

SINGLE CONSTRUCT PLATFORM FOR SIMULTANEOUS DELIVERY OF GENE EDITING MACHINERY AND NUCLEIC ACID CARGO

1. CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of and priority to U.S. Provisional Application No. 63/274,483, filed on November 1, 2021; U.S. Provisional Application No. 63/282,055, filed on November 22, 2021; U.S. Provisional Application No. 63/298,941, filed on January 12, 2022; U.S. Provisional Application No. 63/318,344, filed on March 9, 2022; and U.S. Provisional Application No. 63/352,897, filed on June 16, 2022, each of which is hereby incorporated by reference in its entirety.

2. SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing with 559 sequences, which has been submitted electronically in XML format and is hereby incorporated herein by reference in its entirety. Said XML copy, created on October 31, 2022, is named 50408WO_CRF_sequencelisting.xml, and is 789,348 bytes in size.

3. BACKGROUND OF THE INVENTION

[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 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 OF THE INVENTION

[0005] A single nucleic acid construct is described herein that allows for incorporation of any template into any DNA locus using DNA delivery of a single component DNA. Additionally, a physical portion of the nucleic acid construct is capable of self-circularizing, forming a circular construct that contains a DNA template. Further, the nucleic acid construct

can be packaged and delivered in any viral or non-viral delivery vector including a recombinant adenovirus, helper dependent adenovirus, AAV, HSV, anellovirus, retrovirus, lentivirus, Doggybone™ DNA (dbDNA™), minicircle, plasmid, miniDNA, LNP, or nanoplasmid. Delivery of the nucleic acid construct can also be by fusosome or exosome, (*See, e.g.*, WO2019222403 which is incorporated by reference herein). Delivery of nucleic acid construct can also be by VesiCas (*See, e.g.*, US20210261957A1 which is incorporated by reference herein).

[0006] The present disclosure provides nucleic acid compositions, methods, and an overall 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 which is incorporated herein by reference), transposon-mediated gene editing, or other suitable gene editing or gene incorporation technology packaged into a single nucleic acid construct, (described in some instances as an “installer”). Non-limiting examples of PASTE include those as described in U.S. Patent Publication No. 2022/0154224, which is herein incorporated by reference in its entirety. Described herein are “installer” nucleic constructs that encode for a prime editor system or a gene writer protein, one or more attachment site-containing guide RNA (atgRNA), optionally a nickase guide RNA (ngRNA), an integrase, a nucleic acid cargo, and optionally a recombinase. The integrase may be directly linked, for example by a peptide linker, to the prime editor fusion or gene writer protein. The nucleic acid construct described herein can be used to introduce, delete, or delete and introduce large pieces of DNA (as well as small pieces of DNA) to any genomic site in any organism. The technology described herein can be used broadly in therapeutic, diagnostic, agricultural, research, and for the general inclusion of genetic- and protein-based circuits.

[0007] In one aspect, this disclosure features a nucleic acid construct comprising: a nucleotide sequence encoding a prime editor system; a nucleotide sequence encoding at least a first attachment site-containing guide RNA (atgRNA); a nucleotide sequence encoding at least a first integrase; a nucleic acid cargo; optionally, a nucleotide sequence encoding a nickase guide RNA (ngRNA); and optionally a nucleotide sequence encoding a recombinase.

[0008] In some embodiments, the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.

[0009] 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 gene editor system comprises a fusion protein comprising the nickase and the reverse transcriptase.

[0010] In some embodiments, the first integrase that is encoded by a nucleotide sequence in the nucleic acid construct is fused to the prime editor system, the nickase, or the reverse transcriptase by a linker.

[0011] In some embodiments, the first atgRNA comprises a domain that is capable of guiding the prime editor system to a target sequence; and a reverse transcriptase (RT) template that comprises at least a portion of a first integration recognition site.

[0012] In some embodiments, the RT template comprises the entirety of the first integration recognition site.

[0013] In some embodiments, upon introducing the nucleic acid construct into a cell, the first atgRNA incorporates the first integrase recognition site into the cell's genome at the target sequence.

[0014] In some embodiments, the nucleic acid construct further comprises a second atgRNA.

[0015] 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; 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.

[0016] In some embodiments, upon introducing the nucleic acid construct into a cell, the first pair of atgRNAs incorporate the first integrase recognition site into the cell's genome at the target sequence.

[0017] In some embodiments, the nucleic acid construct further comprises a second integrase recognition site.

[0018] In some embodiments, the second integrase recognition site and the first integrase recognition site are a first cognate pair.

[0019] In some embodiments, nucleic acid construct further comprises a third integrase recognition site.

[0020] In some embodiments, the nucleic acid construct further comprises a fourth integrase recognition site.

[0021] In some embodiments, the third integrase recognition site and the fourth integrase recognition site are a second cognate pair.

[0022] In some embodiments, the second cognate pair has a faster integration rate than the first cognate pair, whereby in the presence of the first integrase the second cognate pair recombines prior to recombination of the first cognate pair.

[0023] In some embodiments, the nucleic acid construct further comprises a nucleotide sequence encoding a second integrase.

[0024] In some embodiments, the first integrase, the second integrase, or both, are selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.

[0025] In some embodiments, the first integrase and the second integrase recognize different integration recognition sites.

[0026] In some embodiments, the nucleic acid construct further comprises at least a first recombinase recognition site.

[0027] In some embodiments, the nucleic acid construct further comprises a second recombinase recognition site.

[0028] In some embodiments, the recombinase is FLP or Cre.

[0029] In some embodiments, the nucleic acid cargo comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.

[0030] In some embodiments, the nucleic acid construct further comprises a sub-sequence of the nucleic acid construct that is capable of self-circularizing to form a self-circular nucleic acid.

[0031] In some embodiments, the sub-sequence of the nucleic acid construct that is capable of self-circularizing includes the nucleic acid cargo, whereby upon self-circularizing the self-circular nucleic acid comprises the nucleic acid cargo.

[0032] In some embodiments, the sub-sequence is flanked by the third integrase recognition site and the fourth integrase recognition site.

[0033] In some embodiments, the sub-sequence includes the second integrase recognition site.

[0034] In some embodiments, self-circularizing is mediated by recombination of the third integrase recognition site and the fourth integration recognition site by the first integrase.

[0035] In some embodiments, the sub-sequence is flanked by the first recombinase recognition site and the second recombinase recognition site.

[0036] In some embodiments, self-circularizing is mediated by recombination of the first recombinase recognition site and a second recombinase recognition site by the recombinase.

[0037] In some embodiments, the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.

[0038] In some embodiments, upon introducing the nucleic acid construct into a cell and after self-circularizing to form the self-circular nucleic acid, the self-circular nucleic acid comprising the second integrase recognition site is capable of being integrated into the cell's genome at the target sequence that contains the first integrase recognition site.

[0039] In some embodiments, self-circularization to form the self-circular nucleic acid is effected by the first integrase and integration of the self-circular nucleic acid is effected by the second integrase.

[0040] In some embodiments, the nucleic acid construct further comprises a 5' inverted terminal repeat (ITR).

[0041] In some embodiments, the nucleic acid construct further comprises a 3' inverted terminal repeat (ITR).

[0042] In another aspect, this disclosure features a vector comprising any of the nucleic acid constructs described herein.

[0043] In some embodiments, the vector is recombinant adenovirus, helper dependent adenovirus, AAV, lentivirus, HSV, anneovirus, retrovirus, Doggybone™ DNA (dbDNA™), minicircle, plasmid, miniDNA, or nanoplasmid.

[0044] In another aspect, this disclosure features a pharmaceutical composition comprising any of the nucleic acid constructs described herein or any of the vectors described herein.

[0045] In another aspect, this disclosure features a method comprising administering an effective amount of any of the pharmaceutical compositions described herein to a patient in need thereof.

5. BRIEF DESCRIPTION OF THE DRAWINGS

[0046] 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:

[0047] **FIG. 1** illustrates a single construct that contains a prime editor fusion protein or gene writer protein, the attachment site-containing guide RNA (atgRNA), a nickase guide RNA (ngRNA), an integrase, a recombinase, recombination target sites, integration target site, a DNA of interest, and flanking ITRs. Recombinase expression leads to self-circularization of a sub-sequence of the single nucleic acid construct. DNA of interest contained within the self-circularized nucleic acid is capable of being integrated into a genomic locus of interest via an integrase.

[0048] **FIG. 2** illustrates a single construct that contains a prime editor fusion protein or gene writer protein, the attachment site-containing guide RNA (atgRNA), a nickase guide RNA (ngRNA), an integrase, integration target sites, a DNA of interest, and flanking ITRs. Integrase expression leads to self-circularization of a subsequence of the single nucleic acid construct. Optionally, the integrase may be directly linked or fused to the prime editor protein or Gene Writer and expression driven from a single promoter. Self-circularization occurs at an integrase recognition target sequence (attB2/attP2). Additionally, a DNA of interest contained within the self-circularized nucleic acid is capable of being integrated into a genomic locus of interest via

the integrase at an orthogonal integration target site (i.e., cognate pairs (e.g., attP1/attB1)). Initial self-circularization, prior to genomic integration, is achieved via the use of att integrase recognition target sites (i.e., attB2/attP2 and attP1/attB1) that are cognate pairs. The orthogonal integrase sites display an integrase-mediated recombination rate difference to allow for template/cargo circularization prior to genomic integration.

[0049] FIGS. 3A-3E show multiplex and orthogonal gene insertion with PASTE. FIG. 3A shows a schematic of AttP mutations tested for improving integration efficiency (SEQ ID NOS 394 and 540-542, respectively, in order of appearance). FIG. 3B shows integration efficiencies of wildtype and mutant AttP sites across a panel of AttB lengths. FIG. 3C shows a 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. 3D shows orthogonality of top 4 AttB/AttP dinucleotide pairs evaluated for GFP integration with PASTE at the ACTB locus. FIG. 3E 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.

[0050] FIGS. 4A-4E show additional characterization of AttP mutants for improved editing and multiplexing. FIG. 4A shows AttP single mutants are characterized for PASTE EGFP integration at the ACTB locus. FIG. 4B shows characterization of integration of a 5 kb payload at the ACTB locus with all 16 possible dinucleotides for AttB/AttP pairs between the atgRNA and minicircle. FIG. 4C shows a schematic of the pooled AttB/AttP dinucleotide orthogonality assay. Each AttB dinucleotide sequence is cotransfected with a barcoded pool of all 16 AttP dinucleotide sequences and BxbINT, and relative integration efficiencies are determined by next generation sequencing of barcodes. All 16 AttB dinucleotides are profiled in an arrayed format with AttP pools. FIG. 4D illustrates relative insertion preferences for all possible AttB/AttP dinucleotide pairs determined by the pooled orthogonality assay. FIG. 4E shows orthogonality of BxbINT dinucleotides as measured by a pooled reporter assay. Each web logo motif shows the relative integration of different AttP sequences in a pool at a denoted AttB sequence with the listed dinucleotide.

[0051] FIG. 5 illustrates a schematic of single atgRNA and dual atgRNA approaches for beacon placement.

[0052] FIG. 6 illustrates the six different C-terminus to N-terminus arrangements (C-to-N) of exemplary nucleic acid programmable DNA binding proteins (napDNAbp), the RT, and the integrase is be fused or linked.

[0053] FIG. 7 illustrates the extrachromosomal circular DND (EccDNA) sensor assay to detect template circularization, beacon placement, and gene insertion. AttP (GT) for genome insertion. AttB'-AG and AttP'-AG at both ends for circularization in presence of Bxb1. EF1 α promoter will drive NanoLuc and GFP expression. Screen for efficient di-nucleotides and configuration. Based on FG- and HD- AdV vector, tested in plasmid and virus format
Abbreviations: Nanoluc = Nanoluc luciferase; GFP = green fluorescent protein; EF1 α = elongation factor 1 alpha promoter; ori = origin of replication; and AmpR = gene encoding an Ampicillin resistance protein.

[0054] FIG. 8 illustrates transfection screening conditions for circularization detection and *ACTB* beacon placement and gene insertion.

[0055] FIG. 9 illustrates EccDNA ddPCR analysis.

[0056] FIG. 10 illustrates EccDNA ddPCR analysis with PE2, atgRNA, ngRNA components co-transfected.

[0057] FIG. 11 illustrates *ACTB* beacon placement analysis.

[0058] FIG. 12 illustrates EccDNA *ACTB* gene insertion analysis at a placed beacon.

[0059] FIG. 13 illustrates transfection screening conditions for circularization detection and *LMNB* beacon placement and gene insertion.

[0060] FIG. 14 illustrates in cell EccDNA circularization detection by GFP detection.

[0061] FIG. 15 illustrates EccDNA ddPCR analysis.

[0062] FIG. 16 illustrates EccDNA *LMNB* beacon placement analysis.

[0063] FIG. 17 illustrates *LMNB* gene insertion analysis at a placed beacon.

[0064] FIG. 18 illustrates a single construct that contains a prime editor fusion protein, dual attachment site-containing guide RNA (atgRNAs) (i.e., atgF and atgR), a tet-inducible integrase, an integration target site, a DNA of interest, and flanking ITRs. Abbreviations: ITR

= inverted terminal repeat; Ad5 ψ = Adenovirus 5 packaging domain; atgR = atgRNA reverse; U6 = U6 promoter; atgF = atgRNA forward; U6 = U6 promoter; PE2 = prime editing complex PE2 (as described herein); tet-off = tetracycline off promoter; EF1 α = elongation factor 1 alpha promoter; mScarlet = a red fluorescent protein; Nanoluc = Nanoluc luciferase; GFP = green fluorescent protein; ori = origin of replication; and AmpR = gene encoding an Ampicillin resistance protein.

[0065] FIGs. 19A-19J show brightfield (FIG. 19A, 19C, 19E, 19G, and 19I) and RFP (FIG. 19B, 19D, 19F, 19H, and 19J) on day 2 following transfection with the single nucleic acid construct depicted in FIG. 18.

[0066] FIGs. 20A-20B illustrates beacon placement (BP) at the *NolcI* locus. FIG. 20A shows raw data from a ddPCR assay at the *NolcI* locus. FIG. 20B shows summary of the data in FIG. 20A. Abbreviation: AIO – all-in-one (also referred to herein as the single nucleic acid construct).

[0067] FIGs. 21A-21B illustrates programmable gene insertion (PGI) at the *NolcI* locus. FIG. 21A shows raw data from a ddPCR assay at the *NolcI* locus. FIG. 21B shows summary of the data in FIG. 21A. Abbreviation: AIO – all-in-one (also referred to herein as the single nucleic acid construct).

[0068] FIG. 22 shows PGI conversion rate ($= \text{PGI \%} / (\text{PGI\%} + \text{BP\%})$) for the data in FIGs. 20A-20B and FIGs. 21A-21B.

[0069] FIGs. 23A-23B show next generation sequence data confirming beacon placement and PGI. FIG. 23A shows next generation sequencing data for beacon placement. FIG. 23B shows next generation sequencing data for PGI.

[0070] FIG. 24 shows next generation sequence data from FIG. 22A and FIG. 22B as PGI conversion rate ($= \text{PGI \%} / (\text{PGI\%} + \text{BP\%})$).

[0071] FIGs. 25A-25L show brightfield (FIG. 25A-25D), RFP (FIG. 25E-25H), and GFP (FIG. 25I-25L) on day 2 following transection with the single nucleic acid construct depicted in FIG. 18 or a four plasmid system.

[0072] FIGs. 26A-26B illustrates beacon placement (BP) at the *human factor IX* (“*hF9*”) locus. FIG. 26A shows raw data from a ddPCR assay at the *hF9* locus. FIG. 26B shows

summary of the data in **FIG. 26A**. Abbreviation: AIO – all-in-one (also referred to herein as the single nucleic acid construct).

[0073] **FIGs. 27A-27B** illustrates programmable gene insertion (PGI) at the *hFP* locus. **FIG. 27A** shows raw data from a ddPCR assay at the *hF9* locus. **FIG. 27B** shows summary of the data in **FIG. 27A**. Abbreviation: AIO – all-in-one (also referred to herein as the single nucleic acid construct).

6. DETAILED DESCRIPTION OF THE INVENTION

6.1. Gene Editors

[0074] 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.

[0075] “**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. 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 or linked to one or more proteins or protein fragment motifs. Gene editors may be fused, linked, complexed, operate in cis or trans 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.

[0076] “**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. Such an enzyme can be a catalytically-impaired Cas9 endonuclease (a nickase). Such an enzyme can be a Casl2a/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. Described herein, are attachment site-containing guide RNA (atgRNA) that both specify the target and encode 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 pegRNA (or atgRNA), whereby a nick or single stranded cut occurs. The reverse transcriptase domain then uses the pegRNA (or atgRNA) 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).

[0077] 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 single nucleic acid construct design 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.

[0078] 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 pegRNA or atgRNAs 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 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 used in PrimeDel, TwinPE (WO2021226558 incorporated by reference herein), 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 single nucleic acid construct described herein. The integrase may be directly linked, for example by a peptide linker, to the prime editor fusion or gene writer protein.

[0079] 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

[0080] 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]	atgactcactatcaggccttgctttggacacggaccgggtccagttcggaccggtggtagccctgaac ccggctacgctgctcccactgcctgaggaagggtgcaacacaactgccttgatGGGACAGGT GGCGGTGGTGTCAACCGTCAAGTTCAAGTACAAGGGTGAGGAAC TTGAAGTTGATATTAGCAAAATCAAGAAGGTTTGGCGCGTTGG TAAAATGATATCTTTTACTTATGACGACAACGGCAAGACAGGT AGAGGGGCGAGTGTCTGAGAAAGACGCCCCCAAGGAGCTGTTGC AAATGTTGGAAGTCTGGGAAAAAGTctggcgggtcaaaaagaaccgccga cggcagcgaattcgagcccaagaagaagaggaaagtc	4

[0081] PE3, PE3b, PE4, PE5, and/or PEmax, which a skilled person can incorporate into the gene editor (and express from a single nucleic acid construct, e.g., any of the single nucleic acid constructs 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).

[0082] The skilled person can readily incorporate into a gene editor single nucleic acid construct (“installer”) 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 a single nucleic acid construct ("installer") described herein.

[0083] Prior to RT-mediated edit incorporation, the prime editor protein (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.

[0084] 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.

[0085] 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 single nucleic acid constructs described herein.

[0086] 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	GGGSGGGSGGGG S	14
C - GGGS	GGAGGTGGCGGGAGC	7	GGGS	15
D - PAPAP	CCCGCACCAGCGCCT	8	PAPAP	16
E - (EAAK) 3	GAGGCAGCTGCCAAGGAAGCC GCTGCCAAGGAGGCGGCCGCA AAG	9	EAAAKEAAKEAAA K	17
F - XTEN	AGTGGGAGCGAGACCCCTGGG ACTAGCGAGTCAGCTACACCCG AAAGC	10	SGSETPGTSESATPES	18
G - (GGS)6	GGGGGGTCAGGTGGATCCGGC GGAAGTGGCGGATCCGGTGA TCTGGCGGCAGT	11	GGSGSGSGSGSGG SGGS	19
H - EAAAK	GAAGCTGCTGCTAAG	12	EAAAK	20
(GGGS) 4	GGCGGCGGCGGCAGCGCGGC GGCGGCAGCGGCGGCGGC AGCGGCGGCGGCGGCAGC	543	GGGSGGGSGGGG SGGGS	551
PAS8	GGCGGCGCGAGCCCGGCGGGC GGC	544	GGASPAGG	552
PAS12	GGCGGCGCGAGCCCGGCGGCG CCGCGCCGGCGGGC	545	GGASPAAPAPAG	553
A(EAAK) 4ALEA(E AAK)4A	GCGGAAGCGGCGAAAGAAGCG GCGAAAGAAGCGGCGAAAGAA GCGGCGAAAGCGCTGGAAGCG GAAGCGGCGGCGAAAGAAGCG GCGGCGAAAGAAGCGGCGGCG AAAGAAGCGGCGGCGAAAGCG	546	AEEAKEAAKEAAKE AAKALEEAAAKEA AAKEAAKEAAK A	554
Camel	GCGCATCATAGCGAAGATCCG GGCGGCGGCGGCAGCGGCGGC GGCGGCAGCGGCGGCGGCGGC AGC	547	AHHSEDPGGGSGG GSGGGGS	555
FRF	GGCGGCGGCGGCAGCGAAGCG GCGGCGAAAGGCGGCGGCGGC AGC	548	GGGGSEAAKGGGG S	556
RFR	GAAGCGGCGGCGAAAGGCGGC GGCGGCAGCGAAGCGGCGGCG AAA	549	EAAKGGGGSEAAA K	557
Modified XTEN (mXTEN)	AGCGGCGGCGAGCAGCGGCGGC AGCAGCGGCGAGCGAAACCCCG GGCACCAGCGAAAGCGCGACC CCGAAAGCAGCGGCGGCGAGC AGCGGCGGCGAGCAGCACC	550	SGGSSGGSSGSETPG TSESATPESGSSG GSST	558

[0087] 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

[0088] The skilled person can incorporate a selected CRISPR enzyme, described below, as part of the prime editor fusion, into a single nucleic acid construct (“installer”) 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 3' 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.

[0089] *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.

[0090] *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.

[0091] *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).

[0092] *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.

[0093] *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.

[0094] 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.

[0095] *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.

[0096] *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.

[0097] 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.

[0098] *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.

[0099] 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 20-hydroxyl groups of U64 and A65 hydrogen bonded with Glu57 and His721, respectively. The C67 nucleobase forms two hydrogen bonds with Val1100.

[0100] 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.

[0101] 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.

[0102] 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, 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.

[0103] 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.

[0104] 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.

[0105] 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 GI: 757015980 WP_01092225 1.1	MDKKYSIGLD LLFDSGETAE LEESFLVEED LRLIYLALAH INASGVDAKA NFKSNFDLAE LLSDILRVNT FFDQSKNGYA KQRTFDNGSI YVGPLARGNS NLPNEKVLPK LLFKTNRKVT IKDKDFLDNE LKRRRYTGWG SLTFKEDIQK MGRHKPENIV VENTQLQNEK SIDNKVLTRS TKAERGGLSE REVKVITLKS YPKLESEFVY TLANGEIRKR QTGGFSKESI KGKSKKLKSV YSLFELENGR DNEQKQLFVE PIREQAENII SITGLYETRI	IGTNSVGWAV ATRLKRTARR KKHERHPIFG MIKFRGHFLI ILSARLSKSR DAKLQLSKDT EITKAPLSAS GYIDGGASQE PHQIHLGELH RFAWMTRKSE HSLLYEYFTV VKQLKEDYFK ENEDILEDIV RLSRKLINGI AQVSGQGDSL IEMARENQTT LYLYYLQNGR DKNRGKSDNV LDKAGFIKRQ KLVSDFRKDF GDYKVYDVRK PLIETNGETG LPKRNSDKLI KELLGITIME KRMLASAGEL QHKKHYLDEII HLFTLTNLGA DLS	ITDEYKVPSK RYTRRKNRIC NIVDEVAYHE EGDLNPDNSD RLENLIAQLP YDDDLNLLA MIKRYDEHHQ EFYKFIKPIL AILRRQEDFY ETITPWNFEE YNELTKVKYV KIECFDSVEI LTLTLFEDRE RDKQSGKTIL HEHIANLAGS QKGQKNSRER DMYVDQELDI PSEEVVKKMK LVETRQITKH QFYKVREINN MIAKSEQEIG EIVWDKGRDF ARKKDWDPKK RSSFEKNPID QKGNELALPS EQISEFSKRV PAAFKYFDTT	KFKVLGNTDR YLQEIFSNEM KLVDSTDKAD VDKLFQILVQ GEKKNGLFGN QIGDQYADLF DLTLLKALVR EKMDGTEELL PFLKDNREKI VVDKGASQAS TEGMRKPAFL SGVEDRFNAS MIEERLKTYA DFLKSDFGAN PAIKKGILQT MKRIEEGIKE NRLSDYDVDH NYWRQLLNAK VAQILDSRMN YHHAHDAYLN KATAKYFFYS ATVRKVL SMP YGGFDSPTVA FLEAKGYKEV KYVNFLYLAS ILADANLDKV IDRKRYTSTK	HSIKKNLIGA AKVDDSFHR KLVDSTDKAD TYNQLFEENP LIALSLGLTP LAAKNLSDAI QQLPEKYKEI VKLNREDLLR EKILTFRIPY FIERMTNFDK SGEQKKAI VD LGT YHDL LKI HLFDDKVMKQ RNF MQLIHDD VKVVD ELVKV LGSQILKEHP IVPQSFLKDD LITQRKFDNL TKYDENDKLI AVVGTALIKK NIMNFFKTEI QVNIVKKTEV YSVLVVAKVE KKDLIIKLPK HYEKLKGSPE LSAYNKHRDK EVL DATLIHQ	(SEQ ID NO: 21)
AJN60021.1 GI: 757015977 J7RUA5.1 WP_05301979 4.1 Staphylococcus aureus	MKRNYILGLD RGARRLKRRR SEEEFSAALL AELQLERLKK YIDLLETRRT YNADLYNALN KEILVNEEDI IAKILTIYQS NLILDELWHT KRSFIQSIKV TNERIEETIR FNYEVDHIIP YETFKKHILN YATRGLMNL HAEDALIIAN KEIFITPHQI IVNNLNGLYD KNPLYKYEE RNKVVKL SLK KKLKKISNQA YREYLENMND KKG	IGITSVGYGI RHRIQVRVKKL HLAKRRGVHN DGEVRGSINR YYEGPGEGSP DLNNLVITRD KGYRVTSTGK SEDIQEELTN NDNQIAIFNR INAIKKYGL TTGKENAKYL RSVSFDNSFN LAKGGRISK RSYFRVNNLD ADFIFKEWKK KHIDKFDYK KDNDKLKKLI TGNYLTKYSK PYRFDVYLDN EFIASFYNN KRPPRIKTI	IDYETRVID LFDYNLLTDH VNEVEEDTGN FKTSDYVKEA FGWKDIKEWY ENKLEYEYK PEFTNLKVYH LNSELTQEEI LKLVPKKVDL PNDIIIELAR IEKIKLHDMQ NKVLVKQEE TKKEYLLEER VKVKSINGGF LDKAKKVMEN YSHRVDKKPN NKSPEKLLMY KDNGPVIKKI GVYKFVTVKN LIKINGEL YR ASKTQSIKKY	AGVRLFKEAN SELSGINPYE ELSTKEQISR KQLLKVQKAY EMLMGHCTYF FQI IENVFKQ DIKDI TARKE EQISNLKGYT SQQKEIPTTL BKNSKDAQKM EGKCLYSLEA SKKGNRTPFQ TSFLRRKWK QMFEEKQAES RELINDTLYS HHDPQTYQKL KYYGNKLNH LDVIKENY VIGVNNDLN STDILGNLYE	VENNEGRRSK ARVKGLSQKL NSKALEEKYV HQLDQSFIDT PEELRSVKYA KKKPTLKQIA IIENAE LLDQ GTHNLSLKAI VDDFILSPVV INEMQKRNRQ IPLEDLNNP YLSSSDSKIS FINRNLVDTR KKERNKGYKH MPEIETE QEY TRKDDKGNTL KLIMEQYGE LDITDDYPNS EVNSKCYEEA RIEVMIDIT VKS KHPQII	(SEQ ID NO: 22)

Table 4- Cas9 orthologs						
AJN60008.1 GI: 757015964 WP_00286448 5.1 Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819	MARILAFDIG	ISSIGWAFSE	NDELKDCGVR	IFTKVENPKT	GESLALPRRL	(SEQ ID NO: 23)
	ARSARKRLAR	RKARLNHLKH	LIANEFKLNY	EDYQSFDESL	AKAYKGSLSIS	
	PYELRFRALN	ELLSKQDFAR	VILHIAKRRG	YDDIKNSDDK	EKGAILKAIK	
	QNEEKLANYQ	SVGEYLYKEY	FQKFKENSKE	FTNVRNKKES	YERCIAQSFL	
	KDELKLIFKK	QREFGFSFSK	KFEEEVLSVA	FYKRALKDFS	HLVGNCSFFT	
	DEKRAPKNSP	LAFMFVALTR	IINLLNNLKN	TEGILYTKDD	LNALLNEVLK	
	NGTLTYKQTK	KLLGLSDDYE	FKGEKGTYFI	EFKKYKEFIK	ALGEHNLSQD	
	DLNEIAKDIT	LIKDEIKLKK	ALAKYDLNQN	QIDSLSKLEF	KDHLNISFKA	
	LKLVTPMLLE	GKKYDEACNE	LNLKVAINED	KKDFLPAFNE	TYKKDEVNTP	
	VVLRAIKEYR	KVLNALLKKY	GKVHKINIEL	AREVGKNHSQ	RAKIEKEQNE	
	NYKAKKDAEL	ECEKLGKIN	SKNILKLRLF	KEQKEFCAYS	GEKIKISDLQ	
	DEKMLEIDHI	YPYSRSFDDS	YMNKVLVFTK	QNQEKLNQTP	FEAFGNDSAK	
	WQKIEVLAKN	LPTKKQKRIL	DKNYKDKEQK	NFKDRNLNDT	RYIARLVVLY	
	TKDYLDLPL	SDDENTKLND	TQKGSKVHVE	AKSGMLTSAL	RHTWGFSAKD	
	RNNHLHHAID	AVIIAYANNS	IVKAFSDFKK	EQESNSAELY	AKKISELDYK	
	NKRKFFEPFS	GFRQKVLDKI	DEIFVSKPER	KKPSGALHEE	TFRKEEEFYQ	
	SYGGKEGVK	ALELGKIRKV	NGKIVKNGDM	FRVDIFKHKK	TNKFYAVPIY	
	TMDFAKVLV	NKAVARSKKG	EIKDWILMDE	NYEFCFSLYK	DSLILIQTKD	
	MQEPFVYYN	AFTSSTVSLI	VSKHDNKFET	LSKNQKILFK	NANEKEVIAK	
	SIGIQNLKVF	EKYIVSALGE	VTKAEFRQRE	DFKK		
Streptococcus thermophilus LMD-9 AJN60026.1 GI: 757015982 WP_01168095 7.1	MSDLVLGLDI	GIGSVGVGIL	NKVTGEIHHK	NSRIFPAAQA	ENNLVRRTNR	(SEQ ID NO: 24)
	QGRRLARRKK	HRRVRLNRLF	EESGLITDFT	KISINLNPHY	LRVKGTLDEL	
	SNEELFIALK	NMVKHRGISY	LDDASDDGNS	SVGDYAIQIV	ENSKQLETKT	
	PGQIQLEERYQ	TYGQLRGDFT	VEKDGGKHRL	INVFPTSAYR	SEALRILQTO	
	QEFNPQITDE	FINRYLEILT	GKRKYHGP	NEKSRTDYGR	YRTSGETLDN	
	IFGILIGKCT	FYPDEFRAAK	ASYTAQEFNL	LNDLNNLTVP	TETKKLSKEQ	
	KNQIINYVKN	EKAMGPAKLF	KYIAKLLSCD	VADIKGYRID	KSGKAEIHTF	
	EAYRKMKTL	TLDIEQMDRE	TLDKLAIVLT	LNTEREIGQE	ALEHEFADGS	
	FSQKQVDELV	QFRKANSSIF	GKGWHNFSVK	LMMELIPELY	ETSEEQMTIL	
	TRLGKQKTTS	SSNKTKEYIDE	KLLTEEIYNP	VVAKSVRQAI	KIVNAAIKEY	
	GDFDNIVIE	ARETNEDDEK	KAIQKIQKAN	KDEKDAAMLK	AANQYNGKAE	
	LPHSVFHGHK	QLATKIRLWH	QQGERCLYTG	KTISIHDLIN	NSNQFEVDHI	
	LPLSITFDSD	LANKVLVYAT	ANQEKGQRT	YQALDSMDA	WSFRELKAFV	
	RESKTLNKK	KEYLLTEEDI	SKFDVRKKFI	ERNLVDTRYA	SRVVLNALQE	
	HFRAHKIDTK	VSVVRGQFTS	QLRRHWGIEK	TRDTYHHHAV	DALIIAASSQ	
	LNLWKKQKNT	LVSYSQDQLL	DIETGELISD	DEYKESVFKA	PYQHFVDTLK	
	SKEFEDSILF	SYQVDSKFNR	KISDATIYAT	RQAKVGKDKA	DETYVLGKIK	
	DIYTQDGYDA	FMKIYKKDKS	KFLMYRHDPQ	TFEKVIEPIL	ENYPNKQINE	
	KGKEVPCNPF	LKYKEEHGYI	RKYSKKGNGP	EIKSLKYYDS	KLGNHIDITP	
	KDSNNKVVLQ	SVSPWRADVY	FNKTTGKYEI	LGLKYADLQF	EKGTGTYKIS	
	QEKYNDIKKK	EGVDSSEFK	FTLYKNDLLL	VKDTETKEQQ	LFRFLSRTMP	
	KQKHVELKLP	YDKQKFEGGE	ALIKVLGNVA	NSGQCKKGLG	KSNISIIYKVR	
	TDVLGNQHII	KNEGDKPKLD	F			
Parvibaculum lavamentivora ns DS-1 AJN60020.1 GI: 757015976 WP_011995013. 1	MERIFGFDIG	TTSIGFSVID	YSSTQSAGNI	QRLGVRIFFE	ARDPDGTPLN	(SEQ ID NO: 25)
	QRRQKRMMR	RQLRRRRIRR	KALNETLHEA	GFLPAYGSAD	WPVVMADPEY	
	ELRRRGLLEEG	LSAYEFGRAI	YHLAQHRHFK	GRELEESDTP	DPDVEDDEKA	
	ANERAAATLKA	LKNEQTTLGA	WLARRPPSDR	KRGIAHRNV	VAEEFERLWE	
	VQSKFHPALK	SEEMRARISD	TIFAQRPVFW	RKNTLGEICRF	MPGEPLCPKG	
	SWLSQQRRL	EKLNNLAIAG	GNARPLDAEE	RDAILSKLQ	QASMSWPGVR	
	SALKALYKQR	GEPGAEKSLK	FNLELGESK	LLGNALEAKL	ADMFGPDWPA	
	HPRKQEIIRHA	VHERLWAADY	GETPDKKRV	ILSEKDRKAH	REAAANSFVA	
	DFGITGEQAA	QLQALKLPTG	WEPYSIPALN	LFLAELEKGE	RFGALVNGPD	
	WEGWRRTNFP	HRNQPTGEIL	DKLPSPASKE	ERERISQLRN	PTVVRTQNEL	
	RKVNNLIGL	YGKPDRIE	VGRDVGKSKR	EREBIQSGIR	RNEKQRKKAT	
	EDLIKNGIAN	PSRDDVEKWI	LWKEGQERCP	YTGDIQGFNA	LFREGRYEVE	
	HIWPRSRSFD	NSPRNKTLCR	KDVNIEKGNR	MPFEAFGHDE	DRWSAIQIRL	

Table 4- Cas9 orthologs						
	QGMVSAKGGT LWPDMGPEAP DALTVACTHP IVVSHRVRKK ELDEIRDPRI QQNLMAQTG PIVQTRADG TDADHSTTTR	GMSPGKVKRF VKVEAVTGQV GMTNKLRYW VSGPLHKETT KEIVAAHVAG NGYADLGSNH ASFVMSLAAG PMPNPILKDD	LAKTMPEDFA TAQLRKLWTL QLRDDPRAEK YGDGTGDIKT RGGDPKKAFP HIAIYRLPDG EAIMIPEGSK AKKVSIDPIG	ARQLNDTRYA NNILADDGEK PALTPPWDTI KSGTYRQFVT PYPCVSPGGP KADFEIVSLF KGIWIVQGVW RVRSND	AKQILAQLKR TRADHRHAI RADA EKAVSE RKKIESLSKG EIRKVRITSK DASRRLAQRN ASGQVVLERD	
Corynebacterium diphtheriae NCTC 13129 AJN60012.1 GI: 757015968 WP_01093396 8.1	MKYHVGIDVG VTRLASSGIA WKVRAELAAS PSDAFKAIRE LQQSDYAREI QPGKNRALKA NLTPKKEPEW VNSRIAPLVD DDD VHAKLDS PSWTPPTPRI GFVTEKRARE QRQNCQCAYC GNTPF AIWAK FQRATMDEEI ARRASGISGK LKQSQAHQOE VVMSNVRLR TREP GFDPKE ELGSSFHAR SMRQAEKKLR TIRRWVRDGF KLFS DGNVTV	TFSVGLAAIE RRTRRLYRRK YIADEKERGE EIKRASGQPV QEICRMQEIG SDAFQRYRIA VTIAEILGID WWKTASALEQ LHLPVGRAAY GEPVGNPAVD MDGDMRRRAA GSPITFSNSE NTSIEGVSVK DARSMESVAW LKFFDGVGKS APQWREFTGK LGNGSAHKET GLPANPERHI VYKITS GKKP DALATGNAEY FSPSKLRLRP VRRDSLGRVR	VDDAGMPIKT RRRLQQLDKF KLSVALRHIA PETATVGQMV QELYRKIIDV ALIGNLRVRV RGQLIGTATM HAMVKALSNA SEDTLVRLTR RVLKTVSRWL RNAKLFQEMQ MDHIVPRAGQ EAVERTRHVW MANELRSRVA RLDRRHHAID DAEHRAAWRV IGKLSKVKLS RVNGTHVYAG AFAMLRVYTI LGWLVDDEL LQMSKEGIKK LESTAHL PVT	LSLVSHIHDS IQRQGWPIE RHRGWRNPYA TLCELGTLKL VFAAESPKGS DGEKRILSVE TDDGERAGAR EVD DFDSP EG RMLS DGV DLY ESATKTWGAP EKLNVQ GKPS GSTNTREN LV TDTGMRSTDF QHFASHGTTV AAVIAFTSDY WCQKMEKLSA SQLSVSDIDK DNIGLF PVSA DLLPYRNQDL VVDTSKIATD ESAPELSKII WKVQ	GLDPDEIKSA LEDYSDPLYP KVSSLYLPDG RGE GGVLSAR ASSRVGKDPL EKNLVFDHLV PPTHDTNRSI AKVQAF FADL TARLQEF GIE ERVII EHVRE RADLWR YQSV AVCHRCNQSK KKFTKAVVER RVYRGS LTAE VAETLAVRSN LLTEDLRDDR ASSEALWCAL GSIALRGGYA FSVELKPQTM QVKAVEAELG DRPGWL PAVN	(SEQ ID NO: 26)
Streptococcus pasteurianus WP_01385204 8.1	MTNGKILGLD RGSRRLNRRK NEELFAALRT QIQLERLEKY DYNKKITAEF FGILIGKCNF ESLVEFAKNT PYRKLKFNLE TEEQISELIK RLEKFKVNKK GDFDKIVIE LPDEVFHGK LPLSLSFDD LKQKGLGKKK ALRELKDKT LKLWEKQDNP TISSKGFEDE KIKDIYSQNG KSEDKNDVK DITPEESRNK YHISQEKYDA RTMPNVNHVY YKVRTDVLGN	IGIASVGVGI KHRVKVRDL ISKRRGISYL GQLRGNFTVY IDDYVEILTQ YPDEYRASKA ATLGP AKLLK SINIDDL SRE VRKSQSTAFN SSKNTKTIDE PRDKNADDEK QLETKIRLWY LANKVLVYAW RDYLLT TENI VSVVRGQFTS MFVDYGNQV ILFSYQVDSK FDTFIKKYNK CNPFE EYRRE VILQSINPWR IKEKEGIGKK ELKPYDKEKF KYFVKKKGDK	TEAKTGKVVH FEKYGIVTDF DDAEDDSTGS DENG EAHR LI LRKYYHGP GN SYTAQEYNFL EIAKILDCKV VIDKLADILT KGWHSFS AKL KEVTDEIYNP KFIDKR NKEN QQGERCLYSG TNQEKGQKTP DKIEVKKKFI QLRRKWKIDK VDKQTGEILS YNRKVSDATI DKTQFLMYQK NGLICKYSKK ADVYFNPETL SEFKFTLYRN DNVQELVEAL PKLDFKNNK	ANSRLFSAAN RNLNLNPYEL TDYAKSIDEN NVFSTSDY EK EKSRTDYGRF NDLNNLKVST DEIKGYREDD LNTEREGIED MNELIPELYA VVAKSVRQTI KKEKDDALKR KPISIQELVH YQVIDSMDAA ERNLVDTRYA SRETYHHHAV VSDDEYKELV YSTRKAKIGK DSL TWENVIE GKGTP IKS LK KYBELMGLKYS DLILIKDIAS GEADKVGRCI KGLNKP NISI	AENNAERRGF RVKGLTEQLK RRL LKNKTPG EARKILETQA RTDGTTL ENI ETGKLSTEQK KGKPD LHTFE AIKRNL PNQF TSDEQMTILT KIINAAVKKY AAAYLNSSDK NSNNFEIDHI WSFREM KDYV SRVVLNSLQS DALIIAASSQ FQPPYQGFVN DKKEEYVLG VILRDYPTTK YYDKKLGNCI DLSFEKGTGN GEQEYRFLS KGLNKP NISI	(SEQ ID NO: 27)

Table 4- Cas9 orthologs						
Neisseria cinerea ATCC 14685 AJN60019.1 GI: 757015975 WP_00367641 0.1	MAAFKPNPMN VPKTGDSLAA GLIKSLPNTF ADKELGALLK HTFNRKDLQA VQKMLGHCTF ERATLMDEPY KAYHAISRAL DRVQPEILEA DHYGKKNTEE IHETAREVVG KDILKLRLYE NNKVLALGSE RILLQKFDED QITNLLRGFW NAFDGKTIDK DTPEKLRTLL KRLDEGISVL KAFAPFPYKY DVFEKGKKYY LYANDLIKLT GVKTALSFOK	YILGLDIGIA ARRLARSVRR WQLRAAALDR GVADNTHALQ ELNLLFEKQK EPTPEKAAKN RKSCLTYAQA EKEGLKDKKS LLKHISFDKF KIYLPPIPAD KSFKDRKEIE QQHGKCLYSG NQNKGNQTPY GFKERNLNDT GLRKVRAEND ETGEVLHQKA AEKLSSRPEA RVPLTQLKLK DKAGNRTQQV LVPIYSWQVA AKKNEFLGYF YQIDELGKEI	SVGWAIVEID LTRRRRAHRL KLTPLEWSAV TGDFRTPAEL EFGNPHVSDG TYTAERFVWL RKLLDLDDTA PLNLSPELQD VQISLKALRR EIRNPVVLRA KRQENRKDR KEINLGRLE EYFNGKDNSR RYINRFLCQF RHHALDAVVV HFPQPWEFFA VHKYVTPLEI DLEKMVNRR KAVRVEQVQK KGILPDRADV VSLNRATGAI RPCRLKKRPP	EEENPIRLID RARRLLKREG LLHLIKHRGY ALNKFEKESG LKEGIETLLM TKLNNLRILE FFKGLRYGKD EIGTAFSLFK IVPLMEQGNR LSQARKVING EKSAAKFREY KGYVEIDHAL EWQEFKARVE QVVMIRVFGK SRAPNRKMSG EPKLYEALKA TGWVWHNHNG QKDEEDWTV DIRTHDSTD VR	LGVRVFERAE VLQAADFDE LSQRKNEGET HIRNQRGDYS TQRPALSGDA QGSEPLTDT NAEASTLMEM TDEDITGRLK YDEACTEIIYG VVRRYGSPAR FPNFVGEPKS PFSRTWDDSF TSRFPKSKKQ GKRRVFASNG ITRFVRYKEM PDGKPEFEEA QGHMETVKSA RLEAHKDDPA IADNATVVRV MDDSEFEKFFV KGKNGIFQSV	(SEQ ID NO: 28)
AJN60009.1 GI: 757015965 St1Cas9+SpCas9	MSDLVLGLDI QGRRLARRKK SNEELFIALK PGQIQLERYQ QEFNPQITDE IFGILIGKCT KNQIINYVKN EAYRKMKTLE FSQKQVDELV TRLGKQKTTS GDFDNIVIE TQLQNEKLYL NKVLTRSDKN ERGGLSELDK KVITLKSCLV LESEFVYGDY NGEIRKRPLI GFSKESILPK SKKLKSVKEL FELENGRKRK QKQLFVEQHK EQAENIHLF GLYETRIDL	GIGSVGVGIL HRRVRLNRLF NMVKHGRISY TYGQLRGDFT FINRYLEILT FYPDEFRAAK EKAMGPAKLF TLDIEQMDRE QFRKANSSIF SSNKTKEYIDE ARENQTTQKG YYLQNGRDMY RGKSDNPVSE AGFIKRQLVE SDFRKDFQFY KVYDVRKMIA ETNGETGEIV RNSDKLIARK LGITIMERS LASAGELQKG HYLDEIIEQI TLTNLGAPAA QLGGD	NKVTGEIIHK EESGLITDFT LDDASDDGNS VEKDGGKHRL GKRKYHGGPG ASYTAQEFNL KYIAKLLSCD TLDKLAYVLT KQNSRERMKR VDQELDINRL EVVKMKMNYW TRQITKHVAQ KVREINNYHH KSEQEIGKAT WDKGRDFATV KDWDPKKYGG FEKNPIDFLE NELALPSKYV SEFSKRVILA FKYFDTTIDR	NSRIFPAAQA KISINLNPYQ SVGDYAIQIVK INVFPSTAYR NEKSRDYG R LNDLNNLTVP VADIKGYRID LNTEREGIQE IEEGIKELGS SDYDVDHIVP RQLLNAKLIT ILDSRMNTKY AHDAYLNAV AKYFFYSNIM RKVLSMPQVN FDSPTVAYS LVVAKVEKGK AKGYKEVKKD NFLYLASHYE DANLDKVL SA KRYTSTKEVL	ENNLVRRRTNR LRVKGLTDEL ENSKQLETKT SEALRILQTQ YRTSGETLDN TETKKLSKEQ KSGEAEIHTF ALEHEFADGS ETSEEQMTIL KIVNAAIKEY QILKEHPVEN QSFLKDDSID QKRFNLTKA DENDKLIREV GTALIKKYPK NFFKTEITLA IVKKT EVQ TG LVVAKVEKGK LI IKLPKYSL KLKGS PEDNE YKHKRDKPIR DATLIHQ SIT	(SEQ ID NO: 29)
Campylobacter lari Cas9 BAK69486.1	MRILGFDIGI RSSRRRLKRR YELRYKALTQ KNNALKLENY LEKELKLILE EEEKCRACKNS DKGSITYKKF GVHSLSRQEL YINLSFKALG FAHELSPV KIEKEQKENQ	NSIGWAFVEN KARLIAIKRI NLETCDLARV QSVGEYFYKE YSAWEFVALT RSCINLHESI DQISTHITLI MILPLMREGK NRAISEYRKV AVNAWALKEC	DELKDCGVRI LAKELKLNYK ILHIAKHRGY FFQKYKKN TK EDFINEIL KV KIINEIKS LE SFKSLKYD KE KDNV KLKT VL RYDEACE IAN LNALLK KYGK ENIGLK KASAK	FTKAENPKNK DYVAADGELP MKNKNEK KSN DN NFIKIRN TKD AFFQRPL KDF KISGEIV PTQ NAENAKL IDF EKYNLS NEQI LKP KT VDEKK DFLPA FCDSI DVGLS KKARE QKEICI YSGN	ESLALPRNA KAYEGSLASV AKKGKILSAL SHLVGACTFF TINEVNLIL RKLVEFKKAL NNLLEIEFND DFLPAFCDSI DVGLSKKARE QKEICIYSGN	(SEQ ID NO: 30)

Table 4- Cas9 orthologs						
	KISIEHLKDE AFGKNIEKWS IATLIAKYTK TWGFDKKDRN ELTSDNYKHQ HSENKIIDKC KFYAIPYAM NDLILLQKKN NAKEGSVKVE GLR	KALEVDHIYP KIQTALQNL EYLNFLLLSE NHLHHALDAI VKFFPEPKSF SYNSKEGLQI DFALGILPNK MQEPEFAYYN SLGIQNLKVF	YSRSFDDSF YKKKNKILDE NENANLKSSE IVAYSTNSII REKILSKIDE ALSCGRVRKI IVITGKDKNN DFSISTSSIC EKYIITPLGD	NKVLVFTKEN NFKDKQQEDF KGSKIHVQTI KAFSDFRKNQ IFVSKPPRKR GTKYVENDTI NPKQWQTIDE VEKHDNKFEN KIKADFQPRE	QEKLNKTPFE ISRNLDNTRY SGMLTSVLRH ELLKARFYAK ARRALHKDTF VRVDIFKKQN SYEFCFSLYK LTSNQKLLFS NISLKTSSKY	
AJN60010.1 GI: 757015966 SpCas9+St1Ca s9	MDKKYSIGLD LLFDSGETAE LEESFLVEED LRLIYLALAH INASGVDAKA NFKSNFDLAE LLSDILRVNT FFDQSKNGYA KQRTFDNGSI YVGPLARGNS NLPNEKVLPA LLFKTNRKVT IKDKDFLDNE LKRRTYTGWG SLTFKEDIQK MGRHKPENIV KAELPHSVFH DHILPLSITF AFVRESKTLA LQEHFRAHAI SSQLNLWKKQ TLKSKEFEDS KIKDIYTQDG INEKGKEVPC ITPKDSNNKV KISQEKYNDI TMPKQKHVE KVRTDVLGNQ	IGTNSVGWAV ATRLKRTARR KKHERHPIFG MIKFRGHFLI ILSARLSKSR DAKLQLSKDT EITKAPLSAS GYIDGGASQE PHQIHLGELH RFAWMTRKSE HSLLEYEFTV VKQLKEDYFK ENEDILEDIV RLSRKLINGI AQVSGQGDSL IEMARETNED GHKQLATKIR DDSLANKVLV NKKKEYLLTE DTKVSVVRGQ KNTLVSYSED ILFSYQVDSK YDAFMKIYK NPFLKYKEEH VLQSVSPWRA KKKEGVSDS LKPYDKQKFE HIIKNEGDKP	ITDEYKVPSK RYTRRKNRIC NIVDEVAYHE EGDLNPDNSD RLENLIAQLP YDDDLNLLA MIKRYDEHHQ EFYKFIKPIL AILRRQEDFY ETITPWNFEE YNELTKVKYV KIECFDSVEI LTLTLFEDRE RDKQSGKTIL HEHIANLAGS DEKKAIQKIQ LWHQQGERCL YATANQEKGO EDISKFDVRK FTSQLRRHWG QLLDIETGEL FNRKISDATI DKSKFLMYRH GYIRKYSKKG DVFYFNKTTGK EFKFTLYKND GGEALIKVLG KLDF	KFKVLGNTDR YLQEIFSNEM KYPTIYHLRK VDKLFIQLVQ QIGDQYADLF QIGDQYADLF DLTLLKALVR EKMKGTEELL PFLKDNREKI VVDKGASQAS TEGMRKPAFL SGVEDRNFAS MIEERLKTYA DFLKSDGFAN PAIKKGILQT KANKDEKDAA YTGTITISIH RTPYQALDSM KFIERNLVD IEKTRDTYHH ISDDEYKESV YATRQAKVGK DPQTFEKVIE NGPEIKSLKY YEILGLKYAD LLLVDKDTETK NVANSQGCKK	HSIKKNLIGA AKVDDSFHFR KLVDSTDKAD TYNQLFEENP LIALSLGLTP LAAKNLSDAI QQLPEKYKEI VKLNREDLLR EKILTFRIPY FIERMTNFDK SAGEQKKAIVD LGTYHDLKLI HLFDDKVMKQ RNFMQLIHDD VKVVDLKV MLKAANQYNG LINNSNQFEV DDAWSFRELK RYASRVVLNA HAVDALIAA FKAPYQHVD DKADETYVLG PILENYPNKQ YDSKLGNHID LQFEKGTGT EQQLFRFLSR GLGKSNISY	(SEQ ID NO: 31)
SpCas9 inactive AJN60011.1 GI: 757015967	MDKKYSIGLD LLFDSGETAE LEESFLVEED LRLIYLALAH INASGVDAKA NFKSNFDLAE LLSDILRVNT FFDQSKNGYA KQRTFDNGSI YVGPLARGNS NLPNEKVLPA LLFKTNRKVT IKDKDFLDNE LKRRTYTGWG SLTFKEDIQK MGRHKPENIV VENTQLQNEK SIDAKVLTRS	IGTNSVGWAV ATRLKRTARR KKHERHPIFG MIKFRGHFLI ILSARLSKSR DAKLQLSKDT EITKAPLSAS GYIDGGASQE PHQIHLGELH RFAWMTRKSE HSLLEYEFTV VKQLKEDYFK ENEDILEDIV RLSRKLINGI AQVSGQGDSL IAMARENQTT LYLYYLQNGR DKARGKSDNV	ITDEYKVPSK RYTRRKNRIC NIVDEVAYHE EGDLNPDNSD RLENLIAQLP YDDDLNLLA MIKRYDEHHQ EFYKFIKPIL AILRRQEDFY ETITPWNFEE YNELTKVKYV KIECFDSVEI LTLTLFEDRE RDKQSGKTIL HEHIANLAGS QKGQKNSRER DMYVDQELDI PSEEVVKKMK	KFKVLGNTDR YLQEIFSNEM KYPTIYHLRK VDKLFIQLVQ QIGDQYADLF QIGDQYADLF DLTLLKALVR EKMKGTEELL PFLKDNREKI VVDKGASQAS TEGMRKPAFL SGVEDRNFAS MIEERLKTYA DFLKSDGFAN PAIKKGILQT MKRIEKGIKE NRLSDYDVDA NYWRQLLNAK	HSIKKNLIGA AKVDDSFHFR KLVDSTDKAD TYNQLFEENP LIALSLGLTP LAAKNLSDAI QQLPEKYKEI VKLNREDLLR EKILTFRIPY FIERMTNFDK SAGEQKKAIVD LGTYHDLKLI HLFDDKVMKQ RNFMQLIHDD VKVVDLKV LGSQILKEHP IVPQSFLKDD LITQRKFDNL	(SEQ ID NO: 32)

