKKLPEELK KIQGINEGIN IEMISDLINK YDNNFGDGKR ELEQNYQVCK DIVDEKKNPS LNFDSPTLAK KIIPLEDNGL EGRHKKGPDF TEFLEDVERC LNTIYFESLF LKDKFDYPII FTHIIGIDRG LNKLEEKSKT VMENLNYGFK TNPITTLDKI AKSFIQKIDN RNPQKNNKWD LRNSVTGTDI IMAIENIMNG	(SEQ ID NO: 170)
WP_055225123.1 Eubacterium rectale	MNNGTNNFQN RQILKDIMDD EQTEYRKAIH TQVIKLFSRF LVYRRIVKSL FYNDICGKVN ESDEEVYQSV VSQKTYRDWE NELVSNYKLC ASELKNVLDV YNLVRNYVTQ YLGIFNAKKN TGVETYPKSA KNFGFDFSDT LFQIYNKDFS SSIKNPIIHK NDKSDKELSD ANKTGFINDR FNTVNGYDYQ MVIKYNALIA	FIGISSLQKT YYRGFISETL KKFANDDRFK ATSEKDYFKN SNDDINKISG SFMNLYCQKN NGFLDNISSK TINTALEIHY SDDNIKAETY IMNAFHWCVS KPYSTKKIKL PDKKIEGNT YILEGYKQNK STYEDISGFY KKSTGNDNLH KGSILVNRTY EAAKLKNVVG ILQYIAKEKD IKLKQQEGAR MEDLSYGFKK	LRNALIPTET SSIDDIDWTS NMFSAKLISD RANCFSADDI DMKDSLKEMS KENKNLYKLQ HIVERLRKIG NNILPGNGKS IHEISHILNN FMTEELVDKD NFGIPTLADG SENKGDYKKM HIKSSKDFDI REVELQGYKI TMYLKNLFSE EAEKDKQFGN HHEAATNIVK LHVIGIDRGE QIARKEWKEI GRFKVERQVY	TQQFIVKNGI LFEKMEIQLK ILPEFVIHNN SSSSCHRIVN LEEIYSYEKY KLHKQILCIA DNYNGYNLDK KADKVKKAVK FEAQELKYNP NNFYAELEEI WSKSKEYSNN IYNLLPGPNK TFCHDLIDYF DWTYISEKDI ENLKDIVLKL IQIVRKNIP DYRYTYDKYF RNLIYVSVID GKIKEIKEGY QKFETMLINK	IKEDELGRGEN NGDNKDTLIK NYSASEKEEK DNAEIFFSNA GEFITQEGIS DTSYEVYPYK IYIVSKFYES NDLQKSITEI EIHLEVESELK YDEIYPVISL AIILMRDNL MIPKVFLSSK KNCIAIHPEW DLLQEKGQLY NGEAEIFFRK NIYQELYKYF LHMPITINFK TCGNIVEQKS LSLVIHEISK LNYLVFKDIS	(SEQ ID NO: 171)

Table 5 - Cas12a orthologs						
	ITENGGLK	YQLTYPDKL	KNVGHQCGCI	FYVPAAYTSK	IDPTTGFVNI	
	FKFKDLTVDA	KREFIKKFDS	IRYDSEKNLF	CFTFDYNNFI	TQNTVMSKSS	
	WSVYTYGVRI	KRRFVNGRFS	NESDTIDITK	DMEKTLEMTD	INWRDGHDLR	
	QDIIDYEIVQ	HIFEIFRLTV	QMRNSLSELE	DRDYDRLISP	VLNENNIFYD	
	SAKAGDALPK	DADANGAYCI	ALKGLYEIKQ	ITENWKEDGK	FSRDKLKISN	
	KDWFDFIQNK	RYL				
WP_055237260.1 Eubacterium rectale	MNNGTNNFQN	FIGISSLQKT	LRNALIPTET	TQQFIVKNGI	IKEDELRGEN	(SEQ ID NO: 172)
	RQILKDIMDD	YRGFISETL	SSIDDIDWTS	LFEKMEIQLK	NGDNKDTLIK	
	EQAEKRKAII	KKFADDDRFR	NMFSAKLISD	ILPEFVIHNN	NYSASEKEEK	
	TQVIKLFSTR	ATSFKDYFKN	RANCFSAADD	SSSSCHRIVN	DNAEIFFSNA	
	LVYRRIVKNL	SNDDINKISG	DMKDSLKEMS	LDEIYSYEKY	GEFITQEGIS	
	FYNDICGKVN	SFMNLYCQKN	KENKNLYKLR	KLHKQILCIA	DTSYEVYPKF	
	ESDEEVYQSV	NGFLDNISSK	HIVERLRKIG	DNNGYNLDK	IYIVSRFYES	
	VSQKTYRDWE	TINTALEIHY	NNILPGNGKS	KADKVKKAVK	NDLQKSITEI	
	NELVSNYKLC	PDDNIKAETY	IHEISHILNN	FEAQELKYNP	EIHLVESELK	
	ASELKNVLDV	IMNAFHWCVS	FMTEELVDKD	NNFYAELEEI	YDEIYPVISL	
	YNLVRNYVTQ	KPYSTKKIKL	NFGIPTLADG	WSKSKEYSNN	AIILMRDNL	
	YLGIFNAKNK	PKKIIIEGNT	SENKGDYKKM	IYNLLPGPNK	MIPKVFLSSK	
	TGVETYPKPSA	YILEGYKQNK	HLKSSKDFDI	TFCRDLIDYF	KNCIAIHPEW	
	KNFGFDFSDT	STYEDISGFY	REVELQGYKI	DWTYISEKDI	DLQEKQGLY	
	LFQIYNKDFS	KKSTGNDNLH	TMYLKNLFSE	ENLKDVLKL	NGEAEIFFRK	
	SSIKNPIIHK	KGSILVNRTY	EAEKDKQFGN	IQIVRKTIPE	NIYQELYKYF	
	NDKSDKELSD	EAAKLKNVVG	HHEAATNIVK	DYRYTYDKYF	LHMPITINFK	
	ANKTSFINDR	ILQYIAKEND	LHVIGIDRGE	RNLIYVSVID	TCGNIVEQKS	
	FNIVNGYDYQ	IKLKQQEGAR	QIARKEWKEI	GKIKEIKEGY	LSLVIHEISK	
	MVIKYNALIA	MEDLSYGFKK	GRFKVERQVY	QKFETMLINK	LNLYVFKDIS	
	ITENGGLK	YQLTYPDKL	KNVGHQCGCI	FYVPAAYTSK	IDPTTGFVNI	
	FKFKDLTVDA	KREFIKKFDS	IRYDSEKNLF	CFTFDYNNFI	TQNTVMSKSS	
	WSVYTYGVRI	KRRFVNGRFS	NESDTIDITK	DMEKTLEMTD	INWRDGHDLR	
	QDIIDYEIVQ	HIFEIFRLTV	QMRNSLSELE	DRDYDRLISP	VLNENNIFYD	
	SAKAGDALPK	DADANGAYCI	ALKGLYEIKQ	ITENWKEDGK	FSRDKLKISN	
	KDWFDFIQNK	RYL				
WP_055272206.1 Eubacterium rectale	MNNGTNNFQN	FIGISSLQKT	LRNALPTTET	TQQFIVKNGI	IKEDELRGEN	(SEQ ID NO: 173)
	RQILKDIMDD	YRGFISETL	SSIDDIDWTS	LFEKMEIQLK	NGDNKDTLIK	
	EQAEKRKAII	KKFADDDRFR	NMFSAKLISD	ILPEFVIHNN	NYSASEKEEK	
	TQVIKLFSTR	ATSFKDYFKN	RANCFSAADD	SSSSCHRIVN	DNAEIFFSNA	
	LVYRRIVKNL	SNDDINKISG	DMKDSLKKMS	LEKIYSYEKY	GEFITQEGIS	
	FYNDICGKVN	SFMNLYCQKN	KENKNLYKLR	KLHKQILCIA	DTSYEVYPKF	
	ESDEEVYQSV	NGFLDNISSK	HIVERLRKIG	DNNGYNLDK	IYIVSKFYES	
	VSQKTYRDWE	TINTALEIHY	NNILPGNGKS	KADKVKKAVK	NDLQKSITEI	
	NELVSNYKLC	PDDNIKAETY	IHEISHILNN	FEAQELKYNP	EIHLVESELK	
	ASELKNVLDV	IMNAFHWCVS	FMTEELVDKD	NNFYAELEEI	YDEIYPVISL	
	YNLVRNYVTQ	KPYSTKKIKL	NFGIPTLADG	WSKSKEYSNN	AIILMRDNL	
	YLGIFNAKNK	PEKKIIIEGNT	SENKGDYKKM	IYNLLPGPNK	MIPKVFLSSK	
	TGVETYPKPSA	YILEGYKQNK	HLKSSKDFDI	TFCRDLIDYF	KNCIAIHPEW	
	KNFGFDFSDT	STYEDISGFY	REVELQGYKI	DWTYISEKDI	DLQEKQGLY	
	LFQIYNKDFS	KKSTGNDNLH	TMYLKNLFSE	ENLKDVLKL	NGEAEIFFRK	
	SSIKNPIIHK	KGSILVNRTY	EAEKDKQFGN	IQIVRKTIPE	NIYQELYKYF	
	NDKSDKELSD	EAAKLKNVVG	HHEAATNIVK	DYRYTYDKYF	LHMPITINFK	
	ANKTSFINDR	ILQYIAKEND	LHVIGIDRGE	RNLIYVSVID	TCGNIVEQKS	
	FNIVNGYDYQ	IKLKQQEGAR	QIARKEWKEI	GKIKEIKEGY	LSLVIHEISK	
	MVIKYNALIA	MEDLSYGFKK	GRFKVERQVY	QKFETMLINK	LNLYVFKDIS	
	ITENGGLK	YQLTYPDKL	KNVGHQCGCI	FYVPAAYTSK	IDPTTGFVNI	
	FKFKDLTVDA	KREFIKKFDS	IRYDSEKNLF	CFTFDYNNFI	TQNTVMSKSS	
	WSVYTYGVRI	KRRFVNGRFS	NESDTIDITK	DMEKTLEMTD	INWRDGHDLR	
	QDIIDYEIVQ	HIFEIFRLTV	QMRNSLSELE	DRDYDRLISP	VLNENNIFYD	
	SAKAGDALPK	DADANGAYCI	ALKGLYEIKQ	ITENWKEDGK	FSRDKLKISN	
	KDWFDFIQNK	RYL				

Table 5 - Cas12a orthologs						
OLA16049.1 Eubacterium sp. 41_20	MNNGTNNFQN RQILKDIMDD EQAEKRKAII TQVIKLFSTRF LVYRRIVKNL FYNDICGKVN ESDEEVYQSV VSQKTYRDWE NELVSNYKLC ASELKNVLDI YNLVRNYVTQ YLGIFNAKNK TGVETYPKPSA KNFGFDFSDT LFQIYNKDFS SSIKNPIIHK NDKSDKELSD ANKTSFINDR FNI VNGYDYQ MVIKYNIAIA ITENGGLLKG FKFKDLTVDA WSVYTYGVRI QDIIDYEIVQ SAKAGYALPK KDWFDFIQNK	FIGISSLQKT YYRGFISETL KKFADDDRFRK ATSFKDIFYKN SNDDINKISG SFMNLYCQKN NGFLDNISSE TINTALEIHY SDDNIKAETY IMNAFHWCSV KPYSTKKIKL PDKKIIIEGNT YILEGYKQNK SAYEDISGFY KKSTGNDNLH KGSILVNRTY EAAKLKNVVG ILQYIAKEKD IKLKQQEGAR MEDLSYGFKK YQLTYIPDKL KREFIKKFDS KRRFVNGRFS HIFEIFKLT DADANGAYCI RYL	LRNALIPTET SSIDDIDWTS NMFSAKLISD RANCFSAADDI DMKDSLKEMS KENKNLYKLR HIVERLRKIG NNILPGNGKS IHEISHILNN FMTEELVDKD NFGIPTLADG SENKGDYKKM HLKSSKDFDI REVELQGYKI TMYLKNLFSE EAEKDKQFNG LHVIGIDRGE QIARKEWKEI GRFKVERQVY KNVGHQCGCI IRYDSEKNLF NESDTIDITK QMRNSLSELE ALKGLYEIKQ	TQQFIVKNGI LFEKMEIQLK ILPEFVIHNN SSSSCHRIVN LEEIYSYEKY KLHKQILCIA DNYNDYNLDK KADKVKKAVK FEAHELKYNP NNFYAELEEI WSKSKEYSNN IYNLLPGPNK TFCHDLIDYF DWTYISEKDI ENLKDIVLKL IQIVRKTIPE RNLIYVSVID GKIKEIKEGY QKFETMLINK FYVPAAYTSK CFTFDYNNFI DMEKTLEMTD DRDYDRLISP ITENWKEDGK	IKEDELRGKN NGDNKDTLIK NYSASEKKEK DNAIEFFSNA GEFITQEGIS DTSYEVYPYKF IYIVSKFYES NDLQKSITEI EIHLEVESELK YDEIYPVISL AIILMRDNL MIPKVFLSSK KNCIAIHPEW DLLQEKGQLY NGEAEIFFRK NIYQVMSKYS TCGNIVEQKS LSLVIHEISK LNYLVFKDIS IDPTTGTVNI TQNTVMSKSS INWRDGHDLR VLNENNIFYD FSRDKLKISN	(SEQ ID NO: 174)

Table 6 - Cas12b (C2c1) orthologs						
Alicyclobacillus macrosporangiid us strain DSM 17980 WP_074948407. 1	MVAVKSIKVK RRGKDGAQEC LYELLVPQSV KEAGDSRWVE RWMP LGKHQ REKHFTGQEH IDDWLKLSEG EDPSFLSRWA FKYDFLFDHF PNDAA SPIDV GFR LPSQRT QTRRVIVRYG VFRVAHRDEL VRSIREQRLD GERMPSDWWD DWRKEVRRNA AGQVVRAERD RAGQWQAVYP HQAQVHDVLV SHYLKQTERD NLQRR LWENF SEDGVSFTVG RDP SGFVNGG	LMLGHLPEIR YMTAEQCRQE GKKGQAQMLA AKARYEANKA VRAWDRDMFQ LVALAQRLEQ EPVDRFDEIL SYNEVLN KLE GKGRHGVRFQ FVGDP AAPGA GTPADDAGEV ETERYLA EHP TPDAHGRQPF RLNRLRSQMA LFQAQVRYLA DKVKIRGIAR SRFAVALREH PCQLV LLEEL GTIPAAAFSSR AAALRP GELI DLSDIRVRCD DAKTRRRSSA RWT AQRAFWG	EGLWHLHEAV LLVRLRDRQK SGFLSPLADP KDPTKQVIAS QSLERLMSWE EMKEASPGFE RKRQAQNPRR DAKQFATFTL RMIVMRDGPV FRGQFGGAKI FLNLSLRVES DTGIPGSRGL FFPIHGM DHL SLRLLVRTGV QHRDASGEAW DVPGGHSLAQ IDNGKKDRLK SEYRFSNDRP FDARTGAPGI PTGDGEFLVT RREGKDGTVV SQGEGDDLSD MVHNRIETLL	NLGVRYYTEW RNGHTGDPGT KSEGGKGT SK LEMYGLRPLF SWNERVGA EF SKSSQAHRIT FGSHDLFLKL PSPCSNPVWA TEVEGIVVPI QYRRSALVRK QSEQAGRRNP TSGLRVMSVD VALHERSHLI LDEQKRDRNW GRMVQA AVRT LDYLERQYRF KLADRI LMEA PSENSQLMVW RCRRVPSIPL PAGRGASGVR LIPRLTNQRV EEQEL LAEAD AERFSVSGAA	LALLRQGNLY DEELLGVARR SGRKPAWMGM DVFTETYKTI ARLVDRRDRF KRALRGADGI AEPVFQPLWR RFENAEGTNI APSRQLDALA GRREEKAYLC PYAAVFHISD LGLRTSAAIS RLPGETESKK ERLQSSMERG LWRQLAKQVR LRSWSAFSVQ LGYVYVTDGR SHRGVLEELI KDAPSIPIWL VHADINA AH KERYSGVIFT DARERSVVL F EKVRG	(SEQ ID NO: 175)
Bacillus hisashii strain C4 WP_095142515. 1	MATRSFILKI EQDPKNPKKV ELVPSSVEKK GDPSWEEEEK IKWMEKSRNQ	EPNEEVKKGL SKAEIQAE LW GEANQLSNKF KWEEDKKKDP SVRRLDKDMF	WKTHEVLN HG DFVLKMQCKN LYPLVDPNSQ LAKILGKLAE IQALERFLSW	IAYYMNILKL SFTHEVDKDE SGKGTASSGR YGLIPLFIPY ESWNLKVKEE	IRQEAIYEHH VFNI LRELYE KPRWYNL KIA TDSNEPIVKE YEKVEKEYKT	(SEQ ID NO: 176)

Table 6 - Cas12b (C2c1) orthologs						
	LEERIKEDIQ QKWLKMDENE HPEYPYLYAT KYRILTEQLH YNQIFLDIEE ESGNVGRIYF DSKGGKLLSG FPIKGTELYA RNVLHFQQFE KDWVAFLLKQL RKFLLRWLSL MHALGYCYDV SRREIPRQVA QDNRFKLNQ ADINAAQNLO FGEYFILKD LKGEKMLLYR DDSSKQSM	ALKALEQYK PSEKYLEVFK FCEIDKKKKD TEKLKKKLT KGGKHAFTYK NMTVNIEPTE IESLEIGLRV VHRASFNIK DITEREKRV HKRLEVEIG PTEPGEVRL RKKKWQAKNP LQGEIYGLQV REGRLTLDKI KRFWTRTHGF GVYEWVNAK DPSGNVFPD KWMAAGVFFG	ERQEQLLRDT DYQRKHPREA AKQQAFTFLA QLDRLIYPTE ESIKFPLKGT SPVSKSLKIH MSIDLGQRQA PGETLVKSRE KWISRQENS EVKHWKSL EPGQRFADQ ACQIILFEDL GEVGAQFSSR AVLKEGDL YKVYCKAYQV LKIKKGS KWMAAGVFFG	LNTNEYRLSK GDYSVYEF DPINHLWVR SGGWEEKGK RDDFPKVNF AAASIFEVVD VLRKAREDN VPLVYQDEL DGRKGLYGI LNHLNALKE SNYNPYEERS FHAKTGSPGI DKGGEKFISL DGQTVYIPES SSSELVSDI KLERILISK	RGLRGWREII KKENHFIWRN FEERSGSNLN DIVLLPSRQF DHLRRYPHKV KPKELTEWIK QKPDIEGKLF KLMNQKLNFL QIRELMYKPY LKNIDEIDRT RLKKMANTII RFENSKLMKW RCSVVTKEL SKDRKCVTTH KDQKQKIIIE LKDSFDLASE TNQYSISTIE	
Candidatus Lindowbacteria bacterium RIFCSPLOWO2 OGH55994.1	MPRDDLILLT KTSDDSAHFL SVILEKAARW IDENQIEFGD LRKELFEKGG GLRSYEENEM QATRDDLIGE CGRDDRDCVS EIEYDLETAC TPVRLDVLVE QTDRKIENGK DQKKLPETAP LVVVSTDVRT SAYAVWELGL EESLRLRSRQ GLRIRRRRCRT DISFWKMAV SIYENRVKLA NPRKLNNEGI RLNFLKEVGD KRLAHLIVAQ ELREVSRAE LMQWAHRQIV PAWRNQEWFK ANAKCSAHDE TGIHSFRPPV LFRDPGSAFR QPQSGPGLAK	NLNSTAKGIR EKLKQHPDLK KEAKQEREK LKLSDCGAEW AIRRKSEEL AKYAAEAGMG LNALRSQDPR AFVAYNEEFA PTAIRLPLLV IREVTRAGRK IFISSKLMA ARSRVFSFS GKSKSSSAPP QKSWRAGVL IRRLNLQNS EKTPLSEAEV KNPPLSAVVD GALCSGYDAD LRLRLRPRA FVKKWSRPR ALGFEPDIRR IDDDAAARPH HFTKQVASLY RRTSNPRSKV PLNADENAAA SGSPFWSPMA NKQYWYEGKI PGGD	ERGKTKEGTD DAFGNLSSGG EQDSSEHGSK KRGMNQAGQ TPEDILPGKA SRSAPRIRRG AYGDARLFDW DEPSSITLTE GKENGGAER VTCFYSYFKQ YRDDLMVSPA CNVVERAPRK LPPGSRLWPA KGSTQTPVYA ILRVSRLLSL RKNCDKAAEI VAPSTIVPDD HRRPATGGLW RPDWREFHRT WPGDRRHIPP GLWKYVDGST TVSAPAHIVV GLKVAMVYAA DHSLKRASED NIGLRLRGV EPAQKKKIGA FWSNVMMAVE	KKKSGRKSSW SKKLEYKYL AAYRRLFDAG RVRSHMGWQR APDQNDWQER TIKWSKLRE LSKTDQRFLW TDERLHPVWP QGTRPLAEY NGVWYVREGE TGDFGSIKIL QLTRKPDVAVV DAVHGDPLR DCTGTGLLCL DKFEKTFEQ LIRWADTDAM GPDRETLLKK HDLDRTLIRE LNDANRIPKG GQLFDRQDAE GEILWQHPET FENLIRYRFQ FSSKFCSRCSG PTADETRPWW EDFRTKVNPA AAPGADVDEA AKIAGASVGA	PMDKAAWETA AGSAPWKESQ CLMPPEFAKY RREKENAVYS PAYGNQMWFI RWLQILKRN DGFDADGKIL FFGESSAVPY ADLASSFQLP IPSGESIQIK WERIELASHV VTIPSGVDQG ILSVDLGHRH PGDGEDTPAE SDVRDRPNKK AKSLAATGNA RQEEEEKFAS ISYGRGQKG RTLGGGLSMG HLEHLRDDRI RRFFAEGAAG SDRPKTENAG SPGARVSRFD LIEGGKEFVC GALKGKLRF GDADESGVVV KPVAAWSGQA	(SEQ ID NO: 177)
Elusimicrobia bacterium RIFOXYA12 OGS02326.1	MNRIYQGRVT TLALAAISGS SILGLEQNAS ERLPFFCPRN TLDISLFSK IDKHGSSLKV DREVTNDPIA IMSPHRFYQD FEGDTRIDL QKQWSELAEK WLKSGTEEW FPPKSETEEK ARFSYSAPRL	KVEVPDGDKE AVGSDEKSII FDIAAKHILR IQSKRSPTSK PKIVFLEDPK PAPGSKPSGL DARVNDKPHF TQKRKEAARK RKLVTDTLW GETTEGKLE AENPLKAWLN EVSKNTHSLT RRDCLRSENN	KGNIKWKKLE LREWAVQVQ TSEAKPEQRA AVSSVQEQKR KARAEHLKQF YPSAIVFKYF DYFTNIALIR LEEKIKAIEG LGESETPDNN VLKHEQTEHQ YKELQYELTD TSLASEHIKN ENLYKAPWLQ	NWSDILWQHH IWEKAKKKAT SALIRLLEEI QEEVRRFHNM DNACKKKHEL PVDITKTVFL EKEKNRAAWF KGGQFKESDS EGKKTEYSIS SDFGSATLYQ KKRPIHFTPA SLQFTAGLIR PMMRALGIDE	MLFQDAVNYY VFEGPQKRLT DKKNHNVVCG QPEEVVKNV VGIKKAFES KATEKLAMGK EFDLAAFI EDDDVDSLPG ERTLIRFPDI HLAKPEFHPI HPVYSPRYFD KTNVGGKAIK EKADRQNFAN	(SEQ ID NO: 178)

Table 6 - Cas12b (C2c1) orthologs						
	TRITLMAKGL LRWPHEKKPK KSRFIGEACD LFGEKGRSAT IVARRAQAAV EKVKETIISL LDQTKAMPN AKRDDSIPDC ELNETCDMHG AVRDKLKEMC GFRAVEVPIPG KPEKPTLLV ISDPRIWEIH DRKPNYFADF NKRHIEDWSN	DDIQLGFPVE NPPEQPWWGI KKWYAKVSRM PLEAEETAEL SRLHRWAWFF LKVQELLPT VLIRGQRLS CPDLLEKLDH AYAKVDNPVS EVFFPLCERR FELKYPWSWL PIAGGPPIFVP PRLRTEKRDG SGKVDWGFNG KQKQ	ANSQELQKEV DSFSCLAVDL GLLRLPGEDV IKLFGANEKD DEAKRSDDAI LLTRLANRVL MDRIEQITEL MKEQRVNQTA FIVIEDLSRY KAGSAWVSLP KDKKDKAGNL ISEVGLSSFG RLFAREKRKY IKNESGLTLV	SNGISWKGQF GQRYAGAFAR KVWRDASKID VMPDNWSKEL REILESDDTD PLRGRSWEWK RKRQALNQS HMLAEALGL RSSQGRSPRE PLLETPAAYS AKEALNIRTV LKPQVVQADI GEEKVEVQPS SGKALWWTIN	NWGGIASLSA LDVSTIEKKG KENGFAFRKE SFPEQNDKLL LKQKVNKNEI KHHQKNDGFI LRRQIGKKAP KLAEPFKDKK NSRLMKWCHR SRFCSRSGVA SEQLKAFNQD NAAINLGLRA KNEKAKKVVD QLQWERCDFI	
Omnitrophica WOR_2 bacterium RIFCSPHIGO 2 OGX36711.1	MNRIYQGRVT ALAAAEGLP SESASFGDAL FCDDLKKKPN SLVKCPDQD LRQHSDVRN LPCPFTEFDYV LARKNNKPIF EEREKRRSEA VKLFGDLDEE QDLQKAVTGF RWVKSVVYAA GAKHIREGLV LQPMMAALGL PVDLDISKLE PGQFWWDNPT AGGQSWFTKL AIALAKQLLH RYKTWLRWSW QLQKLIADAA HLLVSDGQSP KPEFGSSTCG HSLFTEEREK LMKWCHRKIN SEDRPFWRKT VKEDQPLPAI KWLPRKGKEA LNIACYGFAT AWQKKAEEAA	KVEKLKNGKS DKHPINVLRL KKILPPAPEN WDYSEEELAR INPEGARNLV FIHKNKNQFS RAAVPEPKVK KAFTALEKWS EAEKLYMMDE GSEHATGKIY QREHKLDMGY ADTRELAEEEL EVSLAVKDQS SSDARGCFTR ENIGKKARWE IQKEGMYCLA CAAGSLRLPG NENSAELES RLTEQHKELW VDLRNFLEKA AEIPWLRGQR ESLPDPCPD ADMHGEHEVI EKKVLLAEPE IEKQSVAREV RQADINAAVN NKRENAQFGN IQNLQHPFLA VKR	PDDREELKDW RMEEAWEEFP KEVRALAVAL KTGSGDWVAG KEALDHLEGV SLPNGLRILE KVDQEKKESE DQNSKSCWSD NPEWKPKKET GPSRASLRGF TAFFLKLCER ERLQEPVRYT GKYGTCRVRL DSKGNVKEPA QOMNTAYEKN IDLSQRRAAD EDTEVIREGK ARDWLGDNAK DKTLDEIRKV LLHIAIRALP GLSIARIEQL TDKIVRLKEE PGRSPVDFV GIPVIEVFAS FDCLDNLVGK IGLRAIAGPS VDLDLEVKN HASAVFSRQK	QTALWRHHEL RKTVTPAKNL LAEKARTLKP LWSEDALNKI SNGTKKEKND RARGGGININ SATWTELEKR FDKCAFEEAL EGDDVREVPI GKLRNEWVDL DYWDIWRDDT PAEPQFSRRL HYSAPRLIRD VALMSDFVGR KLKQRFHLIW YALLHAGVNR RQIELSGKKG RFSFPEQNDK PYFASWGELA LRENTWRWIE ENFRRAVLSL RVNQTAHLII LEDLSRYTTD YSSKFDARTG GLNGIHLVLP CYHAHPKVRL RLDIDSDVLK GAVARLQWEV	FQDAVSYYTL RDSVRPWLGL QKTSASYWGR DELAESLKL PGPAKKTNNW KTYAGVLFKA ILRIGDDPIE KTLNQDNQKT LKGDPREYKL FTKANDNPRE EVEVKKIREK FMFSDIKGKQ HLSDGSSSMW KRELRLMLNF PGMELKETQE DSKTFVELGQ RNATQSEYDQ LIDLYYGALS NGTNEATVQ NGKDGGKGPL NRLLRHEIGT AQSLGVRLKG KSRSRSENSR APGFRAVEVT QNGGGLFIAA IKGESGTDKG GDNTNLFHDP CRAINSRRLE	(SEQ ID NO: 179)
Phycisphaerae bacterium ST- NAGAB-D1 (transposase) AQT69685.1	MATKSYRARI KKLAQIMLSR LQAERFATVK RELMMQWRKD QLAAEQQPTA FKTVDVVRK FDSFKRPPTL GNWFFNWMPA TDAERKAGIA AFSSDGKTRK SPSIPPLRQV KGASAIIVSLA LVSEQAPEFG EKCTDFGYPV LKYDAETKTF	LTDSRLAAAL SNTFAHGVM NRAIHTKSRG RAAWLKDKNE SKNNRPRKTG QIITKYPQQN TLPSYPYRHPY SLKPDPRLPV GAKLMIKNNR PLRILSIDLG RRHDYDLKRK REHSADLILF ITVKDDINPW WEPGGHLFRC TVHTDKKTPP	DRTHVVFVES QITRDQPTST KVIPSPETLA WQKHPEFMQ KRFARWHLWY KYITRLLDWL WFTMELDQFY SWRAETFETE SELLFTVFEQ IRHIGAFALT RRRHGKPVKG ENLHSLKFSA MTSRICSNCN PHCDHRVNAD LIFKPRPQFD	LKQMINTYLR LDEEWTDLAR VPAKFHWQVC FYNGPYQNFL KWLSNPETII EDNNPELCTL KKADFENGTI GRFPPYLGKK DCPPRVKWK QGTRNDSAWQ QRSNANLQAH FDERWMNRQL LPGFRFSMKK INAAANLANK LWADSVKTRK	MQNGKFGPDH RIHKTGGLF DSASAYIRSN KLCDDDRITS EWRNKASASD ENLRRTYVKK QLLLIDEDDD IGKKLSRPAP TKNRKCPADN TESLKKGIIN RTNMAQDRFK RDMNRRHIVE KNPYREKLPR FFGLGYWNGG QLGPDPF	(SEQ ID NO: 180)

Table 6 - Cas12b (C2c1) orthologs						
Planctomycetes bacterium RBG_13_46_10 OHB62175.1	MSVRSFQARV KQKSLYKSIS DAAKLSSQGI EHNEWLKEKE KWLSDNPDFA PEMKALDNLH NPKPTNPEGY DPRLSLIRPV LIFPDKTPKA LPHGLVSCAV HPYRWTEGPD DRFKKAARTI INRALAGWNR RYSIIRENNR FKSYDWDKRL	ECDKQTMHEHL QSILEANAQN HVVYDKKQVLG KWESEDEHKK AWRGNKAVIN GYERNFVRR RKLILPKKAG KGRRVVRKGG AYLYFTCDIP DLSMRRTTG LGHIACHKRE VNFALNTENA RHLVERVIEM REIRFGYVEK SEKKQKEEIE	WRTHKVFNER ADYLLNSVSI DLPGMMSQMV YLDLREKFEQ PLSEKAQIRI RKTCKNPDPGF DLGSLEMRLL EQGQTKETDS DEPLTETAKK FATLCRYENG IRILRSKRGG ASKNGFYPPRA AKDAGFKRRV LFACPNCGYC TIESKLMMDKL	LPEIIKILFK KGWKPGTAKK CRQSVEAISG FEQSIGGKIT NKAKPNKKNS DHKPTFTLPH TGEKNKGNYP YEFFDKHLKK IQWLETGDTV KIHILRSRNL PVKGEESHID DVLLENLEG FEIPPYGTSQ ANADHNASVN CAMHKISRGS	MKRGECEGQND YRNASFTWAD HIELTKKWEK KRRGRWHLYL VERDEFFKAN PTIHPRWVVF DDWISVKFKA WRPAKLSGVK KKGKKRKKKV WVGKKEGKGC LQKHIDYMGGE LIPDAEKERG VCSKCGALGR LNRFLIEDS ISK	(SEQ ID NO: 181)
Spirochaetes bacterium GWB1_27_13 OHD16008.1	MSFTISYPFK GEDETGKRIY ICLVSSSMEN DYGKILLKRL SYESWCKLMI FEQDSGFACK DNSKLVGDRN RFTMPDFKKH FEENYTFNLA YIKYCDTAKN KSLFAKEKLQ DWMDEKTFQN LNDQNHKIIIR ERKNKIGIFT GIIDWNNVDN GGLSMQRIDF DNRVKQGVNY LTEYTFEAEDR KCLSDNNAPG KSIKINDILP VVLKDKDET KRIKMDKHEK KNKMTKLLRE	LIINKNDEAK IEETEEKKQI KKGFPNAQQA RDNNLIPIFT KERFDKEVKI GKFYITGRML LFEFIIKLEN PLWYRYEDPS KLKKLSENIT TSYGKSGGIR PQIFNTVDNK VKELNEGIRV INDKNYYAKH RQINKLNQLL KILSYENKED LEKLNKLKRR ILMTALGYIK PREENSKLRM IRCHQVTKKD LRGGELFASI FHLKNEPNYL ESTDEGDDYS RNNKKNKSGK	ALLDTHQYMN EEFYLIKKEE SLNIFSPLFD KFTDIKKITA KELENKCEENK KGFDEIKEKW QCLWNGDIDY NSNFRNYKIE FIPKSKNKEF LYFNRNELEN ITSLKIQKKF LSVDLGQRFF IYSKTIKLSG KIIRNDEIDK VINLHKELDK WVARTQKSAE DNEIKNDSKK WSHRQIFNFI LIDNSWFLKI KDGKLHIVQA KNYYSILNFV KNQIALFRDD	EGVKYYLEKL LGRNLTLTLD AESKGYLKE KLSPTALDRM QERDKIFEIL LKEKDRSEQN LKIKRDINKN VVKDENYITI EFIDSNDSEE YKDGKKMDSY GNEEQTNFLS AAVSCFEIMS EDDDLYKERK EKFKELIETT KLEIDFKEFI SIVLTPKFGY KQKEDWVKKN QQKASLWGIL VVKDDAFCDL DINASRNIAC PTNEELTFFK SGIFFDKSLW	LMFRQEKIFI EFKTLMRELY ENNNISLIHK IFAQAIEKLL EKYEERQKT LINILNKYQT QIWLDRPEMP PLITERNNEY DKKDQKSKSQ TVFTLSIRDY YFTQNQITKK EIDNNKLFFN INKNYKLSYQ KRYVKNTYND RECRKPIFRS KLKEHINELK RACQIILMEK VGDVFAPYTS IEINKENVKN RFLSQINPFR VEENKDIKPT VDGKIFWSV	(SEQ ID NO: 182)
Verrucomicrobia ceae bacterium UBA2429 GCA_00234350 5.1	MPLSRIYQGR YQLCLVALAS AASLFEAALA DPKANPTYDY VKLQPDKNFV QALPEKIAAL KPKSVKAPKV RGFVFPAFTS ALLAEAQRRL EGVSNKATGD ELISIVTDYQ RAKDMISAYR AEFGSDGKSL MRWFQPMKA LEPEGLIRHI AAAWWNREEI LGEAGGKLWS EAWEQEAREM ARFHRWSCRI GILDSKSKQP ELKQHLRVSL	TNSLIILPT SDGTRPLSKL KVLEGNEAPE SATARASASG GAEARARLLE PEETQVPRNR VEKVSARRKA LSFWAVPGPH DYMDEKTHDW QHIPKGVYTG RDHVYDVGDV VWLELQEDVV EVSIAVDVVG LDCPEVEMPA GKQAMWYKQF RAKGFSVLSV TALLGLGMLR ALLFGGEEAK HEKPEATGDD LTLVGLRKKL ERLANRELPL	PQEPWDHKAL HEQMKASWDE RARELAGELL LTKLAAVIHA AAHHFIKVAE KASPDLTFFAT NAVTAQAVVIE VPVWKEFDIA KTGDSDEPGH GLRGFYAIKK GLFRALCEPR RLAQPIRFTP KLQPVRAKLE LEKCAVSLMP NGTYKPRTOQ DLGQRDAGAW LPGEDARTGA SRLGPGFDHL VIDYGQVDEL EAQKVEPEKI RHREWVWNEA	ARFDSPLWRH AKTDTEDSWR LDKIEGDIQQ ENVTEEALQK SPPTKLAEVL LLFQHFPSLF EPEIDFAELG AFKEALKTVN IPPRKSDPN DWCELWERKA FWPLWQPLTD AHAENSRRLF FSAPRAARDE DVVKKGGGKW LDTGLHLYWP ALLESRSKA LDDQKRAVE SHSKQNEELL LTKTAEAMLE AAVLKPHAET FEKLEQGNFK	HALFQDAVNY VRLARRLGIP AGRGYWPRFC VAAEMDLWT ARFPDGLALW TAAVLGLSVG DDPIKLARGE QFKLKTSENR FTLIQALTQD DKSQGTPTTE EQEAERIKAG MFSDISGSHG LEGLSGGSES VRLLLNFPAT GLEKAPEAED FSRNRQPFIE FHGKAGRNAL RILSRAQSRL NLKALYTNAG IFQRLGTLP KEENPKWIRG	(SEQ ID NO: 183)

Table 6 - Cas12b (C2c1) orthologs						
	QRGLSMARIE DILNKLDRMK LIPGRKPVDF PFGIPVLEVP DKSSKDAEAA PMVGGGPRQA KLEKAAALTL RLSGGMSLWK	QIENLRKRFRM QQRVNQTAHL IVMEDLSRYL AAYSSRFICAL AMLFQDLHDL DMNAAINLGL EPPKEPTKEL AIKDGAWQRV	SLRRQMSLIP ILAQALGLRL SSQGRAPSEN TGVPGFRAVE NIEAREARKQ RAIASPTCLR AAQKRTNFFL KINDARIAK	GEQVKQGVED RPHLANDAER GRLMKWCHRA VHDGNAEDFR DKKLPLRTL ARPKIRAELE DEKFVGKFD WKNNPPPEPD	KGQRQPEPCE EEKDIHGEYE VLAKLQKMC WKRLIKKA APVAGGPLFI DGKHQAMLG AHVTTSKKL PDDEIQF	
Alicyclobacillus kakegawensis WP_067936067.1	MAVKSIVKVL RGPDGGQCY YEILVLQSIG DQGDPRWVAA DWKPLGKSQ EQEHFAEQSH FEIWKSIPEE ADETFLTRYA HKYEFLFDHL LREEEKSSVA SWRRQPSMLM FRIDDKQRRV SASISVFRVA TETKQLRKLR QGREFTKRLT MGKQVRDWRK FFARASGLVV EASGPREGQW LEELVNQAQV DSLPLWLTEF ADINAAQNLO GKKVLVSQDG VLFDRPSGHI	RLSECPDILA MTAEDCQREL KRQDAQQIAS REKYEQRKAV VRTWDRDMFQ LVKLARALEA ALFSQYDEVI LYNGVLRDLE GPGRHAVRFQ LHLHDTARPD SAAQMLEDAK TVNYNKLSAY KKEEVEALGD EERQAVLRPL PSWREALELE QVKSGAKVKV RADRESHFAV TAQHPPCQLI HDVLVGTVYA LDKHLRLDKNL RRLWQNFIDIT VTFFERSQTG GKGHWIRQRE	GMWQLHRATN LRRLNRQLH SFLSPLVDPN DPSKEILNSL QALERLMSWE DMRAASQGF RQVQAEKRRD KARQFATFTL RLLVVESEGA GFMAEWAGAK QAGDVYLNIS LEEHPDKQIP GRPPHYYPPIH FAQLALLRL LTRLEAYCGR KGYQLDVVGG ALRQHIEENAK ILEELSAYRF AFSSRFDART LRPDDVIPTG ELRLRCDVKM GKPHSEKQTD FWSLVKQRIE	AGVRYYTEWV NGRQDQPGTD SKGGRGEAKS DALGLRPLFA SWNRRVGEEY AKRGTAHQIT FGSHDLFAKL PDACVNPIWT KERDSVVVPV LQYERSTLAR VRVKSPSEVR GAPGLLSGLR GTDDLVAVHE VRCGAADERI VPDDEWSRIV NSLAQIDYLE RDRLKKLADR SDDRPPSENS GAPGVRCRRV EGEFLVSPCG VNNARAKQLF EADARAKSV SHTAERIRVR	SLMRQEILYS ADLLAISRRL GRKPAWQKMR VFTETYSRGSV ARLFQQKMKF RRALRGADRV AEPKYQPLWR RFESSQGSNL APSGQLDKLV KARRDKQGM GQRRPPYAA VMSVDLGLRT RSHLIQMPGE RTRSWQRLTK DRTVIALWRR QQYKFLRRWS ILMEALGYVY KLMAWGHRI PARFVGATVD EEAARVRQVH VNNARAKQLF EADARAKSV GVGSSLD	(SEQ ID NO: 184)
Bacillus sp._V3-13 WP_101661451.1	MAIRSIKLLM IGDKTKEAYQ SIGESGDANQ ELEKKKDEER QSVRKWDKDM EWEIEKIRKF ESASPEELWK IAAYNGLQKK QKKNYVVTLS EISFSDYSSR DVAPLQETRN SFGVASLLEG ELFAIHKRSF ETKKTDPDERK IWKESLVELH RLLISWSKRS TALGFKYDKE MKWAHRSIPR EDLKAGSNTL DSDNNELTVI SQTETIKKYF DISKTIELAQ LKKKILSNKV	KTNSGTDISIY AELINIIRNQ LGNKFLYPLV KAKDPTVKIF FIQAIERLLS EKERNMELEK VVAEQQNKMS LSRTKEQATF KLIWPSEEKW ISLDGVLGGS GRLQSPIGKA MRVMSIDMGQ LLNLPGEVVT KAIHKLMEIV HRIEYPVQOI RTPGEANRIE EKDRYKRWKE TVSMQGEMFG KRLIEDGFIN HADINAAQNL GKGSFVKNN EQQKKYLTMF EL	LRKALWRTHQ QRNNGSSEEH DPNSQSGKGT DNLNKYGLLP WESWNRRVAD NAFAPNDGYF EGFGDPKVFS TLPDAIEHPL IEKENIEIPL RIQFNRKYIK LKVISSDFSK RTSASVSIFE KNNKQQRQER QSYDSWTASQ VSKWRKGLSE TDEPFGSSLL TYPACQIILF LQVGDVRSEY ESELAYLKKG QKRFWQQNSE EQEVYKWEKS RDPSGYFFNN	LINEGIAYYM GSDQEILALL SNAGRKPRWK LFPLFTNIQK EYKQLKEKTE ITSRQIRGWD FLANRENRI WIRYESPGGT APSIQFNRQI NHKELLGEGD VIDYKPKELM VVKELPKDQE RKKRQFVRSQ KEVWEKELNL QHIQNVKDDR ENLNRYLFNL SSRFHAKTGA DIIPSQGGEL VYRVPCQLAR EKMKIKTDTT ETWRPQKEYW	NLLTLYRQEA RQLYELIIPS RLKEEGNPDW DIEWLPLGKR SYYKEHLTGG RVYEKWSKLP WRGHSERIYH NLNLFKLEEK KLKQHVKGKQ IGPVFFNLV DWMNTGSASN QKLFYSINDT IRMLANVLRL LTNMAAFNDE WNIDELEDTR LKQMANLIIM DRSRENSRL PGIRCHALTE FVTLSKRYKK MGEDKLYIPK FDLQDLGFE SIVNNIISKSC	(SEQ ID NO: 185)
Desulfatirhabdium butyrativorans WP_028326052.1	MPLSNNPPVT LTLRGGLDHT GAPSSYIVAS LSAAIRDDAV PMKDPEDKSS	QRAYTLRLRG LADTKVKGGK GDEPAKDRND WVNRSKVFDE ETEQEDKAKD	ADPSDLWSRE GKPDRTPTPE NVVSALEEIL AVKSVGSSLT LVQKAGQWLS	ALWHTHEAVN ERKARRILLA QSRKVAKSEI REEAWMLER SRYGTSEGAD	KGAKVFGDWL LSWLSVESKL DDWKRDCSAS FFGSRDAYLT FCRMSDIYK	(SEQ ID NO: 186)

Table 6 - Cas12b (C2c1) orthologs						
	IAAWADNASQ TPNPVHELLR LNDVESVCGF ARRISLAHKW GAVEPYRIRR GDIQLFEALA VPAYRHPDEL IKPVPLCWQS NITGLFEQAD RWLVTFVSVKL RLPGLRVLSV HIKVKKQGIG QGEEKDAREA DALKELGWIP RHGNRARIAY GWDEAAKQL AKTLSGDIVIT TIRNVGGLSL ALELLREQRV HAVVIENLAN VSAWYTSRQD DARERFLCEL IHADLNAAAN GQPLRQPAQT IGEWKETSAY	GGSTVDDLV QNTSLNKSHL TYRVDKDGQP IKRAEAERHK RAVDGWKEVV EDDALCVWHK LHPVFCDFGK KRLTRDLALG WNGRLQAPRQ QQGPWCAYA DLGHRYAAAC KQTEVDKTTI SNEEIWALHQ APDSSENLSR YLISEVKIRP WDSRIATLAG CKELSDAWKE SRLATITIEFR KQLASRIAEA YRPDETRTRL SRTGAPGIRC NKTWKAKTPA IGLRALTDPD NADGAASKIR QNDVQYRVIR	SELRQHFDTK DDLKKKANTR VSVADYSKYD FEEDAKRIAN AAWSKSDCKS DGEATNEPDF SRWKINYDVH NNHRNDAASA QLEAIAVVRD EQHGLNTNPQ AVWEAVNTET YRRIGADTLF MECKLDRTKP EDGEAKDYRE GGIQEKLDEN YKAPEENGDN RWEDEDQRWK RKVQVGFETR ALGIGSEGK ENRRLMTWSA QDVSVREFMQ EWKKAGFVRI WPGKWWYVPC KGKKNRTAGT MLKEHIKSLD	ESKATNGLDW AESCKSKIGS VDYKWGTARH VPARAREWLD TEDRIAAARA QPLIDYSLAI KNVQAPFYRG VTRADRLGRA NPRLSEQERN YWPHADTNRD VKEACQNVGR DGRPHAPAWA LIDRLIASGW SLAVDDLMFS GRIDLLQDAL VSDVAYRKKQ KLLRWFKDWV LRPDGTRHEI GWDGGKRPRQ SKVHKYLSEA SPFWRKQVKQ PLRGGEIFVS DPVSFESKMD SKEKVYLWRD NRTGDNVEG	IIGLSSYTGH KGQRPYSDAI YIFAVMLDHA SFCBERSVTS LQDDSEIDKF EAEFKKRQFK LCITLWTGSE ASNVTKSDMV LRMCGMIEHI RKVHARLILP DMPKEHDLYL RLDRQFLIKL GLLKRQMARL AVRTLRLALQ ALWHELFSSP QVYREQLRNV LPSGTQANNA GEQFGQKTLT RINDSRFAPC CQLNGLYLCT AEAKHDENKG ADSKSPSAKG YVKGCAAVKV ISAFFLESNE	
Desulfonatronum thiodismutans WP_031386437. 1	MVLGRKDDTA PVHVPESQVA LLDDLKPLK YLGALPEWAT PANQDALKNK LWLELGLLPL ALARAKRDEL GRALRSWTRV DGQEALWKER APGGSNLRTY SNVSFDQIQK TDLASKPGHV KFADQVRPGL DPEKLWAKHE LRLSVLQEDD LWKQCHCHFFH WAVTYLEAVR DRIKTGADMI SRRENSRLMR VRCRHLVEED MLVPWDGGEL QLSVDGSTRY MTPTNAGKKY FFAPDRWLPS	ELRRALWTTT EDALAMAREA GDAQKIGTNY PISKQEFDGK AYNKDDWKKE FTPVFDKTMV AALFLGMEDG REEWLRHGDT DCVTSFSLN QIRKTENGLW GSKMVGRCRY WFKLTLDVRP RVLSVDLGV RSFKITLPG PRTEHLRLFM DKAEKVVAER GLILRWNMGR IQAARGFVPR WSHREIVNEV FHDGLPGMHL FATLNAASQL EMAKAPKARL RAGPGEKSSG EITYWSRIIRR	EHVNLAVAEV QRRNGWPVVG AGPLFDSDT DASHLRFKAT KDKGISSWAV GNLWNRLAVR FAGLREYELR QESRKNICNR DADGLLEKRK ADVLLSPRN QSANQQFEG QAPQGWLDGK SFAACSVFEL NPSRKEEIR EAIVDDPAKS FSRWRTETRP RTYGEVNRQD KNGAGWVQVH GMQGEGLYGLH VGELDWLLPK HVIHADINAA LGALQQLKNG EEDELALDIV IWQVTLERN	ERVLLRCRGR EDEEILLALR RRDEGKDVAC GGDDAFFRVS KYIQKQLQLG LALAHLLSWE RNEAIKQHAF LQDRLRGKFG GYALMTFADA ESAAVEEKT LGGAEILFDR GRPAPPEAK VRGGPDQGT RAAMEELRSL ALNAELFKGF KSSSWQDWRE KKQFGTVASA EPCRLILFED VDTTEAGFSS DKDRTANEAR QNLQRREFWGR DAPFHLTSIP EQAEELAQGR SGRQERAEMD	SYWTLDRRGD YLYEQIVPSC CGPFHEVAGK IEKANAWYED QDPRTVRRK SWNHRAVQDQ EPVDRPYVVS DPDVFHWLAE RLHPRWAMYE NVR LAPSGQL KRIANEQHGA HFKTALS NKS FPAADGRTVD NGDIRRLKAI GDDFRSTPD RRGYAGGKSY LLHHINQLKE LARYRFR TDR RYLASSGAPG RLGGMVRPG CGEAIRIVCN NSQKPENSYV KTFFRDPSGV EMPY	(SEQ ID NO: 187)
Lentisphaeria bacterium DCFZ01000012. 1	MAVELNRIYQ VALAAMALDS AELNTSNGAD PRFCDTAYHG GWGGVKTQTG PLEQYLKID IPNNKYEIGI FSDALKTINQ VDEAATSVLP GYQKLRSMMML	GRVNHVYIFD KDSLFGKFKM IAMNLILEDG NWDVDAKSFS KFFTGDEAKA TFPEISFGRI APQWGDDPVK FRNKQTQERN ILAGDKRWNA EKEEKMRKT	ENQNVQSVDN QIRAVWVDFY GIPSEILNAA EKKGRQRLVD SLKKAISYFL SSHQNINISN LSRGKRGYTF QLKRYCAALN LLQLQKELGI ADDEEISQAL	GDDLFLVHHE RNGQLRPLGK LEHLAEKCTG ALYSLHPVQA QDTGKNPSEL AMWILKFFPD RAFTDLAMWE WMDGESSDKK CNDFTENELM QEIIIKFQSS	LYQDAINYYL HSLIRSLGHA DVSQLGKTF VQELAPEIEI QEYFSVAGKQ QYSVDLIKLN KNWKVFDRAA PPVEPADADA DYGSLRRTIR HRDTIGSVSL	(SEQ ID NO: 188)

Table 6 - Cas12b (C2c1) orthologs						
	FLKLAEPKYF TPADARYSRR PAKVVLAFSA ILMPDWDKNG WPSYQYKKPV GEAGGTQWYA VQACVLCKML RWLWMLNENG IADRVLP LRG IEQLESLRMR MKEQRINQTA AFVVLLEDLSR VGAAAYSSKFS GHRLLTFDAK VADATNLLCN GAGGNFFVLG KLERCCKINQ	CVWHDADKNQ VSDLYALVYK PRLKRDGLID KRRILLNFPI TWYQGKKPFD YRKFSGLLRL IGDATLLGGS LCDKAKTEIS RKWDWVLNPQ CQALNRILMR NLILAQALGL YRFSQDRSSY ANAI PGFRAE ATIIMPRNGG NRVSCDRKKD CSERILTGTS SRLDKFIAKK	NFASVDMVAD NAKECKTGYG KEFSAYYPPV KLDVSAIHQK VVAVDLGQRS PGEDATVIRD DEKTIRSF PK NSDWLVNKN I SNSFVLQQT A KTGEKPATLA RHCLHSESAT ENSRLMKWSH ECSIDQLSFY PVFIPFVPSD CWQVKRSSNF EKSPPVFTSSE EVQNEL	AVRYYSYQEE LRPDGNFVFE LQAFRLREEEA TDHRFENQFY AGAVSRITVS GQRTEELSGN QNDKLLIAFR DNVLKEEKQH HGSGDPHKKI EMRNNPIPD C KRKENG MHGE RKILEKLALL PWRELKDSRE SKDTLIQADI SKMVYPEKLS MAKKYPNLMF	KARLEEPIQI IAQKNAKGYA PKQSFKTTAV FANNTNTCLL TEKREHSVAI AGRLSTEEET RATGRMKQLQ REMLPAIILLQ CGQRLGSFAR CPDILMRLDA YEKIPGVEPA CEVFNVPILQ KALVEQIRKI NASFNIGLRG LSFDPIKKQE GSALWRNEIL	
Laceyella sediminis WP_106341859. 1	MSIRSFKLKI LFIRNEETNE HLYEEIVPSV MKDAGDPNWV ASGYTRTWDR SDVQWMNKL R LDSAASEEAY IDFAELNHLQ QDTRKRNLT LI QKKREVVFYD FFNISVDVRP VGESGRVSSL FYQLEGTELF LSAILRLHKK AKENDLQWNQ ELEATKKLLT ANLIVMTALG KLMEWSHRSI TEADLRNETN DNPRI LTLHA VHKKQGKTFR DPSGTF FKEQ	KTSGVNAEE IEKRSKEEIQ IGKSGNASLK QEYEKYMAER DMFQQA IERL EYEAQQEKS L WQEVATCQTA RELRRAKEDA LDKFI LPDEN YSTNLPHLGT AVEVKNGRLQ GLDSLSEGLR AVHQRSFLLA VNEDERIQAI AIKNAHHQLE RWSKRSREPG YKYDQEQKKW PKLVQM QGEL IIHELIEAGF DINAAQNIQK FVKVEGSDVY EWVEQKT FWG	LRRGLWRTHQ GELLERVHKQ ARFFLGPLVD QTLVRLEEMG LSWESWNRVR EENAFAPNEP MRGEFGDPAI TFTLPDSVDH GSWHEVKKVP LAGAKLQWDR NGLGKALTVL VMSIDL GQRT LPGENPPQKI DKLLQKVASW PVVGKQISLW VVKRIERFET IEVYPACQVV FGLQVADVYA IKEEHRPYLQ RFWHPSMWFR EWAKWSKNRN KVQSMIQAYM	LINDGIAYYM QQRNQWSGEV PNNKTTKDVS LIPLFPMYTD RERRAQFEKK YALTKKALRG YQFLAQKENH PLWVRYEAPG FSLAKSKQFH NFLNKRTQQQ THPDGTKIVT SATVSVEFIT QOMREIRWKE QLNEETATAW RKDLSTGRQG FAKQIQHHIN LFENLRSYRF AYSSRYHGRT QGDLPWSSGG VNCESVMEGE KNTFSSITER KKTIVQRMEE	NWLVL LRQED DDQTLLQTLR KSGTTPKWKK EVGDIHWLPQ THDFASRFSE WERVYHSWMR DIWRGYPERV GTNIHGYDLV RQVWLQEEQK IETETGEIGKV GWKAEQLEKW KEAPDNPYKF RNRIKQQVDQ NQALSQLYSK IAGLSLWSIE QVKENRLKQL SYERSRRENK GAPGIRCHAL ELFATLQKPY IVTVVPKNKT KPPSSMILFR	(SEQ ID NO: 189)
Methylobacteriu m nodulans (long form)	MYEAI VLADD ASDPIDPAVL ARKLGELRKE DRLAFRLAVG TLREYERARK EALHAI IATE NAMERLVERS QITLPLLKAA NEVFEAKLGS GVDDQLARAA LKDTAPT TGV LRQLRGGLNR ELERCSTVAD SMWSVQH LTD KDDRLKTGAD RPRRENSQLM CLDTPAAFS S GYKPGDLVPL LPCGKSAVQG AEWRRLSGAQ	ANAQLANAF L AEANAWLDTD AASGTS AIIK HLLSWESWCT EQIAQLGLPM QTRKRGRFGD RDTALMTLPD DDGRCIDTPL ADLLLNDWHL FHFQSAKGAK AFPLAEFRLW HRQLLRAATV PLWQDTCKRA VRRFLQSWSL LIVQAARGFQ QWAHRGV PDM RYHASTMT PG PGGEV FVCLN QIRWAPLSMG KDRDEAAAAE	GPLTDPNSAG AGRAWLVD TG ALKRDFGVLP RARDEHTARV GERDFLITVR PDLFRWLARP PVAHPRS AQW SFS LAPSDQL RGRIRDRVDA SKHADSVQAG AVHERSFTLE QKGERDAYLT ARLYRTEFGA AGRASGDIRR RNEFGYVWQK VGMQGEIYGI IRCHPLRKRE ANGLSRIHAD KRQAGALGGF DEELQGLEEE	FLEAFNKVDR APPRWRS LAA LFQPSLAPRI QRLEQFSSAH MTRGWDDLRE ENHHVWADGH EAEGGSNLRN QGVVLT KQDK GDIGSAFLKL LRVLSIDLGV LPGENVGAGG DLREAWSAKE VVSEWRSRTR LDRERGGVFA HAPCHVILFE QDRRDPSAR FEDQGFL ELL INAAQNLQRR GYLEPTGHDS LLERSGERVV	PAPSWLDQVP KQDP IWPREF LGSRSSLT PW LKGD LATKVS KWRRSGDKGQ ADAVGV LARV YQLEAVGGEL QQKITYCTNM ALDVAHVLPD RSFATCSVFE QQWRAQADAE LWPFEASLLS SREDRKYAGK KDLLDHIDAL DLSRYRMRTD KHARQPLAAF KRENEGLDLN FWTQHGDAFR GSCQWRKTT E FFRDP SGVVL	(SEQ ID NO: 190)

Table 6 - Cas12b (C2c1) orthologs						
	PTDLWFPSAA	FWSIVRAKTV	GRLRSHLDAQ	AEASYAVAAG	L	
Opitutaceae bacterium WP_009513281.1	MSLNRIYQGR	VAAVETGTAL	AKGNVEWMPA	AGGDEVWQH	HELFQAAINY	(SEQ ID NO: 191)
	YLVALALAD	KNNPVLGPI	SQMDNPQSPY	HVWGSFRRQG	RQRTGLSQAV	
	APYITPGNNA	PTLDEVFRSI	LAGNPTDRAT	LDAALMQLLK	ACDGAGAIQQ	
	EGRSYWPKFC	DPDSTANFAG	DPAMLRREQH	RLLLPQVLHD	PAITHDSPAL	
	GSFDTYSIAT	PDTRTPQLTG	PKARARLEQA	ITLWRVRLPE	SAADFDRLAS	
	SLKKIPDDDS	RLNLQGYVGS	SAKGEVQARL	FALLLFRHLE	RSSFTLGLLR	
	SATPPPKNAE	TPPPAGVPLP	AASAADPVRI	ARGKRSFVFR	AFTSLPCWHG	
	GDNHPTWKS	FDIAAFKYAL	TVINQIEEKT	KERQKECAEL	ETDFDYMHGR	
	LAKIPVKYTT	GEAEPPPIA	NDLRIPLRE	LLQNIKVDTA	LTDGEAVSYG	
	LQRRITRGFR	ELRRIWRGHA	PAGTVFSSEL	KEKLAGELRQ	FQTDNSTTIG	
	SVQLFNELIQ	NPKYWPIWQA	PDVETARQWA	DAGFADDPLA	ALVQEAELQE	
	DIDALKAPVK	LTPADPEYSR	RQYDFNAVSK	FGAGSRSANR	HEPGQTERGH	
	NTFTTEIAAR	NAADGNRWRA	THVRIHYSAP	RLLRDGLRRP	DTDGNEALEA	
	VPWLQPMMEA	LAPLPTLPQD	LTGMPVFLMP	DVTLSGERRI	LLNLPTLEP	
	AALVEQLGNA	GRWQNQFFGS	REDPFALRWP	ADGAVKTAKG	KTHIPWHQDR	
	DHFTVLGVDL	GTRDAGALAL	LNVTQAQPAK	PVHRIIGEAD	GRTWYASLAD	
	ARMIRLPGED	ARLFVRGKLV	QEPYGERGRN	ASLLEWEDAR	NIIILRLGQNP	
	DELLGADPRR	HSYPEINDKL	LVALRRAQAR	LARLQNRSWR	LRDLAESDKA	
	LDEIHAERAG	EKPSPLPPLA	RDDAIKSTDE	ALLSQRDIIR	RSFVQIANLI	
	LPLRGRREW	RPHVEVPDCH	ILAQSDPGTD	DTKRLVAGQR	GISHERIEQI	
	EELRRRCQSL	NRALRHKPG	RPVLGRPAKG	EEIADPCPAL	LEKINRLRDQ	
	RVDQTAHAIL	AAALGVRLRA	PSKDRAERRH	RDIHGEYERF	RAPADFVVE	
	NLSRYLSSQD	RARSENTRLM	QWCHRQIVQK	LRQLCETYGI	PVLAVPAAYS	
	SRFSSRDGSA	GFRVHLTPD	HRHRMPWSRI	LARLKAHEED	GKRLKTVLD	
	EARAVRGLFD	RLDRFNAGHV	PGKPWRLLA	PLPGGPVFP	LGDATPMQAD	
	LNAAINIALR	GIAAPDRHDI	HHRLRAENKK	RILSLRLGTQ	REKARWPGGA	
	PAVTLSTPNN	GASPEDSDAL	PERVSNLFVD	IAGVANFERV	TIEGVSQKFA	
	TGRGLWASVK	QRAWNRVARL	NETVTDNNRN	EEEDDIPM		
Thermomonas hydrothermalis WP_072754838.1	MSEKTTQRAY	TLRLNRASGE	CAVCQNNSCD	CWHDALWATH	KAVNRGAKAF	(SEQ ID NO: 192)
	GDWLLTLRGG	LCHTLVEMEV	PAKGNNPPQR	PTDQERRDRR	VLLALSWLSV	
	EDEHGAPKEF	IVATGRDSAD	DRAKKVEEKL	REILEKRDFQ	EHEIDAWLQD	
	CGPSLKAHIR	EDAVWVNRRA	LFDAVERIK	TLTWEEAWDF	LEPFFGTQYF	
	AIGDVGKDKD	DAEGPARQGE	KAKDLVQKAG	QWLSARFGIG	TGADFMSMAE	
	AYEKIAKWS	QAQNGDNGKA	TIEKLACALR	PSEPPTLDTV	LKCISGPGHK	
	SATREYLKTL	DKKSTVTQED	LNQLRKLAD	DARNCRKKVG	KKGKKPWADE	
	VLKDVENSCE	LTYLQDNSPA	RHREFSVMLD	HAARRVSMAM	SWIKKAEQRR	
	RQFESDAQKL	KNLQERAPSA	VEWLDRFCES	RSMTTGANTG	SGYRIRKRAI	
	EGWSYVVQAW	AEASCDTEDK	RIAAARKVQA	DPEIEKFGDI	QLFEALADE	
	AICVWRDQEG	TQNPSILIDY	VTGKTAENHQ	KRFKVPAYRH	PDELRRHPVC	
	DFGNSRWSIQ	FAIHKEIRDR	DKGAKQDTRQ	LQNRHGLKMR	LWNGRSMTDV	
	NLHWSSKRLT	ADLALDQNP	PNPTEVTRAD	RLGRAASSAF	DHVKIKNVFN	
	EKEWNGRLQA	PRAELDRIAK	LEEQKGTEQA	EKLRLRLRWY	VSFSPCLSPS	
	GPFIVYAGQH	NIQPKRSGQY	APHAQANKGR	ARLAQLILSR	LPDLRLSV	
	LGHREFAAACA	VWETLSSDAF	RREIQGLNVL	AGSGEGDLF	LHVEMTGDDG	
	KRRTVVYRRI	GPDQLLDNTP	HPAPWARLDR	QFLIKLQGED	EGVREASNEE	
	LWTVHKLEVE	VGRTVPLIDR	MVRSGFGKTE	KQKERLKKLR	ELGWSAMPN	
	EPSAETDEKE	GEIRSISRVS	DELMSSALGT	LRLALKRHGN	RARIAFAMTA	
	DYKMPGGQK	YYFHEAKEAS	KNDDETKRDR	NQIEFLQDAL	SLWHDLFSSP	
	DWEDNEAKKL	WQNHIAATLP	YQTPPEISAE	LKRVERNKKR	KENRDKLRTA	
	AKALAENDQL	RQHLHDTWKE	RWESDDQQWK	ERLRLKDWI	FPRGAEDNP	
	SIRHVGGLSI	TRINTISGLY	QILKAFKMRP	EPDDLKRNIP	QKGDDLENF	
	NRRLEARDR	LREQRVKQLA	SRIIEAALGV	GRIKIPKNGK	LPKRPRTTVD	
	TPCHAVVIES	LKTYRPDDL	TRRENRLMQ	WSSAKVRKYL	KEGCELYGLH	
	FLEVANYTS	RQCSRTGLPG	IRCDDVPTGD	FLKAPWWRRA	INTAREKNGG	
	DAKDRFLVDL	YDHLNNLQSK	GEALPATVRV	PRQGGNLFIA	GAQLDDTNKE	
	RRAIQADLNA	AANIGLRALL	DPDWRGRWWY	VPCDKGTSEP	ALDRIEGSTA	
	FNDVRSLEPTG	DNSRRAPRE	IENLWRDPG	DSLESWTSP	TRAYWDTVQS	
	RVIELLRHA	GLPTS				

Table 6 - Cas12b (C2c1) orthologs						
Methylobacterium nodulans WP_043747912.1	MYEAIVLADD ASDPIDPAVL ARKLGELRKE DRLAFRLAVG TLREYERARK EALHAIATE NAMERLVERS QITLPLLKAA	ANAQLANAF AEANAWLDT AASGTSIIK HLLSWESWCT EQIAQLGLPM QTRKRGRFGD RDTALMTLPD DDGRCIDTPL	GPLTDPNSAG AGRAWLVDTG ALKRDFGVLP RARDHESTARV GERDFLITVR PDLFRWLARP PVAHPRSAQW	FLEAFNKVDR APPRWRSLLA LFQPSLAPRI QRLEQFSSAH MTRGWDDLRE ENHHVWADGH EAEGGSNLRN	PAPSWLDQVP KQDPIWPREF LGSRSSLTWP LKGDLATKVS KWRRSGDKGQ ADAVGVLARV YQLEAVGGEL	(SEQ ID NO: 193)
Chloracidobacterium thermophilum WP_058868187.1	MPQQAQPPVT LTLRGGLDHT GAPAGLIIAF LSAAIRDDAV RDAYLAPAKG ATVYEAIAKW PGYKSATRN YSNSILEKVE EAERRKFEE GWKEVVAEWS VCVWHKDGDA FCDFGKSRWD RDLALGNNTG WNGRLQAPRQ WCAYAKIHGL YAAACAVWEA IYRRIGADTL QLEAALGRPV ELMSSAVRTM LQDALVLWHD RKKKQRENRE FKDWVFPRGN GGNGRRGNHA AETKEQFGQS RPDVRVDEPC CQLHGLHLRE AEAKQAANKG ADANSFAAKG KVDGSAVVNP SPLECIEFGE L	QRAYTLRLRG LADTPVKGGK GTEAAEERNR WVNRSKAFDE SEDESSEAKQ DGKASLEMA LNQLAAQTTV SVCGFTYLQD AKKIDQVPEA KSDCKTVEDR AKAPDPQPLI ICFDVHKNMQ DAGASEVTRA QLEAIAARYVE KEDPQYWP LSTEAQFQREI PDGKPHPAW SLIDRLVASG RLALKRHGDR LFSSRGWRDD KLYDVAKALA HASDPTIRKR SDPTIRKVG ALDALEHLRE HAIVIEDLTH VSASYTSRQD DARERLLCDL IQADLNAAAN DQPLRQSAQA WKEYAAYQNE	ADSNDPSWRD GKPDPTDTE KVVAALEEIL AVESIGSSGS EDQAKDLVQK DKAIADLATA TQQDFVSLKD GGPARHSEFA AKDWLDRFCL IAAARALQDD DYALAAEAEF TPFPRALCLT DRLGRAASRA KHDWDQKAEK DTNKNRKGHA KGRTILRGRT ARLDQFLIK WGGSDKQKAR ARIAHYLITD TAKQLWNAHV QDVTLREALH QLINPSGGNG LSLPRLATLT QRVKQLASRI YRPEETRTRR SRTGAPGVRC NARWKDRATA IGLRALTDPD QSGDAAKDKN VQCRVIRILK	ALWQTHEAVN ERKARRILLA KSRGVDQNEI SGSSLTREEP AGQWLSSRF LSEFNPASND KANNDQAECK VILDHAARRV ERSGVSGALE PEIDKFGDIQ KKRHFKVPAY LWTGSEM KRI ASNVTKSDV MRNAIQWLVT RLILSRPLGL DGNALYCHTR LQGEEEGVRE LEGKQLGWD AKLHGKAP DAWKRWEND RRGNHASDPT EFRRKVQVGF AEAALGVGRV ENRQLMTWSS QDVPVKEFMR DWEKAGAVRI WAGKWWYVPC GNKGAGKSKE EQIKGRDKQP	RGAQAFGDWL LSWLSVESKL NAWKKDCSAS WDMLEFFFGS TGKGA DFRM LQGVGLGISG QNTGSKGQRP SLAHTWIKLA PYRIRRRRAVD LFEALAEDDA RHPDALLHPI PLCWQSKRLA NIAGLFEQAD FSARLQPGGP RVLAVDLGHR HKANGKERV ASNEEIWAVH PADKPSLSVD TLDEKGRIDL EPGEDSSGAE DERWKKQLRW IRKRLINPS FTRLKPDGTR RRPVEGKDPK SKVKKYLAEA SPFWRKQVKQ PLQGGEIFVS DPASFRPVRD VNLWRDISS HEGSKEDDIP	(SEQ ID NO: 194)
Desulfovibrio inopinatus WP_027186183.1	MPTRTINLKL YRTVDNEGKE LYERLVPSVN MKLKEEKAPG FVSDQKKKQD DRLAVRAAVA RQYEA KRHST VTILSKLQTQ VLRRLKPYAL LLVDDAHGTW GLAGGAEVLF GKGRVRTPPE ELIEGKPETG ERSIARAEIY VPGQAFPRSL RTRPRPTSRE NRQDARFGS EQRYEPCQLI YGQIVENTAA MLGNTKVESE	VLGKNPENAT AEIPRHAVQE ENNEAGDAQA WEAASQIWIQ ELTKGNAPLI HFTSWESWNH LKSIALADDS LRGKFGDPDL MTFAHPRFHP IEKKIRVPLA HRPYMEHDER VHHFKTALS RAFPVADERS ALKRDIQRLK FQGLGAAPFR MWYKTRSYHG LASRLHHIN LFEDLARYRF GFSSRFHAAT EEKLRLLEK	LRRALFSTHR EALAFAKAAQ ANAWVSPLMS SDEGQSLLNK QQLKEMGLLP RTRAEYNSLK NPYRIGVRS FNWLAQDRHV RWILYEAPGG PSGQIQDLTL SEESLLERPG KSKHTRTLQP MDSPNKLWAK SLLRLGEEDN STPELWRQHC GKSIWMLEYL SLKEDRIKTG RVDPRPRENS GAPGVRCRFL IRPGSLVPWD	LVNQATKRIE RHNGCISTYE AESEGGLSVY PGSPPRWIRK LVNPFRRHLL LRRDEFEAAS RAWNRVREEW HLWSPRDSVT SNLRQYALDC EKLEKKKNRL AVWFKLTL DV GLRVLSDVLG HERSFKLTLP DNRRDALLEQ QTYYDKAEAC DAVRKLLLSW ADSI VQAARG QLMQWNHRAI LERDFDNDLP GGEQFATLHP	EFLLLCRGEA DQEILDVLRQ DKVLDPPP VW LRSQPWQDD DPEGKGVSPW DEFKDDFTLL IDKGATEEQR PLVRINAVDK TENALHITLP YYRSGFQQFA ATQAPPNWLD MRTFASCSVF GETPSRKEEE FFKGWGEEDV LAKHISDWRK SLRGRTYGAI YIPLPHGKGW VAETTMQAE KPYLLRELSW KRQTLCVIHA	(SEQ ID NO: 195)

Table 6 - Cas12b (C2c1) orthologs						
	DMNAAQNLQR NGQGAFELVR PEEDDTGRIT	RFFGRCGEAF DMGSTSQMNR VFRDSSGIF	RLVCQPHGDD FVMKSLGKKK PCNVWIPAKQ	VLRLASTPGA IKPLQDNNGD FWPAVRAMIW	RLLGALQQLE DELEDVLSVL KVMASHSLG	
Desulfonatronum thiodismutans WP_031386437.1	MVLGRKDDTA PVHVPESQVA LLDDLKPLK YLGALPEWAT PANQDALKNK LWLELGLLPL ALARAKRDEL GRALRSWTRV DGQEALWKER APGGSNLRTY SNVSFDQIQK TDLASKPGHV KFADQVRPGL DPEKLWAKHE LRLSVLQEDD LWKQHCHFFH WAYTYLEAVR DRIKTGADMI SRRENSRLMR VRCRHLVEED MLVPWDGGEL QLSVDGSTRY MTPTNAGKKY FFAPDRWLPS	ELRRALWTH EDALAMAREA GDAQKIGTNY PISKQEFDGK AYNKDDWKKE FIPVFDKTMV AALFLGMEDG REEWLRHGD DCVTSFSLN QIRKTENGLW GSKMVGRCRY WFKLTLDVRP RVLSDVLGVR RSFKITLPG PRTEHLRLFM DKAEKVVAER GLILRWNNMRG IQAARGFVPR WSHREIVNEV FHDGLPGMHL FATLNAASQL EMAKAPKARL RAGPGEKSSG EIYWSRIRRR	EHVNLAFAEV QRRNGWPVVG AGPLFDSBTC DASHLRFKAT KDKGISSWAV GNLWNRLAVR FAGLREYELR QESRKNICNR DADGLLEKRR ADVLLSPRN QSANQQFEG QAPQGWLDGK SFAACSVFEL NPSRKEETAR EAIVDDPAKS FSRWRTETRP RTYGEVNRQD KNGAGWVQVH GMQGEYGLH VGELDWLLPK HVIHADINAA LGALQQLKNG EDELALDIV IWQVTLERN	ERVLLRCRGR EDEEILLALR RRDEGKDVAC GGDDAFFRVS KYIQKQLQLG LALAHLLSWE RNESEIKQHAF LQDRLRGKFG GYALMTFADA ESAAVEEKT LGGAEILFDR GRPALPEAK VRGGPDQGT RAAMEELRSL ALNAELFKGF KSSSWQDWRE KKQFGTVASA EPCRLILFED VDTTEAGFSS DKDRTANEAR QNLQRRFWGR DAPFHLSIP EQAEELAQGR SGRQERAEMD	SYWTLDRRGD YLYEQIVPSC CGPFHEVAGK IEKANAWYED QDPRTEVRRK SWNHRVQDQ EPVDRPYVVS DPDVFWHLAE RLHPRWAMY NVR LAPSGQL KRIANEQHG HFKTALSNS FPAADGRTV NGDIRRLKAI GDDRFRSTPD RRGYAGGKSY LLHHINQLKE LARYRFRTR RYLASSGAPG RLLGGMVRPG CGEAI RIVCN NSQKPENSYV KTFRDPSPGV EMPY	(SEQ ID NO: 187)
Tuberibacillus calidus WP_027726362.1	MATKSFILKM HNDEDPDHPV YEELVPSAVG EAGDPSWKDA ISVKWMPKSK ESLQKELKGI EIVKKWMKLD WREYPEFLPL NLNKYRLIMN KFSS EDTGKG NVGRIFLNV LTHEIKESK DNKLFYPVKD KLKFLKNVLN LYSPHSVWVD DIEKTRRLLF ANQIVMTALG NLMRWSRREI KEHELYITEG FATLDERGEL VPSDKDQKEK ELVSEILQEA LLKRKLAERR	KTNNPQLRL VLKKEEIQER KSGEANQISN YEKWEKERQE NQSVRKFDKD STKA FEIMER PSAPQGNLYD YVKYRHAEQR EKEKVVFDR KHNF SYHKG LNIEPMQPF NTLT LGVESL TDLFAVHRTS MQKLEKTDER QLKSIHRKLE RWSMRPENPG YRYDGKRKKW PKQVAQIGGL GQKVRNQKFL VITHADINAA MENLFGIGYL SVMADDELKGN LFDGSSRRG	SLWKTHELFN LWMKVRETQQ KYLPLTDPA DPKLIKILAL MFNQAIERFL VEKAYEAHLR VVKDYQRRHP MKTAKKQATF LICLNADGHY INYELKGT LG RSGNLQTSVG PTGLRVMSVD FNIKLPGEKR EKRVNRWIKD EQLGKEISKW EVKQLQPGER IAXHPACQLV YGLLVGEVGA DSLVENNIE QNLQKRFWTR QPFKQENDVY RKTLFRDPSG LFNGTDSNTN	FGVAYYMDLL KNGFHEVSK SQSGKGTANS QSFGILPLFR SWESWNEKVA EITFSNSTYR RESGDFKLFE TLC DPIRHPL EEQEDVTVPL GARIQFDREH KALKVYVDGY LGQRQAA AIS TERRMLEQQK REREENPVY RQSI SQGRQG PDDARRLEPG THGLYRIRCE KWVKGEKIKG YVFPKDRWYT VE	SLFRQKDLYM DEVLET LRAL GRKPRWKKLK PFTENDHKAV EDYEKTVSIY IGNRAIRGWT LLSRPENQAA WVRYEERSGT APSQQFDDQI LLRRQGVKAG PKVVNF KPKE IFEVVSEKPD RDQAIRDLRSR VQEFEMISKV VYGISLKNIE HLKDDRIKKL YDERSRLENR GAPGIRCRV DLIRDQGGDK SREIKDAVVL KKTSSQSDDK GGRYFGTLEH	(SEQ ID NO: 196)
Bacillus thermoamylovorans WP_041902512.1	MATRSFILKI EQDPKNPKKV ELVPSSVEKK GDPSWEEKK IKWMEKSRNQ LEERIKEDIQ QKWLKMDENE HPEYPYLYAT KYRILTEQLH	EPNEEVKKGL SKAEIQAE LW GEANQLSNKF KWEEDKKKDP SVRRLDKDMF AFKSLEQY EK PSEKYLEVFK FCEIDKKKKD TEKLKKKLT	WKTHEVLNHG DFVLKMQKCN LYPLVDPSNQ LAKILGKLAE IQALERFLSW ERQEQLLRDT DYQRKHPREA AKQQATFTLA QLDRLIYPTE	IAYYMNILKL SFTHEVDKDV SGKGTASSGR YGLIPLFIPF ESWNLKVKEE LNTNEYRLSK GDYSVYEFLS DPINHPWLVR SGGWEEKGKV	IRQEAIYEHH VFNILRELYE KPRWYNLKIA TDSNEPIVKE YEKVEKEHKT RGLRGWREII KKENHFIWRN FEERSGSNLN DIVLLPSRQF	(SEQ ID NO: 197)

Table 6 - Cas12b (C2c1) orthologs						
	YNQIFLDIEE	KGKHAFTYKD	ESIKFPLKGT	LGGARVQFDR	DHLRRYPHKV	
	ESGNVGRIYF	NMTVNIIEPT	SPVSKSLKIH	RDDFPKFVNF	KPKELTEWIK	
	DSKGGKLLSG	IESLEIGLRV	MSIDLGQRQA	AAASIFEVVD	QKPDIEGKLF	
	FPIKGTETYA	VHRASFNIK	PGETLVKSRE	VLKAREDNL	KLMNQKLNFL	
	RNVLHFQQFE	DITEREKRV	KWISRQENSD	VPLVYQDELI	QIRELMYKPY	
	KDWVAFKQL	HKRLEVEIGK	EVKHWKSL	DGRKGLYGIS	LKNIDEIDRT	
	RKFLLRWSLR	PTEPGEVRL	EPGQRFADQ	LNHLNALKED	RLKKMANTII	
	MHALGYCYDV	RKKKWQAKNP	ACQIILFEDL	SNYNPYEERS	RFENSKLMKW	
	SRREIPRQVA	LQGEIYGLQV	GEVGAQFSSR	FHAKTGSPGI	RCSVVTKEKL	
	QDNRFKFNQ	REGRLTLDKI	AVLKEGDLYP	DKGGEKFISL	SKDRKLVTTH	
	ADINAAQNQ	KREWTRTHGF	YKVYCKAYQV	DGQTVYIPES	KDQKQKIEE	
	FGEGYFILKD	GVYEWGNAGK	LKIKKGSSKQ	SSSELVDSI	LKDSFDLASE	
	LKGEKLMLYR	DPSGNVFPD	KWMAAGVFFG	KLERILISK	TNQYSISTIE	
	DDSSKQSM					
Bacillus sp. NSP2.1 WP_026557978.1	MAIRSIKLLK	KTHTGPEAQN	LRKGIWRTHR	LLNEGVAYYM	KMLLLFRQES	(SEQ ID NO: 198)
	TGERPKEELQ	EELICHIREQ	QQRNQADKNT	QALPLDKALE	ALRQYELLV	
	PSSVGQSGDA	QIISRKFSLP	LVDPNSEGGK	GTSKAGAKPT	WQKKKEANDP	
	TWEQDYEKWK	KRREEDPTAS	VITTLLEYGI	RPIFPLYTNT	VTDIWLPLQ	
	SNQFVRTWDR	DMLQQAIRL	LSWESWNKRV	QEEYAKLKEK	MAQLNEQLEG	
	GQEWISLLEQ	YEENRERELR	ENMTAANDKY	RITKRQMKGW	NELYELWSTF	
	PASASHEQYK	EALKRVQQR	RGRFGDAHFF	QYLMEEKNRL	IWKGNPQRIH	
	YFVARNELTK	RLLEEAKQSAT	MTLPNARKHP	LWVRFDARGG	NLQDYILTAE	
	ADKPRSRFV	TFSQLIWSE	SGWMEKKDVE	VELALSRQFY	QQVKLLKNDK	
	GKQKIEFKDK	GSGSTFNHGL	GGAKLQLERG	DLEKEKNFE	DGEIGSVYLN	
	VVIDFEPLQE	VKNRGRVQAPY	GQVLQLIRRP	NEFPKVTTYK	SEQLVEWIK	
	SPQHSAGVES	LASGFRVMSI	DLGLRAAAAT	SIFSVEESSD	KNAADFSYWI	
	EGTPLVAVHQ	RSYMLRLPGE	QVEKQVMEKR	DERFQLHQRV	KFQIRVLAQI	
	MRMANKQYGD	RWDELDSLKQ	AVEQKKSPLD	QTDRTFWEGI	VCDLTKVLP	
	NEADWEQAVV	QIHRKAEEYV	GKAVQAWRKR	FAADERKGIA	GLSMWNIEEL	
	EGLRKLISW	SRRTNPQEV	NRFERGHTSH	QRLTHIQNV	KEDRLKQLSH	
	AIVMTALGYV	YDERKQEWCA	EYPACQVILF	ENLSQYRSNL	DRSTKENSTL	
	MKWAHSIPK	YVHMQAEPYG	IQIGDVRAEY	SSRFYAKTGT	PGIRCKKVRG	
	QDLQGRRFEN	LQKRLVNEQF	LTEEQVKQLR	PGDIVPDDSG	ELFMTLTGDS	
	GSKEVFLQA	DINAHNQK	RFWQRYNELF	KVSCRIVVRD	EVEYLVPKTK	
	SVQAKLGKGL	FVKKSDTAWK	DVYVWDSQAK	LKGKTTFTTE	SESPEQLEDF	
	QEIEEAEAEA	KGTYRTLFRD	PSGVFFPESV	WYPQKDFWGE	VKRKLYGKLR	
	ERFLTKAR					
Alicyclobacillus acidoterrestris WP_021296342.1	MAVSIKVKL	RLDDMPEIRA	GLWKLHKEVN	AGVRYYTEWL	SLLRQENLYR	(SEQ ID NO: 199)
	RSPNGDGEQE	CDKTAEECKA	ELLERLRARQ	VENGHRGPAG	SDELLQLAR	
	QLYELLVPQA	IGAKGDAQQI	ARKFLSPLAD	KDAVGGLGIA	KAGNKPWRVR	
	MREAGEPGWE	EEKEKAETRK	SADRTADVLR	ALADFLKPL	MRVYTDSEMS	
	SVEWKPLRKG	QAVRTWDRDM	FQQAIRMMMS	WESWNQVRGQ	EYAKLVEQKN	
	RFEQKNFVGQ	EHLVHLVNQL	QQDMKEASPG	LESKEQTAHY	VTGRALRGSD	
	KVFEKWGKLA	PDAPFDLYDA	EIKNVQRRNT	RRFGSHDLFA	KLAPEYQAL	
	WREDASFLTR	YAVYNSILRK	LNHAKMFATF	TLPDATAHPI	WTRFDKLGGN	
	LHQYTFLFNE	FGERRHAIRF	HKLLKVENG	AREVDDVTVP	ISMSEQLDNL	
	LPRDPNEPIA	LYFRDYGAEQ	HFTGEFGGAK	IQCRDQLAH	MHRRRGARDV	
	YLNVSVRVQS	QSEARGERRP	PYAAVFRIVG	DNHRAVHFVD	KLSDYLAHP	
	DDGKLGSEGL	LSGLRVMSVD	LGLRTSASIS	VFRVARKDEL	KPNSKGRVPF	
	FFPIKGNLNL	VAVHERSQLL	KLPGETESKD	LRAIREERQR	TLRQLRTQLA	
	YLRLLVRCGS	EDVGRRRERSW	AKLIEQPVDA	ANHMTPDWRE	AFENELQKLL	
	SLHGICSDKE	WMDAVYESVR	RVWRHMGKQV	RDWRKDVRS	ERP KIRGYAK	
	DVVGGSIEQ	IEYLERQYKF	LKWSFFGKV	SGQVIRAEKG	SRAITLREH	
	IDHAKEDRLK	KLADRIIMEA	LGYYVALDER	GKGKVVAKYP	PCQLILLEEL	
	SEYQFNNDP	PSENNQLMQW	SHRGVFQELI	NQAQVHDLV	GTMYYAFASSR	
	FDARTGAPGI	RCRRVPARCT	QEHNPFPFW	WLNKFVVEHT	LDACPLRADD	
	LIPTEGEIEF	VSFFSAEEDG	FHQIHADLNA	AQNLQQLRWS	DFDISQIRLR	
	CDWGEVDGEL	VLIPLRTGKR	TADSYSNKVF	YTNTGVTYEE	RERGGKRRKV	
	FAQEKLSSEE	AELLVEADEA	REKSVVLMRD	PSGIINRGW	TRQKEFWSMV	