Table 4- Cas9 orthologs						
	TKAERGGLSE REVKVITLKS YPKLESEFVY TLANGEIRKR QTGGFSKESI KGKSKKLKSV YSLFELENGR DNEQKQLFVE PIREQAENII SITGLYETRI	LDKAGFIKRQ KLVSDFRKDF GDYKVYDVRK PLIETNGETG LPKRNSDKLI KELLGITIME KRMLASAGEL QHKHYLDEII HLFTLTNLGA DLSQLGGD	LVETRQITKH QFYKVREINN MIAKSEQEIG EIVWDKGRDF ARKKDWDPKK RSSFEKNPID QKGNELALPS EQISEFSKRV PAAFKYFDTT IDRKRYTSTK	VAQILDSRMN YHHAHAAYLN KATAKYFFYS ATVRKVLSP YGGFDSPTVA FLEAKGYKEV KYVNFLYLAS ILADANLDKV IDRKRYTSTK EVLDATLIHQ	TKYDENDKLI AVVGTALIKK NIMNFFKTEI QVNIVKKTEV YSVLVVAKEV KKDLIIKLPK HYEKLKGSPE LSAYNKHRDK EVLDATLIHQ	
AJN60013.1 GI: 757015969 WP_00543065 8.1 Sutterella wadsworthensi s 3_1_45B	MTQSERRFSC PRYVQAQRTA GREALSGLLK LAEQWEEFTA ALSTLRANAN ERLARLLGNL PAKDQNKQHL TLLLSPEKLT NGHEPLPTLA LENCIGGQNV KKILPLLVAN FGGGFNIAYN DEQKRKFANP GETVRAAQCS NGIVDVSIFV RIKKASRGTC SRGNQLKKNQ LKFFDLLNEH IKNLANKIRE KPDIIPIASH IHLQAKPQEE ERCGAIEVSG AALQPLSAEE KLFQLKVELK LTSKLARIWK YQVHAINAKK TPMSEWRKVV ITSPLSLPAS FWYIVVSSNK	SIGIDMGAKY VRHRLRGQKR RRGYSRPNAD NISNVEKFLG VLTGLRQMGH SNLQLRAERW ELIKQIENSE RQYGEIWKTW YQLSYALQRA KTFLDCARRY ILQTDETTQ TAQYREVNKL FSLAQFYTLI RLPAETARPF EENKFEFSAS PYTGDRLAEG RYSLSDLKAN EQDCVRHALF ELQNWCKTTN SIDALCSFAV KSHFDSVAIF KQPKELLEML KRLAALLDAL GEKSFKLNGS RPVMRDLAHA YRGFASAGSN AEDNLSIWIA FKVDKPAEFQ KMNESYNNVS	TGVFYALFDR YTLARKLAFL GEDLTPLENV DPNIPADKEF KPRSEYFKAI YFNAPDIMKD DIIETLCTLD SARLTSAEPT FDRSKALDPY YREADDAKVG KFLDEIWRKQ PRNAQDKELL ETEVSFGSAT DGLVRRRLVDR GEIDHILPRS YRNEIFKTSN LDDGSEARDA NRLHFQAAAT GSADAERDQN KEGIYAEQFL APFFNKPVGD RYCTSRKSLM LTLPVKQDWL PVRREFSLPA VDWSKGILFN PGTEGRRYVR KAVGTELSEL KS	EELPTNLNSK VDDMIKKQE RADVFAAHPA IEFAVAEGLI EADLKKDSRL RGWEPDRFKK PNRTIPPYED LAPAAEILER ALRALAAGSK LWFDNADGLL IKGRETVASR TIRDVAETA TLAVHLENOW QAWBIAKRVS KDKMLSEAEK LIKDARGIVF IAAITAEIED VLELLATQRR NVSDAKNLRL GFDYLDGKTV PIFTLNEKIW LSAHATYRIL SLFMAANGKS RICDSPELAD IDNPSSGGFRI ELQHENLTEC VETTFIQASH LGQPRSEIFI	AMTLVMPETG KRLTDEEWKR FSTYFSEVRS DKTEKKAYQS AKINEAFGGA TLVRAFKKFFH QNNRRPPLDQ STDKRSRVAV SNKLTSARTA ERSDLHPPMK CARIETVRKS DFIAANLGLS RMTIKDAVIN TDIQSKVDFS LETRWLKNE NAEPNLIYAS VVTKLQQTHR TRVNGTQIWM KLAQNQPDFE LGLYPQSCFV IGYETLNAKG KKPAYEFLAK LKKREDVLKP AFGKPCSADE RRTNLFQNEL GGRFITSADV WFEQSVENWA ENVGNAXHIR	(SEQ ID NO: 33)
AJN60014.1 GI: 757015970 WP_01121279 2.1 Legionella pneumophila str. Paris	MESSQILSPI QLSQAQRRAT LNNRGYTYVD FRKILVSKVE KSDITKDNQL QTFGNEFLRM TIPPYEARTN LEHTKLDRDK QGKQLPANLI AFSLCELSNI GRTSLKSKCK DIIQAIQSHL NYWRSQKTEI IKHIPDNSSL EEKFQRIINA LIYCSSQGNR KKYISFHLLT	GIDLGGKFTG RHRVRNKKRN TDLDEYIKDE EKKDDKELKN DSIKKKIPSV LKNFRHLKGS TGMEKDQSL RIISPSKQDE ETQKEMETHF NPPRKQKILP EIEEARKNSG GHNDSQALIY DPEISYASRL LIPIYLEQNR SMNICPYKGA EKKEEHYLL PEQQKAARHA	VCLSHLEAFA QFVKRVALQL TTINLLKELL AVKNIKNFIT CLSNLLGHLS QESLAVRNLI LNPEKLNNLY KRDSYILQRY NSSLVSVLIQ LLVGAILSED NAFKIDYEEA HNPFSLSQLY PADSVRPFDF FEFEESFKKI SIGGQGEIDH HLSPLYLKHQ LFLDYDDEAF	ELPNHANTKY FQHILSRDLN PSESEHNFID GFEKNSVEGH NLQWKNLHRY QQLEQSQDYI PNWRNLIPGI LDLNKKIDKF IASAYNKERE FINNKDKWAK LNHPHNSNNK TILETKRDGF VLARMQRLA KGSSSDKTLE IYPRSLSKKH FGTDNVSDIK KTITKFLMSQ	SVILIDHNNF AKEETALCHY WFLQKMQSSE RHRKVYFENI LAKNPKQFDE SILEKTPPEI IDAHPFLEKD KIKKQLSFLG DAAQGIWFDN FKIFWNTHKI ALIKIIQTIP HKNCVAVTCE YEIAMAKWEQ QAIEKQNIQW FGVIFNSEVN NFISQNVANI QKARVNGTQK	(SEQ ID NO: 34)

Table 4- Cas9 orthologs						
	FLGKQIMEFL	STLADSKQLQ	LEFSIKQITA	EEVHDHRELL	SKQEPKLVKS	
	RQQSFPSHAI	DATLTMSIGL	KEFPQFSQEL	DNSWFINHLM	PDEVHLPVVR	
	SKEKYNKPNI	SSTPLFKDSL	YAEFPIPVWV	KGETFAIGFS	EKDLFEIKPS	
	NKEKLTLLK	TYSTKNPGES	LQELQAKSKA	KWLYFPINKT	LALFLHHYF	
	HKEIVTPDDT	TVCHFINSRL	YYTKKESITV	KILKEPMPVL	SVKFESSKKN	
	VLGSFKHTIA	LPATKDWERL	FNHPNFLALK	ANPAPNPKEF	NEFIRKYFLS	
	DNNPNSDIPN	NGHNIKPKQH	KAVRKVFSLP	VIPGNAGTMM	RIRRKDNKGQ	
	PLYQLQTIDD	TPSMGIQINE	DRLVKQEVLM	DAYKTRNLST	IDGINNSEGQ	
	AYATFDNWL	LPVSTFKPEI	IKLEMKPHSK	TRRYIRITQS	LADFIKTIDE	
	ALMIKPSDSI	DDPLNMPNEI	VCKNKLFGNE	LKPRDGKMKI	VSTGKIVTYE	
	FESDSTPQWI	QTLYVTQLKK	QP			
AJN60015.1 GI: 757015971 WP_00268128 9.1 Treponema denticola ATCC 35405	MKKEIKDYFL	GLDVGTSVSG	WAVTDTDYKL	LKANRKDLWG	MRCFETAETA	(SEQ ID NO: 35)
	EVRRRLHRGAR	RRIERRKKRI	KLLQELFSQE	IAKTDEGFFQ	RMKESPFYAE	
	DKTILQENTL	FNDKDFADKT	YHKAYPTINH	LIKAWIENKV	KPDPRLLYLA	
	CHNIIKKRGH	FLFEGDFDSE	NQFDTSIQAL	FEYLREDMEV	DIDADSQVKV	
	EILKDSCLKN	SEKQSRLNKI	LGLKPSDKQK	KAITNLISGN	KINFADLYDN	
	PDLKDAEKNS	ISFSKDDFDA	LSDDLASILG	DSFELLLKAK	AVYNCSVLKS	
	VIGDEQYLSF	AKVKIYEKHK	TDLTKLKNVI	KKHFPKDYKK	VFGYNKNEKN	
	NNNYSYVGV	CKTKSKLLII	NNSVNQEDFY	KFLKTILSAK	SEIKEVNDIL	
	TEIETGTFLP	KQISKSNAEI	PYQLRKMELE	KILSNAEKHF	SFLKQKDEKG	
	LSHSEKIIML	LTFKIPYYIG	PINDNHKKFF	PDRCWVVKKE	KSPSGKTPPW	
	NFFDHIDKEK	TAEAFITSRT	NFCTYLVGES	VLPKSSLLYS	EYTVLNEINN	
	LQIIIDGKNI	CDIKLKQKIY	EDLFFKYYKI	TQKQISTFIK	HEGICNKTDE	
	VIIIGIDKEC	TSSLKSYIEL	KNIFGKQVDE	ISTKNMLEEI	IRWATIYDEG	
	EGKTILKTKI	KAEGYGYCSD	EQIKKILNLK	FSGWGRLSRK	FLETVTSEMP	
	GFSEPVNIIT	AMRETQNNLM	ELLSSEFTFT	ENIKKINSKF	EDAQKQFSYD	
	GLVKPLFLSP	SVKKMLWQTL	KLVEISHIT	QAPPKKIFIE	MAKGAELEPA	
	RTKTRLKILQ	DLYNNCKNDA	DAFSSEIKDL	SGKIENEDNL	RLRSKLYLY	
	YTQLGKCMYC	GKPIEIGHVF	DTSNYDIDHI	YPQSKIKDDS	ISNRVLCSS	
	CNKNKEDKYP	LKSEIQSKQR	GFWNFLQRNN	FISLEKLNRL	TRATPISDDE	
	TAKFIARQLV	ETRQATKVAA	KVLEKMFPEP	KIVYSKAETV	SMFRNKFIDV	
	KCREINDFHH	AHDAYLNIIV	GNVYNTKFTN	NPWNFIKEKR	DNPKIADTYN	
	YYKVFYDVK	RNNITAWKEG	KTIITVKDML	KRNTPIYTRQ	AACKKGELFN	
	QTIMKKGLGQ	HPLKKEGPF	NISKYGGYNK	VSAAYYTLIE	YEEKGNKIRS	
	LETIPLYLVK	DIQKDQDVLC	SYLTDLLGKK	EFKILVPKIK	INSLKLINGF	
	PCHITGKTND	SFLLRPAVQF	CCSNNEVLYF	KKIIRFSEIR	SQREKIGKTI	
	SPYEDLSFRS	YIKENLWKK	KNDEIGEKEF	YDLLQKKNLE	IYDMLLTCHK	
	DTIYKKRPNS	ATIDILVKGK	EKFKSLIEN	QFEVILEILK	LFSATRNVS	
	LQHIGGSKYS	GVAKIGNKIS	SLDNCILIIYQ	SITGIFEKRI	DLLKV	
AJN60016.1 GI: 757015972 EFE28295.1 Filifactor alocis ATCC 35896	MTKEYYLGLD	VGTSVGVAV	TDSQYNLCKF	KKKDMWGIRL	FESANTAKDR	(SEQ ID NO: 36)
	RLQRGNRRRL	ERKKQRIDL	QEIFSPEICK	IDPTFFIRLN	ESRLHLEDKS	
	NDFKYPLFIE	KDYSDIEYK	EFPTIFHLRK	HLIESEEKQD	IRLIYALHNN	
	IIKTRGHFLI	DGDLQSAKQL	RPILDTFLLS	LQEEQNLSVS	LSQNKDEYE	
	EILKNRSIAK	SEKVKKLKLN	FEISDELEKE	EKKAQSAVIE	NFCKFIVGNK	
	GDVCKFLRVS	KEELEIDSFS	FSEGKYEDDI	VKNLEEKVPE	KVYLFEQMK	
	MYDWNILVDI	LETEEYISFA	KVKQYKHK	NLRLLRDIIL	KYCTKDEYNR	
	MFNDEKEAGS	YTAYVGKLKK	NNKKYWIEKK	RNPEEFYKSL	GKLLDKIEPL	
	KEDLEVLTM	IEECKNHTLL	PIQKNKDNV	IPHQVHEVEL	KKILENAKKY	
	YSFLTETDKD	GYSVVQKIES	IFRFRIPYV	GPLSTRHQEK	GSNVWMVRKP	
	GREDRIYPWN	MEEIIDFEKS	NENFITRMTN	KCTYLIGEDV	LPKHSLLYSK	
	YMLNLNLLNV	KVRGKKLPTS	LKQKVFEFLF	ENKSKVTGKN	LLEYLQIQDK	
	DIQIDDLSGF	DKDFKTSLSK	YLDFFKKQIFG	EEIEKESIQN	MIEDIIKWIT	
	IYGNDEMLK	RVIRANYSNQ	LTEEQMKKIT	GFQYSGWGNF	SKMFLKGISG	
	SDVSTGETFD	IITAMWETDN	NLMQILSKKF	TFMDNVEDFN	SGKVGKIDKI	
	TYDSTVKEMF	LSPENKRAVW	QTIQVAEEIK	KVMGCEPKKI	FIEMARGGEK	
	VKKRTKSRKA	QLLELYAACE	EDCRELIKEI	EDRDERDFNS	MKLFLYYTQF	

Table 4- Cas9 orthologs						
	GKCMYSGDDI KKSNELLSED IARQLVETRQ VNDYHHAKDA VINGEVIWEK FNATIQNKNG HIIGVPIYIA PLRIRIGENEV ENRDHITHEK IDKINVINKM VNQSVTGLYE	DINELIRGNS IQKKMHSFWL STKAIADIFK YLNIVVGNVY CTYHEDTNTY NSTVSLKKGL NMLEHSPSAF DTSFKRAIQL MNQLEYVLLS LNLLRCDNDT NRREL	KWDRDHIYPQ SLLNKKLITK QIYSSEVVYV NKKFTSNPIQ DGGTLDRIK DVKKYGGYFS LEYCEQKGQY KLDQKNYELV KMKKFNKKGM KADLSLIELP	SKIKDDSIDN SKYDRLTRKG KSSLVSDFRK WMKKNRDTNY IVERDNILYT ANTSYFSLIE NVRILVEKIK RNIEKFLEKY ADPSDRIEKS KNAGSFVVVK	LVLVNKTYNA DFTDEELSGF KPLNYLKSRR SLNKVFEHDV EYAYCEKGEL FEDKKGDRAR KNSLLIINGY VEKKGNYPID KPKFIKLEDL NTIGKSKIL	
AJN60017.1 GI: 757015973 WP_01461325 9.1 Staphylococcus pseudintermedius ED99	MGRKPYILSL RRQFRSRRR KNKSLSEVLS YRGHFLFDHL EILKDNEMTK LKGANQSHTF SMAMSKVAAY TPNDQTFSSL LKVLNLTDNA IPYYVGPLVN TNKCSYLINE IKLFAVEHIF KFASKLSSYQ PELTSKQINQ LTNDVYGFQN LTSIFGEPEK IIDVANKLNN VDHIFPRSEI LNKAGLISDS YPNTKVIPMK AYPNVDLDFD ERLISKIKLD EVVYKGLTPY LALYLAQREN ELTVEQQLSL RVRTFFSESN GKGKDVKIAY	DIGTGSVGYA KNRRIKRLGL FLGYESKKYP KIENTLNNDN NDRAKRVKNM ADNYEENLTP DKFRNELKQV CLFDQYLIRP SIPMQINLYE DHTASKFGWM DVLPKNSLLY KKYKTVSHSK DMTKIFGDI LKKLNYSGWG FIKEENQVQS IIMEFATEDQ EQLQKEKLWL KDDSIDNKVL KLHKLMKPEF AKMVSEFRKK NFKWEKVREK MNHFKINYSR QTYVVAIKSV KDEVLDQIV YNVMNNKETN QTHEDFIKAL TSISGLKTTK	CMDKGFNVLK LQELLAPLVQ TIYHLQEALL MHDFVELIET EKKLEQFSIM FLTVEQSEFI KDIFYKADST KKQYSLLIKE AETILRNQQK ERKSNESEIKP QEMEVLNELN FLEIMLNSNH GKRAQIEEII RLSEKLLTHA NKIQHQDIAN QKGGKQKSRK YLSQNGKCMY VIKKMNQTKG TAMDKEGFIQ FDIPKIRQMN WKALGEFNTK KLANIPQQFY NKKGKEKMEY YSLNKGDLly VEKLLIEYDF DELFKVVtas PKSLFKLAES	YHDKDALGVY NPNFYQFQRQ LKDEKFDPEL YENLNNIKLN LLGLKFNEGK ERANKIYLSL RTQFKKIFVS LKKIIPQDSE YHABITDEMI WNFDEVVDRS ATQIRLQTDp RENFMNHGEK QWITIFEDKK YQGHSSIIEL LTTSPALKKG QLWDDNIKKN SGQSIDLDAL DQVPLQFIQ RQLVETROIS DAHHAIDAYL QKSRELFFFK NQTA VSPKTA Q MIDHYVDF INNHPCYFVS IAEKVINEYH ATRS DKIGSR RNEL	LFDGALTAQE FAWKNDNMDF IYMALYHLVK LDYEKTKVIY LFNHADNAEE TLQDILKGKK SKKSLKQYDA LYFEAENDTL EKVLSLIQFR KSATQFIRRM KNRKYRMMPQ LSIFGTQDDK ILVQKLKECY RHSDENFMEI IWSITIKLVRE KLKSVDEYKY LSPNATKHYE PYERIAWKS VHVRDFLKEE NGVVYHGAQL KLEKMEVSQG ELKYESNKS YKFQNGNEKE RKEVINAKQF HYLSNKLKEK KNSMTHRAFL	(SEQ ID NO: 37)
AJN60018.1 GI: 757015974 WP_01456756 1.1 Lactobacillus johnsonii DPC 6026	MTKIKDDYIV AERRLFRTTH DKRKTVSSIV HHIIKYRGNF SNTAEIEKVL AVLGYKTRFD AILLTIKELF IDRKKAKELA DDSRASKEIK PFLKEINPVS QNFAWMIRKE ANSLLYQKFT NYLKENHNLV DFEKIIEWST LLAQLHDHNG EDILNNAYTS PKRSQTRGSK QLGRDAYTGE	GLDIGTDSCG RRIKRRRWRL FPTS AEDKKF LYNTSVKDFK KDKQIFKRDK TIALKEISKD NEVTNLNGIVE LAYDLYVNNR KKIELD SFMP SHLKEAPYKL EGRITPWNFD VLNELNNIRI KVEIKGLADE IFEDKSIYKE QTIIEQLWDS PANKKAIRQV LQKVYKDLST PINIDEIQKY	WVAMNSNNDI KLLEEFFDPY YDDYPTIYHL ASKIDVKSSI VKKIAELFAI ELSDWNFKLS DGNTLSESMI HGQLLQAKKK KQRTNANGVI DELIRFRVPY QKVDRIESAN NGKRISVDLK KKFNSGLTTY KLRSIDWLNE QNNFMQIVTQ IKVVDIVKA ELASKTIAEE DIDHILPQSF	LKLQGKTAIG MAEVDPYFFA RYKLMTEDEK EKLNELYENL KTDNKEQSKR DIDADSKFEA NKYNDHRDDL LGKIKPRSKE PYQLQQLELD YVGPLISPNE KFIKRMTTKD QEIIYENLFKK NRFKNLNIFD KQINALSNIR ADFKDAIAKA ASGKVPKQIA LNEAIKDKKL IKDDALDNRV	SRLFEGGKSA RLKESGLSPL FDLREYVLA GLDLNVEFNI IKDISQVAN LMGNLDENEQ KLLKEVIENH DFYKVVNKNL KIIENQSKYY STKDIQTKKN TYLFGEDVLP HTTVTVKKLE NQIDDLKYRN LQGWGRLSKK NQNLVATSV IEFTRDADEN VQDKYYLYFM LVSRAVNNGK	(SEQ ID NO: 38)

Table 4- Cas9 orthologs						
	SDNVPVKLFG KYKSAGFIRR DLYKSREVND DKEKALFNKT TRDQEMFKMT IVRIDKKKGT NGKRDKKITS SMKAITNNFD EKFIKLSLED LSENAKLIYQ	NEMAANLGMT QLVETSQIIK YHHAIDAYLS RKFSFISQLL VYPRFSHDTV EYKILGIPTR FEIVKSKIPY KDSDENDALI KKDTILKVLE SPTGLFKKSV	IRKMWEEWKN LVSTILQSR AICGNLLYQN KNKSENSKEI KAPRNLIPKK ELVNLKKA KQVIQDGD KAYDEILD GLHDNAVMT KISDL	IGLISKTKYN PNTETITVKA YPNLRPFFVY AKKLKRAYQF MGMSPIYGG EDHYKSYLKE FMLGSSTYVY DKYLPLFDIN IPTIGLSTPL	NLLTDPDHIN KYNHYLREKF GQYKKFSSDP KYMLVSRERE YTNNSDAYMV ILTPRILYNK NAKQLTLSTE KFREKLHSGR GFMQFPNGVI	
Mycoplasma gallisepticum str. F AJN60022.1 GI: 757015978 WP_01457478 9.1	MNNSIKSKPE RRSFRGVRRL HNTVLNLKSE AKYFDKYGYY GLPEKFKEEY NIWDKTIGKC NQEFKKAYLN ENEEKLEKED LEDQYATLKD QKQKDGFLKL KNNDKGWNFE GVKRILREAT KLIKKNGDKI LKEIAFDDIF PYEFISSGNA SKYDIGFLAR TSFLRKNFDD YKLFKNTDGS DSNIERKARY ESAKENKNSK INLAEQIFNE TAIYYDDKKN INSLALMIMR FVKYLKANEI NLIETEKYEN KNRIDEILNS	VTIGLDLGVG IRRRKYKLKR ALNAKIDPKA KGIIDSKE ESLFSYVRNY NIFPDEYRAP KLLDLLIKTN HKWKLKGLKL INFLQSLFVY FEQTDKDD AIKNFDQKFI KVFNAILKQF SEGLKALGIS TKTEKFEIDH GIKWEDYEAY NLNDTRYATI SSYAKDRDK WKKIDPKTGV SRKIENKTNI VKRQFVYRKL YTENKKIKSQ IVHRIKRLKM SIDPTAKKQY GDEYKPWRVL DDSDSKKKKK KLGLDKIVK	SVGWAIVDNE FVNLIWKYNS LSWILHDYLY DNKLEELTK SEGPGSINSV KNSPIAMIFN GEKPIDARQF NTNGKIQYND LGKHLRYSNR ILAQTHSLST DITKKNNNLS SEEYDVTQV EDEIKDILKS IIPYSISFDD CRKFKDGDSS VFRDALEDYA NIHHAVDASI VTEVTDENWK SLFNDTVYSA VNVSLNNDK NVFEKYMLDL LSSELKENKL IRVPLNTLNL TSGTLLIHKK ANRFLMTLST	TNIIHHLGSR YFGFKNKEDI NRGHFYEDNR YKFSNKHWE SPYGIYHLDE EINELSTIRS KKLREETIAE LSSLAKFVHK VDSANLKEFS KAMLLAITRM LKQNKRYLDD IELARELSEE PTKSYKFLW SSSNKLLVLA LLDSTQRSKK NNHLVEDKPM ISIFSNETKT QIRVRNQVSE KKVGIEDQIK LADLFAEKED TKEFPEKFSE SNVIIRSKNQ HLGDHDFDLH DKKLMYISSF ILNDYILLDA	LFSQAKTAED LNNYQEQQKL DFNVYPTKEL EVKKVLSNQT KEGKVVQKYN YSIYLTGWFI SIGKETLKD LKQHLKLDL DSNKLFERIL TNLDNDEDNQ RFINDAILSP KELENTKNYK LQQDHIDPYS ESNQAKSNQT FAKMMKTDTS FKVVCINGSV LFNQLTQFAD IAKVIKYYIQ RKNLKTLDIH ILMYRANPWV FLVKSMRLNK SGTKLSYQDT NMDAYLKKPK QNLDNDVIEIK KDNFDILGLS	(SEQ ID NO: 39)
AJN60023.1 GI: 757015979	MRILGFDIGI RSSRRRLKRR YELRYKALTQ KNNALKLENY LEKELKLILE EEEKRACKNS DKGSITYKKF GVHSLSRQEL YINLSFKALG FAHELSPV KIEKEQKENQ KISIEHLKDE AFGKNIEKWS IATLIAKYTK TWGFDKKDRN ELTSDNYKHQ HSENKIIDKC KFYAIPYAM NDLILQKKN NAKEGSVKVE	NSIGWAFVEN KARLIAIKRI NLETKDLARV QSVGEYFYKE KQKEFGYNYS YSAWEFVALT RSCINLHESI DQISTHITLI MILPLMREGK NRAISEYRKV AVNAWALKEC KALEVDHIYP KIQTALQNL EYLNFLLLSE NHLHHALDAI VKFFPEPKSF SYNSKEGLQI DFALGILPNK MQEPEFAYYN SLGIQNLKVF	DELKDCGVRI LAKELKLNK ILHIAKHRGY FFQKYKKNK EDFINEILKV KIINEIKSLE SFKSLKYDKE KDNVKLKTVL RYDEACEIAN LNALLKKYK ENIGLKASAK YSRSFDDSF YKKKNKILDE NENANLKS IVAYSTNSII REKILSKIDE ALSCGRVRKI IVITGKDKNN DFSISTSSIC EKYIITPLGD	FTKAENPKNK DYVAADGELP MNKNEKKSND NFIKIRNTKD AFFQRPLKDF KISGEIVPTQ NAENAKLIDF EKYNLSNEQI LKPKTVDK VHKIHLELAR NILKLKLWKE NKVLVFTKEN NFKDKQQEDF KGSKIHVQTI KAFSDFRKNQ IFVSKPPRKR GTKYVENDTI NPKQWQTIDE VEKHDNKFEN KIKADFQPRE	ESLALPRRNA KAYEGSLASV AKKGKILSAL NYNNCVLSSD SHLVGACTFF TINEVLNLIL RKLVEFKKAL NNLLEIEFND DFLPAFCDSI DVGLSKKARE QKEICISYSGN QEKLNKTPFE ISRNLNDRY SGMLTSVLRH ELLKARFYAK ARRALHKDTF VRVDIFKKQN SYEFCFSLYK LTSNQKLLFS NISLKTSSKKY	(SEQ ID NO: 30)

Table 4- Cas9 orthologs						
	GLR					
AJN60025.1 GI: 757015981	MSDLVLGLDI QGRRLARRKK SNEELFIALK PGQIQLEERYQ QEFNPQITDE IFGILIGKCT KNQIINLVKN EAYRKMKTLE FSQKQVDELV TRLGKQKTTS GDFDNVVIEM LPHSVFHGK LPLSITFDDS RESKTLNKK HFRAHKIDTK LNLWKKQKNT SKEFEDSILF DIYTQDGYDA KGKEVPCNPF KDSNNKVVLQ QEKYNDIKKK KQKHVELKP TDVLGNQHII	GIGSVGVGIL HRRVRLNRLF NMVKHRGISY TYGQLRGDFT FINRYLEILT FYPDEFRAAK EKAMGPAKLF TLDIEQMDRE QFRKANSSIF SSNKTKEYIDE ARETNEDDEK QLATKIRLWH LANKVLVYAT KEYLLTEEDI VSVVRGQFTS LVSYSQDQLL SYQVDSKFN FMKIYKKDKS LKYKEEHGYI SVSPWRADVY EGVDSQSEFK YDKQKFEGGE KNEGDKPKLM	NKVTGEIHHK EESGLITDFT LDDASDDGNS VEKDGGKHRL GKRKYHGP NEKSRTDYGR KYIAKLLSCD TLDKLAYVLT GKGWHNFVSK KLLTEEIYNP VVAKSVRQAI KAIQKIQKAN QOGERCLYT ANQEKQRT SKFDVRKKFI QLRRHWGIEK DIETGELISD KISDATIYAT KFLMYRHDPQ RKYSKKGNP FNKTTGKYEI FTLYKNDLLL VKDTETKEQQ ALIKVLGNVA	NSRIFPAAQA KISINLNPYQ SVGDYAIQIVK INVFPTSAYR NEKSRTDYGR VADIKGYRID LNDLNNLTVP LNTEREIGIE YQALDSMDDA ERNLVDTRYA TRDTYHHHAV DEYKESVFKA RQAKVGKDKA TFEKVIEPIL EIKSLKYIDS LGLKYADLQF VSKDTETKEQQ NSGQCKKGLG	ENNLVRRRTNR LRVKGLTDEL ENSKQLETKT SEALRILQTQ YRTSGETLDN TETKKLSKEQ KSGKAEIHTF ALEHEFADGS ETSEEQMTIL KIVNAAIKEY AANQYNGKAE NSNQFEVDHI WSFRELKAFV SRVVLNALQE DALIIAASSQ PYQHFDVTLK DETYVLGKIK ENYPNKQINE KLGNHIDITP EKGTTGYKIS LFRFLSRTMP KSNISIIYKVR	(SEQ ID NO: 41)
WP_00266404 8.1 Bergeyella zoohelcum ATCC 43767	MKHILGLDLG QAQTKNADRR DPNKIDKITI QLRGYAGGSL KKLNRRRAIIV LPNKTNRWKK TILRARYESE YRELGLEKGL FQYKQVWEQI QNKKEITFSA GELWDVLGLD IVDDNAEETA LQSKILKLLN TAKEYSHDEL KPNEIRVELA ANIEKVKLWE TNKVISEKSV RKNLLATSIP YLRNHWGLTD KEVLFEQEQL IVACTEPAHI ENERTEIFTQ KAGDRQIKLR SPFLSGFIAN ISTVKIYYKD LEGEIKTKKT YFEERNKAKL NIYVVQKFSK LEIDRLGNIV	TNSIGWALIE TNRGARRLNK LPISKKQEQ EPEKEDIFDE ETEEGNFEGS MENIENQLKE FEAIWNEQVK KYI IKNQVVF NKLIVNTKIE IFKKLKAEDF SINRQIELWN IRISKIKFAR ENVEDPFAKA YKSYKEINLL RDLKNSAKER AQRHLSPTYG NQEKANRTAM EDPVQRQIKD KFKLLLKERY TREEFIKEYK KRLNDLNKVL IEKFRAIEMP GQLHEGTLYG HLKEYNNKKE PSKNKKKKDE SQIKRLYDII LFTLKQGDFV KQIYFIKHTI KVIKR	RNIEEKYKGI RYKQRRNKLI TAFDLVSLRV EQSKDKKNKS TFLENIKVGD KSKEMGREFY HYPFLENLDK YQRELKDQSH AGTNRKGEKK LREGIDFLNG ILYNEKGNEY AYSSLSLKAV AQTYLDNNQS KQGDLRNPLV ATIHKRKNKD QPIPLSDLFD EYFEVGSGLY TOYIAIRVKE EALLESEKFL ENYIRYKKNK QDWLVEHKSE WKGFPQVEQ ISQGEAYRI EAFSAEGIMD EDLSLQKLDR SFFDATNFLK YLPNENEEVI ADIIKKDVEF	IGMGSRIVPM YILQKLDMLP KALTEKVGLE FIAFSKIVFL SLELLINISA ISEFLLELLK KTLEIIVSFI LISDCRYEPN YKYIDRPIPT MSPKDKLKGN DLTSDRTSKV ERILPLVRAG VLSEGGVGNS EQINEALVL QTINNIKIKET SIFTKEQFIA ELNKIVGNEN EAEYDNYKKD LIIKGWSKRI FMPNFEGSNS KLKEI IISHK PLTKFGGSKF LNNKLAQYRN EKAFNEKLYV EEFRNAPDKK LDKESPLYNQ GSQNCYETVE	GAELSKFEQG SQIKLKEDFS DLGKIIYKYN GEPQEEIFKN SKSGDTITIK ENRWAKIRNN FPGEKESQKK EKAIKASHPV ALKEWIFEEL ETKLQLQKSL LEFINKYGNN KYFNNDFSQQ IATILVYDKH IRDIWKNYGI LVKNKKELSL PISRYFDDSF HVNEYFSGVK VKTTTGSITD FDSRKKEYEE DHRHHAIDAL ELLEELSLP PKDKLLQYN ATEKNIQKIV EKGELKPHTP KTGDNYLFAV TFDKDLLFRQ YWGDLKERGK GRSIKENCFFK	(SEQ ID NO: 42)
CBK78998.1	MKQEYFLGLD RMFRTARRRL DAEGNCPPELP	MGTGSLGWAV DRRNWRIQVL YALFVDDNYT	TDSTYQVMRK QEIFSEEISK DKNYHKDYPT	HGKALWGTRL VDPGFFLRMK IYHLRKMLME	FESASTAEER ESKYYPEDKR TTEIPDIRLV	(SEQ ID

Table 4- Cas9 orthologs						
Coprococcus catus GD/7	YLVLLHMMKH AAIQFVEHVL DIFNNKELDE SVLVEILGNS TEEKLNNYSA ITQEIESEIE ENAEKIQQLF VIDVEASAEK GEKISVELKQ FKASLTAYHD PNLTTGQLKG NDNLMQLLSR TLKVVKEIQK ERDWITELNA IDHIYPQSKT QQQGFITKEK LPDTEIVYVK KFTKNAAWFI MQKNNILVTR NKAAGTYFML GLNSPEILLS AILKGVVKYV RLSAQIKTLT GSAGILVMNN	RGHFLLSGDI KDRNLTRSTK SERPKVSFAD VSISEAKIKV YIGMTKKNGK KENFLPKQVT EFRIPYYVGP FIRRMNTNKCT RIYEELFCKY FKERLTDVQL ICSLSYQGWG NYGFTNEVEE VMGNAPKRVF QSDQQLRSDK MDDSLNNRVL YVRLVRSDEL AGNVSNFRQT RNNPGRSYNL KAYEVKGGLF VKSLDKKKE KIKIDTLFKV NRKNENKDAK EKRKAFIGLS NITACKQISV	SQIKEFKSTF KSRLIKQLNA SGYDDYIGIV YQKHQADLKT KVDLKSQCT KDNQVIPYQV LNRVDDGKDG YLVGEDVLPK RKVTRKKLER SQRAKEAIVL RLSKTFLEEI FNTLKKETDL VEMAREKQEG LFLYYIQKGR VKKNYNAIKS SADELAFIE YELLKVREMN KRMFEFDIER DQQIMKKGKG IRTIEFVPLY DGFKMWLSGR LSERDGMTEE NEDQCIVLNE INQSPTGIYE	EQLIQNIQDE KSACEKAILN EABLAEQYYI LKKIVRQYMT QADFYDFLKK HDYELKKILD KFTWSVRKSD DSLVSYSKFMV YLVIEGIAKK NVVLFGDDKK TVPAAGTGEV SYKTVDELYV KRSDSRKKQL CMYSGETIQL DTYPLSLDIQ RQIVETRQST DLHHAKDAYL SGEIAWKAGN QVPIKGNDER LKNQIEINHE TGNQLIFKGA KLLQLYDTFL ILHMFQCQSG KEIDLILK	ELEWHISLDD LLSGGTVKLS IASAKAVYDW KEDYKRVFVD NVIKVIDHKE NLGTRMPFIK ARIYPWNFTE LNELNNLRLN GVEITGIDGD LLKQRLSKMY WNIMTALWQT SPAVKKQIWQ VELYRACKNE DELWDNTKYD KKMMSFWKML KAVATILKEA NIVVGNAFYV KGSIVTVKKV LADIEKYGGY SAIQYLAQER NQLILSHQEA DKLSNTVYSI SANLKLIGGP	NO: 43)
WP_00223516 2.1 Neisseria meningitidis Z2491	MAAFKPNPIN VPKTGDLSAM GLIKSLPNT ADKELGALLK HTFSRKDLQA VQKMLGHCTF ERATLMDEPY KAYHAISRAL DRIQPEILEA DHYGKKNT IHETAREV KDILKLRLYE NNKVLVLGSE RILLQKFDED QITNLLRGFW NAFDGKTIDK DTPEKLRTLL KRLDEGVSVL KAFAEPFYKY DVFEKGDYKY LHPNDLVEVI GVKTALSFQK	YILGLDIGIA ARRLARSVRR WQLRAAALDR GVADNAHALQ ELILLFEKQK EPAEPKAAKN RKSCLTYAQA EKEGLKDKKS LLKHISFDKF KIYLPPIPAD KSFKDRKEIE QQHGKCLYSG NQNKGNQTPY GFKERNLNDT GLRKVRAEND ETGEVLHQKT AEKLSSRPEA RVPLTQLKLL DKAGNRTQQV LVPIYSWQVA TKKARMFYGF YQIDELGKEI	SVGWAMVEID LTRRRRAHRL KLTPLEWSAV TGDFRTPAEL EFGNPHVSGG TYTAERFIWL RKLGLLEDTA PLNLSPELQD VQISLKALRR BIRNPVVLRA KRQEENRKDR KEINLGRLE EYFNGKDNSR RYVNRFLCQF RHHALDAVVV HFPQPWEFFA VHEYVTPLFV DLEKMNVRER KAVRVEQVQK KGILPDRAVV ASCHRG TGNI RPCRLKKRPP	EDENPICLID RARRLKREG LLHLIKHRGY ALNKFEEKESG LKEGIETLLM TKLNNLRILE FFKGLRYGKD ETGTAFFSLFK IVPLMEQGR LSQARKVING EKAANKFREY KGYVEIDHAL EWQEFKARVE EPKLYEALKA TGVVVRNHNH QKDEEDWQL SRAPNRKMSG BPKLYEALKA TGAVVRNHNH QKDEEDWQL NIRIHDLDHK VR	LGVRVFERAE VLQAADFEN LSQRKNEGET HIRNQBGDYS QQRPALSGDA QGSRPLTDT NAEASTLMEM TDEDITGRILK YDEACAEIYG VVRVYGPSPAR FPNFVGPBKS PFSRTWDDSF TSRFPKSKKQ GKKRVFASNG ITRFVRYKEM PDGKPEFEEA QGHMETVSKA RLEAHKDDPA IADNATMVRV IDDSFNFKFS IGKNGILEGI	(SEQ ID NO: 44)
WP_01241442 0.1 Elusimicrobiu m minutum Pei191	MQKNINTKQN DGNVLLPKEI LAEQKLALPV LSLQELGYVL AKKNYRTYGE ILKKQAGKDG PASHKLNEER DYLLTGSELT MPFWARLSEA VLSSSYAPLG	HIYIKQAQKI IMVGSRIFFKA PADLDRKENS YHIAGHRGSS YLYKEFFENK FHKELTEEYI RLWETLNNAR PAQTKKLLGL QQDSFLYDWN KSAMLIILEK	KEKLGDKPYR SAGAADRKLS SEGETSAKRF AIRTTFENDSE EKHKREKISN EKLTKAIGYE YSDPIVDIVT KNTNFEDIIL SCPDEKLLTE IKNDLSYTEA	IGLDLGVGSI RGQRNNHRHT LGDVLQKDIY EAQKENTENK AANNHKFSPT SEKLIPESGF GEITGYEYEQ QGRDKKAQKI KLSNEYHLTE VEEALKEGKL	GFAIVSMEEN RERMRYLWKV ELRVKSLDER KIAGNIKRLM RDLVIKEAEA CPYLKDEKRL FTKEQKQKLF KGYKLIKLES EEIDNAFNEI TKEKQAIKDR	(SEQ ID NO: 45)

Table 4- Cas9 orthologs						
	LPYYGAVLQE	STQKIIAKGF	SPQFKDKGYK	TPHTNKYELE	YGRIANPVVH	
	QTLNELRKLV	NEIIDILGKK	PCEIGLETAR	ELKKSIEDRS	KLSREQNDNE	
	SNRNRIYBIY	IRPQQQVIIT	RRENPRNYIL	KFELLEEQKS	QCPFCGQIS	
	PNDIINNQAD	IEHLFPIAES	EDNGRNNLVI	SHSACNADKA	KRSPWAAFAS	
	AAKDSKYDYN	RILSNVKENI	PHKAWRFNQG	AFEKFIEKNP	MAARFKTDNS	
	YISKVAHKYL	ACLFEPNII	CVKGSLLTAQL	MAWGLQGLM	IPFAKQLITE	
	KESESFNKDV	NSNKKIRLDN	RHHALDAIVI	AYASRGYGNL	LNKMAGKDYK	
	INYSERNWLS	KILLPPNNIV	WENIDADLES	FESSVKTALK	NAFISVKHDH	
	SDNGELVKGT	MYKIFYSERG	YTLTTYKKLS	ALKLTDPQKK	KTPKDFLETA	
	LLKFKGRESE	MKNEIKSAI	ENNKRLFDVI	QDNLEKAKKL	LEENEKSKA	
	EGKKEKNIND	ASIIYQKAISL	SGDKYVQLSK	KEPGKFFAIS	KPTPTTTGYG	
	YDTGDSLVD	LYYDNKGKLC	GEIIRKIDAQ	QKNPLKYKEQ	GFTLFFERIYG	
	GDILEVDFDI	HSDKNSFRNN	TGSAPENRVF	IKVGTTFTEIT	NNNIQIWFNG	
	IIKSTGGQDD	SFTINSMQQY	NPRKLILSSC	GFIKYRSPIL	KNKEG	
WP_00910577 7.1 Treponema sp. JC4	MIMKLEKWRL	GLDLGTNSIG	WSVFSLDKDN	SVQDLIDMGV	RIFSDGRDPK	(SEQ ID NO: 46)
	TKEPLAVARR	TARSQRKLIY	RRKLRRKQVF	KFLQEQGLFP	KTKEECMTLK	
	SLNPYELRIK	ALDEKLEPYE	LGRALFNLA	RRGFKSNRDK	GSREEVSEKK	
	SPDEIKTQAD	MQTHLEKAIK	ENGCRITTEF	LYKNQGENGG	IRFAPGRMTY	
	YPTRKMYBEE	FNLIRSKQEK	YYPQVDWDDI	YKAIFYQRPL	KPQQRGYCIY	
	ENDKERTFKA	MPCSQKLRIL	QDIGNLAYYE	GGSKKRVELN	DNQDKVLYEL	
	LNSKDKVTFD	QMRKALCLAD	SNSFNLEENR	DFLIGNPTAV	KMRSKNRFGK	
	LWDEIPLEEQ	DLIIETIITA	DEDDAVYEVI	KKYDLTQEQR	DFIVKNTILQ	
	SGTSMLEKEV	SEKLVKRLEE	IADLYHEAV	ESLGYKFADQ	TVEKYDLLPY	
	YGKVLPGSTM	EIDLSAPETN	PEKHYGKISN	PTVHVALNQT	RVVVALIKE	
	YGKPSQIAIE	LSRDLKNNVE	KKAEIARKQN	QRAKENIAIN	DTISALYHTA	
	FPGKSFYPNR	NDRMKYRLWS	ELGLGNKCIY	CGKGISGAEL	FTKEIEIEHI	
	LPFSRTLDA	ESNLTVAHSS	CNAFKAERSP	FEAFGTNPSG	YSWQEIQRA	
	NQLKNTSKKN	KFSPNAMDSF	EKDSSFIAEQ	LSDNQYIAKA	ALRYLKCLVE	
	NPSDVWTTNG	SMTKLLRDKW	EMDSILCRKF	TEKEVALLGL	KPEQIGNYKK	
	NRFDHRHAI	DAVVIGLTDR	SMVQKLATKN	SHKGNRIEIP	EFPIILRDLI	
	EKVKNIVVSF	KPDHGAEGKL	SKETLLGKIK	LHGKETFVCR	ENIVSLSEKN	
	LDDIVDEKIK	SKVKDYVAKH	KGQKIEAVLS	DFSKENGIKK	VRCVNRVQTP	
	IEITSGKISR	YLSPEYFAA	VIWEIPGEKK	TFKAQYIRRN	EVEKNSKGLN	
	VVKPAVLENG	KPHPAKQVC	LLHKDDYLEF	SDKGKMYFCR	IAGYAATNNK	
	LDIRPVYAVS	YCADWINSTN	ETMLTGYWKP	TPTQNWVSVN	VLFDKQKARL	
	VTVSPIGRVF	RK				
WP_00246084 8.1 Staphylococcus lugdunensis M23590	MNQKFILGLD	IGITSVGYGL	IDYETKNIID	AGVRLFPEAN	VENNEGRRSK	(SEQ ID NO: 47)
	RGSRRRLKRR	IHRLEVRKKL	LEDYNLLDQS	QIPQSTNPYA	IRVKGLSEAL	
	SKDELVIAL	HIAKRRGIHK	IDVIDSNDV	GNELSTKEQL	NKNSKLLKDK	
	FVCQIQLERM	NEGQVRGEKN	RFKTADIIE	IIQLLNVQKN	FHQLDENFIN	
	KYIELVEMRR	EYFEGPGKGS	PYGWEGDPKA	WYETLMGHCT	YFPDELRVSK	
	YAYSADLFNA	LNDLNNLVIQ	RDGLSKLEYH	EKYHIIENVF	KQKKKPTLKQ	
	IANEINVNPE	DIKGYRITKS	GKPQFTEFKL	YHDLKSVLFD	QSILENEDVL	
	DQIAEILTIY	QDKDSIKSKL	TELDILLNEE	DKENIAQLTG	YTGTHRLSLK	
	CIRLVLEEQW	YSSRNQMEIF	THLNIKPKKI	NLTAANKIPK	AMIDEFILSP	
	VVKRTFGQAI	NLINKIIEKY	GVPEDIIIEI	ARENNSKDKQ	KFINEMQKKN	
	ENTRKRINEI	IGKYGNQNAK	RLVEKIRLHD	EQEGKCLYSL	ESIPLEDLLN	
	NPNHYEVDHI	IPRSVSFDNS	YHNKVLVKQS	ENSKKSNLTP	YQYFNSGKSK	
	LSYNQFKQHI	LNLKSQDRI	SKKKKEYLLE	ERDINKFEVQ	KEFINRNLVD	
	TRYATRELTN	YLKAYFSANN	MNVKVKTING	SFTDYLRKVW	KFKKERNHGY	
	KHHAEDALII	ANADFLFKEN	KKLKAVNSVL	EKPEIESKQL	DIQVDSIEDNY	
	SEMFIIPKQV	QDIKDFRNFK	YSHRVDKKPN	RQLINDTLYS	TRKKNSTYI	
	VQTIKDIYAK	DNTTLKKQFD	KSPEKFLMYQ	HDPRTFEKLE	VIMKQYANEK	
	NPLAKYHEET	GEYLTKYSKK	NNGPIVKSLL	YIGNKLGSHL	DVTHQFKSST	
	KKLVKLSIKP	YRFDVYLTDK	GKFKITISYL	DVLKKDNYYY	IPEQKYDKLK	
	LGAIDKNAK	FIASFYKNDL	IKLDGEIYKI	IGVNSDTRNM	IELDLPLDIRY	

Table 4- Cas9 orthologs						
	KEYCELNNIK KRGH	GEPRIKKTIG	KKVNISIEKLT	TDVLGNVFTN	TQYTKPQLLF	
WP_01168147 0.1 Streptococcus thermophilus LMD-9	MTKPYSIGLD LLFDSGITAE LDDSFLLVDD LRLVYLALAH SLENSKQLEE DFRKCFLNDE LLSGFLTVD FKDDTKNGYA KQRTFDNGSI YVGPLARGNS YLPPEEKVLPK LYFKDKRKVT NDKEFLDDSS SRRHYTGWVK LSFKKKIQKA VMGGRKPESI IPAKLSKIDN QAFKDNSID QRKFDNLTKA DENNRVRTV ASALLKKYPK SLADGRVIER QNHGLDRGKP SFTVLVKGTI IELIIEPKY LLYHAKRISN LNSAFQSWQN PRYRDYTPSS	IGTNSVGWAV GRRLKRTARR KRDSKYPIFG MIKYRGHFLI IVKDKISKLE KASLHFSKES NETEAPLSSA GYIDGKTQNE PYQIHLQEMR DFAWSIRKRN HSLLYETFN DKDIEYLHA NEAIEEIIH LSAKLINGIR QIIGDEDKGN VEMARENQY NALQNDRLYL NKVLVSSASN ERGGLSPEDK KIITLKSTLV LEPEFVYGDY PLIEVNEETG KGLFNANLSS SLFELSDGSR TINENHRKYV HSIDELCSSF LLKDATLIHQ	TTDNYKVPSK RYTRRRNRIL NLVEEKAYHD EGEFNSKNND KKDRILKLP YDEDLETLLG MIKRYNEHKE DFYVYLKLL AILDKQAKFY EKITPWNFED YNELTKVRFI IYGYDGIELK TLTIFEDREM DEKSGNTILD IKEVVKSLPG TNQKKSNSQQ YYLQNGKDMY RGKSDDVPSL AGFIQRQLVE SQFRKDFELY PKYNSFRERK ESVWNKESDL KPKPNSNENL VLEFQGISIL RMLASILSTN ENHKKEFEEL IGPTGSEKRG SVTGLYETRI	KMKVLGNTSK YLQEIFSTEM IQKNFQDFLD YIGDDYSDVF DLALLKEYIR AEFEGADYFL PFLAKNKERI AESMRDYQFL GIEKQFNSSL IKQRLSKFEN YLIDDGISNR SPAIKKGILQ RLKRLEKSLK SATEKVYFYS ATVRRVLSYP VGAKEYLDPK DRINYRKDKL NKRGEIHKGN FYIILEFNEN LFELTSRGSA DLAKLGEG	KYIKKNLLGV ATLDDAFFQR TYNAIFESDL FLKLIVGNQA LKAKKLYDAI NISLKYNEV EKIDREDFLR EKILTRIPY FINRMTSFDL DSKQKKDIVR STYHDLNII IFDKSVLKKL NFMQLTHDDA SIKIVDELVK ELGSKILKEN SNYDIDHIIP YQLLKSCLIS LLDEKFNNKK AHDAYLNAV NIMNIFKKS QVNVVKVEE KYGGYAGISN NFLLEKGYK QIFLSQKFVK YVGAKKNGKL ADFEFLGVKI	(SEQ ID NO: 48)
WP_00929301 0.1 Bacteroides fragilis NCTC 9343 Cas9	MKRILGLDLG KGKSIITNAD GNRTTFETYR TLIDGMIDAR EKQKEYYPEI TEQQEREHKL NTQINNSSGY RQDYLDDEFNM CEFESRQIEV GKKRKRRLK KMTKSEVLKL DFKKPVEKVV EGDNTPTGNG LPHLKEGNRY PVVEKILNQ QTTKAHEEYK SREKLSKEF EKFGESGADN DLRNTQYIAK KYKALGLVEY YFNNKNASLD SIKAKNKVIT EKVNASFDMR WLDKEQMRKV LNEYGNDAKK DGKPILDENG	TNSIGWALVN RTLKRGMRRN LRAKAVTEEI ELYNNNLTPG LTDVLKEELR EGIYSKRKR LGAISDRSKE LWEKQAVYHK DIDGKKKIKT ENYSALFEEL LFDNPQELDL EYIKAVFDLL RLIQKMTELY DVACVYAGYR VNVINVIIDI TLLQTEFGLT DIEHIIPQAR SLEHYLNNEI KALSMLNEIS FEDRDGRQIG PNANEHAIKN GNINKTRKKG KIGTVSKSAY PEKVKIVTLE AFSNLDKNPI RNIPVDFVNT	EAENKDERSS LQRYKLRRRT SLEEFARVLL ELCLQLLDAG GKKRDVWAI EAKRENQWR LYFNKQTVGQ ELTEELKKEI VGNRVISRSS NDAEQLELNG NFKTIDGNKT NWNTDILGFN GFEKEYATIL HSESSLTREE YGKPEIRVE NVSRTDILRY LFDDFSNSKT DLFKSGKISK HRVVATSGSV RIKDWTKRND KYFQNGRAIA GVNKNMQQTP RDALLKRLYE AIYTIRKEIS WLNKEKGISI GNNHHVAVYY	IVKLGVVRVNP LTEVLKEHKL MINKKRGYKS KKFLPDFYRS CAKYFVWKEN VNGLEKELSL YQMEMLDKNP RDIIIFYQRR PLFQEFKIWQ SRRLCQEEKE GYALFQAYSK SNEELDNQPY ANVSFQDDYG IANKVLKDRL LARELKKNK KLYKELESCG LEARSVNIEK TKYNKLKMAE TDKLREDWQL HRHHAMDALT PMPLREFRAE RGQLHLETIY NDNDPKKAFA PDLKVDKVID KRVTISGISN RPVIDKRGQL	LTVDELTNFE ITEDTILSEN SRKAKGVEEG DLQNELDRIW YTEWNKEKKG EQLVTVFQEM NASLRNMVYF LKSQKGLIGF ILNNIEVTVV LLAQELFIRD MIEMSGHEPV YKLWHLLYSF SLSAKATHKI MLLPKNSLHN EREELTKSIA YKTLYSNTYI GNKTAIDFVK QDIPDGFIER IDVMKELNWE VAFTKDVFIQ AKKHLENTLI GSGKQYLTKE GKNSLDKQPI VGVRKILIDR AQSLHVKKDK VVDEAGNPYK	(SEQ ID NO: 49)

Table 4- Cas9 orthologs						
	ELEEVVVSFF KTGFNPKEID ILRGITWIDF	EAVTRANLGL LLDVENYGLI RSSKGLDTIV	PIIDKDYKTT SPNLFVRVQKF KVRVNHIGQI	EGWQFLFSMK SLKNYVFRHH VSVGEY	QNEYFVFPNE LETTIKDTSS DNGVIPYQLH	
AOL40912.1 Veillonella atypica ACS- 134-V-Col7a	METQTSNQLI KMWGSRLFEE TFFIRLHESK LMENGTDDIR ITFNCFDTS KPQKEQVKED TYDEKRDELA DKHKEDLVIL REDFYKYTKK LEELKVILDK NIDNGGFSWA DVLPKSSLLY KNRIELFLKD MAEDIITDIT SKKLLKGIDG NAKLAQQQV FVEVARTNKS LANYDDAALR LTKDDSDNL RKYERLTRIA VKAENVSDFR FFKKNGANRT RVTRRLLEQS MTKIKNAYSI PKAKDISIKY ATYIKLLDKY YFMSKLRTPL NSKTTFDVKP	TSHLKDYPKQ GESAVTRRGF YHYEDKTTGH KLFLAVHHIL AISSISNILM KKTCLKAFANL KVGWDEIHII KSLCLKLDRNV IVEGLADSKD CGPKFPFLHT VRKQAGRVTP SEFMLLNELN NNYITKKHKP IFGESKKMLR CDKAGNGAPK NPHDIIDELA EKKKKDSRQK SKKLYLYYTQ VLCERTANAK PLTADDLSGF HNNNFIVKRS YNLAKMFNYD GALADATIYK IVQYTGKKGE RKLCINQLVK DLLRKENKTL YMKMKGNKVD LGITGSRSTI	DYFVGLDIGT RSMRRRLERR SSKHILFIDE KYRGNFLYEG ESGKTKSDKA VLGLSANLID DDCKSVYDAI YNEMFKSDKK KEYILNEIEL VSDGFSVTEK WNFEKIDRE NVRIDGKALA BITGLDGEIK QTLRNKFGSQ TIELMRNDS LSPAVKRAVW RLSDLYSAIK MGRCAYTGNI KSDIYPIDNR IARQLVETNQ LNHHHHAKDA VICTNAQDGK ASVAAKAKDG BIKEIVPLPI VNGFYLYLGG KASSITTSIY ELSSGTGRSKF GVKIHNLDEF	NSVGWAVTNT KLRLKLLEEL DYTDQDYFTE ATFNSNAFTF KAIERLVDTY LFGSVEDIDD ILMSIKEPGL GLHNYVHYIK QTLPLQRIK LIKMLEFRIP KSAAAFIKNL QGVKQHLIDS NDLTSYRDMV LNDETICKLS YNLMEILGDK QALRIVDEVA KDDVLQSGLQ IDLNLQNTDN IQTQKQPFWA SVKATTTLLR YLNIVVGNVY AWDVKTSMT AYIGMKTYS YLINRNATDI KTNDKIYIDN NINTSTVVS IKMTLEEQSI KIINESITGL	SYELLKFHSH FADAMAQVDS YPTIYHLRDK EDVLKQALVN TVFDEVNTPD DLKKLQIVGD TISQSKVKAF QGRTEETS YFVGLKSG YYVGPLNTHH TNKCTYLFGE IFKQDHKKMT RILGNFVDS KLRIRDWGR FSFMECIEEE HIKKALPSRI DKEFGALKSG YDIDHIYPRS FLKHQGLISE RLYPDIDVVF HEKFTRNFRL VKKMMASNDV VFADVTKYGG ELIDYVKS AIELVVP SNKVGIDVFD YLLEVLNLLT YSNEVTIV	(SEQ ID NO: 50)
WP_01338902 6.1 Ilyobacter polytropus DSM 2926	MKYSIGLDIG REKRGSRRRN RAKSLEAKIS KKIMEEKGYL AKQKELGNTH GNCTFESSEK EKEVYLKKEV IGNYTIKKLN LSREDIEILL NNGIKFKRGE VEVARDLAKS KYRLYQEQEG VLVLSGENRK KFTREDNKDW LTAQLRARWR NCKDVIPSIE REEFYNLLSD SSKKISNQIA PEKAFABPFY DVFRKNGKFY YPNDLIEIEF IDTQDGKIQM	IASVGWSVIN RRKKHRLDRI NKEIAQILLH TIGDMVAKDP FTDDVLEKYI RTAKNTWSGE RYEKLRLKILY LSEKCLKSEIE SEEFSGTLNL LLPVVDKDLI YDDRQTIIEKE RCAYSRKELS KSNLLPKKEYF KSRALNDTRY LNKVRENGDL YHADEETGEV KRYLGFNFYE VSKKPLNSIK KPTNSGKRGF TVPIYVHQT GKSLSIKNDN RIGIKTVKNI	KDKERIEDMG KNILCESGLV IAKRRGFKSF KFNTHVRNKA EVFNSQRNFD RFVFLQKLNN LKEEERFGDL EDKSKLDKII SLKAIKKILP ANPVVLRAIS NKKRELENEK LSEVILDESM DRQGRDWDTF ISRFVANYLE HHALDAAVVA YFEEVKDTRF EGFIEKLRPV LKDLEKMQGR LVRGIKVEEK LKELPNRAIN KLTKTEIPEV KKYQVDVLGN	VRIFQKAENP KKNEIEKIYK RKTDRNADDT GSYLFSSFRK EGPSKPSPPY FRIVGLSGKR NYSKDEKQDK EILTFNKSDK YLEKGLSYNE QTRKVVNAAI TKKFISEEFG TDIDHIIPYS VLNVKAMKIH NALEYRDDSP VTDQKAINNI PMPWSGFDLE FVSRMPNRRV QNVGVYVNGG GKPYKDWDLI NLSEVLGYR VYKVKREKRQ	KDGSSLASSR NAYLKSPWEL GKLLSGIQEN LLEDEVRIKQ SEIGQIAKMI PLTEEERD KTEKTKFISL TIESNLKKLE ACEKADYDYK RKYGTPHTIH IKNVKGKLLL RSMDDSYSNK PRKKSLLKE KKRVFMIPGQ SNISRYKELK LQKRLESEN KGQAHQETIR KNRLEEYDDK QASNGSMVRI DGSFEFLYSF GMDTSTGAAT TF	(SEQ ID NO: 51)
WP_00586426 3.1	MKKIVGLDLG KGNSISQTAD RYGKFLNDIE	TNSIGWALIN RTSYRGIRRL CKLPWKDET	AYINKEHLYG RERHLLRRER GSYKFIFQES	IEACGSRIIP LHRILDLLGF FKEMLANFTE	MDAAILGNFD LPKHYSDSL HHPILIANNK	(SEQ ID)

Table 4- Cas9 orthologs						
Parabacteroides sp. 20_3	KVPYDWTIYY KLVEYYSKLV KEFIVTTDL MTVGAYIYDT RDKQLYIACL IDNCPYEENR KETDVDVTQE YRWNYVEDKP DKQIEKALH MRMGKYWSIE ALPLWLACYL ITETLRTVRD LRIKALLTEF IILGKFNQTD EIEHII PQSR FDKPVVLSV ISKVVKSLLS VLPRFIRLNE DAIIACANR VIDKPVWTF QSKGDSWAIR ILAMLELGYD KPIDTSFDDK NILILNKGGK IYETEEIDKD ILSPNDLVVY LIFALPKATA VDRLGNIQV	LRKKALTQKI EKVEDSGERK ADGSPKDKKE LLQKPDQKIR NELYPNNEPR YIDKESGEIK LLPTEADYVT YPCNETHAQI SFANKNNLSE NIDNGTRIRI VYDRHSEVKD IWQQVGHIDE LNPEFGIENV TLKRPTRESEI YFDDSLSNKV EAYEKLHVES NIVRENEQE LTESTRFTSI NIVNYLNNVS QDTLTALQKI KSMHKEVHG TKRIKNYFEE KIKESITDTG HQPIYKVRVY TKKVIRKRSY PTQEEINKGE EIYCNGENCI GSCILTN	SKEELAWILL GKDTWYNVHL GNIKRSFRAP GKLVRTIERK RNSISTRDFC HASIKCIAKS LFEWLNEKKE IARLGKAHIP EFIEQFKNFP NKIIDGEYDE IVKWKTPKDI IHIELGREMK LRYKLWLEQK ICESEINKLK YSHNRSKMKK AISKNVIPCT NTNNTMIPSM ASKNTKITRR TVSFKQNLRV EVNLRMIKT NKDTWQDINP IQQIMLRHLE EKA EKFTVGQ STIPLNVVIE VVMPIDRDRI QNEYGIGSPQ SKNQKAITGE	NFNQKRGYYQ ENGMIYRRTS KDDDWTLIKK YYKNELYQIL HLFIEDIIFY HPLYQEFLRW IDQKAFFKYP KAFLSKEKEE PFKKEYGSYS NIRERVRQKA DLYLKSFKQH NPADKRARMS YRSPYTGEMI DRSLGYEFIK LLMEDIPDQF GGITDRLKKD PLELQKGFNK DLQTLCHCKD INKTNNHYQH SFNEALKKPQ SKIKVYFTEK TKDNDPTLAF KGNKRTKFVE RQKQGLSSAP YKMVDSSGIT SKNQKAITGE	LRGEEEEETPN NIPLDWEGKT KTEADIDKIK KTQSEFHEEL QRPLKSKKSL QFIVNLRIYR PFGFKKTSN TLWHILYSIE AKAIKKLLPL INLTDITHFR SLRNPIVEQV QQMIKNENTN SILELPEDIG PLSKLFTPAY NHHGEKVELA IERQLNDSRY WGINDVWNKI KRIDHRHAM KTDNNGNYKW YENGKKIVSN AIVEMDLKKK ETKDRYFAVR SPDGIEMNR AAKGTNLFFA EDENGNLPKY ANFIPASTAN MVKEICFPIK	NO: 52)
GAP01010.1 Fructobacillus fructosus KCTC 3544	MVYDVGLDIG FRTRRRRLSR AHYYGGYLF HLVKYRGHFL TEIISGEAGY AKLFGKDISD AYDGFVLIGL EIYEQLALAT KQRTKANSVI LRFRIPYYVG NRDKAARDFI KFDSKTKADL SSLSSYNDLK PLDDDQVKKL FMSIINNDKY VLNDIKKAMK KVLKKTLDGN IPQAFTKDD FINRKKYERL KAVAIRSEIT RDSTNVYNEF SEENTSILRK RSDTALYGGY IVNNPAIKKF DADKIASISE LDSDKRIAVI NSEMIYQSAT	TGSVGWVALD RKWRLRLLE TEEETKKFHR SSQEKITIGS GLNKSMADE KDEAKLWKLK LNGADSVSAA DEATIKNGIS PHQLHLAELQ PLVTKKDVEH ERLTGNDTYL INDLKFARKT KTFDAEYLEN SQTHYTGWGR GVQAWITEQN EEPRRVYLEF KNKMSDDRYF LDNRVLVSRP TKAGKYLDGQ ADMRLLVGKI DRYTNDYLN VMMFKNLT SNVYSAYMTL EKILISKPL NSSDEELLE QTILRGLQID GIFETRVKIS	ENGKLARAKG LFSAEINEID SYPTIYHLRQ TYNPEDLANA ALKLFEFDNN LDDEALEEKS MVQLYDQHRE TARELVEESN KILQNGQYY AGGDADNHV IGEKTLPQNS VLSALKDYL EDNQETLEKI LSEKLLDSKI TGSSKLTDFE AREDQTSVRS LYFQQQGD ENARKSDSFA KTGFIAQRLV KHREINSFHH LRQLSSRDEV KTEKDRGPL VRANGNLLI GQLVNEDGNL YDILTSENVK AAYQAPVKII DLL	KNLVGVRLFD SSFFQRLKYS ELMAQPNKRF IEVYADEKGL QDKVAIKTLL QTILSQTDE DRKLLKSLAQ LSKEVKEDTL PFLLDTFEKE ERNEGFEKSR LRYQLFTVLN KAQKGQDVTI IEIQTVFEDS IDERGQKVS KVNELTTSFA VPRYNQLKEK YTGRPINF YTDEVQKQDG ETRQIIKNVA AFDALLITAA RRLKSFGFV NKETIFSPKS KIPISIANQI IYLASNEYRH NRFPFFKKDI SKKVSDWHKL	TAQTAADRRG YVHPKDEENK DIREIYLAIH SWELNNPEQL AGLTGNQIDF EIELFHAVVQ KAGLKHKRF RRLDENEFLP DGQDNKIEEL VTPWNFDKVF ELNNVRVNGK TGLADESKFN KIASRELSKL LDKDKSTSON NKRGIKQSFA YQSKSLSEEA LSQDYDIDHI SLWTSLLKSG SLIEGEYENS GQYMQNRYPD GTMKGNEDW GKKLIPLNSK EVGNLKINDY NAKQLWLSTT DKLSQVRDEF QQSGGIKLS	(SEQ ID NO: 53)

Table 4- Cas9 orthologs						
Bacillus smithii WP_00335419 6.1	MNYKMGLDIG	IASVGWAVIN	LDLKRIEDLG	VRIFDKAHP	QNGESLALPR	(SEQ ID NO: 54)
	RIARSARRRL	RRRKHRLERI	RRLVSENVL	TKEEMNLLFK	QKKQIDVWQL	
	RVDALERKLN	NDELARVLLH	LAKRRGFKSN	RKSERNSKES	SEFLKNIBEN	
	QSILAQYRSV	GEMIVKDSKF	AYHKRNKLD	YSNMIARDDL	EREIKLIFEK	
	QREFNNPVCT	ERLEEKYLN	WSSQRPFASK	EDIEKKVGF	TFEPKEKRAP	
	KATYTFQSFI	VWEHINKLRL	VSPDETRALT	EIERNLLYKQ	AFSKNKMITY	
	DIRKLLNLS	DIHFKGLLYD	PKSSLKQIEN	IRFLELDSYH	KIRKCIENVY	
	GKDGIRMFNE	TDIDTFGYAL	TIFKDDIV	AYLQNEYITK	NGKRVSNLAN	
	KVYDKSLIDE	LLNLSFSKFA	HLSMKAIRNI	LPYMEQGEIY	SKACELAGYN	
	FTGPKKKEKA	LLLPIVPIA	NPVVMRALTQ	SRKVVNAI	KYGPSVSIHI	
	ELARDLSHSF	DERKKIQKQD	TENRKKNETA	IKQLIEYELT	KNPTGLDIVK	
	FKLWSEQQGR	CMYSLKPIEL	ERLLEPGYVE	VDHILPYSRS	LDDSYANKVL	
	VLTKENREKG	NHTPVEYLGL	GSERWKKFEK	FVLANKQFSK	KKKQNLRLR	
	YEETEEKEFK	ERNLNDTRYI	SKFFANFIKE	HLKFADGDGG	QKVYTINGKI	
	TAHLRSRWDF	NKNREESDLH	HAVDAVIVAC	ATQGMIKKIT	EFYKAREQNK	
	ESAKKKEPIF	PQPWPHFADE	LKARLSKFPQ	ESIEAFALGN	YDRKKLESIR	
	PVFVSRMPKR	SVTGAHQET	LRRCVGIDEQ	SGKIQTAVKT	KLSDIKLDDK	
	GHPFMYQKES	DPRTYEAIRQ	RLLEHNNDPK	KAFQEPLYKP	KNGEPPGVI	
	RTYKIIDTKN	KVVHLDGSKT	VAYNSNIVRT	DVFEKDGKYY	CVPVYTMIDM	
	KGTLNPKAIE	ANKPYSEWKE	MTEEYTFQFS	LFPNDLVIRIV	LPREKTIKTS	
	TNEEIIKDI	FAYYKTIDSA	TGGLLEISHD	RNFSLRGVGS	KTLKRFEKYQ	
	VDVLGNIHKV	KGEKRVGLAA	PTNQKKGKT	DSLQSVSD		
Mycoplasma canis PG 14 EIE39736.1 WP_00479473 0.1	MEKKRKVTLG	FDLGIAVSW	AIVDSETNQV	YKLGSRFLDA	PDTNLERRTQ	(SEQ ID NO: 55)
	RGTRRLRRR	KYRNQKFYNL	VKRTEVFGLS	SREAIENRFR	ELSIKYPNII	
	ELKTKALSQE	VCPDEIAWIL	HDYLNKRGYF	YDEKETKEDF	DQQTVESMPS	
	YKLNEFYKKY	GYFKGALSQP	TESEMKNKD	LKEAFFDFDS	NKEWLKEINY	
	FFNVQKNILS	ETFIEEFKKI	FSFTRDISKG	PGSDNMPSPY	GIFGEFGDNG	
	QGGRYEHIWD	KNIGKCSIFT	NEQRAPKYLP	SALIFNFLNE	LANIRLYSTD	
	KKNIQPLWKL	SSVDKLNILL	NLFNLPISEK	KKKLSTNIN	DIVKKESIKS	
	IMISVEDIDM	IKDEWAGKEP	NVYGVGLSGL	NIEESAKENK	FKFQDLKILN	
	VLINLLDNVG	IKFEFKDRND	IKNLELLDN	LYLFLIYQKE	SNNKDDSIDL	
	FIKNESLNI	ENLKLKLKEF	LLGAGNEFEN	HNSKTHSLSK	KAIDELPKL	
	LDNNEGWNLE	AIKNYDEEIK	SQIEDNSSLM	AKQDKKYLND	NFLKDAILPP	
	NVKVTFQQA	LIFNKIIQKF	SKDFEIDKVV	IELAREMTQD	QENDALKGIA	
	KAQSKSKSLV	EERLEANNID	KSVFNDKYEK	LYKIFLWIS	QDFKDPYTGA	
	QISVNEIVNN	KVEIDHIIPY	SLCFDDSSAN	KVLVHKQSNQ	EKSNSLPYIEY	
	IKQGHSGWNW	DEFTKYVKRV	FVNNVDSILS	KKERLKKSEN	LLTASYDGYD	
	KLGFLLARNLN	DTRYATILFR	DQLNNYAEHH	LIDNKKMKFV	IAMNGAVTSF	
	IRKNMSYDNK	LRLKDRSDFS	HHAYDAAIIA	LFSNKTKTLY	NLIDPSLNGI	
	ISKRSEGYWV	IEDRYTGEIK	ELKKEDWTSI	KNNVQARKIA	KEIEEYLIDL	
	DDEVFFSRKT	KRKTNRQLYN	ETIYGIATKT	DEDGITNYK	KEKFSILDDK	
	DIYLRLLRER	EKFVINQSNP	EVIDQIEII	BSYGKENNIP	SRDEAINIKY	
	TKNKINYNLY	LKQYMRSLTK	SLDQFSEEFI	NQMIANKTFV	LYNPNTKNTTR	
	KIKFLRLVND	VKINDIRKNQ	VINKFNGKNN	EPKAFYENIN	SLGAIVFKNS	
	ANNFKTLSIN	TQIAIFGDKN	WDIEDFKTYN	MEKIEKYKEI	YGIDKTYNFH	
	SFIFPGTILL	DKQNKEFYI	SSIQTVRDII	EIKFLNKIEF	KDENKNQDTS	
	KTPKRLMFGI	KSIMNNYEQV	DISPFGINKK	IFE		
Odoribacter laneus YIT EHP49880.1	METTLGIDL	TNSIGLALVD	QEEHQILYSG	VRIFPEGINK	DTIGLGEKEE	(SEQ ID NO: 56)
	SRNATRRRAK	QMRQYFRKK	LKAKLLELL	IAYDMCPLKP	EDVRRWKNWD	
	KQKSTVRQF	PDTPAFREWL	KQNPYELRKQ	AVTEDVTRPE	LGRILYQMIQ	
	RRGFLSSRKG	KEEGKIFTGK	DRMVGIDETR	KNLQKQTLGA	YLYDIAPKNG	
	EKYRFRTERV	RARYTLRDMY	IREFEIIWQR	QAGHLGLAHE	QATRKKNIFL	
	EGSATNVRNS	KLITHLQAKY	GRGHVLIEDT	RITVTFQLPL	KEVLGGKIEI	
	EEEQLKFKSN	ESVLFWQRPL	RSQKSLLSKC	VFEGRNFYDP	VHQKWIAGP	
	TPAPLSHPEF	EEFRAYQFIN	NIIYGKNEHL	TAIQREAVFE	LMCTESKDFN	
	FEKIPKHLKL	FEKFNFDTT	KVPACTTISQ	LRLKLFHPVW	EEKREEIWHC	

Table 4- Cas9 orthologs						
	FYFYDDNTLL YLKKGAYAYST EVIRKIKDYL NPIVQQGLNE EKQSRQIREN CPYTGKTLNI GELTPYDFYQ NEFISRQLND QSAPDITFPL LTGEVERKVF DGQIVLKGR NNSKLT SQQV TPIKNAPPGV TVESCDPTLI RHHAIDAIVI DVRQSVVPLL YGQRTAPGAT IDITQEFNIP DNINQYVNPR EGRNIVSILQ FRHHLASTLN	FEKLQKDYAL AVLLGGIRNS VHNRFGFAKN LRRTVNKLLA EKKNEAAKVK SHTLGSDNSV KDPSPEKWGA TRYISKKAVE PVSATENHRE RCKGMQEFQT EKGVFVCNQL QLFGRVREGI GGGQIILTGD PIELSAPKTS ALSSQSIFQR VSYKQNPCTL EKSYHIRKDI SNAFFKEGVY NNHHVMYQD INDTFLIGLK NEREEFRIQS	QTNDLEKIKK FGKRFEYFKE DRAFQKLYHH TCREKYGPSF LAEYGLKAYR QIEHIIPYSI SSWEEIEDRA YLSAICSDVK YYVITNEQNE DVSDGKYWRR KQKLTGTPD FRCHNYQCPA VDDKGIFHAD KGENLIEGNI LSTYNARREN CKISKTLYKD RELKTSKHIG RIFLPNKHGE ADGNLKEEIV EEPEVYRND LEAWKRANPV	IRLSESYGNV YEPEIEKAVC SQAITTQAQK KFDHIHVEMG DNIQKYLKYK SLDDSLANKT FRLLPYAKAQ AFPGQLTAE VIRLFPKQGE IKLSSSVTWS GSYWISLPVI SGADGNFWCT DDLHYELPAS WVDEHTGEVR KKRGLDSTEH GKKIHS CGNA KVVDITIRQM PVPIKKIRMK SFWSVIERQN LSTLSKHLR KVQIDEIGRI	SLKAIRRNIP RILKEKNAEG ERLPETGNLR RELRSSKTER EIEEKGTV LCDATFNREK RFIRRKQES RHLWGLNNIL TPRTEKGELL PLFAPKPISA SQTFFKEGESV LDTDTAQPAF LPKGKYYGIF FDPKKNREDQ FPSPWP GFAQ VRGQLHKETV LLKHLQENYH EELGNAERLK QGQPIYQLPR VQKLSCMYYT TFLNGPLC	
Akkermansia muciniphila ATCC BAA- 835 WP_01242103 4.1	MSRSLTFSFD RREYRLRRN ALKGHRTLAP HAQDLMCHKG IVEKEVRRIL RRYRGSLFPG KLSKVPTANC TKASLEKAIS QILSPSVYRI KEADYADTPL GELKAHDGCL DLIQDFADGQ VNRLKRKLPG IVPHSFRQSN ICSLNNYREL MKEIGMTEGM VFGVFKELCP HNGLLRVLA NIREQLMEQR GKKEKNQVKA IKVFKRIMAL KGGVKLDLQR R	IGYASIGWAV IRSRRVRIER IELWHVLRWY TATMAETICR ELSAPLIPGL QLIPRFDNRI PEFYEYRMAR SRLGKETETN AANRLRRGKS KPKYATGRAP YCLLDTDSSV KDRISRVCE KALSANLIRK ALSSLVLTWP VEKLDDKKGH MTQSSHLMKL EAADPD SGKI MRRIPKELIP VIQHV PADMG SKLVGVFPEG KEQNGGKPV AHCAPKNKT	IASASHDDAD IGRLLVQAQI AHNRGYDNNA ELKLEEGKAD TAEIIELIAQ ISRCPTWAQ ILCNIRADGE VSNYFTLHPD VTPNYLLNLL YARTVLKKVV NQHQKERRLD VGKELTTFSA CRIAMD MNWT GVNRMKGQRT EDDRRRKKKR ACKSIKTSLP LKENLRSLTH QVRPVANQRH GALLKETMQR PSKLKALKAA ILKKGMLIHL HECNWREVDL	PSVCGCGTVL ITPEMKETSG SWNSLS SEDG APMEVSTPAY VYEAELKKGN PLSAEIRREL SEEALYLNPA KSRGESGEAL EEILDGEDPT TMTNNHLVRH MDSKKIQREL CPFTGATYGD GYDFVEQEQE KALLMVRGLS DAHIDMIPGA LHHALDACVL YVLNDDGRMM VLSVDGSGED IEIDGNYGVA TSSKDPKHAG ISLLKKYQMK	FPKDDCQAFK HPAPFYLASE GNGEDTERVK KNLNTAF PRL VLLQHG I KLA SEQSARERAE MNQARQEGKL VEVLQRSGIG EKKIEKESKK RPARGEAHPD RMLILDRLK HELENLELEH NPVPDKPNLH HKHQSQNHEA VTAEVRKAWD GLIPIYIIPAH LRDLSASLKE AMVSLSKKKD LDPKPVVIRH VWRIBSIQDS RYPTSYTGTP	(SEQ ID NO: 57)
Dinoroseobact er shibae DFL 12 = DSM 16493 WP_01217707 9.1	MRLGLDIGTS LAVDRRAARA FALRAAGLDE ARLDMEMMAN EEAGYDFYPD KEPEVGLCLF ARPLTREERD TGVRDAIACD AALVDWLTEA TYADAVKACG FGRITNPTVH DIKIRDTTE	SIGWWLYETD MRRRRDRYLR PLPLPHLGRA GARTYGEFLH RRHLEEFHK SGHHGVPPKD QVIHALDNKK PLRASPAHPD HGLDRAHAEA WHHSDGRTGE IGLNQLRRLV AAKKRSEKLE	GAGSDARITG RRATLMKVLA LFHLNQRGFG KRRQKATDPR LWAAQGAHHP PRLPKAHPLT PTKSLSSMVL RFGPRWSILD TAHAPLPDGY CFDRLPYGGE NRIIETHGKP ELEIEDNGRN	VVDGGVRIFS ETGLMPADPA KSNRKTDRGD HVPSVRTRL ELTETLRDLL QRRVLYETVN KLPAKAVLK ADAQWEVISR GRLGLTATTR VLERHVIPGS HQIVVELARD RMLLRLWEDL	DGRDPKSGAS EAKALEALDP NESGKIKDAT IANRGGPDGK FEKIFFQRPL QLRVTADGRE LRDGERFTLE IRRVQSDAEH ILYQLTADV YHPDDDDITR LKKSEEQKRA NPDDAMRRFC	(SEQ ID NO: 58)

Table 4- Cas9 orthologs						
	PYTGTTRISAA TPWQAWGDTF DTQYLARISR AVKAKNRTDH DTPPPWEGR TAYSIVDDIH KTGKDFENAL GGSNHLFEIW MVALERDGR GARPAIASGI	MIFDGS CDVD HWHAI AANLK SYLDTLFTKG RHHAIDA AAVI DDLRLVRLDRI VASRTDLLSL RYFASKPGPY RLPDGEIEAQ VVGHVQKMDI RRVSVDEIGR	HILPYSRTLD NLPENKRWRF GHVWVVPGRF AATDPGLLNR IVSHRADHGR KPAQLLDEPG QAIRRVRIIK VITSFEAHTL ANGLFIVPHN LRDGGTRPI	DSFPNRTLCL APDAMTRFEG TEMLRRHWGL ISRAAGQGEA IDHAARKQGR RSGQVRDPQL PLQAQARVPV EGEKRPHPAA EANADTRNND	REANRQKRNO ENGFLDRALK NSLLSDAGRG AGQSABLIAR DSTAGQLHQE RKALRVATGG PAQDPKAYQ KRLLRVHKGD KSDPFKWIQI	
Wolinella succinogenes DSM 1740 WP_01113928 9.1	MIERILGVDL SPNKARREAR KANRADVWEL KVKKAGVVL LVVEVEAIFV FPEERRAPKS AVSREKVEFR ELEFDKVEAE YYKDEGQKRR GARYDEAVLM LVRKYGAFDR FQVPLTRKNI RSADDSFANK VRTGKGKIDR NSHTKAPVQV VQKLESEYRF VKRLLISRPP FKADRVASAN SLGTNPEAMD MVLSSINNSP SQEGLRPPRK RSACKLVKK	GISSLGWAIV GIRRVLNRRR RHDGLYRLK QNFEAAGCRT AQQEMRSTIA APTA EKFIAT HLRKFLDLSD KKAWISLRGA ELTKLPLEAE LGVPKKEKSA VHFELAREIN LKKRLYIQD VLCLARANQQ LLKKNFDENS RSGKLT SVLR KETHREKERP RARVTGQAE QKNFYETSTI ENFFKFSIFK CEGFTCTPVS EFEC DQGVKF	EYDKDDEAAN VRMNMKKLF GDELARVLIH VGEWLWRERG TDALKAAYRE SKFFSTVIID NEIFKGLHYK AKLREALGNE MVERLVKIGF ILPPLNKTDI TKGEIEDIKE GRCAYTGDVI KTDRTPYEFW EMAFKDRNLN YQWGLESKDR KLAVPLANFR QTAKPYPRIK PRVDVYHKKG DDLISIQTQG MDKKHKDKCK ALDVKKYQID	RIIDCGVRLF LRAGLIQDVD IAKHRGYKFI ANGKKRNKHG IAFFVRPMQR NEGWEQKIE GKPKTAKKRE FYGRFVALGK SDFLKL SLKA DILNPTVIRA SQRKNEKERK ELERLFDEGY GHDAARWNAF DTRYMARAIK ESHTHHA VDA QVKNKKKWR KFHLVPIYLH TPKKPAKIIM LCPEENRIAG PLGYYYEVKQ	TAAETPKKKE LDGEGGMFYS GDDEADEESG DYEISIRDL IEKVMGHCTY RKTLELLDF ATLFDPNPT HADEATKILT IRDILPAMES FAQFRKVANA EAADWIAETS CEIDHILPRS ETRTSAPS NR TYCEQYVFK IIIAFSTQGM NTETVKEGVE APIDEEKFES EMVLNLPNL GYFKNMHGAN RCLQGFLDYW EKRLGTIPQM	(SEQ ID NO: 59)
Parasutterella excrementih minis YIT 11859 WP_00886484 3.1	MGKTHIIGVG SKSRTAVRHR LSGYLKRRGY EAIANSFETT AILDASKYIA WRIIGNISNL DKEWSASQKQ LYLNPALSS SDVLPDSYR VLSGHQLEEF NVIVGQALGV FSEDYKFVKT EAQSKFDNLY CRLPADCVRP IEANSFNFTA RADSHGICAY GNQEKNNIY YFDLLSEKER SIFSKVRQAL KVQPVASHSI VIQNTPRNFS EEKPNRISIG SKVAGSKFTA ENLAKQKGCL	LDLGGTYTGT VRSYKGFDLR ARTEAETDTS NLQSLGHKHR QERAVRWYFN IIQSLEQSGD EYGEKWKSWA LAYILQRAFD LEFANRYQE SPAEGTDFIE ALKEGKTEKE SLAQLYNLIE FDGFIRKAID SLTDLKYIQL TGRPLDDVGE LLSNLAKNYL ACARHALFLN AAWTQETGNE DAMCIYLAAC DKTNIANSPI GKDPFSILSV EEDKAAKILE KKVEYSSKEF	FITSHPSDEA RRLLLLVAEY VLES LDPSVF KEADFKKYIK SKYLS DILQD DAKFEQGQEQ VLDVLAGLDP NKFAGAYPLL RSIALDECSI TAKAKNGLWF EIWNSKVKGR LSKKFAAVIK TERNGF SKVS RNSWEVAKRI KEQKLKKKLE IDHIIPRSLT AAVFGTSDLS SDSEARRAVI LIFDAISVPA SDPFKTKRMG FKETIYAERF LGAYLDKAPS ALHFVTVKQD KFKGSLIIPA	EHRDHSSAFT QLLQKKQTLA SSAPSFTNFF TSFPEYSAKE MKRDSRITRL LDAVKLKNVL DRTIPPYEDQ TEDLTEILKN RRTAEDFENG PENALLERAD STVRSICNAI VLKMVSEVVP LAAHLEN AWR ABEVKKSVD DIQRNEENQE LKKSESIYNS QITNEIESTV DVLGSRRKAS ADSSEMRKRF SQLAIYEPIN LDIIVSRGEI SEKEKLT IYR VAATVSDLIK AVEWGKVLWN	VVNSEKLSFS PEERENLRIA NDSEPLNIQW ILANYEGRR SEAFGSTDNL VRALKYLRSD NNRRPPEDQT TDRKSRIRKIR VVIKNEKLED LHPPMKNKIL ENERKTYGPy FIGKELRLSD MTMTDGS AQC TNGTVKIPVA KRWLSKEERI EVNLIFVSAQ LQLKAAGRLG VNGTQAWFVR AEYRPEFRKP FDNLFTGSCQ FIGYPSNMPF VVKNAKAFELF SKKELSKDSI VFKENTAEEL	(SEQ ID NO: 60)

Table 4- Cas9 orthologs						
	KDENALRKAL FGDFVYQSQD HDRSISMSEF SISLGSSEFK PRVIFNYIVG	EAAPSSSFGT TNNLYSSFPV REVYNKDDLM SVSEVFDNKI GAASSLKEIF	RNLHSAKARV KNGKLDWSSP RIELAQGTSS YTENAEFTKF SEAGKERS	FSLPVVATQS IIHPALQNRN RRYL RVEMPG LPKPREDNKH	GAVRIRRKTA LTAYGYRFVD EKFLAWFGEN NGTIFFEVLVG	
Streptococcus sanguinis SK49 WP_00293358 9.1	MTKFNKNYSI IGSTTFVSAQ YRRLDESFRV ADKNSPVADI CQEVFDLEIS QLLQLLFGLK QNLSEQDYLD NFYDYISGKI KIIERQKDY SSDILDDNDE VLPKRSLIYE KRVNAKSLIK ELLDSEIYQK SKLRYSGWGR SFEPKIKDIQ GYPPHNIVIE HSDEQLQSEK SFLPIDSIDN TKYQRLTTSE TKIITLKSQ KLAPELIYGH NRDLPIIKDV KTKALCLDTS DYEKFNPNPF LFESKSNLHK FILKHKEEFD TIKYFYDNFI IHQSITGLYE	GLDIGVSSVG PAKGTRVFRV LGDKSEDLQI REVYMAISHI DESKNIADIF TKFKDCFELE ILLSGMLTYT IFGRKTQNGF TGIEGAEYFL PFLLENKDKL DTRNGKIRPW EVMLQNELTR YLSDNHKDLN ELEEIIVKVI YSAKLLDIR SKSTIEDDIF MARENMTTEE LYLYYLQNGK KVLTHRENNQ RTPDGVLTES ITNFRNTFHI HAHFNREHEN IYNSQINFVK IYGGYGPMNS QFLNDTSENG ATQFKLTKTQ NISNQLSAFS KMFSFVKSGA TRIDLSKLG	YAVVTEDYRV NRRRIDRRNH KQPFPGDKEL LKYRGHFLT KSENREQEV EEDLNFSKE GATHARFSAT DVKETKGYV NKISDGTFLR LSILTFKIPY VYKDKYGA AIEIVSGVEK VFDDKSIKN DEDTGFNLLQ DEIKKLAGSP GQKAKTRKT DMYTLDKTGS QKLNNIPDKE MKAGFIERQL AKIRELNDYH KATLRKHLYS RTMIKKGAFY ALSIIIIAER FLKKNNINRV NELFFHMKRL QKMLGNTTSL PKDINDFFDN D	PAFKFKVLGN RITYLRDIFQ ETAYHKKYPT DKINPNINM NYDENLENFL MVERYEEHRK GYITNMVLT KLRTSDNGAI YVGPLAKGSN TRDAFITNLI HFFDSELQRN GKSFNSTLKT YLTFFFHGLE FLRNDENRN AIKRGILNSI KLESALKNIE PAPLYLDQLD TVANMKPFEW VETRQIIKHV HAHDAYLAVV NIMRFFNNPD NQNPVGKFNK FNEKKGKIET LGFYRIPKYS LTKSNLMDLK KNLIKGYNER KCTVARMRPK	TEKEKIKKNL KEIEKVDKNF IYHLRKHLAD QNSWIDFIES LLKKDKSIFK GSLEEDFSDV DLQRFKFFIK NPQKQKTIQQ PNQIHAYELE SRFAWIKRAT GNDIILLNEK IINGLFKNNS YNDLKTIFSE ILDEEKINQL LTKLISDNTL KIVDELVQII NSLLENGKVP QYEVHDIIPIY KLYNAKLISQ ARILDNRFS VGQTLKLVYP SKVSKDIWDC QLAANNRYPL VKEFHDIFII LMQKIDGTRM SKSAIKESQN KIKEIDIRDE PDKLLNATL	(SEQ ID NO: 61)
Actinomyces sp. oral taxon 180 str. F0310 AOL41039.1	MLHCIAVIRV VGLRSIGFCA ARARRRGRAE IEDETERROC VEDRTGLQFS KYMQSDLVAE LDELQALALDP EELNRVARYL YPPVDDTIVR DEEVNELISS DLSQAITRLY GVPQTVNIEH PFRSDQVRY NLAACHSCN TEKAMGKRKS DRPEGCAAVQ RNHAMDALVI QDRWESWIGD KARRGSKRPR ADLNRGLKLR GPKEKVYAL NGSAQIGWL	PPSEEPGFFE VEVDDDEHPI KQRLKKLDVL LSVAMAHIA EEVTQGEI LRQICRTQV AAKQPRERA LNHTESESP VMSAEVDWLA ATAEDMLKLE GVDPGWVPT TREGLKSASL EILDLDQDCAC SAKGGALFQ EVISRLKTEM VNAYSGRLTA ALMTPAIART VEYACDRLNE PQRYVLGDAL GKRISADFPI LRVCAIDLCC VLGDEIQIDP	THADSCALCH RILNSVHVH LEELGWGVSS HRGWRNSFSK TLEHGDGVT SETTFEKLVL HPAQKFVKV WDDVARKLEV DWDWCANDES LLAKKLPSGR APIEAPVGNP LEEEERERWER LYCGNEINFQ WVKRGDCPSG PYEEFDGRSM CARRAAHVDK IAVREDRREA LIDADKIPVT PADVINRVT DYFPTDSPAL IDCDDLFEVE TKFQKQSIGK	HGCMTYAAND DAGTGGPGET NELLD SHAPW VDTLLLEQAP IRGFVRKGGK SIFHSKEPAP ATLANMRIRE PRHRLRGSSR RGHMIDAISN VAYS LKTLRE SVDRLKQVA FEARREIRQK TFEVDHIIPR VSENAIKRV ESVAWMAIEL RVRLIRLKG QQLTRAFESW ENLRLRNSGK PGLWTALVRA AVQGGYVGLE LKPSSISMRT FLKECGPVSS	KAIRYRVGID ESLRKRSQVA HIRKRLVSEY SDRMQGLKER ATKVHGVLEG SAARQREVRG QSAGERSLTS ASLETGGGLT GCGSEPDDVE VTAAILETGD RWLKFASKRW EMYKRLGISG VDASSDSRRT RSWSKDRLGL KKRIEYGFNS DGHKKNRFD KNFLGSEERM LHADQPESLK PGFDSQLGLP FHARLYRII ADAKLKEAMG WRVSALDTPS	(SEQ ID NO: 62)

Table 4- Cas9 orthologs						
	KITLKPRLLS QSGLRVIRR	NEPLLKTSRV NALGEVRTSP	GGHESDLVVA KSGLPISLNL	ECVEKIMKKT R	GWVVEINALC	
Rhodovulum sp. PH10 WP_00838698 3.1	MGIRFAFDLG GQSLATMRRV TDPYHLRAKA EGSKRLAETL FYPTREMVRA GCCTFEPSE LASALLAGKS DHYGAAWHGL SIPLADGYGR WHHSDEIRDGV HIGLNQLRKV EENERRTAIL AELFTSEVEI PAWNDIVARA KIYLGKICDP DAVVIGVLTR STITVAVKPE EVDVRDRAL RILKPDASVV EVNRPGRPE ANVRVLSPHN VVTLLRSNV	TNSIGWAVWR PRQSRKRRDR LDESRLPHEM AATNCRTLGE EFERLWTAQS RLPRALPSVE LTFKAVRCTL SFAEKDTFVG LGRTANTEIL ILDRLPYYGE VNRLIAAHGR AEHGQORDTAE DHILPVSLTL AKLPPNKRWR DRVYVTPGTL GMIQRIAHDA HGKGGALHED RARLGALAAP TIADRRTGVP WEVKKLGGKL DGGKLQDRHA	TGPGVFGEDT FVLRRLDLLA GRVIFHLNQR FLWSRHRGTP RFAPDLLTPE ARGIYERLAH KILPHALVNF KLLDEADEER AALVEETDET ILQRHVVPGS PDQIVVELAR NKIRLRLFEE DDSLANRVLC FDPAALERFE TGLLRARWGL ARAEDQDLDR TSYGLVPDTP FRDESGRVRD YRAVAPGENH VMRLHKGDMV DADDPFRWDL	AASLDGSGVL ALRKAGLFPV RGFRSNRKAD RTRSPTRIRM RHEEIAIGILF LRITTPGPVSD EEAGEKGLDG LIRRLVTENR GTVVTYAEAV GEPEEKNEAA ELKLNREQKE QARANAGIAL RREANREKRR REGGFLGRQL NSILSDSNFK VFRDVPVPFE PNAALGNLVV AKGLAQALEA HVDIVQMRDG ELSDKDGQRR ATIPLLKDRG	IFKDGRNPKD DVEEGRRLAA RQDREKGKIA EGEGAKALYA RQRDLAPPKI RGLTRPERDV ALTAKLLSKP LSEDAARRCA RRAGERTGRN RWGRLANPTV RLDRENKRN CPYTGRAIGI QTPFQAFGAT NETKYL SRLA NRSDDRHHAV DFRDHVRERV RKPIRSLTAG FGAENGIRRV SWRGFAASVF VKVVQQIEIS CVAVRVDPFG	(SEQ ID NO: 63)
Bifidobacteriu m bifidum S17 WP_01336299 5.1	MSRKNYVDDY TAESRRMART PDDENNADCW IYLAIHHMVK SDIEQDDENK QRAIVKAIFA QIVDSGLLDD ANWNLIKLQL KANVFFKKMI NELKQIIKKQ DLQKSDNSEN LGEPTLPKNS LCQRLFMKGQ CRIFGEKYVN VNTHYTGWGR ITDKQLGFSD DISKAVGKRP ADELNACSDK QAVTQNDSD SRVKFERLTR VIGLNAELTK ADGEVSDGAQ EQKRVNPRTG AATDPKGIKG QEYSLGDRP IKSAGELNNA CLYPWHRFTE VGMTGSWRRM NSKYRTNEKR	AISLDIGNAS TRRRYSRRRW YGSVLFDSNE YRGNFLVEGT LVAPANGQLA GLEGNKDLDA EEREFLLDRM RTKENEKDIN ETSDLSETDK GKYYPFLDLA HWMVRKKKGE LLYQEYEVLN TVTKRTAENL EHRDLMEKIV LSRKLTLTKA WIEQENLGAE SRIFLELADD DLQDDRLFLY NRVLVARAEN QNEFSEREKE EMHYRLGFSH NSYNQYLRDY EVVWSEEDKD IPFDGAKQDT SDDELKRVLA QQLWLPYEEY EELADLHVAF NNKSGYTFSD LPTLSGASQP	VGWSAFTPNY RLRLLDALFD QDKRFYEKYP LESSNAFKED DALCATRGSR KIFVSKEFSS GQYNAIALKQ ENYGILVGWK NRLIHDIEED FEKDGKQINK ITPWNFDEM ELNNVRLSVR LRKEYGRTYE ELQTVFEDKE GECKISDDFA NGSSLMEVVD IQPSGRTISR YTQLGKDMYT ARKTDSFTYM RFVQRLSVET KNRDINDYHH LRGYREKLSA YLRKVMNYRK SLYGGFSSAK CYFDDLSQK DKLPEDEKKN EDEFIFQSPS	RLVRAKGHEL QALSEIDPSF TIYHLRKALM ELLKLLGRIT SMRVDNALEA ENKKILGIYF LLGRSTSVSD IDSGQRKSVR KLFPIQRDSD IEGLLTFRVP DKDASGRKFI TGNHWNDRR LSGLSDESKF TLLHQLRQLE PRKHSIIIEIM DLRVSPKVKR KSRLQDLYRN GEELDLDRLS PQIADRMNRF RQIMKNVATL AQDALCVGIA EDRKQGRAFG MLVTQKVGDD PAYAVLIESK ILLRHVPKM GSLEKDDLK GLFEKRVTVG	IGVRLFDPAD LARRKYSWVH EDDSQHDIRE RYEMSEGEON LSAVNDLSRE NKSDYEEKCV SKCASYDAHR GESAYENMRK NGVIPHQLHQ YFVGPLVPE ERLVGTDSYL MRLGREETL TSSLSTYGKM GISEADCALL RAEDRNLM EI GIIQSIRLID ANLGKEFKGI SAYDIDHIIP WQIILLNGLI MRQRYGNSAA GQFAANRGFF FIVGSMRSQD FGALYDETRY GKTRLNVNVTM LIRYGGGLMV LLDSILGSVQ ADAKTANLSI ELKRKAKKEV	(SEQ ID NO: 64)

Table 4- Cas9 orthologs						
Barnesiella intestinihominis YIT 11860 WP_00886324 5.1	MKNILGLDLG	LSSIGWSVIR	ENSEEQELVA	MGSRVVSLLTA	AELSSFTQGN	(SEQ ID NO: 65)
	GVSINSQRTQ	KRTQRKGYDR	YQLRRTLLRN	KLDTLGMLPD	DSLSYLPKLQ	
	LWGLRAKAVT	QRIELNELGR	VLLHLNQKRG	YKSIKSDFSG	DKKITDYVKT	
	VKTRYDELKE	MRLTIGELFF	RRLTENAFFR	CKEQVYPRQA	YVEEFDCIMN	
	CQRKFYPDIL	TDETIRCIRD	EIIYYQRPLK	SCKYLVSRCE	FEKRFYLNAA	
	GKKTEAGPKV	SPRTSPLFQV	CRLWESINNI	VVKDRRNEIV	FISAEQRAAL	
	FDNFLNTHEKL	KGSDLLKLLG	LSKTYGYRLG	EQFKTGIQGN	KTRVEIERAL	
	GNYPDKRLL	QFNLQEESSS	MVNTETGEII	PMISLSFEQE	PLYRLWHVLY	
	SIDDDREQLQS	VLRQKFGIDD	DEVLERLSAI	DLVKAGFGNK	SSKAIRRIIP	
	FLQLGMNYAE	ACEAAGYNHS	NNYTKAENEA	RALLDRLPAI	KKNELRQPVV	
	EKILNQMVNV	VNALMEKYGR	FDEIRVELAR	ELKQSKEERS	NTYKSINKNQ	
	RENEQIAKRI	VEYGVPTRSR	IQKYKMWEES	KHCCIYCGQP	VDVGDFLRGF	
	DVEVEHIIPK	SLYFDDSFAN	KVCSCRSCK	EKNRRTAYDY	MKSKGEKALS	
	DYVERVNTMY	TNNQISKTKW	QNLLTPVDKI	SIDFIDRQLR	ESQYIARKAK	
	EILTSICYNV	TATSGSVTSF	LRHVWGWDTV	LHDLNFDYRK	KVGLTEVIEV	
	NHRGSIIRRE	QIKDWSKRFD	HRHHAIDALT	IACKQAYIQ	RLNNLRAEEG	
	PDFNKMSLER	YIQSQPHFSV	AQVREAVDRI	LVSFRAGKRA	VTGPKRYIRK	
	NRKRISVQSV	LIPRGALSEE	SVYGVIVHWE	KDEQGHVIQK	QRAVMKYPIIT	
	SINREMLDKE	KVVDKRIHRI	LSGRLAQYND	NPKEAFAPKV	YIDKECRIPI	
	RTVRCFAKPA	INTLVPLKKD	DKGNPVAWVN	PGNNHHVAIY	RDEDGKYKER	
	TVTFWEAVDR	CRVGIPAIVT	QPDITWDNII	QRNDISENVL	ESLPDVKWQF	
	VLSLQQNEMF	ILGMNEEDYR	YAMDQQDYAL	LNKYLRYVQK	LSKSDYSFRY	
	HTETSVEDKY	DGKPNLKLMS	QMGKLRVSI	KSLGLGNPHK	VHISVLGEIK	
	EIS					
Aminomonas paucivorans DSM 12260 WP_00629985 0.1	MIGEHRVGGC	LFDDHWTWN	GAFRLPNTVR	TFTKAENPKD	GSSLAEPRRQ	(SEQ ID NO: 66)
	ARGLRRLRR	KTQRLEDLRR	LLAKEGVLSL	SDLETLFRET	PAKDPYQLRA	
	EGLDRPLSFP	EWVRVLYHIT	KHRGFQSNRR	NPVEDGQERS	RQEEEGKLLS	
	GVGENERLLR	EGGYRTAGEM	LARDPKFQDH	RNRAGDYSH	TLRSRLLEE	
	ARRLFQSQRT	LGNPHASSNL	EEAFLHLVAF	QNPFAFGEDI	RNKAGHCSLE	
	PDQIRAPRRS	ASAETFMLLQ	KTGNLRLIHR	RTGEERPLTD	KEREQIHLA	
	WKQEKVTHKT	LRRHLEIPEE	WLFTGLPYHR	SGDKAEKLF	VHLAGIHEIR	
	KALDKGPDPA	VWDTLRSRD	LLDSIADTLT	FYKNEDEILP	RLESGLSPE	
	NARALAPLSF	SGTAHLSLSA	LGKLLPHLEE	GKSYTQARAD	AGYAAPPDR	
	HPKLPPLLEA	DWRNPVVFRA	LTQTRKVVNA	LVRRYGPPWC	IHLETARELS	
	QPAKVRRIE	TEQQANKEKK	QQAEREFLDI	VGTAAPGPD	LKMRLWREQG	
	GFCPYCEEYL	NPTRLAEPGY	AEMDHILPYS	RSLDNGWHNR	VLVHGKDNRD	
	KGNRTPFEAF	GGDTARWDRL	VAWVQASHLS	APKKRNLLRE	DFGEEAEREL	
	KDRNLTDTRF	ITKTAATLLR	DRLTFHPEAP	KDPVMTLNGR	LTAFLRKQWG	
	LHKNRKNGDL	HHALDAAVLA	VASRSFVYRL	SSHNAAWGEL	PRGREANGF	
	SLPYPAFRSE	VLARLCPTRE	EILLRLDQGG	VGDEAFRNG	LRPVFVSRA	
	SRRLRGKAHM	ETLRSKPWKD	HPEGPRTASR	IPLKDLNLEK	LERMVGKDRD	
	RKLYEALRER	LAAFGGNGKK	AFVAPFRKPC	RSGEGLVRS	LRIFDSGYSG	
	VELRDGGEVY	AVADHESMVR	VDVYAKNRF	YLVVPYVADV	ARGIVKNRAI	
	VAHKSEEEWD	LVDGSFDFRF	SLFPGDLVEI	EKKDGAYLGY	YKSCHRGDGR	
	LLLDRHRMP	RESDCGTFYV	STRKDVLSMS	KYQVDPLGEI	RLVGSEKPPF	
	VL					
Ralstonia syzygii R24 CCA84553.1	MAEKQHRWGL	DIGTNSIGWA	VIALIEGRPA	GLVATGSRIF	SDGRNPKDGS	(SEQ ID NO: 67)
	SLAVERRGPR	QMRRRDRYL	RRRDRFMQAL	INVGLMPGDA	AARKALVTEN	
	PYVLRQRGLD	QALTLPFEGR	ALFHLNQRRG	FQSNRKTDR	TAKESGKVN	
	AIAAFRAGMG	NARTVGEALA	RRLEDGRPVR	ARMVGQKDE	HYELYIAREW	
	IAQEFDALWA	SQQRFAEVL	ADAARDRLRA	ILLFQRKLLP	VPVGKCFLEP	
	NQPRVAAALP	SAQRFRLMQE	LNHLRVMTLA	DKRERPLSFQ	ERNDLAQLV	
	ARPKCGFDML	RKIVFGANKE	AYRFTIESER	RKELKGCDA	AKLAKVNALG	
	TRWQALSLDE	QDRLVCLLLD	GENDAVLADA	LREHYGLTDA	QIDTLLGLSF	
	EDGHMRLGRS	ALLRVLDAL	SGRDEQGLPL	SYDKAVVAAG	YPAHTADLEN	
	GERDALPYYG	ELLWRYTQDA	PTAKNDAERK	FGKIANPTVH	IGLNQLRKL	