Table 6 - Cas12b (C2c1) orthologs						
	NQRIEGLVK	QIRSRVPLQD	SACENTGDI			
Alicyclobacillus hesperidum WP_074693942.1	MTVRSIRVKL	AVGSPQYRDV	RRGLWKTHEI	MNQGVRYYCE	WLVLMRQEPI	(SEQ ID NO: 200)
	YDEDEHGLTV	VQRTREDIQA	ELLSRLRTLQ	SAHQHSGDMG	TDEELLSLMR	
	QLYEQLVPSS	VDKNKSGDAR	MIARNFFNPL	TNPNSQGGIG	ISNAGRKPKW	
	LLKKLSGDPT	WEEDYKKAME	QKQESSVSFL	LLELRRFGLH	PIFLPYTDTV	
	LEVSWAPKKA	RQWVRKWDYD	LFQQSIERML	SWESWTRRVK	ERFEKLVESE	
	KKFYDENFAT	DPEFIKLAET	LEGELQASSQ	GFVAVDEHAF	QIRPRSMRGF	
	DRVADEWCKL	ADDAPIEEYE	AAIKRVQARL	GRNFGSYVLF	AHLAKPEYWS	
	LWRSDFTKIL	RFARLRALQR	AVARAKRHAR	LTLPDALHHP	IWIYDAKGG	
	NIYSYRLLIP	EKRSKRYVE	FSSLIMPGE	NRWAEHRNIR	VPLAFSRQWE	
	RLHFSIMEDG	SLCVQYRDPG	VDEPLRAELG	GAKIQFDRRY	LIRRSSTLSA	
	GECGPVYLVN	SVDVNPAPHR	DVQVLQSAKL	VSVSRDTNRI	YLRPENLSAY	
	WKSQGDGTLF	LRVMSVDLGV	RSSAAVVICR	LEHRDSVVSS	GRRTATIYRI	
	AGTDEFVAVQ	ERAFLLRLPG	EGKGTNEDAP	LRDVYAQLGT	IRQGIQILRS	
	LLRLCDTKTP	DERQEALHGL	AQSLEPSGAW	KDELHPLHVM	LQGVVHDSVD	
	NWKQKVISVH	RQMERILGHA	VREWKVARKN	AGKPPIRRGA	GGLSLRRIRQ	
	LEQERRTLVA	WSNHAREPGQ	VVRIKRGTVQ	AQWLVERVNH	LKEDRLKKLA	
	DLIMITALGY	VYDETKPSGH	KWDKRYPPCQ	IILMEDLSRY	RFQSDRPPSE	
	NSQLMAWSHR	RLEILKLQA	DLHKLIVGTV	FPAFSSRFDA	QSGAPGVRCR	
	SVKKQDIENA	AQGKGWLARE	LQRLNWTLEW	LQPNDLIPTG	DGELFVTPAC	
	CDRQKGKIV	HADLNAAQNL	QRRFWGGHAE	SLCRVTCDVV	ERDGRRYAVP	
	RISNAFADSF	YKVFQGGVVF	STDEEDVYRW	MVGEKISSRG	RSRGRTSDEE	
	AEAETWIDEA	REQQGKVIAL	FRDASGQIHG	GDWLVAKVFW	GWVERLVTAR	
	LLSRMSEREA	AAHKE				
Alicyclobacillus acidiphilus WP_067623834.1	MAVKSMKVKL	RLDNMPEIRA	GLWLKHTEVN	AGVRYYTEWL	SLLRQENLYR	(SEQ ID NO: 201)
	RSNGDGEQE	CYKTAECKA	ELLERLRARQ	VENGHCGPAG	SDELLQLAR	
	QLYELLVPQA	IGAKGDAQOI	ARKFLSPLAD	KDAVGGLGIA	KAGNKPRWVR	
	MREAGEPGWE	EKAKAEARK	STDRTADVLR	ALADFGKPL	MRVYTDSDMS	
	SVQWKPLRKG	QAVRTWDRDM	FQQAIERMMS	WESWNQVRGE	AYAKLVEQKS	
	RFEQKNFVGQ	EHLVQLVNQL	QQDMKEASHG	LESKEQTAHY	LTGRALRGSD	
	KVFEKWEKLD	PDAPFDLYDT	EIKNVQRRNT	RRFGSHDLFA	KLAEPKYQAL	
	WREDASFLTR	YAVYNSIVRK	LNHAKMFATF	TLPDATAHPI	WTRFDKLGGN	
	LHQYTFLENE	FGEGRHAIKF	QKLLTVEDGV	AKEVDDVTVP	ISMSAQDLDD	
	LPRDPHELVA	LYFQDYGAEQ	HLAGEFGGAK	IQYRRDQLNH	LHARRGARDV	
	YLNLSVRVQS	QSEARGERRP	PYAAVFRLVG	DNHRAVFHFD	KLSDYLAEHP	
	DDGKLGSEGL	LSGLRVMSVD	LGLRTSASIS	VFRVARKDEL	KPNSEGRVFF	
	CFPIEGNENL	VAVHERSQLL	KLPGETESKD	LRAIREERQR	TLRQLRTQLA	
	YLRLLVRCGS	EDVGRRRERSW	AKLIEQPMDA	NQMTPDWREA	FEDELQKLKS	
	LYGICGDREW	TEAVYESVRR	VWRHMGKQVR	DWRKDVRSGE	RPKIRGYQKD	
	VVGNSIEQI	EYLERQYKFL	KSWSFSGKVS	GQVIRAEKGS	RFATTLREHI	
	DHAKEDRLKK	LADRIIMEAL	GYVYALDDER	GKGKWWAKYP	PCQLILLEEL	
	SEYQFNNDRP	PSENNQLMQW	SHRGVFQELL	NQAQVHDLIV	GTMYAAFSSR	
	FDARTGAPGI	RCRRVPARCA	REQNPEPFPW	WLNKFVAEHK	LDGCPRLADD	
	LIPTGEGEFF	VSPFSAEEGD	FHQIHADLNA	AQNLQRRLWS	DFDISQIRLR	
	CDWGEVDGEP	VLIPTTTGKR	TADSYGNKVF	YTKTGVTYYE	RERGKKRRKV	
	FAQEELSEEE	AELLVEADEA	REKSVVLMRD	PSGIINRGDW	TRQKEFWSMV	
	NQRIEGLVK	QIRSRVRLQE	SACENTGDI			
Alicyclobacillus macrosporangiid us SFU30094.1	MAVKSIVKVL	MLGHLPEIRE	GLWHLHEAVN	LGVRYYTEWL	ALLRQGNLYR	(SEQ ID NO: 202)
	RGKDGAQECY	MTAEQCRQEL	LVRLRDRQKR	NGHTGDPGTD	EELLGVARRL	
	YELLVPQSVG	KKGQAQMLAS	GFLSPLADPK	SEGGKGTSSK	GRKPAWMGMK	
	EAGDSRWVEA	KARYEANKAK	DPTKQVIASL	EMYGLRPLFD	VFTETYKTIR	
	WMPLGKHQGV	RAWDRDMFQQ	SLERLMSWES	WNERVGAEEA	RLVDRDRFR	
	EKHFTGQEHL	VALAQRLQE	MKEASPGFES	KSSQAHRITK	RALRGADGII	
	DDWLKLSEGE	PVDRFDEILR	KRQAQNPRRF	GSHDLFLKLA	EPVFQPLWRE	
	DPSFLSRWAS	YNEVLNKLED	AKQFATFTLP	SPCSNPVWAR	FENAEGTNIF	
	KYDFLFDHFG	KGRHGVRFQR	MIVMRDGVPT	EVEGIVVPIA	PSRQLDALAP	
	NDAASPIDVF	VGDPAAPGAF	RQFGGAKIQ	YRRSALVRKG	RREEKAYLCG	
	FRLPSQRRTG	TPADDAGEVF	LNLSLRVESQ	SEQAGRNP	YAAVFHISDQ	
	TRRVIVRYGE	IERYLAEHDP	TGIPGSRGLT	SGLRVMSVDL	GLRTSAAISV	

Table 6 - Cas12b (C2c1) orthologs						
	FRVAHRDELT RSIREQRLDR ERMPDWDWL WRKEVRRNAD QQVVAERDS AGQWQAVYPP QAQVHDVLVG HYLKQTERDA LQRRLENWFD EDGVSTFTVD DPSGFVNGGR	PDAHGRQPF LNRLRSQMAS FQAQVRYLAQ KVKIRGIARD RFAVALREHI CQLVLEELS TIPAAFSSRF AALRPGEIIP LSDIRVRCR AKTRRRSSAS WTAQRAFWGM	FPIHGMDHLV LRLLVRTGVL HRDASGEAWG VPGGHS LAQL DNGKKDRLKK EYRFSNDRPP DARTGAPGIR TGDGEFLVTP REGKDGTVVL QGEGDDLSDE VHNRIETLLA	ALHERSHLIR DEQKRDRNWE RMVQAAVRTL DYLERQYRFL LADRI LMEAL SENSQLMVWS CRRVPSIPLK AGRGASGVRV IPRLTNQRVK EQELLAEADD ERFSVSGAAE	LPGETESKKV RLQSSMERGG WRQLAKQVRD RSWSAFSVQA GYVYVTDGRR HRGVLEELIH DAPSIPIWLS VHADINA AHN ERYSGVIFTS ARERSVVLFR KVRG	
Sulfobacillus thermosulfidooxi dans PSR34340.1	RQSREDASPQ VYELIVPSSV KAEGNPLWEE FDPKKPGQFV REQFTAEPDA LLERWQKTLG WTVDPFTLQR SGNRYHIHLP IREAPTDSVV PDSIYLSLTL KPLYVMSVDL SVLLTMPGER EFLARLSKLE SSRIQNLRS WQQGRSHTVA RRYPSCHELMV GILVGTMYAG DINTLRPHDV QDEVFRVSCR IKISDMLEED VNSLVNKGVR	IIISASDLKA GKSGDSKTIA KFRQWKDRKD RPWDRSMFQQ ALYRVAQSLE KNGQSATLLD YAAFNDLTQR TTGQPSSVTF DNFPLVVEDQ NLDTTKPSGL GIRSAAAVSI REQRDRRYEQ ETSPELWEPV RSAYRMSGGL LGRHILELKR FEDLTRYRFR FSSRFDAVTK LPTGDGEYLL LALNSERVVA MDIPEDTAEL DKILGGIPVD	DLLYHARQQQ RKFLSPLTDP NDPTPFVLNQ AIERLMSWES EEMRKEHQGF DIRRVQSDLG LQRAKRVANL DRILWPDGDG SARTILRASW FHMQQNGRVW FSVQLKTGIE QRQGLRELRT YRSLNDSKMA SLDHVQDLER DRVKKVANYL TDRPRSENRO APGVRGATVR SVVRHRDSYR TTPPSYNKRY LRGNSVTLFR NSSAHAE	KEHVPRTIGS DSAGGRDQSA LADYGLLPLI WNQVRQEWEE ATDAPEAFRI DKFGSAPLYQ TLPDAVAHPI GWYERKRVTV GGAKLEYDRN IRKDVVMQYY EHRLTYPVAD DMRGMNDLLR PAAEWERLVV IRGIIASWTN IMTALGYAYD LMRWTHQELI QILRTRGMVR LKQVHADINA GKGFFEKGDN DPSGTIAGGN WLEAKEFWGR	DAEVLGALRQ SGRKPTWTKM RLFTDVGENI ALTQKHSAFY RRVALKGFDR KLVDERWQRL WSRYEGPNAS FLRPSHQVDR RLPRQLKKGV NEIPGDNVQF CPGLVAVHER GAYVDGDRRE YCHRQVEQSL HPRI PGSVVR SKRRARGEKVV AVTGIQAEPH LKEIADVG AHNLQRRLLWT GVYIWKTTGGK WLEAKEFWGR	(SEQ ID NO: 203)
Spirochaeta sp. LUC14_002_19_ P3 OQX29950.1	MGLLLPSLSR KRAEAERQKF AVLGWEALVE APCVWQKDGE FGQSRWSISF LNSNLALSLE RLQAPRKQLE YAEFSGEAP LGHRAAACA RRIGGNFLPD EKVFGRSIPL YSLCVDLSMT DFIVEALMLW AKIIKQKERK ELRSLRDWIL PEGEPMDGTM KPENNKTPKR KVKKYLFESE PFRQREIAHA EIFVSALTGN PHSKCSGAAC KSTINLWRDI	TVNVTIHLIL EADLYKIDRV AWDQKDCLSV PNAQIILLYV DIHEFFKNGE SPEDAIEVSR ALAKLEEDKS SRAAVIKGKY VWETLSSESM GNEHPAPWAK VDRLVRAGWG LAVNTVRFAL YELATDSRW ELLKKEFIPL PKGTRGKKRT AVGEGFGQKI PFTAVDEPCH QLHGLYLFV EENMAQENPC TLQADLNAAA LDNWRVGLKN SVLPLTEGQW	HPRKKGSRHR PQEARDWLDE EDRIAAARDL DAGEAEYKRS KNPVNIHALT ATRLGRAAVG AEALAKALRN TVIHQDKTRR EKKCREAGCL LDRQFIIDLQ EKNKQPEILQ RRHACRARIA GSWEANFWDE AEGLENSRR IRHVGGLSLS LDDLETMKEQ AVVIENLTHY QASYTSRQDS NRYLIALHNK NIGLQALLDP NLYTGVRTPL	EYAVMLDHAV FCRERTESTG QDNPGMDKFG HYKVPAYRHP MGLVSKKRIV ASQDRAVNIA RIKWFITFSP RPLAKLHLCR PPAPEDLYLH GEEGCTRMAL ELKQKGWVPL YYMEGGAIPE NFDKKLAETQ ISIASQWRMV RLAVIKSLYR RVKQLASRVV RPENKRTRRE RTGAPGVRCS WKNREYDKTA DWPGRWYVYP PGKNKGSTSG	RKIFLAHNWI SIDGYHIRRK DIWLYEALAS DPLLHPICFD KTELKWSKR GLFESAGWNG KLQPHGPWME MPGLRVLSVD LKKKNKTAVY AGEIWQVHCM EVSKTNTGYH GGLPENSGNK DAVNEREGDK WNEEDAIWQS VQKSFYTRMK EALGTGRIK NRQLMTWSSS ELSVKKFLES PPLRIPHWGG AVKGC DGRR EDVHKSNAVE	(SEQ ID NO: 204)
Bacillus hisashii strain C4 v4	MATRSFILKI EQDPKNPKKV ELVPSSVEKK GDPSWEEKK IKWMEKSRNQ	EPNEEVKKGL SKAEIQAELEW GEANQLSNKF KWEEDKKKDP SVRRLDKDMF	WKTHEVLNHG DFVLKMQCKN LYPLVDPNSQ LAKILGKLAE IQALERFLSW	IAYYMNILKL SFTHEVDKDE SGKGTASSGR YGLIPLFIPY ESWNLKVKEE	IRQEAIYEHH VFNILRELYE KPRWYNLkia TDSNEPIVKE YEKVEKEYKT	(SEQ ID NO: 205)

Table 6 - Cas12b (C2c1) orthologs						
mutant of WP_095142515. 1 K846R S893R E837G	LEERIKEDIQ	ALKALEQYEK	ERQEQLLRDT	LNTNEYRLSK	RGLRGWREII	
	QKWLKMDENE	PSEKYLEVFK	DYQRKHPREA	GDYSVYEFLS	KKENHFIWRN	
	HPEYPYLYAT	FCEIDKKKKD	AKQQATFTLA	DPINHPLWVR	FEERSGSNLN	
	KYRILTEQLH	TEKLKKKLTV	QLDRLIYPTE	SGGWEEKGKV	DIVLLPSRQF	
	YNQIFLDIEE	KGKHAFTYKD	ESIKFPLKGT	LGGARVQFDR	DHLRRYPHKV	
	ESGNVGRIYF	NMTVNIEPTE	SPVSKSLKIH	RDDFPKVVNF	KPKELTEWIK	
	DSKGKKLKSG	IESLEIGLRV	MSIDLGQRQA	AAASIFEVVD	QKPDIEGKLF	
	FPIKGTELYA	VHRASFNIK	PGETLVKSRE	VLRKAREDNL	KLMNQKLNFL	
	RNVLHFQQFE	DITEREKVRT	KWISRQENS	VPLVYQDELI	QIRELMYKPY	
	KDWVAFKQL	HKRLEVEIGK	EVKHWKSKLS	DGRKGLYGIS	LKNIDEIDRT	
	RKFLLRWSLR	PTEPGEVRL	EPGQRFADQ	LNHLNALKED	RLKKMANTII	
	MHALGYCYDV	RKKKWQAKNP	ACQIILFEDL	SNYNPYGERS	RFENSRLMKW	
	SRREIPRQVA	LQGEIYGLQV	GEVGAQFSSR	FHAKTGSPGI	RCRVVTKEKL	
	QDNREFKNLQ	REGRLTLDKI	AVLKEGDLYP	DKGGEKFISL	SKDRKCVTTH	
	ADINAAQNLO	KRFWTRTHGF	YKVYCKAYQV	DGQTVYIPES	KDQKQKIIEE	
	FGEYFILKD	GVYEWVNAGK	LKIKKGSSKQ	SSSELVSDI	LKDSFDLASE	
	LKGEKMLYR	DPSGNVFPD	KWMAAGVFFG	KLERILISK	TNQYSISTIE	
	DDSSKQSM					

Table 7- Cas12c (C2c3) orthologs						
OspCas12c	MTKLRRHQKK	LTHDWAGSKK	REVLGSNGKL	QNPLLMPVKK	GQVTEFRKAF	(SEQ ID NO: 206)
AWU30132.1	SAYARATKGE	MTDGRKNMFT	HSFEPFKTKP	SLHQCELADK	AYQSLHSYLP	
	GSLAHFLLSA	HALGFRIFSK	SGEATAFQAS	SKIEAYESKL	ASELACVDLS	
	IQNLTISTLF	NALTTSVRGK	GEETSADPLI	ARFYTLTGTG	PLSRDTQGPE	
KZX85786.1	RDLAEVISRK	IASSFGTWKE	MTANPLQSLQ	FFEEELHALD	ANVSLSPAFD	
	VLIKMNDLQG	DLKNRTIVFD	PDAPVFEYNA	EDPADIIIKL	TARYAKEAVI	
	KNQNVGNYVK	NAITTTNANG	LGWLLNKGLS	LLPVSTDDEL	LEFIGVERSH	
	PSCHALIELI	AQLEAPELFE	KNVFSCTRSE	VQGMIDSAVS	NHIARLSSSR	
	NSLSMDSEEL	ERLIKSFQIH	TPHCSLFIGA	QSLSQQLESL	PEALQSGVNS	
	ADILLGSTQY	MLTNSLVEES	IATYQRTLNR	INYLSGVAGQ	INGAIKRKAI	
	DGEKIHLPA	WSELISLPFI	GQPVIDVESD	LAHLKNQYQT	LSNEFDTLIS	
	ALQKNFDLNF	NKALLNRTQH	FEAMCRSTKK	NALSKPEIVS	YRDLLARLTS	
	CLYRGSIVLR	RAGIEVLKKH	KIFESNSELR	EHVHERKHV	FVSPDLRKA	
	KLLRLTDSRP	DLLHVIDEIL	QHDNLENKDR	ESLWLVRSY	LLAGLPDQLS	
	SSFNLPIIT	QKGDRLIDL	IQYDQINRDA	FVMLVTSFAK	SNLSGLQYRA	
	NKQSFVVRT	LSPYLGSKLV	YVPKDKDWLV	PSQMFEGRFA	DILQSDYMW	
	KDAGRLCVID	TAKHLSNIKK	SVFSSEEVLA	FLRELPHRTF	IQTEVRGLGV	
	NVDGIAFNNG	DIPSLKTFSN	CVQVKVSRTN	TSLVQTLNRW	FEGGKVSPPS	
	IQFERAYYKK	DDQIHEDAAK	RKIRFQMPAT	ELVHASDDAG	WTPSYLLGID	
	PGEYGMGLSL	VSINNGEVL	SGFIHINSLI	NFASKKSNHQ	TKVVPQQYK	
	SPYANYLEQS	KDSAAGDIAH	ILDRLIYKLN	ALPVFEALSG	NSQSAADQVW	
	TKVLSFYTWG	DNDAQNSIRK	QHWFGASHWD	IKGMLRQPPT	EKKPKPYIAF	
	PGSQVSSYGN	SQRCSCCGRN	PIEQLREMAK	DTSIKELKIR	NSEIQLFDGT	
	IKLFNPDPT	VIERRRHNLG	PSRIPVADRT	FKNISPPSSLE	FKELITIVSR	
	SIRHSPEFIA	KKRIGIGSEYF	CAYSDCNSSL	NSEANAAANV	AQKFQKQLFF	
	EL					
QFN42172.1	MRSNYHGGRN	ARQWRKQISG	LARRTKETVF	TYKFPLETD	AEIDFDKAVQ	(SEQ ID NO: 207)
	TYGIAEGVGH	GSLIGLVCAF	HLSGFRLFSK	AGEAMAFRNR	SRYPTDAFAE	
	KLSAIMGIQL	PTLSPEGLDL	IFQSPPRS	GIAPVWSENE	VRNRLYTNWT	
	GRGPANKPDE	HLLEIAGEIA	KQVFPKFGGW	DDLASDPDKA	LAAADKYFQS	
	QGDFFPSIASL	PAAIMLSPAN	STVDFEGDYI	AIDPAAETLL	HQAVSRCAAR	
	LGRFRPDLQ	NKGPFVSSLQ	DALVSSQNG	LSWLFVGVGF	HWKEKSPKEL	
	IDEYKVPADQ	HGAVTQVKSF	VDAIPLNPLF	DTTHYGEFRA	SVAGKVRSWV	
	ANYWKRLLDL	KSLLATTEFT	LPESISDPKA	VSLFSGLLVD	PQGLKKVADS	
	LPARLVSAAE	AIDRLMGVGI	PTAADIAQVE	RVADEIGAFI	GQVQQFNNQV	
	KQKLENLQDA	DDEEFLKGLK	IELPSGDKEP	PAINRISGGA	PDAAAEISEL	

Table 7- Cas12c (C2c3) orthologs						
	EEKLQRLDDA	RSEHFQTISE	WAEENAVTLD	PIAAMVELER	LRLAERGATG	
	DPEEYALRLL	LQRIGRLANR	VSPVSAGSIR	ELLKPVMFEE	REFNLFHNR	
	LGSLYRSPYS	TSRHQPFSSID	VGKAKAIDWI	AGLDQISSDI	EKALS GAGEA	
	LGDQLRDWIN	LAGFAISQRL	RGLPDTVPNA	LAQVRCPPDV	RIPPLLAMLL	
	EEDDIARDVC	LKAFNLYVSA	INGCLFGALR	EGFIVRTRFQ	RIGTDQIHVY	
	PKDKAWEPD	RLNTAKGPIN	AAVSSDWIEK	DGAVIKPVET	VRNLSSTGFA	
	GAGVSEYLVQ	APHDWYTPLD	LRDVAHLVTG	LPVEKNITKL	KRLTNRTAFR	
	MVGASSFKTH	LDSVLLSDKI	KLGDFTIIID	QHVRQSVTYG	GKVKISYEPE	
	RLQVEAAVPV	VDTRDRTPVE	PDTLFDHIVA	IDLGERSVGF	AVFDIKSCLR	
	TGEVKPIHDN	NGNPVVGTV	VPSIRRLMKA	VRSHRRRRQP	NQKVNQTYST	
	ALQNYRENV	GDVCNRIDTL	MERYNAFPVL	EFQIKNFQAG	AKQLEIVYGS	
QFN42158.1	MKKFELKQNF	RNNYSGKTLR	NFRQTLAQIA	NKKSSDSILT	IKFKLDCSKT	(SEQ ID NO: 208)
	GKLPKYENLI	SLYDTIEDIK	KGTLSYYLFT	LIVSGFKFFG	SASQAKAFST	
	KDIFKDNDY	NQFKIQSHLD	LPDFVPSKIY	QRLKKNVRST	NGKDNAFKAS	
	VIVAERYKEI	GKLNKDESS	EHQCEELFKK	IGTALETRES	SWQDLINNC	
	TGCEIIDEIL	NDSFGTLPST	KKMVLASTTQ	SSDGEQDGIA	IAYDPDSTFI	
	KSDELLNPYF	AVATILKSMP	PEIQQDKKSA	YVKANLTTPT	HNALSWIFGK	
	GLTLFQTEST	EKLCAMFNVS	DKRVIEQVQD	AAKAVKLPAE	LDLNHCTLKF	
	QDFRSSLGGH	LDSWTTNYLK	RLDELNDLLL	NLPKNLSLPD	IFMIDGKDFI	
	EYSGCNRDEI	QQMIDFVVNE	QNRKLLQESL	NALLGKGNQ	ICSDDISTVK	
	DFSEIVNSLH	SFVQQIDNSL	EQSSNEANSI	FSELKKKIEK	NEKWDIWKN	
	LKKIPKLNKL	SGGVPDAWKE	IREIEQKFHE	ISENQKKHFT	EVMEWIDAGN	
	GTIDIFESRF	KYDELLKSK	KNNLQSADEL	AFRSVLNKL	RFARQGNLIV	
	CEKIKNWFE	QNIFDSSKDF	NRYFINQKGF	IFKHPSSKDD	NSPYNLSANL	
	LEKRYEVTNT	VGALLEQCES	DPAIVNDPFS	MRSLEVEFRAL	WFSINISGIS	
	KEQHIFPKIA	QPKLDDSTYQ	ESVSPTLKRY	LEKEQITSS	LNSIFTVYKS	
	LLSGLSIRLS	RNSFYLRTKF	SWIGNNSLIY	CPKETTWKIP	AAYFKSDLWN	
	EYKDKQILIV	NEEYDVDVVK	TFESVYKIVK	SKDNNEKNRI	LPLLKQLPHD	
	WMFKLPFGAS	NAEKCKVLKL	EKNKCKFKPL	SVSKDSLARL	SGPSTYFNQI	
	DEIMMNDSE	LSEMTLLADE	PVRQQMSNGK	IEIIPDDYVM	SLAIPITRSL	
	KKGNTESEFP	KNIVSIDQGE	AGFAYAVFKL	SDCGNERAEP	IATGLIPIPS	
	IRRLIHSVKK	YRGKKQRIQN	FNQKFDSTMF	TLRENVTDGI	CGLIVALMKK	
	YNAFPILKQ	VGNLESGSKQ	LMLVYKAVNS	KFLAAKVDMQ	NDQRRSWWYQ	
	GNSWNTPIRL	ISNPQSNK	NIVKNINGKK	YEELKIYPGY	SVSAYMTSCI	
	CHVCGRNALE	LLKNDDSTGK	VKKYQINQDG	EVTIGGEVIK	LYRKPDRLTP	
	VKNLAKKGNR	ERTYASINER	APMSKDTTQS	RYFCVFKNCP	CHNKEQHADV	
	NAAINIGRRF	LKDCILDDNK	EKD			
QFN42173.1	MNARDWRKHV	GVLAAQHKET	TRTYTFPLDT	TGSAIDFDAA	LQAYNAVEGV	(SEQ ID NO: 209)
	GYGSLGLAC	AVHLSGFRLF	STGKEAATFR	NRARYPNAAF	QAALRKELGT	
	TITTLTPETL	DRLFSSRPKR	RNGVPLPWNQ	DSIRDRLYN	WVKPRPGDTP	
	DAVLFQIATG	IAQEITEDVS	SWTDLAKNSD	RGLKAAHRYF	ARVGGFPAFD	
	NLTTPATVQP	TDTTIDYDPN	APFHLVSHAD	QTLIHQSISL	CAHRIRQEDP	
	ALDPNKSQFI	KQLQNNFLSQ	TFYGLSWLFG	AGYVHFRECT	ANDLAIQYGI	
	PNNCRDGIHQ	IKSFADAILP	NTFFEKKHYR	KDSRSVGKKA	KSWISNYWQR	
	LLQLQTVVDD	HTWVTLPOEL	TEAQFKPLFR	GLLVDAVELM	AIAERLPQRL	
	ADCRDSLDC	MGKGPQAATK	NDVEIVEKVR	EEIESFVGQI	EQLGNQLRHQ	
	LENENNDQVH	RDNLHQLKNR	LPLDLRRPQA	LNKISGGVPD	VAKSIRGLET	
	QLDQVLKERR	SHFGRLTKWA	KECGITLDPL	QPLIESEKQR	VAERGSAHDA	
	KELAIRLLQ	RIGRLGHRLS	PTNATAIQEL	LRPVFAVKRE	FNLFFHNMHG	
	ALYRSPYSTS	RHQPFQINVD	VAHGTDWIGT	IETLIQNLFT	QIQDDALLRD	
	LVQLEGFVFS	HKLRALPGVI	PSELARPNNL	QQMGLPALLL	VLLQADQVHR	
	ETVLRVFNLY	GSAINGYLFQ	ALRPGFIVRA	GFQRLETKKL	RYVPKAQSWQ	
	YPDRLHHAKS	AIKNSLSAGW	IKKNHQGAIL	PQKTLTALVK	QKSLKDTGVP	
	EYLVQAPHDW	YVPIDLRGPA	PIEGLTVGT	EGPELTQLGP	MKDDCAFR	
	GPSSFYSKID	AGLLPQDVKY	GDMTLIFDQH	YQSSISFANG	TFSIQYQPTS	
	LQVKAAPV	DKRPRDTRNN	SHLYDRIVAI	DLGERKIGYA	IFDLKQVLKS	
	EQLEPMREDG	KPLIGSISIR	SIRGLMKAVQ	THRNRRQPNY	RIDQTYSKAL	
	MHYRESVIGD	VCNAIDTLCA	RYGGFPVLES	SVRNFEVGS	QLKTVYGSVS	
	RRYTWSAVDA	HKNQRQQYWL	GGTKDKIPIW	THPYLMTREW	DEKNSKWSNR	

Table 7- Cas12c (C2c3) orthologs						
	SKPLKMHFGV DGTIRLNSGY QPPKSTRAKD SRQALSTR	EVHPAGTSQI IDTTEIKRAR TTQSRYTCLY	CHQCKRNPIG RKKIRLPENK VDCGHECHAD	ALWNVADTVV PLTGSHTKSH ENAAINIGRK	LDDQGQLDLD VRAVARNLR YLQERIHIEA	
<u>QFN42174.1</u>	MVAGLKKIKR ETEEAGIDFD FCNQGRYPNQ SKNAIRNRLY PEKGLSAAADR NTLLHQAVAR VGFKYWKEMS EFRASVAGKM LLIDSKGLRQ SAFIGQVEQF SGGSPDPAT TLEAQRTER FTEKREFNLF AETIMKGLSQ EGFIVRTKFQ EGSVIDPVET RLRGLPDTVP ITGLKRQKKQ VSYNGRVKIT RVGYAVYDIK QRQPNQKVNQ FEAGSRQLEM WNERTRSYSG EINADGSLEL KHVSAVAKRN WKYLTERRIAL	DGVTMKSNIH KAVQTYGIAE AFAEKLRLNEL TNWTGKGAGT YFQAQGDFFS CAGRILQEQP VDQLADDYKV RSWVKNYWKR VTDKLP SRLK NNQLEQRLN IAELEEKLDL GAEGDQEEFA FHNRMGSLYR AGDELSLRQL RLERDVLSYV VTALSDTGFS GELALVRSAD TAFRMVGPSS FEPDRLHVEA ACLRTGDIKP TYSTALMNYR VYGSVLHRYT SLSALTLYPG DDGTICLYEG LRRPTVSMMS SESKDKAGV	GGVKARAWRK GISQGS LIGL SVTLPLKLSQ NPDEHLLLEIA LTGLPPSVPL NLS PDKNRFI KSTDLDALKQ LLDLKSQ LGT KAEDTIDRLM PLEGDDDETFL LMSARKEHYE LRLLLQRIGR SPYSTSRHQ EEDEVSREVC PKTKLWNPQ DDGIPEYLVQ DVRIPPMLAL FKSHLDSTLL AVPVIDKVR LEDGDGKPIV ENVIGDVCNR YSKIDAHTAK VMVHPAGTSQ YDYSPEEYKK GDTTQARYVC	RIGGLARRQK VCAFHLSGFR SLDVL FQSSP EDIAAEIDS TPQNSTVAFE NQ LQDELVSS VKS FIDAIPL ANINLPEGLD GDGNPTSDDI TIAEWASANK LANRLSPQGA FTIDVAVAKN LKAFLNYVSA RLDTARGPIH APHDWYLRDW TPIDLRDISK SEEVKLGDF GSVAVPSIRR IDTLMKEYNA RKEYWYTGEY RCHQCKRNPM AKREKRRLDP LYTDCDFTGH	ETVFTYKFPL LFSKADETKA KSKNGVAPEW LDGWKDL EEH GDPVCLNP SD QNNGLSWLFG NPLFDTPHYG EQRAENLFSG EQVETVAAEI FKKPPAINRI VTLDPM EAMT TAIRDLLRPV TDWMDALDGI INGCLFRALR SALAAWINK INISGFSLSQ PVSGLPVKK LIFDQYYKQR HLLAIDLGEK LMKAVRSHRQ FPVLESSVMN WDHPYLM AHK VEIKQLTGQV NVPLSGRHQA ADENAAINIG	(SEQ ID NO: 210)

Table 8 - Cas12e (CasY) orthologs						
APG80656.1 GI: 1110962136 QFN42175.1	MSKRHPRI SG IVSAINDDYV YTAPGLLKNV NGSLDKLKKD KLFRDFFGIS KEYAQKLDKN KAMMDITDAW GKLSSWLQNY SLLESIEKIV ALIKERLEAE YGDSKRELYK FFIKRLQKIF RSRKSAAIDK YIDLIELHKT GRFLEMFSQS FQSAKITTPK FGYELTRTGQ YVRSSYYQTQ ALEPVSGSER GDSAKILDQN VHSLRNRIHH IDADKNLQTT ELIGTVRVIK RQNSCLFICP IKVLGQMKKI	VKGYRLHAQR GLYGLSNFDD AEVRGGSYEL IIDCFKAEYR EQSENDKPSF EGSLEMWEYI RGQEQEEELE INQTVKIKED PDDSADDEKP KKKKPKKRKK KYKNAAIYTD SVYRRFNTDK NRVRLPSTEN ALALLLAVTE IVFSELRLGLA EMSRFLDLA GIDGGVAENA FLEWFLHRPK VFVSQPFTIF FISDPQLKTL LALKHKAKIV VWGKLAVASE GGTLIDAIKD FCRANADADI	LEYTGKSGAM LYNAEKRENEE TKTLKGSHLY ERHKDQCENK TNPLNLTCCL GIGNSGTAFS KRLRILAAALT LKGHKDLKK DIPAIATYRR KSDAEDEKET ALWKAVEKIY WKPIVKNSFA IAKAGIALAR TQLDISALDF GLMSRKEFIT PAEFATSLEP LRLEKSPVKK YELEVS RFEE ISASYTSQFC FMRPPIFDEN QASQTIALLR	RTIKYPLYSS KVYSVLDFWY DELQIDKVIK ADDIKNAKAD NFLGEGFLGR IKLREP KFDN AKEMINRFGE FLSDGRLTLN IDFKELFPHL KSAFSSSLKN PYCDIVSLAE ELSVAGFDWK VENGTVKDFM RSATQTMNGK ESLSEKSLK REIKCKQYKT SFLIDEKKVK RYLGIDIGEY QRRGT FAMP GKQIKKKVYA GACKKLWRAE DTPFPKYRDF YVKEEKKVED	PSGGRTVPRE DCVQYGAVFS FLNKKEISRA AGASLGERQK NRGEVLFNKL LRENKITELEK HWGGYRSDIN SDTKEEAVVS RFVQREDVQE AKPLKLVPNF SFFDTDFDKD NEVLYKPKQS DLLKKEEHEE KTRDGNLVLE QAE LLYIPHE LKQMRYPHY LGRGQNKIVL TRWNYDALTV GIA YTALEIT TKIARIRESL TLKKADVSE MQVDETITTQ CDKHHISKKM YFERFRKLKN	(SEQ ID NO: 211)

6.4. Protospacer Adjacent Motif

[0310] As used herein, the term “protospacer adjacent sequence” or “protospacer adjacent motif” or “PAM” refers to an approximately 2-6 base pair DNA sequence (or a 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11-, 12-long nucleotide sequence) that is an important targeting component of a Cas9 nuclease. Typically, the PAM sequence is on either strand, and is downstream in the 5' to 3' direction of Cas9 cut site. The canonical PAM sequence (i.e., the PAM sequence that is associated with the Cas9 nuclease of *Streptococcus pyogenes* or SpCas9) is 5'-NGG-3' wherein “N” is any nucleobase followed by two guanine (“G”) nucleobases. Different PAM sequences can be associated with different Cas9 nucleases or equivalent proteins from different organisms. In addition, any given Cas9 nuclease may be modified to alter the PAM specificity of the nuclease such that the nuclease recognizes alternative PAM sequence.

[0311] For example, with reference to the canonical SpCas9 amino acid sequence, the PAM specificity can be modified by introducing one or more mutations, including (a) D1135V, R1335Q, and T1337R “the VQR variant”, which alters the PAM specificity to NGAN or NGNG, (b) D1135E, R1335Q, and T1337R “the EQR variant”, which alters the PAM specificity to NGAG, and (c) D1135V, G1218R, R1335E, and T1337R “the VRER variant”, which alters the PAM specificity to NGCG. In addition, the D1135E variant of canonical SpCas9 still recognizes NGG, but it is more selective compared to the wild type SpCas9 protein.

[0312] It will also be appreciated that Cas9 enzymes from different bacterial species (i.e., Cas9 orthologs) can have varying PAM specificities and some embodiments are therefore chosen based on the desired PAM recognition. For example, Cas9 from *Staphylococcus aureus* (SaCas9) recognizes NGRRT or NGRRN. In addition, Cas9 from *Neisseria meningitis* (NmCas9) recognizes NNNNGATT. In another example, Cas9 from *Streptococcus thermophilis* (StCas9) recognizes NNAGAAW. In still another example, Cas9 from *Treponema denticola* (TdCas9) recognizes NAAAAC. These examples are not meant to be limiting. It will be further appreciated that non-SpCas9s bind a variety of PAM sequences, which makes them useful to expand the range of sequences that can be targeted according to the invention. Furthermore, non-SpCas9s may have other characteristics that make them more useful than SpCas9. For example, Cas9 from *Staphylococcus aureus* (SaCas9) is about 1 kilobase smaller than SpCas9, so it can be packaged into adeno-associated virus (AAV). Further reference may be made to

Shah et al., “Protospacer recognition motifs: mixed identities and functional diversity,” RNA Biology, 10(5): 891-899 (which is incorporated herein by reference). Gasiunas used cell-free biochemical screens to identify protospacer adjacent motif (PAM) and guide RNA requirements of 79 Cas9 proteins. (Gasiunas et al., A catalogue of biochemically diverse CRISPR-Cas9 orthologs, Nature Communications 11:5512 doi.org/10.1038/s41467-020-19344-1) The authors described 7 classes of gRNA and 50 different PAM requirement.

[0313] Oh, Y. et al. describe linking reverse transcriptase to a *Francisella novicida* Cas9 [FnCas9(H969A)] nickase module. (Oh, Y. et al., Expansion of the prime editing modality with Cas9 from *Francisella novicida*, bioRxiv 2021.05.25.445577; doi.org/10.1101/2021.05.25.445577). By increasing the distance to the PAM, the FnCas9(H969A) nickase module expands the region of a reverse transcription template (RTT) following the primer binding site.

6.5. Prime Editors

[0314] “**Prime editor fusion protein**” describes a protein that is used in prime editing. Prime editing uses CRISPR enzyme that nicks or cuts only single strand of double stranded DNA, i.e., a nickase; and a nickase can occur either naturally or by mutation or modification of a nuclease that makes double stranded cuts. Such an enzyme can be a catalytically-impaired Cas9 endonuclease (a nickase). Such an enzyme can be a Cas12a/b, MAD7, or variant thereof. The nickase is fused to an engineered reverse transcriptase (RT). The nickase is programmed (directed) with a prime-editing guide RNA (pegRNA). The skilled person in the art would appreciate that the pegRNA both specifies the target site and encodes the desired edit. Advantageously the nickase is a catalytically-impaired Cas9 endonuclease, a Cas9 nickase, that is fused to the reverse transcriptase. During genetic editing, the Cas9 nickase part of the protein is guided to the DNA target site by the pegRNA, whereby a nick or single stranded cut occurs. The reverse transcriptase domain then uses the pegRNA to template reverse transcription of the desired edit, directly polymerizing DNA onto the nicked target DNA strand. The edited DNA strand replaces the original DNA strand, creating a heteroduplex containing one edited strand and one unedited strand. Afterward, optionally, the prime editor (PE) guides resolution of the heteroduplex to favor copying the edit onto the unedited strand, completing the process (typically achieved with a nickase gRNA).

[0315] As used herein, “PE1” refers to a PE complex comprising a fusion protein comprising Cas9(H840A) and a wild type MMLV RT having the following N-terminus to C-

terminus structure: [NLS]-[Cas9(H840A)]- [linker]-[MMLV_RT(wt)] + a desired atgRNA (or PEgRNA). In various embodiments, the prime editors disclosed herein is comprised of PE1.

[0316] As used herein, “PE2” refers to a PE complex comprising a fusion protein comprising Cas9(H840A) and a variant MMLV RT having the following N-terminus to C-terminus structure:

[NLS]-[Cas9(H840A)]- [linker]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)] + a desired atgRNA (or PEgRNA). In various embodiments, the prime editors disclosed herein are comprised of PE2. In various embodiments, the prime editors disclosed herein is comprised of PE2 and co-expression of MMR protein MLH1dn, that is PE4.

[0317] As used herein, “PE3” refers to PE2 plus a second-strand nicking guide RNA that complexes with the PE2 and introduces a nick in the non-edited DNA strand. The induction of the second nick increases the chances of the unedited strand, rather than the edited strand, to be repaired. In various embodiments, the prime editors disclosed herein are comprised of PE3. In various embodiments, the prime editors disclosed herein are comprised of PE3 and co-expression of MMR protein MLH1dn, that is PE5.

[0318] As used herein, “PE3b” refers to PE3 but wherein the second-strand nicking guide RNA is designed for temporal control such that the second strand nick is not introduced until after the installation of the desired edit. This is achieved by designing a gRNA with a spacer sequence with mismatches to the unedited original allele that matches only the edited strand. Using this strategy, mismatches between the protospacer and the unedited allele should disfavor nicking by the sgRNA until after the editing event on the PAM strand takes place.

6.6. Guides for Prime Editing

[0319] Anzalone et al., 2019 (Nature 576:149) describes prime editing and a prime editing complex using a type II CRISPR and can be used herein. A prime editing complex consists of a type II CRISPR PE protein containing an RNA-guided DNA-nicking domain fused to a reverse transcriptase (RT) domain and complexed with a pegRNA. The pegRNA comprises (5' to 3') a spacer that is complementary to the target sequence of a genomic DNA, a nickase (e.g. Cas9) binding site, a reverse transcriptase template including editing positions, and primer binding site (PBS). The PE-pegRNA complex binds the target DNA and the CRISPR protein nicks the PAM-containing strand. The resulting 3' end of the nicked target hybridizes to the

primer-binding site (PBS) of the pegRNA, then primes reverse transcription of new DNA containing the desired edit using the RT template of the pegRNA. The overall structure of the pegRNA is like that of a typical type II sgRNA with a reverse transcriptase template/primer binding site appended to the 3' end. The structure leaves the PBS at the 3' end of the pegRNA free to bind to the nicked strand complementary to the target which forms the primer for reverse transcription.

[0320] Guide RNAs of CRISPRs differ in overall structure. For example, while the spacer of a type II gRNA is located at the 5' end, the spacer of a type V gRNA is located towards the 3' end, with the CRISPR protein (e.g. Cas12a) binding region located toward the 5' end. Accordingly, the regions of a type V pegRNA are rearranged compared to a type II pegRNA. The overall structure of the pegRNA is like that of a typical type II sgRNA with a reverse transcriptase template/primer binding site appended to the 3' end. The pegRNA comprises (5' to 3') a CRISPR protein-binding region, a spacer which is complementary to the target sequence of a genomic DNA, a reverse transcriptase template including editing positions, and primer binding site (PBS).

[0321] In typical embodiments, an atgRNA comprises a reverse transcriptase template that encodes, partially or in its entirety, an integration recognition site (also referred to as an integration target recognition site) or a recombinase recognition site (also referred to as a recombinase target recognition site). The integration target recognition site, which is to be placed at a desired location in the genome or intracellular nucleic acid, is referred to as a "beacon" site or an "attachment site" or a "landing pad" or "landing site." An integration target recognition site or recombinase target recognition site incorporated into the pegRNA is referred to as an attachment site containing guide RNA (atgRNA).

6.7. Attachment Site-Containing Guide RNA (atgRNA)

[0322] As used herein, the term "attachment site-containing guide RNA" (atgRNA) and the like refer to an extended single guide RNA (sgRNA) comprising a primer binding site (PBS), a reverse transcriptase (RT) template sequence, and wherein the RT template encodes for an integration recognition site or a recombinase recognition site that can be recognized by a recombinase, integrase, or transposase. In some embodiments, the RT template comprises a clamp sequence and an integration recognition site. As referred to herein an atgRNA may be referred to as a guide RNA. An integration recognition site or recombinase target recognition

site incorporated into the pegRNA is referred to as an attachment site containing guide RNA (atgRNA).

[0323] As used herein, the term “cognate integration recognition site” or “integration cognate” or “cognate pair” refers to a first integration recognition site (e.g., any of the integration recognition sites described herein) and a second integration recognition site (e.g., any of the integration recognition sites described herein) that can be recombined. Recombination between a first integration recognition site (e.g., any of the integration recognition sites described herein) and a second recognition site (e.g., any of the integration recognition sites described herein) is mediated by functional symmetry between the two integration recognition sites and the central dinucleotide of each of the two integration recognition sites. In some cases, a first integration recognition site (e.g., any of the integration recognition sites described herein) that can be recombined with a second integration recognition site (e.g., any of the integration recognition sites described herein) are referred to as a “cognate pair.” A non-limiting example of a cognate pair include an attB site and an attP site, whereby a serine integrase mediates recombination between the attB site and the attP site.