Table 4- Cas9 orthologs						
	NALIQRYGKP DAGVPDNREN HILPFSKTL RAAVLPRNKR CSKDDVYVSP DAVVIGATDR ARCVVSHKPD LSNVRCDAPL LPIRPRGEDG NQAAVRRFRN MSENKITLAE IGRVLDPGIK	AQIVVELARN RLRMRLFEEL DSFANKVLAQ NRFAETAMQD GRLTAMLRK AMLQQVATLA HGPEGGLHND QAELEPIFEQ KTLPLDSAPYK DPAFRFRYTA HFEGGTLKQR GD	LKAGLEEKER GQGNGLGTPC HDANRYKGNR WLHNETDFLA WGLNRVLDGV ARAREQDAER TAYGIVAGPF DDARAREVAL AYKGDSDNYCY GGRPLLMLRC DADKDDPFKY	IKKQQTANLE IYSGRQISLQ GPFEAFGANR RQLTDTAYLS MEEQGRPAVK LIGDMPTPWP EDGRYVRVHR TALAERYRQR ELFINERGRW INDYIAVGTA LTKSPGALRD	RNERIRQKLQ RLFSNDVQVD DGYAWDDIRA RVARQYLTAI NRDDHRHHAI NFLEDVRAAV VSLFDLKPGR KVVLEELMSV DGELISTFRA AERTIFRVVK LGARRIFVDL	
Catenibacteriu m mitsukai DSM 15897 WP_00650669 6.1	IVDYCIGLDL ASRSIRRRYN KYLGTDYKDN YLALHHIVKY EKKNLEILEI NVTKMILCEP VYSWVELQTI YFDMFRNDSE NDIELENFLL LSILTFRIPI EGFIKRMRSY KNDIYNELFM PWIDFTNIFG VKQILKLKYK NDKDLGYAQM ITKVMKCRPK VLEELKGFDN PQSLVKDDSF ISPCKFYSLI AAIRANLSHE DSKAVYSEYM EIKKCFYYKD RSDVHKYGGF KINYIEEKEG NEKQCALIAD EKFESMNENY RLKTKNHNLN	GTGSVGWAVV KRRERIRLLR YNLFIDEDFN RGNFLYEGQK LKKPLSKKAK IKQGDSEIKL MGATHTDNAS KSKGYNYIN KQNSRTNGSV YFGPLNETSE CTYFPDEEVL KNKTVTTEKKL KIDQSNFDLI DWSRLSKKLL IEEATSCPED YIYIEFERSE TKKISSDSLF DNRVLVVPSE KTEYTERDEE FRVKNHIYKN KMFRKNKNDQ CYCTTKLDQK SGLQYTIIVAI LSDVRIIKDN IYNAIYQDY VVISKEEKAN NIVFISQSPT	DMNHRLMKRN AILQDMVLEK DYTTYHXYPT FNMDASNIED VDEVMTLIAP KFSDSDNYDDQ ISEAMVSRYN RPSKAPVDEF PYQMQLDEMI HAWIKRLEGK PKNSLIVSKY KNWLNNQCC ENIYDLTVF DGIVADNRF GKFTYEEVER EAKERTESKI LYFTQLGKCM NQRKLDLVLV RFINRQLVET RDINDYHHAH KRWDGDFVIN SGQLFNLTVL EGQKKKGKKT IPVNQMIEMD DNLDDILMIQ I IKQMLIVMH GIYTKKYKL	GKHLWGSRLF DPTFFIRLEH IYHLRKALCE KLSDI FTQFT EKDYKSAFKE FSEVEKDLGE KHHDDLKLLK YKYVKKCIEK KIIDNQAEEY ENQRILPWN EVYNELNKIR SKDAEIKGFQ EDKKIMKRRL SSVTVLDVLE LAGSPALKRG KKLENVYKDL YSGKKLDIDS PFDIRDKMYR RQITKNVTQI DAYIVALIGG SMNYPYEV SNDADADKGV ELVKKISGVP GGEYLLTSPT LYIELTNKMK RGPQNGNIVY	SNAETAANRR TSFLDEEDKA STEKADPRLI SFNNIPYEDD LVTGIAGNKM YVEFVDALHN DCIKNNVPNK VDTPEAKQIL PILKEKREQL QDIVDVEDATA VDDKLEVDV KENQFSTSLT KKKYALPDDK MSRLNLMETI IWQSLQIVEE DEQTKKEYKS LDKYQIDHIV FWKLLFDHEL IEDHYSTTKV FMRDRYPNMH KLIWNPDLIN TKAVVPVNKN HHLKAASINE EYVNARQLVL VLYPAYRGIA DDFKISDRIG	(SEQ ID NO: 68)
Mycoplasma synoviae 53 AOL40776.1	MLRLYCANNL KEKIVIGFDL RRLLRKKFK KINALKEEIK HYEKYGFFRG INEIVKVFEN DIDPETNKF IKNEFTTEFK KKYLLNFALE FSSLKPLYVF ERNEWFKKLL REAFNHIPL NFLKEALVPL NLEKLLNAT LWFEQDRKDP TNQLKKNKTV NSIFKNKKPN	VLNNVQNLWK GVASVGWSIV REKFHKLILK PSELVWILRD SVKLDNKL QSYLTESFKE DYSNIWDKTI WWLNAEQKSE NFNLNEEILK IKFLKEKKLD KELYPKIKNN LLSNNEGKNY SVKTSVLQAI NSNIKILKEK YTQLDIKINE WEYYQNESDP LKFKNITKKL	YLLLLIFDKK NAETKEVIDL NAEIFGLQSR YLQNRGYFYK KKDKAKEKDE EYLKLFNYVR GKCSLFEET ILKFVFTTELF NRDLENDTVL LNYLLGLENT NLEIENVED ESLKHSNEEL KIFNQIIKNF LDQTEKFDDF IEDETEIDHV EAKWNKFVAW FDPYKDLGFL	IIFLFIKVI GVRLFSEPEK NEILNVYKDQ NEKLTDEFVS EEESDAKKES PFNGPGSKN RAPKNLPSAL NWKDKKYS GLKGVKYYEK EILYFLDSIY IFEITDQEK KKRTEKAELK GKKYEISQV TKKKFIDKIE IPYKSADDS AKRIYLVQKS	LIRRYMENNN ADYRRAKRTT SSKYRNILKL NSFPSKKLHE EELIFSINKQW SRTAYGVFST IFNLQNEICT KFKNLQDKI SNATADAAL LAISYSSDLK ESFSKTHSL AQONQKYLK IEMARELTKP NSVVFNRKLF WFNKLVLVKS DKESKDNSEK TKVFRDQLNN	(SEQ ID NO: 69)

Table 4- Cas9 orthologs						
	YSKHHSKDDE QFFHHAVDAS VKKADKDYWK QLFNETLYGV NGSRKRILTE FSQDFINKCK VHKDSKNQIK DIFDFDSYNK QQIYYISTLE GINLNWTKKL	NKLFKVVCMN IIAIFSLTLK DQHNFLKIRE KEFENNFKYL KYIVDEILQI KENSILITFD SFYESYKCVG EKLLNEINLK SSSLRAKIKLLNKMDKGKAVS ENNN	GSITSFLRKS TLYNKLRVYE NAIEIKNVLN EKNLFSRKD LENEEFKDSK AIKHNDPKKV FIWLKNKNDL RPNKKFNSI MKKITNPDEY	MWRKNEEQVY SYDVQRREDG NVDFQONQVRY LRKFILEDLN SDINALNKYM IKIKNLKFFR EESIFVPINS NEIEFVKFVK MKKITNPDEY	RFNFWKKDRD VYLINKETGE SRKANTKLNT EESEKNKKNE DSLPSKFSEF EDATLKNKQA RVIHFQDKDK PGALLNFEN KIIEHVNPL	
Flavobacterium branchiophilum FL-15 WP_01408415 1.1	MAKILGLDLG AAERTGYRSA DYPLNPEFLK GFLSNRLDQS SNEKNGYAKD DVLGKVKGKI SEFITICKVQ GKCSFEKSKS KLVPKFYRKN VAACQVAASL ATSDVYLYEF EGLLYSHAVF EIIINGLLKEY EQQETIFNEL EKGTAEEKEK SIKNPMAMRA QDYQNDNKKF IYEEGKNISI KKQSMPIELN KKIRRRHYLT LKSIFYKKVES IDAITIACMT KIEEILVSH SGKTIYKLDG RKDLESIKGS MNEEKRIAIN AMAIYELDGK KRNRKRDVVK RPSGKVDEYG NQFTA FVEGI	TNSIGWAVVE RKIKYRRKLR WLSTDEESNV AEGILEEHCP LDEGDKKLV KDISQAMLDG GIDQINEEEK RCAISHPDFF DFNFDVLAK KNAIGEDWKT AIDKLGLDEK VANIEINIVDE KSENEDEGKRV LPIFIQQLKD KIKNRLKKLY LHQLRKVLNA REDAIKEIKK CDIIGSNPAY NHLEILPRIA LKR DYLGQKY VKGGMVAEFR KEKYDVLHA YTPDNVKKQA EGKKLPRLQQ DVESIVDEVV KVRIYANSVK RDFELINIFN RGQQVVFYDK VIMLRYFKEA DFKVLPSGKF	RENIDFSLID KYETLKVLSL NPYFFRDRAS KIEAIVEDLI LYKSLLAILK NYKTLGQYFY INEKKFDGLA EYRMWTYLNT LIEKGSSFGF KSFKYQTINS NAKAFSKTKL NIWKDEKQRD YYSKEAEQSF YEFIKIQRLD HPSDIEKFCK LILEGQIDEK LYFEDCKKEV DIEHTIPRSR HWKEEADNLT DRFIWEEPVK KIWGIQESFI WTLEDQQNKK KKIVRVRGKK GDTIRGSLHQ KEKIKEAIA NPLHIKEHSL LAKLIKQGGQ EVENPKDISE RKADDIKQDN EKI	KGVRIFSEGV NRMCPLSIEE KHKVSLFELG SIDEISTNIT SLYSKEKIRN KDLYKAIFFO IKIGTQSDKK YKSSKKNDDF NKEQVSRITVD KKDFASLSLS YIKTQISEII ENDLKKKLVL QKVLIFLKGK KIIKDEFGNE TIIHIEMARE EPTEDDILRY SQDNSQMNKT REIEIISRSI GFKNSQIPDT DENGMKHYKV EARSIEASK QFVAEVERDV DSIYGAIKNP KVL LSSNAQ LSKSKHVHKQ FYPLHKKKEI IVDFKGRIYI FKPDGVFKLG	KSEKGISSR VEEWKKSFGK RAFYHIAQRR SEIERLNNK QYTSREEHYL RPLKSQKGLI LRFLTQDEKL YWFNYKPTDT YKDLWHLLTV AINKILPYLK ENYTTLEKSRF FYKSNEIENK NETGQIFCTE KIVLGSPLTP LNDANKRKG QLWMEQNRSE LCSQRFNREV KAAATKEIKD GIITKYAQAY KDRSKHTHT PWKTFKEDLL NGKAVPKKAA LNTDEIKYVI KRNKLVGTVW KVYQONDENY KGKIVFVPIE IEGLSIQRI ENKPTRKMNH	(SEQ ID NO: 70)
Eubacterium yurii subsp. margaretiae ATCC 43715 EFM38267.1	MENKQYYIGL RRTKRTRSRR SENNRQRYTL ILLNLFSHRGH NFEKILSQKG YNIEDIQDEN GLLSNIIGNS TDKNYSAYVG IEILEKIKLG GEKNLTVSEM WNFEDKVDVK NIKIDGEEKIS GTDTVCNAYL KEEIVEKYGD IIQSLWETNF SSPVKRMIWQ	DVG TNSVGWA SEREKARKAM FNDATFTDKD FLNASLKG DG KSRTKILEEL KKIKIGFRES KYLCEARVEA SVNSNIAKER EFLKKQLTAS IIQLFEFQIP GSRKEFIEKM VEAKQKIYND SSIGKFTGVF EIDKDKIKRI NLMELLSSRF SMKIVDEIQT	VTDTSYNLLR LKELFADEIN YYEKYKTIFH DIQGMDFYN SEELSISKKD DYEESLKVK YENHHKDLLK RSVDKRKIED NGVIPNQLQS YYVGPLDKNP VRKCTYISDE LFVKGKKVSQ KEEINKQSIV LGFKFSNWGN TYMDELEKRV VIGYAPKRIF	AKGKDMWGAR RVDPSFFIRL LRSALINSDE DLVESCEYFE KSKYNLIKLI EIIGDEYFDL IKELLKKYDK LYKYIEDTAL RELRAILKKA KKDNKANSWA HTLPKQSLLY KDIKKELISL DMIEDIIFLK LSKSFLELEG KKLEKPLSEW VEMTRSEGEK	LFEKANTAAE EESKFFLDDR KFDVRLVFLA IELPRITNID SGLEASVVEL VERAKSVHDM KAYNDMFRKM KNIPDDNKDK ENYLPFLKEK KIKQGGRIPL EKFVMVNEIN NIMDKDSVLS TVYGDEKRFV ADVGTGEVRS TIEDLDDMYL VRTKSRKDRL	(SEQ ID NO: 71)

Table 4- Cas9 orthologs						
	KELYNIGIKED DKLMDNLYD SCQGFWKILH KCMAQILQKS LNIVVGN SYF VAWIGQSEGN KFGRNKVEKA LVEYVKRKT IRLCLPFIST KIEKAVSMSK KYNKKDLSFG GGSKSTGKCR	SKQWVKELDS IDHIYPRSFV DQGFMSNEKY MGEDVDVVS VKFTRNPANF SGTIVIVKKT DKTPKKPNLQ IRSLEAIPVY NSLVKIDGYL FDEIDREKNP DFLKNKKSFK CKKNITNYKE	KDESYFRSKK KDDSLDNLVL SRLTRKTQEF KARLVSEFRH IKDARKNPDN MAKNSPLITK AYRPIKTSDE LGRKDSLSEE YYLGGKNDDR VLTEEKNIEL BEIDLEKQCK FKLIQQSITG	MYLYYLQKGR VKKEINNRRQ SDEBKLSFIN KFELFKSRLI PVYKYHMDRF KVEEGHGSIT RLCNILRYGG KLLNYFRYNL IQLYNAYQLK YNKIQDKFEN VLYNIIFNLS LYSCEKDLMT	CMYSGEVIEL NDPITPQIQ RQIVETGQAT NDFHHANDAY FERDVKSKSE KETIVGVKEI RTSISISGYC NDGGKDSVSD MKKEEVEYIR TVFSKRMSLV NLKEVDLSDI I	
Acidovorax ebreus WP_01265517 6.1	MAQHVFGLDI QARLLRRRLY QGLDRLLTPL HTKALMQAKN ATQRRRLGNPH FEKGEYRAPK YQTETSKYKT LVKLPAWHEL GAEVVQQLRQ RYDEAVAQIP FRDDADI PRN ERNKVKRAQE YSQQPLDIQR RTAFEYLTSE KGFIDRNLND LRARWGLTKV GFPDPETGEI QRLGSYTTED EKILLTSLKL KAFGPDNLFH VDIFTKAGKF VYPNDYVKVT GVKTALSVEK	GIASVGWAIL RRARWLTQLS EWARVIYHQC YRSAAEMVLA ASDFFEKLIL ASFSVERHVW LKNAFIKAGL RKAFKAAGHE LALPEPAASI EYGHHSQRIE PVVLRALNQA EFRDRNDRAR VLDDHNYAQV PDGEDGERWR TRYICKFFKN RGDSDRHHAL LNPAAFDRAR LGRRLTLFVS QDFDKLLNPE KPDKNNQPTG HLVPVYVHHR LKKEQQSGYY FNVDVLGRIY	GEQRIIDLGV RLLKRKGLIA KHRGFHWTSK EFPDAQNRKR GDGDRKSGLF LTRLNNLRIV WGDGVRFGGL ALWQOISTPA AVLEKISFDK PGAACHLYLP RKVVNALIRE SEFERDFGYK DHALPYSRSY TFVAWVQGNK YVEEHLQLAA DAAVVAACHT QHFPEPWTHF RAPQRRSGGA SNDHFVEPHR PVVRSIKLVR VTGLPNRAIV SGADRSTGAM LAPPETRSGL	RCFDKAETAK DAKLFAKAPS ABEAKADSDA GQYDKALSRV WQQKPALSGA VDGRSRPLNE AYPSQAQIDA LDGDP TLLDQ FSSLSLKALR PFYEAQRKYA YGSPIAVNIE PKAAAFEKWM DDSKNNKVLV RADGDTARRC GMVKALADYS AHELKARLFT VHKETIYAQP NERLYAAIRQ GKQGTGPIRG AFKDEDEWTL NLWAHDRAAS A	EGDPLNLTRR YGDSAWELRR EGGRVKQGLA LLGEELALLF DLLKMLGKCT AERQAALLLP EKT KD PEDQF IATVLSVYKD RIVPLMQSGL GKGDHIGSMQ MARDLSRPLD LYREQLGQCA LTHENQNKGN LRKNYGVDES VVVNGQLTAF RRKEISFLQE DDLAALREDM ESLKQQGGVI RLEQFGGRAD GLAKNDSMLR IDESFAFLFS VGKDGLIRGI	(SEQ ID NO: 72)
Porphyromona s sp. oral taxon 279 str. F0450 WP_00943351 8.1	MLMSKHVLGL EAFNAGAAFT IQLP LLELWE VGDEGEKKKD GGISYRIKDQ MQRPLKSCKH KSSPLFQLCC AALKKLLKEK ETRMNTVQLT RRALITQLGM GLGYSEACAA NQMINLVNAL VAAKIRECGL EHIIPKSVLY NDLLKEKKIS IRRVSASEGG ELHITNWSKR QEAQEQQATE TRGRNIYRKK RYPLHDLKGE SVRCYAKTLS	DLGVGSIGWC ASQERTARRT LRERAATAGR SNSAYLAGIR IFSRQCYIDE LVSLCEFEKQ IYEAVNNIRL ALIADQLTSK DEETGEVTER KEEDLDGGLL VGYRHSNSPT KAEYGIDEVR YPTKPRIQKY DDSYGNKTCA YSKHQRLRWL VTARLRSLWG MDHRHHAIDA TGRLSNLERW MADGREVSCV VVDPHLRELI LDKAIPMCFD	LIALDAQGDP MRRGFARYQL RLTLP ELGRV ANDEKLQAEH YDQIMAVQRV ERV MRVQDD TRPNGSPCDI SGLKGNSTRV EVAVVTDSYV DQLYRLDFVK SEEITERTLL VELARELKMS MLWKEAGRQC CRRCNKEKGN KEDIPSDFLE YGKILHTLNL LVVACTRQSY LTQRPHFSVR QRGVLVPRGE TTYNQELKSR EKGEPTAFVK	AEILGMGSRV RRYRLRRELE LCHINQKRGY KTVGQYFAEQ HYPDILTDEF GKGGWQLVER TPEERAKIVA ALASALQPYP RKPLYRLWHI PGYGNSAKF EKIPLLRNE REERERMARN LYCGRSIEEE RTALEYIRAK RQLRLTQYIS DRYDSMGETE IQRLNRLSSE TVSDKVAEIL LMEASFYGKI EKGAPIPPLC SASNHHALY	VPLNNATKAI KVGMLPDAAL RHVKS DAAAI LRQNQSESPT IRMLRDEVIF RVKFGPKVAP HLQSSASLSF QYHHLLDMEL LYSIEEREAM ICKLLPQLQQ LRQPLVEKIL NKDREERNKG QCLREGGMEV GREAEYMKRI RQAMAILQQG RVSREGEATE FGREDKKKED ISYRPGQRVV LSQGRVRIVK LDKDKKQEV RTPKGLLVES	(SEQ ID NO: 73)

Table 4- Cas9 orthologs						
	IVTFWDAVDR VDSLQQDEM HLETSVADDDK	ARYGIPLVIT VIGLSDEELQ NTSGRIPKFEH	HPREVMEQVL RALEAQNYRK RVQSLKAYEE	QRGDIPEQVL ISEHLYRVQK RNIRKVRVDL	SLLPPSDWVF MSSSYVFRY LGRISLL	
Mycoplasma ovipneumoniae SC01 WP_01032092 2.1	MHNKKNITIG RSTRRNIRRK ISFFTEIYKK DLEENNVADK WVNEIKLFE DEKGKVFCKY TDLKTTNKKI DFGFEKSDID INNILEFLPY DFVKSIFSNT RKWEEQKSKL KYSKENIIDA NKGKYLSDLE IEHIIPYSMS EKAKELFINK NLNDRYSTK FRAPVQKNN IANLLTLADN KIIQEKYEEA IKKKLVTSKN EKYNGFSNKT QFTYHDNINN LYSLVYKVYE SDFAKPIPVN SKRFNRDTAY IKEHTDDEKL	FDLGIASIGW AYRNQRFINL CAAKHSNILE YEGIEHPSIL VQEINPEFSE DNIWDKTIGK WQLSSNDRNE DQDTIEGRKI LDAICIIILDR KFNFKKIGNF GKTDKTKYL IIIESPREKN TKPAKLLDRL YDNSQANKIL YKKNKKLDSY LFYHALVEHF GPNENLNNNK KTDKKFLLHD KKHTAIKFSR EELKKYFENP GNAFVEYMND KSFKRFYKNI NGKESYKSIP CKPVFVLKKG GLKPLKLSVV MCTIK	AIIDSTTSKI ILKYKDLFEL VKVKALDSKI LYDFFKKNGF KFLNLFTSVR CSFFVEENRS LDELLEKVE IKEEPTTKLE EKSRGQDEVL SLKAIREFLP NPRIFQDEII DKKTIEEIKK RFYHQDGDID TEKAENLKKG VDLDEDSAKN ENNEFFTYID PEKIEKNREN ENYKENIETG KTRTILNGGL FGKKADGKSE LALKEPTLKA RIIEYKSIPI VTSQMRNFGI SILKKKSLDI KPVAEPSTNP	LDWGTTRTFEE KNISDIQRAN EKLDLIWILH FKSNSSIPKD DYAKGPGSEH PVNYPSEYIF KAKIISISLK VTKHLLATY KKLTEKNIFE KMFEQNKNSE SPGKNTFEQ RNKKGKGT LYTLDKINID KLIASEYIKR RFRFLTQDY ENSSKHVKI NEHHAVDAAI ELVKIPKFEV SDETLYGFKY YTVLMAQSHL EIESAKSVEK KFKILSKHDG DEFDFLDENL DDFKETKETE TFKEYIPIHL	RKTANERRAF KKDTENYEKI DYLENRGFFY LGGYSFSNLQ SASEYGIQK NLLNQLINLS KNEIKKIILK SHSSDSNWIN VLKIDREKQL YLKWKDEEIR AVLVNLQIIK EKLFIQILNLE QLINGSQKYE NGDEFYNKY DEFQVEFLAR STIKGHVTKY VAIIGNKNPQ DKLAKVEDLK DEKEDKYFKI SEFNKLKEIF LLYYNFKPSD GKSFKDTLFS YNKEKLDIYK EGNYFIST DELGNEYPVK	(SEQ ID NO: 74)
Wolinella succinogenes WP_01113943 1.1	MLVSPISVDL SKRNNLRNKL DEKKKDALES FEAVFASATH ENQGNLPRAK LKLRRYFND KDLKSKEIME NWRDIAKRLN KKPNELLVKT LEDNSNMLKR VGNKKLSAYC DDVILNEWSQ KRCTAENRFR DKLGYELAKI VINSKKDRDG EIDHILPRSH SAQWIEEQVA TDQSARVNGT ARQNPLLAKE LDSWNPASNE GERYKPIVVQ DLISLKENQY VENCRIYTKK KDRFVIDKSS KSIRIKARKT ANILLEANSK ADFDTHGIRL	GGKNTGFFSF VKRLFLLILQ DTLKEFLSEK SPNKKLELKD YFEELGEYIA KEMEKGDIWI LLTTTEPVMT HGKFNDLAD LGQSDANRLY CNHNPPHKKK KNIEELRKTH KIANFFDIDD SETAFYNDET KAKELEGMEA KKLAKAKENA TLKIYGTVMF NIKGYKTFSV QKYLAKKIQE EKQAPSSHAI PLTKGLSTNQ EGGYYIGYPA IKIFSINKQT VDVKFAPKYI LNEYYPQDKN FSLLAESGVQ TSLSLPKSGI IVENTSVKVG	TDSLDSNSQSG EHHGLSIDVL LQSIDRSDSV ELKSEYGENA TNEKVSFFD PNKLHKITER IPPYDDMNNR STVKGYSSEDS GFAQNYIELI QIHNLVAGIL GNTFKIDIEE KHRQRFNNLF GEFHKKATAT KEIKVPIILE EDRLKDKDKR PEGNLIYVHQ LSAEQQKAFR KLTKMLPNKH DAVMAFVARY KIEKMIKSGD TVKKGYELKN ISELSNRYFN HETKYPFYDD EYKLDVDTQP GKVFRKRKI SIEKQLDKKL NFPISAIKDS	TVIYDESFVL PDEIRGLFNK EDFLNQIASN KELLAGLRVT SNSLKLDMT FVRSWHPKND GAVKCQTLRL TLLHRLDTS RQKVRAGIWW GVKLDEAKFA LRRKKDPAELS SMAQLHTVID CQRLPADTQR QNAFEYEESL IKAFSSGICP KCNQAKADSI YALFLQNDNE LSFEFILADA QKVFKDGTTP YGQKNMREVF CKVVTSKNDI MNYKNLVERD WRRFDEAWRY IWDDFCRWYF PTGYAYQALP DVIKKTDVIRG AKRMIFRALF	SQVGRRSKRH RGYTYAGFEL AESFKDYKKG KEILDEFDKQ KLIGNISNYQ ADRQRAELM GEEYLDKHL KEIDIYELRG PVKNKDDSLN EFEKELWSAK KEEKAKLRLT TPRSGFSSTC PFSGKIERYI RKSKTGSNDR YCGDTIGDDG YKLSDIKAGV AYKKVVDWLR TEVSELRRQY NADEVAKLAM GKSIFGENAI AKLEKIKNQ KEIVGLLEFI LQENQNTSS LDRYKTANDK MDNNVIAGDY LAIDNNSFFN EKEKGKRRKK	(SEQ ID NO: 75)

Table 4- Cas9 orthologs						
	TTISFKESGP SEHIQKLLK	VQDYLKVFLK	KIVKIQLRD	GSISNIVVRK	NAADFTLSFR	
Streptococcus mutans UA159 WP_00226354 9.1	MKKPYSIGLD LLFDSGNTAE LEDSEFLVTE LRLVYLALAH LQEQNVQVEE DFKKHFEELE LLSGILTVD FSDVSKDGYA KQRTFDNGSI YVGPLARGKS YLPNQKVLPK VFKVYRKVTK KILDKDFLDN KLERRHYTGW DALSFKEEIA IMGHQPENIV VENSQLOQND SIDNRVLTSS TKAERGGLTD RQVKIVTLKS YPQLEPEFVY IIWKKDEHIS RKTKKFYWD MEKMTFERDP ELQKGNIEIV SNFSKKYTLA FKFFDKNIDR	IGTNSVGWAV DRRLKRTARR KRGERHPIFG IIKFRGHFLI ILTDKISKSA KAPLQFSKDT VGTKAPLSAS GYIDGKTQNE PHQIHLQEMR DFAWLSRKSA HSLLEYKFTV DKLMDFLEKE SKNEKILEDI GRLSAELIHG KAQVIGETDN VEMARENQFT LFLYYLQNGR KENRGKSDDV DDKAGFIKQ NLVSNFRKEF GDYPHFHGHK NIKKVLSYPQ KKYGGFDSPI VAFLEKGYR PNHLGTLTYH EGNLEKIKEL KRYTSTTEIL	VTDDYKVPK RYTRRRNRIL NLEEEVKYHE EGKFDTRNND KKDRVLKLF YEEEELEVLLA MIQRYNEHQ AFYKYLKGLL AIIRRQAEFY DKITPWNFDE YNELTKVKYK FDEFRIVDLT VLTLTFLFEDR IRNKESRKTI LNQVVSDIAG NQGRNSQQR DMYTGEELDI PSKDVRKMK LVETRQITKH ELYKVRIND ENKATAKKFF VNIVKKVEEQ VAYSILVIAD NVQEENIIL AKNIHKVDEP YAQNNGEDLK NATLIHQSI	KMKVLGNTDK YLQEIFSEEM NFPTIYHLRQ VQRLFQEFLLA NEKSNRFAE QIGDNYAELF DLAQLKQFIR NKIEGSGYFL PFLADNQDRI IVDKESSAEA TEQGKTAFED GLDKENKVF EMIRKRLNY LDYLIDDGNS SPAIKKGILQ LKGLTDSIKE DYLQYDIDH SYWSKLLSAK VARILDERFN YHHAHDAYLN YSNIMNFFKK TGGSFESIL IEKGSKSKLK PKYSLFKLEN KHLDYVDKHK ELASSFINLL GLYETRIDL	SHIEKNLLGA GKVDDSFHR YLADNPEKVD VYDNTFENSS FLKLIVGNQA LSAKKLYDSI QKLSDKYNEV DKIEREDFLR EKLLTFRIPY FINRMTNYDL ANMKQEIFDG ASYGTYHDL SDLLTKEQVK NRNFMQLIND SLKIVDELVK FGSQILKEHP IIPQAFIKDN LITQRKLDNL TETDENNKKI AVIGKALLGV DDVRTDKNGE PKGNSDKLIP TVKALVGVTI GRKRLLASAR DEFKELLDV TFTATGAPAT KLGGD	(SEQ ID NO: 76)
Prevotella timonensis CRIS 5C-B1 WP_00812271 8.1	MNKRIILGLD SSKAAERSGY QKQYPKSDEL LYHLTQRRGF YKLYDAQGNK FQLPLKSQRH LELRPLTYEE AYKFNRYMTQ EMVDDVWNV LKAIRKFLPF LVFNYPKHR QRNEFGILQL VEYARELND TDVLKFQLWE QMNLTLCNDR QRTFAGMDKA TGIGLISRYA EKKCRDNHSH KPWATFTEDV GSLHLDITYG IADKNFKQAI YKQQYHVMNE FSSIVPTHST VVGIEKDGR VDGIDFTIDI	GTNSLGWAV KAIRKQYFRR MLWQRTSDEE LSNRLDTSAD VRIRQRYTDR GVGRCTFERG REKIEPLFFR GVPGCPTIAQ YSFSSVEK LRKGMYYTHA EVQGTIEMLI GSPRTNAIRN NKRRAIADRQ EQNHHCCLYTG FNREVKKAKL VKDIRIQKRH GLYLKSLFHQ HCMDAITIAC LNIYKNLLV AIERDGEIRY AEPIYMNEEK NNYLLAIYEG KYGLPLKTKL KFYHQEARK LGKVTLE	DWDEHAQSYE RLRKIQVLKV GKNPYYDRHR NKEDGVVKS NKHYQHEFDA KPRCADSHPD KSKPNFDFED LKSIFGDDWK EFAHKLQLD SFFANIPTIV KDFLANNFEL PMAMRSLHIL KEQDKQHKKY EQIGITDFIG PTELANHEEI KLQMEIDYWR ADSRNKSNNV IGKREYDLMA HDTNNMPKH VVRPLSSFT GILIKKVRCE LVKNKVREF LMGQLVLMFE EGLPIFSTPY	LIKYGDVIFQ LVKYHLCPYL CLHEKLDLTV ISQLSTEMEE ICEKQELSSE YEEFRMLCFV IAXALAGKKN TGIAETTYLI EESAELFAKI GKEIWNKEQN PAGATDKLYH RRVNVQLLKE VVKGVATAEF EYRMEETFK TKKYVQTSIG KPELENIVD AKSVKQPINI EIVSYIEAAK ENPDEIQVDN KNNDYAPIF	EGVKIEKGIE SDDDLRQWHL EADRYTLGRA AGCEYLGDYF LIEDLQRAIF NNIQVKGP YAWIHDKEER QKKNKSLSLQ KLSHSFAALS RKYIMENVGE PSMIETYPNA SIIDENTEVH ETGKDIEPTQ IPQSVGGDST YEQLVKERDK VPEGFSRRQG RKMWGLQSEY QGRGSKPKFS KVLAQGD ETVKRTIKEA RQHRDLKKE YYKRSQDRNI TKDLVKRLYK RQSINNINIL	(SEQ ID NO: 77)

Table 4- Cas9 orthologs						
Clostridium cellulolyticum H10 ACL77411.1	MKYTLGLDVG	IASVGWAVID	KDNNKIIDLG	VRCFDKAEES	KTGESLATAR	(SEQ ID NO: 78)
	RIARGMRRRI	SRRSQRLRLV	KKLFVQYEII	KDSSEFNRI	DTSRDGWKDP	
	WELRYNALS	ILKPYELVQV	LTHITKRRGF	KSNRKEDLST	TKEGVVITSI	
	KNNSEMLRTK	NYRTIGEMIF	METPENSNGR	NKVDEYIHTI	AREDLLNEIK	
	YIFSIQRKLG	SPFVTEKLEH	DFLNIWEFQR	PFASGDSILS	KVGKCTLLKE	
	ELRAPTSCT	SEYFGLLQSI	NNLVLVEDNN	TLTLNNDQRA	KIIEYAHFKN	
	EIKYSEIRKL	LDIEPEILFK	AHNLTHKNPS	GNNESKKFYE	MKSYHKLKST	
	LPRTDIWGLH	SNKESLDNLF	YCLTVYKNDN	EIKDYLDQANN	LDYLIBEYIAK	
	LPTFNKFKHL	SLVAMKRIIP	FMEKGYKYSD	ACNMAELDFT	GSSKLEKCNK	
	LTVEPIIENV	TNPVVIRALT	QARKVINAI	QKYGLPYMVN	IELAREAGMT	
	RQDRDNLKKE	HENNRKAREK	ISDLIRQNGR	VASGLDILKW	RLWEDQGGRC	
	AYSGKPIPVC	DLLNDSLTIQ	DHIYPYSRSM	DDSYMKNVLL	LTDENQNKRS	
	YTPYEVWGST	EKWEDFEARI	YSMHL PQSKE	KRLNLRNFIT	KDLDSFISRN	
	LNDTRYISRF	LKNYIESYLQ	FSNDSPKSCV	VCVNGQCTAQ	LRSRWGLNKN	
	REESDLHHAL	DAAVIACADR	KIIKEITNYY	NERENHNYKV	KYPLPWHFSR	
	QDLMETLAGV	FISRAPRRKI	TGPAHDETIR	SPKHFNKGLT	SVKIPLTTVT	
	LEKLETMVKN	TKGGISDKAV	YNVLKNRLIE	HNNKPLKAFA	EKIYKPLKNG	
	TNGAIIRSIR	VETPSYTGVE	RNEGKGISDN	SLMVRVDVFK	KKDKYYLVPI	
	YVAHMIKKEL	PSKAIIVPLKP	ESQWELIDST	HEFLFSLYQN	DYLVIKTKKG	
	ITEGYRSCH	RGTGSLSLMP	HFANNKNVKI	DIGVRTAIS	EKYNVDILGN	
	KSIVKGEPRR	GMEKYNFSKS	N			
Francisella tularensis subsp. novicida U112 WP_00303894 1.1	MNFKILPIAI	DLGVKNTGVF	SAFYQKGTSL	ERLDNKNKGV	YELSKDSYTL	(SEQ ID NO: 79)
	LMNNRTARRH	QRRGIDRKQL	VKRLFKLIWT	EQLNLEWDKD	TQQAISFLFN	
	RRGFSFITDG	YSPEYLNIVP	EQVKAILMDI	FDDYNGEDDL	DSYLKLATEQ	
	ESKISEIYNK	LMQKILEFKL	MKLCTDIKDD	KVSTKTLEKI	TSYEFELLAD	
	YLANYSESLK	TQKFSYTDKQ	GNLKELSYHH	HDYNIQEFEL	KRHATINDRI	
	LDTLTDDLD	IWNFNFEKFD	FDKNEEKLN	QEDKDHIQAH	LHHFVFAVKN	
	IKSEMASGGR	HRSQYFQET	NVLDENNHQE	GYLKNFCENL	HNKYSNLSV	
	KNLVNLIGNL	SNLELKPLRK	YFNDKIHAKA	DHWDEQKFTE	TYCHWILGEW	
	RVGVKDQDKK	DGAKYSYKDL	CNELKQKVTK	AGLVDFLLEL	DPCRTIPPYL	
	DNNNRKPPKC	QSLILNPKFL	DNQYPNWQQY	LQELKKLQSI	QNYLDSFETD	
	LKVLKSSKDQ	PYFVEYKSSN	QQIASGQRDY	KDLDAKILQF	IFDRVKASDE	
	LLLNEIYFQA	KKLKQKASSE	LEKLESSKKL	DEVIANSQLS	QILKSQHTNG	
	IFEQGTFLHL	VCKYYKQRQR	ARDSRLYIMP	EYRYDKKLHK	YNNTRGRFDD	
	NQLTTCYCNH	PRQKRYQLLN	DLAGVLQVSP	NFLKDKIGSD	DDLFSKWL	
	EHIRGFKKAC	EDSLKIQKDN	RGLLNHKINI	ARNTKGKCEK	EIFNLICKIE	
	GSEDKKGNK	HGLAYELGVL	LFGEPPNEASK	PEFDRKIKKF	NSIYSAFIQ	
	QIAFAERKGN	ANTCAVCSAD	NAHRMQQIKI	TEPVEDNKKD	IILSAKAQRL	
	PAIPTRIVDG	AVKKMATILA	KNIVDDNWQN	IKQVLSAKHQ	LHIPIITESN	
	AFEFEPAALAD	VKGKSLKDRR	KKALERISPE	NIFKDKNNRI	KEFAKGISAY	
	SGANLTDGDF	DGAKEELDHI	IPRSHKKYGT	LNDEANLICV	TRGDNKNKGN	
	RIFCLRDLD	NYKLKQFETT	DDLEIEKKIA	DTIWDANKKD	FKFGNYSFI	
	NLTPQEQAFA	RHALFLADEN	PIKQAVIRAI	NNRNRTFVNG	TQRYFAEVLA	
	NNIYLRAKKE	NLNTDKISFD	YFGIPTIGNG	RGIAEIRQLY	EKVDSIDIQAY	
	AKGDKPQASY	SHLIDAMLAF	CIAADEHRND	GSIGLEIDKN	YSLYPLDKNT	
	GEVFTKDIFS	QIKITDNEFS	DKKLVRKKAI	EGFNTHRQMT	RDGIYAENYL	
	PILIHKELNE	VRKGYTWKNS	EEIKIFKGGK	YDIQQLNNLV	YCLKFVDKPI	
	SIDIQISTLE	ELRNILT'TNN	IAATAEYIII	NLKTQKLHEY	YIENYNTALG	
	YKYSKEMEF	LRSLAYRSE	VKIKSIDDVK	QVLDKDSNFI	IGKITLPFKK	
	EWQRLYREWQ	NTTIKDDYEF	LKSFFNVKSI	TKLHKKVRKD	FSLPISTNEG	
	KFLVKRKTWD	NNFIYQILND	SDSRADGTPK	FIPAFDISKN	EIVEAIDISF	
	TSKNIFWLPK	NIELQKVDNK	NIFAIDTSKW	FEVETPSDLR	DIGIATIQYK	
	IDNNSRPKVR	VKLDYVIDDD	SKINYFMNHS	LLKSRYPDKV	LEILKQSTII	
	EFESSGFNKT	IKEMLGMKLA	GIYNETSNN			

Table 4- Cas9 orthologs						
Azospirillum sp. B510 AOL40891.1	MARPAFRAPR	REHVNGWTPD	PHRISKPFPI	LVSWHLLSRV	VIDSSSGCFP	(SEQ ID NO: 80)
	GTSRDHTDKF	AEWECVQPY	RLSFDLGTNS	IGWGLLNLDR	QGKPREIRAL	
	GSRI FSDGRD	PQDKASLAVA	RRLARQMRRR	RDRYLTRRTR	LMGALVRFGF	
	MPADPAARKR	LEVAVDPYLA	RERATRERLE	PFEIGRALFH	LNQRRGYKPV	
	RTATKPDEEA	GKVKEAVERL	EAAIAAAGAP	TLGAWFAWRK	TRGETLRARL	
	AGKGKEAAYP	FYPARRMLEA	EFDTLWAEQA	RHHPDLLTAE	AREILRHRIF	
	HQRPLKPPPV	GRCTLYPDDG	RAPRALPSAQ	RLRLFQELAS	LRVIHLDLSE	
	RPLTPAERDR	I VAFVQGRPP	KAGRKPGKVQ	KSVPF EKLRG	LLELPPGTGF	
	SLES DKRPEL	LGDETGARIA	PAFGPGWTAL	PLEEQDALVE	LLLTEAEPER	
	AIAALTARWA	LDEATAAKLA	GATLPDFHGR	YGRRAVAELL	PVLERETR GD	
	PDGRVRPIRL	DEAVKLLRGG	KDHSDFSREG	ALLDALPYYG	AVLERHVAFG	
	TGNPADPEEK	RVGRVANPTV	HIALNQLRHL	VNAILARHGR	PEEIVIELAR	
	DLKRS AEDRR	REDKRQADNQ	KRNEERKRLI	LSLGERPTPR	NLLKLRLWEE	
	QGPVENRRCP	YSGETISMRM	LLSEQVDIDH	ILPFSVSLDD	SAANKVVCLR	
	EANRIKRNRS	PWEAFGHDSE	RWAGILARAE	ALPKNKRWRF	APDALEKLEG	
	EGGLRARHLN	DTRHLSRLAV	EYLRVCVCPKV	RVSPGRLTAL	LRRRWGIDAI	
	LAEADGPPPE	VPAETLDPSP	AEKNRADHRH	HALDAVVIGC	IDRSMVQRVQ	
	LAAASAEREA	AAREDNIRRV	LEGFKEEPWD	GFRAELERRA	RTIVVSHRPE	
	HGIGGALHKE	TAYGPVDPPE	EGFNLVVRKP	IDGLSKDEIN	SVRDPRLRRA	
	LIDRLAIRRR	DANDPATALA	KAAEDLAAQP	ASRGIRRVRV	LKKE SNPIRV	
	EHGGNPSGPR	SGGPFHKLLL	AGEVHHVDVA	LRADGRRWVG	HWVTLFEAHG	
	GRGADGAAAP	PRLGDGERFL	MRLHKGDCCLK	LEHKGRVRVM	QVVKLEPSSN	
	SVVVVEPHQV	KTDRSKHVKI	SCDQLRARGA	RRVTVDPLGR	VRVHAPGARV	
	GIGGDAGRTA	MEPAEDIS				
Peptoniphilus duerdenii ATCC BAA-1640 WP_008901059. 1	MKNLKEYYIG	LDIGTASVGW	AVTDESYNIP	KFNGKKMWGV	RLFDDAKTAE	(SEQ ID NO: 81)
	ERRTQRGSRR	RLNRRKERIN	LLQDLFATEI	SKVDPNFFLR	LDNSDLYRED	
	KDEKLKSKYT	LFNDKDFKDR	DYHKKYPTIH	HLIMDLIEDE	GKKDIRLLYL	
	ACHYLLKNRG	HFI FEGQKFD	TKNSFDKSIN	DLKIHLRDEY	NIDLEFNNED	
	LIEIITDTTL	NKTNKKKELK	NIVGDTKFLK	AISAIMIGSS	QKLVDLFEDG	
	EFEETTVKSV	DFSTTAFDDK	YSEYEEALGD	TISLLNILKS	IYDSSILENL	
	LKDADSKDG	NKYISKAFVK	KFNKHGKDLK	TLKRI IKKYL	PSEYANIFRN	
	KSINDNYVAY	TKSNITSNKR	TKASKFTKQE	DFYKF IKKHL	DTIKETKLNS	
	SENE DLKLID	EMLT DIEFKT	FIPKLKSSDN	GVIPIQLKLM	ELKKILDNQS	
	KYYDFL NESD	EYGT VKDKVE	SIMEFRIPYY	VGPLNPDSKY	AWIKRENTKI	
	TPWNFKDIVD	LDSSREEFID	RLIGRCTYLK	EEKVL PKASL	IYNDFMVLNE	
	LNNLKLNEFL	ITEEMKKAIF	EELFKTKKKV	TLKAVSNLLK	KEFNLTGDIL	
	LSGTDGDFKQ	GLNSYIDFKN	IIGDKVDRDD	YRIKIEEIIK	LIVLYEDDKT	
	YLKKKIKSAY	KNDFTDDEIK	KIAALNYKDW	GRLSKRFLTG	IEGVDKTTGE	
	KGSIIYFMRE	YNLNLME LMS	GHYTFTEEVE	KLNPVENREL	CYEMVDELYL	
	SPSVKRMLWQ	SLRVVDEIKR	IIGKDPKKIF	IEMARAKEAK	NSRKESRKNK	
	LLEFYKF GKK	AFINEIGEER	YNYLLNEINS	EEESKFRWDN	LYLYYTQLGR	
	CMYSLEPIDL	ADLKSNNIYD	QDHIYPKSKI	YDDSL ENRVL	VKKNLNHEKG	
	NQYPIPEKVL	NKNAYGFWKI	LFDKGLIGQK	KYTRLTRRTP	FEERELAEFI	
	ERQIVETRQA	TKETANLLKN	ICQDSEIVYS	KAENASRFRQ	EFDI IKCRTV	
	NDLHHMHDAY	LNIVVGNVYN	TKFTKNPLNF	IKDKDNVRSY	NLENMFKYDV	
	VRGSYTAWIA	DDSEGNV KAA	TIKKVKRELE	GKNYRFRTRMS	YIGTGGLYDQ	
	NLMRKGKGQI	PQKENTNKS N	IEKYGGYNKA	SSAYFALIES	DGKAGRERTL	
	ETIPI MVYNQ	EKYGNTEAVD	KYLKDNLELQ	DPKILKDKIK	INSLIKLDGF	
	LYNIKGKTGD	SLSIAGSVQL	IVNKEEQKLI	KKMDKFLVKK	KDNKDIKVTS	
	FDNIKEEELI	KLYKTLSDKL	NNGIYSNKR N	NQAKNISEAL	DKFKEISIEE	
	KIDVLNQIIL	LFQSYNNGCN	LKSIGLSAKT	GVVFIPKKLN	YKECKLINQS	
	ITGLFENEVD	LLNL				
Lactobacillus coryniformis	MGYRIGLDVG	ITSTGYAVLK	TDKNGLPYKI	LTLD SVIYPR	AENPQTGASL	(SEQ ID NO: 82)
	AEPRRIKRGL	RRRTTRTKFR	KQRTQQLFIH	SGLLSKPEIE	QILATPQAKY	
	SVYELRVAGL	DRRLTNSELF	RVLYFFIGHR	GFKSNRKAEL	NPENEADKKQ	
	MGQLLNSIEE	IRKAI AEKGY	RTVGELYLKD	PKYNDHKR NK	GYIDGYLSTP	

Table 4- Cas9 orthologs						
subsp. torquens KCTC 3535 WP_01001440 6.1	NRQMLVDEIK EGPGAGPYAG QNTTGDWHT DARFNLVNYG LDHIGTALT MQNIIPYLEM TIKIVEQIIR AAELTELGIP IIPYSISWDD NIIRNSRKRO NPELEKKQRV KFIQQVTKYS FPLPWPEFLD HKISGPAHLD ENSNNGDKVV AKKVSPLPVL KAIARGKSLW VPIKNFYGYF GRYHKVHEKK	QILDKQREL DQIKKMVGKD LNGLDRQAI SLQTQKEIET YSSDKRRRRY GQVYSEATTN RYGKPDGINI VNGQNIIRYK SYTNKVLTS KLLKQKLSDE LPLNGEVT QHGEVKNNQA ELLARISDNP TIRSAKLFD YSAIRQALEA KNKAAADNGS YQITESDQFC DGADIATASI RQLFVKRDE	NEKLTDIFYA IFEPTEDRAA DAVFAKAEKP VEKKTRFVDF FAEELNLP TGYDFRKKQI ELARELGRNF LHKEQNGVDP KCNREKGNRI ELKDWKQONI IRSRWGFLKV LWHD AEIKDA VEMMKSRSWN KGIVLSRVSI HNGSGELAFP MVRIDVFNTG FSLYPGDMVH LVRAHDSSYT RQFVKRDE	TYLLGDENRA KATYTFQYFN TKTYKPTDFG KAYHDLVKVL LIEKLLPLNF SKDTIREBIT KERGDIQKRQ YTGDQIPFER PMVYLANNEQ NDTRFITRVL REDGDLHHAI EYAAEAQRM TYTPIEIAKL TKLKINKKGQ DGYLEYVDHG KKFVFPPIYI IESKTGIKPK ARSIGIAGLL	GIFQAQRDFD LLQKMTSLNY ELRKLLKLPD PEEMWQSRQL SKFGHLSIKS NPVVRRAVTK DKNRQTNDKI AFSEGEYVDH RLNALTNIAD YNYFRQAIEF DATVIAAITP ADLFNKIFNG KPVFVVRLAN VATGDGIYDP TKKLVRKVRV KDTVEQVLPN YSNKENNTSV KFEKYQVDYF	
Ignavibacteriu m album JCM 16511 WP_01456187 3.1	MEFKKVLGLD MKAFIDGKNG LPEDFPLDNP EIEQVIEEIN IRIIYLFNQR VSITKVKEVV TFLVTQKLEK LKAFKEKRGY RKLATVLYPS ECRYEKRKGI KKKINIDETQ ISKKEEETH LNRLWHS VAKAIANLPP LMLFDKNLIQ ELVSDQDLEK NSPDELGEYI ELGRELKNNS NKIFSSFTVN VKKYILWLTQ LVICELGVNK FKYQKAKYKN FTIGSITSEL HFDSLINPQL EEKREYFLSL FNKYLKWEYK SLKEAIKIQA RKQDDEKLNK YKAKRMPLNK KYIIEPHKKN PIKYITRLDG PSISVLKAIE SSNETIINWD GEFGSQNISE	IGTNSIGCAL LPQVITPAGK KRIKETISTE FRRKTGKGNK RGFKSSRKDL ELKTDGRKKG GKFKQNKPD KIRQYPVNRW QSKFFGPKLI DGEIYGLKCI LYINENIKEK RINLFANRKE NDYADKEKTE LKKEYGSYSA LTNNQRKVLN QVLKSFLEKK RKKLPNNSLR EERKKTSESQ PNPDSPDLIE KCRSPYTGKI AKGDRLAANF LLATEPPEDF KITWGLNGVW DKEGLKRLDH CNHKIRDFKL NGKWQKVFAI IWEEVKNDPV FINEAKVSAG KEYIEKLSLQ NPVNRLLEH DINEEIEFRG HKNKIDFFAP DNEFISNRIY YSVDDPPIRI	LSLPKSIQDY RRQKRGSRRL GKFSFRISDY NPMIKLLPED TETAILDYDE KRFKVILEDS PKEEDWALCT RYKKELEFIW EFENSVDLHI PKSSPLYQEF LFELFNSKDS LKGNETKSR KSILSSLGWK LAIRKMLVVM KYLTLAEVQ NESDYLKGLK NPIVEQVIRE EKNFQEKERA KFRIWKNQSG IPLSKLFD ISESNGKCKF IERQINDTRY KDILRPRFKR RHHALDATII PWENFTSEVK QIKNDRWKAV RKKKEKYIYD VNKNLNTTNK KMFNDLPNFI ILEYHNNPKE AVFETDKGSN NRLGFSRIIL QVKKFTGNSC KDVCIKIRVD	GKGRLEWLT KHRYKLRRSR VPISDESYRE WVVYYLRKKA FAKRLAEKEK RIEPIEIERK TALDNRMGSK TKQCQLNP ISEDIIYYQR RIWQDIHNIK LSEKDILELI RKVFKKLGFD NRNGKWEKSK RDGKYWQHPD KRSTLIKQKL TYQAGYLIYG TIFIVRDVWK RKLLKELLNS LTDEELNKKL VYEIEHIIPR GEVEYTLK IGRKLAEELLT LESINKKLI AATTREHVR SKLLSCVVS RRSMFKEPIG DYAQKVI TIYNLEGRFY LEKSILDNYP AFSTEGLEKL VYFVMYENNQ SPGDLVYVPT YFLKNDIASL RLGNVRPL	SRVPIPLDADY LIRVFKTLNW FYREFGYPEN LIKPTTKEEL YSAENYETKF EKPDWEGKEY HPGEFFDEL NNLNINKBIL DLKSQKSLIS VIRKESEVNG SLNIINSGIK GEYILNHP NYDVFNLP QIAKDQENTS NEIEHNPYKL KHSEKDVP SFGIIDEIHI SNFEHYDENG KDEKIPTIE SKMKNDSTNN GDYLYQCKDT PVVKDSKNII FQEDDPN LNSLNAADND KESKPILSDP TVWIKKIKEV VQELGLSSSM EKIKVAEYVL EILKELES NKKAINKIGK TKDREFLKPN NDQYVLKDN ILSYSASNGV	(SEQ ID NO: 83)

Table 4- Cas9 orthologs						
uncultured delta proteobacterium HF0070_07E19 ADI19058.1	MSSKAIDSLE TAEELNSTGN DELEEVVHQ KLPPDDES REGDYSNLMA RSPKLGNC PRRAQIDAC LQQQERKTLA LADLLPEASD QESLPEIGVP IRVELARDLG LRYRLWKEQN LCFTEENAQK AEEWKSALN KDRDKHTHHA TFRQDVLDVE LADLERLVEK RPILSKVTLL EYRVVLQRM EFQRSADSEV VQVNLGKVI	QLDLFKPQ KLISKAERG SNRTLWDVRA ADEEQGKINR RKLVFEEALQ IPSELRAPSS SQRSTSSYWQ NFRNWKQLEK KAITQLCELD PAGDRVPPFD KGRRLERIK CTCPYSGRMI SNRTPFEYLD DTRYLTSALA VDAIVVACTT EEIFITRQPK DASNRNLYEH DLMRGEENVH IRGYRTFNI K	TLGLDLG RKRRIRMLD BAVERKLTQ ATSRLREELK ILAFQRKQGH APSTEWFKFL IRRDFQIPDE IIGTGHPIQT FTTAAKISLE EMYNPVVNRV LDQLDKSKQN PVNSVLSEDT AADFORLEAI DHLRHHLPS DHLRHHLPS PAIVQQVTLY KVSGGIQTKD LKCLEESGD TELSGGRRYD VFQKGVYPDQ ANGQLKISTY	IGWAILSGER RKARRGRHIR ELAAVLFLV ASDCKTIGQF ELSKDFEKTY QNLGNLQISN YRFNLVNYER LDEAARLITL AMYRILPHMN LSQSRKLINA DQRAEDFRAE QIDHILPISQ SGNWPEAKRN KIQTVNGRIT HQDIRRYKKL TLRKHRSPD QPTKAFKAPF SMAQGRLDIY GPEIEQNYTF LEGRQDFDFF	IANAGVYLFE YLLEREGLPT RHRGYFPNTK LAQNRDRQRN LDVLMGQSRG AYREEWSIDA RDPDVLQBEY IKDDEKLSQD QGMGFFDACQ VIDEYGMPEAK FQQAPRGDQS SFDNSLSNKKV KLLHKSFGKV GYLRKQWGLE GEKRPTWPPE RQRVALTKVK YMPSPGPEAKQ RYKPGGKRKD LFSLYFDDLV GANRLAHFAK	(SEQ ID NO: 84)
Ruminococcus albus 8 WP_00284692 6.1	MGNYYLGLDV QQCRRSRGLR QLRNKGLSEK EEHDSIVLIS YLITRHLMLV NERFRFTGY RGESVFDTSF KCFKYIKVVK RREKLKALNI KYQAKFNKEI NAIIDKYGYP ECIKCDEVHA HIVPYSLILD MFKSKKCSK NLRFRDSYES AELKKLTYLD EQSKKACIDN NIYEQFWKDD YRGTTTGEEA KTIFEQADYK DDNNFSLND FNYPDDYYTH RFISDNKPCA LLKEKN	GIGSIGWAVI RLYRRKSHRK LTPEEIAACL NLMNEGGYCT EVDLILENQS LDSIGKCQFF ANDLINSALK PLFEKYGYDW GLDDGLINEL PDIDENAKPQ AAVNIETADE RHLIEKYKLW NTINNKA KYQYLMLPDL SDEDDLKIRD HAADAI LYKLFRMDRR KDKKSCEELY IRVKEIDGKL DVG DYLKKT SSYYCIELYK VMYIFPGDYL KDKRVSITTK	NIEKKRIEDF LRLKNYLSII IHICNNRGYK PAEMICNCRE KYYGILDDKT LDSIGKCQFF NGSMDKRELK DKLIENYTD DNNVLNRIGI TKLKLSTAN KLPPFKNEDD LNKTFEDRAI EAQEGKC AEENQKKGQR NDQELLGGWR HYRVFPVKSR CRPEYVVL TAEKLLSGHG RENVASLYKG IKLKRKSISE QHFFTSSGK DSKGDNNLQ RIKSTSKKSG DIVIKLAVDL	NVRIFKSGEI GLTTSEKIDY DFYEVNVEDI FDEPNSVYRK IAKIKDIIFA VIADYAFVN ATAKSYHIDI DNNVLNRIGI VSYKMQGSI CEFFKNPVVF DTKRNDNQK GETITKEDML TPLMYMNEAQ SRNLNDTRYI FTSMFRRWWL EKLKLNKMYH RLTPIIPNLS DPKFASSLSM ITAESINSIY RVNKVTVIEK IAMS EQLKFEGYFI MGKVQGENNG	QEKNRNSRAS YYETADNNVI EDPDERNEYK FHNSAASKNH QRDFEIGPGK LDSQYTYTNN SDKNSDTSLT VLSQAQTPKR EAFCEGDLYG RSINETRKL ENDRIVKEII RDKDKLFEVD AADYRVVNT CKYLVNLYLRK NEKTWGRYDK QAGKRITPEY EEVDKRLWDK PVISLKP TDDKILIDSL VPSRWLRKEI KTKKLYLKP SVKNVNENSF KGISCGEPLS	(SEQ ID NO: 85)
Lactobacillus faracinis KCTC 3681 WP_01001894 9.1	MTKKEQPYN KETRLNRSTR DPDRKRDYK ALHNILKYRG DWNHISDILI NTIFKTS ITLNEILNGE YDDYINKPKY SENGVVPYQL LQSAEKNPFA GEPVLPKNSL	GLDIGTSSVG RRYRRRKNRI LFIDGPTYDK NFTYEHQKFN GRGNATQKSS DEEKLNFSGK SYFSMAKV GTKEYTSLK NKIEME WMERKSNGHA IYQRYEVLNE	WAVTNDNYDL NWLNEIFSEE EYYREFPTIF ISNLNNLSK NILKDFTL DIESKLDDLD YENHAIDLCK KFLKVALPTN NQSQYYPFLK RPWNFDEIVD LNNIRITENL	LNKKKNLWG LAKTDPSFLI HLRKELILNK ELIELNQOLI ETKKLLKEVI SILDDDQFTV LRDMWHTTKN LAKEAEKIS ENKEKLLSIL REKSSNKFIR KTNPIGSRLT	VRLFEEAQT RLQNSWVSKK DKADIRLIYL KYDISFPDDC NLILGNVAHL LDAANRIYST EEAVEQSRQA KGTYLVKPRN SFRIPYVGP RMTVTDSYLV VETKQRIYNE	(SEQ ID NO: 86)

Table 4- Cas9 orthologs						
	LFKKYKKVTV FGSSFMEDNK SNMRYKGWGR MSNDKYDFKN EIVKFMGHAP QLREYLVPNK SDYQVDHILP WTYMLNNMI DNMYKNQGT HAQDAYLASF SRKNGFIISP EIKTGQFYDQ IDNNKIKPVA IIYSKKVGLL YDKNILVEIL EELITLLHAN TRIHIE	KKLTKWLIAQ NYDQIEELIE LSKKILMDIT YIENHNLNKN KHIFIEVTRE KIQEELKKHK RTYIPDDSLE GLKKFKNLTR CIQARANLST LGTYRLRRFP LVNGTTQYDR TIYSPKNPKY IPIRLINDLK SLNSDREVAN QELITKMKKF STSAHLIFNN	GGYKNPILIG WLTIFEDKQI TETNTPQLLQ EDQNISDLVN TKKSEITTSR NDLSSERIML NKALVLAKEN RVITDKDKLG AFRKALSGQD TNEMLLMNGE NTGEIIWNVG KKLIAQKKDM DKKTLQNWLE RQQLILPPEH YPFYKGEREF TEKKAFGKKT	LSQKDEFNST LNEKLHSSKY LSNYSILDLM DIHVSPALKR EKRIKRLQSK YFLQNGKSLY QRKADDLLN FIHRQLVQTS DTYHFKHPEL YNKFYGOVKE FRDKILKIFN DPNIYGGFSG ENVKHKKSIQ SALLRLLQIP LIANIENFNQ HGLTLNNTDF	LTTYLDMKKI SYTPDQIKKI WATNNNFISI GITQSIKIVQ LLNKANDFKP SEESLNINKL SNVIDRNLER QMVKGVANIL VKNRNVNDFH LYSKKKKLPD YHQC�VTRKT DNKSSITIVK IIKNNVPIGQ DEDLQILAF ATTSEKVNSL IYQSVTGLYE	
Eubacterium dolichum DSM 3991 WP_00480045 7.1	MMEVFMGRLV RRTKRGGRR ERLNGEELAT KSGKYVCEIQ SAIITIISRK FPNEPRAPKL ITPKQLAKEV LEDHVFYDEI YHSLSFKALR TAILSPVAKR ERQKFEMRN KLLIDDPNAY KGRFTKGSLA KEFINRNLVD NLKKDRDEDY SSGDDSMYYD IDGKEKVVK VVNYFFEEMS GPAITQMKYF YKFVTIRYKD LIGITKAEGE KPIHCYCKMR	LGLDIGITSV KRRRVTRRED ALLHLCKHRG KERLRTNGHI RMYDGGPGGP AYSAEFLNLL GVSLEQIRGF AEILTKTKDI LINEEMLKTE AQRETFFKVVN KQVADIIGDD EVDHIIPISI QYKAYCLKLK TSYASRVVLN GHHAIDALII DRYFAFIASL YKDIYDPKFT KSEKYFTKDK DGV LGNHIDI VRWSEKKKKY ALIYDPDETWH LMPTISKKIV	GFGIIDLDES MLHLLKQAGI SSVETIEDDE RGHENNFKTR LSPTPYGRYT NDLNNLSIEG RIDTKGSPLL EGRKKQISEL LNQMQSITLF RLREIYGEFD RKINAKLREK SLDDSITNKV EKNIKTNGY TLTTYFKQNE ASMPKMRLLS KAIKVRKFSH ALAEDILNNA KGRIKISGMN SAHYQVRDCK VIDQQDYAMK NFNFFFHAGE RIDKYATDVV	EIVDYGVRLEF ISTSFHPLNN AKAKEAGETK AYVDEAFQIL YFGQKEPIDL EKLTSQKAM SELTGYKMIR SSDLNEESVH GLKQNNELSV SIVVEMAREK LVLYQEODGK LVTHRENQEK RKKVEQYLLN IPTKVFTVKG TIFSRYKIED KIDTKPNRSV YQEKYLMALH PLSLYRDEHG VVLQQISPYR KAEKKIDDTY TPEILKFTAT GNLYKVKKNT	KEGTAAENET PYDVRVKGLN KVLSDNDQLL SHQDLSNELK IEKMRGKCSL ILKIVHEK GK EVLEKSND E QLAGLTKFTA KGMKNIQADD NSEEQRKAIR TAYSLEPIDL GNLTPISAFV ENDIYKYDIQ SLTNAFRRKI IYDESTGEVF ADETIYSTRV DPQTFDQIVK MLKKYSKKGD TDFYYSKENG EFQFSMRDE NNDKSNKIEV LKFEEF	(SEQ ID NO: 87)
Nitratifactor salsuginis DSM 16511 ADV46720.1	MKKILGVDLG SIRSTQKSMR NRWQLRAVDA LNDPEKESEK RSYRNHDDYE ACITDQEMPT GYEVTQEDRE DERIVKGKKE QRSKTPQEAL ALPLFLEGYD INNHAVKSLA MREREKALDK QLLDGSADIE DYRERIGMLS VAQVLKRYYP HAEDALILST ADYIDEAFRR	ITSFGYAILQ RLIEKRKKRI WKRPLSPQEL KADERRQVYN KMIRREDIEE IDESLFGKCT KVIEWVEKKI PRTFVPFFFL DRLRALMAGK EKEVQRILGF SWALGLIADL IIGKYKKEFP HIVPQSLGGL EKGLIDWKKR FPDPELRKNG LTRGWQNR LH FMSKGEESLF	ETGKDLYRCL RCVAQTMERY FAIFAHMAKH ALRHLEELRK EIEKVLLRQA FYKDELAAPA AQGKNLKKIT ADIAKFKELF GIDTDDRELL DDREDYSRYP SWRYGPFDEI SIDKRLARKI STDYNTIVTL KNLLAQSLDE IGVRMIPGKV RMLRDNYGKS YRDMFDTIRS	DNSVVMRNNP GILDYSETMK RGYKSIATED KYGGETIAQT ELGALGLPEE YSYLYDLYRL HKDLRKILGL ASIQKHPDAL ELFKNKRSGT KSLRHLHLRE ILETTRDALP QLWERQKGLD KSVNAAKGNR IYTENTHSGK TSKTRSL LGI EAE LKELWKK ISYWVDKKPL	YDEKSGESSQ INDPKNNPIK LIYELELELG IHRAVEAGDL QVSELIDELK YKKLADLNID APEQKIFGVE QIFRELAEIL RELSHRYILE GNLFEKEENP EKIRKEIDKA LYSGKVINLS LPGDWLAGNP IRATSYLEAL KSKSRETNFH YMPHIEGLTL SASSHKETVY	(SEQ ID NO: 88)

Table 4- Cas9 orthologs						
	SSRHEVPTLR IGNTNDESYR SPIEVTGKLL FRRMDVNNAH SKYIALFNPR RSYEVFGTLP	KNILEAFDSL AVEERATQIA RKMRFVYDKL ELQKERSGIL FPANPKAQPS EGSIEWFKEE	NVIKDRHKLT QILTRYQLMD NAMQIDRGLV CYLNEMLFIF KFTSDSKIKQ SGYGRVEDDP	TEEFMKRYDK AQNDKEIDEK ETDKNMLGIH NKKGLIHYGC VGIGSATGII HH	EIRQKLWLHR FQQALKELIT ISKGPNEKLI LRSYLEKGQG KAHLDLGDHV	
Rhodospirillum rubrum ATCC 11170 WP_01138821 2.1	MRPIEPWILG KDHTSRQAER LALRREAVSR PAKATAKATA RGQPVRMRYN LVLDQRPLRS ADEPRPLTDE AKQAARGTAG ALIGDPRGPT GIGYDEAVTL EDDGAAAEAY ELKAGADERK RRQNNLCYPY NKTKGDKTPW ARKSAEDEDQ LLRLAWDITP VQARLAALGR VTTRGLINQI SIRPAHRDGG KIAEIIDGFA GMTANTVIAG RPRWTHRVLV VKKMAVDGRL RKEKIRTTSC	LDIGTDSLW GAFRRARRQT PLAPDALWAA PAKEADDEAG QSDRDGVVAP KGAGPCAFLP EHAKALALLS NLTEAILAPL RVTEDETAEA ALGLHSHSRP YGRIGNISVH RMIAEQAERE STPIGHADLL EAFHDKPGWI GFLPRQLTDT GPAPRDLLPT SRVADAGLAD NQASGAGRIL SLHAATVFGV SPRMAKRFGA RSDGDGEDAG RFDRTQPAPP ALWPARLATG TALGRLRLSK	AVFSCEEKGP RTWPWRRDRL LLHLAHRGF FWEGAEAAALR TRALIAEELA GEDRALRALP TARFVEWPAL IPGWSGWDLD VADAIQIVLP RQERLARLPY IALNETRKIV RENAEIDVEL GDAYDIDHVI AQRDDFLARL GYIARVALRY PRDALRDDTA ALGLTLASLG DLRKWPRTNF RNRPDARVLV LLARYQAAHP LITPFRANPK PPPENARLVM KATALYAQLS KAT	PTAKELLGGG IALFQAAGLT RSNRIDKRER QRMAASGAPT EIVARQSSAY TVQDFIIRQT RRALGLKRGV RKDRVFSDLW TGRASLSAKA YAAALPDVGL NALLHRHGPI RKSDRWMANA PLARGGRDSL LSLVTNEPNA ARRFLDGLTP GGGKNRADHR EPPYPTFRAE QRKPVEKLF EVPPALAALA AAVRTMGNAV RLRRGDLVYW CPNINLNGDQ	VRLFDSGRDA PPAAETRQIA AAAKALAKAK VGALLADDLD PGLDWPVAVTR LANLRLPSTS KFTAETERNG AARQDRSALL ARAIQAQAMAP DGDPVGPPPA LRLVMVETTR RERRQVRVLA DNMVLCSQSDA RFADDAGERV VVATNGRLTG PPLAKAVEGA HHFIDAAMIA VMKQWDHIHP DANAKPLPAD VARDPAFGPR YEVWEIQVKG PLESGDRLFL GYCVQSABGI	(SEQ ID NO: 89)
Finegoldia magna ATCC 29328 WP_01229014 1.1	MKSEKKYYIG NTRIFRSGRR KKVSGKYSLF NQMMKRRGHF EILADVNEKR IISDEDILEK YDFLILQSIL VFKEDNGNGY DEDVKYLLGK FNNKKEEFTK RPWNFEKIVD LNNLRINGKP DNSEKNLEIA KYIEETYPDL INFLRNSSDN SVKKMIWQVI LYKAIKKDER IDLDKLFDS PSSIRNPNKI VETRQSTKAI HAHDAFLNIV KGKVIWTPEK GKIYLPKKD HLLSKFYEDK KSNDGSITVK DKIYQQFKAG	LDVGTNSVGW RNDRKGMRLO NDKNFSQKQY LIDGQISHVT TDKKNNLKEL IKEDNKEDFV KGKSTLSDAQ VSYIGYYLNK IEQENFLLKQ IEKIRKTFTF LHKSEEEFIK LDTDVKLKL SNMKSYIDFN SSSQIQKIIN LMQIIGSQNY RVTEBITKVM DSQYEKLLTG HYDKDHIIPQ YNYWKYLMK KELFEKFYQK AGDVWNREFT GRKLIVDTLN DRLQDVSKYG NTVLDYAINV PNEQMYWRVD KYKNRRTTDT	AVTDEFYNIL ILREIFEDEI FEKFPTIFHL DDKPLKEQLI IKGQDFNKQE LTGDSYEENL VERYDEHKKD NKKITAKKKI ISSINSVIPH RIPYYVGPLN RMLNQCTYLP EELFKKKTKV NILEDKFDVE LKYKDWGRLS SMKKDDSIIN EFISKEKYNR SKIIPVKASL SNPINYVKEN KPSVLISNES GYKAINGAFF LQLQDPKIII EISNLKKIEN IIEKYEIIDL	RAKGKDLWGV KKVDKDFYDR RKYLMEEHGK LLINDLLKIE GNILNSIFES QYFEEVLQEN LEILKKVIKK SNIEFTKYVK QIHLFELDKI DYHKNNGGNA EETVLPKSSI TLKSIRDYMV MVEDLIEKIT RKLLDGIKGT KKYIPQEISY MAKSEEEKKT SRKLYLYYTQ NLVLVNKNAN LIRNTPLTNE ASDLRKDMNT REGDKVKYSL HVKKGELFNA FLVEHTKSKK DKINYRTEII KYYKDAILTE DTLDNKQLYQ	RLFEKADTAA LDESKFWAED VDIRYYFLAI LEEELMDSIF IVTGKAKIKN ITLFNTLKST YDEDGKLFKQ GILEKQCDC LENLAKNYP WIFRNKGEKI LYSEYMLVNE RNNFADKEDF IHTGNKKLLK KKETEKTDTV EVVENLYVSP TISRKNKLLD MGRDMYTGEK QTTKGNIYPV ELGGFINRQL LKSREVNDLH SKDFTPRKRS TIAGKKDYKK RIRSIELFPL IDNFSYLIST EDRKIMESYI LLVAFISLSY	(SEQ ID NO: 90)

Table 4- Cas9 orthologs						
	KTSNNAVDFT	VIGLGTECGK	PRITNLPDNT	YLVYKSITGI	YEKRIRIK	
Eubacterium rectale ATCC 33656 WP_01274255 5.1	MNYTEKEKLF	MKYILALDIG	IASVGWAILD	KESETVIEAG	SNIFPEASAA	(SEQ ID NO: 91)
	DNQLRRDMRG	AKRNNRRLKT	RINDFIKLWE	NNNLSIPQFK	STEIVGLKVR	
	AITEEITLDE	LYLILYSYLK	HRGISYLEDA	LDDTVSGSSA	YANGLKLNK	
	ELETHYPCEI	QQRNLNTIGK	YRQSQIINE	NGEVLDSLNV	FTIGAYRKEI	
	QRVFEIQKKY	HPELTDEFCD	GYMLIFNRKR	KYYEGPGNEK	SRTDYGRFTT	
	KLDANGNYIT	EDNIFEKLI	KCSVYPDEL	AAAASYTAQE	YNVNLNLDNL	
	TINGRKLEEN	EKHEIVERIK	SSNTINMRKI	ISDCMGENID	DFAGARIDKS	
	GKEIFHKFEV	YNKMRKALLE	IGIDISNYSR	EELDEIGYIM	TINTDKEAMM	
	EAFQKSWIDL	SDDVKQCLIN	MRKTNGALFN	KWQSFSLKIM	NELIPEMYAQ	
	PKEQMTLLTE	MGVTKTQEE	FAGLKYIPVD	VVSEDFNPV	VRRSVRISFK	
	ILNAVLKKYK	ALDTIVIEIMP	RDRNSEEQKK	RINDSQKLE	KEMEYIEKKL	
	AVTYGIKLS	SDFSSQKQLS	LKLKLWNEQD	GICLYSGKTI	DPNDIINNPO	
	LFEIDHIIPR	SISFDDARSN	KVLVYRSENQ	KKGNQTPYYY	LTHSHSEWSF	
	EQYKATVMNL	SKKKEYAISR	KKIQNLLYSE	DITKMDVLKG	FINRNINDTS	
	YASRLVLNTI	QNFFMANEAD	TKVKVIKGSY	THQMRCNLKL	DKNRDESYSH	
	HAVDAMLIGY	SELGYEAYHK	LQGEFIDFET	GEILRKDMWD	ENMSDEVYAD	
	YLYGKKWANI	RNEVVKAEN	VKYWHYVMRK	SNRGLCNQTI	RGTREYDQKQ	
	YKINKLDIRT	KEGIKVFAL	AFSKKDSRE	RLLVYLNDRR	TFDDLCKIYE	
	DYSDAANPFV	QYEKETGDII	RKYSKKHNGP	RIDKLKYKDG	EVGACIDISH	
	KYGFEGKSKK	VILESLVPYR	MDVYYKEENH	SYLVGVKQKS	DIKFEKGRNV	
	IDEEAYARIL	VNEKMIQPGQ	SRADLENLGF	KFKLSFYKND	IIIEYKDGKI	
	YTERLVSRMT	PKQRNYIETK	PIDKAKFEKQ	NLVGLGKTKF	IKKYRYDILG	
	NKYSCSEKFP	TSFC				
Corynebacteriu m diphtheriae C7 (beta) AEX66236.1 WP_01431843 1.1	MKYHVGIDVG	TFSVGLAAIE	VDDAGMPIKT	LSLVSHIHDS	GLDPDKIKSA	(SEQ ID NO: 92)
	VTRLASSGIA	RRTRRLYRRK	RRRLQQLDKF	IQRQGWPIE	LEDYSDPLYP	
	WKVRAELAAS	YIADEKERGE	KLSVALRHIA	RHRGWRNPYA	KVSSLYLPDE	
	PSDAFKAIRE	EIKRASGQPV	PETATVGQMV	TLCELGLTKL	RGEGGVLSAR	
	LQQSDHAREI	QEICRMQEIG	QELYRKIIDV	VFAAESPKGS	ASSRVGKDPL	
	QPGKNRALK	SDAFQRYRIA	ALIGNLRVRV	DGEKRILSVE	EKNLVFDHLV	
	NLAPKKEPEW	VTIAEILGID	RGQLIGTATM	TDDGERAGAR	PPTHDTNRSI	
	VNSRIAPLVD	WWKTASALEQ	HAMVKALSNA	EVDDFDSPEG	AKVQAFADL	
	DDDVHAKLDS	LHLPVGRAAY	SEDTLVRLTR	RMLADGVLDY	TARLQEFIE	
	PSWTPPAPRI	GEPVGNPAVD	RVLKTVSRWL	ESATKTWGAP	ERVIEHVRE	
	GFVTEKRARE	MDGDMRRRAA	RNAKLFQEMQ	EKLNVOGKPS	ADLWRYQSV	
	QRQNCQCAVC	GSPITFSNSE	MDHIVPRAGQ	GSTNTRENLV	AVCHRCNQSK	
	GNTPFPAIWAK	NTSIEGVSVK	EAVERTRHVV	TDTGMRSTDF	KKFTKAVVER	
	FQRATMDEEI	DARSMSVAW	MANELRSRVA	QHFASHGTTV	RVYRGLTAE	
	ARRASGISGK	LEFLDGVGKS	RLDRRHHAID	AAVIAFTSDY	VAETLAVRSN	
	LKQSQAHREQ	APQWREFTGK	DAEHRAAWRV	WCQKMEKLSA	LLTEDLRDDR	
	VVMSNVRLR	LGNGSAHEET	IGKLSKVKL	SQLSVSDIDK	ASSEALWCAL	
	TREPDFDPKD	GLPANPERHI	RVNGTHVYAG	DNIGLFPVSA	GSIALRGGYA	
	ELGSSFFHAR	VYKITSGKKP	AFAMLRVYTI	DLLPYRNQDL	FSVELFKPQTM	
	SMRQAEKKLR	DALATGNAEY	LGWLVDDEL	VVDTSKIATD	QVKAVEAELG	
	TIRRWVRDGF	FGDTRLRLRP	LQMSKEGIKK	ESAPELSKII	DRPGWLPVAV	
	KLFSEGNVTV	VRRDSLGRVR	LESTAHLPTV	WKVQ		
Roseburia inulinivorans DSM 16841 WP_00788930 5.1	MNAEHGKEGL	LIMEENFQYR	IGLDIGITSV	GWAVLQNNQ	DEPVRTDLG	(SEQ ID NO: 93)
	VRIFDVAENP	KNGDALAAPR	RDARTTRRL	RRRRHRLERI	KFLQENGLI	
	EMDSFMERY	KGNLPDVYQL	RYEGLDRKLK	DEELAQVLIH	IAKHRGFRST	
	RKAETKEKEG	GAVLKATTEN	QKIMQEKGYR	TVGEMLYLDE	AFHTECLWNE	
	KGYVLTPRNR	PDDYKHTILR	SMLVEEVHAI	FAAQRAGNQ	KATEGLEEAY	
	VEIMTSQRSF	DMGPGQLPDG	KPSPYAMEGF	GDRVGKCTFE	KDEYRAPKAT	
	YTAEFLVALQ	KINHKLIDE	FGTGRFFSEE	ERKTIIGLLL	SSKELKYGTI	
	RKKLNIDPSL	KFNLSNYSK	KEGETEEERV	LDTEKAKFAS	MFWTYEYSKC	
	LKDRTEMPV	GEKADLFDRI	GEILTAYKND	DSRSSRLKEL	GLSGEEDGL	
	LDLSPAKYQR	VSLKAMRMQ	PYLEDEGLIYD	KACEAAGYDF	RALNDGNKKH	

Table 4- Cas9 orthologs						
	LLKGEEINAI MSKNFQDRTN QGGYCLYSGK QKGNRTPYEY FKERNLNDTK KRWGLMQKDR FPDEETGEIL FLLTHADVQR SAHDATIRSA ALVRQLLHLG GGTGIAANGE EWRVMDANF DIATASIAGF FH	VNDITNPVVK LEKEMKKRQQ KIPLEELFDG FGADEKRWED YITRVVYNMI STDRHHAMDA NRDNFTREQW ELDYPGWMYG RDYETRGVVI NDGKKAFAED MVRIDVFREN VFSLYSRDLI ANDSNFKFRG	RSVSQTIKVI ENERAKQQII GYDIDHILPY YEASVRLLR RQNLELEPFN VVIACCTDGM DEKFGVKVPL EEESPIEEGR TKVPLTDLKL FHKPKADGTE GKYYFVPVYT HVKSCKDIKT LGIQSLEIFE	NATIQKYGSP ELGKQNPTGQ SITFDDSYRN DYKKQQKLLK HPEKKKQVWA IHKISRYMQG PWNSFRDELD YINYIRPLFV NKDNEIEGYY GPVVRKVKIE ADVVRKVLPN NLVNGGLLLQ KCQVDILGNI	QAVNIELARE DILKYRLWND KVLVTAQENR KNFTTEERKE VNGAVTSYLR RELAYS RNFK IRLLNEDPKN SRMPNHKVTG DKDSDRLLYQ KKQTS GVMVR RAATHTKPYS KEIFAYYTGA SVVRHENRQE	
Alicyclophilus denitrificans K601 WP_01351712 7.1	MRSRLRYRLAL SSLAVTRRAA NPYALRAKGL AIGQLRQQMA KSYDLYIDRA PVKPGRCTLL QRDAVVTALE LARKELFGAA QAIVDVSLPE FEYDASEVED RYGKIANPTV EAQRRQADNQ DRRCPSYGVQ KRNRTPWDAR KDFLARALND GLDGEKNRND MPMPWPYRD RKADGSAGRE TVNDKGKWE SVRLEVDGAE KYAGSLQKAK	DLGSTSLGWA RAMRRRRDRL QEALLPGEFA EQGSRTVGEY MIEQEFDALW PEEERAPLAL TKAKLSFEQI WSGFDEALQD GYGNLSRKAL LVHPETGEIR HIGLNQVRMV RRNARIRRSI ISAAMLLSDE AEFEAQGWSY TRYLSRVAAE HRHHAVDACV HVERAVRHIW ISNLIRIHEA EVISTFRAYG RTMRIVKISG TRRVTISPIGEVRDPGFKG	LFRLDACNRP LKRKTRMQAK RALFHINQRR LWTRLQQGQG AAQAAFNP TL PSTQRFRIHQ RKLLKLSGSV EIVWQLVTEE ARIVPALRAA SVFKQLPYYG VNALIRRYGR AEVLGIGEER VEVEHILPFS EDILQRAERM YLRLVCPGTR IGVTDQGLMQ VSHRPDHGFE TQPLRHGVSA VVRAGGMGRL SNGQIFMAPI FDTGSNLC LD	TAVIKAGVRI LVEHGFFPAD GFKSNRKT DK VRARYREKPY FHEAARADLK EVNHLRLLDE QFNLEDAKRT GEGALIAWLQ VITYDKAVQA KALQRHVAFG PTEVVIELAR VRGSDIQKWI KTLDDSLNNR PLRKRYRFAP VIPGQLTALL RFATAS AQAR GAMMEETS YG DGQPLAYKGY RNPHEGQNGR HEANVDARNT FYHDAQGKLC	FSDGRNP KD G AGKRKALEQL KDND SGVLKK TTEEGKKRID DTLLHQ RPLR NLREVALTLA ELKGNATSAA THTGVDEARA AGFDHHSQLG SGKPEDPDEK DLKQSRQKQV CWEELSFDAA TVAMRQANRI DGYERWL GDD RGKFG LNDVL GDGLTRLVDG IRKDGS IKQR VGGSNYCIEI KLIMRLVIGD DKQDAFTYTS	(SEQ ID NO: 94)
Sphaerochaeta globosa str. Buddy WP_01360784 9.1	MSKKVSRRYE SDLVFVSSRI LSYSEQDVPD NHRGSSSVRT TNGLIGYRNS DRIVSAILFE IKMPMQEGAA SIIIVLQGKEK KRLSKYLMKH TYTEALERGM KEKRDSEGGF PQEITVELAR IERFRILLEDQ VVAHRQCNDI EEYAKYLQSK TSILRKAWNL YCSRSLVQMI ILEFMDLHAF DIGVKIGDYE KESLVQAAAV NKRTKTFVVK	EQAQEICQRL FIPSTGAAER PARLRFEDAV FLDEEKSSDD ESVKS KGPV NEKIVPEAGC KRYQSASFSD AIQKIAGFRF LLLTENEVVD ETGEFQELSV KPNTNSDEEK ELKVGAEKRE AFVCPYCLEH KGKRTPYAAF GFVSRFESDN QGIDDLGSR NTMSEQGKRA VSMKTDNDAN EVASAIRGRI LEESNRKLIE EPSNEVKGFA	GSRPYSIGLD RQKRGQRNSL VRANPYELRL KKLEEQQAMT PTRDIISNEI CPYFPDEKKL EQRHILFHIA RRLEEKSFWK ALKTVSLIGD WEQQSLLPYY YGRIANPVVH DIIKQQTQKE ISVADIAAGR SNTSAWGPIM SYIAKAAKEY HWSKDADTSP VEIEAMIPIP GALLKDTVYS TDKQPKWYPM SGKKPIQLSE FDTGSNLC LD	LGVGSIGVAV RHRANRLKFL KGLNEQLTSL EQLAKEKGIS DVLLQTQKQF PRCHFLNEER RSGTDITPKL RLSEEQKDDF YGPIGKTATQ GQILTGSTQA QTLNELRKLM KEAVLAYS KY ADV DHI FPRD HYLDETPGMW LRCLFNPNNV GYASEPNLSF ILGADTQGED EMKD KIEQLQ KTISKKALEL FYHDAQGKLC	AAAYDPIKKQP WKLLAERNLM ELGYALYHIA TFIEVLTA FN YQEILSDEYC RLWEAINNAR VQKEFPALKT FSAWNTNTPDD LLMKHLEDGL LMGKYWHS AF NELITILGAK CEPNNLDKRY DTADNSYGNK RKRKFETNE TAVGSLKGME HHGLDAIVAL EAQRELFRKK SKNEAALQKY VGGYYLISN GEIIRKI QAM	(SEQ ID NO: 95)

Table 4- Cas9 orthologs						
	NPSYKPAYMK KPGRTFVIII QLSSAGLVRY	QGYSLYVRLY TFTEMGSGYQ VSPLLVDKIE	QGDVCELTRAS IYFSNLAKSK KDEVALCGE	DLTEAESNLA KGQDTSFTLT	KTTHVRLPNA TIKNYDVRKV	
Fusobacterium nucleatum subsp. vincentii ATCC 49256 WP_00588864 9.1	MKKQKFSDDY AAERRVQRNS EDKNSKEKFT ALHSIFKSRG EKLEKICDS TDEYKKEEVE LNNILSDSNY NNYPAYIGLN FNEILNKIEL EENGVS TKDK NFEQKVDIEK VQVNDEF LNE IKDSFNSNYV KKIKNEYGDI SVMEALRRTN SLKRAILQTL LKLYDSCGN YTGREIDLDR NEYPVKKEIQ VNVQR TKEV HAKDAYLNIV ENSLEIVKKN LYDGKDNKLN KNEKNLIKYL LKNKKQLYLE TIDAVKERYN LKLWEKSLIL SSLGNNKELL	LGFDIGTNSV RRRLKRRKWR LFNDDNYKDY HFLFEGQNLK GKGLKDKEKE KEKISFREQI ISEAKVKLYE KEKDKKEVVE KTILPKQRIS LLKTFKFRIP SAEEFIKMT ENKRKIIDEL SYIKFKDIFG LNKDEIKKIN YNLMELLSSK KIYEEIKKIT DIANFSIDIK LLQNNDTYDI EKMKSFWRF GKILQQIEPE AGNVYNTKFT MEKNTVNITR EKYGYTSLK VSQKLLNPK KKYEQILKNA IEFNEMYDKF REFLKIFNKN EESVTGLFVK	GWCVTDL DYN LNLLEEIFSD DFYKQYPTIF EIKNFETLYN FKGIFNSDKQ YEDDKPIIYS EHKKDLKNLK KSRLKIDDLI DNGTLPYQIH YYVGPLNSYH NKCTYLN GED FKENKKVSEK EKLNLDIYKE SFKFNTWGR L FTLQESIDNE GRVPKKVFIE EMKNSLSSYD DHIYPRSKVI KEKNFISDEK IKIVYSKAEI EKP YRYLQEI FIKEEKGE L AAYFIYVEHE I IKKIYKEQT LKFVEDNQGE LEKLSSKDYK TYGKYEIKDS KIKL	VLRFNKKDMW EIMKIDSNFF HLRDEL IKNP NLISFLEDNG LVAIFKLSVG ILGEKIELLD YIIRKYNKEN KVIKGYLPKP EVELEKILEN KDKGGSWIV VIPKDSFLYS KFKEYLLVNQ ISEKSILWKC SEKLLTGIEF NKMNEVS YR MARGGDESMK NNSLRQKKLY KDDSF DN LVL YKRLTGKDDF ASSFREMDFD KENYDVKKIY NLNPIKKGET KKNKKVKTFE LIIDSYPYTF TEENYKFIYL NYINNKL YTN QTKEKLFSFP	GSRLFDEAKT RRLKESSLWL EKKDIRLIYL INKSIDKDNI SSVSLNDLFD IAKSFYDFMV YDKLFDKNE ERIEEKDKTI QSKYYDFLNY RKEEGKILPW EYIILNELNK IANRTVELKG LYGDDKKIFE INLETGECYS DLIEESYVSP NKKIPARQEQ LYYLQFGKCM VLKNENAEKS ELRGFMARQL IKVRELNDTH NYDIKNAWDK SNEIISIKPK RITRIDSTLI TGVD SNKKVE KKRNNNEKNE FLNSKEKFKK EDTGRIRLGQ	(SEQ ID NO: 96)
Pasteurella multocida subsp. multocida str. Pm70 WP_01090703 3.1	MQTTNLSYIL TGESLALSRR QAWELRVAGL LSGVAQNHQL LAELNLLFAQ THEKNEFKAA PYEKS KLYA ALENQGLKDT NALLVSLNFD KTSQLLPAIP LGKSFKERRE YEQQHGKCLY SENQNKGNQT DNKFIDRN LN RWGLIKAREN YEMVDQESGE NHQFVQPLFV LLENMVNKER EQVQKSGVLV NKAI VAHKNE ATGNISLKEH QRQPVR	GLDLG IASVG LARSTRRLIR ERRLSAIEWG LQSDDYRTPA QH QFGNPHCK KHTYSAERFV QVRKLLGLSE WQDLAKKPD L KFIELSLKSL AQEIRNPVVL IQKQEDNRT SGKEINI HRL PYEWLQGKIN DTRYIARFLS NNRH HALDAI IISPHFPEPW SRAPTRKMSG EPALYAGLKA RENNGVADNA DEWEEMDEGA DGEISKGKDG	WAVVEINENE RRAHRLLLAK AVLLHLIKHR ELALKKFAKE EHQQYMTTEL WLTCLNNLRI QAIFKHLRYS LDEIGTAFSL RKILPLMEQG RTLSQARKVI KRESAVQKFK NEKGYVEIDH SERWKNFVAL NYIQENLLLV VVACATPSMQ AYFRQEVNIR QGHMETIKSA RLAEFNQDPA SIVRTDVF I K KFKFSLFPND VYRVGVKLAL	DPIGLIDVGV RFLKREGILS GYLSKRKNES EGHIRNQGA LMWQKPALSG LEDGAERALN KENAESATFM YKTDEDIQQY KRYDQACREI NAIRQY GSP ELFSDFSSEP ALPFSRTWDD VLGSQCSAAK GKNKKNVFTP QKITRFIRFK VFDNHPD TVL KRLAEGISVL KAFATPFYKQ NNKF LVP IY LV ELKTKKEY SFEKYQVDEL	RIFERA E VPK TIDLEKGLPN QTNNKELGAL YTHTFNR L DL EAILKMLGKC EEERQLLINH ELKAWHAIRK LTNKVPNSVI YGHYGEANQ ARVHIETGRE KSKDILKFRL SFNNKVLVLA KQRLLTQVID NGQITALLRS EVHPYKIENR KEMLPDRPQA RIPLTQLKPN GGQQVKAIRV TWQVAKGILP FFGYYIGLDR GKNRQICRPQ	(SEQ ID NO: 97)

Table 4- Cas9 orthologs						
Alcanivorax pacificus W11- 5 WP_00873826 9.1	MRYRVGLDLG PKKAARRLAR TLRAMAVNER RLEEEEMVALA LFRQYLG LPA VKEEIRNAIF ADLRWGMGRR GKGLNMDRAA ADLGSPEQLD MGFEGGRSSY ATPAQGGRRS EMAREMKGGV LWIEQGHQCP HRECNDKGN EGEALTDESI RRRWGLDTVI RTDRRPDKRI VKVELPMPIL ERSGEKVDWL FEKRVSKGMT HYKMLLNDGF KGDTVKNIKT GGRNIHLLRL	TASVGA AVFS QQRQIDRRRA IELGQLRAVL SVQNKDSVTV IKGRPDCLPN FQRPLKSPAD AEMLNDHQKA LGGRDDL SGN TDDWSCR FMG SIKALKALTE KLEPPPLTGN YCESNISLEQ RTPYQAFGHD DEFADRQLHE PQVRFESGMP DHRHHLVDAI TIRDIALEAV VSRKSLTDLA PQQALREPIE AYMEVPCKEG GRVYTIKQIL CAE	MDEQGNPMEL SRLRRIAIVS LRMGKKRGY GEYLAARVEH MYALRHQIEH KVGRCSLQTN VIRELLNQOK TTLA AWRKLG KNGRPRNFSD WMIAPHWRET EVVDVALRQV KRFASERKKA ALSGAYTNFE DRRWRIEQR SSWLAKVTTQ VDEEGAEIT VTALTSRSLY RSVRISHKPD PEKKSIDVDK FQGNILRKVR ILYGVPNLVR GDGGGKLILT	IWHYERLFSE RRLGIAPGRN GTFKAVRKVG GLPSK LKVA EFERIWATQS LPRAPRAQIA ELSFRKIYKE LEDRWQELDE EFVAFMNELR PETHRVDEEA AQSIENGKT HILPRTL TQI ANALPKKSSR WLSSLGSDVY PEEF EK FRLQ QQYAKAWKVA RYPDGRFFEA VRANISRIVG CFYSKADDCV PSEAVGIKRA PVTETKPADL	PLVPDMGQLK DSGVHGN DVP EAGEVASGAS NNEY YAPEYA QFHDVMKDHG AQNFRIEKQM LERAGCPGPE VTQIQVINFL MTDGFDRLSK AIRECYPESL LGSVPAQIVV PTPARILRYQ GRKRSELVLA KTRLLLLKDF VSRGSLT AEL WEGHRVTREM DEKQRHGRVD TAYGIAQR LD EAIRLHISNI RIEHSRRRGH PESGDFIRFY LSAKWGRLKV	(SEQ ID NO: 98)
Mycoplasma mobile 163K AAT27519.1	MYFYKNKENK FETVDDSDDK DKNPKILVRN LDKSELIVLL FPIDKIEFY AMEEKSF FSR KIFKTLWESK ERLKRKILLS IINAYHSLTT GSYEINQNNN SEIYSEIAKL NQKNNFSHSN IEKIVVEVTR KLWLLRQQQG VIVNKL DNAK NDKGKFIKLY YSVSVNGKQ EEGKNDVKDD KETRELD DLE KNTQLFNETL KENELTKLHI LQEGKISFNE DQRQNF SFHN LKTQKEKLQI LGKKIKDQKQ	LNKKVVLGLD LLNETRRKKR FIEKYINPFS YFYLSLRGAF KISGKIRSTI IRNYSEGPGN IGKCSYDKKL RFLNKDSKSA IFKKHLINFE VLHIFDAISN REFSGTSSLS LFEKTWVEDL SSNNKHERKK YDAYSLRKIE KSNDSLAKQF TIDNLDEFDN TSNLRNQIAF VLIKDRSFNG KNNIKFKEKA YSGKYDKGKN FNHDKNLYET WVPILDNDFK TLYWVQIWVY INEEPILKIN IQKPVKKYFP	LGIASVGWCL GQRRNRRLRF KNLELKYKSV FDNPEDTKSK NLKFGHQDYL EKSFSKYGLY YRAPKNSFSA VEKILKEENI NYLISNENDL ILNKFSTIQD FGAYYKFIPN IASPTVKRSL SDFINRNLSD VGIKNNKETE SFDNFL LINA TIKKVEKLN LDNR TDKIKK LKIIWNEVKI IIRKIRYIKF KNQKDQYCFI KGDLFENE EK NWKKVNLTYM	TDISQKEDNK TRKRDFIKYL TNLPIGFHNL EMNKNEIEIF KEIKQVF EKQ ANENG NPELI KVFDITNKLT KFENLSEIAY RIRILEGYFE LISEGSKNYS RQTMNLLKEI YDIDHIVPRS TSYITNALVN REWKRP EGFK IISQYFRSFK LDELNEKLNQ EYGIKREEVT SSEEKETDEI SIDARNSKFE ELFYIVGRDE GEIFKK	FPIILHGVR L IDNNIELEF RKA AINEKYK DNESIKNAE NIDFMNYEKF INEKGQKIYT DWKHKN EYIS NKDDNKINLP SEKLFVPNEK FSNLKKDVKS TISYEEKALQ FKYSEKNLE LPNENTTLK ISFDDSFNL YLRNANGKAF HLTFSNSKYK SINSNDFLIR RIERLNVNYR MRFSRMVITK IEEFFDEDKL NYFKYFVNKK IFSQSNFLKI KDEIKINYEK KPQKLEIKYI	(SEQ ID NO: 99)
gamma proteobacteriu m HTCC5015	MTKNYISPIA QEGRRGKRHQ LNRRGYTYIS NKLVRFRAEG NVLKSTVDGH NFQLRLLRKY	IDLGAKFTGV VRGYKRRKMA EEVDEESMN KIPSNKNEFK KSRSEYLANI FNDPNMSGVS	ALYQYLEGAD KRLWLILDS SPLPFSEMP KLLDTALDGK KEDIKSNEEL YWDEKRLEKY	CTQEVAKGLL EYGIKREEVT DYFNSSAPLL YKDEKKELSE EKQISSKEID FYQWVQGWHT	VDDRGNVTWS EPLLKFINGL EQLAKLLSDK AWGNIL IASE GFYNLVGHLS KGGTDEAEKK	(SEQ ID NO: 100)

Table 4- Cas9 orthologs						
WP_00828423 9.1	NIILKTKGAP HYPKWKEWVG SITDAEKRND ENKLFKCGK IERRNVRGWC KNSGLIASKI CTYENIWRMQ DGQMASIIIEH EIKRGRGSAK GEIDHIIPRS YLKKQFSTSD FVPELKSEVT AHQIDPWSVS TASRLKTISS DGIYAERFLP VGAQSLSDWQ TQRKTELRAK IGIPAKSAWE KVRKDFSLPV DFKSPALVPQ FVSSLELAPG WNREIPSLLI	LLKTLKSLSA QLVKQNDNAY LAGYKRLKLS TPPRKEKLKS RLASQVQKTY GEALNIEPDE EEKVESLLTN IARKIAQHKI AKKAKELGEK LTGRTKKTIVF VNLIKKKIKT SLLAVKNITR KQRKMLASAE TSFEKTGWRS ILIDENGLMA ERIDGRYLYM FSDDSGKKMK NLLDEPLLET VDSVSGGFRV IAESKSVTPI SQNRFYIRFT PPRSNLFLE	DLTIPPYEDQ LNENVTLANA LGSEVDEFLL TLLSAVLGKN GVYLKEYGLQ VSRFASPHSL QLLSEIHGER AQINDVPKEF SKAGWVSKTE NSEANLIYCS TIQRFTEGGE VNGTQAWLAK PIWAKKDPQP ALIPDLIKVD GYDIDNSLKA SIDKVKAFDY TLDAIRKSLK YWGTMPPQE KRKTPNGYNY SSELVHLDKN VDEDQFERHF TGQKITFEYI	NNRRPPKCQS LHRIVERSRS LVKNIVDETK LSDDEQSSFI QLHKLEAGKK AQIFNIEGD KVPLKSAMCT SIDIPIIIES RIKTSSEGIC SKGNHDKGNR KLRSFSELSR KIASLLAEHL AASHVVDVAVC ALDRRPKYRR KGADVVFESL LQEKVGEKDI LTVTVNEIGK IWEKVYRKHF QLLAIDGYSA EIVYFDEWRK KSALRVNGIQ ANGANAEVKK	VLLSDEKLTM IDPYQLRLLI EAREGLWFET EEFWKSGTPK LDDKPLALLY VAGFNKTCRA RLSADSTRPF NQFSFTAEL PYTGAPLGGG VYVIEQLNDK EDQKAFRHAL DKQGRDYTLS TFLEALEQPH SPFLLFKGEE AAELLSNTHF RKEKCGFSGT PRNIPNQAHR VGFKKEGDNV IDISDSLDKQ DLDTVNKTFD AYSLRRA	
Planococcus antarcticus DSM 14505 ANU10858.1	MKNYTIGLDI LKRGMRRRYN RSLTEVLSSL GHFLQEGNWS LLLNRNLTKT YKETKTQVQL YVAKAKVSAF HKVEVLCKLD KAIDSNQFLQ VKQILAFRIP AEAFISMRMK FRHRLSYEMK KEVFGTQKEN EILHLKIQEK LDHMEQYSSV SPALKRGIWR ELTKTTLKND DIQNLMSYEV IEANQQYAMR TRQIIKHVRD MDALFAAALI FGNLRLQSPV PKVKQAKYVS IEHHQFLKEP FISTRELHNA PIYTKSSIQD ISKPKPPEEV	GVASVGWVCI RRKKRLQLLQ RISSRKYPTI EAASAEGMDD DQSKELTAMF SNENVEEVME KQYQKDMASL QFNKEAKFAE KQKGIQNAAI YYIGPLVKDT TCTYLKGQEV CWIIDNVFKQ AFATSLSGYI YPSITDVQRQ FMEVLKNKGF SVKIVEELVS PDLKSFIGEI DHILPQNFVK TLWERLHELK LLDERFSKSD QSILGKYGKN TGEEVSGVEY PKTEKFVHDE ESQLAKFLAE RQFEISYELM RVQKFVDQTL AIGYESITGL	DENYKILNYN SLFDSYITDS YHLRSDLIES QLELVTRYA GKEYEPFCKL LLEEBESALL KNLLDKTFGE TFYKDLKKLL PHQNSLYEAE TQSPFSWVER LPKSSLTYER YKTVSTKRLL SMKSILGAVV KLALVKLPGW GLEKKIQKMN IFGEPANIVL KSQGDQRFNE DDSLDNLALV LISSGKLGR IHLVKAGIVS FLAFDLSKKD MKHVYFELPW VKNHSICLVE KETNSPIIHA EKVKQLSERS YDFKSFEIGF KYRKPRSVVG	NRHAFGVHEF GFFSKTDSQH NKKMDLRLVY ELENLSPLDL VAGLGVSLHQ EAVQPFYQQV KVYRSYFISD EDKSKTSIGT KILRNQQAHY KGDAPIPTWN FEVLNELNGI DDNPAMTEEL GRFSRLLIDG QHQVDGTKKI EVAREDGEKK QRFWLYVTQQ MPEANQRKNQ KKPSFDEVVK KFRRFSEIPK RQKQWRSVKG FTFMKKEKEV RIIRTIPKYQ SVEELKIVFG EELKKAVAAN TKR	ESAESAAGR FWKNNNEFEN LALHNLVKYR SESQWKAET LFPSSQALA VLYELLKGET KNSQREYQKS TEKDEMLRII PFITTEWIEK FDEQIDKAAS QLRRTGAESD ELYDEHTGEI IYWIADVFE LPLDEQGGQSV RYEDIEELAG RTKSRKDQWE GKCLYTGKAL VGQNKMPLEI DKFIARQLVE IRDYNNKHHA SNKEFFLFKN GMFYKESIFS QETKFIDLKV KIWIEHFPYY LLIDQMNDNY AQRSDTFGSR	(SEQ ID NO: 101)
Prevotella sp. C561 WP_00901330 3.1	MTQKVLGLDL REYSLAAQRS TYDKRKGLFR DFEQPIERYK AGAMKASEEK RSQFQHEIET VGKCTLERSK	GTNSIGSAVR AHRRSRGLNE EYPIDDKDFN LGRALYHIAQ LSKGLSTYMK IFKFQQGLSV PRCAIGHPLF	NLDLSDDLQW VRRRLWATL AWILLDFNGD HRGFKSSKGE EHNLLTVGAA ESELyerLIS EKFRATLIN	QLEFFSSDIF NLLIKHGFCP GRPDYSSPYQ TLSQQETNSK FAQLEDEGVR EKKNVGTIFY NIKVRMSVDT	RSSVNKESNG MSSESLMRWC LRRELVTROF PSSTDEIPDV VRNNNDYRAI KRPLRSQRGN LDEQLPMKLR	(SEQ ID NO: 102)