[0324] In typical embodiments, an atgRNA comprises a reverse transcriptase template that encodes, partially or in its entirety, an integration recognition site (also referred to as an integration target recognition site) or a recombinase recognition site (also referred to as a recombinase target recognition site). The integration target recognition site, which is to be placed at a desired location in the genome or intracellular nucleic acid, is referred to as a “beacon,” a “beacon” site or an “attachment site” or a “landing pad” or “landing site.” An integration target recognition site or recombinase target recognition site incorporated into the pegRNA is referred to as an attachment site containing guide RNA (atgRNA).

[0325] During genome editing, the primer binding site allows the 3' end of the nicked DNA strand to hybridize to the atgRNA, while the RT template serves as a template for the synthesis of edited genetic information. The atgRNA is capable for instance, without limitation, of (i) identifying the target nucleotide sequence to be edited and (ii) encoding new genetic information that replaces (or in some cases adds) the targeted sequence. In some embodiments, the atgRNA is capable of (i) identifying the target nucleotide sequence to be edited and (ii) encoding an integration site that replaces (or inserts/deletes within) the targeted sequences.

[0326] In some embodiments, the co-delivery system described herein includes a polynucleotide sequence encoding an attachment site-containing guide RNA (atgRNA) packaged in an LNP. In some embodiments, the co-delivery system described herein includes a vector comprising a polynucleotide sequence encoding an atgRNA. In some embodiments, the atgRNA comprises a domain that is capable of guiding the prime editor fusion protein to a target sequence, thereby identifying the target nucleotide sequence to be edited; and a reverse transcriptase (RT) template that comprises a first integration recognition site. In some embodiments, the atgRNA comprises a domain that is capable of guiding the prime editor fusion protein (or prime editor system) to a target sequence, thereby identifying the target nucleotide sequence to be edited; and a reverse transcriptase (RT) template that comprises at least a portion first integration recognition site.

[0327] In some embodiments, the co-delivery system described herein includes a polynucleotide sequence encoding a first attachment site-containing guide RNA (atgRNA) and a polynucleotide nucleotide sequence encoding a second attachment site-containing guide RNA (atgRNA) packaged into the same LNP. In some embodiments, the co-delivery system described herein includes a polynucleotide sequence encoding a first attachment site-containing guide RNA (atgRNA) packaged into a first LNP and a polynucleotide nucleotide sequence encoding a second attachment site-containing guide RNA (atgRNA) packaged into a second LNP.

[0328] In some embodiments, the co-delivery system described herein includes a vector comprising a polynucleotide sequence encoding a first attachment site-containing guide RNA (atgRNA), a polynucleotide sequence encoding a second atgRNA, or both.

[0329] In some embodiments, the co-delivery system described herein includes a polynucleotide sequence encoding a first attachment site-containing guide RNA (atgRNA) packaged into a first LNP and a vector comprising a polynucleotide sequence encoding a second atgRNA.

[0330] In some embodiments, where the co-delivery system contains a first atgRNA and a second atgRNA, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, where the at least first pair of atgRNAs have domains that are capable of guiding the gene editor protein or prime editor fusion protein to a target sequence, the first atgRNA further includes a first RT template that comprises at least a portion of the first integration recognition

site; and the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

[0331] In some embodiments, the first atgRNA's reverse transcriptase template encodes for a first single-stranded DNA sequence (i.e., a first DNA flap) that contains a complementary region to a second single-stranded DNA sequence (i.e., a second DNA flap) encoded by a second atgRNA comprising a second reverse transcriptase template. In certain embodiments, the complementary region between the first and second single-stranded DNA sequences is comprised of more than 5 consecutive bases of an integrase target recognition site. In certain embodiments, the complementary region between the first and second single-stranded DNA sequences is comprised of more than 10 consecutive bases of an integrase target recognition site. In certain embodiments, the complementary region between the first and second single-stranded DNA sequences is comprised of more than 20 consecutive bases of an integrase target recognition site. In certain embodiments, the complementary region between the first and second single-stranded DNA sequences is comprised of more than 30 consecutive bases of an integrase target recognition site. Use of two guide RNAs that are (or encode DNA that is) partially complementary to each other and comprised of consecutive bases of an integrase target recognition site are referred to as dual, paired, annealing, complementary, or twin attachment site-containing guide RNAs (atgRNAs). In certain embodiments, use of two guide RNAs that are (or encode DNA that is) full complementarity to each other and comprised of consecutive bases of an integrase target recognition site are referred to as dual, paired, annealing, complementary, or twin attachment site-containing guide RNAs (atgRNAs).

[0332] In some embodiments, upon introducing the nucleic acid construct into a cell, the first atgRNA incorporates the first integration recognition site into the cell's genome at the target sequence.

[0333] Table 9 includes atgRNAs, sgRNAs and nicking guides that can be used herein. Spacers are labeled in capital font (SPACER), RT regions in bold capital (**RT REGION**), AttB sites in bold lower case (**attB** site), and PBS in capital italics (*PBS*). Unless otherwise denoted, the AttB is for Bxb1.

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 46 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaa gccggccTGAGCTGCGAGAA	212
ACTB N-term PBS 13 RT 29 AttB 46 atgRNA with v2 scaffold	GCTATTCTCGCAGCTCACCAgtttgagagctatgctggaaacagcatagcaagtt caaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGC GGCGATATCATCATCCATGGccggatgatcctgacgacggagaccgccgt cgtcgacaagccggccTGAGCTGCGA GAA	213
ACTB N-term PBS_13_RT_29 _with TP901-1 minimal AttB f atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc GAGTCGGTGCGACG AGCGCGGCGATATCATCATCCATGGcacaattaacatctcaatcaag gtaaaTGCTTGAGCTGCGAGAA	214
ACTB N-term PBS_13_RT_29 _with TP901-1 minimal AttB rc atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc GAGTCGGTGCGACG AGCGCGGCGATATCATCATCCATGGagcattaccttgattgagatgt taattgtgTGAGCTGCGAGAA	215
ACTB N-term PBS_13_RT_29 _with PhiBT1 minimal AttB f atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc GAGTCGGTGCGACG AGCGCGGCGATATCATCATCCATGGcaggttttgacgaaagtgatc cagatgatccagTGAGCTGCGAGAA	216
ACTB N-term PBS_13_RT_29 _with PhiBT1 minimal AttB rc atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc GAGTCGGTGCGACG AGCGCGGCGATATCATCATCCATGGctggatcatctggatcatttcg tcaaaaacctgTGAGCTGCGAGAA	217
ACTB N-term Nicking guide 1 +48 guide	GAAGCCGGCCTTGCACATGCGgttttagagctagaatagcaagttaaataaggc tagtccggttat caactgaaaaagtggcaccgagtcggtgc	218
ACTB N-term PBS_18_RT_16 _with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCGgttttagagctagaatagcaagttaaataaggc tagtccggttatcaactgaaaaagtggcaccgagtcggtgc ATATCATCATCCATG GtaccgttcgtatagcatacattatacgaagttatTGAGCTGCGAGAATAGCC	219
ACTB N-term PBS_13_RT_29 _with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCGgttttagagctagaatagcaagttaaataaggc tagtccggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGA TATCATCATCCATGGtaccgttcgtatagcatacattatacgaagttatTGAG CTGCGAGAA	220
ACTB N-term PBS 13 RT 34 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc TCGACGACGAGCGCG	221

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
	GCGATATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGAGAA	
ACTB N-term PBS 13 RT 26 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GAGCGGGCGATATC ATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGAGAA	222
ACTB N-term PBS 13 RT 23 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc CGCGGGCGATATCATC ATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGAGAA	223
ACTB N-term PBS 13 RT 20 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GGCGATATCATCATC CATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGAGAA	224
ACTB N-term PBS 13 RT 16 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc ATATCATCATCCATG GccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGAGAA	225
ACTB N-term PBS 18 RT 34 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc TCGACGACGAGCGCG GCGATATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGAGAATAGCC	226
ACTB N-term PBS 18 RT 29 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGGCGAT ATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaa gccggccTGAGCTGCGAGAATAGCC	227
ACTB N-term PBS 18 RT 16 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc ATATCATCATCCATG GccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGAGAATAGCC	228
LMNB1 N-term PBS 13 RT 39 atgRNA	GCTGTCTCCGCCGCCGCCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc CTGCCCCATCCGCGGC GGCACGGGGGTTCGAGTCGCCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccCGGGCGGCGGAGA	229
LMNB1 N-term PBS 13 RT 34 atgRNA	GCTGTCTCCGCCGCCGCCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc CATCCGCGGGCGGCAC GGGGGTTCGAGTCGCCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccCGGGCGGCGGAGA	230
LMNB1 N-term PBS 13 RT 29 atgRNA	GCTGTCTCCGCCGCCGCCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GCGGGCGGCACGGGG GTCGAGTCGCCATGccggatgatcctgacgacggagaccgccgtcgtcgac aagccggccCGGGCGGCGGAGA	231

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
LMNB1 N-term PBS 13 RT 24 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GGCACGGGGGTCCG AGTCGCCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggc cCGGGCGGCGGAGA	232
LMNB1 N-term PBS 13 RT 19 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GGGGGTCCGAGTCC CCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccCGGG CGGCGGAGA	233
LMNB1 N-term PBS 18 RT 39 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc CTGCCATCCGCGGC GGCACGGGGGTCCGAGTCCCATGccggatgatcctgacgacggag accgccgtcgtcgacaagccggccCGGGCGGCGGAGACAGCG	234
LMNB1 N-term PBS 18 RT 34 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc CATCCGCGGCGGCAC GGGGGTCCGAGTCCCATGccggatgatcctgacgacggagaccgccgt cgtcgacaagccggccCGGGCGGCGGAGACAGCG	235
LMNB1 N-term PBS 18 RT 29 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGG GTCCGAGTCCCATGccggatgatcctgacgacggagaccgccgtcgtcgac aagccggccCGGGCGGCGGAGACAGCG	236
LMNB1 N-term PBS 18 RT 24 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GGCACGGGGGTCCG AGTCGCCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggc cCGGGCGGCGGAGACAGCG	237
LMNB1 N-term PBS 18 RT 19 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GGGGGTCCGAGTCC CCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccCGGG CGGCGGAGACAGCG	238
LMNB1 N-term Nicking guide 1 +46	GCGTGGTGGGGCCGCCAGCGgttttagagctagaatagcaagttaaataagg ctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	239
ACTB N-term PBS 13 RT 29 AttB 42 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGCGAT ATCATCATCCATGGggatgatcctgacgacggagaccgccgtcgtcgacaagc cggTGAGCTGCGAGAA	240
ACTB N-term PBS 13 RT 29 AttB 40 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGCGAT ATCATCATCCATGGgatgatcctgacgacggagaccgccgtcgtcgacaagcc gTGAGCTGCGAGAA	241
ACTB N-term PBS 13 RT 29 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGCGAT ATCATCATCCATGGatgatcctgacgacggagaccgccgtcgtcgacaagccT GAGCTGCGAGAA	242

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 36 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGG gatcctgacgacggagaccgccgtcgtcgacaagc TG AGCTGCGAGAA	243
LMNB1 N-term PBS 13 RT 29 AttB 44 atgRNA v2	GCTGTCTCCGCCGCCGCCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGG GTTCGAGTCGCCATG cggtatgacgacggagaccgccgtcgtcgaca agccggcCGGGCGGCGGAGA	244
LMNB1 N-term PBS 13 RT 29 AttB 42 atgRNA v2	GCTGTCTCCGCCGCCGCCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGG GTTCGAGTCGCCATG ggatgacgacggagaccgccgtcgtcgacaa gccggCGGGCGGCGGAGA	245
LMNB1 N-term PBS 13 RT 29 AttB 40 atgRNA v2	GCTGTCTCCGCCGCCGCCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGG GTTCGAGTCGCCATG gatgacgacggagaccgccgtcgtcgacaag cgcCGGGCGGCGGAGA	246
LMNB1 N-term PBS 13 RT 29 AttB 38 atgRNA v2	GCTGTCTCCGCCGCCGCCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGG GTTCGAGTCGCCATG atgacgacggagaccgccgtcgtcgacaagc cCGGGCGGCGGAGA	247
NOLC1 N-term PBS 18 RT 29 AttB 46 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGA ATGCCGGCGTCCGCC ccggatgacgacggagaccgccgtcgtcgac aagccggcTCCTCCAGGCAATACGCG	248
NOLC1 N-term PBS 13 RT 29 AttB 46 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGA ATGCCGGCGTCCGCC ccggatgacgacggagaccgccgtcgtcgac aagccggcTCCTCCAGGCAAT	249
NOLC1 N-term PBS 13 RT 29 AttB 44 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGA ATGCCGGCGTCCGCC ccggatgacgacggagaccgccgtcgtcgaca agccggcTCCTCCAGGCAAT	250
NOLC1 N-term PBS 13 RT 29 AttB 42 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGA ATGCCGGCGTCCGCC cgatgacgacggagaccgccgtcgtcgacaa gccggTCCTCCAGGCAAT	251
NOLC1 N-term PBS 13 RT 29 AttB 40 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGA ATGCCGGCGTCCGCC gatgacgacggagaccgccgtcgtcgacaag cgcTCCTCCAGGCAAT	252
NOLC1 N-term PBS 13 RT 29 AttB 38 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGA ATGCCGGCGTCCGCC atgacgacggagaccgccgtcgtcgacaagc cTCCTCCAGGCAAT	253

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
NOLC1 nicking guide -43	GAGCCGAGCACGAGGGGATACgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	254
ACTB N-term PBS 13 RT 20 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GGCGATATCATCATC CATGGatgatcctgacgacggagaccgccgtcgtcgacaagccTGAGCTGCGA GAA	255
ACTB N-term PBS 13 RT 15 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc TATCATCATCCATGGa tgatectgacgacggagaccgccgtcgtcgacaagccTGAGCTGCGAGAA	256
ACTB N-term PBS 13 RT 10 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc TCATCCATGGatgatcctg acgacggagaccgccgtcgtcgacaagccTGAGCTGCGAGAA	257
ACTB N-term PBS 9 RT 20 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GGCGATATCATCATC CATGGatgatcctgacgacggagaccgccgtcgtcgacaagccTGAGCTGCG	258
ACTB N-term PBS 9 RT 15 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc TATCATCATCCATGGa tgatectgacgacggagaccgccgtcgtcgacaagccTGAGCTGCG	259
ACTB N-term PBS 9 RT 10 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc TCATCCATGGatgatcctg acgacggagaccgccgtcgtcgacaagccTGAGCTGCG	260
LMNB1 N-term PBS 13 RT 20 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc CGGGGGTTCGAGTC GCCATGatgatcctgacgacggagaccgccgtcgtcgacaagccCGGGCGGCG GAGA	261
LMNB1 N-term PBS 13 RT 15 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GTCGAGTCGCCATG atgatectgacgacggagaccgccgtcgtcgacaagccCGGGCGGCGGAGA	262
LMNB1 N-term PBS 13 RT 10 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc AGTCGCCATGatgatcct gacgacggagaccgccgtcgtcgacaagccCGGGCGGCGGAGA	263
LMNB1 N-term PBS 9 RT 20 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCCAgtttagagctagaatagcaagttaaataaaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc CGGGGGTTCGAGTC GCCATGatgatcctgacgacggagaccgccgtcgtcgacaagccCGGGCGGCG	264

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
LMNB1 N-term PBS 9 RT 15 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaacttgaaaaagtggcaccgagtcggtgc GTCGCAGTCGCCATG atgatcctgacgacggagaccgccgtcgtcgacaagccCGGGCGGCG	265
LMNB1 N-term PBS 9 RT 10 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaacttgaaaaagtggcaccgagtcggtgc AGTCGCCATGatgatcct gacgacggagaccgccgtcgtcgacaagccCGGGCGGCG	266
SUPT16H N-term PBS 13 RT 24 Bxb1- GT_Initial length	GAGAAGCGGGCGTCCGGGGCTAgttttagagctagaatagcaagttaaataag gctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc TCTTTGTCCAGAGT CACAGCCATAccggatgatcctgacgacggagaccgccgtcgtcgacaagccgggc cCCCCGGACGCCGC	267
SRRM2 N-term PBS 13 RT 24 Bxb1 Initial length	GGGCACGGGGCCCATGTACAAgttttagagctagaatagcaagttaaataagg ctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc GGCGTCGGCAGCCC GATCCCGTTGccggatgatcctgacgacggagaccgccgtcgtcgacaagccgggc cTACATGGCCCCGT	268
DEPDC4 N-term PBS 18 RT 24 Bxb1 Initial length	GTGTCAAGTGGGGCGGGGCTAgttttagagctagaatagcaagttaaataag gctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc GCTGGCTCCTCCCC TGGCACCATAccggatgatcctgacgacggagaccgccgtcgtcgacaagccgggc cCCCCGCCCCACCTGACAC	269
NES N-term PBS 13 RT 29 Bxb1 Initial length	GAGTGGGTCAGACGAGCAGGAgttttagagctagaatagcaagttaaataa ggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc CGACTCCTCCCCC ATGCAGCCCTCCATCccggatgatcctgacgacggagaccgccgtcgtcgaca agccgggcTGCTCGTCTGACC	270
SUPT16H nicking guide - 53	GCAGCCACCCGCTCTCGGCCCgttttagagctagaatagcaagttaaataag gctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	271
SRRM2 N-term nicking guide 1 +87	GTGTAGTCAGGCCGCTCACCCgttttagagctagaatagcaagttaaataag gctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	272
DEPDC4 N-term Nicking guide 1 +59	GCTGACAAGTCTACGGAACCTgttttagagctagaatagcaagttaaataag gctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	273
NES N-term Nicking guide 2 +79	GCTCCTCCAGCGCCTTGACCgttttagagctagaatagcaagttaaataaggct agtccgttatcaacttgaaaaagtggcaccgagtcggtgc	274
HITI_ACTB_g uide	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaacttgaaaaagtggcaccgagtcggtgc	275
HITI_SUPTH1 6_guide	AGAAGCGGGCGTCCGGGGCTAgttttagagctagaatagcaagttaaataagg ctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	276

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
HITI_SRRM2_guide	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	277
HITI_NOLC1_guide	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	278
HITI_DEPDC4_guide	TGTCAGGTGGGGCGGGGCTAgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	279
HITI_NES_guide	AGTGGGTCAGACGAGCAGGAgtttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	280
HITI_LMNB1_guide	GCTGTCTCCGCCGCCGCCAgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	281
HDR Cas9 ACTB guide	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	275
HDR Cas9 LMNB1 guide	GGGGTCGCAGTCGCCATGGCgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	282
ACTB N-term PBS 13 RT 29 AttB original length atgRNAs for dinucleotides	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCATCATCCATGGccggatgatcctgacgacggagXXcgcgctcgtcgacaagccggccTGAGCTGCGAGAA XX: CG,GC,AT,TA,GG,TT,GA,AG,CC,TC,CT,AA,TG,GT,CA,AC	283
ACTB N-term PBS 13 RT 29 atgRNA with AttB 46 GT for fusion	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCATCATCCATGccggatgatcctgacgacggagACcgcgctcgtcgacaagccggccTGAGCTGCGAGAA	284
ACTB N-term PBS 13 RT 29 atgRNA with AttB 46 CT for multiplexing	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCATCATCCATGccggatgatcctgacgacggagAGcgcgctcgtcgacaagccggccTGAGCTGCGAGAA	285
NOLC1 N-term PBS 18 RT 29 atgRNA with AttB 46 GA for multiplexing	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCGGCGTCCGCCccggatgatcctgacgacggagTCcgcgctcgtcgacaagccggccTCCTCCAGGCAATACGCG	286

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
LMNB1 N-term PBS 18 RT 29 atgRNA with AttB 46 AG for multiplexing	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GCGGCCGCACGGGG GTCGCAGTCGCCATG ccggatgatcctgacgacggagCTcgccgtcgtcga caagccggccCGGGCGGCGGAGACAGCG	287
EMX1 Cas9 guide 1	GTCACCTCCAATGACTAGGGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc	288
EMX1 Cas9 guide 2	GGGCAACCACAAACCCACGAgttttagagctagaatagcaagttaaataagg ctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	289
ACTB N-term PBS 13 RT 29 AttB 56 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGG ctatg ccggatgatcctgacgacggagtcg ccgctcgtcgc acaagccggccctagcTGAGCTGCGAGAA	290
ACTB N-term PBS 13 RT 29 AttB 51 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGG tgc ccggatgatcctgacgacggagtcg ccgctcgtcgaca agccggccctaTGAGCTGCGAGAA	291
ACTB N-term PBS 13 RT 29 AttB 46 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGG ccggatgatcctgacgacggagtcg ccgctcgtcgacaa gccggccTGAGCTGCGAGAA	292
ACTB N-term PBS 13 RT 29 AttB 41 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATA TCATCATCCATGG gagatgatcctgacgacggagtcg ccgctcgtcgacaagccgT GAGCTGCGAGAA	293
ACTB N-term PBS 13 RT 29 AttB 36 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGG tgatcctgacgacggagtcg ccgctcgtcgacaagcTGA GCTGCGAGAA	294
ACTB N-term PBS 13 RT 29 AttB 31 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGG atcctgacgacggagtcg ccgctcgtcgacaTGAGCT GCGAGAA	295
ACTB N-term PBS 13 RT 29 AttB 26 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGG cctgacgacggagtcg ccgctcgtcgTGAGCTGCG AGAA	296
ACTB N-term PBS 13 RT 29 AttB 21 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATA TCATCATCCATGG tgacgacggagtcg ccgctcgtcTGAGCTGCGAGAA	297

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 16 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGacgacggagtcgcgTGAGCTGCGAGAA	298
ACTB N-term PBS 13 RT 29 AttB 11 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGgacggagtcgTGAGCTGCGAGAA	299
ACTB N-term PBS 13 RT 29 AttB 6 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGcggagtTGAGCTGCGAGAA	300
ACTB N-term PBS_18_RT_34 _with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc TCGACGACGAGCGC GGCGATATCATCATCCATGGtaccgttcgtatagcatacattatacgaagt tatTGAGCTGCGAGAATAGCC	301
ACTB N-term PBS_18_RT_29 _with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGA TATCATCATCCATGGtaccgttcgtatagcatacattatacgaagttatTGAG CTGCGAGAATAGCC	302
ACTB N-term PBS_13_RT_34 _with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc TCGACGACGAGCGC GGCGATATCATCATCCATGGtaccgttcgtatagcatacattatacgaagt tatTGAGCTGCGAGAA	303
ACTB N-term PBS_13_RT_16 _with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCGgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc ATATCATCATCCATG GtaccgttcgtatagcatacattatacgaagttatTGAGCTGCGAGAA	304
ACTB N-term Nicking guide 2 +93 guide	CCCCACGATGGAGGGGAAGAgtttagagctagaatagcaagttaaataagg ctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	305
LMNB1 N-term Nicking guide 2 +87 guide	CCTTCTCCTGGAGCCGCGACgttttagagctagaatagcaagttaaataaggc tagtccgttatcaactgaaaaagtggcaccgagtcggtgc	306
ACTB N-term PBS 13 RT 29 AttB 46 N191352_143_ 72 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGcattatatgttcttacagtatggcggccggattgtaaaaa catataatgTGAGCTGCGAGAA	307

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 46 N684346_90_6 9 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGcgttatagggtattacagtatggcggctcggtactgcaatac cctataac <i>TGAGCTGCGAGAA</i>	308
ACTB N-term PBS 13 RT 29 AttB 46 N675015_95_5 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGgtatcattttcatatagttagcacctgcacactatatgaaa atgataca <i>TGAGCTGCGAGAA</i>	309
ACTB N-term PBS 13 RT 29 AttB 46 N189929_49_5 4 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGgtctactatctgtatatgcgacacatgtggcataaagaca tagtagaca <i>TGAGCTGCGAGAA</i>	310
ACTB N-term PBS 13 RT 29 AttB 46 N203911_4518 6_6 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGcatcgaccctgacgcattgaggaggcgctccatgcgtc tgacctcatt <i>TGAGCTGCGAGAA</i>	311
ACTB N-term PBS 13 RT 29 AttB 46 N687663_53_2 9 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGgttagtaccataatgacaaaaggctcatcctttatcattgg gtactaac <i>TGAGCTGCGAGAA</i>	312
ACTB N-term PBS 13 RT 29 AttB 46 N687611_90_6 8 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGcgttattaaaaccggttcggtctgtcaaagcggtcatcggtt ttataaac <i>TGAGCTGCGAGAA</i>	313
ACTB N-term PBS 13 RT 29 AttB 46 N190156_234_12 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGggcgtgatggctgtaacctcaacatgacgacgaacacg acctcgcgcc <i>TGAGCTGCGAGAA</i>	314
ACTB N-term PBS 13 RT 29 AttB 46 N191533_224_76 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGtctacatcttgaatatatacaagtataactttgaattatatca gtttata <i>TGAGCTGCGAGAA</i>	315

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 46 N208621_9_15 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaacttgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGaattatatctaaaagcactaagctccgcatactgctttta gatataataTGAGCTGCGAGAA	316
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus _AH187_38 bp_Att	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaacttgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGgatattgggaagtgaatcagtacaaccgccacagtaccT GAGCTGCGAGAA	317
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus _AH187_38 bp_Att_rc	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaacttgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGggtactgtggcggtgtactgattcacttccccatattcTGA GCTGCGAGAA	318
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus _lugdunensis _N920143_38b p_Att	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaacttgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGtggtggttacagggtgccacattagtgtaccatttatgTG AGCTGCGAGAA	319
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus _lugdunensis _N920143_38b p_Att_rc	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaacttgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGcataaatggtacaactaatgtggcacctgtaccaccaT GAGCTGCGAGAA	320
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytoto xicus_NVH_ 391- 98_38bp_Att	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaacttgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGgttgttttccagatccagttggtcctgtaaatataagTGA GCTGCGAGAA	321
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytoto xicus_NVH_ 391- 98_38bp_Att_rc	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaacttgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGtttatatttacaggaccaactggatctggaaaaaacT GAGCTGCGAGAA	322

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus _AH187_Att 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGtactgtggcggtgtactgattcacttccccatatTGAG CTGCGAGAA	323
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus _AH187_Att 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGtactgtggcggtgtactgattcacttccccataTGAGC TGCGAGAA	324
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus _AH187_Att 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGactgtggcggtgtactgattcacttccccatTGAGCTG CGAGAA	325
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus _AH187_Att _rc 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGatattggggaagtgaatcagtacaaccgccacagtacTG AGCTGCGAGAA	326
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus _AH187_Att _rc 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGtatggggaagtgaatcagtacaaccgccacagtaTGAG CTGCGAGAA	327
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus _AH187_Att _rc 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGatggggaagtgaatcagtacaaccgccacagtTGAGC TGCGAGAA	328
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus _lugdunensis _N920143_Att 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGataaatggtaactaatgtggcacctgtaccacctGA GCTGCGAGAA	329

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus _lugdunensis _N920143_Att 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGtaaattggtacaactaatgtggcacctgtaccacTGAG CTGCGAGAA	330
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus _lugdunensis _N920143_Att 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGaaattggtacaactaatgtggcacctgtaccacTGAGCT GCGAGAA	331
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus _lugdunensis _N920143_Att_ rc 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGgggtggtacaggtgccacattagttgtaccatttatTGA GCTGCGAGAA	332
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus _lugdunensis _N920143_Att_ rc 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGgggtggtacaggtgccacattagttgtaccatttatTGAGC TGCGAGAA	333
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus _lugdunensis _N920143_Att_ rc 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGgtggtacaggtgccacattagttgtaccattttTGAGCT GCGAGAA	334
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytoto xicus_NVH_ 391-98_Att 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGttatatttacaggaccaactggatctggaaaaacaaTGA GCTGCGAGAA	335
ACTB N-term PBS 13 RT 29 AttB 46	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccgttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGAT ATCATCATCCATGGtatatttacaggaccaactggatctggaaaaacaTGAG CTGCGAGAA	336

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
Bacillus_cytotoxicus_NVH_391-98_Att 34bp		
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_Att 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGatatttacaggaccaactggatctggaaaaacTGAGC TGCGAGAA	337
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_Att_rc 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGttgttttccagatccagttggctcctgtaaatataTGAG CTGCGAGAA	338
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_Att_rc 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGgtttttccagatccagttggctcctgtaaatataTGAGCT GCGAGAA	339
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_Att_rc 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggct agtccggttatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGAT ATCATCATCCATGGgtttttccagatccagttggctcctgtaaatatTGAGCTG CGAGAA	340
Bacillus_cereus_AH187_Att_rc_36 LMNB1 PBS 9 RT 10 AttB 36 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtccggttatcaactgaaaaagtggcaccgagtcggtgc AGTCGCCATGatattggggaagtgaatcagtacaaccgccacagtac CGGGCGGCG	341
Bacillus_cereus_AH187_Att_rc_36 NOLC1 PBS 18 RT 29 AttB 36 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggc tagtccggttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGA ATGCCGGCGTCCGCCatattggggaagtgaatcagtacaaccgccacagtacT CCTCCAGGCAATACGCG	342

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
Bacillus_cereus_AH187_Att_rc_36 SUPT16H PBS 13 RT 24 AttB 36 atgRNA	GAGAAGCGGCGTCCGGGGCTAgtttagagctagaatagcaagttaaataag gctagtcggttatcaactgaaaaagtggcaccgagtcggtgc TCTTTGTCCAGAGT CACAGCCATA Aatatggggaagtgaatcagtacaaccgccacagtac CCCCGG ACGCCGC	343
Bacillus_cereus_AH187_Att_rc_36 SRRM2 PBS 13 RT 24 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgtttagagctagaatagcaagttaaataagg ctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GGCGTCGGCAGCCC GATCCCCGTTG atatggggaagtgaatcagtacaaccgccacagtac TACATGG CCCCCGT	344
Bacillus_cereus_AH187_Att_rc_36 DEPDC4 PBS 18 RT 24 AttB 36 atgRNA	GTGTCAGGTGGGGCGGGGCTAgtttagagctagaatagcaagttaaataag gctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GCTGGCTCCTCCCC TGGCACCAT Aatatggggaagtgaatcagtacaaccgccacagtac CCCCGC CCCACCTGACAC	345
Bacillus_cereus_AH187_Att_rc_36 NES PBS 13 RT 28 AttB 36 atgRNA	GAGTGGGTCTCAGACGAGCAGGAgttttagagctagaatagcaagttaaataa ggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc CGACTCCTCCCC ATGCAGCCCTCCAT Catatggggaagtgaatcagtacaaccgccacagtac T GCTCGTCTGACC	346
B. cereus LMNB1_PBS 9 RT 20 AttB 36 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggc tagtcggttatca acttgaaaaagtggcaccgagtcggtgc CGGGGGTTCGCAGTCGCCATGata tggggaagtgaatcagtacaaccgccacagtacCGGGCGGGC	347
B. cereus LMNB1_PBS 13 RT 20 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgtttagagctagaatagcaagttaaataagg ctagtcggttatcaactgaaaaagtggcaccgagtcggtgc CGGGGGTTCGCAGTCG CCATG atatggggaagtgaatcagtacaaccgccacagtac CGGGCGGGCGGA GA	348
B. cereus LMNB1_PBS 13 RT 29 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgtttagagctagaatagcaagttaaataagg ctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGG GTTCGCAGTCGCCAT Gatatggggaagtgaatcagtacaaccgccacagtac C GGGCGGCGGAGA	349
B. cereus NOLC1_PBS 13 RT 29 AttB 36 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggc tagtcggttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAA TGCCGGCGTCCGCC atatggggaagtgaatcagtacaaccgccacagtac TCC TCCAGGCAAT	350
B. cereus NOLC1_PBS	GGGCACGGGGCCATGTACAAgtttagagctagaatagcaagttaaataagg ctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GGCGAATGCCGGCG	351

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
13 RT 20 AttB 36 atgRNA	TCCGCCat atggggaagtgaatcagtagacaaccgccacagtagTCCTCCAGGCAAT	
B. cereus NOLC1_PBS 18 RT 20 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAGtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GGCGAATGCCGGCGTCCGCCat atggggaagtgaatcagtagacaaccgccacagtagTCCTCCAGGCAATACGCG	352
B. cereus SRRM2_PBS 9 RT 24 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAGtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GGCGTCGGCAGCCCCGATCCCGTTG atggggaagtgaatcagtagacaaccgccacagtagTACATGGCC	353
B. cereus SRRM2_PBS 9 RT 10 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAGtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GATCCCGTTG atggggaagtgaatcagtagacaaccgccacagtagTACATGGCC	354
B. cereus SRRM2_PBS 13 RT 10 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAGtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GATCCCGTTG atggggaagtgaatcagtagacaaccgccacagtagTACATGGCCCCGT	355
Screen validation guides ACTB_1_11_2 4_38	GCTATTCTCGCAGCTCACCAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcgcgcgcgatcatcatcatcatgatgatcctgacgacggagaccgcccgtcgtcgacaagccgctgagctgagcgag	356
Screen validation guides ACTB_1_16_1 8_43	GCTATTCTCGCAGCTCACCAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcgcgatcatcatcatcatgatgatcctgacgacggagaccgcccgtcgtcgacaagccgctgagctgagcgagaatag	357
Screen validation guides LMNB1_1_8_2 6_38	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcgcgcgacgggggtcgcagtcgcca	358
Screen validation guides NOLC1_1_15_ 16_43	CGGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcgcaatgccggcgctccgcccgatgatcctgacgacggagaccgcccgtcgtcgacaagccggtcctccaggcaatac	359
Screen validation	CGGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcgcgcgatcatcatcatcatgatgatcctgacgacggagaccgcccgtcgtcgacaagccgctcctccaggcaatac	360

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
guides NOLC1_1_14_10_38		
Screen validation guides SERPIN_13_32_38	GGGAAATGCATCTTGCACAAgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgcagcccctccatgctctctagctgtgccattgggctgtcgacgacggcggtctccgtcgcaggatcattgcaagatgcatt	361
Screen validation guides DEPDC4_8_10_38	GTGTCAGGTGGGGCGGGCTAgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgcggcaccataatgatcctgacgacggagaccgctgtcgacaagcccccgccc	362
SERPIN Nicking guide - 107 guide	GTGGGGACAGCCCCGTCTCTgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	363
SERPIN Nicking guide - 91 guide	GCTCTTGGGAAAAAAACCCTAgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	364
SERPIN Nicking guide - 90 guide	GTCTTGGGAAAAAAACCCTAAgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	365
SERPIN Nicking guide - 84 guide	GAAAAAAACCCTAAGGGCTGgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	366
SERPIN Nicking guide - 67 guide	GCTGAGGATCCTTGTGAGTGTgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	367
SERPIN Nicking guide - 66 guide	GTGAGGATCCTTGTGAGTGTgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	368
SERPIN Nicking guide - 63 guide	GGATCCTTGTGAGTGTGGGgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	369
SERPIN Nicking guide - 62 guide	GATCCTTGTGAGTGTGGGTgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	370
SERPIN Nicking guide - 49 guide	GTTGGGTGGGAACAGCTCCCgttttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	371

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
SERPIN Nicking guide - 46 guide	GGGTGGGAACAGCTCCCAGGgttttagagctagaatagcaagttaaataagg ctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	372
SERPIN Nicking guide +34 guide	GCTTCTGTGCAGCAGTTTCCGgttttagagctagaatagcaagttaaataagg ctagtccgttatc aactgaaaaagtggcaccgagtcggtgc	373
SERPIN Nicking guide +48 guide	GTTTCCCTGGCCACTAAATAGgttttagagctagaatagcaagttaaataagg ctagtccgttatc aactgaaaaagtggcaccgagtcggtgc	374
SERPIN Nicking guide +49 guide	GTTCCCTGGCCACTAAATAGTgttttagagctagaatagcaagttaaataagg ctagtccgttatc aactgaaaaagtggcaccgagtcggtgc	375
SERPIN Nicking guide +71 guide	GATTAGATAGAAGCCCTCCAgttttagagctagaatagcaagttaaataagg tagtccgttatc aactgaaaaagtggcaccgagtcggtgc	376
SERPIN Nicking guide +72 guide	GATTAGATAGAAGCCCTCCAgttttagagctagaatagcaagttaaataag gctagtccgttat caactgaaaaagtggcaccgagtcggtgc	377

6.8. Integrases/Recombinases and Integration/Recombination Sites

[0334] In typical embodiments, the co-delivery system described herein contains an integrase and/or a recombinase. In some embodiments, the co-delivery system includes an integrase and/or a recombinase packaged in a LNP. In one embodiment, the co-delivery system includes a polynucleotide encoding an integrase and/or a recombinase. In some embodiments, the co-delivery system includes an integrase or a recombinase packaged in a vector (e.g., a viral vector). In some embodiments, the co-delivery system includes at least a first integrase (e.g., a first integrase and a second integrase) and/or at least a first recombinase (e.g., a first recombinase and a second recombinase).

[0335] In some embodiments, the integration enzyme (e.g., the integrase or recombinase) is selected from the group consisting of Dre, Vika, Bxb1, ϕ C31, RDF, ϕ BT1, R1, R2, R3, R4, R5, TP901-1, A118, ϕ FCI, ϕ C1, MR11, TG1, ϕ 370.1, W β , BL3, SPBc, K38, Peaches, Veracruz, Rebeuca, Theia, Benedict, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Airmid, Benedict, Hinder, ICleared, Sheen, Mundrea, BxZ2, ϕ RV, retrotransposases encoded by a

Tc1/mariner family member including but not limited to retrotransposases encoded by LI, Tol2, Tel, Tc3, Himar 1 (isolated from the horn fly, *Haematobia irritans*), Mos1 (Mosaic element of *Drosophila mauritiana*), and Minos, and any mutants thereof. As can be used herein, Xu et al describes methods for evaluating integrase activity in *E. coli* and mammalian cells and confirmed at least R4, ϕ C31, ϕ BT1, Bxb1, SPBc, TP901-1 and W β integrases to be active on substrates integrated into the genome of HT1080 cells (Xu et al., 2013, Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. BMC Biotechnol. 2013 Oct 20;13:87. doi: 10.1186/1472-6750-13-87). Durrant describes new large serine recombinases (LSRs) divided into three classes distinguished from one another by efficiency and specificity, including landing pad LSRs which outperform wild-type Bxb1 in episomal and chromosomal integration efficiency, LSRs that achieve both efficient and site-specific integration without a landing pad, and multi-targeting LSRs with minimal site-specificity. Additionally, embodiments can include any serine recombinase such as BceINT, SSCINT, SACINT, and INT10 (see *Ionnidi et al., 2021; Drag-and-drop genome insertion without DNA cleavage with CRISPR directed integrases*. bioRxiv 2021.11.01.466786, doi.org/10.1101/2021.11.01.466786). In some embodiments, the integration site can be selected from an attB site, an attP site, an attL site, an attR site, a lox71 site a Vox site, or a FRT site.

[0336] It will be appreciated that desired activity of integrases, transposases and the like can depend on nuclear localization. In certain embodiments, prokaryotic enzymes are adapted to modulate nuclear localization. In certain embodiments, eukaryotic or vertebrate enzymes are adapted to modulate nuclear localization. In certain embodiments, the invention provides fusion or hybrid proteins. Such modulation can comprise addition or removal of one or more nuclear localization signal (NLS) and/or addition or removal of one or more nuclear export signal (NES). Xu et al compared derivatives of fourteen serine integrases that either possess or lack a nuclear localization signal (NLS) to conclude that certain integrases benefit from addition of an NLS whereas others are transported efficiently without addition, and a major determinant of activity in yeast and vertebrate cells is avoidance of toxicity. (Xu et al., 2016, Comparison and optimization of ten phage encoded serine integrases for genome engineering in *Saccharomyces cerevisiae*. BMC Biotechnol. 2016 Feb 9; 16:13. doi: 10.1186/s12896-016-0241-5). Ramakrishnan et al. systematically studied the effect of different NES mutants developed from mariner-like elements (MLEs) on transposase localization and activity and concluded that nuclear export provides a means of controlling transposition activity and

maintaining genome integrity. (Ramakrishnan et al. *Nuclear export signal (NES) of transposases affects the transposition activity of mariner-like elements Ppmar1 and Ppmar2 of moso bamboo*. Mob DNA. 2019 Aug 19;10:35. doi:10.1186/s13100-019-0179-y). The methods and constructs are used to modulate nuclear localization of system components of the invention.

[0337] In typical embodiments, the integrase used herein is selected from below (Table 10).