Table 4- Cas9 orthologs						
	LDLYNECFLA CPITARFRKM YDAEEPEAVL KILMLGLKYS SIVNGLIAKY LMDANEQTDI EREQWKLYH LRRRVNQLLD KIKKILEEFY FWLEQGGQCM AHYNRFIKKN VEYWKQARR KNSQLVDTRV DRSLHSHHAI GLIQELEGLK VRRRLRKKGY LMKEGRPQVN MKGQFPAETS IKEQTYKSEK SDIEDIPEFI RLYKINRFES LNFLIMGENR	FVRTEFKFED LGEEWESFRV AFAQETLRLE DAVILAKVPE KSVSEERYFA LQKVRDRYQD PSQITIYRPV DGVISPDETR PKRDGISTDD YTGRNTINLSN HIPTDMPNYD AQNKDRKDQC ITRHAVLYLK DATTLLTIIPV NKLQMEVKVC IVGGVDNPRW FKDACEQGIY EYKRYFYAAV TDKKGNRLML QSNLVLMTTH DFVIKNGKII	IRKYLEKRLG EGQKERQAHS RKKAEELVRI LVDVSDEELL DHNYEYLLDE FFRSHERKFV SVGKDRSVLR MREMHLWKME SIFPHVDVQR SAKRDRMLEL RIGHNVSEIG QTGDALRGEI LKYKKSADS MIKKKGKNGMP GDLYAMCCYT DYKLRTGDMI LSTSKERGRS FNHR	IHFSYNDKTI KNNISFHRVS WSAMPQGYAM SIKDYLYLVE SDEKDIIRQI ESPKLGESFE LGNPDIGAII HKASYYGAIT GFASWDDLK DIKLHHIRHV NGKIREFRIY LLYKDNPAEL LGKTVDYQNL	NYKDSTSVAG YSIEDIWHFC LSQKAI RNIN AQVNYDKRIN ENSLGARRWS NYLTKKFP MV NPTVLRVLNT YNRIRHDENE TYNKDIKKYK SSDMNLTLCD VERVERLN RN FTVTEVTDGF LGIQSVDEKK LSFSGSEDRT HHIKNQALTP QFAKDDEGKV AIVDKELFAL RCEAPQSGLK SLYDYVCHRK YDLDNVNLRS PESIRSSVKS	
Alicyclobacillus hesperidum URH17-3-68 WP_00644656 6.1	MAYRLGLDIG ASLALPRREA SSPDVYALRV LLKAVQENEN NKAGEYRHTV NFDDGPGGDS ERFQLLTKVV LRKLDMKDT FIQDLSIETL FGHLSLKAIR GEVTNPVVR EKAQKENQKN KLDVTRLFEP YLRDKPEQKA TRYITRALMN DAGDKHHAAD GITQYVDKET NIRAIGLLTY GTEIQVVVSK AKAAFKEPVY RIDVFAKGGK FSLYPNDAMV DNSGELEGLG	ITSVGWAVVA RSARRRTRRR AGLDRCLIAE LMQSKGYRTV SRQAIVDEV PYGHGSVSPD NLRIRYQQED ESFAGLTYGL DQIGWILSVW KILPFLLEDGY ALSQAIKVVN NEQIKDEIRE GYAEVDHIIP KFIALVESIP HIQAWLLFDE AVVVACIGDS GEVFTWESFD SETEQIDPIF VALTELKLT KISKDGSVRT YYFVPIYVAD IKPSREVDIT IQGLEVFKEY	LEKDESGLKP RHRLWRVKRL EIARVLIHIA AEMLVSEATK KIFAAQRALG GVRQSIYERM GGRYPCELTQ NRSRNKTEDT KSDDNRRKKL SYDVACELAG AVIRKHGSPE ILGSAHVTGR YGISFDDSYD LSQKKKNHLL TASTRSKR VV LIQRVTKYDK ERKFLPNEPL VSRMPTRKVT DGEIKDYFRP PVRKVKIQEK VLKREL PNRL YKDRKEPVGC VVGPLGDTHP	VRIQDLGVRI LEQHGISLME HRRGFQSNRK TDAEGKLVHG NDVMSEEL ED VGSCTFETGE TERARVIDCA VFVEMKFYHE STLGLSDNVI YQFQGKTEYV SIHIELAREL DIVKYKLFKQ NKVLVKTEQN MDKRAIDLEQ CVNGAVTAYM FKRNALADRN EPWPFPRDEL GAAHKETIRS EDDPRLYN TL LTLGVPVHGG ATAHKPYSEW RIMYFVSANI VYKERRMPFR	FDKAEDSKTG QIEALYAQRT SEIKSDSAGK KKHGYVSNVR SYLKILCSQR KRAPRSSYSF YEQTKITYGK VRKALQRAGV EELLPLNGSK KQRLPLPLGE SKNLDERRKI QQEFCMYSGE RQKGRNSTPLE EGFRERNLSD RARWGLTKDR RYVQQVSKSE LARLSDDPSK PRIVKVDDNK RERLVQFGGD RGIAENGGMV RVVDDSYQFK ASASISLRTH VERKMN	(SEQ ID NO: 103)
Lactobacillus rhamnosus GG WP_01456997 7.1	MTKLNQPYGI ADRRGSRTTR DKDRFKYEKR HHILKSRGHF AKADDFKAKL AITGLKTKFN FEQIQELYRA HELAKTIRGL MGQANFMLKQ RWKMPYQLDE PYNFDEKVDR	GLDIGSNSIG RRLSRTRWRL LFNDRD AEF LTPGAAKDFN LDEQATPSDT LALGTEVDEA RLLNGIVPAG YDRYINGDDA RTKANGAIP I LLNFHIPYYV EASANTFIQR	FAVVDANSHL SFLRDFFAPH YEDYPSMYHL TDKVDLEDIF QKALVNLLLS DASNWQFSMG MSLSQAKVAD KPFLREDFVK QLQQRELDQI GPLITPKQQA MKTDTDTYLIG	LRLKGETAIG ITKIDPDFFL RLHLMTHTHK PALTEAYA QV SDGEKEIVKK QLDDKWSNIE YGQHKEDLEL ALTKEVTAHP IANQSKYYDW ESGENVFAWM EDVLPKQSL L	ARLFREGQSA RQKYSEISPK ADPREIFLAI YPDLELTFDL RKQVLTEFAK TSMTDQGTEI FKTYLKKLND NEVSEQLLNR LAAPNPVEAH VRKDPSGNIT YQKYEVLNEL	(SEQ ID NO: 104)

Table 4- Cas9 orthologs						
	NNVRINNECL RGLADEKRFL IFETKLAEIE LMLSNNHLMQ LRQVLRVVED NAAQELIDSA LSHYDIDHIL WRKWHEAGLI AAAQYPNTKI YLLKRYPKLA VWDKERDIRE KQLIPKKQGY KIASANLKSR LFQNAKYGLF ALQVYKAID ETLRQILIGL SGLFKRRIPL	GTDAQKRLIR SSLSTYHQLK WLDPPKINEL ILADETLKET IKHAANGQDP VRGELEDKIA PQSLIKDDSL SGRKLRLNML IAVKAGLSHQ PFFTYGEFAK LDRIYNFKRM PTQVYGGYTQ EKGKQLLNEI MVNSDTYYRN QVEKFFKLYD QANGTRSNVK ADL	EVFERHSSVT EILHEAIDDP SGIRYRGWQ MTELNQDKLK SWLFIETADG DKASFTDRLV DNRVLVNATI RPDEIDKFAK LREELDFPKN VDVKKFREFN LITHEVYFET ESGSYNALVR VVKQLAKRRK YQELWLSREN INQFRAKLS NLGIKTDLGL	IKQVADNLVA TKLLDIENII FSRKLLDGLK TDDIEDVIND TGTAGKRTQS LYFMQGGDI NREKNNVFAS GFVARQLVET RDVNHYYHAF FIGALTHAKK ADLFKQTIYA VAEADTTAYQ NWKPSANSFK QKLLKKLFSI AIERFEKLPI LQVGSIGIKLD	HGDFARRPEI TWSTVFEDHT LGNGHTVIQE AYTSPSNKKA RQKQIQTVYA YTGAFLNIDQ TLFAGKMKAT RQIIKLTEQI DAFLAARIGT NIIAKDTGEI AKDSKERGGS VIKISAQNAS IVIPRFGMGT KYEKTQMNHD NTDGNKIGKT KDTQIVYQSP	
Enterococcus faecalis TX0012 WP_00240890 1.1 EFT93846.1	MYSIGLDLGI RRLIRRKTNR IYKVTLHILK QLQRLKENNR TDDWIALFVQ PATNIFDKLI QKEYILTELM GYRNRKIFA LSESVKLLVL LEQFQLKSDV KYGKEQIRYI WSQEFQTTI HIIPISISLD WDDYQKWVES VLNTLQSFFT LCAVTSFVKV YERKTYQVKW KTEVKTLSG KTYERLLSIA KKNNGPLIRS RYDIYQDLET RFCFSLQKND FKQKPYNNGA	SSVGWSVIDE LKDAAKILAA KRGISYLDDEV VKTGINAQGN PGIAEEAGLI GKDFQGELRA TKEFTFRGVN EAGIDLATLP DRYKELSQSI RKRYSEYKKL TIEMPRDDNE QKNRRFLAKL DSINNKVVLV RHFSHKKENN NQETKVRVNV SRYHYAVKEE PNFREQLKPV KQKITTDYET ETTPDFQEVE LKYDYGKLNK KAYYTVQLYY GLEIEWKDSQ INLNIACYGK	RTGNVIDLGV VGFYEDKSLK DTEAAKESQD YQLNVFKVSA YRKRPPYHGP SGLSLSAQQY DVVKLLGVKK TETIDCLAKV STQSWHRFSL PTKDVLAIEY EDEKKRIKEL LYYYEQDGIC SKANQVKGQQ LLETNRNIFDS GSFTHTLRKK TGEKVMREID NLHPRIKFHS IGKIKDIYTL EKNQKVKRVK HINITKDSQG SDLRFVEGKY RYDVRFYNFQ EGKKLRKFNT	RLFSANKSEK NSCPYQLRVK YKEQVRKNAQ YANELATILK GNEANNSPYG NLLNDLTNLK ERLSGWRLDK LTLNTEREGI KTLHLLIPEL NPTVNKTVSQ HAKNSQRKND AYTGLPISPE TPYDAWMDGS EQVEKFLARN WGADLDKTRE FETGEIVNEM QVDRKANRKL DGWEAFKKKQ RSPFAVYCEE RPVEKTKNGR GITEKEYMKK SANSINFKGL DILGKKHYLF	NLERRTNRGG GLTEPLSRGE LLTKYTPGQI TQQAFYPNEL RWSDFQKTGE IDGEVPLSSE KKGKPEIHTLK ENTLAFELPE MNATSEQNTL AFKVIDALLV SQSYFMQKSG LLVSDSTEID FKKINGKFSN LNDTRYASRL THHHHAVDAT SYWEFKKSKK SDATIYSVRE DKLLMKDLDE NDIPAIQKYA KVTLLQSLKPY VAEQTKGQVV EQEMMPAENQ YEKEPKNIK	(SEQ ID NO: 105)
Candidatus Puniceispirillum marinum IMCC1322 WP_01304741 3.1	MRRGLDLGT ATRREARLTR IRKKALDEAI MKLGEGART DTLWAKQAAF SKALPSFQRF LTFKAMRAIL AWDRLDDEEQ PDGHGSLSKK LDYYGKALAG RLHGKPDDEVV IDSRESRQKF SLTLDDSMAN PYAKRWRFPL WVVTGRLLTSL MTSRGLLQKV	NSIGWCLLDL RRDRFIQRQ DPYEMGRAIF IGEFLADRQA NPSLYIEAAR RIYQELSRLA RKQGVVDYPV DSFILMLQDD AIDRILPVLR HVMGASGKFE IEIGRDLPNG QLWEQLAKEP KTVCFRQANR DAMKRFEADG LRGFWGLNSI SKAARRSEDL	GDDGEPVSIF KNLINALVKY HINQRRGFKS TNDTVRARRL ERLKEIVFFQ WIDHDGVAHR GFNLESDNRD QKGDDEVRSI DQGLIYYDAV DSDEKRYGTI ADGKRELERF VDRCCPFTGK DKGNRAPFDA GFLERQLNDT LRGHNTDDGT DLTRLFEGRI	RTGARIFSDG GLMPADEIQR NRKSADNEAG SGTNALYEFY RKLKPQEVGR ITASLALRDH HLIGNLTSCI LTQQYGLSDD KEAGLGEANL SNPTVHIALN QKEGRAKNER MMSISDLFSD FGNSPAGYDW RYISRYTTEY PAKKSRRDHR DPWDGFRDEV	RDPSLGLSLK QALAYKDPYP VVKQSIADLE PDRYMLEQEF CIFLSEDEDRI LFDELEHKKK MRDAKMMIGS VAEDCLDVRL YDPYAALSDK QVRVAVNELI ARDELKKLGH KVEIEHLLPF QEILGRSQNL ISTIIPKNKI HHIDAIVVG KKHIDAIVS	(SEQ ID NO: 106)

Table 4- Cas9 orthologs						
	HRPRKKSQGA PLIKSALLNE GVAYKGYKGD TARLIMRLRI DKNDTFKLTS	LHNDTAYGIV TAGLSGKSFE GNAYMDIYQD NDLLKLQDGE KSPGKLQSAS	EHAENGASTV NAVQKWCADN PTSSKWKGEI IEEIYRVQRL ARKVHISPTG	VHRVPITSLG SIKSLRIVET VSRFDANQKG SGSKILMAPH LIREG	KQSDIEKVRD VSIIPITDKE FIPSWQSQFP TEANVDARDR	
Oenococcus kitaharae DSM 17330 EHN59352.1	MARDYSVGLD RGYRTRRRRL SNQAHFAAGL IHHLVKYRGH EKKLSAALTD LMAIFGLDSS LKAADFGLTL GLYEKSKKFD FLKKIDDDQF ENSHELNKIT NDTITPWNLS VLQELNNIRV EDTQIEGLAD QTVFEDKKVL NADKIDNQTF NTTDGDEQDV EFARKTQESH YFLQQGKDMY RRKSDSSNYL RFIARSLVET NRDINDYHHA HAARLNAQSQ DYLHKVMNFR PVDLYGGFSS GKYPKAEMVL NTLLKTKEDN ALVKLQAHRD RLFSPSHFTL	IGTSSVGWAA SRRHWRLRLL LFDSKEQDKD FLIEGDVKAD KHGSKSQRAE AISKDEQKNY KMLLGDDKT GINAAYLALQ LPIQRTKNNG ALINFRVPYY QVVLDLDSGR SGRRLDIRAK VNGKNFNAL RRQLDQLDGL DVPRMNQSI YSLIDELAGP LTNSRKNQLS SGEKLNDNL PALIDKMRPF RQIIKNVASL FDALLFSTVG RVNPFQFVVG KITVTRRLKD AKPAYAALIK SHIPYQQLVK LVDILHNRLD DFNALNYEDK ADTDEFIFQS	IDNKYHLIRA NDIFAGPLTD FYRKYPTIYH SAFVHTFAD TAETAFDILD KFSFDDADID SAAMVRRFNE SDNEDDRKKA TIPHQLHRNE VGPLVEEEQK RFIERLTGTD QDAFEHLFKV TTYNYLVSVL SDHNREKLSR DTLYNTKMNL KEIKRGIVQS TLLKNAGLSE SNYDIDHIIP WSVLSKQGLL IDSHFGGETK QYTENSGLMK SMRNAAPGKL QKGQLYDESR FKNKFRLVNV KDGALVTISS LPEMTIESAF QQTLEIRILDA VTGLFSTQKT	KSKNLIGVRL FGDENFLARL LRLALMNDDQ AIQRYAESNN LQSKKQIQAI EKIADSEALL HQKDWEYIKS KKIFQDEISS LEQIEKQGI IADDGKNIPD TYLIGEPTLP QKTVSATNLK GREFVENPSN KHYTGWGRIS MEIINNAEDD FRILDDITKA LVTQVSQYDA QAYTKDNSLD SKHKFANLTR AVAIRSSLTA KGQLSDSAGN NPETGEITPE YPSVLHDAKS LRQWTYSDKN ATELHNFQEL YKAFDSILSF LHASPASSDL VAQLYQETK	FDSAVTAEKR KYSWVHPQDQ KHDLEEVYLA SDENLLGKID LKSUVGNQAN SDTEFEFLCD HIRNAKNAGN ADIPDDVKAD YYPFLKDTYQ PTNHWVVRKS KNSLLYQKFD DFLVQAGYIS EELLEETEL KKLLTTKIVQ FGVRAWIDKQ VGYAPKRIVL AALQNDRLYL NRVLVSNITN TRDFDDMEKE DMRRYVDIPK QYNRYIKEWI ENADWSIADL KASINFDKHK SEDIYLEQIR WLPLADYKLI AFNRYALHQN KKINLSSGFG	(SEQ ID NO: 107)
Helicobacter mustelae 12198 WP_01302238 9.1	MIRTLGIDIG ESLALPRTLA LFQTSKDFLS NDSGKIKKAI VGRSELREEL YQRPLKGFSD LTNRHGLCFS GLDYSGKNPE KELESQQLSK VEILQERGIF VVNALLEKYG LEVGLPNTY PLSRSLDDSQ NFSPSKKRKL LPDGKKEHIR MIQKYTTYLK QNIIPSHHVS NQGIVDNGAM GIIKDWLEMD ELEHLSKYAL LGEVLEHKPR	IASIGWAVIE RSARRRNARK PWELRERALLY AENSKRIKEE KTIFQIQQEL KIGKCSHIKK QEDMCVYLKG KAVFLSLPST EQIQTIKDAK SKPQPKNRQL GFHYFHIELT NNRLKCKLWK SNKVLCLTSS TQKNFKERNE IISGSMSTSM DKETHRLKSH HKVTGELHQE VRVDIFKSKD ENYEFCSFLF KNEDEEKMFT NRQNIALKTT	GEYTDKGLN KGRIQQVKHY RVLDKEELAR QCKTIGEMMY KSPWVNEELI GENSPYRACK ILQEAQKNEK FKLNKITQDR LNFSKHINLS LPPLSELAKE RDVCKAKSAR QOEYCLYSG NQEKSNTPT EDFLARNLVD RSFWGVQEK QKAQILREGD TVRTKEFYQ KGKFYAVPIY KNDCIKIQT DTDKEKNKTM PKHV	KEIVASGVRV LSKALGLDLE VILHIAKRRG KLYFQKSLNV YKLLGNPDAQ HAPSAEEFVA GLTYSKLLKL KTQDKIANIL LEALYHLLPL ESYFDIPNPV MQLEKINKKN EKITIDHLKD EWLGSDEKKW TG YIGRVTK RDHHLHHAQD HKLSLRWPM AFGGEEGVKK TYDFAIGKLP EMQEAVALAIY TRESCGIQGL KVFQVKLSV	FTKAENPKNK CFVQGEKLAT YDDITYGVED RNKKESYNRC SKQEREGLIF LTKSINFLKN LDLPSDFEFL GANKDWEAIL MREGKRYDEG LRRALSEFRK KSENDAAASQL QRALQIDHAF DMYVGRVYSS YIKHLSLFLP AIIIIACIEPS NFKDKIQESI ALKFGKIREI NKAIVQGKKN KSTNSAKATI KVFQVKLSV	(SEQ ID NO: 108)

Table 4- Cas9 orthologs						
Bradyrhizobium sp. BTAi1 WP_012044026.1	MKRTSLRAYR QSKQSNAAGR ECLDPYGLRA IKAAAAGRLQT YPTRAMVDAE CSLDPATSDQ EDQDKVARAL ILSARRHFGP AQRALSALLP GKLSPTGYLP RVVNELIRWH LLRKLGLPAS HLIPFSISWD ERAARLPRGK THPHRIHVLP AALTDQALLR HGLQARLHED DLVRAHVAGE LVPIRDKAGR DAPAAQPIMR FQTRHDDPED DATRIGWTRP	LGVDLGANSL RLARSARRRR KALDEVLPLH SMQACGARTL FEAIWAAQAP DVGDFRCAWS LQTDQLSFDE AWHDRSLDRQ DGYCRLGLRA YYGQWLQNDV GPPAEITVEL THNLLKLRLW DSAANKVVCN RWRFGPGARA GRLTALLRAT RMANAHDDTR TAYGTVEHPE RQQGKTLKAA IYKSYNAGEN VFKGDMRLID SFRWLFASFD KKWP	GWFFVWLDDH DRYLQRRGKL HVGRALFHLN GEFLNRRHQL HHPTMTAEAH HPLAQRFRIW IRGLLGLPSD IDIVALLES IKRVLPLMEA VGSDDERDTN TRDLKLSPRR DEQGDVASEC RYANREKGNR QFEELGDFQA WELNDLLPGS RKIEVLLPWP TEDGANLVYR VLSFAQRRDI AFVDILQAES HAGA EKFKI KLREWNALV RIDTLGQPWR	GQPEGLGPGG MGLLVKHGLM QRRGLFANRA RATVRARSPV DTIREAIFSQ QDVRNLAVVE ARFNLESRR LDEAAIIASL GRTYAEASAA ERRWGRLPNP LAELEREQAE PYTGEAIGLE TPFEAFGHRQ RLLNETSWLA DDRAAKSRKD TFRIDLETRL KTFVDISEKE AGHPNGIRHV GRWIARATT VRLSPSNNLL RKRGLTGSE	VRIFPDGRNP PADEPARKRL IEQGDKDASA GGDVQARYEF RAMKRPSIGK TGPTSSRLGK DHLKGATGA GT'THSLDEAA AGYDHALLPG TVHIGIGQLR NQRKNDKRTS RLVSDVDID GRPYDWADIA RVAKQYLAAY HRHHAIDALV KAMLVSHKPD IDRIRDRRLR RLTKSIKPDY FQANQANESH YLVEHHQAGV RKRGLTGSE	(SEQ ID NO: 109)
Acidaminococcus sp. D21 WP_009016219.1	MGKMYLGLD RSFRTSRRRL ETDKHIFND LVAHRGHFLN ATLLSRNSVN NKLFPQESND HALKDGRITIS TKNYVAYS EMIQLRTDNS KDAISNQDKL VDLDKSEAF YPIVDVKQQ SNYNTYHFFK TADDDVKHIS LMQLLSECYT LAVVNDIRKA FQQEVDFLEK SFYNIDHIYP WDAYYNHGLI LKRKFDPTEI VYHERFNRRW KQNKNTIHFT TAAYYLLVRF EILQKDKQV AMVPLIVPHK LFLQLQHNP CRSSGCDLKS SQNLLEYL	IGTNSVGYAV DRRQQRVKLV PTYTDKEYYS EVDKDNIGDV DKYKALKSLI ASFTLNDKED ESKVKLYEQH VKEVKGTLPK FMPKQVSGEN LSIMTFRIPY IRRMNTCTY VFGLFEKKRR SLMERGVLTR LRKHDFGRLS FSDEITKLQE CGTAPKRIFI ILENKSDGQL QSMVKDDSLD SLKKYQRLTR VYSKAGLSSD FMVNQPYSVK RFSFDRKEGL TLEDKKTQHK INIMFPMGTR IECYIKAMES YNKFFSTQFD INGSQAARI	TDPSYHLLKF QEIFAPVIS DYPTIHHLIV LSFDAFYPEF FGSQKPEDNF AIEEILGTLT HHDLTQLKYF NKATQEEFCK RVIPYQLYYY FVGPLRKDNS YPGEDVPLD VTVKDIQNLL DDVERIVERM KMFLTGLKGV AAYAKAQLSL EMARDGESKK QSDALYLYFA NKVLVQSEIN STPFTDDEKW FRHEFGLVKS TKTLFTHSIK FDIQPLKAST LMMIPVEGLY HIKLSNMISI FARKFKENNK VLTNGRSTFT MISADLTGLS	KGEPMWGAHV IDPRFFIRLH DLMESSEKHD LAFLSDNGVS DANISEDGLI PDECEWIAHI VKTYLAKEYD YVLGKVKNIE ELKTILNKA EHAWLERKAG SLIYEKFMIL LSLGALDKHG TYSDDTKRVR HKETGERASI NDFLDSMYIS KRSVTRREQI QLGRDMYTG GEKSSRYPLD DFINRQLVET RNINDLHHAK NGNFVAWNGE GLVPRKAGLD KARIDHDKEF DGFYLSIGGK LRIVEKFDKI KLSPEEQVQT KKYSDIRLVE	FAAGNQAER ESALWRDDVA PRLVYLAVAW PWVCESKALQ QLLAGKKVKV RRLFDWAIMK DIFRNVDS CSEADKVD SYLPFLTQCG KIYPWNFNDK NEINNIRIDG KLTGIDTTIH LWLNNNYGT LDFMWNTDN NAVKRPIYRT KNLYRSIRKD PIKLEHIKDQ AAIRNKMKPL RQSTKALAIL DAFLAIVTGN EDLGRIVKML VVKYGGYDKS LTDYAQT SSKGKSVLCH TVEDNLNLYE LLNILSIFKT QSASGLFVSK	(SEQ ID NO: 110)
Methylosinus trichosporium OB3b	MRVLGLDAGI GAKLKSEQRR GLDPWRIRAE SKMKRAVNET REGDYRSLL LQDSEKLVGP	ASLGWALIEI TFRGQRRVVR ALDRLLGPVE REKLARFGSA RDDLAEMRA CPFEVDEKRA	EESNRGELS RRRQRMNEVR LAVALGHIA AKMLVEDESF LFTAQARFQS PKRGYSFELF	GTIIGAGTWM RILHSHGLLP HRGFKSNSKG VLRQTPTKNG AIATADLQTA RFLSRLNHVT	FDAPEEKTQA SSDRDALKQP AKTNDPADD ASEIVRRFRN FTKAAFFQRP LRDGKQERTL	(SEQ ID NO: 111)

Table 4- Cas9 orthologs						
WP_00361103 4.1	TRDELALAAA RSGKAAEGTA ISEGLAQAGC GMTYDKACCA PTARKALIES KRNRQKDKLR ISPSQLVATD WFKAEKTDTE DTRWACRLLA WMKKSTKGDR LVKNVTPPWP EQRVYERRKV PKGDPPIFKVR KYEYYLVPIY TYLQVISSKG KKFNVDRFGR	DFGAAAKVSF RLRSVIVDAL NAPLVDALTA ADYDHTASRE IKQVKAIVER GLFEKEVGRP DAVQVDHILP WDAFIVRVEA EALKQLYPKG IPDDRHHALD GFREQAVEAV AELKLADLDR LVTKSKVNIA PHDIATMKTP EIFEGYYRGM KHEVERELRT	TALRKKLKL GELAWGALLC AASDGRFDPF RGAFDVGGHG YGVDPRIHVE PQDGARGKEE WSRFADDSYA LADMKGFKKR EKDKDGKERR AIVIAATTES EKVFVARAER VKDAERNARL LDTGNPKRPG PIRAVQAYKP NRSVGAIQLS WRGETWRGKA	ETTVFVG VKA SPEKLDKIAE TGAGHISSKA REALKRILQE LARDVGKSIE LLRFELWSEQ NKTLCMAKAN NYKL RNAEEA RVFSRPGALT LLQRATREVQ RRARGKAHDA IEKLRNWIEA TVDRGEMARV EDEWPEMDSS AHSNSSDVVQ YI	DEESKLDVVA VISFRSDIGR ARNILSGLRQ ERISRELVGS EREEITRGIE MGRCLYTDY QDKKGRTPYE AAKFNRNLN DRLRRAWGLQ EIEDKGLHYD TIRHIAVREG GSPKDDPPLS DVFRKASKKG YEFCSLVPM GIGARTLTFE	
Actinomyces colecianis DSM 15436 WP_00654647 9.1	MDNKNYRIGI KTNTTRLAMS YTPWLVRAL EQPPSDQYLA KTDTRPENKK ADSPKGASGE GSGADERLDV QTADGERASA AEKLTGTAA KRMINGEDL LMAAEAEWGA ADHINATSGV LGSTNTRDNL IADGFASSKE QYYFDEKHTG AMDAAATVAML FESWSENMRV KVGDAWSLTE GPLDKVGIFG APDLLRYRNE ELLLDLSET GEDVSEGSKS VSWQG	DVGLNSIGFC GVARRTRRLF AQTPIRDEND LKERVEAKTL PGFLGGKLMQ LVGYDVLPGQ ETQKRVEYFL KPPVDVTNVA EVEVAEFLQN FEARVNEFGV PLSVNIEHVR RGSDVTRYLA VATCERCNKS HRELQKGVKD TKVRVFRGSL RNSVAKTLVL LVEKFNLALY IDRASTPALW KAAASLLVRG DLFNVELPPQ SGQIAELQQD IIAGQGWRPA	AVEVDQHDTP RKRKRRLAAL LHEKLAIABR LQMPGATPA SDNANELRKI HGKRRAEKAH LNAKPTADIT FATCKIKPLK LSDEDNEKLD SEDWRPPAEP EGFISKRQAV IQRQNGECLY KSNKPFAVWA RLKRKVSDPE TSAARKASGF RGNIRASERA NDEVSIFSSL CALTRQPDFT GSVDIGSAIH SVSMRYAEPK FPGTTHWTVA VNVVFGSAMP	LGFLNLSVYR DRFIEAQGWT HIAHRGWRS EMVVALDLSV AKIQGLDDAL PAFQRYRIAS WSDVAEEIGV EWWLNADYEA SFSPLIGRAA IGARVGNPAV EIDRENQKRY CGTAITFVNS AECGIPGVSV IDNRSMESVA ESRVNFIGGN IGAAETWKSF RLQLGNGKAH KPTYGMVRVF VREAIREGKA GFFSPSRLRL EVIRRDGLGR	HDAGIDPNGK LPDHADYKDP PWVPRSLHV DVNLRPKNRE LRELIELVFA IVSNLRIRHL ERNLLMGAT RCVMVSALSH YSVDSLRLT DRVLKAVNRY QRNQAVRSQI EMDHI VPRAG AEALKRVDFW WMARELAHRV GKTRLD RRHH RGENVADRQI DDTITKLQMH RTIIVNGTHY KPTYGMRVVF EYLGLVVGVD RPVYLAQEGL KRRFSYSGLP	(SEQ ID NO: 112)
Caenispirillum salinarum AK4 WP_00954133 0.1	MPVLSPLSPN TLFPSAWSNE RSPEDLKDLT IRLAELQEVD RHGTQPTCGE FDAIRAIQAP FQGRITITREA RAMPGLGLSH DNSPVCFAQE ARGADSAALA WAAEAILDAW RTMAAARTPN EERNPFLRTW LAREAKHGV KERAALRLRL DGPDNLVLAH KGKGHKARTP	AAQRRRWSL NGTYVAHGAA RTPPKADPRA PPPESDADDA IMAGRLRETP RHPDLPWDSL IDRGLTVDPL GELTAPERDT RDTSGGGITV RLLAEGAHSV ANAPTEGFYD RAAQQRGPLP TGNAATDHIL RRNEIAKENR AQRQEFFCPY KDCNNAGKK REDKDFMDRV	ALDIGEGSIG DRAVRGQQQR IFFLRADAAR APAAATEDEDG AGAQPVTAR RRLVLDQAPI IQALRIRET LVRALMHDPD NPTDPLMARW PPVAAAAPPA VTRGLFGFAP AYESVIPSQL NQVRKTANEV ENEGRRKKES CAERPCLRAT TPHEHAGDLL GWRFEEDARA	WVAEVD AEG HDSRRRRLAG RPLDGPPELFR TRRAAADERA DGLRVGGGVA AVPPATPCLF GNLRLHERIT GLAAKDGRIP IDGWVDLPLK ATAAILES DI GEIVLEDLRR ITSLRRAHKG ITKYGNRRGW AALDTFCQDN DLFSPAETEI DSPALAALWQ KAEENQERRG	RVLQLTGTGV LARLCAPVLE VLHHMAAHRG FRRLMAEHMH VPTRALIEQE LEELRRRGET EPDGRQRYVP YTRLRLKIGY ARSLYVRDVV MQPGRYSVCP ARGALLAHL RAADWSAADP DPLPSRITVE TVSWQAGGLP DHVIERRMGG GWRKENADRL RRMLHDTARA	(SEQ ID NO: 113)

Table 4- Cas9 orthologs						
	TRLARLYLAA	AVMPEDPAEI	GAPPVETPPS	PEDPTGYTAI	YRTISRVQPV	
	NGSVTHMLRQ	RLLQRDKNRD	YQTHHAEDAC	LLLLAGPAVV	QAFNTEAAQH	
	GADAPDDRVP	DLMPTSDAYH	QRRRARALGR	VPLATVDAAL	ADIVMPESDR	
	QDPETGRVHW	RLTRAGRGLK	RRIDDLTRNC	VILSRPRRPS	ETGTPGALHN	
	ATHYGRREIT	VDGRTDTVVT	QRMNARDLVA	LLDNAKIVPA	ARLDAAAPGD	
	TILKEICTEI	ADRHDRVVDP	EGTHARRWIS	ARLAALVPAH	AEAVARDIAE	
	LADLDALADA	DRTPEQEARR	SALRQSPYLG	RAISAKKADG	RARAREQEIL	
	TRALLDPHWG	PRGLRHLIMR	EARAPSLVRI	RANKTDAFGR	PVPDAAVWVK	
	TDGNAVSQLW	RLTSVVTDDG	RRIPLPKPIE	KRIEISNLEY	ARLNLGDEGA	
	GVTGNNAAPP	PLRQDIDRLT	PLWRDHGTAP	GGYLGTAVGE	LEDKARSALR	
	GKAMRQTLTD	AGITAEAGWR	LDSEGAVCDL	EVAKGDTVKK	DGKTYKVGVI	
	TQGIFGMPVD	AAGSAPRTPE	DCEKFEEQYG	IKPWKAKGIP	LA	
Coriobacterium glomerans PW2 WP_01370957 5.1	MKLRGIEDDY	SIGLDMGTSS	VGWAVTDERG	TLAHFKRKPT	WGSRLFREAQ	(SEQ ID NO: 114)
	TAAVARMPRG	QRRRYVRRRW	RDLQLQKLFE	QQMEQADPDF	FIRLRQSRLL	
	RDDRAEEHAD	YRWPLFNDCK	FTERDYYQRF	PTIYHVRSWL	METDEQADIR	
	LIYLALHNIV	KHRGNFLREG	QSLSAKSARP	DEALNHLRET	LRVWSSERGF	
	ECSIADNGSI	LAMLTHPDL	PSDRRKKIAP	LFDVKSDDAA	ADKKLGIALA	
	GAVIGLKTEF	KNIFGDFPCE	DSSIYLSNDE	AVDAVRSACP	DDCAELFDRL	
	CEVYSAYVLQ	GLLSYAPGQT	ISANMVEKYR	RYGEDLALLK	KLVKIYAPDQ	
	YRMFFSGATY	PGTGIYDAAQ	ARGYTKYNLG	PKKSEYKPSE	SMQYDDFRKA	
	VEKLFAKTDA	RADERYMMM	DRFDKQQFLR	RLKTSDNCSI	YHQLHLEELK	
	AIVENQGRFY	PFLKRDADKL	VSLVSFRIPY	YVGPLSTRNA	RTDQHGENRF	
	AWSERKPGMQ	DEPIFPWNWE	SIIDRSKSAE	KFILRMTGMC	TYLQQEPVLP	
	KSSLLYEEFC	VLNELNGAHW	SIDGDDEHRF	DAADREGIIE	ELFRRKRTVS	
	YGDVAGWMER	ERNQIGAHVC	GGQGEKGFES	KLGSYIFFCK	DVFKVERLEQ	
	SDYPMIERII	LWNTLFEDRK	ILSQRLEKEY	GSRLSAEQIK	TICKKRFTGW	
	GRLESEKFLTG	ITVQVDEDSV	SIMDVLRGEC	PVSGKRGGRAM	VMMEILRDEE	
	LGFQKKVDDF	NRAFFAENAO	ALGVNELPGS	PAVRRSLNQS	IRIVDEIASI	
	AGKAPANIFI	EVTRDEDPKK	KGRRTKRRYN	DLKDALEAFK	KEDPELWREL	
	CETAPNDMDE	RLSLYFMQRG	KCLYSGRAID	IHQLSNAGIY	EVDHIIPRTY	
	VKDDSLNKA	LVEYREENQRK	TDMLLIDPEI	RRRMSGYWRM	LHEAKLIGDK	
	KFRNLLRSRI	DDKALKGFIA	RQLVETGQMV	KLVRSLLEAR	YPETNISVK	
	ASISHDLRTA	AELVKCREAN	DFHHAHDAFL	ACRVGLFIQK	RHPCVYENPI	
	GLSQVVRNYV	RQQADIFKRC	RTIPGSSGFI	VNSFMTSGFD	KETGEIFKDD	
	WDAEAEVEGI	RRSLNFRQCF	ISRMPFEDHG	VFWDATIYSP	RAKKTAAALPL	
	KQGLNPSRYG	SFSREQFAYF	FIYKARNPRK	EQTLFEFAQV	PVRLSAQIRQ	
	DENALERYAR	ELAKDQGLEF	IRIERSKILK	NQLIEIDGDR	LCITGKEEVR	
	NACELAFQD	EMRVIRMLVS	EKPVSRECVI	SLFNRILLHG	DQASRRLSKQ	
	LKLALLSEAF	SEASDNVQRN	VVLGLIAIFN	GSTNMVNLSD	IGGSKFAGNV	
	RIKYKELAS	PKVNVHLIDQ	SVTGMFERRT	KIGL		

[0106] 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.

[0107] 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

[0108] 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.

[0109] *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 20-hydroxyl group of U(-19) on the scissile phosphate and promote the formation of the 2',3'-cyclic phosphate product.

[0110] *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.

[0111] 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.

[0112] *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.

[0113] 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)

[0114] 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).

[0115] 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_C0015G0004 [Candidatus Peregrinibacteria bacterium GW2011_GWA2_3_10]	MSNFFKNFTN	LYELSKTLRF	ELKPVGDTLT	NMKDHLEYDE	KLQTFLKDQN	(SEQ ID NO: 116)
	IDDAYQALKP	QFDEIHEEFI	TDSLESKKAK	EIDFSEYLDL	FQEKKELNDS	
	EKKLRNKIGE	TFNKAGEKWK	KEKYPQYEWK	KGSKIANGAD	ILSCQDMLQF	
	IKYKNPEDEK	IKNYIDDTLK	GFFTYFGGFN	QNRANYYETK	KEASTAVATR	
	IVHENLPKFC	DNVIQFKHII	KRKKDGTVEK	TERKTEYLN	YQYLKNNNKI	
	TQIKDAETEK	MIESTPIAEK	IFDVYFSSC	LSQKQIEEYN	RIIGHYNLLI	
	NLYNQAKRSE	GKHLSANEEK	YKDLPKFKTL	YKQIGCGKKK	DLFYTIKCDT	
	EEEANKSRNE	GKESHSVEEI	INKAQEAINK	YFKSNNDCE	INTVPDFINY	
	ILTKENYEGV	YWSKAAMNTI	SDKYFANYHD	LQDRLKEAKV	FQKADKKSED	
	DIKIPEAIEL	SGLFGVLDL	ADWQTTLFKS	SILSNEDKLL	IITDSQTPSE	
	ALLKMIFNDI	EKNMESFLKE	TNDIITLKKY	KGNKEGTEKI	KQWFDYTLAI	
	NRMLKYFLVK	ENKIKGNSLD	TNISEALKTL	IYSDDAEWFK	WYDALRNYLT	
	QKPQDEAKEN	KLKLNFDNPS	LAGGWDVNKE	CSNFCVILKD	KNEKKYLAIM	
	KKGENTLQFK	EWTEGRGKNL	TKKSNPLFEI	NNCEILSKME	YDFWADVSKM	
	IPKCSTQLKA	VVNHFKQSDN	EFIFPIGYKV	TSGEKFREEC	KISKQDFELN	
	NKVFNKNELS	VTAMRYDLSS	TQEKQYIKAF	QKEYWELLFK	QEKRDTKLTN	
	NEIFNEWINF	CNKYSELLES	WERKYKDALT	NWINFCKYFL	SKYPKTTLFN	
	YSFKESENYN	SLDEFYRDVD	ICSYKLNINT	TINKSILDRL	VEEGKLYLFE	
	IKNQDSNDGK	SIGHKNLHT	IYWNAIFENF	DNRPKLNGEA	EIFYRKAIKSK	
	DKLGIVKGKK	TKNGTEIKN	YRFSKEKFIL	HVPITLNFCS	NNEYVNDIVN	
	TKFYNFNLH	FLGIDRGEKH	LAYYSLVNKN	GEIVDQGTIN	LPFTDKDGNQ	
	RSIKKEKYFY	NKQEDKWEAK	EVDWNYNDL	LDAMASNRDM	ARKNWQRIGT	
	IKEAKNGYVS	LVIRKIADLA	VNNERPFIIV	LEDLNTGFKR	SRQKIDKSVY	
	QKFELALAKK	LNFLVDKNAK	RDEIGSPTKA	LQLTPPVNNY	GDIENKKQAG	
	IMLYTRANYT	SQTDPATGWR	KTIYKAGPE	ETTYKKDGKI	KNKSVKQDII	
	ETFTDIGFDG	KDYFFEYDKG	EFVDEKTGEI	KPKKWRLYS	ENGKSLDRFR	
	GEREKDYEW	KIDKIDIVKI	LDDLFFVNFDK	NISLLKQLKE	GVELTRNNEH	
	GTGESLRFAL	NLIQQRNTG	NNERDNDFIL	SPVRDENGKH	FDSREYWDKE	
	TKGEKISMPS	SGDANGAFNI	ARKGIIMNAH	ILANSDSKDL	SLFVSDEEWD	
	LHLNNKTEWK	KQLNIFSSRK	AMAKRKK			
KKR91555_(modified) hypothetical protein UU43_C0004G0003 [Parcubacteria (Falkowbacteria) bacterium GW2011_GWA2_4_1_14]	MLFFMSTDIT	NKPREKGVFD	NFTNLYEFSK	TLTFGLIPLK	WDDNKKMIVE	(SEQ ID NO: 117)
	DEDFSVLRKY	GVIEEDKRIA	ESIKIAKFYL	NILHRELIGK	VLGSLKFEKK	
	NLENYDRLLG	EIEKNKNEN	ISEDKKKEIR	KNFKKELSIA	QDILLKKVGE	
	VFESNGSGIL	SSKNCLDELT	KRFTRQEVDK	LRRENKDIGV	EYPDVAYREK	
	DGKEETKSFF	AMDVGYLDDF	HKNRKQLYSV	KGKKNLSLGR	ILDNFEIFCK	
	NKKLYEKYKN	LDIDFSEIER	NFNLTLEKVF	DFDNYNERLT	QEGLDEYAKI	
	LGGESNKQER	TANIHGLNQI	INLYIQKKQS	EQKAEQKETG	KKKIKFNKKD	
	YPTFTCLQKQ	ILSQVFRKEI	IIESDRDLIR	ELKFFVEESK	EKVDKARGII	
	EFLLNHEEND	IDLAMVYLPK	SKINSFVYKV	FKEPQDFLSV	FQDGASNLDF	
	VSFDKIKTHL	ENNKLTYKIF	FKTLIKENHD	FESFLILLQQ	EIDLLIDGGE	
	TVTLLGGKKES	ITSLDEKKNR	LKEKLGWFEG	KVRENEKMKD	EEEGEFCSTV	
	LAYSQAVLNI	TKRAEIFWLN	EKQDAKVGED	NKDMIFYKKF	DEFADDGFAP	
	FFYFDKFGNY	LKRRSRNTTK	EIKLHFGNDD	LLEGWDMNKE	PEYWSFILRD	
	RNQYYLGIGK	KDGEIFHKKL	GNSVEAVKEA	YELNEADFY	EKIDYKQLNI	
	DRFEGIAFPK	KTTEEAFFRQ	VCKKRADEFL	GGDTYEFKIL	LAIKKEYDDF	
	KARRQKEKDW	DSKFSKEKMS	KLIEYYITCL	GKRDDWKRFN	LNFRQPKYEY	
	DRSDFVRHIQ	RQAYWIDPRK	VSKDYVDKKV	AEGEMFLFKV	HNKDFYDFER	
	KSEDKKNHTA	NLFTQYLLEL	FSCENIKNIK	SKDLIESIFE	LDGKAEIRFR	
	PKTDDVKLKI	YQKKGKDVY	ADKRDGNKEK	EVIQHRRFAK	DALTLLHLKIR	
	LNFGKHVNLF	DFNKLNVNTE	FAKVPVKILG	MDRGENNLIY	YCFLDEHGEI	
	ENGKCGSLNR	VGEQITLED	DKKVKPVDY	FQLLVDRREG	RDWEQKNWQK	
	MTRIKDLKKA	YLGNVVSWIS	KEMLSGIKEG	VVTIGVLEDL	NSNFKRTRFF	
	RERQVYQGFE	KALVNLKGYL	VDKKYDNYRN	VYQFAPIVDS	VEEMEKNKQI	
	GTLVYVPASY	TSKICPHPKC	GWRRERLYMKN	SASKEKIVGL	LKSDGIKISY	

Table 5 - Cas12a orthologs						
	DQKNDRFYFE FVDGTDGSIT NLIMQIRNYD YNIARKGFMQ BEN	YQWEQEHSKD NKLKSLKKG KEKSGSEADY LCRIRENPQE	GKKKKYSGVD GIELDNINQQ IQCPSCLFDS PMKLITNREW	KVFSNVSRMR IVNQKQKELGV RKPEMNGKLS DEAVREWDIY	WDVEQKKSID EFFQSIIFYF AITNGDANGA SAAQKIPVLS	
KDN25524_(modified) hypothetical protein MBO_03467 [Moraxella bovoculi 237] > WP_052585281.1 type V CRISPR-associated protein Cpf1 [Moraxella bovoculi]	MLFQDFTHLY VILDDYHRDF AVLRKEIVKP SPKLAHLAHF DNLQILTITK AYNTLLGGIS SDGMSVSFLP GIYVEHKNLN AKAKLTKEKD LAGVDNPIQK KELLDNALNV VRDYLKQKPF LLDKAHKKVF PSKELVEIKD FKKDKKGREV VDQYNIYKKI EFSKKLQDIG YLFQIYNKDF KASLDMNETT PITMNFVQOG KGEILEQCSL IKELKSGYLS FENALIKKLN VPAWNTSKID HIDYAKFTDK KSLFARHHIN EDFILSPVAN SDDLNVKLA	PLSKTVRFEL IADMMEVKL IGNGGKYKAG EKFSTYFTGF QKHSALYDQI GEAGSPKIQG SKFADDSEMC ELSKQAFGDF KFIKGVHSLA IHNNHSTIKG AHFAKLTTK STEKYKLNFG DNAPNTGKSI KGRQRSDDER PISEKDLFDK DSNDRKKNEN CYVDVNELEFT SPKAGHKPNL IHRAGEVLEN MTIKEFNKKV NDITTASANG HVVHQISQLM HLVLKDKADD PETGFVDLLK AKNSRQIWTI EKQPNLVMDI DEGVFFNSAL IDNQTWLNFA	KPIDRTLEHI TKLAEFYDVY YDRLFGAKLF HDNRKNMYSY INELTASGLD INELINSHHN QAVNEFYRHY ALLGRVLDGY SLEQAIEHYT FLERERPAGE TTLDNQDGNF NPILLNGWDL YQKMIYKYLE LKLYRFILEC INGIFSSKPK FYNHHPKFKK EIETRLNLYK HTLYFKALFS KNPDNPKKRQ NQSIOQYDEV TQMTTPYHKI LKYNIAIVVLE EIGSYKNALQ PRYENIAQSQ CSHGDKRYVY CQNNDKFHK ADDTQPQNAD QNR	HAKNFLSQDE LKFRKNPKDD KDGKELGDLA EDKHTAIAYR VSLASHLDGY QHCHKSERIA ADVFAKVQSL YVDVVNPEFN ARHDDSVQA RALPKIKSGK YGEFGVLYDE NKEKDNFGVI VRKQFPKVFF LKIHPKYDKK LEMEDFFIGE DLVRYYYESM ISFCNINADY EDNLADPIYK FVYDIIKDKR NVIGIDRGER LDKREIERLN DLNFGFKRGR LTNNFTDLKS AFFGKFDKIC DKTANQNKA SLMYLLKTLL ANGAYHIALK	TMADMHQKVK ELQKQLKDLQ KFVIAQEGES LIHENLPRFI HKLLTQEGIT KLRPLHKQIL FDGFDDHQKD ERFAKAKTDN GKLGQYFKHG NPMTQLRQL LAKIPTLYNK LQKDGQYILA SKEAIAINYH FEGAIGDIQL FKRYNPQDL CKHEEWEESE IDELVEQQGL LNGEAQIFYR YTQDKFMLHV HLLYLTVINS ARVGWGEIET FKVEKQIYQN IGKQTGFLFY YNADKDYFEP AKGINVNDL ALRYSNASSD GLWLLNELKN	(SEQ ID NO: 118)
KKT48220_(modified) hypothetical protein UW39_C0001G004 4 [Parcubacteria bacterium GW2011_GWC2_4_17]	MENIFDQFIG AKFYFDSLHQ LRKRDKNAV YQEREVDGKK NQPSLFVEEK ATRIADNFII IEHIKNESY GEVEKEKQLI EFSEYEGYI KKFISIAEIK YLFRDIEREN GLGIFQMMKY NFITKKPFDE HKNHRKLFQS FKPSQEILRI QQQDTGKYQN NKDIYNTKNG REKITTQKNQ NKIINQRIS NGVNYRDKLI YSAIVVLEDL PGGVNLGYQL SNSASLDKIK WAIFASAPRI EIIKKEKAGD GEVIAVDEGV	KYSLSKTLRF KFIDAALASD IDRLQKEIND ITFSKADLKQ ENPGQKRYIF FHQNTKVFRD NKIIGRINKK EKTREKTEED LKNKAINITIS NAVEELDGI GEKLLGYDSC FSLDDKDRKN DKIKLNFENG MGNAKGDNAN KKEKTFKRES IKEFTDDVQK SKNLHTLYFE CILDKGDRA SDNEMRVNVI EREKERLKNR NMRFKQIRGG SAPFVSFEKM EAVKKFDAIG RRQKGEDGYW LQGEKELDGR DFIASPVKPF	ELKPVGKTED KTSELSFQNF AEDIIQKEKE KGADFLTAAG DSFDKFAGYL KYKNNHTDLG IKEYRDQKAK VLIERFKEFI RRWFVSDRDF FKAVFYDKKI LKI AKQLGIF TPGQLSTNFY ALLKGWDENK RYQKMIYKQI KNFSLRDLHA YGYKISFRDI HILSAENLND KYRRYTEKKI GIDRGEKNLL QSWKPVVKIK IERSVYQFE RKQTGILFYT WDGKEQSYFF KYDRVKVNEE LRNFWSFIY FTTPNPYIPS	FLKINKVFEK ADVLEKQNKI KIYKDVRTL ILKVLKYEF TKFQQTKKNL FDEENIFEIE DTKLTKSDPF ENNEERFTAA ELKLPQQKSK IAQGGSKLEQ PQEKEAREKA AEYDGYKDF EYDFMGVILK ADASKDVPR LIEYYRNCIP DDEYINQALN PVFKLSGMAE MFHMSLVNLT YYSVVKQNGE DLKKGYSISV KALIDKLGYL QAEYTSKTD KYNPNYLADE FEKLLKVWNF LFNLVLELRN NLCWLAVENA	DQTIDDSYNQ ILDKKREMG DNEAESWKTY EEKEKEFQAK YAADGTSTAV RYKNCLLQRE FFKNLDKQIL KKLMNAFCNG NKSEKNEPKV FLVIWKYEF TAVIKNYADA EFIKYYNEFR KEGRLYLGIM LLTSKKAMEK QYSNWSFYDF EGKMYLFEVV IFQRQPSVNE GKGEIKQVQF IEQASLNEI IHKICQLIEK VFKDNRLRA ITGFRKNVYI KYKNSTVSKE VNPKATDIKQ SFSLQIKIKA DANGAYNIAR	(SEQ ID NO: 119)

Table 5 - Cas12a orthologs						
	KGVMILKKIR DH	EHAKKDPEFK	KLPNLFISNA	EWDEAARDWG	KYAGTTALNL	
WP_031492824_(modified) hypothetical protein [Succinivibrio dextrinosolvens]	MSSLTKFTNK KIIVDDFLRD VSKFETFDLF FKLVLPSTYK AYRIVHDNFP NYFTVGAYDY SELKSKLKNR EQCKDNNVIF LRNDIEDSAN TKFSDLLSCT YNTLLIFNCK DKFKLNFNSP TQAIADNTDN SDKEKFASPL NEWIAFCKEF CPIKTSFIEN LNNPTIMLNG RNEIYQYENK PLTINYKEGD EILDSVSFNT LKEGYLSAIV KMLINKLNYF PAAYTSKIDP SFDLDSLSKK FEMKKLLNEY VTNGKEDVLI ERNNLVREEK	YSKQLTIKNE FINKALNNTQ SSYSIKKDEK TTNQDKLKII KFLDNIRCFN FLSQNGIDFY HAFKMAVLFK NLLNLIKRIA SKQGNKELAK LHKVASEKLV SFNKNNGFYV QLGEGFSKSK CIFKMNYFLL VIKKSTFLLA LKTYYKAATIF LIDNGDLYLF GAELFYRKES FIDTSLDEAK TKQFNNEVLS VTNKSSKIEQ HEICLLMIKH VSKKESDWNK TTGFANVLNL GFSSFVKFSK KVSFDLENNL SPVKNAKGEF DTKKIMAISN	LIPVGKTLEN IGNWRELADA IIDDDNDVEE SSFDNFSTYF VWQTECPQLI NNIIGGLPAF QILSDREKSF FLSDDDELGI KIKTNKGDVE KVNEDGWPKH DYDRGINELS ENDCLTLLFK KDAKKFIPKC TAHVKGKKGN DITTLKKAEE RINNKDFSSK IEQKNRITHK KVLPNVIKKE FLRGNPDINI TVDYEEKLAV NATVVLENLN PSGLLNGQL SKVRNVDAIK SKWNVYTFGE IPNLTSANLK FVSGTHNKTL VDWFEYVQKR	IKENGLIDGD LNKEDEDNIE BELDLGKKTS RGFFENRKN VKADNYLKSK AGHEKIQGLN VIDEFESDAQ FIEGKYLSSV KAISKYEFSL LKNNEEKQKI SVVYLYNKTR KDDNYYVGII SIQLKEVKAH IKKFQKEYSK YADIVEFYKD STGTKNLHTL AGSILVNKVC ATHDITKDKR IGIDRGERNL REKERIEAKR AGFKRIRGGL SDQFESFEKL SFFSNFNEIS RIIKPKNKQG DTFWKELFFI PQDCDANGAY RGVL	EQLNENYQKA KLQDKIRGII SFKYIFKKNL FTKKPISTSI NVIADKSLA BFINQEQQKD VIDAVKNFYA SQKLYSDWSK SELNSIVHDN KEPLDALLEI NYCTKKPYNT RKGAKINFDD FKKSEDDYIL ENPTEYRNSL VDNLQYKLEF YLQAI FDERN KDGTSLLDDKI FTSDKFFFFHC IYVTVINQKG SWDSISKIAT SEKSVYQKFE GIQSGFIFYV YSKKEALFKF YREDKRINLT FKTTLQLRNS HIALKGLMIL	(SEQ ID NO: 120)
KKT50231_(modified) hypothetical protein UW40_C0007G000 6 [Parcubacteria bacterium GW2011_GWF2_4 4_17]	MKPVGKTEDE TSELSFQNF EDIIQKEKEK GADFLTAAGI SFDKFAGYLT YKNNHTDLGF KEYRDQKAKD LIERFKEFIE RWFVSDRDFE KAVFYDKKII KIAQQLGIFP PGQLSTNFYA LLKGWDENKE YQKMIYQIA NFSRLDLHAL GYKISFRDID ILSAENLNDP YRRYTEKKIM IDRGEKNLLY SWKPVVKIKD ERSVYQQFEK QQTGILFYTQ DGKEQSYFFK YDRVKVNEEF RNFVHWSFIYL TTPNPYIPSN LPNLFISNAE	LKINKVFEKD DVLEKQNKII IYKDVRTLFD LKVLYEFPE KFQQTKKNLY DEENIFEIER TKLTKSDFFP NNEBRFTAAC LKLPPQKSKN AQGGSKLEQF QEKEAREKAT EYDGYKDFE YDFMGVILKK DASKDVPRLL IEYYRNCIPQ DEYINQALNE VFKLSGMAEI FHMSLVLTNG YSVVKQNGEI LKKGYISHVI ALIDKLGILV AEYTSKTDPI YNPNLADEK EKLLKVWNFV FNLVLELRNS LCWLAVERNAD WDEAARDWGK	QTIDDSYNQA LDKKREMGAL NEAESWKTTY AADGTSTAVA YKNCLLQREI FKNLDKQILG KLMNAFCNGE KSEKNEPKVK LVIWKYEFY AVIKNYADAG FIKYYNEFRN EGRLYLGIMH LTSKKAMEKF YSNWSFYDFQ GKMYLFEVVN FQRQPSVNER KGEIKQVQFN IEQASLNEIN HKICQLIEKY FKDNRDLRAP TGFRKNVYIS YKNSTVSKEW NPKATDIKQE FSLQIKIKAG ANGAYNIARK YAGTTALNLD	KFYFDSLHOK RKRDKNAVGI QEREVDGKKI TRIADNFIIF EHIKNENSYN EVEKEKQLIE FESEYEGIYL KFISIAEIKN LFRDIERENG LGIFQMMKYF FITKKPFDED KNHRKLFQSM KPSQEILRIK FQDTGKYQNI KDIYNTKNKS EKITTQKNQC KIINQRISSS GVNYRDKLIE SAIVVLEDLN GGVLNGYQLS NSASLDKIKE AIFASAPRI IIEKKEAGDL EVIADVDEGVD GVMILKKIRE H	FIDAALASDK DRLQKEINDA TFSKADLKQK NPGQKRYIFD HQNTKSVFRDK KIIGRINKKI KTREKTEEDV KNKAINITSR AVEELDGDIF EKLGLYDSC LDDKDRKNT KIKLNFENGA GNAKGDNANR KEKTFKRESK KEFTDDVQKY KNLHTLYFEH ILDKGDRAYK DNEMRVNVIG REKERLKNRQ MRFKQIRGGI APFVSFEKMR AVKKFDAIGW RQKGEDGYWK QGEKELDGRL FIASPVKPPF HAKKDPEFKK	(SEQ ID NO: 121)
WP_004356401_(modified)	MKVMENYQEF VRADNVSYVK DCKSDEEEVK	TNLFQLNKT KEIDKKHKIF KTALRNKCTS	RFELKPIGKT IEETLSSFSI IQRAMREAI	CELLEEGKIF SNDLLKQYFD QAFLKSPQKK	ASGSFLEKDK CYNELKAFKK LLAIKNLIEN	(SEQ ID NO: 122)

Table 5 - Cas12a orthologs						
hypothetical protein [Prevotella disiens]	VFKADENVQH	FSEFTSYFSG	FETNRENFYVS	DEEKSTSIAY	RLVHDNLPIF	
	IKNIYIFEKL	KEQFDAKTLS	EIFENYKLYV	AGSSLDEVFS	LEYFNNTLTQ	
	KGIDNYNAVI	GKIVKEDKQE	IQGLNEHINL	YNQKHKDRRL	PPFISLKKQI	
	LSDREALSWL	PDMFKNDSEV	IKALKGFYIE	DGFENNVLTP	LATLLSSLDK	
	YNLNGIFIRN	NEALSLSQN	VYRNFSIDEA	IDANAELQTF	NNYELIANAL	
	RAKIKKETKQ	GRKSFEKYEE	YIDKKVKAID	SLSIQEINEL	VENYVSEFNS	
	NSGNMPRKVE	DYFSLMRKGD	FGSNDLIENI	KTCLSAAEKL	LGTKYQETAK	
	DIFKKDENS	LIKELLDATK	QFQHFQKPLL	GTGEEADRDL	VFYGDFLPLY	
	EKFEELTLLY	NKVRNRLTQK	PYSKDKIRLC	FNKPKLMTGW	VDSKTEKSDN	
	GTQYGGYLF	KKNEIGEYDY	FLGISSKAQL	FRKNEAVIDG	YERLDYYQPK	
	ANTIYGSAYE	GENSYKEDDK	RLNKVIIAYI	EQIKQTNIKK	SIIESISKYP	
	NISDDDKVTP	SSLLEKIKKV	SIDSYNGILS	FKSFQSVNKE	VIDNLLKTIS	
	PLKNKAFLD	LINKDYQIFT	EVQAVIDEIC	KQKTFIYFPI	SNVELEKEMG	
	DKDKPLCLFQ	ISNKDLSFAK	TFSANLRKKR	GAENLHTMLF	KALMEGQDN	
	LDLGSGAIFY	RAKSLDGNKP	THPANEAIKC	RNVANKDKVS	LFTYDIYKNR	
	RYMENKFLFH	LSIVQNYKAA	NDSACLNSSA	TEYIRKADDL	HIIGIDRGER	
	NLLYYSVIDM	KGNIVQDLS	NIIRNNDLET	DYHDLDDKRE	KERKANQNW	
	EAVEGIKDLK	KGYSQAVHQ	IAQLMLKYNA	IIALEDLGQM	FVTRGQKIEK	
	AVYQQFEKSL	VDKLSYLVDK	KRPYNELGGI	LKAYQLASSI	TKNNSDKQNG	
	FLFYVPAWNT	SKIDPVTGFT	DLRPPKAMTI	KEAQDFFGAF	DNISYNDKGY	
	FEFETNYDKF	KIRMKSAQTR	WTICTFGNRI	KRKKDKNYWN	YEEVELTEEF	
	KKLFKDSNID	YENCNLKEEI	QNKDNRKFFD	DLIKLLQLTL	QMRNSDDKGN	
	DYIISPANA	EGQFFDSRNG	DKKLPLDADA	NGAYNIARKG	LWNIRQIKQT	
	KNDKKLNLIS	SSTEWLDFVR	EKPYLK			
CCB70584 (modified) Protein of unknown function [Flavobacterium branchiophilum FL-15]	MTNKFTNQYS	LSKTLRFELI	PQKGTLEFIQ	EKGLLSQDKQ	RAESYQEMKK	(SEQ ID NO: 123)
	TIDKFHKYFI	DLALSNAKLT	HLETYLELYN	KSAETKKEQK	FKDDLKKVQD	
	NLRKEIVKSF	SDGDAKSIFA	ILDKKELITV	ELEKWFENNE	QKDIYFDEKF	
	KTFTTYFTGF	HQNRKNMYSV	EPNSTAIAYR	LIHENLPKFL	ENAKAFKIK	
	QVESLQVNFR	ELMGEFGDEG	LIFVNELEEM	FQINYNDVL	SQNGITTYNS	
	IISGFTKNDI	KYKGLNEYIN	NYNQTKDKKD	RLPKLKQLYK	QILSDRISLS	
	FLPDAFTDGK	QVLKAFIDFY	KINLLSYTIE	QOEESQNLLL	LIRQTIENTLS	
	SFDTQKIYK	NDTHLTITISQ	QVFGDFSVFS	TALNYWYETK	VNPKFETEYS	
	KANEKKREIL	DKAKAVFTKQ	DYFSIAFLQE	VLSEYILTLD	HTSDIVKKHS	
	SNCIADYFKN	HFVAKKENET	DKTFDFIANI	TAKYQCIQGI	LENADQYED	
	LKQDQKLIDN	LKFFLDAILE	LLHFQKPLHL	KSESITEKDT	AFYDVFENY	
	EALSLLTPLY	NMVRNYVTQK	PYSTEKIKLN	FENAQLLNGW	DANKEGDYLT	
	TILKKDGNFY	LAIMDKKHNK	AFQKFPEGKE	NYEKMVYKLL	PGVNMKLPKV	
	FFSNKNIAFY	NPSKELLENY	KKETHKKGDT	FNLEHCHTLI	DFFKDSLNNK	
	EDWKYFDFQF	SETKSYQDLS	GFYREVEHQG	YKINFKNIDS	EYIDGLVNEG	
	KLFLFQIYSK	DFSPFSKGKP	NMHTLYWKAL	FEEQNQLQNI	YKLNQQAEIF	
	FRKASIKPKN	IILHKKIKKI	AKKHFIKKKT	KTSEIVPVQT	IKNLNMYQQG	
	KISEKELTQD	DLRYIDNFSI	FNEKNKTIDI	IKDKRFTVVK	FQFHVPIITMN	
	FKATGGSYIN	QTVLEYLQNN	PEVKIIGLDR	GERHLVYLTL	IDQQGNILKQ	
	ESLNTITDSK	ISTPYHKLDD	NKENERDLAR	KNWGTVENIK	ELKEGYISQV	
	VHKIATLMLE	ENAIIVMEDL	NFGFKRGRFK	VEKQIYQKLE	KMLIDKLNYL	
	VLKDKQPQEL	GGLYNALQLT	NKFESFQKMG	KQSGFLFYVP	AWNTSKIDPT	
	TGFVNYFYTK	YENVDKAKAF	FEKFEAIRFN	AEKKYFEFEV	KKYSDFNPKA	
	BGTQQAWTIC	TYGERIETKR	QKQNNKFVS	TPINLTEKIE	DFLGKNQIVY	
	GDGNCIKSQI	ASKDDKAFFE	TLLYWFKMTL	QMRNSETRTD	IDYLLISPMN	
	DNGTFYNSRD	YEKLENPTLP	KDADANGAYH	IACKGLMLLN	KIDQADLTCK	
	VDLISISNRDW	LQFVQKNK				
WP_005398606 (modified) hypothetical protein [Helcococcus kunzii]	MFEKLSNIVS	ISKTIKFKLI	PVGKTLENIE	KLKLEKDFE	RSDFYPILKN	(SEQ ID NO: 124)
	ISDDYYRQYI	KEKLSDLNLD	WQKLYDAHEL	LDSSKKESQK	NLEMIQAQYR	
	KVLFNLSGE	LDKSGEKNSK	DLIKNNKALY	GKLFKKQFIL	EVLPDFVNNN	
	DSYSEEDLEG	LNLYSKFTTR	LKNFWETRKN	VFTDKDIVTA	IPFRAVNNF	
	GFYYDNKIF	NKNIEYLENK	IPNLENEKE	ADILDDNRSV	KDYFTPNGFN	
	YVITQDGIDV	YQAIRGGFTK	ENGEKVQGIN	EILNLTQQQL	RRKPETKNVK	
	LGVLTKLRKQ	ILEYSESTSF	LIDQIEDDND	LVDRIKFNV	SFFESTEVSP	
	SLFEQIERLY	NALKSIKKEE	VYIDARNTQK	FSQMLFGQWD	VIRRGYTVKI	

Table 5 - Cas12a orthologs						
	TEGSKEEKKK	YKEYLELDET	SKAKRYLNIR	EIEELVNLVE	GFEEVDVFSV	
	LLEKFKMNNI	ERSEFEAPIY	GSPIKLEAIK	EYLEKHLEEY	HKWKLLLIGN	
	DDLDTDETFY	PLLNEVISDY	YIIPLYNLTR	NYLTRKHSKD	DKIKVNFDFP	
	TLADGWSESK	ISDNRSIILR	KGGYYLGLIL	IDNKLLINKK	NKSKKIYEIL	
	IYNQIPEFSK	SIPNYPFTKK	VKEHFKNNVS	DFQLIDGYVS	PLIITKEIYD	
	IKKEKKYKGD	FYKDNNTNKN	YLYTIYKWIE	FCKQFLYKYK	GNPKESYKEM	
	YDFSTLKDTS	LYVNLNDFYA	DVNSCAYRVL	FNKIDENTID	NAVEDGKLLL	
	FQIYNKDFSP	ESKGKKNLHT	LYWLSMFSEE	NLRTRKLKLN	GQAEIFYRKK	
	LEKKPIIHKE	GSILLNKIDK	EGNTIPENIY	HECYRYLNKK	IGREDLSDEA	
	IALFNKDVLK	YKEARFDIIK	DRRYSESQFF	FHVPITFNWD	IKTNKNVQNI	
	VQGMIKDGEI	KHIIGIDRGE	RHLLYYSVID	LEGNIVEQGS	LNTLEQNRFD	
	NSTVKVDYQN	KLRTREEDRD	RARKNWTNIN	KIKELKDGYL	SHVVHKLRL	
	IIKYEIVIM	ENLNQGFKRG	RFKVERQVYQ	KFELALMNKL	SALSFKKEYD	
	ERKNLEPSGI	LNPIQACYPV	DAYQELQGGN	GIVFYLPAAV	TSVIDPVTGF	
	TNLFRLKSIN	SSKYEEFIKK	FKNIYFDNEE	EDFKFIFNYK	DFAKANLVIL	
	NNIKSKDWKI	STRGERISYN	SKKKEYFYVQ	PTEFLINKLK	ELNDNENID	
	IIPLIDNLEE	KAKRKILKAL	FDTFKYSVQL	RNYDFENDYI	ISPTADDNGN	
	YYNSNEIDID	KTNLPNNGDA	NGAFNIARKG	LLLKDRIVNS	NESKVDLKIK	
	NEDWINFIIS					
WP_021736722_(modified) CRISPR-associated protein Cpf1, subtype PREFRAN [Acidaminococcus sp. BV3L6]	MTQFEGFTNL	YQVSKTLRFE	LIPQGKTLKH	IQEQGFIEED	KARNHDHYKEL	(SEQ ID NO: 125)
	KPIIDRIYKT	YADQCLQLVQ	LDWENLSAAI	DSYRKEKTEE	TRNALIEEQA	
	TYRNAIHDFY	IGRTDNLDTA	INKRHAIEYK	GLFKAELFNG	KVLKQLGTVT	
	TTEHENALLR	SFDKFTTYFS	GFYENRKNVF	SAEDISTAIP	HRIVQDNFPK	
	FKENCHIFTR	LITAVPSLRE	HFENVKKAIG	IFVSTSIEEV	FSFPFYNQLL	
	TQTQIDLYNQ	LLGGISREAG	TEKIKGLNEV	LNLAIQKNDE	TAHIIASLPH	
	RFIPLFKQIL	SDRNTLSFIL	EEFKSDEEVI	QSFCYKKTLL	RNENVLETAE	
	ALFNELNSID	LTHIFISHKK	LETISSALCD	HWDTLRNALY	ERRISELTGK	
	ITKSAKEKVQ	RSLKHEDINL	QEIISAAGKE	LSEAFKQKTS	EILSHAHAA	
	DQPLPTTLKK	QEEKELKLSQ	LDSLLGLYHL	LDWFAVDESN	EVDPEFSARL	
	TGIKLEMEPS	LSFYNKARNY	ATKKPYSVEK	FKLNFQMPTL	ASGWDVNKEK	
	NNGAILFVKN	GLYYLGIMPK	QKGRYKALS	EPTEKTSEGF	DKMYDYFPD	
	AAKMIPKCST	QLKAVTAHFQ	THTPILLSN	NFIEPLEITK	EIYDLNNEPK	
	EPKKFQTAYA	KKTGDQKGYR	EALCKWIDFT	RDFLSKYTKT	TSIDLSSLRP	
	SSQYKDLGEY	YAEINPLLYH	ISFQRIAEKE	IMDAVETGKL	YLFQIYNKDF	
	AKGHHGKPNL	HTLYWTGLFS	PENLAKTSIK	LNGQAEFYR	PKSRMKRMAH	
	RLGEKMLNKK	LKDQKTPIPD	TLYQELYDYV	NHRLSHDLSD	EARALLPNVI	
	TKEVSHELIK	DRRFTSDKFF	FHVPITLNYQ	AANSPSKFNQ	RVNAYLKEHP	
	ETPIIGIDRG	ERNLIYITVI	DSTGKILEQR	SLNTIQQFDY	QKKLDNREKE	
	RVAARQAWSV	VGTIKDLKQG	YLSQVIHEIV	DLMIHYQAVV	VLENLNFQFK	
	SKRTGIAEKA	VYQQFEKMLI	DKLNCLVLKD	YPAEKVGGVL	NPYQLTDQFT	
	SFAKMGTSQS	FLFYVPAPYT	SKIDPLTGTV	DPFVWKTIKN	HESRKHFLG	
	FDLHLYDVKT	GDFILHFKMN	RNLFSQGRGLP	GFMPAWDIVF	EKNETQFQDAK	
	GTPFIAGKRI	VPVIENHRFT	GRYRDLYPAN	ELIALLEEKG	IVFRDGSNIL	
	PKLLENDSDH	AIDTMVALIR	SVLQMRNSNA	ATGEDYINSP	VRDLNGVCDF	
	SRFQNPWEPM	DADANGAYHI	ALKGQLLLNH	LKESKDLKLQ	NGISNQDWLA	
	YIQELRN					
WP_004339290_(modified) hypothetical protein [Francisella tularensis]	MSIYQEFVKN	YLSKTLRFE	LIPQGKTLEN	IKARGLILDD	EKRADYKKA	(SEQ ID NO: 126)
	KQIIDKYHQF	FIEEILSSVC	ISEDLLQNY	DVYFKLKKSD	DDNLQKDFKS	
	AKDTIKKQIS	KYINDSEKFK	NLFNQNLIDA	KKGQESDLIL	WLKQSKDNGI	
	ELFKANSDIT	DIDEALEIHK	SFKGWTTFYK	GFHENRKNVY	SSNDIPTSII	
	YRIVDDNLPK	FLENKAKYES	LKDCAPEAIN	YEQIKDLAE	ELTFDIDYKT	
	SEVNQRVFSL	DEVFEIANFN	NYLNQSGITK	FNTIIGGKFV	NGENTKRKGI	
	NEYINLYSQQ	INDKTLKKYK	MSVLFKQILS	DTESKSFVID	KLEDDSDVVT	
	TMQSFYEQIA	AFKTVEEKSI	KETLSLLFDD	LKAQKLDLSK	IYFKNDKSLT	
	DLSQQVFDDY	SVIGTAVLEY	ITQQVAPKNL	DNPSKKEQDL	IAKKTAKAY	
	LSLETIKLAL	EEFNKHRDID	KQCRFEEILS	NFAAIPMIFD	EIAQNKNLA	
	QISIKYQNGQ	KKDLLQASAE	EDVKAIKDLL	DQTNLLHRL	KIFHISQSED	
	KANILDKDEH	FYLVFEECYF	ELANIVPLYN	KIRNYITQKP	YSDEKFKLNF	
	ENSTLASGWD	KNKESANTAI	LFIKDDKYYL	GIMDKKHNNI	FSDKAIEENK	

Table 5 - Cas12a orthologs						
	GEGYKKIVYK QIADASKDIQ NLMIIDGKTV CKKGRKDRNG VNRQLLSLKR KHLPENIYRI KETKSYLKNE ARFSRKDLVD FIDYYKDRLD YYDFEFELKP SNEYSDFNDF TNHIGSQGYK LTFENISQDY INSLVNEGKL YLFQIYSKDF SAYSKGRPNL HTLYWKALFD ERNLQDVVYK LNGEAEFYR QQSIPKKITH PAKETIANKN KDNPKKESVF EYDLIKDKRF TEDKFFFHCP ITINFKSSGA NKFNDINLL LKEKANDVHI LSIDRGERHL AYYTLVDGKG NIIKQDNFNI IGNDRMKTNV HDKLAIEKD RDSARKDWKK INNIKEMKEG YLSQVVHEIA KLVIEYNAIV VFEDLNFQFK RGRFKVEKQV YQKLEKMLIE KLNLYLVFKDN EFDKTGGVLR AYQLTAPFET FKKMGKQTGI IYYVPAGFTS KICPVTGFVN QLYPKYESVS KSQEFFSKFD KICYNLDKGY FEFSFDYKNF GDKAAGKWT IASFGSRLIN FRNSDKNNHW DTREVYPTKE LEKLLKDYSI EYGHGECIKA AICGESDKKF FAKLTSVLNT ILQMRNSKTG TELDYLLISPV ADVNGNFFDS RQAPKNMPQD ADANGAYHIG LKGLMLLDRI KNNQEGKKLN LVIKNEEYFE FVQNRNN					
WP_022501477 type V CRISPR- associated protein Cpf1 [Eubacterium sp. CAG:76]	MNKAADNYTG GNYDEFIALS KVQKTLRNEL KPTPPTAEHI KQGGIISEDE YRAQQSLELK KIADEYYRNY ITHKLNIDIN LDFYNLFDAL BEKYKNDKD NRDKLDLVEK SKRGEIAKML SADDNFKSMF EAKLITKLLP DYVERNITGE DKEKALETLA LFKGFTTYFK GYFKTRKNMF SGEGGASSIC HRIVNVNASI FYDNLKTFMR IQEKAGDEIA LIEBELTEKL DGWRLEHIFS RDYYNEVLAQ KGIDYNNQIC GDINKHMNLY CQQNKFKANI FKMMKIQKQI MGISEKAFEI PPMYQNDDEEV YASFNEFISR LEEVKLTDR L INILQNINII NTAKIYNAR YVTNVSSYVY GGWGVDSAI ERYLYNTIAG KGQSKVKKIE NAKKDNKFMS VKELDSIVAE YEPDYFNAPY IDDDDNVAKA FGGQGVLYGF NKMSELLADV SLYTIDYNSD DSLIENKESA LRIKKQLDDI MSLYHWLQTF IDEVVEKDN AFYAELEDIC CELENNVTLY DRIRNYVTCK PYSTQKFKLN FASPTLAAGW SRSKEPDNNA IILLRNKYY IAFNVNNKP DKQIKKGSEE QRLSTDYKMK VYNLLPGPNK MLPKVFISD TGRDYNPSS YILEGYEKNR HIKSSGNFDI NYCHDLIDY KACINKHPEW KNYGFKFKET NQYNDIGQFY KDVEKQGYSI SWAYISEEDI NKLEDEGKIY LFEIYNKDL AHSTGRDNLH TMYLKNIFSE DNLKNICIEL NGEAEFYRK SSMKSNITHK KDTILVNKTY INETGVRVSL SDEDYMKVYN YNNNYVIDT ENDKNLIDII EKIGHRKSII DIVDKRYTE DKYFLYLPI ITINYGIEDENV NSKIEYIAK QDNMNVIGID RGERNLIYIS VIDNKGNIIE QKSFNLVNNY DYKNKLKNME KTRDNARKNW QEIGKIKDVK SGYLSGVISK IARMVIDYNA IIVMEDLNKG FKRGRFKVER QVYQKFENML ISKLNLYLVFK ERKADENGGI LRGYQLTYIP KSIKNVGKQC GCIFYVPAAY TSKIDPATGF INIFDFKKYS GSGINAKVKD KKEFLMSMNS IRYINECSEE YEKIGHRELF AFSFDYNNFK TYNVSSPVNE WTAYTYGERI KKLYKDRWL RSEVLNLTEN LIKLMEQYNI EYKDGHDIRE DISHMDETRN ADFICSLFEE LKYTVQLRNS KSEAEDENYD RLVSPILNSS NGFYDSSDY ENENNTHTM PKDADANGAY CIALKGLYEI NKIKQNWSDD KKFKENELYI NVTEWLDYIQ NRRFE	(SEQ ID NO: 127)				
WP_014550095 type V CRISPR- associated protein Cpf1 [Francisella tularensis]	MSIQEFVVK YLSKTLRFE LIPQGTLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF FIEELSSVC ISEDLQNYSD DVYFKLKKSD DDNLQKDFKS AKDTIKKQIS EYIKDSEKFK NLFNQNLIDA KKGQESDLIL WLKQSKDNGI ELFKANSIT DIDEALEIIE SFGWTTTYFK GFHENRKNVY SSNDIPTSI YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE ELTFDIDYKT SEVNQRVFSL DEVFEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRGKI NEYINLYSQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA AFKTVEEKSI KETLSLLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY ITQQVAPKNL DNPSKKEQDL IAKKTEKAKY LSLETIKLAL EEFNKHRDID KQCRFEEILA NFAAIPMIFD EIAQNKNLNA QISIKYQNG KKDLLQASAE DDVKAIKDLL DQTNLLHRL KIFHISQSED KANILDKDEH FYLVFEECYF ELANIVPLYN KIRNYITQKP YSDEKFKLNF ENSTLANGWD KNKEPDNTAI LFIKDDKYYL GVMNKKNNKI FDDKAIKENK GEGYKKIVYK LLPGANKMLP KVFFSAKSIK FYNPSEDILR IRNHSTHTKN GNPQKGYEKF EFNIEDCRKF IDFYKESISK HPEWKDFGFR FSQTQRYNSI DEFYREVENQ GYKLTENIS ESYIDSVVNQ GKLYLFQIYN KDFSAYSKGR PNHTLYWKA LFDERNLQDV VYKLNGEAEL FYRKKSIPKK ITHPAKAI NKNKDNPKKE SFFEYDLIKD KRFTEDKFFF HCPITINFKS SGANKFNDEI	(SEQ ID NO: 128)				

Table 5 - Cas12a orthologs						
	NLLLKEKAND TNYHDKLAAI AIVVFEDLNF VLRAYQLTAP SVSKSQEFFS LINFRNSDKN KKFFAKLTSI PQDADANGAY	VHILSIDRGE EKDRDSARKD GFKRGRFKVE FETFKKMGKQ KFDKICYNLD HNWDTREVYP LNTILQMRNS HIGLKGLMLL	RHLAYYTLVD WKKINNIKEM KQVYQKLEKM TGIIYYVPAG KGYFEFSFDY TKELEKLLKD KTGTELDYL I DRIKNNQEGK	GKGNI IKQDT KEGYLSQVVH LIEKLNLYLVF FTSKICPVTG KNFGDKAAKG YSIEYGHGEC SPVADVNGNF KLNLVIKNEE	FNIIGNDRMK EIAKLVI EHN KDNEFDKTGG FVNQLYPKYE KWTIASFGSR IKAAICGESD FDSRQAPKNM YFEFVQNRNN	
WP_003034647 type V CRISPR- associated protein Cpf1 [Francisella tularensis]	MSIYQEFVNK KQIIDKYHQF AKDTIKKQIS ELFKANS DIT YRIVDDNLPK SEVNQRV FSL NEYINLYSQQ TMQSFYEQIA DLSQQVFDDY LSLETIKLAL QISLKYQNGG KANILDKDEH ENSTLANGWD GEGYKKIVYK GNPQKGYEKF DEFYREVENQ PNLHTLYWKA NKNKDNPKKE NLLLKEKAND TNYHDKLAAI AIVVFEDLNF VLRAYQLTAP SVSKSQEFFS LINFRNSDKN KKFFAKLTSV PQDADANGAY	YSLSKTLRFE FIEEILSSVC EYIKDSEKFK DIDEALEI IK FLENKAKYES DEVFEIANFN INDKTLKKYK AFKTVEEKSI SVIGTAVLEY EEFNKHRDID KKDLLQASAE FYLVFEECYF KNKEPDNTAI LLPGANKMLP EFNIEDCRKF GYKLT FENIS LFDERNLQDV SVFEYDLIKD VHILSIDRGE EKDRDSARKD GFKRGRFKVE FETFKKMGKQ KFDKICYNLD HNWDTREVYP LNTILQMRNS HIGLKGLMLL	LIPQGKTLEN ISEDLLQNY S NLFNQNLIDA SFKGWTTYFK LKD KAPEAIN NYLNQSGITK MSVLFKQILS KETLSLLFDD ITQQVAPKNL KQCRFEEILA EDVKA IKDLL ELANIVPLYN LFIKDDKY YL KVFFSAKSIK IDFYKESISK ESYIDSVVNQ VYKLNGEAE L KRFTEDKFFF RHLAYYTLVD WKKINNIKEM KQVYQKLEKM TGIIYYVPAG KGYFEFSFDY TKELEKLLKD KTGTELDYL I DRIKNNQEGK	IKARGLILDD DVYFKLKKSD KKGQESDLIL GFHENRKNVY YEQIKKDLAE FNTIIGGK FV DTESKSFVID LKAQKLDLSK DNPSKKEQDL NFAAIPMIFD DQTNNLLHRL KIRNYITQKP GVMNKKNNKI FYNPSEDILR HPEWKDFGFR GKLYLFQIYN FYRKQSIPKK HCPITINFKS GKGNI IKQDT KEGYLSQVVH LIEKLNLYLVF FTSKICPVTG KNFGDKAAKG YSIEYGHGEC SPVADVNGNF KLNLVIKNEE	EKRADYKKA DDNLQKDFKS WLKQSKDNGI SSDDIPTSI I ELTFDIDYKT NGENTKRKGI KLEDDSDVVT IYFKNDKSLT IAKKT E KAKY EIAQNKDNL A KIFHISQSED YSDEKFKLNF FDDKA I KENK IRNHSTHTKN FSDTQRYNSI KDFSAYS SKGR ITHPAKEAIA SGANKFNDEI FNIIGNDRMK EIAKLVI EHN KDNEFDKTGG FVNQLYPKYE KWTIASFGSR IKAAICGESD FDSRQAPKNM YFEFVQNRNN	(SEQ ID NO: 129)
WP_003040289.1 type V CRISPR- associated protein Cpf1 [Francisella tularensis subsp. novicida U112]	MSIYQEFVNK KQIIDKYHQF AKDTIKKQIS ELFKANS DIT YRIVDDNLPK SEVNQRV FSL NEYINLYSQQ TMQSFYEQIA DLSQQVFDDY LSLETIKLAL QISIKYQNGG KANILDKDEH ENSTLANGWD GEGYKKIVYK GSPQKGYEKF DEFYREVENQ PNLHTLYWKA NKNKDNPKKE NLLLKEKAND TNYHDKLAAI AIVVFEDLNF VLRAYQLTAP SVSKSQEFFS LINFRNSDKN KKFFAKLTSV	YSLSKTLRFE FIEEILSSVC EYIKDSEKFK DIDEALEI IK FLENKAKYES DEVFEIANFN INDKTLKKYK AFKTVEEKSI SVIGTAVLEY EEFNKHRDID KKDLLQASAE FYLVFEECYF KNKEPDNTAI LLPGANKMLP EFNIEDCRKF GYKLT FENIS LFDERNLQDV SVFEYDLIKD VHILSIDRGE EKDRDSARKD GFKRGRFKVE FETFKKMGKQ KFDKICYNLD HNWDTREVYP LNTILQMRNS	LIPQGKTLEN ISEDLLQNY S NLFNQNLIDA SFKGWTTYFK LKD KAPEAIN NYLNQSGITK MSVLFKQILS KETLSLLFDD ITQQIAPKNL KQCRFEEILA DDVKA IKDLL ELANIVPLYN LFIKDDKY YL KVFFSAKSIK IDFYKQSISK ESYIDSVVNQ VYKLNGEAE L KRFTEDKFFF RHLAYYTLVD WKKINNIKEM KQVYQKLEKM TGIIYYVPAG KGYFEFSFDY TKELEKLLKD KTGTELDYL I SPVADVNGNF	IKARGLILDD DVYFKLKKSD KKGQESDLIL GFHENRKNVY YEQIKKDLAE FNTIIGGK FV DTESKSFVID LKAQKLDLSK DNPSKKEQEL NFAAIPMIFD DQTNNLLHKL KIRNYITQKP GVMNKKNNKI FYNPSEDILR HPEWKDFGFR GKLYLFQIYN FYRKQSIPKK HCPITINFKS GKGNI IKQDT KEGYLSQVVH LIEKLNLYLVF FTSKICPVTG KNFGDKAAKG YSIEYGHGEC SPVADVNGNF	EKRADYKKA DDNLQKDFKS WLKQSKDNGI SSNDIPTSI I ELTFDIDYKT NGENTKRKGI KLEDDSDVVT IYFKNDKSLT IAKKT E KAKY EIAQNKDNL A KIFHISQSED YSDEKFKLNF FDDKA I KENK IRNHSTHTKN FSDTQRYNSI KDFSAYS SKGR ITHPAKEAIA SGANKFNDEI FNIIGNDRMK EIAKLVI EYN KDNEFDKTGG FVNQLYPKYE KWTIASFGSR IKAAICGESD FDSRQAPKNM	(SEQ ID NO: 130)

Table 5 - Cas12a orthologs						
	PQDADANGAY	HIGLKGMLLL	GRIKNNQEGK	KLNLVIKNEE	YFEFVQNRNN	
KKQ38174 hypothetical protein US54_C0016G0015 [Candidatus Roizmanbacteria bacterium GW2011_GWA2_3 7_7]	MKSFDSTNL	YLSKTLKFE	MRPVGNTQKM	LDNAGVFEKD	KLIQKKYGKT	(SEQ ID NO: 131)
	KPYFDRHLRE	FIEEALTGVE	LIGLDENFRT	LVDWQKDCKN	NVAMKAYENS	
	LQRLRTEIGK	IFNLKAEDWV	KNKYPILGLK	NKNTDILFEE	AVFGILKARY	
	GEEKDTFIEV	EEIDKTGKSK	INQISIFDSW	KGFTGYFKKF	FETRKNFYKN	
	DGTSTAIATR	IIDQNLKRFI	DNLSIVESVR	QKVDLAETEK	SFSISLSQFF	
	SIDFYNKCLL	QDGIDYYNKI	IGGETLKNGE	KLIGLNELIN	QYRQNNKDQK	
	IPFFKLLDKQ	ILSEKILFLD	EIKNDTELIE	ALSQFAKTAE	EKTKIVKKLF	
	ADFVENNSKY	DLAQIYISQE	AFNTISNKWT	SETETFAKYL	FEAMKSGKLA	
	KYEKKDNSYK	FPDFIALSQM	KSALLSISLE	GHFWKEKYK	ISKFQEKTNW	
	BQFLAIFLYE	FNSLFSDKIN	TKDGETKQVG	YYLFAKDLHN	LILSEQIDIP	
	KDSKVTIKDF	ADSVLTIYQM	AKYFAVEKKR	AWLAEYELDS	FYTQPDGTGYL	
	QFYDNAYEDI	VQVYNKLRNY	LTKKPYSEEK	WKLNFENSTL	ANGWDKNKES	
	DNSAVILQKG	GKYVLGLITK	GHNKIFDDR	QEKFIVGIEG	GKYEKIVYKF	
	FPDQAKMFPK	VCFSAKGLEF	FRPSEEILRI	YNNAEFKKGE	TYSIDSMQKL	
	IDFYKDCCLK	YEGWACYTFR	HLKPTEEYQN	NIGEFFRDVA	EDGYRIDFQG	
	ISDQYIHEKN	EKGELHLFEI	HNKDNWLDKA	RDGKSKTTQK	NLHTLYFESL	
	FSNDNVVQNF	PIKLNGQAEI	FYRPKTEKDK	LESKKDKKGN	KVIDHKRYSE	
	NKIFFHVPLT	LNRTKNDSYR	FNAQINNFLA	NNKIDINIIGV	DRGEKHLVYY	
	SVITQASDIL	ESGSLNELNG	VNYAEKLGKK	AENREQARRD	WQDVQGIKDL	
	KKGYISQVVR	KLADLAIKHN	AIIILEDLNM	RFKQVRGGIE	KSIYQQLEKA	
	LIDKLSFLVD	KGEKNPEQAG	HLLKAYQLSA	PFETFQKMGK	QTGIIFYTQA	
	SYTSKSDPVT	GWRPHLYLYK	FSAKKAKDDI	AKFTKIEFVN	DRFELTYDIK	
	DFQQAKEYPN	KTVWKVCSNV	ERFRWDKNLN	QNKGGYTHYT	NITENIQELF	
	TKYGIDITKD	LLTQISTIDE	KQNTSFFRDF	IFYFNLICQI	RNTDDSEIAK	
	KNGKDDFILS	PVEPFDFS RK	DNGNKLPENG	DDNGAYNIAR	KGIVILNKIS	
	QYSEKNENCE	KMKWGDLYVS	NIDWDNFVTQ	ANARH		
WP_022097749 type V CRISPR- associated protein Cpf1 [Eubacterium elgens CAG:72]	MNGNRSIVYR	EFVGVTPVAK	TLRNELRPVG	HTQEHIIQNG	LIQEDELROE	(SEQ ID NO: 132)
	KSTELKNIMD	DYYREYIDKS	LSGLTDLDFT	LLFELMNSVQ	SSLSKDNKKA	
	LEKEHNKMRE	QICHTLQSDS	DYKNMFNAKL	FKEILPDFIK	NYNQYDVKDK	
	AGKLETLALF	NGFSTYFTDF	FEKRKNVFTK	BAVSTSIAYR	IVHENSILFL	
	ANMTSYKKIS	EKALDEIEVI	EKNQDKMGD	WELNQIFNPD	FYNMVLQSG	
	IDFYNEICGV	VNAHMNLYCQ	QTKNNYNLFK	MRKLHKQILA	YTSTSFVEVPK	
	MFEDDMSVYN	AVNAFIDEET	KGNIIGKLDK	IVNKYDELDE	KRIYISGDFY	
	ETLSCFMSGN	WNLITGCVEN	FYDENIHAKG	KSKEEKVKKA	VKEDKYKSIN	
	DVNDLVEKYI	DEKERNEFKN	SNAKQYIREI	SNIITDTETA	HLEYDEHISL	
	IESEEKADEI	KKRLDMYMMN	YHWVKAFIVD	EVLDRDEMFI	SDIDDIYNIL	
	ENIVPLYNRV	RNYVTQKPYT	SKIKLNFQS	PTLANGWSQS	KEFDNNAIIL	
	IRDNKYLAIF	FNAKNKPKDK	IIQGNSDKKN	DNDYKMMVYN	LLPGANKMLP	
	KVFLSKKGIE	TFKPSDYIIS	GYNNAKHKT	SENFDISFCR	DLIDYFKNSI	
	EKHAERWKYE	FKFSATDSYN	DISEFYREVE	MQGYRIDWTY	ISEADINKLD	
	BEGKIYLFQI	YNKDFAENST	GKENLHTMYF	KNIFSEENLK	NIVIKLNGQA	
	ELFYRKASVK	NPVKHKKDSV	LVNKTYKNQL	DNGDVVRIPI	PDDIYNEIYK	
	MYNGYIKESD	LSEAAKEYLD	KVEVRTAQKD	IVKDYRYTVD	KYFIHTPITI	
	NYKVTARNNV	NDAVKYIAQ	NDDIHVIGID	RGERNLIYIS	VIDSHGNIVK	
	QKSYNILLNNY	DYKKKLVEKE	KTREYARKNW	KSIGNIKELK	EGYISGVVHE	
	IAMLMEVYNA	IIMEDLNYG	FKRGRFKVER	QVYQKFESML	INKLNYFASK	
	GKSVDEPGGL	LKGYQLTYVP	DNIKNLGKQC	GVIFYVPAAF	TSKIDPSTGF	
	ISAFNFKSIS	TNASRKQFFM	QFDEIRYCAE	KDMFSFGFDY	NNFDTYNITM	
	GKTQWTVYTN	GERLQSEFNN	ARRTGKTKSI	NLTETIKLLL	EDNEINYADG	
	HDVRIDMEKM	YEDKNSEFFA	QLLSLYKLTV	QMRNSYTEAE	EQEKGISYDK	
	IISPVINDEG	EFFDSNDYKE	SDDKECKMPK	DADANGAYCI	ALKGLYEVLLK	
	IKSEWTEGDF	DRNCLKLPHA	EWLDFIQNKR	YE		
WP_021739647 hypothetical protein [Eubacterium ramulus]	MIKKTIDTVL	NVRPIFVGIO	HLYFYEGPCR	FGEGDELMPE	YDAMMNQEMN	(SEQ ID NO: 133)
	AAYVNEVVQH	ETEGVHIMDP	IYVERDDWFR	SPEAMYEKMA	EDIDKVDYFL	
	FHFHIGRIGDI	YLEFAERYKK	PVGAAAPGLCC	DGIGNTAAVK	NRGLEAYAFM	
	SWDEFDTWMR	VLRVRKCLKN	TRVLLAVRWD	SNRSYSSYDN	FINQSDVTNK	
	WGIQFRHVN	HELLDQTHPV	DPTTNPSTPG	RKALNINDED	MKEIEKITDE	
	LIANAEEACTM	EPDMVKKTIQ	AYYTVQKLLD	AYDCNAFTAP	CPDLCSTRRF	

Table 5 - Cas12a orthologs						
	SEEKFTLCMT AIVFDKEGHM RYRLAPFAYS AGYEMKNCDQ LGIEAVIA	HSLNDENGIS IPFHKFNDNT GFGATIRYDF GVFFNVADKV	SACEYDINSV IEDIADKTNL AQDIGQVITM DFYHKQQYFG	IGKVIMTNLS YMTFHSTPNR IRISPDATKI NHTVLAYGDY	GKAPYMGNTN NLKGLKAEKE FIAKGTISGG VEELKMLAE	
gi 800943167 WP_045971446.1 type V CRISPR- associated protein Cpf1 [Flavobacterium sp. 316]	MKNFSNLYQV IDEFHKYFID LRKEIVSSFT FTTYFTGFHQ AELQPKIEKL RTATEGKQKI PDAFTEGQV DVNKIYLNK NEAQEKLDK CIADYFTTHF QDQKLIDNIK LNVVTPLYNM LKRDNVFLA SNKNIAFFNP WKYFDFQFSE YLFQIYNKDF KASIKKKNI VPITMNFKAT GTILLQESLN GYISQVVKHI DKLNYLVKLD SKIDPTTGFV DFNPKAADTR KVGLDLSKEL PVANEKGIFY LKKVKLAISN	SKTVRFELKP LALNNKKLSY EGEAKAIFSV NRKNMYSAEA YKEFEAYLNV QGLNEIINLY LKAVFEFYKI TSLTTIAQQ EIKIKFVKQDY IAKKENEADK FFLDAILEVV VRNVYVTKPY IMDKKHNTF SKEILDNYKN TKTYQDLSGF SPYAKGKPNM THKAHQPIAA GNSYINQDVL VIQDEKTHTP TKMMIEHNAI KQPHELGGLY NYFYTKYENV QEWITCTHGE KTQLIAQNEK DSRKATASLP REWLQYVQVQ	IGNTLENIKN LNEYIALYTQ LDKKELITIE NSTAIAYRLI NSISELFEID NQTKPKNERL NLLSYKQDGV FGDFSVPFSA FSIAFLQEVV TFNFIANIKA HFIKPLHLKS STEKIKLNFE QQFTEDDENY NTHKKGATFN YKEVEHQGYK HTLYWKALFE KNPLTPTAKN AYLKDNPEN YHTLLDNKEI VVMEDLNFGF NALQLTNKFE EKAFTFFSKF RIETKRQKEQ SFPEELFHL IDADANGAYH QKK	KSLKNDNSIR SAEAKKEDKF LEKWKNENNL HENLPKFIE YFNEVLTKQ PKLKQLYKQI EESQNLLELI LQYRYETVVN ADYVKTLDEN KYQCIQIGILE ESITEKDNF NAQLLNGWDA EKIVYKLLPG LKDCHALIDF INFKKVSVSQ TQNLNVIYK TFAYDLIKDK IIGLDRGERH ARDKARKNWG KRGRFKVEKQ SFQKMGKQSG DSILYNKTKG NNNFVSTTIQ KLTLQMRNSE IAKKGLWIME	AESYQKMKKT KADFKKVQDN AVYLDESFKS SKAFKSSQI ITVYNNIIGG LSDRISLSFL QQVVKNLGNQ PKYTAEYQKA LDWKQKYTPS QADDYEDELK YDVFNYYEBA NKEKDYLTTI VNMKLPKVV FKDSLKNKHE IDTLIEEGKM LNGQAEIFFR RYTVDFQFQ LVYLTLLIDQ SIESIKELKE IYQKLEKMLI FLFYVPAWNT YFEFVKNYS LTEQFVNFFE SHTIEDYLIS QINKTNSADD	(SEQ ID NO: 134)
WP_044110123.1 type V CRISPR- associated protein Cpf1 [Prevotella brevis]	MKQFTNLYQL IDDYHKDYIE ASALKGDDRY ENRKNMYSDE YKNFEEYLVN GINEFVNLYN LNAVKEYCTS GEWSIIPNAI CISELNGIDI REKNLSQDKL QLWEVIDSIT NASIILRKNG VPKCTTQMK VKKFQIGYLN IGCYNDLNSF STYSKGTPNM HPANKPIDNK NGDINMQVRE IVNNYKGETEY HKITQLMIKY DKNKDANEIG GFFVNLDDTRY TRTQWTLCTY NLKEAILLQA FFDSRKVKEG EWLQFAQKKQ	SKTLRFELKP NVLNNFKLNG KTIDKKELIR AKSTAIAYRL NSIDEIFSLD QQLANKKDRL YFDNVEESVK EQRLRSDNPK ADYYATLGAI TVMQLKDLDL PLYNKVRNYC MYYLGIMKKE VKEHFSNSND NTGDSFGYNH YGAVNLLLYN HTLYWKMLFD NVNNPKQSN YIKTTDDLHF KTDYHTLLSE NAIVLLEDLN GLLHAYQLTN ENVAKAQAFF GTRIFTRNA NKNFYTDILH LPVNADANGA YLNK	IGTLEHINA EYLQAYFDLY QDMKTFLKKD IHDNLPKFID YYNIVLTQTQ PKLKPLFKQI VLLTGISDYD KTNEKEEKYS NTESKQEPSI DFKALQHF TRKPFSTEKI YRNILTKPMP DYTLFEKDKF AVEIWSFCL LTYRKVSVDY ESNLNDVVYK FEYDLIKDKR IGIDRGERHL RDKKRKEERS MGFKRGRQKV DPKLPNKNSK KKFDSIRYNK EKNNSWDSRE LMKLTLQMRN YNIARKGLWL	NGFIDNDAHR SQDTKDKQFK TDKALLDEFY NIAVFKKIAN IEVYNSIIGG LSDRVQLSWL LSKIYITNDL DRISKLLKLP PTSIQVHYNA PLLNGDEAE KVNFEQAQLL SDGDCYDKV IAPVVITKEI KFLKAYKSTS IHQLVDEDKM LNGQAEVFYR YTVDKFMFHV LYICVINGKG SWQTIEGIKE ESSVYQQFEK QSGFLFYVPA EYDRFEFKFD IDLTTEWKT SVTGTDIDYM AQQIKNANDL	AESYKKVKKL DIQDKLRKSI EFTTYFTGYH TSVADNFSTI RTLEDDTKIQ QEEFNTGADV ALTDVSRQMF KSYSLGYINE LKPILDTPY KDEKFGELM DGWDENKES YKFFKDIITM FDLNNVLYNG IYDFSSIEKN YLFMIYNKDF KKSITYQHPT PITLNFKGMG EIVEQYSLNE LKSGYLSQVI ALIDKLNLYV WNTSKIDPVT YSNFTAKAED FTQHNIPLNA VSPVANECEG SDVKLAITNK	(SEQ ID NO: 135)
WP_036388671.1	MLFQDFTHLY AILDDYHRDF	PLSKTMRFEL IADMMGEVKL	KPIGKTLEHI TKLAEFYDVY	HAKNFLSQDE LKFRKNPKDD	TMADMYQKVK GLQKQLKDLQ	(SEQ ID NO: 136)

Table 5 - Cas12a orthologs						
type V CRISPR-associated protein Cpf1 [Moraxella caprae]	AVLRKEIVKP SPKLAHLAHF DNLQILATIK AYNTLLGGIS SDGMGVSFLLP GIYVEHKNLN AKAKLTKEKD LAGVDNPIQK KELLDNALNV VRDYLSQLKPF LLDKAHKKVF PSAELLDKYA PTSSYQDLSD FSPKAHGKPN TIHRAGEVLE GMTIKEFNKK LNDITTASAN SHVVHQISQL NHLVLKDEAD DPETGFVDLL KAKNSRQIWK NDKQPNLVM NDEGMFFNSA AIDNQTLWNF	IGNGGKYKAG EKFSTYFTGF QKHSALYDQI GEAGSRKIQQ SKFADDSEMC ELSKQAFGDF KFIKGVHSLA IHNNHSTIKG AHFAKLLTTK STEKYKLNFG DNAPNTGKNV QGTHKKGNF FYREVEPQGY LHTLYFKALF NKNPDNPCKR VNQSIQQYDE GTQMTTPYHK MLKYNAIVVL DEIGSYKNAL KPRYENIAQS ICSHGDKRYV ICQNDKEFH LADDTQPQNA AQNR	YDRLFGAKLF HDNRKNMYS INELTASGLD INELINSHHN QAVNEFYRHY ALLGRVLDGY SLEQAIEHYT FLERERPAGE TTLDNQDGNF NPTLLNGWDL YQKMIYKLLP NLKDCHALID QVVFVDINAD SKDNLANPIY QFVYDIKDK VNVIGIDRGE ILDKREIERL EDLNFGFKRG QLTNFTDLK QAFFGKFDKI YDKTANQNG KSLIYLLKTL DANGAYHIAL	KDGKELGDLA EDKHTAITYR VSLASHLDGY QHCHKSERIA ADVFAKVQSL YVDVVNPEFN ARHDDDESQA RALPKIKSGK YGEFGALYDE NKEKDNFGII GPNKMPLPKVF FFKAGINKHP YINELVEQQQ KLNGEAQIFY RYTQDKFMLH RHLLYLTVIN NARVGWGEIE RFKVEKQIYQ SIGKQTGFLF CYNADKDYFE ATKGINVNDE LALRYSNASS KGLWVLEQIK	KFVIAQEGES LIHENLPRFI HKLLTQEGIT KLRPLHKQIL FDGFDDHQKD ERFAKAKTDN GKLGQYFKHG NPMTQLRQL LAKIPTLYNK LQKDGCCYLA FAKSNLDYYN EWQHFQFKFS LYLFQIYNKD RKASLDMNET VPITMNFVQ SKGEILEQRS TIKELKSGYL NFENALIKKL YVPAWNTSKI FHIDYAKFTD LKSLFARHHI DEDFILSPVA NSDDLNVKVL	
WP_020988726.1 type V CRISPR-associated protein Cpf1 [Leptospira inada]	MEDYSGFVNI KKI IDKYHRA KNIPDKERLE KFCETDEERK LPKFLDNLKI VLNQKGIDAY ILFKQILGDR KKFLSSFNRY DPPKKIKSPL EAYFAEFKSK VGKIKDFLDS LYNKVRNYLT YYLGVMKDKN NLSIYNPSKS FDFKFSKTSS QIYNKDFSIF INYDEKKKRE LNFNLKVNEF QSGKGRPEIN SKLMVENNAI NKPTPEGGVL DFLHPAYENI RVWVICTTNV PEILRKNDAV FFNSLEASDD NKDWLEFVWE	YSIQKTLRFE YIEEVFDSVL ALSEKLRKML QVSNFKSFTT IESIQRRFKD NTILGGKSEE ETKSFIPFAF ELDGIYLAND KYEKEKEKWL DDAKQDFDL IKSLQFFLKP GKIYSKEKFK NTILSDIPKV ILKIREAKSF YENISEFYRE SKGKPNLHTI GHHPELFELK LKRNDINII YKEKLQEKEI VLEDLNIGF KAYQLTDEFQ EKAKQWINKF ERYFTSKTAN FFKSLLFYIK EPKDADANGA RNR	LKPVGKTLLEH HQKKKKDKTR VGAFKGEFSE YFTGFHSNRQ FPWSDLKKNL SGEKIQGLNE PDDQSVLNSI NSLASISTFL KQKYTTISFL ERIEEAYAIV LLSAEIFDEK LNFENSTLLK KPNELFYEKM KEGKNFKLKD VERQGYNLDF YFRSLFSKEN KYPILKDKRY GIDRGERNLL ERDKARKSWG KRGRQKVERQ SFEKLSKQTG DSIRFNSKMD SSIQYNSIQI TTLSLRQNG YHIALKGLMN	IEKKGFLKKD FSTQFIKEIK EVAEKYKNLF NIYSDEKKST KKIDKNIKLT YINLYRQKNV TEFAKYLKLD FDDWSFIKKS NDAIESYSKS EPLLGAEYPR DLGFYNQLEG GWDENREVAN VYKLIPTPHM CHKFIDFYKE KKVSKFYIDS LKDVCCLKNG SEDKFQFHLP YLVMINQKGE TVENIKELKE VYQKFEKMLI FLFYVPSWNT WFETADTRK TEKLKELFVD KKGEEKDFI LLVLNETKEE	KIRAEDYKAV BFSELYKTE SKELIRNEIE AIGYRIHQN EYFSIDGFVN IDRKNLPNVK KKKKSIAEL VSFKYDES QDEKRVKIRL DRNLKADKKE YYEIDSIGH LCVIFREDQK QLPRIIFSSD SISKNEWD LVEDGKLYLF BAEMFFRKKS ISLNFKSKER ILKQTLILDSM GYLSIVIHQI DKLNLFLVFE SKIDPRTGFI FSENLMGKN IPFSNGQDLK LSPVVD NLSRPKWIK	(SEQ ID NO: 137)
WP_023936172.1 type V CRISPR-associated protein Cpf1 [Porphyromonas crevioricanis]	MPWIDLKDF RVKKI IDTYH DKALDKIRAV FVLSTEAESL IAYRLIHENL EDIFSLNYYI GREDRLPLFR ILGRTOQLMT DHEQAPKRIT	NLYPVSKTLR KVFISSLEN LRGLIVGAFT PFSVEEATRS PKFIDNILVF HVLSQAGIEK PLYKQILSDR SISEYDLSRI AKYERDRIKA	FELKPVGKTL MAKMGIE GVCGRRENTV LKEFDSFTSY QKIKEPIAKE YNALIGKIVT EQLSYLPESF YVRNDSQLTD LKGEESISLA	ENIEKAGILK KAMLQSFCEL QNEKYESLFK FAGFYENRKN LEHIRADFS EGDGEMKGLN EKDEELLRAL ISKKMLGDWN NLNSCIAFLD	EDEHRAESYR YKKDHRTEGE EKLIKEILPD IYSTKPQSTA GGYIKKDERL EHINLYNQQR KEYFDHIAED AIYMARERAY NVRDCRVDY	(SEQ ID NO: 138)