Table 10 - Integrases

Data base	SRA accession	bioproject acc	nucleotide accession	protein accession or ORF ID	internal protein ID	Proposed names	Alternative names	organism/source	description	Sequence	SEQ ID NO:	Length	Group
ENA	SRS1205298	PRJB26277	NA	NA	N189929_49_54	SsuINT	NA	human gut metagenome	stool sample from male in USA	MEKNRAVLYLRLSKEDVDKVNKGDDSSSIKSQRLLLTDFALERGFKIVGVYSDDDSEGLYDDRPDFERMMTDAKLDEFDIIIKTQSRFSRNMEHIEKYLHHDLPNLGIRFIGAVDGVDTESDENKKS RQINGLVNEWYCEDLSKNIRS AFKAKMKDGGQLGSSCPYGYKKDPQNHNLVDDYAAKVVKQIFNLYLEGYGKAKIGSILSSEGILIPITLYKKDILKQNYHNSKALDTTQNWSYQTIHTILNNEVYLGHLIQNKVNTMSYKDKNKRI LPKEKWIIVRNTHPEIITEEMFQDVQKLQKNRTRSVENIEPNGLFSGLIFCADCKHAMSRYARRGEKGFVGYVCKTYKTQGKNFCESHSIDYDELEEAFLFSIKNEARSILQQEEIDELRKVQAYDET KSYEMQLENIKSMEKIEKYKKKTYDNYMDDLISRDDYKKYVTEYDKEIGGLKQQQELINSKTDLEKEISTQYDEWVEAFINYVDIDKLTREIVIELIEKIEVNKDG SINIYYKFKNPYIS	378	527	INTc
ENA	ERS396461	PRJEB26280	NA	NA	N190156_234_12	SsuINT	NA	human gut metagenome	stool sample from Spain	MNTVIYARYSAGPRQTDQSIDGQLRVCTEFCKQRLTVVDTYCDRHISGRTERPEFQRLIADAKAHKFEAVVVYKTDRFARNKYDSAIYKRELRRNGIQIFYAAEAIPEGPEGIILESLMEGLAEYYS AELAQKIKRGLNESALKCQSLGSGRPLGYTVDEQKHQFQIDPESQAVKTIFEMYIKGESNAACDYLNARGLRTSQGNLFNKN SINRIIKNRKYIGEYRYNDIVVEGGMPIIISKETFCMAQAEMERRRTHRAPHVSPKAEYLLAGKLFCHCKGPMQGVSGTGKSGNKWY YYYCANTRGKERTCDKKQVSRDRLEKAVVDFTVR YILQENVLEELSKKVYAAQERQNTASEIAFYEKKLAENKKAIANILRAIESGAMTQALPARLQELENEQTVIQGELSYLK GARLAFTEDQILFALLQHLDPRPGESERDYHRRITIDFVSEVYLYDDRMLIYFNISADGKLLKHADLSAIESGVFDA GLISSSSRASSFSTRCALI	379	510	INTd

ENA	ERS 1015 837	PRJE B2 6832	NA	NA	N1913 52_14 3_72	Ssc1N T	NA	human gutmetag enome	stool sample from China	MNEKNLEIGAAYIRVSTDDQT ELSPDAQLRVILEAAKKDGIIP QEFVFMEDRGRSGRRADNRPE FQRMISTARQNPSPFRYL YLW KFSRFARNQEESAFYKGILRKK CGVTIKSVSEPIEMGMFGRLVE MIIWSDDEFYSVNLSGEVLRG MTQKALEHGYQLTPCLGYDA VGHGRPYVINEEQYQIVFIHR SFFDGKDMTWIAREANRRGY HTRRGPNPFDTRAVRIILNSFY VGLVKWNDVTFQGTHECRES VTSVFSANQERLNRHPRGRR QASSCKHWLSGLLKCICGAS LGYNQTKDLTKRGHAFQCWK YTKGIHPGSCSVSSLKAEAAVL ESLQMILETGEVEYTYEQREK HLDDNKLTLIQKSLERLDTKEL RIREAYESGIDTLEDFKTNKAR LQREDQLMEELEELHSQEEP EDVPKGKILIERIQNVYDLLQSP DVDNDDKGNVRSIIKKIVYIK ESKTFCFYYYV	380	482	INT d
ENA	ERS 1289 677	PRJE B2 6924	NA	NA	N1915 33_22 4_76	Ssc2I NT	NA	human gutmetag enome	stool sample from China	MERTIKVIQPGTVKIPTKKRVA AYARVSSGKDAMLHLSAQVS YYSNMIQKNEWSYVGIYADE AITGKDRRVEFNRLIQDCTDG KIDMIITKSISRFAFNLTMLLEV VRKLKNINVDVYFEKENIHSIS GDGELMLTILASFAQEESRSVS ENCKWRIRKGFEQGELINLRFL YGYRINKGKIEIYEKAEIVRM IFDDYLNAGECTRIGNKLKRM KVNKL RGGMWNSER VVDIHK NEKYTGALLQKKYVKDHLS KKLVRNKGILTQYYAEGTHPA IIDIKTFEIAQKIMEANRTKFQ KCGSNRYLFTSKIECGICGKNY RHKDREGKSTWVCANHLKYG NSRCIAKPLNEEKLKLINEAL ELKYFDEEIFIRNIKRIVTGNO TIEFILKDGVIEEGMI	381	406	INT c
ENA	ERS 265 5827	PRJ EB2 8245	NA	NA	N2039 11_45 186_6	Ssd1N T	NA	human gut metageno me	stool sample from Denmark	MKKIKIDRAIQERPATRKQTRN EKIRQSLTEHVDVQVIPAITDR EGYEKPKLRVCAYCRVSTDM DTQALSYELQVQNYTDYIRGN DEWRFAGIYADRGISGTSKHK RDEFNRMIEDCKAGKIDLIITK AVTRFARNVLDICISTIRMLKQL EHPVAVYFETERINTLDTTSET YLGLISLFAQGESESKSESLKW SYIRRWKRGTGIYPAWSLLGY EMGEDGKWOIVEAEAEELVRII YDMYLNQYSSPQIAELTRSGV PTATNQTVWSSGGVLGILRNE KYCGNVLCQKTMVDFVSHK AIKNTGQKTQYFIEGHDPILR SDWDRVQQMIDEKYYRKRRG RRTKPRIVLKGCLAGFTQIDLD WDEDDIARIFYSTTPAAEVATP AMADHIEIHKVGEN	382	401	INT c

ENA	SRS 2949 42	PRJ EB3 0046	NA	NA	N2086 21_9_15	SmcI NT	NA	human gut metageno me	sample from 72- year-old male from China	MKTAAAYIRVSTDDQVEYSPD SQIKLIRDYAKRNDYILPDEFIF RDDGISGKSAKHRPEFTKMIAL AKSPEHPFDAILVWKFSRFARN QEEISIVFKNLRKIGVEVRSVSE PISEDPFGSLVERHIEWTDEYYII NLSGEVKRGMLEKISRGPVV PPPVGYKMENGQYIPDENAHFI KEIFEAYAAEGGARHIAQRLA AQGCLTKRGNPIDNRFVDYVL HNPVYIGKLRWSVNSHAASSR HYDSADIIVFDGTHEPLISSEL WESVQKRLHEVKTLYPKYQR REQPVSFMLKGLVRCSSCGST LCYCRTSEPSLQCHSYARGSCR QSHSINIATANEAVIKGLQLAV DKLDFAIAPAKPHYSADAPGT NKLLAAEYKKMERIKAAAYAN GTDITLEEY AANKKKISAEIARL EAELQQESNVK PINKK AFAKR VSEIKYISDPHNSEAAKNQAL RTVISYIIFDRAATTFNIFHF	383	476	INT d
Met aSU B	NA	NA	NA	NA	N6750 15_95_5	UhmI NT	NA	urban human microbio me	NA	MKIAIYARKSKYSPTGESVENQ IQLCKEYLQAKYKSETLEIDY KDEGYSGGNTNRPDFKKLIAQI EDYDMLICYRLDRISNVADFS STLTLLQNNKCDVFSIKEQFDT TSPMGRAMIYISSVFAQLERET IAERIRDNMMELAKMGRWL GTIPMGFDSEPTIFIDENMKERS MTKLIPNVEELKVIELIYEKYL QLGSMGKVVTYLLQNNIKTKK GKDFTLGSIKVILTNPIYVKAN QEVVNHKLTQGITICGDVDGK KALLTYNKTGTISNDVGTKTIV KDKSEWIAAVANHKGIIPADK WLQAQNIKDKNKDSFALGRS NTTIASRVLRCDKCESTMVGT HGHINPVTGKKHYYYNCTLKK RSKGVRCDNKAAPAAEVDEAI LITLENMFKAASSIIDNLKAKN KARRIEMISSNRVDVINKIIEDK TKQIDNLVNKLSDDDLTDILF KKIKGLKAEIKELEDELLTLTS DNIKLNEDEVVLDFTKLEK CSHRTL DILEQQQIVDALIPLVT WNGDTEVLNIYPLGSPELELKE AESKKK	384	550	INT d
Sega ta- Pasol li	NA	PRJ NA4 2243 4	NA	NA	N6843 46_90_69	SacI NT	NA	human gut metageno me	stool sample from adult in China	MKEKVSEKRTGAIYIRVSTDK QEELSPDAQLRLLLDYAKKDSI DVPKEYIFQDNGISGRKANKRP AFQNMIALAKSKEHPIDTIIVW KFSRFARNQEEISIVYKSLKKN NVDVVSVSEPLIDGPFGLIERI IEWMDEYYSIRLSGEVMRGMT QNAMRGHYQSDAPIGYTSPGD KKPPVINPDTVQIPLMIKDMFL SGSTQLQIARKLNDSGYRTKR GNLWDARGVRYVLENPFYIGK SRWNYTERGRRLKPADEVYA DGNWEALWDEDTFKEIQKRL ALNMRKSKSRDISAAKHWLSG LLICSSCGGTAFGGAHNMRG FQCWKYSKGFCSESHYISTGPI EKMVLEYLEAVMHSPALSYTV ISSSV DASSKLSDLERQLQKID AKEKRIKAAAYLNEIDTLEEK ANKTAL EEEERTVEKEIEELTL SDVKYSKEDLDKMKQNISDL LRVLRDESADYIQKGNMMRN VVDHIVFNKNTSLDVLKLV V	385	493	INT d

Sega ta- Paso Li	ERR 1136 864	PRJ EB11 532	NA	NA	N6876 11_90 _68	RsaIN T	NA	human gut metageno me	rectal swab from adult in Isreal	MKITKKQPLRPRGRSEDKRQS TKNVIRDAYINGPQKEVQIIPA KRDMEAETEEKKLRLVCAYCR VSTDEDTQASSYELQVQNYTR MIRENPEWEFAGIFADEGISGT SVLHREHFLEMIEKCKAGEIDL IITKQVSRFARNVLDLNYIFM LRKLDPPVGVYFETEKLNTLD KSSDMVITVLSLVAQSESEQKS NSLKWSFKRRRAQGLGIYPSW ALLGYRLDDEKNWEIVEDEAD IVRTIYSLYLDGYSSTQIAELLT KSGIPTVKGLSVWSSGSVLGIL KNEKFCGDALCQKTVTIDFFT HKS VKNNGIEPQYFVEGHHIPII EKNDWLLAQQIRKERRYKRKR STHRKPRIVVKGALSGFMIVDT SWDEEYVDSLISATQKPEPAP VIAEEDENFIVIEKE	386	404	INT c
Segat a- Paso lli	ERR 1136 737	PRJE B115 32	NA	NA	N6876 63_53 _29	Rsa2I NT	NA	human gut metageno me	rectal swab from adult in Isreal	MADIQPVKNGALYIRVSTHLQ EELSPDAQKRLLMYAEAHNII VLKEHIYIDSGISGRSARQRPQF NNMIAEAKSKEHPFDVILVWK YSRFARNQEESIVYKSMKRE NVDVISVSEPISSDPFGSLIERII EWMDEYYIRLSGEVSRGMAE NAMRGNYQARPPLGYRIPGYR QTPVIVPEEAELIQLIFDLYTEK KMGIFEIVRYLNEHGYQTGHK KPFQRRSVTYILKNPTYIGKTI WNQHDQDHKLRLDKSEWIIAD GKHEPIISKEQFDKAQKRIEST YKPAYRKPTSVCHHWLSSLLK CSSCGRTL VVKRTASKKKDRM YVNFQCYGYQKGICNTNQSIS AIKLEPVIMHALEDAMTSGKIH FDVLNPTLDSSQKQFLTRLN EIEKKEERIKRAYRDGIDTLEE YKENKSIIQTEKEMLLKKIEHIE EPALSPEEAKPIMMDRIKNVYE IITNPDIGMEEKNKAARSIIKEI VDRATGSVNIFFYLAHCP	387	498	INT d
NC BI	NA	NA	NC 0 02656. 1	NP 0 75302. 1	NA	BxbIN T	Bxb1 integras e	Mycobact erium phage Bxb1	NA	MRALVVIRLSRVTDATTSPER QLESCQQLCAQRGWDVVGVVA EDLDVSGAVDPFDRKRPNLA RWLA FEEQPFVIVAYRVDRL TRSIRHLQQLVHWAEDHKKL VSATEAHFDTTTTFAAVVIAL MGTVAQMELEAIKERNRSAA HFNIRAGKYRGSLLPPWGYLPT RVDGEWRLVPDPVQRERILEV YHRVVNDNHEPLHLVAHDLNR RGVLSPKDYFAQLQGREPQGR EWSATALKRSMISEAMLGYAT LNGKTVRDDDGA PLVRAEPIL TREQLEALRAELVKTSRAKPA VSTPSLLLRVLFCVAVCGEPAYK FAGGGRKKHPRYRCRSMGFPHK CGNGTVAMA EWDAFCEEQVL DLLGDAERLEK VVWAGSDSA VELAEVNAELVDL TSLIGSPAY RAGSPQREALDARIAALAAARQ EELEGLEARPSGWEWRETGQR FGDWWREQDTAAKNTWLRS MNVRLTFDVRGGLTRTIDFGD LQEYEQHLRLGVSVERL HTGMS*	388	501	INT a

NC BI	NA	NA	NC_0 02747.1	NP_11 2664.1	NA	TP9IN T	TP901- lintegr ase	Lactococ us phage TP901-1	NA	MITKKVAIYTRVST'TNQAEEGF SIDEQIDRLTKYAEAMGWQVS DTYTDAGFSGAKLERPAMQRL INDIENKAFDTVLVYKLDRLSR SVRDTLYLVKDVFTKNKIDFIS LNESEDTSAMGSLFLTILSAIN EFERENIKERM'TMGKLGRAKS GKSMMWTKTAFGYHHNRKT GILEIVPLQATTIVEQIFTDYLSGI SLTKLRDKLNEIGHIGKDIPWS YRTLRTLDNPVYCGYIKFKD SLFEGMHKPIIPEYETLYKVQKE LEERQQQTYERNNNPRPQAK YMLSGMARCGYCGAPLKIVL GHKRKDGSRTMKYHCANRFP RKTGKITVYNDNKKCDSGTYD LSNLENTVIDNLIGFQENNDL LKIINGNNQPILODTSFQKQISQI DKKIQKNSDLYLNDFTTMDL KDRTDSLQAEKKLLKAKISEN KFNDSTDVFEVLVKTQLGSIPIN ELSYDNKKKIVNNLVSKVDVT ADNVDIIFKFQLA*	389	486	INT d
NC BI	NA	NA	NC_0 04664.2	NP_8 13744.2	NA	Bt1IN T	PhiBTin tegrase	Streptom yces virus phiBT1	NA	MSPFIAPDVPEHLLDTRVFLY ARQSKGRSDGSDVSTEAQLAA GRALVASRNAQGGARVWVAG EFVDVGRSGWDPNVTRADFER MMGEVRAGEGDVVVVNELSR LTRKGAHDALEIDNELKKHGV RFMSVLEPFLDTSTPIGVAIFAL IAALAKQDSDLKAERLKGAKD EIAALGGVHSSAPFGMRAVR KKVDNLVISVLEPDEDNPDHV ELVERMAKMSFEGVSDNAIAT TFEKEKIPSPGMAERRATEKRL ASIKARRLNGAEKPMWRAQT VRWILNHPAIGGFAFERVKHG KAHINVIRRDPPGKPLTPHTGI LSGSKWLELQEKRSKNSDR KPGAEEPTLLSGWRFLGCRIC GGSMGQSQGGKRNGDLAEG NYMCANPKGHGGLSVKRSEL DEFVASKVWARLRTADMEDE HDQAWIAAAERFALQHDLA GVADERREQQAHLNVRRSIK DLQADRKAGLYVGREELETW RSTVLQYRSYEAECTTRLAEL DEKMNGSTRVPSEWFSGEDPT AEGGIWASWDVYERREFLSFF LDSVMVDRGRHPETKKYIPLK DRVTLKWAEELKEEDEASEAT ERELAAL*	390	595	INT a

NC BI	NA	NA	NC_011658.1	WP_00286206.1	NA	BceINT	NA	Bacillus creus AH187	NA	MYPYDVPDYAGSYRPESLDVC IYLKRSRKDVEEERRAIEEGSS YNALERHRKRLFAIAKAENHN IIDIFEEVASGESIQERPQMQL LRKLEGNEIDGVLVIDLDRGR GDMLDAGMIDRAFRYSSTKIIT PTDVYDPDESWEVFGIKSLI SRQELKSITKRLQNGRIDSVE GKHIGKKPPYGYLKDENRLY PDPEKAWIVKKIFELMCDGKG ROMIAAELDRLGIDPPVTKRG AWDSSTITSIIKNEVYTGIVW GKFKHKRNGKYTRHKNPQE KWIMYENAHEPIISKELFDAAN EAHSSRHKPAVITSKLTNPLA GILKCKLCGYTMLIQTRKDRP HNYLRCNNPACKGKQKQSVF NLVEEKLLYSLQQIVDEYQAQ KVEEVEIDDSKLISFKEKAISK EKELKELQAQKGNLHDLLEQG IYTVEIFLERQKNLVERITSIN DIEVLQKEIETEIQIHEHNKTEFI PALKTVIESYHKTTNIELKNQL LKTILSTVTYRHPDWKTNEF EIQVYFKIS*	391	529	INT c
NC BI	NA	NA	NC_009674.1	WP_012095429.1	NA	BcyINT	NA	Bacillus cytotoxicus NVH391-98	NA	MYPYDVPDYAGSAVGIIYRVS TQEQAEGHSIESQKKLASYS EIQGWDDYRFYIEEGISGKNTN RPKLKLLMEHIEKGKINILLVY RLDRLTRSVIDLHKLLNQLQEH GCAFKSATETYDTTANGRMS MGIVSLLAQWETENMSERIKL NLEHKVLVEGERVGAIPYGF LSDDEKLVKNEKSAILDMVE RVENGWSVNRIYNLNTNN DRNWSPNGVLRLLRNPALYG ATRWNDKIAENTHEGIISKERF NRLQQILADRSIHRRDVKGT YIFQGVLRCPVCDQTLVNRFI KKRKDGTEYCGVLYRCQPCIK QNKYNLAIGEARFLKALNEYM STVEFQTVEDVIPKKSEREM ESQLQQIARKREKYQKAWASD LMSDDEFELKLMVETRETYDEC KQKLESCEDPIKIDETYLKEIV YMFHQTFNDLESEKQKEFISKF IRTIRYTVKEQQPIRPDKSKTG KKGKQKVIITEVEFYQS*	392	487	INT d
NC BI	NA	NA	NC_017353.1	WP_014533238.1	NA	SluINT	NA	Staphylococcus lugdunensis N920143	NA	MYPYDVPDYAGSKVAIYTRVS SAEQANEGYSIHEQKKKLISYC EIHDWNEYKVFTDAGISGGSM KRPALQKLMKHLSSFDLVVY KLDRLTRNVRLDMLLEEFQ YNVSFKSATEVFDTTSAIGKLF ITMVGAMAEWERETIRERSLF GSRAAVREGNYIREAPFCYDNI EGKLHPNEYAKVIDLIVSMFK KGISANEIARRLNSSKVHVPNK KSWNRNSLIRLMRSPVLRGHT KYGDMLIENTHEPVLSEHDYN AINNAISSKTHKSKVKHHAIFR GALVCPQCNRRHLHYAGTVK DRKGYKYDVRRYKCETCSKN KDVKNVSFNESEVENKFNLL KSYELNKFHIRKVEPVKKIEYD IDKINKQKINYTRSWSLGYIED DEYFELMEEINATKKMIEEQTT ENKQSVSKEQIQSINNFILKGW EELTIKDKEELISTVDKIEFNFI PKDKKHK TNTLDINNHFKFS*	393	473	INT d

[0338] Sequences of insertion sites (i.e., recognition target sites) suitable for use in embodiments of the disclosure are presented below (Table 11). FIGs. 14A-14E shows analysis of effect of variant AttP sites on integration efficiency.

Table 11				
Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bxb1_AttP_GT_original_site	GTGGTTTGTCTGGTCAAC CACCGCGGTCTCAGTGGT GTACGGTACAAACCCA	SEQ ID NO: 394	TGGGTTTGTACCGTACACCACTGAGA CCGCGGTGGTTGACCAGACAAACCAC	SEQ ID NO: 473
Bxb1_AttP_CG_site	GTGGTTTGTCTGGTCAAC CACCGCGcCTCAGTGGTG TACGGTACAAACCCA	395	TGGGTTTGTACCGTACACCACTGAGC GCGCGGTGGTTGACCAGACAAACCAC	474
Bxb1_AttP_GC_site	GTGGTTTGTCTGGTCAAC CACCGCGcCTCAGTGGTG TACGGTACAAACCCA	396	TGGGTTTGTACCGTACACCACTGAGG CCGCGGTGGTTGACCAGACAAACCAC	475
Bxb1_AttP_AT_site	GTGGTTTGTCTGGTCAAC CACCGCGatCTCAGTGGTG TACGGTACAAACCCA	397	TGGGTTTGTACCGTACACCACTGAGA TCGCGGTGGTTGACCAGACAAACCAC	476
Bxb1_AttP_TA_site	GTGGTTTGTCTGGTCAAC CACCGCGtaCTCAGTGGTG TACGGTACAAACCCA	398	TGGGTTTGTACCGTACACCACTGAGT ACGCGGTGGTTGACCAGACAAACCAC	477
Bxb1_AttP_GG_site	GTGGTTTGTCTGGTCAAC CACCGCGgCTCAGTGGTG TACGGTACAAACCCA	399	TGGGTTTGTACCGTACACCACTGAGC CCGCGGTGGTTGACCAGACAAACCAC	478
Bxb1_AttP_TT_site	GTGGTTTGTCTGGTCAAC CACCGCGttCTCAGTGGTG TACGGTACAAACCCA	400	TGGGTTTGTACCGTACACCACTGAGA ACGCGGTGGTTGACCAGACAAACCAC	479
Bxb1_AttP_GA_site	GTGGTTTGTCTGGTCAAC CACCGCGgaCTCAGTGGTG TACGGTACAAACCCA	401	TGGGTTTGTACCGTACACCACTGAGT CCGCGGTGGTTGACCAGACAAACCAC	480
Bxb1_AttP_AG_site	GTGGTTTGTCTGGTCAAC CACCGCGagCTCAGTGGTG TACGGTACAAACCCA	402	TGGGTTTGTACCGTACACCACTGAGC TCGCGGTGGTTGACCAGACAAACCAC	481
Bxb1_AttP_CC_site	GTGGTTTGTCTGGTCAAC CACCGCGccCTCAGTGGTG TACGGTACAAACCCA	403	TGGGTTTGTACCGTACACCACTGAGG GCGCGGTGGTTGACCAGACAAACCAC	482
Bxb1_AttP_TC_site	GTGGTTTGTCTGGTCAAC CACCGCGtcCTCAGTGGTG TACGGTACAAACCCA	404	TGGGTTTGTACCGTACACCACTGAGG ACGCGGTGGTTGACCAGACAAACCAC	483
Bxb1_AttP_CT_site	GTGGTTTGTCTGGTCAAC CACCGCGctCTCAGTGGTG TACGGTACAAACCCA	405	TGGGTTTGTACCGTACACCACTGAGA GCGCGGTGGTTGACCAGACAAACCAC	484
Bxb1_AttP_AA_site	GTGGTTTGTCTGGTCAAC CACCGCGaaCTCAGTGGTG TACGGTACAAACCCA	406	TGGGTTTGTACCGTACACCACTGAGT TCGCGGTGGTTGACCAGACAAACCAC	485
Bxb1_AttP_CA_site	GTGGTTTGTCTGGTCAAC CACCGCGcaCTCAGTGGTG TACGGTACAAACCCA	407	TGGGTTTGTACCGTACACCACTGAGT GCGCGGTGGTTGACCAGACAAACCAC	486

Table 11				
Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bxb1_AttP_AC_site	GTGGTTTGTCTGGTCAAC CACC ^G GacCTCAGTGGTG TACGGTACAAACCCA	408	TGGGTTTGTACCGTACACCACTGAGG TCGCGGTGGTTGACCAGACAAACCAC	487
Bxb1_AttP_TG_site	GTGGTTTGTCTGGTCAAC CACC ^G GtgCTCAGTGGTG TACGGTACAAACCCA	409	TGGGTTTGTACCGTACACCACTGAGC ACGCGGTGGTTGACCAGACAAACCAC	488
Bxb1_AttB_46_GT_original_site	GGCCGGCTTGTGCGACGAC GGCGGTCTCCGTCGTCAG GATCATCCGG	410	CCGGATGATCCTGACGACGGAGACCG CCGTCGTCGACAAGCCGGCC	489
Bxb1_AttB_46_AA_site	GGCCGGCTTGTGCGACGAC GGCG ^a aCTCCGTCGTCAGG ATCATCCGG	411	CCGGATGATCCTGACGACGGAGTTCG CCGTCGTCGACAAGCCGGCC	490
Bxb1_AttB_46_GA_site	GGCCGGCTTGTGCGACGAC GGCG ^{ga} aCTCCGTCGTCAGG ATCATCCGG	412	CCGGATGATCCTGACGACGGAGTCCG CCGTCGTCGACAAGCCGGCC	491
Bxb1_AttB_46_CA_site	GGCCGGCTTGTGCGACGAC GGCG ^{Ga} aCTCCGTCGTCAGG ATCATCCGG	413	CCGGATGATCCTGACGACGGAGTGCG CCGTCGTCGACAAGCCGGCC	492
Bxb1_AttB_46_TA_site	GGCCGGCTTGTGCGACGAC GGCG ^{Ga} aCTCCGTCGTCAGG ATCATCCGG	414	CCGGATGATCCTGACGACGGAGTACG CCGTCGTCGACAAGCCGGCC	493
Bxb1_AttB_46_AG_site	GGCCGGCTTGTGCGACGAC GGCG ^{ag} agCTCCGTCGTCAGG ATCATCCGG	415	CCGGATGATCCTGACGACGGAGCTCG CCGTCGTCGACAAGCCGGCC	494
Bxb1_AttB_46_GG_site	GGCCGGCTTGTGCGACGAC GGCG ^{gg} ggCTCCGTCGTCAGG ATCATCCGG	416	CCGGATGATCCTGACGACGGAGCCCG CCGTCGTCGACAAGCCGGCC	495
Bxb1_AttB_46_CG_site	GGCCGGCTTGTGCGACGAC GGCG ^{Gg} gCTCCGTCGTCAGG ATCATCCGG	417	CCGGATGATCCTGACGACGGAGCGCG CCGTCGTCGACAAGCCGGCC	496
Bxb1_AttB_46_TG_site	GGCCGGCTTGTGCGACGAC GGCG ^{Gg} gCTCCGTCGTCAGG ATCATCCGG	418	CCGGATGATCCTGACGACGGAGCACG CCGTCGTCGACAAGCCGGCC	497
Bxb1_AttB_46_AC_site	GGCCGGCTTGTGCGACGAC GGCG ^{Gac} acCTCCGTCGTCAGG ATCATCCGG	419	CCGGATGATCCTGACGACGGAGGTTCG CCGTCGTCGACAAGCCGGCC	498
Bxb1_AttB_46_GC_site	GGCCGGCTTGTGCGACGAC GGCG ^{Ggc} gcCTCCGTCGTCAGG ATCATCCGG	420	CCGGATGATCCTGACGACGGAGGCCG CCGTCGTCGACAAGCCGGCC	499
Bxb1_AttB_46_CC_site	GGCCGGCTTGTGCGACGAC GGCG ^{Gcc} ccCTCCGTCGTCAGG ATCATCCGG	421	CCGGATGATCCTGACGACGGAGGGC GCCGTCGTCGACAAGCCGGCC	500
Bxb1_AttB_46_TC_site	GGCCGGCTTGTGCGACGAC GGCG ^{Gtc} tcCTCCGTCGTCAGG ATCATCCGG	422	CCGGATGATCCTGACGACGGAGGAC GCCGTCGTCGACAAGCCGGCC	501
Bxb1_AttB_46_AT_site	GGCCGGCTTGTGCGACGAC GGCG ^{Gat} atCTCCGTCGTCAGG ATCATCCGG	423	CCGGATGATCCTGACGACGGAGATCG CCGTCGTCGACAAGCCGGCC	502

Table 11				
Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bxb1_AttB_46_CT_site	GGCCGGCTTGTCGACGAC GGCGctCTCCGTCGTCAGG ATCATCCGG	424	CCGGATGATCCTGACGACGGAGAGC GCCGTCGTCGACAAGCCGGCC	503
Bxb1_AttB_46_TT_site	GGCCGGCTTGTCGACGAC GGCGttCTCCGTCGTCAGG ATCATCCGG	425	CCGGATGATCCTGACGACGGAGAAC GCCGTCGTCGACAAGCCGGCC	504
Bxb1_AttB_38_GT_site	GGCTTGTCGACGACGGCG GTCTCCGTCGTCAGGATC AT	426	ATGATCCTGACGACGGAGACCGCCGT CGTCGACAAGCC	505
Bxb1_AttB_38_AA_site	GGCTTGTCGACGACGGCG aaCTCCGTCGTCAGGATCA T	427	ATGATCCTGACGACGGAGTTCGCCGT CGTCGACAAGCC	506
Bxb1_AttB_38_GA_site	GGCTTGTCGACGACGGCG gaCTCCGTCGTCAGGATCA T	428	ATGATCCTGACGACGGAGTCCGCCGT CGTCGACAAGCC	507
Bxb1_AttB_38_CA_site	GGCTTGTCGACGACGGCG caCTCCGTCGTCAGGATCA T	429	ATGATCCTGACGACGGAGTGCGCCGT CGTCGACAAGCC	508
Bxb1_AttB_38_TA_site	GGCTTGTCGACGACGGCG taCTCCGTCGTCAGGATCA T	430	ATGATCCTGACGACGGAGTACGCCGT CGTCGACAAGCC	509
Bxb1_AttB_38_AG_site	GGCTTGTCGACGACGGCG agCTCCGTCGTCAGGATCA T	431	ATGATCCTGACGACGGAGCTCGCCGT CGTCGACAAGCC	510
Bxb1_AttB_38_GG_site	GGCTTGTCGACGACGGCG ggCTCCGTCGTCAGGATCA T	432	ATGATCCTGACGACGGAGCCCGCCGT CGTCGACAAGCC	511
Bxb1_AttB_38_CG_site	GGCTTGTCGACGACGGCG cgCTCCGTCGTCAGGATCA T	433	ATGATCCTGACGACGGAGCGCGCCGT CGTCGACAAGCC	512
Bxb1_AttB_38_TG_site	GGCTTGTCGACGACGGCG tgCTCCGTCGTCAGGATCA T	434	ATGATCCTGACGACGGAGCACGCCGT CGTCGACAAGCC	513
Bxb1_AttB_38_AC_site	GGCTTGTCGACGACGGCG acCTCCGTCGTCAGGATCA T	435	ATGATCCTGACGACGGAGGTTCGCCGT CGTCGACAAGCC	514
Bxb1_AttB_38_GC_site	GGCTTGTCGACGACGGCG gcCTCCGTCGTCAGGATCA T	436	ATGATCCTGACGACGGAGGCCGCCGT CGTCGACAAGCC	515
Bxb1_AttB_38_CC_site	GGCTTGTCGACGACGGCG ccCTCCGTCGTCAGGATCA T	437	ATGATCCTGACGACGGAGGGCGCCGT CGTCGACAAGCC	516
Bxb1_AttB_38_TC_site	GGCTTGTCGACGACGGCG tcCTCCGTCGTCAGGATCA T	438	ATGATCCTGACGACGGAGGACGCCGT CGTCGACAAGCC	517
Bxb1_AttB_38_AT_site	GGCTTGTCGACGACGGCG atCTCCGTCGTCAGGATCA T	439	ATGATCCTGACGACGGAGATCGCCGT CGTCGACAAGCC	518

Table 11				
Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bxb1_AttB_38_CT_site	GGCTTGTCGACGACGGCG ctCTCCGTCGTCAGGATCA T	440	ATGATCCTGACGACGGAGAGCGCCGT CGTCGACAAGCC	519
Bxb1_AttB_38_TT_site	GGCTTGTCGACGACGGCG ttCTCCGTCGTCAGGATCA T	441	ATGATCCTGACGACGGAGAACGCCGT CGTCGACAAGCC	520
Cre Lox 66 site	TACCGTTCGTATAATGTA TGCTATACGAAGTTAT	442	ATAACTTCGTATAGCATACATTATAC GAACGGTA	521
Cre Lox 71 site	ATAACTTCGTATAATGTA TGCTATACGAACGGTA	443	TACCGTTCGTATAGCATACATTATAC GAAGTTAT	522
TP901-1 minimal AttB site	TTTACCTTGATTGAGATGT TAATTGTG	444	CACAATTAACATCTCAATCAAGGTAA A	523
TP901-1 minimal AttP site	GCGAGTTTTTATTTTCGTTT ATTTCAATTAAGGTAAC AAAAAATCCTTT	445	AAAGGAGTTTTTTAGTTACCTTAATT GAAATAAACGAAATAAAAACTCGC	524
PhiBT1 minimal AttB site	CTGGATCATCTGGATCAC TTTCGTCAAAAACCTG	446	CAGGTTTTTGACGAAAGTGATCCAGA TGATCCAG	525
PhiBT1 minimal AttP site	TTCGGGTGCTGGGTGTT GTCTCTGGACAGTGATCC ATGGGAACTACTCAGCA CCA	447	TGGTGCTGAGTAGTTTCCCATGGATC ACTGTCCAGAGACAACAACCCAGCAC CCGAA	526
Bacillus_cereus_AH1 87_Int30_38 bp_Att	gatatggggaagtgaatcagtacaaccgc cacagtacc	448	ggtactgtggcggtgtactgattcacttccccatac	527
Staphylococcus_lugdunensis_N920143_Int1 2_38bp_Att	tgggtggtacaggtgccattagtgtac catttatg	449	cataaatggtacaactaatgtggcacctgtaccacca	528
Bacillus_cytotoxicus_NVH_391-98_Int13_38 bp_Att	gttgtttttccagatccagttggctctgtaa tataag	450	cttatattacaggaccaactggatctggaaaaacaac	529
Bacillus_cereus_AH1 87_Int30_Att_30	tggggaagtgaatcagtacaaccgccac ag	451	ctgtggcggtgtactgattcacttcccca	454
Bacillus_cereus_AH1 87_Int30_Att_28	ggggaagtgaatcagtacaaccgccaca	452	tgtggcggtgtactgattcacttcccc	455
Bacillus_cereus_AH1	gggaagtgaatcagtacaaccgccac	453	gtggcggtgtactgattcacttccc	456

Table 11				
Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
87_Int30_Att_26				
Bacillus_cereus_AH1 87_Int30_Att_rc_30	ctgtggcgggtgtactgattcacttcccca	454	tgggaagtgaatcagtacaaccgccacag	451
Bacillus_cereus_AH187_Int30_Att_rc_28	tgtggcgggtgtactgattcacttccc	455	gggaagtgaatcagtacaaccgccaca	452
Bacillus_cereus_AH187_Int30_Att_rc_26	gtggcgggtgtactgattcacttccc	456	gggaagtgaatcagtacaaccgccac	453
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_30	ttttccagatccagttggctcctgtaaata	457	tatttacaggaccaactggatctggaaaaa	460
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_28	ttttccagatccagttggctcctgtaaat	458	atttacaggaccaactggatctggaaaa	461
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_26	tttccagatccagttggctcctgtaaa	459	tttacaggaccaactggatctggaaa	462
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_rc_30	tatttacaggaccaactggatctggaaaaa	460	ttttccagatccagttggctcctgtaaata	457
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_rc_28	atttacaggaccaactggatctggaaaa	461	ttttccagatccagttggctcctgtaaat	458
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_rc_26	tttacaggaccaactggatctggaaa	462	tttccagatccagttggctcctgtaaa	459
N680429_560_31_50bp	CATTATATGTTTTTACAAT CCGGGCCGCCATACTGTA AGAACATATAATG	463	cattatatgttcttacagatggcgcccgattgtaaaaacat ataatg	530

Table 11				
Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
N191607_8_101_50bp	CGTTATAGGGTATTGCAG TACCGACCGCCATACTGT AATACCTTATAACG	464	cgttatagggtattacagtatggcggtcgggtactgcaatacc tataacg	531
N674992_1_1308_50bp	TGTATCATTTTCATATAGT GTGCAGGTGCTAACTATA TGAAAATGATACA	465	tgtatcattttcatatagttagcacctgcacactatatgaaaatg ataca	532
N684613_54_96_50bp	TGTCTACTATGTCTTTATG CCACATGTGTGTCGCATATA CAGATAGTAGACA	466	tgtctactatctgtatatgcgacacatgtggcataaagacatag tagaca	533
N252616_12_1_74_50bp	AATGAGGTCAGACGCATG GAGCGCCGCCTCCGCATG CGTCAGGGTCGATG	467	catcgaccctgacgcacgcggaggcggcgctccatgcgtct gacctcatt	534
N683040_22_2_19_50bp	GTTAGTACCCAAATGATA AAAGGATGACCTTTTGTC ATTTGGGTACTAAC	468	gtagtagcccaaatgacaaaaggctacacctttatcatttgggta ctaac	535
N687537_17_3_59_50bp	GTTTATAAAACCGATGCC GCTTTGACAGAAGCGGAA CGGGTTTATAAAG	469	cttataaaacccgttccgcttctgtcaaagcgcatcggtttta taaac	536
N183629_47_40_50bp	GGCCGCGAGGTCGTGTTT GTCGTCATGTTGAGGTTC ACGACCATCACGCC	470	ggcgtgatggctgtgaacctcaacatgacgacgaacacgac ctcgcggcc	537
N191533_22_4_76_50bp	TATAAACTGATATAATTC AAAGTTATAACTTGATAT ATTCAAGATGTAGA	471	tctacatcttgaatatatcaagtataactttgaattatatcagttt ata	538
N682356_18_8_20_50bp	TATTATATCTAAAAGCAG TATGGCGGAGCTTAGTGC TTTAGATATAATT	472	aattatatctaaaagcactaagctccgccatactgcttttagata taata	539

6.9. Co-delivery of gene editor and donor DNA template

[0339] This disclosure features methods of delivering (e.g., co-delivery or dual delivery) a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, where the methods includes delivering to a (i) gene editor construct and a (ii) template polynucleotide, and (iii) at least a first attachment site-containing guide (atgRNA).

[0340] This disclosure also features a method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, where the method includes: delivering a lipid nanoparticle (LNP) comprising a gene editor polynucleotide (e.g., a gene editor polynucleotide construct) and a vector comprising a template polynucleotide and at least a first attachment site-containing guide RNA (atgRNA). In some embodiments, the first atgRNA comprises (i) a domain that is capable of guiding the prime editor system to a target

sequence; and (ii) a reverse transcriptase (RT) template that comprises at least a portion of a first integration recognition site. In some embodiments, where the vector comprises a polynucleotide encoding a first atgRNA, the RT template comprises the entirety of the first integration recognition site. In some embodiments, where the vector comprises a polynucleotide encoding a first atgRNA, the vector also includes a sequence encoding a nicking guide RNA (ngRNA).

[0341] This disclosure also features a method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, where the method includes: delivering a lipid nanoparticle (LNP) comprising a gene editor polynucleotide (e.g., a gene editor polynucleotide construct) and a vector comprising a template polynucleotide and a first attachment site-containing guide RNA (atgRNA) and a second attachment site-containing guide RNA (atgRNA). In some embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence, the first atgRNA further includes a first RT template that comprises at least a portion of the at least first integration recognition site; and the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site. In such embodiments, the first atgRNA and second atgRNA include at least a 6bp overlap (e.g., 6bp of complementarity).

[0342] This disclosure also features a method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, where the method includes: delivering into a cell a lipid nanoparticle (LNP) comprising: (i) a gene editor polynucleotide (e.g., a gene editor polynucleotide construct), and (ii) a first attachment site-containing guide RNA (atgRNA); and a vector comprising: (i) a template polynucleotide, and (ii) a second atgRNA. In some embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence, the first atgRNA further includes a first RT template that comprises at least a portion of the a first integration recognition site; and the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site. In such embodiments,

the first atgRNA and second atgRNA include at least a 6bp overlap (e.g., 6bp of complementarity).

[0343] This disclosure also features a method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, where the method includes delivering: a lipid nanoparticle (LNP) comprising: (i) a gene editor polynucleotide (e.g., a gene editor polynucleotide construct), (ii) a first attachment site-containing guide RNA (atgRNA), and (iii) a second atgRNA; and a vector comprising (i) a template polynucleotide. In some embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence, the first atgRNA further includes a first RT template that comprises at least a portion of the at least first integration recognition site; and the second atgRNA further includes a second RT template that comprises at least a portion of the at least first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site. In such embodiments, the first atgRNA and second atgRNA include at least a 6bp overlap (e.g., 6bp of complementarity).

[0344] This disclosure also features a method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, where the method includes delivering: a lipid nanoparticle (LNP) comprising: (i) a gene editor polynucleotide (e.g., a gene editor polynucleotide construct) and (ii) a first attachment site-containing guide RNA (atgRNA); and a vector comprising: (i) a template polynucleotide, and (ii) a nicking atgRNA. In some embodiments, the first atgRNA comprises (i) a domain that is capable of guiding the prime editor system to a target sequence; and (ii) a reverse transcriptase (RT) template that comprises at least a portion of a first integration recognition site. In some embodiments, where the vector comprises a polynucleotide encoding a first atgRNA, the RT template comprises the entirety of the first integration recognition site.

[0345] In some embodiments, where the method includes delivering an LNP and a first vector, the LNP and the first vector are delivered at least 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 weeks apart. In some embodiments, where the method includes delivering an LNP and a second vector, the LNP and the second vector are delivered at least 1 day, 2 days, 3 days, 4 days, 5

days, 6 days, 7 days, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 weeks apart.

[0346] This disclosure also features a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising: a lipid nanoparticle (LNP) comprising a gene editor polynucleotide (e.g., a gene editor polynucleotide construct) and a vector comprising a template polynucleotide and at least a first attachment site-containing guide RNA (atgRNA). In some embodiments, the first atgRNA comprises (i) a domain that is capable of guiding the prime editor system to a target sequence; and (ii) a reverse transcriptase (RT) template that comprises at least a portion of a first integration recognition site. In some embodiments, where the vector comprises a polynucleotide encoding a first atgRNA, the RT template comprises the entirety of the first integration recognition site. In some embodiments, where the vector comprises a polynucleotide encoding a first atgRNA, the vector also includes a sequence encoding a nicking guide RNA (ngRNA).

[0347] This disclosure also features a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising: a lipid nanoparticle (LNP) comprising a gene editor polynucleotide (e.g., a gene editor polynucleotide construct) and a vector comprising a template polynucleotide and a first attachment site-containing guide RNA (atgRNA) and a second attachment site-containing guide RNA (atgRNA). In some embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence, the first atgRNA further includes a first RT template that comprises at least a portion of the a first integration recognition site; and the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site. In such embodiments, the first atgRNA and second atgRNA include at least a 6bp overlap.

[0348] This disclosure also features a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising: a lipid nanoparticle (LNP) comprising: (i) a gene editor polynucleotide (e.g., a gene editor polynucleotide construct), and (ii) a first attachment site-containing guide RNA (atgRNA); and a vector comprising: (i) a template polynucleotide, and (ii) a second atgRNA. In some embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of

atgRNAs have domains that are capable of guiding the prime editor system to a target sequence, the first atgRNA further includes a first RT template that comprises at least a portion of the a first integration recognition site; and the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site. In such embodiments, the first atgRNA and second atgRNA include at least a 6bp overlap.

[0349] This disclosure also features a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising: co-delivering: a lipid nanoparticle (LNP) comprising: (i) a gene editor polynucleotide (e.g., a gene editor polynucleotide construct), (ii) a first attachment site-containing guide RNA (atgRNA), and (iii) a second atgRNA; and a vector comprising (i) a template polynucleotide. In some embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence, the first atgRNA further includes a first RT template that comprises at least a portion of the a first integration recognition site; and the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site. In such embodiments, the first atgRNA and second atgRNA include at least a 6bp overlap.

[0350] This disclosure also features a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising: a lipid nanoparticle (LNP) comprising: (i) a gene editor polynucleotide (e.g., a gene editor polynucleotide construct) and (ii) a first attachment site-containing guide RNA (atgRNA); and a vector comprising: (i) a template polynucleotide, and (ii) a nicking atgRNA. In some embodiments, the first atgRNA comprises (i) a domain that is capable of guiding the prime editor system to a target sequence; and (ii) a reverse transcriptase (RT) template that comprises at least a portion of a first integration recognition site. In some embodiments, where the vector comprises a polynucleotide encoding a first atgRNA, the RT template comprises the entirety of the first integration recognition site.

[0351] In typical embodiments, the LNP comprising a gene editor polynucleotide construct is capable delivering to a cell cytoplasm the gene editor polynucleotide construct. In some embodiments, the LNP comprising a gene editor polynucleotide construct is capable delivering

to a cell nucleus the gene editor polynucleotide construct. In some embodiments, the LNP comprises a gene editor protein and associated guide nucleic acids. In some embodiments, the LNP comprises a gene editor protein and associated guide nucleic acids that are capable of localizing to cell nucleus.

[0352] In some embodiments, a gene editor polynucleotide construct is delivered to a cell by a fusosome. In some embodiments, a gene editor polynucleotide construct is delivered to a cell cytoplasm by a fusosome. In some embodiments, the fusosome comprises a gene editor protein and associated guide nucleic acids.

[0353] In some embodiments, a gene editor polynucleotide construct is delivered to a cell by an exosome. In some embodiments, a gene editor polynucleotide construct is delivered to a cell cytoplasm by an exosome. In some embodiments, the exosome comprises a gene editor protein and associated guide nucleic acids.

[0354] In some embodiments, the prime editor or Gene Writer protein fusion, either of which may have a fused/linked integrase, is incorporated (i.e., packaged) into LNP as protein. Further, associated atgRNA and optional ngRNAs may be co-packaged with gene editor proteins in LNP.

[0355] In some embodiments, the gene editor polynucleotide construct comprises (a) a polynucleotide sequence encoding a prime editor fusion protein or a Gene WriterTM protein, (b) a polynucleotide sequence encoding an attachment site-containing guide RNA (atgRNA), (c) optionally, a polynucleotide sequence encoding a nickase guide RNA (ngRNA), (d) a polynucleotide sequence encoding an integrase, (e) and optionally, a polynucleotide sequence encoding a recombinase.

[0356] In some embodiments, the prime editor or Gene Writer protein fusion, either of which may have a fused/linked integrase, is expressed as a split construct. In typical embodiments, the split construct is reconstituted in a cell. In some embodiments, the split construct can be fused or ligated via intein protein splicing. In some embodiments, the split construct can be reconstituted via protein-protein inter-molecular bonding and/or interactions. In some embodiments, the split construct can be reconstituted via chemical, biological, or environmental induced oligomerization. In certain embodiments, the split construct can be adapted into one or more nucleic acid constructs described herein.

6.9.1. Gene Editor Polynucleotide

[0357] In some embodiments, the systems described include a gene editor polynucleotide that is delivered to a cell using the methods described herein. In some embodiments, the gene editor polynucleotide is delivered as a polynucleotide (e.g., an mRNA). In some embodiments, the gene editor polynucleotide is delivered as a protein. In some embodiments, the gene editor polynucleotide or protein is packaged, and thereby vectorized, within a lipid nanoparticle (LNP). In some embodiments, the gene editor polynucleotide or protein is packaged in a LNP and is co-delivered with a template polynucleotide (i.e., nucleic acid “cargo” or nucleic acid “payload”) packaged into a separate vector (e.g., a viral vector (e.g., an AAV or adenovirus)) or a second lipid nanoparticle (LNP).

[0358] In some embodiments, the gene editor polynucleotide is delivered to the cells as a polynucleotide. For example, the gene editor polynucleotide is delivered to the cells as an mRNA encoding the gene editor polynucleotide (e.g., the gene editor protein or the prime editor system). In some embodiments, the mRNA comprises one or more modified uridines. In some embodiments, the mRNA comprises a sequence where each of the uridines is a modified uridine. In some embodiments, the mRNA is uridine depleted. In some embodiments, the mRNA encoding the nickase comprises one or more modified uridines. In some embodiments, the mRNA encoding the reverse transcriptase comprises one or more modified uridines. In some embodiments, the mRNA encoding the nickase comprises one or more modified uridines, and the mRNA encoding the reverse transcriptase comprises one or more modified uridines. In some embodiments, where the integrase is encoded in an mRNA, the mRNA comprises modified uridines. In some embodiments, a modified uridine is a N1-Methylpseudouridine-5'-Triphosphate. In some embodiments, a modified uridine is a pseudouridine. In some embodiments, the mRNA comprises a 5' cap. In some embodiments, the 5' cap comprises a molecular formula of $C_{32}H_{43}N_{15}O_{24}P_4$ (free acid).

[0359] In some embodiments, the gene editor polynucleotide (e.g., a gene editor polynucleotide construct) comprises a polynucleotide sequence encoding a primer editor system (e.g., any of the prime editor systems described herein). In some embodiments, the prime editor system comprises a nucleotide sequence encoding a nickase (e.g., any of the Cas proteins or variants thereof (e.g., nickases) and nickases described herein, see Tables 4-8) and a nucleotide sequence encoding a reverse transcriptase (e.g., any of the reverse transcriptases described herein). In some embodiments, the nucleotide sequence encoding the nickase and the

nucleotide sequence encoding the reverse transcriptase are positioned in the construct such that when expressed the nickase is linked to the reverse transcriptase. In some embodiments, the nickase is linked to the reverse transcriptase by in-frame fusion. In some embodiments, the nickase is linked to the reverse transcriptase by a linker. In some embodiments, the linker is a peptide fused in-frame between the nickase and reverse transcriptase.

[0360] In some embodiments, the gene editor polynucleotide (e.g., a gene editor polynucleotide construct) further comprises a polynucleotide sequence encoding at least a first integrase (e.g., any of the integrases described herein, e.g., as described in Table 10 and also in Yarnall et al., *Nat. Biotechnol.*, 2022, doi.org/10.1038/s41587-022-01527-4 and Durrant et al., *Nat. Biotechnol.*, 2022, doi.org/10.1038/s41587-022-01494-w, each of which are herein incorporated by reference in their entireties). In some embodiments, the linked nickase-reverse transcriptase are further linked to the first integrase.

[0361] In some embodiments, the gene editor polynucleotide construct further comprises a polynucleotide sequence encoding at least a first recombinase (e.g., any of the recombinases described herein).

6.9.2. Vector

[0362] In some embodiments, the systems and methods described herein include a vector that is capable of co-delivering a template polynucleotide, one or more attachment site-containing gRNA, one or more integrases, one or more recombinases, a gene editor polynucleotide, one or more integration recognition sites, one or more recombinase recognition sites, or a combination thereof.

[0363] Non-limiting examples of vectors that can be used in the methods or systems described herein include the vectors described in FIGs. 3-6.

6.9.2.1 AtgRNA and/or ngRNA

[0364] In some embodiments, the vector includes a polynucleotide sequence encoding an attachment site-containing guide RNA (atgRNA). In such embodiments, the polynucleotide sequence encoding the attachment site-containing guide RNA (atgRNA) is operably linked to a regulatory element (e.g., a U6 promoter) that is capable of driving expression of the atgRNA. In such embodiments, the atgRNA comprises (i) a domain that is capable of guiding the prime editor system to a target sequence; and (ii) a reverse transcriptase (RT) template that comprises

at least a portion of a first integration recognition site. In some embodiments, where the system, and thereby the vector, include a polynucleotide encoding only a first atgRNA, the RT template comprises the entirety of the first integration recognition site. In such embodiments, the vector or the LNP includes a polynucleotide sequence encoding a nicking gRNA.

[0365] In some embodiments, the vector includes a polynucleotide sequence encoding a first attachment site-containing guide RNA (atgRNA) and a polynucleotide sequence encoding a second attachment site-containing guide RNA (atgRNA). In such embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence, the first atgRNA further includes a first RT template that comprises at least a portion of the a first integration recognition site; the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site. In such embodiments, the first atgRNA and second atgRNA include at least a 6bp overlap.

6.9.2.2 Template Polynucleotide

[0366] In typical embodiments, the vector includes a template polynucleotide and a sequence that is an integration cognate of an integration recognition site site-specifically incorporated into the genome of a cell. For example, the vector includes a template polynucleotide and a second integration recognition site that is a cognate pair with the first integration recognition site site-specifically incorporated into the genome of the cell. In such embodiments, the sequence that is an integration cognate (e.g., a second integration recognition site) enables integration of the template polynucleotide or portion thereof when contacted with an integrase and the site-specifically incorporated first integration recognition site.

[0367] In typical embodiments, the vector comprising a template polynucleotide is a recombinant adenovirus, a helper dependent adenovirus, an AAV, a lentivirus, an HSV, an anneovirus, a retrovirus, a DoggyboneTM DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid. In preferred embodiments, the vector is capable of localizing to the nucleus.

[0368] In certain embodiments, the template polynucleotide is delivered to the cytoplasm and localizes to the nucleus. In certain embodiments, the template polynucleotide is delivered to

the cytoplasm by LNP. In certain embodiments, the donor template polynucleotide construct comprises a recognition sequence that is recognized by a DNA binding protein (DNA binding domain) or a transcription factor binding domain. In certain embodiments, the donor template polynucleotide construct is delivered to the nucleus by an integrase or recombinase.

[0369] In certain embodiments, the template polynucleotide is delivered to the mitochondria. In certain embodiments, the donor template polynucleotide construct comprises a mitochondria targeting sequence.

[0370] In certain embodiments, the vector comprising a template polynucleotide is AAV. In some embodiments, the AAV contains a 5' inverted terminal repeat (ITR). In some embodiments, the AAV contains a 3' inverted terminal repeat (ITR). In some embodiments, the AAV contains a 5' and a 3' ITR. In some embodiments, the 5' and 3' ITR are not derived from the same serotype of virus. In some embodiments, the ITRs are derived from adenovirus, AAV2, and/or AAV5.

[0371] In certain embodiments, the vector comprising a template polynucleotide is single stranded AAV (ssAAV). In certain embodiments, the vector comprising a donor template polynucleotide construct is self-complementary AAV (scAAV).

[0372] In some embodiments, a vector comprises an attachment site-containing guideRNA (atgRNA), a nicking-guideRNA (ngRNA), and template polynucleotide. In typical embodiments, the vector comprising an attachment site-containing guideRNA (atgRNA), a nicking-guideRNA (ngRNA), and template polynucleotide is recombinant adenovirus, helper dependent adenovirus, AAV, lentivirus, HSV, anellovirus, retrovirus, Doggybone™ DNA (dbDNA), minicircle, plasmid, miniDNA, exosome, fusosome, or nanoplasmid. In preferred embodiments, the vector is capable of localizing to the nucleus. In typical embodiments, the attachment site-containing guideRNA (atgRNA) sequence and the nicking-guideRNA (ngRNA) sequence contain a terminal poly dT.

[0373] In some embodiments, a vector comprises an attachment site-containing guideRNA (atgRNA), and donor template. In typical embodiments, the vector comprising an attachment site-containing guideRNA (atgRNA) and donor template is recombinant adenovirus, helper dependent adenovirus, AAV, lentivirus, HSV, anellovirus, retrovirus, Doggybone™ DNA (dbDNA), minicircle, plasmid, miniDNA, exosome, fusosome, or nanoplasmid. In preferred

embodiments, the vector is capable of localizing to the nucleus. In typical embodiments, the attachment site-containing guideRNA (atgRNA) sequence contain a terminal poly dT.

[0374] In typical embodiments, the template polynucleotide is capable of being integrated into a genomic locus that contains an integrase target recognition site or a recombinase target recognition site.

[0375] In certain embodiments, the template polynucleotide comprises at least one of the following: a gene, a gene fragment, an expression cassette, a logic gate system, or any combination thereof. In some embodiments, the template polynucleotide comprises at least one intron or exon.

[0376] In typical embodiments, the template polynucleotide further comprises at least one integrase target recognition site or a recombinase target integrase site. In certain embodiments, at least one integrase target recognition site or a recombinase target integrase site is placed within the donor template vector inverted terminal repeat.

6.9.2.3 Integrase- or recombinase-mediated self-circularization of a subsequence of a vector delivered as part of the co-delivery system

[0377] In some embodiments, the delivery system (e.g., co-delivery system) includes a vector having a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid. In some embodiments, the vector comprises a physical portion or region of the vector that is capable of self-circularizing to form a circular construct. As used herein, the term “sub-sequence” refers to a portion of the vector that is capable of self-circularizing, where the sub-sequence is flanked by integration recognition sites or recombinase recognition sites positioned to enable self-circularization. As used herein, the term “self-circular nucleic acid” refers to a double-stranded, circular nucleic acid construct produced as a result of recombination of a cognate pair of integrase or recombinase recognition sites present on the vector. Recombination occurs when the vector is contacted with an integrase or a recombinase under conditions that allow for recombination of the cognate pair of integrase or recombinase recognition sites.

[0378] In some embodiments, the sub-sequence of the vector includes a first recombinase recognition site and a second recombinase recognition site, wherein the first and second recombinase recognition sites are capable of being recombined by a recombinase. In some embodiments, the sub-sequence of the vector includes a first recombinase recognition site, a

second recombinase recognition site, and a second integration recognition site (e.g., the second integration recognition site is a cognate pair of the first integration recognition site), where the first and second recombinase recognition sites flank the integration recognition site. In such cases, the first recombinase recognition site, the second recombinase recognition, and a recombinase enable the self-circularizing and formation of the circular construct.

[0379] In some embodiments, the sub-sequence of the vector includes a third integration recognition site and a fourth integration recognition site, wherein the third and fourth integration recognition sites are a cognate pair. In some embodiments, the subsequence of the vector includes the second integration recognition site, the third integration recognition site, the fourth integration recognition site, where the third and fourth integration recognition sites flank the second integration recognition site (where the second integration recognition site is a cognate pair of the first integration recognition site). In such cases, the third integration recognition site, the fourth integration recognition site, and an integrase enable self - circularization and formation of the circular construct. In such cases, the third integration recognition site and/or the fourth integration recognition sites cannot recombine with the first integration recognition site and/or the second integration recognition site due, in part, to having different central dinucleotides than the first and second integration recognition sites.

[0380] In some embodiments where the subsequence includes three or more integration recognition sites, each integration recognition site or each pair of integration recognition is capable of being recognized by a different integrase. In some embodiments where the subsequence includes three or more integration recognition sites, each integration recognition site or each pair of integration recognition comprises a different central dinucleotide.

[0381] In some embodiments, self-circularizing is mediated at the integration recognition sites or recombinase recognition sites. In some embodiments, the self-circularizing is mediated by an integrase or a recombinase.

[0382] In some embodiments, upon introducing the vector into a cell and after self-circularizing to form the self-circular nucleic acid, the self-circular nucleic acid comprising the second integration recognition site is capable of being integrated into the cell's genome at the target sequence that contains the first integration recognition site.

[0383] In some embodiments, following self-circularization, the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of an

additional nucleic acid cargo. In such cases, the additional nucleic acid cargo includes a sequence that is a cognate pair with one or more of the additional integration recognition sites in the self-circular nucleic acid. For example, integration of the self-circular nucleic acid into the genome of a cell results in integration of the one or more additional integration recognition sites into the genome along with the nucleic acid cargo. The integrated one or more additional integration recognition sites serve as an integration recognition site (beacon) for placing the additional nucleic acid cargo. Upon contacting the cell harboring the integrated nucleic acid cargo and the one or more additional integration recognition sites with an integrase and the second additional nucleic acid cargo that includes a sequence that is an integration cognate to the one or more additional integration recognition sites the additional nucleic acid cargo is integrated into the cell's genome.

[0384] In typical embodiments, the self-circularized nucleic acid comprises a DNA cargo. In some embodiments, the DNA cargo is a gene or gene fragment. In some embodiments the DNA cargo is an expression cassette. In some embodiments, the DNA cargo is a logic gate or logic gate system. The logic gate or logic gate system may be DNA based, RNA based, protein based, or a mix of DNA, RNA, and protein. In some embodiments, the nucleic acid cargo is a genetic, protein, or peptide tag and/or barcode.

6.9.2.4 A second vector

[0385] In some embodiments, the system or methods described herein include a second vector. In some embodiments, where the gene editor polynucleotide encodes a prime editor system comprising a nickase (e.g., any of the Cas proteins or variants thereof (e.g., nickases) and nickases described herein, see Tables 4-8) and a reverse transcriptase (e.g., any of the reverse transcriptase described herein), the second vector comprises a polynucleotide sequence encoding an integrase (e.g., any of the integrases described herein, e.g., as described in Table 10 and also in Yarnall et al., *Nat. Biotechnol.*, 2022, doi.org/10.1038/s41587-022-01527-4 and Durrant et al., *Nat. Biotechnol.*, 2022, doi.org/10.1038/s41587-022-01494-w, each of which are herein incorporated by reference in their entireties).

[0386] In some embodiments, where the gene editor polynucleotide encodes a prime editor system comprising a nickase and a reverse transcriptase, the second vector comprises a polynucleotide sequence encoding at least a first recombinase. In some embodiments, where the gene editor polynucleotide encodes a prime editor system comprising a nickase, a reverse transcriptase, and an integrase, the second vector comprises a polynucleotide sequence

encoding at least a first recombinase. In some embodiments, where the gene editor polynucleotide encodes a prime editor system comprising a nickase, a reverse transcriptase, and an integrase, the second vector comprises a polynucleotide sequence encoding at least a second integrase.

[0387] In some embodiments, the second vector includes a template polynucleotide and a sequence that is an integration cognate of an integration recognition site site-specifically incorporated into the genome of a cell. For example, the second vector includes a template polynucleotide and a second integration recognition site that is a cognate pair with the first integration recognition site site-specifically incorporated into the genome of the cell. In such embodiments, the sequence that is an integration cognate (e.g., a second integration recognition site) enables integration of the template polynucleotide or portion thereof when contacted with an integrase and the site-specifically incorporated first integration recognition site.

[0388] In some embodiments, the second vector is a vector selected from: adenovirus, AAV, lentivirus, HSV, anellovirus, retrovirus, Doggybone™ DNA (dbDNA), minicircle, plasmid, miniDNA, exosome, fusosome, or nanoplasmid.

[0389] In some embodiments, the polynucleotide sequence encoding the prime editor system is encoded on at least two different vectors. In one embodiment, a first vector comprises a polynucleotide sequence encoding a nickase and a second vector comprises a polynucleotide sequence encoding a reverse transcriptase. In such cases, the first vector and second are delivered concurrently.

[0390] In some embodiments, the polynucleotide sequence(s) encoding the prime editor system is encoded on at least two (non-contiguous) polynucleotide sequences. In one embodiment, a first polynucleotide sequence encodes a nickase and a second polynucleotide sequence encodes a reverse transcriptase. In such cases, the first vector and second are delivered concurrently (e.g., in a first LNP).

6.9.3. Split Lipid Nanoparticles (LNPs)

[0391] Also provided herein are methods of co-delivering a system capable of site-specifically integrating at least a first integration recognition site into the genome of a cell, where the method includes delivering to a cell a mixture of a first LNP and a second LNP (“split LNPs”). In one embodiment, the method includes co-delivering to a cell a first gene editor

polynucleotide construct and a first attachment site-containing guide RNA (atgRNA) are packaged, and thereby vectorized, within the first LNP, and a second gene editor polynucleotide construct and a second attachment site containing guide RNR (atgRNA) are packaged, and thereby vectorized, within the second LNP, where the first atgRNA and the second atgRNA are an at least first pair of atgRNA. The at least first pair of atgRNAs comprise domains that are capable of guiding the prime editor system to a target sequence. The first atgRNA further includes a first RT template that comprises at least a portion of a first integration recognition site. The second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site. The first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site. In such embodiments, the first atgRNA and second atgRNA include at least a 6bp overlap.

[0392] In some embodiments, where the method includes delivering a first LNP (e.g., a first LNP comprising a first gene editor polynucleotide construct and a first atgRNA) and a second LNP (e.g., a second LNP comprising a second gene editor polynucleotide construct and a second atgRNA), the first LNP and the second LNP are mixed prior to delivering to a cell. In some embodiments, the first LNP and the second LNP are mixed at a ratio of first LNP to second LNP of 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 1:1, 1:0.75, 0.75:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1. In some embodiments, the first LNP and the second LNP are mixed at a ratio of 1:1.

[0393] In some embodiments, a first LNP comprising a first gene editor polynucleotide construct and a first attachment site-containing guide RNA (atgRNA1) comprises a ratio of ratio of gene editor polynucleotide construct (e.g., mRNA) to atgRNA1 of 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 1:1, 1:0.75, 0.75:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1. In some embodiments, the first LNP comprises a ratio of mRNA to atgRNA1 of 2:1.

[0394] In some embodiments, a second LNP comprising a second gene editor polynucleotide construct and a second attachment site-containing guide RNA (atgRNA2) comprises a ratio of gene editor polynucleotide construct (e.g., mRNA) to atgRNA2 of 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 1:1, 1:0.75, 0.75:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1. In some embodiments, the second LNP comprises a ratio of mRNA to atgRNA2 of 2:1.

[0395] In some embodiments, where the method includes delivering a first LNP (e.g., a first LNP comprising a first gene editor polynucleotide construct and a first atgRNA) and a second

LNP (e.g., a second LNP comprising a second gene editor polynucleotide construct and a second atgRNA), the first LNP and the second LNP are mixed such that the ratio of gene editor polynucleotide construct (e.g., mRNA) to first atgRNA (atgRNA1) to second atgRNA (atgRNA2) is 1:0.25:0.25, 1:0.5:0.5, 1:0.75:0.75, or 1:1:1.

[0396] In some embodiments, the method of co-delivering to a cell a mixture of LNPs includes co-delivering three or more LNPs, four or more LNPs, five or more LNPs, six or more LNPs, seven or more LNPs, eight or more LNPs, nine or more LNPs, or ten or more LNPs.

[0397] Also provided herein is a system capable of site-specifically integrating at least a first integration recognition site into the genome of a cell, the system comprising: a first gene editor polynucleotide construct and a first attachment site-containing guide RNA (atgRNA) are packaged, and thereby vectorized, within the first LNP, and a second gene editor polynucleotide construct and a second attachment site containing guide RNR (atgRNA) are packaged, and thereby vectorized, within the second LNP, where the first atgRNA and the second atgRNA are an at least first pair of atgRNA. The at least first pair of atgRNAs comprise domains that are capable of guiding the prime editor system to a target sequence. The first atgRNA further includes a first RT template that comprises at least a portion of a first integration recognition site. The second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site. The first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site. In such embodiments, the first atgRNA and second atgRNA include at least a 6bp overlap.

[0398] In some embodiments, the system comprises a first LNP (e.g., any of the first LNPs described herein) and a second LNP (e.g., any of the second LNPs described herein) at a ratio of first LNP to second LNP of 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 1:1, 1:0.75, 0.75:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1. In some embodiments, the system comprise the first LNP and the second LNP at a ratio of 1:1.

[0399] In some embodiments, the system comprises a first LNP having a ratio of a first gene editor polynucleotide construct to a first attachment site-containing guide RNA (atgRNA1) of 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 1:1, 1:0.75, 0.75:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1. In some embodiments, the system includes a first LNP having a ratio of mRNA (i.e., mRNA encoding the gene editor protein) to atgRNA1 of 2:1.

[0400] In some embodiments, the system comprise a second LNP having a ratio of a second gene editor polynucleotide construct to a second attachment site-containing guide RNA (atgRNA2) of 1:10, 1:9, 1:8, 1:7, 1:6, 1:5, 1:4, 1:3, 1:2, 1:1, 1:0.75, 0.75:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1. In some embodiments, the system includes a second LNP having a ratio of mRNA (i.e., mRNA encoding the gene editor protein) to atgRNA2 of 2:1.

[0401] In some embodiments, the system comprises a ratio of gene editor polynucleotide construct (e.g., mRNA encoding the gene editor protein) to first atgRNA (atgRNA1) to second atgRNA (atgRNA2) of 1:0.25:0.25, 1:0.5:0.5, 1:0.75:0.75, or 1:1:1.

[0402] In some embodiments, the system comprises a mixture of LNPs comprising three or more LNPs, four or more LNPs, five or more LNPs, six or more LNPs, seven or more LNPs, eight or more LNPs, nine or more LNPs, or ten or more LNPs.

[0403] In some embodiments, where a split LNP (e.g., a mixture of two LNPs packaged with different cargo) is being used to site-specifically integrate the at least first integration recognition site into the genome, a vector comprising a template polynucleotide and a sequence that is an integration cognate (i.e., cognate to an integration recognition site site-specifically incorporated into the genome of a cell) can be delivered to the cell concurrently with the split LNPs or after delivery of the split LNPs. For example, after delivering the split LNPs to the cell, a vector that includes a template polynucleotide and a second integration recognition site that is a cognate pair with the first integration recognition site is delivered to the cell. In such embodiments, the sequence that is an integration cognate (e.g., a second integration recognition site) enables integration of the template polynucleotide or portion thereof when contacted with an integrase and the site-specifically incorporated first integration recognition site.

6.9.4. Vector Delivery of a template polynucleotide

[0404] In certain aspects the invention involves vectors, e.g. for delivering or introducing in a cell, but also for propagating these components (e.g. in prokaryotic cells). A used herein, a "vector" is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-

stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Vectors for and that result in expression in a eukaryotic cell can be referred to herein as "eukaryotic expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0405] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). With regards to recombination and cloning methods, mention is made of U.S. patent application Ser. No. 10/815,730, published Sep. 2, 2004 as US 2004-0171156 A1, the contents of which are herein incorporated by reference in their entirety.

[0406] Vector delivery, e.g., plasmid, viral delivery: The CRISPR enzyme, for instance a Type V protein such as C2c1 or C2c3, and/or any of the present RNAs, for instance a guide RNA, can be delivered using any suitable vector, e.g., plasmid or viral vectors, such as adeno

associated virus (AAV), lentivirus, adenovirus or other viral vector types, or combinations thereof. Effector proteins and one or more guide RNAs can be packaged into one or more vectors, e.g., plasmid or viral vectors. In some embodiments, the vector, e.g., plasmid or viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector choice, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

[0407] Such a dosage may further contain, for example, a carrier (water, saline, ethanol, glycerol, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, etc.), a diluent, a pharmaceutically-acceptable carrier (e.g., phosphate-buffered saline), a pharmaceutically-acceptable excipient, and/or other compounds known in the art. The dosage may further contain one or more pharmaceutically acceptable salts such as, for example, a mineral acid salt such as a hydrochloride, a hydrobromide, a phosphate, a sulfate, etc.; and the salts of organic acids such as acetates, propionates, malonates, benzoates, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, gels or gelling materials, flavorings, colorants, microspheres, polymers, suspension agents, etc. may also be present herein. In addition, one or more other conventional pharmaceutical ingredients, such as preservatives, humectants, suspending agents, surfactants, antioxidants, anticaking agents, fillers, chelating agents, coating agents, chemical stabilizers, etc. may also be present, especially if the dosage form is a reconstitutable form. Suitable exemplary ingredients include microcrystalline cellulose, carboxymethylcellulose sodium, polysorbate 80, phenylethyl alcohol, chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallate, the parabens, ethyl vanillin, glycerin, phenol, parachlorophenol, gelatin, albumin and a combination thereof. A thorough discussion of pharmaceutically acceptable excipients is available in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991) which is incorporated by reference herein.

[0408] In an embodiment herein the delivery is via an adenovirus, which may be at a single booster dose containing at least 1×10^5 particles (also referred to as particle units, pu) of

adenoviral vector. In an embodiment herein, the dose preferably is at least about 1×10^6 particles (for example, about 1×10^6 - 1×10^{11} particles), more preferably at least about 1×10^7 particles, more preferably at least about 1×10^8 particles (e.g., about 1×10^8 - 1×10^{11} particles or about 1×10^9 - 1×10^{12} particles), and most preferably at least about 1×10^{10} particles (e.g., about 1×10^9 - 1×10^{10} particles or about 1×10^9 - 1×10^{12} particles), or even at least about 1×10^{10} particles (e.g., about 1×10^{10} - 1×10^{12} particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1×10^{14} particles, preferably no more than about 1×10^{13} particles, even more preferably no more than about 1×10^{12} particles, even more preferably no more than about 1×10^{11} particles, and most preferably no more than about 1×10^{10} particles (e.g., no more than about 1×10^9 particles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1×10^6 particle units (pu), about 2×10^6 pu, about 4×10^6 pu, about 1×10^7 pu, about 2×10^7 pu, about 4×10^7 pu, about 1×10^8 pu, about 2×10^8 pu, about 4×10^8 pu, about 1×10^9 pu, about 2×10^9 pu, about 4×10^9 pu, about 1×10^{10} pu, about 2×10^{10} pu, about 4×10^{10} pu, about 1×10^{11} pu, about 2×10^{11} pu, about 4×10^{11} pu, about 1×10^{12} pu, about 2×10^{12} pu, or about 4×10^{12} pu of adenoviral vector. See, for example, the adenoviral vectors in U.S. Pat. No. 8,454,972 B2 to Nabel, et. al., granted on Jun. 4, 2013; incorporated by reference herein, and the dosages at col 29, lines 36-58 thereof. In an embodiment herein, the adenovirus is delivered via multiple doses.

[0409] In an embodiment herein, the delivery is via an AAV. A therapeutically effective dosage for in vivo delivery of the AAV to a human is believed to be in the range of from about 20 to about 50 ml of saline solution containing from about 1×10^{10} to about 1×10^{50} functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In an embodiment herein, the AAV dose is generally in the range of concentrations of from about 1×10^5 to 1×10^{50} genomes AAV (sometimes referred to herein as “vector genomes” or “vg”), from about 1×10^8 to 1×10^{20} genomes AAV, from about 1×10^{10} to about 1×10^{16} genomes, or about 1×10^{11} to about 1×10^{16} genomes AAV. A human dosage may be about 1×10^{13} genomes AAV. Such concentrations may be delivered in from about 0.001 ml to about 100 ml, about 0.05 to about 50 ml, or about 10 to about 25 ml of a carrier solution. Other effective dosages can be readily established by one of ordinary skill in the art through routine trials establishing dose response curves. See, for example, U.S. Pat. No. 8,404,658 B2 to Hajjar, et al., granted on Mar. 26, 2013, at col. 27, lines 45-60.

[0410] The promoter used to drive nucleic acid-targeting effector protein coding nucleic acid molecule expression can include: AAV ITR can serve as a promoter: this is advantageous for eliminating the need for an additional promoter element (which can take up space in the vector). The additional space freed up can be used to drive the expression of additional elements (gRNA, etc.). Also, ITR activity is relatively weaker, so can be used to reduce potential toxicity due to over expression of nucleic acid-targeting effector protein. For ubiquitous expression, can use promoters: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc. For brain or other CNS expression, can use promoters: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc. For liver expression, can use Albumin promoter. For lung expression, can use SP-B. For endothelial cells, can use ICAM. For hematopoietic cells can use IFNbeta or CD45. For Osteoblasts can use OG-2.

[0411] The promoter used to drive guide RNA can include: Pol III promoters such as U6 or H1 Use of Pol II promoter and intronic cassettes to express guide RNA Adeno Associated Virus (AAV).

[0412] Nucleic acid-targeting effector protein and one or more guide RNA can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Pat. No. 8,454,972 (formulations, doses for adenovirus), U.S. Pat. No. 8,404,658 (formulations, doses for AAV) and U.S. Pat. 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 examples, for AAV, the route of administration, formulation and dose can be as in U.S. Pat. No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in U.S. Pat. No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in U.S. Pat. No. 5,846,946 and as in clinical studies involving plasmids. Doses may 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. The viral vectors can be injected into the tissue of interest. For cell-type specific

genome modification, the expression of nucleic acid-targeting effector can be driven by a cell-type specific promoter. For example, liver-specific expression might use the Albumin promoter and neuron-specific expression (e.g., for targeting CNS disorders) might use the Synapsin I promoter.

[0413] In terms of *in vivo* delivery, AAV is advantageous over other viral vectors for a couple of reasons: Low toxicity (this may be due to the purification method not requiring ultra centrifugation of cell particles that can activate the immune response) and Low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

[0414] AAV has a packaging limit of 4.5 or 4.75 Kb. This means that nucleic acid-targeting effector protein (such as a Type V protein such as C2c1 or C2c3) as well as a promoter and transcription terminator have to be all fit into the same viral vector. Therefore embodiments of the invention include utilizing homologs of nucleic acid-targeting effector protein (such as a Type V protein such as C2c1 or C2c3) that are shorter.

[0415] As to AAV, the AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. The herein promoters and vectors are preferred individually.

[0416] Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and psi2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line may also be infected with adenovirus as a helper. The helper virus promotes replication

of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

[0417] Millington-Ward et al. (Molecular Therapy, vol. 19 no. 4, 642-649 April 2011) describes adeno-associated virus (AAV) vectors to deliver an RNA interference (RNAi)-based rhodopsin suppressor and a codon-modified rhodopsin replacement gene resistant to suppression due to nucleotide alterations at degenerate positions over the RNAi target site. An injection of either 6.0×10^8 vp or 1.8×10^{10} vp AAV were subretinally injected into the eyes by Millington-Ward et al. The AAV vectors of Millington-Ward et al. may be applied to the system of the present invention, contemplating a dose of about 2×10^{11} to about 6×10^{11} vp administered to a human.

[0418] Dalkara et al. (Sci Transl Med 5, 189ra76 (2013)) also relates to in vivo directed evolution to fashion an AAV vector that delivers wild-type versions of defective genes throughout the retina after noninjurious injection into the eyes' vitreous humor. Dalkara describes a 7 mer peptide display library and an AAV library constructed by DNA shuffling of cap genes from AAV1, 2, 4, 5, 6, 8, and 9. The rcAAV libraries and rAAV vectors expressing GFP under a CAG or Rho promoter were packaged and deoxyribonuclease-resistant genomic titers were obtained through quantitative PCR. The libraries were pooled, and two rounds of evolution were performed, each consisting of initial library diversification followed by three in vivo selection steps. In each such step, P30 rho-GFP mice were intravitreally injected with 2 ml of iodixanol-purified, phosphate-buffered saline (PBS)-dialyzed library with a genomic titer of about 1×10^{12} vg/ml. The AAV vectors of Dalkara et al. may be applied to the nucleic acid-targeting system of the present invention, contemplating a dose of about 1×10^{15} to about 1×10^{16} vg/ml administered to a human.

[0419] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are

sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SW), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al., *J. Virol.* 66:1635-1640 (1992); Sommnerfelt et al., *Virol.* 176:58-59 (1990); Wilson et al., *J. Virol.* 63:2374-2378 (1989); Miller et al., *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700). In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. 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).

[0420] Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and yr2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line may also be infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid

is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

[0421] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. Cells taken from a subject include, but are not limited to, hepatocytes or cells isolated from muscle, the CNS, eye or lung. Immunological cells are also contemplated, such as but not limited to T cells, HSCs, B-cells and NK cells.

[0422] Another useful method to deliver proteins, enzymes, and guides comprises transfection of messenger RNA (mRNA). Examples of mRNA delivery methods and compositions that may be utilized in the present disclosure including, for example, PCT/US2014/028330, US8822663B2, NZ700688A, ES2740248T3, EP2755693A4, EP2755986A4, WO2014152940A1, EP3450553B1, BR112016030852A2, and EP3362461A1. Expression of CRISPR systems in particular is described by WO2020014577. Each of these publications are incorporated herein by reference in their entireties. Additional disclosure hereby incorporated by reference can be found in Kowalski et al., “Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery,” Mol Therap., 2019; 27(4): 710-728.

[0423] In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEK_n, HEK_a, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, CIR, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr^{-/-}, COR-L23,

COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT cell lines, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassas, Va.)). In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences.

[0424] In some embodiments, one or more vectors described herein are used to produce a non-human transgenic animal or transgenic plant. In some embodiments, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. In certain embodiments, the organism or subject is a plant. In certain embodiments, the organism or subject or plant is algae. Methods for producing transgenic plants and animals are known in the art, and generally begin with a method of cell transfection, such as described herein.

[0425] In one aspect, the invention provides for methods of modifying a target polynucleotide in a prokaryotic or eukaryotic cell, which may be in vivo, ex vivo or in vitro. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal or plant (including micro-algae) and modifying the cell or cells. Culturing may occur at any stage *ex vivo*. The cell or cells may even be re-introduced into the non-human animal or plant (including micro-algae).

[0426] In plants, pathogens are often host-specific. For example, *Fusarium oxysporum f. sp. lycopersici* causes tomato wilt but attacks only tomato, and *F. oxysporum f. dianthii* Puccinia graminis f. sp. tritici attacks only wheat. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible. There can also be Horizontal Resistance, e.g., partial resistance against all

racess of a pathogen, typically controlled by many genes and Vertical Resistance, e.g., complete resistance to some races of a pathogen but not to other races, typically controlled by a few genes. In a Gene-for-Gene level, plants and pathogens evolve together, and the genetic changes in one balance changes in other. Accordingly, using Natural Variability, breeders combine most useful genes for Yield. Quality, Uniformity, Hardiness, Resistance. The sources of resistance genes include native or foreign Varieties, Heirloom Varieties, Wild Plant Relatives, and Induced Mutations, e.g., treating plant material with mutagenic agents. Using the present invention, plant breeders are provided with a new tool to induce mutations. Accordingly, one skilled in the art can analyze the genome of sources of resistance genes, and in Varieties having desired characteristics or traits employ the present invention to induce the rise of resistance genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

[0427] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A “disease-associated” gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

6.9.5. Lipid Nanoparticle Delivery

[0428] In some embodiments, the delivery system is packaged in one or more LNPs and administered intravenously. In some embodiments, the co-delivery system is packaged in one or more LNPs and administered intrathecally. In some embodiments, the co-delivery system is packaged in one or more LNPs and administered by intracerebral ventricular injection. In some embodiments, the co-delivery system is packaged in one or more LNPs and administered by

intracisternal magna administration. In some embodiments, the co-delivery system is packaged in one or more LNPs and administered by intravitreal injection.

[0429] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, *Science* 270:404-410 (1995); Blaese et al., *Cancer Gene Ther.* 2:291-297 (1995); Behr et al., *Bioconjugate Chem.* 5:382-389 (1994); Remy et al., *Bioconjugate Chem.* 5:647-654 (1994); Gao et al., *Gene Therapy* 2:710-722 (1995); Ahmad et al., *Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787). In some embodiments, the LNP formulations are selected from LP01 (Cas No. 1799316-64-5), ALC-0315 (Cas No. 2036272-55-4), and cKK-E12 (Cas No. 1432494-65-9). In some embodiments, the LNP formulation is LP01 (i.e., LNP #F1). In some embodiments, the LNP formulation is ALC-0315 (i.e., LNP #F2). In some embodiment, the LNP formulation is cKK-E12 (i.e., LNP #F3).

[0430] In some embodiments, LNP doses range from about 0.1 mg/kg to about 100 mg/kg (or any of the values or subranges therein). In some embodiments, LNP doses is about 0.1 mg/kg, about 0.2 mg/kg, about 0.3 mg/kg, about 0.4 mg/kg, about 0.5 mg/kg, about 0.6 mg/kg, about 0.7 mg/kg, about 0.8 mg/kg, about 0.9 mg/kg, about 1.0 mg/kg, 1.5 mg/kg, about 2 mg/kg, about 2.5 mg/kg, about 3 mg/kg, about 3.5 mg/kg, about 4 mg/kg, about 4.5 mg/kg, about 5 mg/kg, about 6 mg/kg, about 7 mg/kg, about 7 mg/kg, about 8 mg/kg, about 9 mg/kg, about 10 mg/kg, about 15 mg/kg, about 20 mg/kg, about 25 mg/kg, about 30 mg/kg, about 35 mg/kg, about 40 mg/kg, about 45 mg/kg, or about 50 mg/kg or more.

[0431] In another embodiment, LNP doses of about 0.01 to about 1 mg per kg of body weight administered intravenously are contemplated. Medications to reduce the risk of infusion-related reactions are contemplated, such as dexamethasone, acetaminophen, diphenhydramine or cetirizine, and ranitidine are contemplated. Multiple doses of about 0.3 mg per kilogram every 4 weeks for five doses are also contemplated.

[0432] The charge of the LNP must be taken into consideration. As cationic lipids combined with negatively charged lipids to induce nonbilayer structures that facilitate intracellular delivery. Because charged LNPs are rapidly cleared from circulation following intravenous injection, ionizable cationic lipids with pKa values below 7 were developed (see, e.g., Rosin et al, *Molecular Therapy*, vol. 19, no. 12, pages 1286-2200, December 2011). Negatively charged

polymers such as RNA may be loaded into LNPs at low pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times. Four species of ionizable cationic lipids have been focused upon, namely 1,2-dilinoyleoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoyleoxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoyleoxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoyleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA). It has been shown that LNP siRNA systems containing these lipids exhibit remarkably different gene silencing properties in hepatocytes in vivo, with potencies varying according to the series DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP employing a Factor VII gene silencing model (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, December 2011). A dosage of 1 µg/ml of LNP in or associated with the LNP may be contemplated, especially for a formulation containing DLinKC2-DMA.

[0433] In some embodiments, the LNP composition comprises one or more one or more ionizable lipids. As used herein, the term "ionizable lipid" has its ordinary meaning in the art and may refer to a lipid comprising one or more charged moieties. In some embodiments, an ionizable lipid may be positively charged or negatively charged. In principle, there are no specific limitations concerning the ionizable lipids of the LNP compositions disclosed herein. In some embodiments, the one or more ionizable lipids are selected from the group consisting of 3-(didodecylamino)-N1,N1,4-tridodecyl-1-piperazineethanamine (KL10), N1-[2-(didodecylamino)ethyl]-N1,N4,N4-tridodecyl-1,4-piperazinediethanamine (KL22), 14,25-ditridecyl-15,18,21,24-tetraaza-octatriacontane (KL25), 1,2-dilinoyleoxy-N,N-dimethylaminopropane (DLin-DMA), 2,2-dilinoyleyl-4-dimethylaminomethyl-[1,3]-dioxolane (DLin-K-DMA), heptatriaconta-6,9,28,31-tetraen-19-yl 4-(dimethylamino)butanoate (DLin-MC3-DMA), 2,2-dilinoyleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLin-KC2-DMA), 1,2-dioleyloxy-N,N-dimethylaminopropane (DODMA), 2-({8-[(3)-cholest-5-en-3-yloxy]octyl}oxy)-N,N-dimethyl-3-[(9Z,12Z)-octadeca-9,12-dien-1-yloxy]propan-1-amine (Octyl-CLinDMA), (2R)-2-({8-[(3)-cholest-5-en-3-yloxy]octyl}oxy)-N,N-dimethyl-3-[(9Z,12Z)-- octadeca-9,12-dien-1-yloxy]propan-1-amine (Octyl-CLinDMA (2R)), and (2S)-2-({8-[(3)-cholest-5-en-3-yloxy]octyl}oxy)-N,N-dimethyl-3-[(9Z,12Z)-- octadeca-9,12-dien-1-yloxy]propan-1-amine (Octyl-CLinDMA (2S)). In one embodiment, the ionizable lipid may be selected from, but not limited to, an ionizable lipid described in International Publication Nos. WO2013086354 and WO2013116126.

[0434] In some embodiments, the lipid nanoparticle may include one or more (e.g., 1, 2, 3, 4, 5, 6, 7, or 8) cationic and/or ionizable lipids. Such cationic and/or ionizable lipids include, but are not limited to, 3-(didodecylamino)-N1,N1,4-tridodecyl-1-piperazineethanamine (KL10), N1-[2-(didodecylamino)ethyl]-N1,N4,N4-tridodecyl-1,4-piperazinediethanamine (KL22), 14,25-ditridecyl-15,18,21,24-tetraaza-octatriacontane (KL25), 1,2-dilinoleyloxy-N,N-dimethylaminopropane (DLin-DMA), 2,2-dilinoleyl-4-dimethylaminomethyl-[1,3]-dioxolane (DLin-K-DMA), heptatriaconta-6,9,28,31-tetraen-19-yl 4-(dimethylamino)butanoate (DLin-MC3-DMA), 2,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLin-KC2-DMA), 2-({8-[(3.beta.)-cholest-5-en-3-yloxy]octyl}oxy)-N,N-dimethyl-3-[(9Z,12Z)-octadeca-9,12-dien-1-yloxy]propan-1-amine (Octyl-CLinDMA), (2R)-2-({8-[(3.beta.)-cholest-5-en-3-yloxy]octyl}oxy)-N,N-dimethyl-3-[(9Z-,12Z)-octadeca-9,12-dien-1-yl oxy]propan-1-amine (Octyl-CLinDMA (2R)), (2S)-2-({8-[(3βcholest-5-en-3-yloxy]octyl}oxy)-N,N-dimethyl-3-[(9Z-,12Z)-octadeca-9,12-dien-1-yl oxy]propan-1-amine (Octyl-CLinDMA (2S)).N,N-dioleoyl-N,N-dimethylammonium chloride ("DODAC"); N-(2,3-dioleoyloxy)propyl-N,N--N-triethylammonium chloride ("DOTMA"); N,N-distearyl-N,N-dimethylammonium bromide ("DDAB"); N-(2,3-dioleoyloxy)propyl)-N,N,N-trimethylammonium chloride ("DOTAP"); 1,2-Dioleoyloxy-3-trimethylaminopropane chloride salt ("DOTAP.Cl"); 3-.beta.-(N--(N',N'-dimethylaminoethane)-carbamoyl)cholesterol ("DC-Chol"), N-(1-(2,3-dioleoyloxy)propyl)-N-2-(sperminecarboxamido)ethyl)-N,N-dimethyl- ammonium trifluoroacetate ("DOSPA"), dioctadecylamidoglycyl carboxyspermine ("DOGS"), 1,2-dioleoyl-3-dimethylammonium propane ("DODAP"), N,N-dimethyl-2,3-dioleoyloxy)propylamine ("DODMA"), and N-(1,2-dimyristyloxyprop-3-yl)-N,N-dimethyl-N-hydroxyethyl ammonium bromide ("DMRIE"). Additionally, a number of commercial preparations of cationic and/or ionizable lipids can be used, such as, e.g., LIPOFECTIN.RTM. (including DOTMA and DOPE, available from GIBCO/BRL), and LIPOFECTAMINE.RTM. (including DOSPA and DOPE, available from GIBCO/BRL). KL10, KL22, and KL25 are described, for example, in U.S. Pat. No. 8,691,750.

[0435] In some embodiments, the LNP composition comprises one or more amino lipids. The terms "amino lipid" and "cationic lipid" are used interchangeably herein to include those lipids and salts thereof having one, two, three, or more fatty acid or fatty alkyl chains and a pH-titratable amino head group (e.g., an alkylamino or dialkylamino head group). In principle, there are no specific limitations concerning the amino lipids of the LNP compositions disclosed herein. The cationic lipid is typically protonated (i.e., positively charged) at a pH below the pKa of the cationic lipid and is substantially neutral at a pH above the pKa. The cationic lipids

can also be termed titratable cationic lipids. In some embodiments, the one or more cationic lipids include: a protonatable tertiary amine (e.g., pH-titratable) head group; alkyl chains, wherein each alkyl chain independently has 0 to 3 (e.g., 0, 1, 2, or 3) double bonds; and ether, ester, or ketal linkages between the head group and alkyl chains. Such cationic lipids include, but are not limited to, DSDMA, DODMA, DOTMA, DLinDMA, DLenDMA, .gamma.-DLenDMA, DLin-K-DMA, DLin-K-C2-DMA (also known as DLin-C2K-DMA, XTC2, and C2K), DLin-K-C3-DMA, DLin-K-C4-DMA, DLen-C2K-DMA, y-DLen-C2-DMA, C12-200, cKK-E12, cKK-A12, cKK-O12, DLin-MC2-DMA (also known as MC2), and DLin-MC3-DMA (also known as MC3).

[0436] Anionic lipids suitable for use in lipid nanoparticles include, but are not limited to, phosphatidylglycerol, cardiolipin, diacylphosphatidylserine, diacylphosphatidic acid, N-dodecanoyl phosphatidylethanolamine, N-succinyl phosphatidylethanolamine, N-glutaryl phosphatidylethanolamine, lysylphosphatidylglycerol, and other anionic modifying groups joined to neutral lipids.

[0437] Neutral lipids (including both uncharged and zwitterionic lipids) suitable for use in lipid nanoparticles include, but are not limited to, diacylphosphatidylcholine, diacylphosphatidylethanolamine, ceramide, sphingomyelin, dihydrosphingomyelin, cephalin, sterols (e.g., cholesterol) and cerebroside. In some embodiments, the lipid nanoparticle comprises cholesterol. Lipids having a variety of acyl chain groups of varying chain length and degree of saturation are available or may be isolated or synthesized by well-known techniques. Additionally, lipids having mixtures of saturated and unsaturated fatty acid chains and cyclic regions can be used. In some embodiments, the neutral lipids used in the disclosure are DOPE, DSPC, DPPC, POPC, or any related phosphatidylcholine. In some embodiments, the neutral lipid may be composed of sphingomyelin, dihydrosphingomyelin, or phospholipids with other head groups, such as serine and inositol.

[0438] In some embodiments, amphipathic lipids are included in nanoparticles. Exemplary amphipathic lipids suitable for use in nanoparticles include, but are not limited to, sphingolipids, phospholipids, fatty acids, and amino lipids.

[0439] The lipid composition of the pharmaceutical composition may comprise one or more phospholipids, for example, one or more saturated or (poly)unsaturated phospholipids or a

combination thereof. In general, phospholipids comprise a phospholipid moiety and one or more fatty acid moieties.

[0440] A phospholipid moiety can be selected, for example, from the non-limiting group consisting of phosphatidyl choline, phosphatidyl ethanolamine, phosphatidyl glycerol, phosphatidyl serine, phosphatidic acid, 2-lysophosphatidyl choline, and a sphingomyelin.

[0441] A fatty acid moiety can be selected, for example, from the non-limiting group consisting of lauric acid, myristic acid, myristoleic acid, palmitic acid, palmitoleic acid, stearic acid, oleic acid, linoleic acid, alpha-linolenic acid, erucic acid, phytanoic acid, arachidic acid, arachidonic acid, eicosapentaenoic acid, behenic acid, docosapentaenoic acid, and docosahexaenoic acid.

[0442] Particular amphipathic lipids can facilitate fusion to a membrane. For example, a cationic phospholipid can interact with one or more negatively charged phospholipids of a membrane (e.g., a cellular or intracellular membrane). Fusion of a phospholipid to a membrane can allow one or more elements (e.g., a therapeutic agent) of a lipid-containing composition (e.g., LNPs) to pass through the membrane permitting, e.g., delivery of the one or more elements to a target tissue.

[0443] Non-natural amphipathic lipid species including natural species with modifications and substitutions including branching, oxidation, cyclization, and alkynes are also contemplated. For example, a phospholipid can be functionalized with or cross-linked to one or more alkynes (e.g., an alkenyl group in which one or more double bonds is replaced with a triple bond). Under appropriate reaction conditions, an alkyne group can undergo a copper-catalyzed cycloaddition upon exposure to an azide. Such reactions can be useful in functionalizing a lipid bilayer of a nanoparticle composition to facilitate membrane permeation or cellular recognition or in conjugating a nanoparticle composition to a useful component such as a targeting or imaging moiety (e.g., a dye).

[0444] Phospholipids include, but are not limited to, glycerophospholipids such as phosphatidylcholines, phosphatidylethanolamines, phosphatidylserines, phosphatidylinositols, phosphatidyl glycerols, and phosphatidic acids. Phospholipids also include phosphosphingolipid, such as sphingomyelin.

[0445] In some embodiments, the LNP composition comprises one or more phospholipids. In some embodiments, the phospholipid is selected from the group consisting of 1,2-dilinoleoyl-

sn-glycero-3-phosphocholine (DLPC), 1,2-dimyristoyl-sn-glycero-phosphocholine (DMPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), 1,2-diundecanoyl-sn-glycero-phosphocholine (DUPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1,2-di-O-octadecenyl-sn-glycero-3-phosphocholine (18:0 Diether PC), 1-oleoyl-2-cholesterylhemisuccinoyl-sn-glycero-3-phosphocholine (OChemPC), 1-hexadecyl-sn-glycero-3-phosphocholine (C16 Lyso PC), 1,2-dilinolenoyl-sn-glycero-3-phosphocholine, 1,2-diarachidonoyl-sn-glycero-3-phosphocholine, 1,2-didocosahexaenoyl-sn-glycero-3-phosphocholine, 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (ME 16:0 PE), 1,2-distearoyl-sn-glycero-3-phosphoethanolamine, 1,2-dilinoleoyl-sn-glycero-3-phosphoethanolamine, 1,2-dilinolenoyl-sn-glycero-3-phosphoethanolamine, 1,2-diarachidonoyl-sn-glycero-3-phosphoethanolamine, 1,2-didocosahexaenoyl--sn-glycero-3-phosphoethanolamine, 1,2-dioleoyl-sn-glycero-3-phospho-rac-(1-glycerol) sodium salt (DOPG), sphingomyelin, and any mixtures thereof.

[0446] Other phosphorus-lacking compounds, such as sphingolipids, glycosphingolipid families, diacylglycerols, and .beta.-acyloxyacids, may also be used. Additionally, such amphipathic lipids can be readily mixed with other lipids, such as triglycerides and sterols.

[0447] In some embodiments, the LNP composition comprises one or more helper lipids. The term "helper lipid" as used herein refers to lipids that enhance transfection (e.g., transfection of an LNP comprising an mRNA that encodes a site-directed endonuclease, such as a SpCas9 polypeptide). In principle, there are no specific limitations concerning the helper lipids of the LNP compositions disclosed herein. Without being bound to any particular theory, it is believed that the mechanism by which the helper lipid enhances transfection includes enhancing particle stability. In some embodiments, the helper lipid enhances membrane fusogenicity. Generally, the helper lipid of the LNP compositions disclosure herein can be any helper lipid known in the art. Non-limiting examples of helper lipids suitable for the compositions and methods include steroids, sterols, and alkyl resorcinols. Particularly helper lipids suitable for use in the present disclosure include, but are not limited to, saturated phosphatidylcholine (PC) such as distearoyl-PC (DSPC) and dipalmitoyl-PC (DPPC), dioleoylphosphatidylethanolamine (DOPE), 1,2-dilinoleoyl-sn-glycero-3-phosphocholine (DLPC), cholesterol, 5-

heptadecylresorcinol, and cholesterol hemisuccinate. In some embodiments, the helper lipid of the LNP composition includes cholesterol.

[0448] In some embodiments, the LNP composition comprises one or more structural lipids. As used herein, the term "structural lipid" refers to sterols and also to lipids containing sterol moieties. Without being bound to any particular theory, it is believed that the incorporation of structural lipids into the LNPs mitigates aggregation of other lipids in the particle. Structural lipids can be selected from the group including but not limited to, cholesterol, fecosterol, sitosterol, ergosterol, campesterol, stigmasterol, brassicasterol, tomatidine, tomatine, ursolic acid, alpha-tocopherol, hopanoids, phytosterols, steroids, and mixtures thereof. In some embodiments, the structural lipid is a sterol. As defined herein, "sterols" are a subgroup of steroids consisting of steroid alcohols. In certain embodiments, the structural lipid is a steroid. In some embodiments, the structural lipid is cholesterol. In certain embodiments, the structural lipid is an analog of cholesterol.

[0449] The lipid component of a lipid nanoparticle composition may include one or more molecules comprising polyethylene glycol, such as PEG or PEG-modified lipids. In some embodiments, the LNP composition disclosed herein comprise one or more polyethylene glycol (PEG) lipid. The term "PEG-lipid" refers to polyethylene glycol (PEG)-modified lipids. Such lipids are also referred to as PEGylated lipids. Non-limiting examples of PEG-lipids include PEG-modified phosphatidylethanolamine and phosphatidic acid, PEG-ceramide conjugates (e.g., PEG-CerC14 or PEG-CerC20), PEG-modified dialkylamines and PEG-modified 1,2-diacyloxypropan-3-amines. For example, a PEG lipid can be PEG-c-DOMG, PEG-DMG, PEG-DLPE, PEG-DMPE, PEG-DPPC, or a PEG-DSPE lipid. In some embodiments, the PEG-lipid includes, but not limited to 1,2-dimyristoyl-sn-glycerol methoxypolyethylene glycol (PEG-DMG), 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[amino(polyethylene glycol)] (PEG-DSPE), PEG-disteryl glycerol (PEG-DSG), PEG-dipalmitoyl, PEG-dioleoyl, PEG-distearyl, PEG-diacylglycamide (PEG-DAG), PEG-dipalmitoyl phosphatidylethanolamine (PEG-DPPE), or PEG-1,2-dimyristyloxylpropyl-3-amine (PEG-c-DMA). In some embodiments, the PEG-lipid is selected from the group consisting of a PEG-modified phosphatidylethanolamine, a PEG-modified phosphatidic acid, a PEG-modified ceramide, a PEG-modified dialkylamine, a PEG-modified diacylglycerol, a PEG-modified dialkylglycerol, and mixtures thereof. In some embodiments, the lipid moiety of the PEG-lipids includes those having lengths of from about C.sub.14 to about C.sub.22,

preferably from about C.sub.14 to about C.sub.16. In some embodiments, a PEG moiety, for example a mPEG-NH.sub.2, has a size of about 1000, 2000, 5000, 10,000, 15,000 or 20,000 daltons. In some embodiment, the PEG-lipid is PEG2k-DMG. In some embodiments, the one or more PEG lipids of the LNP composition comprises PEG-DMPE. In some embodiments, the one or more PEG lipids of the LNP composition comprises PEG-DMG.

[0450] In some embodiments, the ratio between the lipid components and the nucleic acid molecules of the LNP composition, e.g., the weight ratio, is sufficient for (i) formation of LNPs with desired characteristics, e.g., size, charge, and (ii) delivery of a sufficient dose of nucleic acid at a dose of the lipid component(s) that is tolerable for in vivo administration as readily ascertained by one of skill in the art.

[0451] In certain embodiments, it is desirable to target a nanoparticle, e.g., a lipid nanoparticle, using a targeting moiety that is specific to a cell type and/or tissue type. In some embodiments, a nanoparticle may be targeted to a particular cell, tissue, and/or organ using a targeting moiety. In particular embodiments, a nanoparticle comprises a targeting moiety. Exemplary non-limiting targeting moieties include ligands, cell surface receptors, glycoproteins, vitamins (e.g., riboflavin) and antibodies (e.g., full-length antibodies, antibody fragments (e.g., Fv fragments, single chain Fv (scFv) fragments, Fab' fragments, or F(ab')₂ fragments), single domain antibodies, camelid antibodies and fragments thereof, human antibodies and fragments thereof, monoclonal antibodies, and multispecific antibodies (e.g., bispecific antibodies)). In some embodiments, the targeting moiety may be a polypeptide. The targeting moiety may include the entire polypeptide (e.g., peptide or protein) or fragments thereof. A targeting moiety is typically positioned on the outer surface of the nanoparticle in such a manner that the targeting moiety is available for interaction with the target, for example, a cell surface receptor. A variety of different targeting moieties and methods are known and available in the art, including those described, e.g., in Sapra et al., *Prog. Lipid Res.* 42(5):439-62, 2003 and Abra et al., *J. Liposome Res.* 12:1-3, 2002.

[0452] In some embodiments, a lipid nanoparticle (e.g., a liposome) may include a surface coating of hydrophilic polymer chains, such as polyethylene glycol (PEG) chains (see, e.g., Allen et al., *Biochimica et Biophysica Acta* 1237: 99-108, 1995; DeFrees et al., *Journal of the American Chemistry Society* 118: 6101-6104, 1996; Blume et al., *Biochimica et Biophysica Acta* 1149: 180-184, 1993; Klivanov et al., *Journal of Liposome Research* 2: 321-334, 1992; U.S. Pat. No. 5,013,556; Zalipsky, *Bioconjugate Chemistry* 4: 296-299, 1993; Zalipsky, *FEBS*

Letters 353: 71-74, 1994; Zalipsky, in *Stealth Liposomes* Chapter 9 (Lasic and Martin, Eds) CRC Press, Boca Raton Fla., 1995). In one approach, a targeting moiety for targeting the lipid nanoparticle is linked to the polar head group of lipids forming the nanoparticle. In another approach, the targeting moiety is attached to the distal ends of the PEG chains forming the hydrophilic polymer coating (see, e.g., Klibanov et al., *Journal of Liposome Research* 2: 321-334, 1992; Kirpotin et al., *FEBS Letters* 388: 115-118, 1996).

[0453] Standard methods for coupling the targeting moiety or moieties may be used. For example, phosphatidylethanolamine, which can be activated for attachment of targeting moieties, or derivatized lipophilic compounds, such as lipid-derivatized bleomycin, can be used. Antibody-targeted liposomes can be constructed using, for instance, liposomes that incorporate protein A (see, e.g., Renneisen et al., *J. Bio. Chem.*, 265:16337-16342, 1990 and Leonetti et al., *Proc. Natl. Acad. Sci. (USA)*, 87:2448-2451, 1990). Other examples of antibody conjugation are disclosed in U.S. Pat. No. 6,027,726. Examples of targeting moieties can also include other polypeptides that are specific to cellular components, including antigens associated with neoplasms or tumors. Polypeptides used as targeting moieties can be attached to the liposomes via covalent bonds (see, for example Heath, *Covalent Attachment of Proteins to Liposomes*, 149 *Methods in Enzymology* 111-119 (Academic Press, Inc. 1987)). Other targeting methods include the biotin-avidin system.

[0454] In some embodiments, a lipid nanoparticle includes a targeting moiety that targets the lipid nanoparticle to a cell including, but not limited to, hepatocytes, colon cells, epithelial cells, hematopoietic cells, epithelial cells, endothelial cells, lung cells, bone cells, stem cells, mesenchymal cells, neural cells, cardiac cells, adipocytes, vascular smooth muscle cells, cardiomyocytes, skeletal muscle cells, beta cells, pituitary cells, synovial lining cells, ovarian cells, testicular cells, fibroblasts, B cells, T cells, reticulocytes, leukocytes, granulocytes, and tumor cells (including primary tumor cells and metastatic tumor cells). In particular embodiments, the targeting moiety targets the lipid nanoparticle to a hepatocyte.

[0455] The lipid nanoparticles described herein may be lipidoid-based. The synthesis of lipidoids has been extensively described and formulations containing these compounds are particularly suited for delivery of polynucleotides (see Mahon et al., *Bioconj Chem.* 2010 21:1448-1454; Schroeder et al., *J Intern Med.* 2010 267:9-21; Akinc et al., *Nat. Biotechnol.* 2008 26:561-569; Love et al., *Proc Natl Acad Sci USA.* 2010 107:1864-1869; Siegwart et al., *Proc Natl Acad Sci USA.* 2011 108:12996-3001).

[0456] The characteristics of optimized lipidoid formulations for intramuscular or subcutaneous routes may vary significantly depending on the target cell type and the ability of formulations to diffuse through the extracellular matrix into the blood stream. While a particle size of less than 150 nm may be desired for effective hepatocyte delivery due to the size of the endothelial fenestrae (see e.g., Akinc et al., *Mol Ther.* 2009 17:872-879), use of lipidoid oligonucleotides to deliver the formulation to other cells types including, but not limited to, endothelial cells, myeloid cells, and muscle cells may not be similarly size-limited.

[0457] In one aspect, effective delivery to myeloid cells, such as monocytes, lipidoid formulations may have a similar component molar ratio. Different ratios of lipidoids and other components including, but not limited to, a neutral lipid (e.g., diacylphosphatidylcholine), cholesterol, a PEGylated lipid (e.g., PEG-DMPE), and a fatty acid (e.g., an omega-3 fatty acid) may be used to optimize the formulation of the mRNA or system for delivery to different cell types including, but not limited to, hepatocytes, myeloid cells, muscle cells, etc. Exemplary lipidoids include, but are not limited to, DLin-DMA, DLin-K-DMA, DLin-KC2-DMA, 98N12-5, C12-200 (including variants and derivatives), DLin-MC3-DMA and analogs thereof. The use of lipidoid formulations for the localized delivery of nucleic acids to cells (such as, but not limited to, adipose cells and muscle cells) via either subcutaneous or intramuscular delivery, may also not require all of the formulation components which may be required for systemic delivery, and as such may comprise the lipidoid and the mRNA or system.

[0458] According to the present disclosure, a system described herein may be formulated by mixing the mRNA or system, or individual components of the system, with the lipidoid at a set ratio prior to addition to cells. In vivo formulations may require the addition of extra ingredients to facilitate circulation throughout the body. After formation of the particle, a system or individual components of a system is added and allowed to integrate with the complex. The encapsulation efficiency is determined using a standard dye exclusion assays.

[0459] In vivo delivery of systems may be affected by many parameters, including, but not limited to, the formulation composition, nature of particle PEGylation, degree of loading, oligonucleotide to lipid ratio, and biophysical parameters such as particle size (Akinc et al., *Mol Ther.* 2009 17:872-879; herein incorporated by reference in its entirety). As an example, small changes in the anchor chain length of poly(ethylene glycol) (PEG) lipids may result in significant effects on in vivo efficacy. Formulations with the different lipidoids, including, but not limited to penta[3-(1-laurylaminopropionyl)]-triethylenetetramine hydrochloride (TETA-

5LAP; aka 98N12-5, see Murugaiah et al., *Analytical Biochemistry*, 401:61 (2010)), C12-200 (including derivatives and variants), MD1, DLin-DMA, DLin-K-DMA, DLin-KC2-DMA and DLin-MC3-DMA can be tested for in vivo activity. The lipidoid referred to herein as "98N12-5" is disclosed by Akinc et al., *Mol Ther.* 2009 17:872-879). The lipidoid referred to herein as "C12-200" is disclosed by Love et al., *Proc Natl Acad Sci USA.* 2010 107:1864-1869 and Liu and Huang, *Molecular Therapy.* 2010 669-670.

[0460] The LNPs of the present disclosure, in which a nucleic acid is entrapped within the lipid portion of the particle and is protected from degradation, can be formed by any method known in the art including, but not limited to, a continuous mixing method, a direct dilution process, and an in-line dilution process. Additional techniques and methods suitable for the preparation of the LNPs described herein include coacervation, microemulsions, supercritical fluid technologies, phase-inversion temperature (PIT) techniques.

[0461] In some embodiments, the LNPs used herein are produced via a continuous mixing method, e.g., a process that includes providing an aqueous solution a nucleic acid described herein in a first reservoir, providing an organic lipid solution in a second reservoir (wherein the lipids present in the organic lipid solution are solubilized in an organic solvent, e.g., a lower alkanol such as ethanol), and mixing the aqueous solution with the organic lipid solution such that the organic lipid solution mixes with the aqueous solution so as to substantially instantaneously produce a lipid vesicle (e.g., liposome) encapsulating the nucleic acid molecule within the lipid vesicle. This process and the apparatus for carrying out this process are known in the art. More information in this regard can be found in, for example, U.S. Patent Publication No. 20040142025. The action of continuously introducing lipid and buffer solutions into a mixing environment, such as in a mixing chamber, causes a continuous dilution of the lipid solution with the buffer solution, thereby producing a lipid vesicle substantially instantaneously upon mixing. By mixing the aqueous solution comprising a nucleic acid molecule with the organic lipid solution, the organic lipid solution undergoes a continuous stepwise dilution in the presence of the buffer solution (e.g., aqueous solution) to produce a nucleic acid-lipid particle.

[0462] In some embodiments, the LNPs used herein are produced via a direct dilution process that includes forming a lipid vesicle (e.g., liposome) solution and immediately and directly introducing the lipid vesicle solution into a collection vessel containing a controlled amount of dilution buffer. In some embodiments, the collection vessel includes one or more elements

configured to stir the contents of the collection vessel to facilitate dilution. In some embodiments, the amount of dilution buffer present in the collection vessel is substantially equal to the volume of lipid vesicle solution introduced thereto.

[0463] In some embodiments, the LNPs are produced via an in-line dilution process in which a third reservoir containing dilution buffer is fluidly coupled to a second mixing region. In these embodiments, the lipid vesicle (e.g., liposome) solution formed in a first mixing region is immediately and directly mixed with dilution buffer in the second mixing region. These processes and the apparatuses for carrying out direct dilution and in-line dilution processes are known in the art. More information in this regard can be found in, for example, U.S. Patent Publication No. 20070042031.

6.10. Genes and Targets

[0464] This disclosure provides compositions and co-delivery methods for correcting or replacing genes or gene fragments (including introns or exons) or inserting genes in new locations. In certain embodiments, such a method comprises recombination or integration into a safe harbor site (SHS). A frequently used human SHS is the *AAVS1* site on chromosome 19q, initially identified as a site for recurrent adeno-associated virus insertion. Another locus comprises the human homolog of the murine *Rosa26* locus. Yet another SHS comprises the human H11 locus on chromosome 22. In some cases, a complete gene may be prohibitively large and replacement of an entire gene impractical. In certain embodiments, a method of the invention comprises recombining corrective gene fragments into a defective locus.

[0465] The methods and compositions can be used to target, without limitation, stem cells for example induced pluripotent stem cells (iPSCs), HSCs, HSPCs, mesenchymal stem cells, or neuronal stem cells and cells at various stages of differentiation. In certain embodiments, methods and compositions of the invention are adapted to target organoids, including patient derived organoids.

[0466] In certain embodiments, methods and compositions of the invention are adapted to treat muscle cells, not limited to cardiomyocytes for Duchene Muscular Dystrophy (DMD). The dystrophin gene is the largest gene in the human genome, spanning ~2.3 Mb of DNA. DMD is composed of 79 exons resulting in a 14-kb full-length mRNA. Common mutations include mutations that disrupt the reading frame or generate a premature stop codon. An aspect of DMD that lends it to gene editing as a therapeutic approach is the modular structure of the

dystrophin protein. Redundancy in the central rod domain permits the deletion of internal segments of the gene that may harbor loss-of-function mutations, thereby restoring the open reading frame (ORFs). In some embodiments, the methods and systems described herein are used to treat DMD by site-specifically integrating in the genome a polynucleotide template that repairs or replaces all or a portion of the defective DMD gene.

[0467] The following are non limiting diseases that may be treated utilizing the methods and compositions of the present disclosure:

Inherited Retinal Diseases:

- Stargardt Disease (ABCA4)
- Leber congenital amaurosis 10 (CEP290)
- X linked Retinitis Pigmentosa (RPGR)
- Autosomal Dominant Retinitis Pigmentosa (RHO)

Liver Diseases:

- Wilson's disease (ATP7B)
- Alpha-1 antitrypsin (SERPINA1)

Intellectual Disabilities:

- Rett Syndrome (MECP2)
- SYNGAP1-ID (SYNGAP1)
- CDKL5 deficiency disorder (CDKL5)

Peripheral Neuropathies:

- Charcot-Marie-Tooth 2A (MFN2)

Lung Diseases:

- Cystic Fibrosis (CFTR)
- Alpha-1 Antitrypsin (SERPINA1)

Blood disorders:

- Sickle Cell
- Hemophilia,
- Factor VIII or
- Factor IX

- CFTR (cystic fibrosis transmembrane conductance regulator)

[0468] Over 2500 mutations have been identified associated with various diseases and defects.

[0469] The most common cystic fibrosis (CF) mutation F508del removes a single amino acid. In some embodiments, recombining human CFTR into an SHS of a cell that expresses CFTR F508del is a corrective treatment path. In some embodiments, the methods and systems described herein are used to CF by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing CF. Proposed validation is detection of persistent CFTR mRNA and protein expression in transduced cells.

[0470] **Sickle cell disease (SCD)** is caused by mutation of a specific amino acid - valine to glutamic acid at amino acid position 6. In some embodiments, SCD is corrected by recombination of the HBB gene into a safe harbor site (SHS) and by demonstrating correction in a proportion of target cells that is high enough to produce a substantial benefit. In some embodiments, the methods and systems described herein are used to sickle cell disease by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing the disease. In some embodiments, validation is detection of persistent HBB mRNA and protein expression in transduced cells.

[0471] **DMD - Duchenne Muscular Dystrophy**

[0472] The dystrophin gene is the largest gene in the human genome, spanning ~2.3 Mb of DNA. DMD is composed of 79 exons resulting in a 14-kb full-length mRNA. Common mutations include mutations that disrupt the reading frame or generate a premature stop codon. An aspect of DMD that lends it to gene editing as a therapeutic approach is the modular structure of the dystrophin protein. Redundancy in the central rod domain permits the deletion of internal segments of the gene that may harbor loss-of-function mutations, thereby restoring the open reading frame (ORFs).

[0473] In some embodiments, recombination will be into safe harbor sites (SHS). A frequently used human SHS is the *AAVS1* site on chromosome 19q, initially identified as a site for recurrent adeno-associated virus insertion. In some embodiments, the site is the human homolog of the murine *Rosa26* locus (pubmed.ncbi.nlm.nih.gov/18037879). In some embodiments, the site is the human H11 locus on chromosome 22. Proposed target cells for

recombination include stem cells for example induced pluripotent stem cells (iPSCs) and cells at various stages of differentiation. In some cases, a complete gene may be prohibitively large and replacement of an entire gene impractical. In such instances, rescuing mutants by recombining in corrected gene fragments with the methods and systems described herein is a corrective option.

[0474] In some embodiments, correcting mutations in exon 44 (or 51) by recombining in a corrective coding sequence downstream of exon 43 (or 50), using the methods and systems described herein is a corrective option. Proposed validation is detection of persistent DMD mRNA and protein expression in transduced cells.

[0475] F8 (Factor VIII)

[0476] A large proportion of severe hemophilia A patients harbor one of two types of chromosomal inversions in the FVIII gene. The recombinase technology and methods described herein are well suited to correcting such inversions (and other mutations) by recombining of the FVIII gene into a SHS.

[0477] In some embodiments, correcting factor VIII deficiency by recombining the FVIII gene into an SHS is a corrective path. In some embodiments, the methods and systems described herein are used to correct factor VIII deficiency by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing the FIX deficiency. Proposed validation is detection of persistent FVIII mRNA and protein expression in transduced cells.

[0478] Factor 9 (Factor IX)

[0479] Hemophilia B, also called factor IX (FIX) deficiency is a genetic disorder caused by missing or defective factor IX, a clotting protein.

[0480] In some embodiments, the methods and systems described herein are used to correct factor IX deficiency by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing the FIX deficiency. Proposed validation is detection of persistent FiX mRNA and protein expression in transduced cells.

6.11. Methods of treatment

[0481] In another aspect, methods of treatment are presented. The method comprises administering an effective amount of the pharmaceutical composition comprising the nucleic acid construct or vectorized nucleic acid construct described above to a patient in need thereof. In some embodiments, the system (e.g., any of the systems described herein) are delivered to a cell *ex vivo* and the cell is then administered to the subject. In some embodiments, the systems (e.g., any of the systems described herein) are delivered to a patient, thereby delivering to a cell *in vivo*.

[0482] DNA or RNA viral vectors can be administered directly to patients (*in vivo*) or they can be used to treat cells *in vitro*, and the modified cells may optionally be administered to patients (*ex vivo*). Conventional viral based systems to be used herein could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0483] In some embodiments, the co-delivery system described herein (e.g., a gene editor construct packaged in a LNP and a donor template packaged in a vector) is administered intravenously. In some embodiments, the co-delivery system described herein (e.g., a gene editor construct packaged in a LNP and a donor template packaged in a vector) is administered intrathecally. In some embodiments, the co-delivery system described herein (e.g., a gene editor construct packaged in a LNP and a donor template packaged in a vector) is administered by intracerebral ventricular injection. In some embodiments, the co-delivery system described herein (e.g., a gene editor construct packaged in a LNP and a donor template packaged in a vector) is administered by intracisternal magna administration. In some embodiments, the co-delivery system described herein (e.g., a gene editor construct packaged in a LNP and a donor template packaged in a vector) is administered by intravitreal injection.

[0484] Methods of non-viral delivery of the donor DNA template described herein include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™

and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

6.11.1.1 mRNA delivery

[0485] Another useful method to deliver proteins, enzymes, and guides comprises transfection of messenger RNA (mRNA). Examples of mRNA delivery methods and compositions that may be utilized in the present disclosure including, for example, PCT/US2014/028330, US8822663B2, NZ700688A, ES2740248T3, EP2755693A4, EP2755986A4, WO2014152940A1, EP3450553B1, BR112016030852A2, and EP3362461A1. Expression of CRISPR systems in particular is described by WO2020014577. Each of these publications are incorporated herein by reference in their entireties. Additional disclosure hereby incorporated by reference can be found in Kowalski et al., “Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery,” Mol Therap., 2019; 27(4): 710-728.

6.12. Additional Embodiments

[0486] Embodiment 1. A method of co-delivering to a cell a gene editor polynucleotide construct and a template polynucleotide construct, the method comprising co-delivering:
a lipid nanoparticle (LNP) comprising a gene editor polynucleotide construct; and
a vector comprising a donor template polynucleotide construct.

[0487] Embodiment 2. The method of embodiment 1, wherein the gene editor polynucleotide construct is capable of localizing to a cell cytoplasm.

[0488] Embodiment 3. The method of embodiment 1, wherein the donor template polynucleotide construct is capable of localizing to a cell nucleus.

[0489] Embodiment 4. The method of embodiment 1 or embodiment 2, wherein the gene editor polynucleotide construct comprises:

- a polynucleotide sequence encoding a prime editor fusion protein or a Gene Writer™ protein;
- a one or more polynucleotide sequence encoding an attachment site-containing guide RNA (atgRNA);
- optionally, a polynucleotide sequence encoding a nickase guide RNA (ngRNA);

a polynucleotide sequence encoding an integrase;
optionally, a polynucleotide sequence encoding a recombinase.

[0490] Embodiment 5. The method of embodiment 4, wherein the integrase that is encoded by a polynucleotide sequence in the gene editor polynucleotide construct is fused to the prime editor fusion protein or the Gene WriterTM protein encoded by a gene editor polynucleotide construct, and wherein the fusion is optionally by a linker.

[0491] Embodiment 6. The method of any of embodiment 4 or embodiment 5, wherein the one or more atgRNA encodes an integrase target recognition site or a recombinase recognition site.

[0492] Embodiment 7. The method of any of the previous embodiments, wherein the vector comprising a donor template polynucleotide construct, the vector is recombinant adenovirus, helper dependent adenovirus, AAV, lentivirus, HSV, anellovirus, retrovirus, DoggyboneTM DNA (dbDNA), minicircle, plasmid, miniDNA, exosome, fusosome, or nanoplasmid.

[0493] Embodiment 8. The method of any of the previous embodiments, wherein the donor template is capable of being integrated into a genomic locus that contains an integrase target recognition site or a recombinase target integrase site.

[0494] Embodiment 9. The method of any of the previous embodiments, wherein the donor template comprises at least one of the following: a gene, a gene fragment, an expression cassette, a logic gate system, or any combination thereof.

[0495] Embodiment 10. The method of any of the previous embodiments, wherein the donor template further comprises at least one integrase target recognition site or a recombinase target integrase site.

[0496] Embodiment 11. The method of any of the previous embodiments, wherein the donor template is capable of self-circularization to form a circularized nucleic acid.

[0497] Embodiment 12. The circularized nucleic acid of embodiment 11, wherein the self-circularizing is mediated by an integrase or recombinase.

[0498] Embodiment 13. A pharmaceutical co-delivery composition comprising:

- (a) a lipid nanoparticle (LNP) comprising a gene editor polynucleotide construct (i) capable of localizing to a cell cytoplasm; and
- (b) a vector comprising a donor template polynucleotide construct (ii) capable of localizing to a cell nucleus.

[0499] Embodiment 14. A pharmaceutical co-delivery composition of embodiment 13, wherein the gene editor polynucleotide construct comprises:

- a polynucleotide sequence encoding a prime editor fusion protein or a Gene WriterTM protein;
 - a polynucleotide sequence encoding an attachment site-containing guide RNA; optionally, a polynucleotide sequence encoding a nickase guide RNA (ngRNA);
 - a polynucleotide sequence encoding an integrase;
 - optionally, a polynucleotide sequence encoding a recombinase; and
- wherein the donor template polynucleotide construct is packaged in recombinant adenovirus, helper dependent adenovirus, AAV, lentivirus, HSV, anellovirus, retrovirus, Doggybone DNA (dbDNA), minicircle, plasmid, miniDNA, exosome, fusosome, or nanoplasmid.

[0500] Embodiment 15. A method comprising administering an effective amount of the pharmaceutical composition of embodiment 13 or embodiment 14, to a patient in need thereof.

7. EXAMPLES

7.1. Example 1: Delivery of gene editor polynucleotide sequence packaged in LNP and donor template packaged in AAV

[0501] A gene editor polynucleotide construct is packaged into a LNP (FIG. 1), wherein the gene editor polynucleotide sequence comprises a polynucleotide sequence encoding a prime editor protein linked to an integrase via peptide linker a polynucleotide sequence encoding an attachment site-containing guide RNA (atgRNA); a polynucleotide sequence encoding a nickase guide RNA (ngRNA).

[0502] A donor template polynucleotide construct is packaged in an AAV vector (FIG. 2).

[0503] Co-administration of the gene editor construct packaged LNP and the donor template packaged AAV co-delivers the gene editor construct to a cell cytoplasm and the donor template to a cell nucleus. By use of programmable genome editing to place integrase landing

site at a desired location in the genome, the direct activity of the associated integrase to the specific genomic site is guided. Gene editor construct expression, with template co-delivery, results in integration of template “cargo” at a precisely defined target location.

7.2. Example 2: Delivery of gene editor polynucleotide sequence packaged in LNP and donor template capable of self-circularization packaged in AAV

[0504] A gene editor polynucleotide construct is packaged into a LNP (FIG. 1), wherein the gene editor polynucleotide sequence comprises a polynucleotide sequence encoding a prime editor protein linked to an integrase via peptide linker a polynucleotide sequence encoding an attachment site-containing guide RNA (atgRNA); a polynucleotide sequence encoding a nickase guide RNA (ngRNA).

[0505] A donor template polynucleotide construct is packaged in an AAV vector (FIG. 2).

[0506] Co-administration of the gene editor construct packaged LNP and the donor template packaged AAV co-delivers the gene editor construct to a cell cytoplasm and the donor template to a cell nucleus. Integrase-mediated self-circularization of donor template occurs at integration target recognition sites within the AAV genome (FIG. 3). By use of programmable genome editing to place an orthogonal integrase landing site (i.e., distinct att site from att sites used for self-circularization) at a desired location in the genome, the direct activity of the associated integrase to the specific genomic site is guided. Gene editor construct expression, with template co-delivery and integrase-mediated circularization of template, results in integration of template “cargo” at a precisely defined target location.

7.3. Example 3: Delivery of gene editor polynucleotide sequence packaged in LNP and atgRNA, ngRNA, and donor template co-packaged in AAV

[0507] A gene editor polynucleotide construct is packaged into a LNP (FIG. 4), wherein the gene editor polynucleotide sequence comprises a polynucleotide sequence encoding a prime editor protein linked to an integrase via peptide linker.

[0508] A polynucleotide sequence encoding an attachment site-containing guide RNA (atgRNA), a polynucleotide sequence encoding a nicking guide RNA (ngRNA), and donor template are packaged in an AAV vector (FIG. 4).

[0509] Co-administration of the gene editor construct packaged LNP and the atgRNA, ngRNA, donor template packaged AAV co-delivers the gene editor construct to a cell.

Integrase-mediated self-circularization of donor template occurs at integration target recognition sites within the AAV genome (FIG. 3). By use of programmable genome editing to place an orthogonal integrase landing site (i.e., distinct att site from att sites used for self-circularization) at a desired location in the genome, the direct activity of the associated integrase to the specific genomic site is guided. Gene editor construct expression, with atgRNA, ngRNA, and template co-delivery and integrase-mediated circularization of template, results in integration of template “cargo” at a precisely defined target location.

7.4. Example 4: Delivery of gene editor polynucleotide sequence and ngRNA packaged in LNP and atgRNA and donor template co-packaged in AAV

[0510] A gene editor polynucleotide construct and a nicking guide RNA (ngRNA) are packaged into a LNP (FIG. 5), wherein the gene editor polynucleotide sequence comprises a polynucleotide sequence encoding a prime editor protein linked to an integrase via peptide linker.

[0511] A polynucleotide sequence encoding an attachment site-containing guide RNA (atgRNA) and donor template are packaged in an AAV vector (FIG. 5).

[0512] Co-administration of the gene editor construct and ngRNA packaged LNP and the atgRNA, donor template packaged AAV co-delivers the gene editor construct to a cell. Integrase-mediated self-circularization of donor template occurs at integration target recognition sites within the AAV genome (FIG. 3). By use of programmable genome editing to place an orthogonal integrase landing site (i.e., distinct att site from att sites used for self-circularization) at a desired location in the genome, the direct activity of the associated integrase to the specific genomic site is guided. Gene editor construct expression, with atgRNA, ngRNA, and template co-delivery and integrase-mediated circularization of template, results in integration of template “cargo” at a precisely defined target location.

7.5. Example 5: Intramolecular circularization of plasmid and packaged AAV genomes

[0513] Three self-complementary AAV (scAAV) genomes were designed and generated to verify recombinase/integrase-mediated intramolecular circularization of a DNA cargo from within a linear AAV genome (FIGs. 6A-6B). Circularization of a scAAV genome is mediated by one of Cre, FLPe (thermostable mutant), or Bxb1. Further, the scAAV genomes are comprised of a DNA cargo of interest (“payload”) and an attP site (GT central dinucleotide for

circularization orthogonality) for gene insertion into a genome placed attB beacon site. Expected recombinase/integrase-mediated intramolecular circularization products are illustrated in FIG. 7. A universal ddPCR probe capable of binding any linear or circularized AAV genome was designed, wherein the universal ddPCR probe is designed to only give signal upon cognate recombinase/integrase mediated circularization (FIGs. 8A-8B). Circularization products are amplified by use of a circle junction PCR primer set that is designed to amplify only circular products due to primer direction constraints. To confirm Bxb1 mediated circularization specifically, an attR scar quencher-fluorophore probe was designed. In addition, a template reference primer set was designed and generated to quantify total template DNA (linear or circular confirmation) (FIGs. 8A-8B).

[0514] Intracellular circularization of either plasmid or packaged AAV genomes were screened in HEK293 cells (35K cells per well) (FIG. 9). Plasmids (25 fmol pDNA = 1X or 50 fmol pDNA = 2X) encoding one of Cre, FLPe, or Bxb1 were transfected by Lipofectamine 3000. Plasmid genome substrates were transfected at a dose of 1E10 copies per well using Lipofectamine 3000 (FIG. 9). Additionally, AAV genomes were packaged in AAV-DJ capsids and delivered at a dose of 3E5 genomes per cell or 1E10 genomes per well. Circularization ddPCR analysis was conducted three days post transfection.

[0515] FIG. 10 demonstrates circularization of AAV pDNA and packaged AAV genomic DNA for both 1X Bxb1 and 2X Bxb1 conditions (confirmed by use of attR ddPCR primer set). Further, replicates that lacked either Bxb1 or AAV pDNA substrate demonstrated insignificant circularization. All three of the Cre-, FLPe-, and Bxb1-targeted AAV pDNA substrates demonstrated circularization upon cognate recombinase/integrase introduction, as confirmed by using the universal ddPCR probe (FIG. 11). Moreover, Cre-, FLPe-, and Bxb1-mediated circularization of packaged AAV DJ genomes substrates were demonstrated and confirmed using the universal ddPCR probe (FIG. 12).

[0516] As shown in FIG. 13, the Bxb1-mediated attR scar probe provided similar percent circularization quantification compared to the universal probe.

7.6. Example 6: In vitro beacon placement in primary mouse hepatocytes and primary human hepatocytes using mRNA and AAV for co-delivery

[0517] This example assessed the efficiency of in vitro beacon placement in primary human hepatocytes using mRNA delivering of a polynucleotide encoding a gene editor

polynucleotide construct and AAV to deliver the first and second atgRNA. See FIG. 15 for a non-limiting example of a dual atgRNA-mediated insertion of an integration recognition site.

[0518] In the mouse experiments, the mRNA and AAV were delivered into the primary mouse hepatocytes (PMH) using (i) concurrent delivery (“co-dose”), (ii) AAV delivery followed by a “1-day delay” before delivery of the mRNA, or (iii) AAV delivery followed by a “2-day delay” before delivery of the mRNA. Beacon placement was then assessed using next-generation sequencing of DNA isolated from cells subjected to the delivery conditions mentioned above. The mRNA encoding the gene editor polynucleotide construct was delivered in various amounts per well: 2000 ng, 1000 ng, 500 ng, 250 ng, 125 ng, 62.5 ng, and 31.25 ng. AAV encoding the first and second atgRNA (see Table 12). The primary mouse hepatocyte data is shown in FIG. 16 and the human primary hepatocyte data is shown in FIG. 17.

Table 12 atgRNAs			
SEQ ID NO:	Target	Name	Sequence
559	Mouse Nole1	AAV-mNole1-F (AAVG023)	GACGCGTTTTACCCGGAGCAGTTTAAGAGCTAT GCTGGAACAGCATAGCAAGTTTAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCACGACGGAGACCGCCGTCGTCGAC AAGCCTCCGGGTAAAACG
560	Mouse Nole1	AAV-mNole1-R	ACAAGGGGATAAAGGTCGCTGTTTAAGAGCTAT GCTGGAACAGCATAGCAAGTTTAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCACGACGGCGGTCTCCGTCGTCAGGA TCATGACCTTTATCCCC
561	Human Factor IX	AAV-hF9-F (AAVG048)	CTTGTATGCCCCGAGAAGTGGTTTTAGAGCTAG AAATAGCAAGTTAAATAAGGCTAGTCCGTTAT CAACTTGAAAAAGTGGCACCGAGTCGGTGCAC GACGGAGACCGCCGTCGTCGACAAGCCTTCTCG GGGCATA
562	Human Factor IX	AAV-hF9-R (AAVG048)	TATATATACTTGCTAGGGCTGTTTTAGAGCTAG AAATAGCAAGTTAAATAAGGCTAGTCCGTTAT CAACTTGAAAAAGTGGCACCGAGTCGGTGCAC GACGGCGGTCTCCGTCGTCAGGATCATCCTAGC AAGTATA

[0519] As shown in FIG. 16, in primary mouse hepatocytes (PMH) delivering the first atgRNAs (SEQ ID NO: 543) and the second atgRNA (SEQ ID NO: 544) using AAV at day 0 and then delivering the mRNA encoding the gene editing polynucleotide construct at day 2 (“2 day delay”) resulted in greater than 10% beacon placement for each amount of mRNA tested. Surprisingly, a 2 day delay resulted in greater beacon placement than either no delay (“co-dose”) or a 1 day delay.

[0520] As shown in FIG. 17, in primary human hepatocytes (PHH), using AAV to deliver the first atgRNA (SEQ ID NO: 545) and the second atgRNA (SEQ ID NO: 546) and mRNA to deliver the gene editing polynucleotide construct resulted in about 17% beacon placement.

[0521] Taken together, this data showed robust ex vivo beacon placement in primary mouse and primary human hepatocytes.

7.7. Example 7: In vivo beacon placement with mRNA + AAV guide

[0522] In vivo beacon placement in mice was assessed using AAV to deliver the first and second atgRNAs and mRNA to delivery the gene editing polynucleotide construct.

[0523] In these experiments, mice were administered AAV containing the first atgRNA (SEQ ID NO: 543; Table 12) and the second atgRNA (SEQ ID NO: 544) targeting the *Nolc1* locus at 3E11 to 1E12 vector genomes (vg) per animal two 2 weeks prior to administration of the mRNA containing the gene editing polynucleotide construct (see FIG. 18). mRNA was delivered using various LNP formulations (e.g., LP01 (LNP #F1), ALC-0315 (i.e., LNP #F2), and cKK-E12 (i.e., LNP #F3)) at concentrations ranging from 5 mg/kg to 0.5 mg/kg via intravenous injection (see FIG. 18). After delivery of the mRNA, liver tissue was harvested, genomic DNA was isolated, and beacon efficiency was assessed by NGS. As shown in FIG. 18, three conditions resulted in vivo beacon placement efficiency greater than 10%.

[0524] Taken together, this data provided proof-of-concept for successful in vivo beacon placement using AAV to deliver the first and second atgRNA and LNPs to deliver the mRNA encoding the gene editor polynucleotide construct.

7.8. Example 8: In vivo integration in mice using AAV to deliver the template polynucleotide and Adenovirus to deliver BxB1

[0525] In vivo integration efficiency in AttP mice was assessed using adenovirus to deliver an integrase (e.g., Bxb1) and an AAV to deliver the template polynucleotide.

[0526] For these experiments, the adenovirus (i.e., adenovirus containing polynucleotide encoding the integrase) and the AAV (i.e., AAV containing the template polynucleotide and an attB site) were administered to mice containing dual AttP sites integrated in to the Rosa26 locus (B6.RosaBxb-GT/GA; female, Strain# 036152). The Rosa26 locus included a first AttP site comprising a GT dinucleotide and a second AttP site comprising a GA dinucleotide. The AAV was a scAAV8 containing a vector having a template polynucleotide and a 38 bp GT AttB site. The Adenovirus was an adenovirus-type 5 (Ad5) containing a polynucleotide encoding Bxb1 (“Bxb1 AdV”) (SEQ ID NO: 563; Table 14). Mice were administered the adenovirus and AAV according to the experimental details in Table 13.

Table 13. Experimental Details for assessment of in vivo integration efficiency							
Group	n	Bxb1 AdV dose (vg/animal)	Cargo AAV Dose (vg/animal)	Route	Volume (ul)	Conc. (vg/ml)	Time points
1	1F, 2M	vehicle		IV	100		Liver punches at 10 days post-dose
2	5	3E10	1E12	IV	100	3E11+1E13	
3	5	1E11	1E12	IV	100	1E12+1E13	

Table 14. Adenovirus Vector	
Vectors	Sequence
Bxb1 AdV (SEQ ID NO: 563)	TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACG GTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGT CAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACCTATGCGGCATCAGAGCAGATT GTACTGAGAGTGACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAA AATACCGCATCAGGCGCCATTTCGCCATTTCAGGCTGCGCAACTGTTGGGAAGGGCGA TCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAG GCGATTAAAGTTGGGTAACGCCAGGGTTTCCAGTCACGACGTTGTAACGACGG CCAGTGAATTCGAGCTCTCGCTATTACTTGGCCACTCCCTCTCTGCGCGCTCGCTCG CTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGG CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGG TTCCTCACTGCCCCGAGATCTACTAGTGGCTTGTCGACGACGGCGGTCTCCGTCGTC AGGATCATTAGGTCAGTGAAGAGAAGAACAAAAGCAGCATATTACAGTTAGTTG TCTTCATCAATCTTAAATATGTTGTGTGGTTTTCTCTCCCTGTTCCACAGTTATG GGCAACAGCTTCAGCACCCAGCGCCTTCGGCCCTGTGGCCTTTTCTCTGGGCCTCCTG

Table 14. Adenovirus Vector

Vectors	Sequence
	CTCGTGCTGCCTGCCGCTTTTCCAGCTCCTGTGTTACCCCTGGAAGATTTCGTGGGA GATTGGCGGCAGACCGCCGGCTACAACCTGGACCAAGTGCTGGAACAGGGCGGAG TGTCCAGCCTGTTTCAGAACCTGGGCGTCTCCGTGACCCCTATCCAGCGGATCGTGC TGAGCGGCGAGAACGGCCTGAAAATCGACATCCATGTGATTATCCCCTACGAGGGC CTGAGCGGAGATCAGATGGGCCAGATCGAGAAAATCTTCAAGGTGGTGTACCCCG TCGACGACCACCACTTCAAGGTGATCCTGCACTACGGCACCCCTGGTGATCGACGGC GTTACCCCTAACATGATCGACTACTTCGGCAGACCCTATGAGGGAATTGCCGTGTT CGACGGCAAGAAAATCACCGTGACCGGCACACTGTGGAACGGCAACAAGATCATC GATGAGCGCCTGATCAACCCAGACGGCAGCCTGCTGTTTCAGAGTGACAATCAATGG CGTGACAGGCTGGAGACTTTGTGAAAAGAATCCTGGCCGGTTCTGGCGAGGGCAGA GGATCTCTGCTGACATGCGGCGATGTGGAAGAGAATCCTGGACCTGCTATGAAAAT CGAGTGCGAGAATTACAGGCACACTGAACGGAGTTGAATTCGAGCTGGTCGGCGGA GGCGAGGGCACACCTGAGCAGGGCAGAATGACCAACAAGATGAAAAGCACCAAG GGCGCCCTGACCTTTTCTCCTTACCTGCTGAGCCACGTGATGGGCTATGGCTTCTAC CACTTCGGCACCTACCCAGCGGCTATGAAAACCCCTTCTGTCATGCTATCAACAA CGGAGGCTACACCAATACCAGAATCGAGAAGTACGAGGACGGCGGCGTGTGCAC GTGTCCTTCAGCTACAGATACGAGGCCGGCAGAGTGATCGGCGACTTCAAGGTGGT GGGCACAGGATTTCCAGAAGATAGCGTGATCTTACCGACAAGATCATCCGGAGC AACGCCACCGTGGAACACCTGCACCCCATGGGCGATAATGTGCTGGTGGGCTCCTT TGCTAGAACATTCTCCCTGCGGGACGGCGGATACTACAGCTTCGTGGTGCAGACC ACATGCACCTCAAGTCTGCCATCCACCCTTCTATCCTGCAGAACGGCGACCTATGT TCGCCTTCCGGCGGGTGGAAGAACTCCACAGCAACACCGAGCTGGGCATCGTGGA ATACCAGCACGCCTTTAAGACCCCTATCGCCTTCGCCAGAAGCAGAGCCAGGTGAG AGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT TTGCCCTTCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGCTCTTC CTAATAAAATGAGAAAATTGCATCGCATTGTCTGAGTAGGTGTCACTTCTATTCTGG GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGC ATGCTGGGGATGCGGTGGGCTCTATGGACTAGTAGATCTCACTGCCCGCCCACTCC CTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAGGTGCGCCGACGCC CGGGCTTTGCCCGGGCGGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGATGCAT TAATGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTC ATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGC CGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA ATTGCGTTGCGCTCACTGCCCGCTTTCAGTCGGGAAACCTGTCGTGCCAGTCGAT TAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCGCTCTTCCGC TTCCTCGCTCACTGACTCGCTGCGCTCGGTCTCGGCTGCGGCGAGCGGTATCAGC TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAG AACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGC TGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAGAAAAATCGACGCTCA AGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCG CCTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCA GTTCCGTGTAGGTGCTTCTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTTCA CCCGACCGCTGCGCCTTATCCGGTAACATATCGTCTTGAGTCCAACCCGGTAAGACA CGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT GTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG GACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG GTAGCTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGCA AGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCT ACGGGGTCTGACGCTCAGTGGAACGAAAACCTACGTTAAGGGATTTTGGTCATGAG ATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAATAAATGAAGTTTTAAATC AATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTG AGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCTGACTCCCCG TCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATG ATACCGCGAGACCCACGCTACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGC

Table 14. Adenovirus Vector	
Vectors	Sequence
	CGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTA TTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC GTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCA TTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAA AAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGT ATGCGGCGACCGAGTTGCTCTTGCCCCGGCGTCAATACGGGATAATACCGCGCCACA TAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCT CAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCCAAC TGATCTTCAGCATCTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG GCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGAAATGTTGAATACTCATA CTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGA TACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCC CCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATA AAAAATAGGCGTATCACGAGGCCCTTTCGTC

[0527] Ten days after administration of the AdV and AAV viruses, liver punches were collected and genomic DNA was isolated. ddPCR of the genomic DNA was used to assess integration efficiency.

[0528] As shown in FIG. 19, administering the AAV and AdV resulted in *in vivo* integration of the donor polynucleotide template into the AttP mice. In particular, 3E10 vg/animal BxB1 AdV resulted in about 7% *in vivo* integration efficiency (see FIG. 19). Administering increased amounts of BxB1 AdV, 1E11 vg/animal, resulted in higher integration efficiency, about 11%, in AttP mice than with lower amount of 3E10 vg/animal (see FIG. 19).

[0529] Overall, this data establishes proof-of-concept for *in vivo* integration using an adenovirus to deliver and drive expression of Bxb1 and an AAV to deliver the template polynucleotide to be integrated into a mammalian genome, in this case, the mouse genome.

7.9. Example 9: *In vivo* beacon placement in neonatal mice using split LNP

[0530] *In vivo* beacon placement was assessed in neonatal mice following administration of a single dose of a mixture of two LNPs. The first LNP contained mRNA encoding a prime editing system and a first synthetic atgRNA (atgRNA1). The mRNA and atgRNA1 were included at 1:1 ratio in the first LNP. The second LNP contained mRNA encoding a prime editing system and a second synthetic atgRNA (atgRNA2). The mRNA and atgRNA2 were included at a 1:1

ratio in the second LNP. Each of the first and second atgRNAs targeted the mouse *Nolc1* locus and each encoded a portion of an integration recognition site (a “beacon”). AtgRNA1 and atgRNA2 together included a 6bp overlap. The first and second LNPs were combined 1:1 as mixture prior to administration. The first atgRNA and second atgRNA are provide in Table 15, where the atgRNA include one or more 2’O-methyl modifications and one or more phosphorothioate linkages.

Table 15. atgRNAs			
SEQ ID NO:	Target	Name	Sequence
564	Mouse Nolc1	mNolc1-F (synthetic guide, 6bp overlap)	mG*mA*mC*rGrCrGrUrUrUrArCrCrCrGrGrArGrCrArGrUrUrUrArGrAmGmCmUmAmGmAmAmUmAmGmCrArArGrUrUrArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrAmAmCmUmUmGmAmAmAmAmGmUmGmGmCmAmCmCmGmAmGmUmCmGmGmUmGmCrArGrArCrCrGrCrCrGrUrCrGrUrCrGrArCrArArGrCrCrUrCrCrGrGrGrUrArArA*mA*mC*mG
565	Mouse Nolc1	mNolc1-R (synthetic guide, 6bp overlap)	mA*mC*mA*rArGrGrGrGrArUrArArArGrGrUrCrGrCrUrGrUrUrUrUrArGrAmGmCmUmAmGmAmAmUmAmGmCrArArGrUrUrArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrAmAmCmUmUmGmAmAmAmAmGmUmGmGmCmAmCmCmGmAmGmUmCmGmGmUmGmCrCrGrGrUrCrUrCrCrGrUrCrGrUrCrArGrGrArUrCrArUrGrArCrCrUrUrUrArUrC*mC*mC*mC

[0531] The LNP mixture was administered to the neonatal mice (2-5 day old CD-1 mice) according to the experimental details in Table 16.

Table 16. Experimental details for in vivo beacon placement in neonatal mice.							
Group	n	Treatment	Dose (mg/kg)	Route	Volume (ml/kg)	Conc. (mg/ml)	Time points
1	5	vehicle		IV	5		Whole liver on

2	3	LNP	1	IV	5	0.2	day 8 post-dose (168 hours)
3	4	LNP	3	IV	5	0.6	
4	5	vehicle		IV	5		Liver punches (one 8mm punch from each lobe) at 6 weeks post-dose
5	5	LNP	1	IV	5	0.2	
6	5	LNP	3	IV	5	0.6	

[0532] Eight days after administration of the LNP mixture in vivo beacon placement was assessed. In particular, at day 8 post administration, liver samples (either whole liver for groups 1-3 or liver punches from each lobe for groups 4-6 (see Table 13)) were collected and genomic DNA was isolated. Beacon placement was detected using ddPCR and NGS.

[0533] As shown in FIG. 20A, ddPCR revealed about 1% beacon placement (in *NolcI* alleles) following administration of a 3mg/kg dose of the LNP mixture. Confirmation of beacon placement using NGS showed about 7% beacon placement (in *NolcI* alleles) following administration of a 3 mg/kg dose of the LNP mixture (see FIG. 20B). In order to determine what percentage of the integrated beacons included the expected integration recognition site (“perfect beacon”), an NGS-based assay was used to make this assessment. As shown in FIG. 20C, about 1% of the integrated beacons contained the expected integration recognition site.

[0534] Neonates were also assessed at six weeks after administration of the LNP mixture. Beacon placement was detected using ddPCR and NGS. As shown in FIG. 21A., at six weeks post administration, ddPCR revealed about 4% beacon placement (in *NolcI* alleles) for a 3mg/kg dose of the LNP mixture. Confirmation of beacon placement using NGS showed about 15% beacon placement (in *NolcI* alleles) for a 3 mg/kg dose of the LNP mixture (see FIG. 21B). Assessment of the percent of integrated beacons that included the expected integration recognition site (“perfect beacon”) revealed that about 3.5% of beacons were comprised of perfect beacons (see FIG. 21C).

[0535] Overall, this data demonstrated successful in vivo site-specific integration of an integration recognition site. In particular, this data showed that a split LNP approach can be

used for site-specifically integrating an integration recognition site in vivo in a mammalian genome, in this case neonatal mice.

7.10. Example 10: In vivo beacon placement in mice using split LNP

[0536] In vivo beacon placement was assessed in adult mice using a single dose mixture of two LNPs. The first LNP contained mRNA encoding a prime editing system and a first synthetic atgRNA (atgRNA1). The mRNA and atgRNA1 were included at different ratios (e.g., 1:0.5, 1:1, and 1:2) ratio in the first LNP. The second LNP contained mRNA encoding a prime editing system and a second synthetic atgRNA (atgRNA2). The mRNA and atgRNA2 were included at different ratios (e.g., 1:0.5, 1:1, and 1:2) ratio in the second LNP. Here, the first and second atgRNAs targeted mouse Factor IX (“mF9”) locus and each encoded a portion of an integration recognition site (“beacon”). Similar to Example 9, atgRNA1 and atgRNA2 together included a 6bp overlap and were combined 1:1 as mixture prior to administration. The first atgRNA and second atgRNA are provide in Table 17, where the atgRNA include one or more 2’O-methyl modifications and one or more phosphorothioate linkages.

Table 17. atgRNAs			
SEQ ID NO:	Target	Name	Sequence
566	Mouse Factor IX	mF9-F (synthetic guide, 6bp overlap)	mA*mG*mU*rGrArCrArGrUrGrCrCrArGrGrArUrCrArGrGrUrUrUrArGrAmGmCmUmAmGmAmAmAmUmAmGmCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrAmAmCmUmUmGmAmAmAmAmAmGmUmGmGmCmAmCmCmGmAmGmUmCmGmGmUmGmCrArGrArCrCrGrCrCrGrUrCrGrUrCrGrArCrArArGrCrCrArUrCrCrUrGrGrCrArCmU*mG*mU
567	Mouse Factor IX	mF9-R (synthetic guide, 6bp overlap)	mG*mU*mU*rGrArCrArUrCrArUrGrUrCrUrGrGrArGrUrGrUrUrUrUrArGrAmGmCmUmAmGmAmAmAmUmAmGmCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrAmAmCmUmUmGmAmAmAmAmAmGmUmGmGmCmAmCmCmGmAmGmUmCmGmGmUmGmCrCrGrGrUrCrUrCrCrGrUrCrGrUrCrArGrGrArUrCrArUrCrCrArGrArCrArUrGrAmU*mG*mU

[0537] In particular, the LNP mixture was administered to female CD-1 mice 6-8 weeks old according to the experimental details in Table 18.

Table 18. Experimental details for in vivo beacon placement in adult mice							
Group	n	Treatment (ratio mRNA: atgRNA1: atgRNA2)	Dose (mg/kg)	Route	Volume (ml/kg)	Conc. (mg/ml)	Time points
1	5	vehicle		IV	5		Terminal : liver punches on day 8
2	5	1:0.25:0.25*	3	IV	5	0.6	
3	5	1:0.5:0.5**	3	IV	5	0.6	
4	5	1:1:1***	3	IV	5	0.6	

*1:0.25:0.25 = mRNA:atgRNA1 1:0.5; mRNA:atgRNA2 1:0.5; LNPs mixed 1:1

**1:0.5:0.5 = mRNA:atgRNA1 1:1; mRNA:atgRNA2 1:1; LNPs mixed 1:1

***1:1:1 = mRNA:atgRNA1 1:2; mRNA:atgRNA2 1:2; LNPs mixed 1:1

[0538] Eight days after administration of the LNP mixture *in vivo* beacon placement was assessed. In particular, at day 8 post administration, liver samples (i.e., liver punches of each lobe (see Table 14)) were collected and genomic DNA was isolated. Beacon placement was detected using ddPCR and NGS.

[0539] As shown in FIG. 22A, ddPCR revealed about 0.8% beacon placement (in *mF9* alleles) following administration of a 1:0.25:0.25 ratio of mRNA:atgRNA1:atgRNA2. Confirmation of beacon placement using NGS showed about 14% beacon placement (in *mF9* alleles) following administration of the 1:0.25:0.25 ratio of mRNA:atgRNA1:atgRNA2 (see FIG. 22B). Similar to Example 9, an NGS-based assay was used to determine what percentages of the integrated beacons included the expected integration recognition site ("perfect beacon"). As shown in FIG. 22C, about 0.02% of the beacons placed in the *mF9* locus were "perfect" beacons.

[0540] Overall, this data showed successful *in vivo* site-specific integration of an integration recognition site in adult mice. In particular, this data showed that the ratio of mRNA

to atgRNA is an important consideration in determining efficacy of in vivo site-specific integration of an integration recognition site.

8. EQUIVALENTS AND INCORPORATION BY REFERENCE

[0541] All references cited herein are incorporated by reference to the same extent as if each individual publication, database entry (e.g. Genbank sequences or GeneID entries), patent application, or patent, was specifically and individually indicated incorporated by reference in its entirety, for all purposes. This statement of incorporation by reference is intended by Applicants, pursuant to 37 C.F.R. §1.57(b)(1), to relate to each and every individual publication, database entry (e.g. Genbank sequences or GeneID entries), patent application, or patent, each of which is clearly identified in compliance with 37 C.F.R. §1.57(b)(2), even if such citation is not immediately adjacent to a dedicated statement of incorporation by reference. The inclusion of dedicated statements of incorporation by reference, if any, within the specification does not in any way weaken this general statement of incorporation by reference. Citation of the references herein is not intended as an admission that the reference is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents.

[0542] It is an object of the invention not to encompass within the invention any previously known product, process of making the product, or method of using the product such that Applicant reserves the right and hereby disclose a disclaimer of any previously known product, process, or method. It is further noted that the invention does not intend to encompass within the scope of the invention any product, process, or making of the product or method of using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. 112(a)) or the EPO (Article 83 of the EPC), such that Applicant reserves the right and hereby disclose a disclaimer of any previously described product, process of making the product, or method of using the product. It may be advantageous in the practice of the invention to be in compliance with Art. 53(c) EPC and Rule 28(b) and (c) EPC. Nothing herein is to be construed as a promise. It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow

for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

[0543] While the invention has been particularly shown and described with reference to a preferred embodiment and various alternate embodiments, it is understood by persons skilled in the relevant art that various changes in form and details can be made therein without departing from the spirit and scope of the invention.

WHAT IS CLAIMED IS:

1. A method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the method comprising:
delivering to a cell:
 - (a) a lipid nanoparticle (LNP) comprising:
 - (i) a gene editor polynucleotide; and
 - (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) at least a first attachment site-containing guide RNA (atgRNA).
2. The method of claim 1, wherein the gene editor polynucleotide is capable of localizing to a cell cytoplasm.
3. The method of claim 1, wherein the template polynucleotide is capable of localizing to a cell nucleus.
4. The method of claim 1 or 2, wherein the gene editor polynucleotide comprises:
a polynucleotide sequence encoding a prime editor system.
5. The method of claim 4, wherein the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.
6. The method of claim 5, wherein the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.
7. The method of claim 6, wherein the nickase is linked to the reverse transcriptase by in-frame fusion.
8. The method of claim 6, wherein the nickase is linked to the reverse transcriptase by a linker.
9. The method of claim 8, wherein the linker is a peptide fused in-frame between the nickase and reverse transcriptase.

10. The method of any one of claims 1-9, wherein the gene editor polynucleotide further comprises:
 - a polynucleotide sequence encoding at least a first integrase.
11. The method of claim 10, wherein the linked nickase-reverse transcriptase are further linked to the first integrase.
12. The method of any one of claims 1-9, further comprising delivering a second vector.
13. The method of claim 12, wherein the second vector comprises a polynucleotide sequence encoding at least a first integrase.
14. The method of any one of claims 10-13, wherein the first integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.
15. The method of any one of claims 1-14, wherein the gene editor polynucleotide further comprises a polynucleotide sequence encoding a recombinase.
16. The method of claim 15, wherein the recombinase is FLP or Cre.
17. The method of any one of claims 1-16, wherein the first atgRNA comprises:
 - (i) a domain that is capable of guiding the prime editor system to a target sequence; and
 - (ii) a reverse transcriptase (RT) template that comprises at least a portion of an at least first integration recognition site.
18. The method of claim 14, wherein the RT template comprises the entirety of the first integration recognition site.
19. The method of any one of claim 1-15, wherein the vector further comprises a second atgRNA.
20. The method of claim 19, wherein the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein
 - the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence;
 - the first atgRNA further includes a first RT template that comprises at least a portion of an at least first integration recognition site;

the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and
the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

21. The method of any one of claims 1-18, wherein the vector further comprises a nicking gRNA.
22. The method of any one of claims 1-18, wherein the LNPs further comprises a nicking gRNA.
23. The method of any one of claims 1-21, wherein the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.
24. The method of any one of claims 1-23, wherein the template polynucleotide comprises a second integration recognition site.
25. The method of claim 24, wherein the second integration recognition site is a cognate pair with the first integration recognition site.
26. The method of any one of claims 1-23, wherein the template polynucleotide comprises at least a third integration recognition site.
27. The method of claim 26, wherein the template polynucleotide further comprises at least a fourth integration recognition site.
28. The method of claim 26, wherein the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.
29. The method of any one of claims 1-28, wherein the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.
30. The method of claim 29, wherein the sub-sequence of the vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.

31. The method of any one of claims 26-30, wherein the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.
32. The method of claim 31, wherein self-circularizing is mediated by recombination of the third integration recognition site and a fourth integration recognition site by the integrase.
33. The method of any one of claims 29-32, wherein the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.
34. The method of any one of claims 1-33, wherein the vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an annelovirus, a retrovirus, a DoggyboneTM DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid.
35. The method of any one of claims 1-34, wherein the LNP and the vector are concurrently delivered.
36. The method of any one of claims 1-34, wherein the LNP and the vector are delivered separately.
37. The method of claim 36, wherein the LNP and the vector are delivered at least 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 weeks apart.
38. The method of any one of claims 1-37, wherein the cell is *in vivo*.
39. A method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the method comprising:
delivering to a cell:
 - (a) a lipid nanoparticle (LNP) comprising:
 - (i) a gene editor polynucleotide, and
 - (ii) a first attachment site-containing guide RNA (atgRNA); and
 - (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) a second atgRNA.

40. A method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the method comprising:
delivering:
- (a) a lipid nanoparticle (LNP) comprising:
 - (i) a gene editor polynucleotide,
 - (ii) a first attachment site-containing guide RNA (atgRNA), and
 - (iii) a second atgRNA; and
 - (b) a vector comprising:
 - (i) a template polynucleotide.
41. A method for delivering a system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the method comprising:
delivering:
- (a) a lipid nanoparticle (LNP) comprising:
 - (i) a gene editor polynucleotide, and
 - (ii) a first attachment site-containing guide RNA (atgRNA); and
 - (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) a nicking atgRNA.
42. The method of any one of claims 39-41, wherein the gene editor polynucleotide comprises:
a polynucleotide sequence encoding a prime editor system.
43. The method of claim 42, wherein the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.
44. The method of claim 43, wherein the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.
45. The method of claim 44, wherein the nickase is linked to the reverse transcriptase by in-frame fusion.

46. The method of claim 44, wherein the nickase is linked to the reverse transcriptase by a linker.
47. The method of claim 46, wherein the linker is a peptide fused in-frame between the nickase and reverse transcriptase.
48. The method of any one of claims 39-47, wherein the gene editor polynucleotide construct further comprises:
a polynucleotide sequence encoding at least a first integrase.
49. The method of claim 48, wherein the linked nickase-reverse transcriptase are further linked to the integrase.
50. The method of any one of claims 39-49, further comprising delivering a second vector.
51. The method of claim 50, wherein the second vector comprises a polynucleotide sequence encoding at least a first integrase.
52. The method of any one of claims 48-51, wherein the first integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.
53. The method of any one of claims 39-52, wherein the gene editor polynucleotide construct further comprises a polynucleotide sequence encoding a recombinase.
54. The method of claim 53, wherein the recombinase is FLP or Cre.
55. The method of any one of claims 41-54, wherein the first atgRNA comprises:
(i) a domain that is capable of guiding the prime editor system to a target sequence; and
(ii) a reverse transcriptase (RT) template that comprises at least a portion of an at least first integration recognition site.
56. The method of claim 55, wherein the RT template comprises the entirety of the first integration recognition site.
57. The method of any one of claims 39, 40 or 42-54, wherein the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein

the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence;

the first atgRNA further includes a first RT template that comprises at least a portion of the first integration recognition site;

the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and

the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

58. The method of any one of claims 39-57, wherein the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.
59. The method of any one of claims 39-58, wherein the template polynucleotide comprises a second integration recognition site.
60. The method of claim 59, wherein the second integration recognition site is a cognate pair with the first integration recognition site.
61. The method of any one of claims 39-60, wherein the template polynucleotide comprises at least a third integration recognition site.
62. The method of claim 61, wherein the template polynucleotide further comprises at least a fourth integration recognition site.
63. The method of claim 62, wherein the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.
64. The method of any one of claims 39-63, wherein the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.
65. The method of claim 64, wherein the sub-sequence of the vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.
66. The method of claim 64 or 65, wherein the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.

67. The method of claim 66, wherein self-circularizing is mediated by recombination of the third integration recognition site and the fourth integration recognition site by the integrase.
68. The method of any one of claims 65-67, wherein the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.
69. The method of any one of claims 39-68, wherein the vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an annelovirus, a retrovirus, a DoggyboneTM DNA (dbDNA), a minicircle, a plasmid, a miniDNA, a exosome, a fusosome, or a nanoplasmid.
70. The method of any one of claims 39-69, wherein the LNP and the vector are concurrently delivered.
71. The method of any one of claims 39-69, wherein the LNP and the vector are delivered separately.
72. The method of claim 71, wherein the LNP and the vector are delivered at least 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 weeks apart.
73. The method of any one of claims 39-72, wherein the cell is *in vivo*.
74. A method of co-delivering a system capable of site-specifically integrating at least a first integration recognition site into the genome of a cell, the method comprising: co-delivering to a cell:
- (a) a first lipid nanoparticle (LNP) comprising:
 - (i) a first gene editor polynucleotide, and
 - (ii) a first attachment site-containing guide RNA (atgRNA); and
 - (b) a second lipid nanoparticle (LNP) comprising:
 - (i) a second gene editor polynucleotide, and
 - (ii) a second attachment site-containing guide RNA (atgRNA),
- wherein the first atgRNA and the second atgRNA are an at least first pair of atgRNAs.

75. The method of claim 74, further comprising mixing the first LNP and the second LNP prior to co-delivering to the cell.
76. The method of claim 75, wherein the first LNP and the second LNP are mixed at a ratio of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.
77. The method of any one of claims 74-76, wherein the first gene editor polynucleotide construct, the second gene editor polynucleotide construct, or both comprise:
a polynucleotide sequence encoding a prime editor system.
78. The method of claim 77, wherein the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.
79. The method of claim 78, wherein the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.
80. The method of claim 79, wherein the nickase is linked to the reverse transcriptase by in-frame fusion.
81. The method of claim 79, wherein the nickase is linked to the reverse transcriptase by a linker.
82. The method of claim 81, wherein the linker is a peptide fused in-frame between the nickase and reverse transcriptase.
83. The method of any one of claims 74-82, wherein the first gene editor polynucleotide, construct, the second gene editor polynucleotide construct, or both, further comprise:
a polynucleotide sequence encoding an integrase.
84. The method of claim 83, wherein the linked nickase-reverse transcriptase are further linked to the integrase.
85. The method of any one of claims 74-84, wherein the first gene editor polynucleotide, the second gene editor polynucleotide, or both, further comprise:
a polynucleotide sequence encoding a recombinase.

86. The method of claim 85, wherein the linked nickase-reverse transcriptase are further linked to the recombinase.
87. The method of any one of claims 74-86, wherein the first gene editor polynucleotide and the second gene editor polynucleotide are the same.
88. The method of any one of claims 74-87, wherein the first gene editor polynucleotide is mRNA, the second gene editor polynucleotide is mRNA, or both the first and second gene editor polynucleotides are mRNA.
89. The method of claim 88, wherein the first LNP comprises a ratio of mRNA to atgRNA of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.
90. The method of claim 88 or 89, wherein the second LNP comprises a ratio of mRNA to atgRNA of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.
91. The method of any one of claims 74-82, further comprising delivering an integrase.
92. The method of claim 91, wherein delivering the integrase comprises co-delivering the integrase with (a) and (b).
93. The method of claim 91 or 92, wherein the method comprises delivering a polynucleotide sequence encoding the integrase.
94. The method of claim 93, wherein the polynucleotide sequence is encoded in a first vector.
95. The method of claim 94, wherein the first vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an anneovirus, a retrovirus, a DoggyboneTM DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid..
96. The method of claim 93, wherein the first vector further comprises a template polynucleotide and a sequence that is an integration cognate with the first integration recognition site.
97. The method of any one of claims 74-96, further comprising delivering a recombinase.

98. The method of claim 97, wherein delivering the recombinase comprises co-delivering the recombinase with (a) and (b).
99. The method of claim 97 or 98, wherein the method comprises delivering a polynucleotide sequence encoding the recombinase.
100. The method of claim 99, wherein the polynucleotide sequence is encoded in the first vector.
101. The method of any one of claims 74-100, further comprising delivering a second vector.
102. The method of claim 101, wherein the second vector comprises a template polynucleotide and a sequence that is an integration cognate with the first integration recognition site.
103. The method of claims 101, wherein the second vector is a vector selected from: an adenovirus, an AAV, a lentivirus, an HSV, an anneovirus, a retrovirus, Doggybone™ DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid.
104. The method of any one of claims 96-103, wherein the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.
105. The method of any one of claims 96-104, wherein the template polynucleotide comprises a second integration recognition site.
106. The method of claim 105, wherein the second integration recognition site is a cognate pair with the first integration recognition site.
107. The method of any one of claims 96-106, wherein the template polynucleotide comprises at least a third integration recognition site.
108. The method of claim 107, wherein the template polynucleotide further comprises at least a fourth integration recognition site.

109. The method of claim 108, wherein the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.
110. The method of any one of claims 96-109, wherein the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.
111. The method of claim 110, wherein the sub-sequence of the vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.
112. The method of claim 110 or 111, wherein the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.
113. The method of claim 112, wherein self-circularizing is mediated by recombination of the third integration recognition site and a fourth integration recognition site by the integrase.
114. The method of any one of claims 110-113, wherein the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.
115. The method of any one of claims 74-114, wherein
the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence,
the first atgRNA further includes a first RT template that comprises at least a portion of a first integration recognition site;
the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and
the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.
116. The method of claim 115, wherein the first integration site is an AttB sequence, a FRT sequence, or a VOX sequence.
117. The method of any one of claims 74-116, wherein the first atgRNA, the second atgRNA or both are synthetic.

118. The method of any one of claims 91-117, wherein the integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.
119. The method of any one of claims 74-118, wherein the cell is *in vivo*.
120. A system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising:
- (a) a lipid nanoparticle (LNP) comprising:
 - (i) a gene editor polynucleotide construct; and
 - (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) at least a first attachment site-containing guide RNA (atgRNA).
121. The system of claim 120, wherein the gene editor polynucleotide construct comprises a polynucleotide sequence encoding a prime editor system.
122. The system of claim 121, wherein the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.
123. The system of claim 122, wherein the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.
124. The system of claim 123, wherein the nickase is linked to the reverse transcriptase by in-frame fusion.
125. The system of claim 123, wherein the nickase is linked to the reverse transcriptase by a linker.
126. The system of claim 125, wherein the linker is a peptide fused in-frame between the nickase and reverse transcriptase.
127. The system of any one of claims 120-126, wherein the gene editor polynucleotide construct further comprises:
- a polynucleotide sequence encoding at least a first integrase.

128. The system of claim 127, wherein the linked nickase-reverse transcriptase are further linked to the first integrase.
129. The system of any one of claims 120-126, further comprising a second vector.
130. The system of claim 129, wherein the second vector comprises a polynucleotide sequence encoding at least a first integrase.
131. The system of any one of claims 127-130, wherein the first integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.
132. The system of any one of claims 120-131, wherein the gene editor polynucleotide construct further comprises a polynucleotide sequence encoding a recombinase.
133. The system of claim 132, wherein the recombinase is FLP or Cre.
134. The system of any one of claims 120-133, wherein the first atgRNA comprises:
- (i) a domain that is capable of guiding the prime editor system to a target sequence; and
 - (ii) a reverse transcriptase (RT) template that comprises at least a portion of an at least first integration recognition site.
135. The system of claim 134, wherein the RT template comprises the entirety of the first integration recognition site.
136. The system of any one of claim 120-133, wherein the vector further comprises a second atgRNA.
137. The system of claim 136, wherein the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein
- the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence;
 - the first atgRNA further includes a first RT template that comprises at least a portion of the first integration recognition site;

the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and
the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

138. The system of any one of claims 120-135, wherein the vector further comprises a nicking gRNA.
139. The system of any one of claims 120-135, wherein the LNP further comprises a nicking gRNA.
140. The system of any one of claims 120-139, wherein the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.
141. The system of any one of claims 120-140, wherein the template polynucleotide comprises a second integration recognition site.
142. The system of claim 141, wherein the second integration recognition site is a cognate pair with the first integration recognition site.
143. The system of any one of claims 120-142, wherein the template polynucleotide comprises at least a third integration recognition site.
144. The system of claim 143, wherein the template polynucleotide construct further comprises at least a fourth integration recognition site.
145. The system of claim 143, wherein the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.
146. The system of any one of claims 120-145, wherein the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.
147. The system of claim 146, wherein the sub-sequence of vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.

148. The system of claim 146 or 147, wherein the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.
149. The system of claim 148, wherein self-circularizing is mediated by recombination of the third integration recognition site and the fourth integration recognition site by the integrase.
150. The system of any one of claims 146-149, wherein the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.
151. The system of any one of claims 120-150, wherein the vector is a recombinant adenovirus, a helper dependent adenovirus, or an adeno-associated virus.
152. A system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising:
- (a) a lipid nanoparticle (LNP) comprising:
 - (i) a gene editor polynucleotide, and
 - (ii) a first attachment site-containing guide RNA (atgRNA); and
 - (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) a second atgRNA.
153. A system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising:
- (a) a lipid nanoparticle (LNP) comprising
 - (i) a gene editor polynucleotide,
 - (ii) a first attachment site-containing guide RNA (atgRNA), and
 - (iii) a second atgRNA; and
 - (b) a vector comprising:
 - (i) a template polynucleotide.
154. A system capable of site-specifically integrating a template polynucleotide into the genome of a cell, the system comprising:
- (a) a lipid nanoparticle (LNP) comprising
 - (i) a gene editor polynucleotide, and

- (ii) a first attachment site-containing guide RNA (atgRNA); and
 - (b) a vector comprising:
 - (i) a template polynucleotide, and
 - (ii) a nicking gRNA.
155. The system of any one of claims 152-154, wherein the gene editor polynucleotide comprises:
a polynucleotide sequence encoding a prime editor system.
156. The system of claim 155, wherein the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.
157. The system of claim 156, wherein the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.
158. The system of claim 157, wherein the nickase is linked to the reverse transcriptase by in-frame fusion.
159. The system of claim 157, wherein the nickase is linked to the reverse transcriptase by a linker.
160. The system of claim 159, wherein the linker is a peptide fused in-frame between the nickase and reverse transcriptase.
161. The system of any one of claims 152-160, wherein the gene editor polynucleotide further comprises:
a polynucleotide sequence encoding at least a first integrase.
162. The system of claim 161, wherein the linked nickase-reverse transcriptase are further linked to the first integrase.
163. The system of any one of claims 152-162, further comprising a second vector.
164. The system of claim 163, wherein the second vector comprises a polynucleotide sequence encoding at least a first integrase.

165. The system of any one of claims 161-164, wherein the first integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.
166. The system of any one of claims 152-165, wherein the gene editor polynucleotide further comprises a polynucleotide sequence encoding a recombinase.
167. The system of claim 166, wherein the recombinase is FLP or Cre.
168. The system of any one of claims 152-167, wherein the first atgRNA comprises:
- (i) a domain that is capable of guiding the prime editor system to a target sequence; and
 - (ii) a reverse transcriptase (RT) template that comprises at least a portion of an at least first integration recognition site.
169. The system of claim 168, wherein the RT template comprises the entirety of the first integration recognition site.
170. The system of any one of claims 152, 153 or 155-169, wherein the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein
- the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence;
 - the first atgRNA further includes a first RT template that comprises at least a portion of the first integration recognition site;
 - the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and
 - the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.
171. The system of any one of claims 152-170, wherein the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.
172. The system of any one of claims 152-171, wherein the template polynucleotide comprises a second integration recognition site.

173. The system of claim 172, wherein the second integration recognition site is a cognate pair with the first integration recognition site.
174. The system of any one of claims 152-173, wherein the template polynucleotide comprises at least a third integration recognition site.
175. The system of claim 174, wherein the template polynucleotide construct further comprises at least a fourth integration recognition site.
176. The system of claim 175, wherein the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.
177. The system of any one of claims 152-176, wherein the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.
178. The system of claim 177, wherein the sub-sequence of vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.
179. The system of claim 177 or 178, wherein the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.
180. The system of claim 179, wherein self-circularizing is mediated by recombination of the third integration recognition site and the fourth integration recognition site by the integrase.
181. The system of any one of claims 178-180, wherein the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.
182. The system of any one of claims 152-181, wherein the vector is recombinant adenovirus, helper dependent adenovirus, or an adeno-associated virus.
183. A system capable of site-specifically integrating at least a first integration recognition site into the genome of a cell, the system comprising:
- (a) a first lipid nanoparticle (LNP) comprising:
 - (i) a first gene editor polynucleotide, and

- (ii) a first attachment site-containing guide RNA (atgRNA); and
- (b) a second lipid nanoparticle (LNP) comprising:
 - (i) a second gene editor polynucleotide, and
 - (ii) a second attachment site-containing guide RNA (atgRNA).

184. The system of claim 183, wherein the first atgRNA and the second atgRNA are an at least first pair of atgRNAs.
185. The system of claim 184, wherein the first LNP and the second LNP are mixed at a ratio of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.
186. The system of any one of claims 183-185, wherein the first gene editor polynucleotide, the second gene editor polynucleotide, or both comprise:
a polynucleotide sequence encoding a prime editor system.
187. The system of claim 186, wherein the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.
188. The system of claim 187, wherein the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the gene editor polynucleotide such that when expressed the nickase is linked to the reverse transcriptase.
189. The system of claim 188, wherein the nickase is linked to the reverse transcriptase by in-frame fusion.
190. The system of claim 188, wherein the nickase is linked to the reverse transcriptase by a linker.
191. The system of claim 190, wherein the linker is a peptide fused in-frame between the nickase and reverse transcriptase.
192. The system of any one of claims 183-191, wherein the first gene editor polynucleotide, the second gene editor polynucleotide, or both, further comprise:
a polynucleotide sequence encoding an integrase.

193. The system of claim 192, wherein the linked nickase-reverse transcriptase are further linked to the integrase.
194. The system of any one of claims 183-193, wherein the first gene editor polynucleotide, the second gene editor polynucleotide, or both, further comprise: a polynucleotide sequence encoding a recombinase.
195. The system of claim 194, wherein the nickase-reverse transcriptase are further linked to the recombinase.
196. The system of any one of claims 183-195, wherein the first gene editor polynucleotide and the second gene editor polynucleotide are the same.
197. The system of any one of claims 183-196, wherein the first gene editor polynucleotide is mRNA, the second gene editor polynucleotide is mRNA, or both the first and second gene editor polynucleotides are mRNA.
198. The system of claim 197, wherein the first LNP comprises a ratio of mRNA to atgRNA of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.
199. The system of claim 197 or 198, wherein the second LNP comprises a ratio of mRNA to atgRNA of 1:0.25, 1:0.5, 1:0.75, 1:1, 0.75:1, 0.5:1, or 0.25:1.
200. The system of any one of claims 183-191, further comprising an integrase.
201. The system of claim 200, wherein the system comprises a polynucleotide sequence encoding the integrase.
202. The system of claim 201, wherein the polynucleotide sequence is encoded in a first vector.
203. The system of claim 202, wherein the first vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an anneovirus, a retrovirus, a Doggybone™ DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid.

204. The system of claim 202 or 203, wherein the first vector further comprises a template polynucleotide and a sequence that is an integration cognate with the first integration recognition site.
205. The system of any one of claims 183-204, further comprising delivering a recombinase.
206. The system of claim 205, wherein delivering the recombinase comprises co-delivering the recombinase with (a) and (b).
207. The system of claim 205 or 206, wherein the system comprises delivering a polynucleotide sequence encoding the recombinase.
208. The system of claim 207, wherein the polynucleotide sequence is encoded in the first vector.
209. The system of any one of claims 183-208, further comprising delivering a second vector.
210. The system of claim 209, wherein the second vector comprises a template polynucleotide and a sequence that is an integration cognate with the first integration recognition site.
211. The system of claims 209 or 210, wherein the second vector is a vector selected from: an adenovirus, an AAV, a lentivirus, a HSV, an anneovirus, a retrovirus, a DoggyboneTM DNA (dbDNA), a minicircle, a plasmid, a miniDNA, an exosome, a fusosome, or a nanoplasmid.
212. The system of any one of claims 204-211, wherein the template polynucleotide comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.
213. The system of any one of claims 204-212, wherein the template polynucleotide comprises a second integration recognition site.
214. The system of claim 213, wherein the second integration recognition site is a cognate pair with the first integration recognition site.

215. The system of any one of claims 204-214, wherein the template polynucleotide comprises at least a third integration recognition site.
216. The system of claim 215, wherein the template polynucleotide further comprises at least a fourth integration recognition site.
217. The system of claim 216, wherein the third integration recognition site and the fourth integration recognition site are selected from attB, attB2, attP, or attP2.
218. The system of any one of claims 204-217, wherein the vector further comprises a sub-sequence that is capable of self-circularizing to form a self-circular nucleic acid.
219. The system of claim 218, wherein the sub-sequence of the vector that is capable of self-circularizing includes the template polynucleotide, whereby upon self-circularizing the self-circular nucleic acid comprises the template polynucleotide.
220. The system of claim 218 or 219, wherein the sub-sequence is flanked by the third integration recognition site and the fourth integration recognition site.
221. The system of claim 220, wherein self-circularizing is mediated by recombination of the third integration recognition site and the fourth integration recognition site by the integrase.
222. The system of any one of claims 217-220, wherein the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.
223. The system of any one of claims 183-222, wherein
the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence;
the first atgRNA further includes a first RT template that comprises at least a portion of a first integration recognition site;
the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and
the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site.

224. The system of claim 223, wherein the first integration site is an AttB sequence, a FRT sequence, or a VOX sequence.
225. The system of any one of claims 183-224, wherein the first atgRNA, the second atgRNA or both are synthetic.
226. The system of any one of claims 192-225, wherein the integrase is selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.
227. The system of any one of claims 194-226, wherein the recombinase is FLP or Cre.
228. A cell comprising the delivery system or co-delivery system of any one of claims 120-227.
229. A pharmaceutical composition comprising the delivery system or co-delivery system of any one of claims 120-227.
230. A method of treating a patient in need thereof, the method comprising administering an effective amount of the system of any one of claims 120-227, the cell of claim 228, or the pharmaceutical composition of claim 229.
231. A method of treating a patient in need thereof, the method comprising:
Administering:
(a) an effective amount of the LNP, the first vector, or the second vector of any one of claims 120-227 as a first dose; and
(b) an effective amount of the LNP, the first vector, or the second vector of any one of claims 120-227 as a second dose.
232. The method of claim 231, wherein the first dose and the second dose are separately administered by multiple administrations.
233. The method of claim 232, wherein the first dose and the second dose are administered at least 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, or 7 days apart.
234. The method of claim 231, wherein the first dose and the second dose are administered at least 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 weeks apart.