Table 5 - Cas12a orthologs						
	LSTLGQKEGP	HGLSNLVENV	FASYHEAEQL	LSFPYPEENN	LIQDKDNVVL	
	IKNLLDNISD	LQRFLKPLWG	MGDEPDKDER	FYGEYNYIRG	ALDQVIPLYN	
	KVRNYLTRKP	YSTRKVKNLF	GNSQLLSGWD	RNKEKDNSCV	ILRKGQNFYL	
	AIMNNRHKRS	FENKVLPEYK	EGEPYFEKMD	YKFLPDPNKM	LPKVFLSKKG	
	IEIYEPSPKL	LEQYGHGTHK	KGDTFSMDDL	HELIDFFKHS	IEAHEDWKQF	
	GFKFSDTATY	ENVSSFYREV	EDQGYKLSFR	KVSESYVYSL	IDQGKLYLFQ	
	IYNKDFSPCS	KGTPNLHTLY	WRMLFDERNL	ADVITYKLDGK	AEIFFREKSL	
	KNDHPHTPAG	KPIKKKSRQK	KGEESLFEYD	LVKDRRYTMD	KFQFHVPIITM	
	NFKCSAGSKV	NDMVNAHIRE	AKDMHVGID	RGERNLLYIC	VIDSRGTILD	
	QISLNTINDI	DYHDLLESRD	KDRQQERRNW	QTIEGIKELK	QGYLSQAVHR	
	IAELMVAYKA	VVALEDLNMG	FKRGRQKVES	SVYQQFEKQL	IDKLNLYLVDK	
	KKRPEDIGGL	LRAYQFTAPF	KSFKEMGKQN	GFLFYIPAWN	TSNIDPTTGF	
	VNLFHAQYEN	VDKAKSFFQK	FDSISYNPKK	DWFEFAFDYK	NFTKKAEGSR	
	SMWILCTHGS	RIKNFRNSQK	NGQWDSEefa	LTEAFKSLFV	RYEIDYTADL	
	KTAIVDEKQK	DDFVDLLKLF	KLTVQMRNSW	KEKDLDYLIS	PVAGADGRFF	
	DTREGNKSLLP	KDADANGAYN	IALKGLWALR	QIRQTSEGGK	LKLAINSKEW	
	LQFVQERSYE	KD				
WP_009217842.1 type V CRISPR- associated protein Cpf1 [Bacteroidetes oral taxon 274]	MRKFNEFVGL	YPISKTLRFE	LKPIGKTLEH	TQRNKLLEHD	AVRADDYVKV	(SEQ ID NO: 139)
	KKIIDKYHKC	LIDEALSGFT	FDTEADGRSN	NSLSEYYLYY	NLKKRNEQEQ	
	KTFKTIQNNL	RKQIVNKLTD	SEYKRIDKK	ELITTDLPDF	LTNESEKELV	
	EKFKNFTTYF	TEFHKNRKNM	YSKEEKSTAI	AFRLINENLP	KFVDNIAAFE	
	KVSSPLAEK	INALYEDFKE	YLNVEEISRV	FRLDYDELL	TQKQIDLYNA	
	IVGGRTEEDN	KIQIKGLNQY	INEYNQQQTD	RSNRLPKLKP	LYKQILSDRE	
	SVSWLPPKFD	SDKNLLIKIK	ECYDALSEKE	KVFDKLESIL	KSLSTYDLSK	
	IYISNDSQLS	YISQKMFGRW	DIISKAIED	CAKRNPQKSR	ESLEKFAERI	
	DKKLKTIDSI	SIGDVDECLA	QLGETYVKRV	EDYFVAMGES	EIDDEQTDTT	
	SFKKNIEGAY	ESVKELLNNA	DNITDNNLMQ	DKGNVEKIKT	LLDAIKDLQR	
	FIKPLLKGGD	EADKDGVFYF	EFTSLWTKLD	QVTPLYNMVR	NYLTSKPYST	
	KKIKLNFENS	TLMDGWDLNK	EPDNTTVIFC	KDGLYYLGIM	GKKYNRVFVD	
	REDLPHDGEC	YDKMEYKLLP	GANKMLPKVF	FSETGIQRFL	PSEELLGKYE	
	RGTHKKGAGF	DLGDCRALID	FFKKSIERHD	DWKKFDFKFS	DTSTYQDISE	
	FYREVEQQGY	KMSFRKVSVD	YIKSLVEEGK	LYLFQIYNKD	FSAHSGKTPN	
	MHTLYWKMLF	DEENLKDVVY	KLNGEAEVFF	RKSSITVQSP	THPANSPIKN	
	KNKDNQKKES	KFEYDLIKDR	RYTVDKFLFH	VPITMNFKSV	GGSNINQLVK	
	RHIRSATDLH	IIGIDRGERH	LLYLTVIDSR	GNIKEQFSLN	EIVNEYNGNT	
	YRTDYHELLD	TREGERTEAR	RNWQTIQNI	ELKEGYLSQV	IHKISELAIK	
	YNAVIVLEDL	NFGFMRSRQK	VEKQVYQKFE	KMLIDKLNLY	VDKKKPVAE	
	GGLLRAYQLT	GEFESFKTLG	KQSGILFYVP	AWNTSKIDPV	TGFVNLFDTH	
	YENIEKAKVF	FDKFKSIRYN	SDKDWFEFV	DDYTRFSPKA	EGTRRDWTIC	
	TQGKRIQICR	NHQRNNEWEG	QEIDLTKAFK	EHFEAYGVDI	SKDLREQINT	
	QNKKEFFEEL	LRLRLTLQ	RNSMPSSDID	YLISPVANDT	GCFFDSRKQA	
	ELKENAVLPM	NADANGAYNI	ARKGLLAIRK	MKQEENDSAK	ISLAISNKEW	
	LKFAQTKPYL	ED				
WP_036890108.1 type V CRISPR- associated protein Cpf1 [Porphyromonas crevioricanis]	MDSLKDFTNL	YPVSKTLRFE	LKPVGKTLEN	IEKAGILKED	EHRAESYRRV	(SEQ ID NO: 140)
	KKIIDTYHKV	FIDSSLENMA	KMGIENEIKA	MLQSFCELYK	KDHRTEGEDK	
	ALDKIRAVLR	GLIVGAFTGV	CGRRENTVQN	EKYESLFKEK	LIKEILPDFV	
	LSTEAESLPF	SVEEATRSK	EFDSFTSYFA	GFYENRKNY	STKQPSTAI	
	YRLIHENLPK	FIDNILVFQK	IKETIAKELE	HIRADFSAGG	YIKKDERLED	
	IFSLNYYIHV	LSQAGIEKYN	ALIGKIVTEG	DGEMKGLNEH	INLYNQQRGR	
	EDRLPLFRPL	YKQILSDREQ	LSYLPESFEK	DEELLRALKE	FYDTHAEDIL	
	GRTQQLMTSI	SEYDLSRIYV	RNDSQLTDIS	KKMLGDWNAI	YMARERAYDH	
	EQAPKRITAK	YERDRIKALK	GEESISLANL	NSCIAFLDNV	RDCRVDTYLS	
	TLGQKEGPHG	LSNLVENVFA	SYHEAEQLLS	FPYPEENNLI	QDKDNVVLK	
	NLLDNISDLQ	RFLKPLWGMG	DEPDKDERFY	GEYNYIRGAL	DQVIPLYNKV	
	RNYLTRKPYS	TRKVKNLFGN	SQLLSGWDNR	KEKDNSCVIL	RKGQNFYLAI	
	MNNRHKRSFE	NKMLPEYKEG	EPYFEKMDYK	FLPDPNKMMLP	KVFLSKKGIE	
	IYKPSPKLLE	QYGHGTHKKG	DTFSMDDLHE	LIDFFKHSIE	AHEDWKQFGF	
	KFSDTATYEN	VSSFYREVED	QGYKLSFRKV	SESYVYSLID	QGKLYLFQIY	
	NKDFSPCSKG	TPNLHTLYWR	MLFDERNLAD	VITYKLDGKAE	IFFREKSLKN	

Table 5 - Cas12a orthologs						
	DHPHPAGKP	IKKKSQRQKKG	EESLFEYDLV	KDRRYTMDKF	QFHPVITMNF	
	KCSAGSKVND	MVNAHIREAK	DMHVIGIDRG	ERNLLYICVI	DSRGITILDQI	
	SLNTINDIDY	HDLLESRDKD	RQQEHRNWQT	IEGIKELKQG	YLSQAVHRIA	
	ELMVAYKAVV	ALEDLNMGFK	RGRQKVESSV	YQQFEKQLID	KLNYLVDKKG	
	RPEDIGLLR	AYQFTAPFKS	FKEMGKQNGF	LFYIPAWNTS	NIDPTTGFVN	
	LFHVQYENV	KAKSFFQKFD	SISYNPKKDW	FEFAFDYKNF	TKKAEGSRSM	
	WILCTHGSRI	KNFRNSQKNG	QWDSEEFALT	EAFKSLFVRY	EIDYTADLKT	
	AIVDEKQKDF	FVDLLKLFKL	TVQMRNSWKE	KDLDYLISPV	AGADGRFFDT	
	REGNKSPLPKD	ADANGAYNIA	LKGLWALRQI	RQTSEGGKLG	LAINSKEWLQ	
	FVQERSYEKD					
WP_036887416.1 type V CRISPR- associated protein Cpf1 [Porphyromonas crevioricanis]	MDSLKDFTNL	YPVSKTLRFE	LKPVGKTLEN	IEKAGILKED	EHRAESYRRV	(SEQ ID NO: 141)
	KKIIDTYHKV	FIDSSLENMA	KMGIEINEIKA	MLQSFCELYK	KDHRTEGEDK	
	ALDKIRAVLR	GLIVGAFTGV	CGRRENTVQN	EKYESLFKEK	LIKEILPDFV	
	LSTEAESLPP	SVEEATRSLK	EFDSFTSYFA	GFYENRKNY	STKQSTATA	
	YRLIHENLPP	FIDNILVFQK	IKUPIAKELE	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	FPYPEENNLI	QDKDNVVLK	
	NLLDNISDLQ	RFLKPLWGMG	DEPDKDERFY	GEYNYIRGAL	DQVIPLYNKV	
	RNYLTRKPYS	TRKVKLNFGN	SQLLSGWDRN	KEKDNSCVIL	RKGQNFYLA	
	MNNRHKRSFE	NKVLPEYKEG	EPYFEKMDYK	FLPDPNKMLP	KVFLSKKGIE	
	IYKPSPKLLE	QYGHGTHKKG	DTFSMDLHE	LIDFFKHSIE	AHEDWKQFGF	
	KFSDTATYEN	VSSFYREVED	QGYKLSFRKV	SESYVYSLID	QGKLYLFQIY	
	NKDFSPCSKG	TPNLHTLYWR	MLFDERNLAD	VIYKLDGKAE	IFFREKSLKN	
	DHPHPAGKP	IKKKSQRQKKG	EESLFEYDLV	KDRHYTMDKF	QFHPVITMNF	
	KCSAGSKVND	MVNAHIREAK	DMHVIGIDRG	ERNLLYICVI	DSRGITILDQI	
	SLNTINDIDY	HDLLESRDKD	RQQERRNWQT	IEGIKELKQG	YLSQAVHRIA	
	ELMVAYKAVV	ALEDLNMGFK	RGRQKVESSV	YQQFEKQLID	KLNYLVDKKG	
	RPEDIGLLR	AYQFTAPFKS	FKEMGKQNGF	LFYIPAWNTS	NIDPTTGFVN	
	LFHAQYENV	KAKSFFQKFD	SISYNPKKDW	FEFAFDYKNF	TKKAEGSRSM	
	WILCTHGSRI	KNFRNSQKNG	QWDSEEFALT	EAFKSLFVRY	EIDYTADLKT	
	AIVDEKQKDF	FVDLLKLFKL	TVQMRNSWKE	KDLDYLISPV	AGADGRFFDT	
	REGNKSPLPKD	ADANGAYNIA	LKGLWALRQI	RQTSEGGKLG	LAINSKEWLQ	
	FVQERSYEKD					
WP_023941260.1 type V CRISPR- associated protein Cpf1 [Porphyromonas crevioricanis]	MDSLKDFTNL	YPVSKTLRFE	LKPVGKTLEN	IEKAGILKED	EHRAESYRRV	(SEQ ID NO: 142)
	KKIIDTYHKV	FIDSSLENMA	KMGIEINEIKA	MLQSFCELYK	KDHRTEGEDK	
	ALDKIRAVLR	GLIVGAFTGV	CGRRENTVQN	EKYESLFKEK	LIKEILPDFV	
	LSTEAESLPP	SVEEATRSLK	EFDSFTSYFA	GFYENRKNY	STKQSTATA	
	YRLIHENLPP	FIDNILVFQK	IKUPIAKELE	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	FPYPEENNLI	QDKDNVVLK	
	NLLDNISDLQ	RFLKPLWGMG	DEPDKDERFY	GEYNYIRGAL	DQVIPLYNKV	
	RNYLTRKPYS	TRKVKLNFGN	SQLLSGWDRN	KEKDNSCVIL	RKGQNFYLA	
	MNNRHKRSFE	NKVLPEYKEG	EPYFEKMDYK	FLPDPNKMLP	KVFLSKKGIE	
	IYKPSPKLLE	QYGHGTHKKG	DTFSMDLHE	LIDFFKHSIE	AHEDWKQFGF	
	KFSDTATYEN	VSSFYREVED	QGYKLSFRKV	SESYVYSLID	QGKLYLFQIY	
	NKDFSPCSKG	TPNLHTLYWR	MLFDERNLAD	VIYKLDGKAE	IFFREKSLKN	
	DHPHPAGKP	IKKKSQRQKKG	EESLFEYDLV	KDRRYTMDKF	QFHPVITMNF	
	KCSAGSKVND	MVNAHIREAK	DMHVIGIDRG	ERNLLYICVI	DSRGITILDQI	
	SLNTINDIDY	HDLLESRDKD	RQQERRNWQT	IEGIKELKQG	YLSQAVHRIA	
	ELMVAYKAVV	ALEDLNMGFK	RGRQKVESSV	YQQFEKQLID	KLNYLVDKKG	
	RPEDIGLLR	AYQFTAPFKS	FKEMGKQNGF	LFYIPAWNTS	NIDPTTGFVN	
	LFHAQYENV	KAKSFFQKFD	SISYNPKKDW	FEFAFDYKNF	TKKAEGSRSM	
	WILCTHGSRI	KNFRNSQKNG	QWDSEEFALT	EAFKSLFVRY	EIDYTADLKT	

Table 5 - Cas12a orthologs						
	AIVDEKQKDF REGNKS LPKD FVQERSYEKD	FVDLLKLFKL ADANGAYNIA	TVQMRNSWKE LKGLWALRQI	KDL DY LISPV RQTSEGGK LK	AGADGRFFDT LAISNKEWLQ	
WP_037975888.1 type V CRISPR- associated protein Cpf1 [Synergistes jonesii]	MANSLKDFTN VTKLIDDYHR EKIISMFTSQ DKFSTYFTGF ERAPEVIDKA VSGKAGDKKL VPEALKDDEQ SSISAFSNEL ADIDAAVKSY SKAFRGDDGK YDSVKDVIPV AII LMKDGKY PKVFLPKRAP NKHKDWRTFG EEGRLFLFQI ELFFRRKSID ASLNAGARGL QGNINSDVNL IGGMDYHAKL DNNAVIVMED PCGILKGLQL NDVSSKEKQK WTNGKRIVRE SRDKSLASEI AALPIDADAN FMQTRKFDF	IYQLSKTLRF KFIHETL DPA AVYKKLFKKE HENRKNIYSK QKELSGILGG RGVNEFSNLY AINAVDGLRS FKNWSFIEDA SEQVSADISS TEIIKAYLDS YNAVRNYTTQ YLGVIDKNNK KRYNPSREIL FDFS DTD SYE YNKDFADGAQ KPAVHAKGSM IESGLVRITE FLRNNKDVNI NQKEKERDTA LNIGFKRGRF TEKFESFTKL DFIGKLDSIR KDKDGKFRMN FYVFKNTLQM GAYHIALKGS	ELKPIGKTEE HFDWNPLAEA LFSELLPEMI KDTSTSI AFR KTLDDIFSIE RQQHPEVASL ELERNDIFNR LKEFKENEFN APSASYFAKF LNDILHCLKP KPFSTEKFKL LRADDLADDG LIKENKTFIK DISDFYMEVQ GSPNLHTLYW KVNRRDIDGN VKHEL VKDKR IGIDRGERNL RKSWK TIGTI KVEKQVYQKF GKQCGIIFYI FDAKRDMFTF DRLLTEDMKN RNSKRDTGED LVLDAIDEKL	HINRKLIIMH LIQSGSKNNK KSELVSDLEK IVHQNF PKFL SFNNVLTQDK RIKMVPLYKQ IKRLFGKNNL GARSAGKKA E TNLIETAAEN FETEDISDID NFENPALAKG SAYGYMKMNY DKNFNRDCH DQGYKLTFTF KAIFSEENLK PIDEGTYVEI YTIDKYFFHV VYVSLIDRDG KELKEGYLSQ EKMLIDK LNY PAGYTSKIDP EFDYDKFR TY ILNKYALAYK FIISPVLNAK KEDGRIDYKD	DEKRGEDYKS ALPAEQKEMR QAQLDAVKSF ANVRAYTLIK IDYYNQIIGG ILSDRTT LSF YSLDKIWIKN KWLKSKYFSF GRKFSYFAAE TEFYSAFAEI WDKNKEQNNT KFIPTPHMEI KLIDFFKDSI LSAEKIDKWF DVVLKLNGEA CGYANGKRDM PFTINFKAQG HIK LQKDFNI VVHEIVRLAV LVFKDAGYDA TTGFVNLFNI QTSYRK K WAV AGEDILPDVI GRFFDSRKTD MAVSNPKWFE	(SEQ ID NO: 143)
WP_081839471.1 type V CRISPR- associated protein Cpf1 [Synergistes jonesii]	MENMANSLKD YKSVTKLIDD EMREKII SMF KSFDKFSTYF LIKERAPEVI IGGVSGKAGD LSFVPEALKD IKNSSISAFS FSFADIDAAV AAESKAFRGD AEIYDSVKDV NNTAII LMKD MELPKVFLPK DSINKHKDWR KWVEEGRLFL GEAELFFRRK RDMASLNAGA AQGGQGNINS FNIIGGMDYH LAVDNNAVIV YDAPCGILKG FNINDVSSKE WAVWTNGKRI DVISRDKSLA KTDAALPIDA WFEFMQTRKF	FTNIYQLSKT YHRKFIHETL TSQAVYKKLF TGFHENRKNI DKAQKELSGI KKLRGVNEFS DEQAINAVDG NELFKNWSFI KSYSEQVSAD DGKTEIIKAY IPVYNAVRNY GKY YLGVIDK RAPKRYNPSR TFGFDFS DTD FQIYNKDFAD SIDKPAVHAK RGLIESGLVR VNLFLRNKND AKLNQKEKER MEDLNIGFKR LQLTEKFESF KQKDFIGKLD VREKDKDGKF SEIFYVFKNT DANGAYHIAL DF	LRFELKPIGK DPAHFDWNPL KKELFSELLP YSKKDTSTSI LGKTLDDIF NLYRQQHPEV LRSELERNDI EDALKEFKEN ISSAPSASYF LDSLNDILHC TTQKPFSTEK NNKL RADDLA EILLIKENKT SYEDISDFYM TKLGKQCGII SIRFDAKRDM RMNDRLLTED LQMRNSKRDT KGS LVLDAID	TEEHINRKLI AEALIQSGSK EMIKSELVSD AFRIVHQNF SIESFNNVLT ASLRIMVPL FNRIKRLFGK EFNGARSAGK AKFTNLIETA LKP FETEDIS FKLN FENPAL DDGSAYGYMK FIKDKNFNRT EVQDQGYKLT LYWK AIFSEE DGNPIDEGTY DKRYTIDKYF RNLVYVSLID GTIKELKEGY QKFEKMLIDK FYIPAGYTSK FTFEFDYDKF GEDFIISPVL EKLKEDGRID	IMHDEKRGED NNKALPAEQK LEKQAQLDAV KFLANVRAYT QDKIDYYNQI YKQILSDRTT NNLYSLDKIW KAEKW LKSKY AENGRKFSYF DIDTEFYSAF AKGWDKNKEQ MNYKFIPTPH DCHKLIDFFK FTRL SAEKID NLKDVVLKLN VEICGYANGK FHVPFTINFK RDGHIK LQKD LSQVVHEIVR LNYLVFKDAG IDPTTGFVNL RTYQTSYRK K AYKAGEDILP NAKGRFFDSR YKDMAVSNPK	(SEQ ID NO: 144)
WP_006283774.1 type V CRISPR- associated protein	MQINNLKIIY HRADSYKKVK KRIEKTEKDK SDEERTLIKE	MKFTDFTGLY KIIDEYHKAF FAKIQDNLRK FKDFTTYFKG	SLSKTLRFEL IEKSLSNFEL QIADHLKGDE FYENRENMY S	KPIGKTLENI KYQSEDKLDS SYKTIFSKDL AEDKSTAISH	KKAGLLEQDQ LEEYLMYYSM IRKNLPDFVK RIIHENLPKF	(SEQ ID NO: 145)

Table 5 - Cas12a orthologs						
Cpf1 [Prevotella bryantii B14]	VDNINAFSKI	ILIPELREKL	NQIYQDFEEY	LNVESIDEIF	HLDYFSMVMT	
	QKQIEVYNAI	IGGKSTNDKK	IQGLNEYINL	YNQKHKDCKL	PKLKLLFKQI	
	LSDRIAISWL	PDNFKDDQEA	LDSIDTCYKN	LLNDGNVLGE	GNLKLLLENI	
	DTYNLKGIFI	RNDLQLTDIS	QKMYASWNVI	QDAVILDLKK	QVSRKKKESA	
	EDYNDRLLKKL	YTSQESFSIQ	YLNDCLRAYG	KTENIQDYFA	KLGAVNNEHE	
	QTINLFAQVR	NAYTSVQAIL	TTPYPENANL	AQDKETVALI	KNLLDSLKRL	
	QRFIKPLLGK	GDESDKDERF	YGDFTPLWET	LNQITPLYNM	VRNYMTRKPY	
	SQEKIKLNFE	NSTLLGGWDL	NKEHDNTAI I	LRKNGLYYLA	IMKKSANKIF	
	DKDKLDNSGD	CYEKMVYKLL	PGANKMLPKV	FFSKSRIDEF	KPSENI IENY	
	KKGTHKKGAN	FNLADCHNLI	DDFKSSISKH	EDWSKFNHFH	SDTSSYEDLS	
	DFYREVEQQG	YSISFCDVSV	EYINKMVEKG	DLYLFQIYNK	DFSEFSKGTG	
	NMHTLYWNSL	FSKENLNNII	YKLNGQAEIF	FRKKS LNYKR	PTHPAHQAIA	
	NKNKCNEKKE	SIFDYDLVKD	KRYTVDKFQF	HVPITMNFKS	TGNTNINQQV	
	IDYLRTEDDT	HIIGIDRGER	HLLYLVVIDS	HGKIVEQFTL	NEIVNEYGGN	
	IYRTNYHDL	DTREQNREKA	RESWQTIENI	KELKEGYISQ	VIHKITDLMQ	
	KYHAVVLED	LNMGMFMRGR	KVEKQVYQKF	EEMLINKLNY	LVNKKADQNS	
	AGGLLHAYQL	TSKFESFQKL	GKQSGFLFYI	PAWNTSKIDP	VTGFFVNLFD	
	RYESIDKAKA	FFGKFDSIRY	NADKDWFEFA	FDYNNFTTKA	EGTRTNWTIC	
	TYGSRI RTFR	NQAKNSQWDN	EEIDLTKAYK	AFFAKHGINI	YDNIKEATAM	
	ETEKSPFEDL	LHLLKLTLQM	RNSITGTTTD	YLISPVHDSK	GNFYDSRICD	
	NSLPANADAN	GAYNIARKGL	MLIQQIKDST	SSNRKFKSPI	TNKDWLIFAQ	
	EKPYLND					
WP_024988992 type V CRISPR- associated protein Cpf1 [Prevotella albensis]	MNIKNFTGLY	PLSKTLRFEL	KPIGKTKENI	EKNIGILTKDE	QRAKDYLIVK	(SEQ ID NO: 146)
	GFIDEYHKQF	IKDRLWDFKL	PLESEGEKNS	LEEYQELYEL	TKRNDQAQEA	
	FTEIKDNLRS	SITEQLTKSG	SAYDRIFKKE	FIREDLVNFL	EDEKDKNIVK	
	QFEDFTTYFT	GFYENRKNMY	SSEKSTAI A	YRLIHQNLPK	FMDNMRSFAK	
	IANSSVSEHF	SDIYESWKEY	LVNSIEEIEF	QLDYFSETLT	QPHIEVYNYI	
	IGKKVLEDGT	EIKGINNEYVN	LYNQQQKDKS	KRLPFLVPLY	QKILSDREKL	
	SWIAEEFDSD	KKMLSAITES	YNHLHNVLMG	NENESLRNLL	LNKIDYNLEK	
	INITNDLSLT	EISQNLFGRY	DVFTNGIKNK	LRVLTPRKKK	ETDENFEDRI	
	NKIFKTQKSF	SIAFLNKL PQ	PEMEDGKPRN	IEDYFITQGA	INTKSIQKED	
	IFAQIENAYE	DAQVFLQIKD	TDNKLSONKT	AVEKIKTLLD	ALKELQHFIF	
	PLLGSGEENE	KDELFYGSFL	AIWDELDTIT	PLYNKVRNWL	TRKPYSTEKI	
	KLNFNDNAQLL	GGWDVNKEHD	CAGILLRKND	SYLLGIINKK	TNHIFDITIT	
	PSDGECYDKI	DYKLLPGANK	MLPKVFFSKS	RIKEFEPSEA	IINCYKKGTH	
	KKGKNFNLT	CHRLINFFKT	SIEKHEDWSK	FGFKFSDTET	YEDISGFYRE	
	VEQQGYRLTS	HPVSASYIHS	LVKEGKLYLF	QIWNKDFSQF	SKGTPNLHTL	
	YWKMLFDKRN	LSDVVYKLN	QAEVFYRKSS	IEHQNR I IHP	AQHPI TNKNE	
	LNKKHTSTFK	YDI IKDRRYT	VDKFQFHVPI	TINFKATGQN	NINPIVQEVI	
	RQNGITHIIG	IDRGERHLLY	LSLIDLKGN I	IKQMTLNEII	NEYKGVTYKT	
	NYHNLLEKRE	KERTEARHSW	SSIESIKELK	DGYMSQVIHK	ITDMMVKYNA	
	IVVLEDLNGG	FMRGRQKVEK	QVYQKFEEKL	IDKLNLYLVDK	KLDANEVGGV	
	LNAYQLTNKF	ESFKKIGKQS	GFLFYIPAWN	TSKIDPITGF	VNLFNTRYES	
	IKETKVFWSK	FDI IRYNKEK	NWFEFVFDYN	TFTTKAEGTR	TKWTLCTHGT	
	RIQTFRNPEK	NAQWDNKEIN	LTESFKALFE	KYKIDITSNL	KESIMQETEK	
	KFFQELHNLL	HLTLQMRNSV	TGTDIDY LIS	PVADEDGNFY	DSRINGKNFP	
	ENADANGAYN	IARKGLMLIR	QIKQADPQKK	FKFETITNKD	WLKFAQDKPY	
	LKD					
WP_039658684.1 type V CRISPR- associated protein Cpf1 [Smithella sp. SC_K08D17]	MQTLFENFTN	QYPVSKTLRF	ELIPQGKTKD	FIEQKGLLKK	DEDRAEKYKK	(SEQ ID NO: 147)
	VKNIIIDEYHK	DFIEKSLNGL	KLDGLEKYKT	LYLKQEKDDK	DKKAFDKEKE	
	NLRKQIANAF	RNNEKFKT LF	AKELIKNDLM	SFACEEDKKN	VKEFEAFTTY	
	FTGFHQNRAN	MYVADEKRTA	IASRLIHENL	PKFIDNIKIF	EKMKEAPEL	
	LSPFNQTLKD	MKDVIKGTTL	EEIFSLDYFN	KTLTQSGIDI	YNSVIGGRTP	
	BEGKTIKIGL	NEYINTDFNQ	KQTDKKKRQP	KFKQLYKQIL	SDRQSLSFIA	
	BAFKNDTEIL	EAIEKFYVNE	LLHFSNEGKS	TNVLDAIKNA	VSNLESFNLT	
	KMYFRSGASL	TDVSRKVFE	WSI INRALDN	YYATTYPIKP	REKSEKYEER	
	KEKWLKQDFN	VSLIQTAIDE	YDNETVKGKN	SGKVIADYFA	KFCDDKETDL	
	IQKVNIEGYIA	VKDLLNTPCP	ENEKLGSKND	QVKQIKAFMD	SIMDIMHFVR	
	PLSLKDTDK	KDETFYSLFT	PLYDHLTQTI	ALYNKVRNYL	TQKPYSTEKI	

Table 5 - Cas12a orthologs						
	KLNFFENSTLL GGWDLNKETD NTAIILRKDN LYYLGIMDKR HNRIFRNVPK ADKKDFCYEK MVYKLLPGAN KMLPKVFFSQ SRIQEFTPSA KLLNENYANET HKKGDNFNIN HCHKLIDFFK DSINKHEDWK NFDPRFSATS TYADLSGFYH EVEHQGYKIS FQSVADSFID DLVNEGKLYL FQIYNKDFSP FSKGKPNLHT LYWKMLFDEN NLKDVVYKLN GEAEVIFYRKK SIAEKNTTTH KANESIINKN PDNPKATSTF NYDIVKDKRY TIDKFQFHIP ITMNFKAEGI FNMNQRVNQF LKANPDINII GIDRGERHLL YYALINQKGK ILKQDTLNVI ANEKQKVDYH NLLDKKEGDR ATARQEWGVI ETIKELKEGY LSQVIHKLTD LMIENNIAIV MEDLNFQFKR GRQKVEKQVY QKFEKMLIDK LNYLVDKNKK ANELGGLLNA FQLANKFESF QKMKGQNGFI FYVPAWNTSK TDPATGFIDF LKPRYENLNQ AKDFFKFDSD IRLNSKADYF EFAFDKFNFT EKADGGRTKW TVCTTNEEDRY AWRALNNNR GSQEKYDITA ELKSLFDGKV DYKSGKDLKQ QIASQESADF FKALMKNLSI TSLRHNNGE KGDNEQDYIL SPVADSKGRF FDSRKADDDM PKNADANGAY HIALKGLWCL EQISKTDDLK KVKLAISNKE WLEFVQTLKG					
WP_037385181 type V CRISPR- associated protein Cpf1 [Smithella sp. SCADC]	MQTLFENFTN QYPVSKTLRF ELIPQGKTKD FIEQKGLLKK DEDRAEKYKK VKNIIDEYHK DFIEKSLNGL KLDGLEEYKT LYLKQEKDDK DKKAFDKEKE NLRKQIANAF RNNEKFKTLF AKELIKNDLM SFACEEDKKN VKEFEAFTTY FTGFHQNRAN MYVADEKRTA IASRLIHENL PKFIDNIKIF EKMKEAPEL LSPFNQTLKD MKDVIKGTTL EEIFSLDYFN KTLTQSGIDI YNSVIGGRTP BEGKTKIKGL NEYINTDFNQ KQTDKKKRQP KFKQLYKQIL SDRQSLSFIA EAFKNDTEIL EAIEKFYVNE LLHFSNEGKS TNVLDIAKNA VSNLESFNLT KIYFRSGTSL TDVSRKVFE WSIIINRALDN YYATTYPIK REKSEKYBER KEKWLKQDFN VSLIQTIDE YDNETVKGKN SGKVIVDYFA KFCDDKETDL IQKVNIGYIA VKDLLNTPYP ENEKLGSNKD QVKQIKAFMD SIMDIMHFVR PLSLKDTDE KDETFSYSLT PLYDHLTQTI ALYNKVRNYL TQKPYSTEKI KLNFFENSTLL GGWDLNKETD NTAIILRKDN LYYLGIMDKR HNRIFRNVPK ADKKDFCYEK MVYKLLPGAN KMLPKVFFSQ SRIQEFTPSA KLLNENYANET HKKGDNFNIN HCHKLIDFFK DSINKHEDWK NFDPRFSATS TYADLSGFYH EVEHQGYKIS FQSVADSFID DLVNEGKLYL FQIYNKDFSP FSKGKPNLHT LYWKMLFDEN NLKDVVYKLN GEAEVIFYRKK SIAEKNTTTH KANESIINKN PDNPKATSTF NYDIVKDKRY TIDKFQFHVP ITMNFKAEGI FNMNQRVNQF LKANPDINII GIDRGERHLL YYALINQKGK ILKQDTLNVI ANEKQKVDYH NLLDKKEGDR ATARQEWGVI ETIKELKEGY LSQVIHKLTD LMIENNIAIV MEDLNFQFKR GRQKVEKQVY QKFEKMLIDK LNYLVDKNKK ANELGGLLNA FQLANKFESF QKMKGQNGFI FYVPAWNTSK TDPATGFIDF LKPRYENLNQ AKDFFKFDSD IRLNSKADYF EFAFDKFNFT EKADGGRTKW TVCTTNEEDRY AWRALNNNR GSQEKYDITA ELKSLFDGKV DYKSGKDLKQ QIASQELADF FRTLKMYLSV TSLRHNNGE KGETEQDYIL SPVADSMGKF FDSRKAGDDM PKNADANGAY HIALKGLWCL EQISKTDDLK KVKLAISNKE WLEFVQTLKG					(SEQ ID NO: 148)
WP_039871282.1 type V CRISPR- associated protein Cpf1 [Prevotella bryantii B14]	MKFTDFTGLY SLSKTLRFEL KPIGKTLNI KKAGLLEQDQ HRADSYKKVK KIIDEYHKAF IEKSLSNFEL KYQSEDKLDS LEEYLMYYSM KRIEKTEKDK FAKIQDNLRK QIADHLKGDE SYKTIFSKDL IRKNLPDFVK SDEERTLIKE FKDFTTYFKG FYENRENMYA AEDKSTAISH RIIHENLPKF VDNINAFSKI ILIPELREKL NQIYQDFEY LNVESIDEIF HLDYFSMVM TQKQIEVYNAI IGGKSTNDKK IQGLNEYINL YNQKHDKCKL PKLKLFLKQI LSDRIAISWL PDNFKDDQEA LDSIDTCYKN LLNDGNVLGE GNLKLLLENI DTYNLKGIFI RNDLQLTDIS QKMYASWNI QDAVILDLKK QVSRKKKESA EDYNDRLKKL YTSQESFSIQ YLNDCLRAYG KTENIQDYFA KLGAVNNEHE QTINLFAQVR NAYTSVQAIL TTPYPENANL AQDKETVALI KNLLDSLKRL QRFIKPLLK GDESDDERF YGDFTPLWET LNQITPLYNM VRNYMTRKPY SQEKIKLNFE NSTLLGGWDL NKEHDNTAII LRKNGLYYLA IMKKSANKIF DKDKLDNSGD CYEKMVYKLL PGANKMLPKV FFSKSRIDEF KPSENIENY KKGTHKKGAN FNLADCHNLI DFFKSSISKH EDWSKFNHFF SDTSSYEDLS DFYREVEQQG YSISFCDVSV EYINKMVEKG DLYLFQIYNK DFSEFSKGTP NMHTLYWNSL FSKENLNII YKLNQQAEIF FRKKSILNYKR PTHPAHQAIK NKNKCNKKEKE SIFDYDLVKD KRYTVDKFQF HVPITMNFKS TGNTNINQQV IDYLRTEDDT HIIGIDRGER HLLYLVIDS HGKIVEQFTL NEIVNEYGGN IYRTNYHDL DTREQNREKA RESWQTIENI KELKEGYISQ VIHKITDLMQ KYHAVVVLLE LNMGMFRGRQ KVEKQVYQKF EEMLINKLNY LVNKKADQNS AGLLHAYQL					(SEQ ID NO: 149)

Table 5 - Cas12a orthologs						
	TSKFESFQKL	GKQSGFLFYI	PAWNTSKIDP	VTGFVNLFDT	RYESIDKAKA	
	FFGKFDSIRY	NADKDWFEFA	FDYNNFTTKA	EGTRTNWTIC	TYGSRIRTFR	
	NQAKNSQWDN	EEIDLTKAYK	AFFAKHGINI	YDNIKEAIAM	ETEKSPFEDL	
	LHLLKLTLQM	RNSITGTTTD	YLISPVHDSK	GNFYDSRICD	NSLPANADAN	
	GAYNIARKGL	MLIQQIKDST	SSNRKFSPFI	TNKDWLIFAQ	EKPPLYND	
EKE28449.1 hypothetical protein ACD_3C00058G00 15 [uncultured bacterium (gcode 4)]	MFKGDAFTGL	YEVQKTLRFE	LVPIGLTQSY	LENDWVIQKD	KEVEENYGI	(SEQ ID NO: 150)
	KAYFDLIHKE	FVRQSLENAW	LCQLDDFYEK	YIELHNSLET	RKDKNLAKQF	
	EKVMKSLKKE	FVSFFDAKWN	EWKQKFSFLK	KWWIDVLNEK	EVLDLMAEFY	
	PDEKELFDKF	DKFFTYFSNF	KESRKNFYAD	DGRAWAIAIR	AIDENLITFI	
	KNIEDFKKLN	SSFREFVNDN	FSEEDKQIFE	IDFYNNCLLQ	PWIDKYNKIV	
	WWYSLENWEK	VQWLNEKINN	FKQNQNKSNS	KDLKFPKMKL	LYKQILGDKE	
	KKVYIDEIRD	DKNLIDLIDN	SKRRNQIKID	NANDIINDFI	NNNAKFELDK	
	IYLTRQSINT	ISSKYFSSWD	YIRWYFWTGE	LQEFVVSFYDL	KETFWKIEYE	
	TLENIFKDCY	VKGINTESQN	NIVFETQGIY	ENFLNIFKFE	FNQNISQISL	
	LEWELDKIQN	EDIKKNEKQV	EVIKNYFDSV	MSVYKMTKYF	SLEKWKRRVE	
	LDTDNNFYND	FNEYLEGFEI	WKDYNLVRNY	ITKKQVNTDK	IKLNFDSNSQF	
	LTWWDKDKEN	ERLGIILRRE	WKYYLWILKK	WNTLNFQDYL	QKEWEIFYEK	
	MNYKQLNNVY	RQLPRLLFPL	TKKLNELKWD	ELKKYLSKYI	QNFYWNBEIA	
	QIKIEFDIFQ	ESKEKWEKFD	IDKLRKLIEY	YKKWVLALYS	DLYDLEFIKY	
	KNYDDLSIFY	SDVEKKMYNL	NFTKIDKSLI	DGKVKSWELY	LFQIYNKDFS	
	ESKKEWSTEN	IHTKYFKLLF	NEKNLQNLV	KLSWWADIFF	RDKTENLKFK	
	KDKNGQEILD	HRRFSQDKIM	FHISITLNAN	CWDKYWFNQY	VNEYMNKERD	
	IKIIWIDRWE	KHLAYCYVID	KSWKIFNNEI	WTLNELNWN	YLEKLEKIES	
	SRKDSRISWW	EIENIKELKN	GYISQVINKL	TELIVKYNAI	IVFEDLNIWF	
	KRWRQKIEKQ	IYQKLELALA	KKLNYLTQKD	KKDDEILWNL	KALQLVPKVN	
	DYQDIWNYKQ	SWIMFYVRAN	YTSVTCPCNW	LRKNLYISNS	ATKENQKKS	
	NSIAIKYNDW	KFSFSYEIDD	KSWKQKQSLN	KKKFIVYSDI	ERFVYSPLEK	
	LTKVIDVNKK	LLELFRDFNL	SLDINKQIQE	KDLDSVFFKS	LTHLFNLILQ	
	LRNSDSKDNK	DYISCPSCYY	HSNNWLQWFE	FNWDANWAYN	IARKGIILLD	
	RIRKNQEKPD	LYVSDIDWDN	FVQSNQFPNT	IPIQNIQEKQ	VPLNIKI	
WP_018359861.1 type V CRISPR- associated protein Cpf1 [Porphyromonas macacae]	MKTQHFFEDF	TSLYSLSKTI	RFELKPIGKT	LENIKKNGLI	RRDEQRLLDY	(SEQ ID NO: 151)
	EKLKKVIDEY	HEDFIANILS	SFSFSEEILQ	SYIQNLSESE	ARAKIEKTM	
	DTLAKAFSED	ERYKSIFKKE	LVKKDIPVWC	PAYKSLCKKF	DNFTTSLVPF	
	HENRNLYTS	NEITASIPYR	IVHVNLPKFI	QNIEALCELQ	KKMGADLYLE	
	MMENLRNVWP	SFVKTDPDDL	NLKTYNHLMV	QSSISEYNRF	VGGYSTEDGT	
	KHQGINWIN	IYRQRNKEMR	LPGLVFLHKQ	ILAKVDSSSF	ISDTLENDDQ	
	VFCVLRQFRK	LFWNTVSSKE	DDAASLKDLF	CGLSGYDPEA	IYVSDAHLAT	
	ISKNIIDRWN	YISDAIRRKT	EVLMPRKKE	VERYAEKISK	QIKKRQSYSL	
	AELDDLLAHY	SEESLPAGFS	LLSYFTSLGG	QKYLVSDEGEV	ILYEESGNIW	
	DEVLIAFRDL	QVILDKDFTE	KKLGKDEEAV	SVIKKALDSA	LRLRKFFDLL	
	SGTGAEIRRD	SSFYALYTDR	MDKLKGLLKM	YDKVRNYLTK	KPYSIEKFKL	
	HFDNPSLLSG	WDKNKELNNL	SVIFRQNGYY	YLGIMTPKGG	NLFKTLPLKG	
	AEEMFYEKME	YKQIAEPMLM	LPKVFFPKKT	KPAFAPDQSV	VDIYNKKTFF	
	TGQKGFNKKD	LYRLIDFYKE	ALTVHEWKLF	NFSFSPTEQY	RNIGEFFDEV	
	REQAYKVMV	NVPASYIDEA	VENGKLYLFQ	IYNKDFSPYS	KGIPNLHTLY	
	WKALFSEQNO	SRVYKLCGGG	ELFYRKASLH	MQDTTVHPKG	ISIHKKNLNK	
	KGETSLFNVD	LVKDKRFTED	KFFFHVPISI	NYKNKKITNV	NQMVRDYIAQ	
	NDDLQIIGID	RGERNLLYIS	RIDTRGNLLE	QFSLNVIESD	KGDLRTDYQK	
	ILGDREQERL	RRRQEWKSIE	SIKDLKDGYM	SQVVHKICNM	VVEHKAIVVL	
	ENLNLFSFMKG	RKKVEKSVYE	KFERMLVDKL	NYLVVDKKNL	SNEPGGLYAA	
	YQLTNPLFSF	EELHRYPQSG	ILFFVDPWNT	SLTDPSTGFV	NLLGRINYTN	
	VGDARKFFDR	FNAIRYDGKG	NILFDLDSR	FDVRVETQRK	LWTLTTFGSR	
	IAKSKSGKW	MVERIENLSL	CFLELFEQFN	IGYRVEKDLK	KAILSQDRKE	
	FYVRLIYLFN	LMMQIRNSDG	EEDYILSPAL	NEKNLQFDSR	LIEAKDLPVD	
	ADANGAYNVA	RKGLMVQORI	KRGDHEIHR	IGRAQWLRYV	QEGIVE	
WP_013282991 type V CRISPR- associated protein	MLLYENYTKR	NQITKSLRLE	LRPQGKTLRN	IKELNLLEQD	KAIYALLERL	(SEQ ID NO: 152)
	KPVIDEGIKD	IARDTLKNCE	LSFEKLYEHF	LSGDKKAYAK	ESERLKKEIV	
	KTLIKNLPEG	IGKISEINSA	KYLVGVLYDF	IDKTHKDSSE	KQNILSDILE	
	TKGYLALFSK	FLTSRITLLE	QSMPCRVIEN	FEIYAANIPK	MQDALERGAV	

Table 5 - Cas12a orthologs						
Cpf1 [Butyrivibrio proteoclasticus]	SFAIEYESIC	SVDYNNQILS	QEDIDSYNRL	ISGIMDEDGA	KEKGINQTIS	
	EKNIKIKSEH	LEEKPFRIK	QLHKQILEER	EKAFTIDHID	SDEEVVQVTK	
	EAFEQTKEQW	ENIKKINGFY	AKDPGDITLF	IVVGPNQTHV	LSQLIYGEHD	
	RIRLLLEEYE	KNTLEVLPRR	TKSEKARYDK	FVNAVPPKVA	KESHTFDGLQ	
	KMTGDDRLFI	LYRDELARNY	MRIKEAYGTF	ERDILKSRRG	IKGNRDVQES	
	LVSFYDELTK	FRSALRIINS	GNDEKADPIF	YNTFDGIFEK	ANRTYKAENL	
	CRNYVTKSPA	DDARIMASCL	GTPARLRTHW	WNGEENFAIN	DVAMIRRGDE	
	YYYFVLTPDV	KPVDLKTKE	TDAQIFVQRK	GAKSFLGLPK	ALFKCILEPY	
	FESPEHKNDK	NCVIEEYVSK	PLTIDRRAYD	IFKNGTFKKK	NIGIDGLTEE	
	KFKDDCRYLI	DVYKEFIAVY	TRYSCFNMSG	LKRADEYNDI	GEFFSDVDTR	
	LCTMEWIPVS	FERINDMVDK	KEGLLFLVRS	MFLYNRPKRP	YERTFIQLFS	
	DSNMEHTSML	LNSRAMIQYR	AASLPRRVTH	KKGSILVALR	DSNGEHIPMH	
	IREAIYKMK	NFDISSEDFI	MAKAYLAEHD	VAIKKANEDI	IRNRRYTEDK	
	FFLSLSYTKN	ADISARTLDY	INDKVEEDTQ	DSRMAVIVTR	NLKDLTYVAV	
	VDEKNNVLEE	KSLNEIDGVN	YRELLKERTK	IKYHDKTRLW	QYDVSSKGLK	
	EAYVELAVTQ	ISKLATKYNA	VVVVESMSST	FKDKFSFLDE	QIFKAPEARL	
	CARMSDLSFN	TIKEGEAGSI	SNPIQVSNNN	GNSYQDGVIT	FLNNAYTRTL	
	CPDTGFVDVF	DKTRLITMQS	KRQFFAKMKD	IRIDDGEMLF	TFNLEEYPTK	
	RLLDRKEWTV	KIAGDGSYFD	KDKGEYVYVN	DIVREQUIIPA	LLEDKAVFDG	
	NMAEKFLDKT	AISGKSVELI	YKWFANALYG	IITKKDGEEKI	YRSPITGTEI	
	DVSKNTTYNF	GKKFMFKQEI	RGDGDFLDAF	LNMQAQDIA	V	
WP_048112740.1 type V CRISPR- associated protein Cpf1 [Candidatus Methanoplasma termitum]	MNNDYDEFTKL	YPIQKTIRFE	LKPQGRTEH	LETFFNFEEED	RDRAEKYKIL	(SEQ ID NO: 153)
	KEAIDEYHKK	FIDEHLTNMS	LDWNSLKQIS	EKYYKSREEK	DKKVFLSEQK	
	RMRQEIVSEF	KKDDRFDKLF	SKKLFSSELLK	EEIYKKGNHQ	EIDALKSFDK	
	FSGYFIGLHE	NRKNMYSDDG	EITAISNRIV	NENFPKFLDN	LQKYQEARKK	
	YPEWIIKAES	ALVAHNIKMD	EVFSLEYFNK	VLNQEGIQRY	NLALGGYVTK	
	SGEKMMGLND	ALNLAHQSEK	SSKGRIHMT	LFKQILSEKE	SFSYIPDVFT	
	EDSQLLPSIG	GFFAQIENDK	DGNIFDRALE	LISYAEYDT	ERIYIRQADI	
	NRVSNVIFGE	WGTLGGLMRE	YKADSINDIN	LERTCKKVDK	WLDSKEFALS	
	DVLEAIKRTG	NNDAFNEYIS	KMRTAREKID	AARKEMKFIS	EKISGDEESI	
	HIKTLTLDV	QQFLHFFNLF	KARQDIPLDG	AFYAEFDEVH	SKLFAIVPLY	
	NKVRNYLTKN	NLNTKKIKLN	FKNPTLANGW	DQNKVYDYAS	LIFLRDGNYY	
	LGIINPKRKK	NIKFEQSGSN	GPFYRKMVYK	QIPGPNKNLP	RVFLTSTKKG	
	KEYKPSKEII	EGYEADKHIR	GDKFDLDFCH	KLIDFFKESI	EKKHDKWKFN	
	FYFSPTESYG	DISEFYLDVE	KQGYRMHFEN	ISAETIDEYV	EKGDLFLFQI	
	YNKDFVKAAT	GKKDMHTIYW	NAAFSPENLQ	DVVVKLNGEA	ELFYRDKSDI	
	KEIVHREGEI	LVNRTYNGRT	PVPDKIHKKL	TDYHNGRTKD	LGEAKEYLKD	
	VRYFKAHYDI	TKDRRYLNDK	IYFHVPLTLN	FKANGKKNLN	KMVIEKFLSD	
	EKAHIIIGIDR	GERNLLYYSI	IDRSGKIIDQ	QSLNVIDGFD	YREKLNQREI	
	EMKDARQSWN	AIGKIKDLKE	GYLSKAVHEI	TKMAIQYNAI	VVMEELNYGF	
	KRGRFKVEKQ	IYQKFENMLI	DKMNYLVFKD	APDESPGGVL	NAYQLTNPLE	
	SFAKLKGQTG	ILFYVPAAYT	SKIDPTTGTV	NLFNTSSKTN	AQERKEFLQK	
	FESISYSAKD	GGIFAFAFDY	RKFGTSKTDH	KNVWTAYTNG	ERMRYIKEKK	
	RNELFDPSKE	IKEALTSSGI	KYDGGQNILP	DILRSNNNGL	IYTMYSFFIA	
	AIQMRVYDGG	EDYIISPIKN	SKGEFFRTDP	KRRELPIAD	ANGAYNIALR	
	GELTMRAIAE	KFDPDSEKMA	KLELKHKDW	BFMQTRGD		
WP_027407524.1 type V CRISPR- associated protein Cpf1 [Anaerovibrio sp. RM50]	MVAFIDEFVG	QYPVSKTLRF	EARPVPETKK	WLESQCSVL	FNDQKRNEYY	(SEQ ID NO: 154)
	GVLKELDDY	YRAYIEDALT	SFTLDKALLE	NAYDLYCNRD	TNAFSSCCEK	
	LRKDLVKAFG	NLKDYLLGSD	QLKDLVKLKA	KVDAPAGKKG	KKIEVDSRLI	
	NWLNNNAKYS	AEDREKYIKA	IESFEGFVTY	LTNYKQAREN	MFSSSEKSTA	
	IAFRVIDQNM	VTYFGNIRIY	EKIKAKYPEL	YSALKGFEEK	FSPTAYSEIL	
	SQSKIDEYNY	QCIGRPIDDA	DFKGVNSLIN	EYRQKNGIKA	RELPMVSMY	
	KQILSDRDNS	FMSEVINRNE	EAIECAKNGY	KVSALFNL	LQLYKKIFTE	
	DNYGNIYVKT	QPLTELSQAL	FGDWSILRNA	LDNGKYDKDI	INLAELEKYF	
	SEYCKVLAD	DAAKIQDKFN	LKDYFIQKNA	LDATLPDLK	ITQYKPHLDA	
	MLQAIRKYKL	FSMYNGRKKM	DVPENGIDFS	NEFNAIYDKL	SEFSILYDRI	
	RNFATKKPYS	DEKMKLSFNM	PTMLAGWDYN	NETANGCFLF	IKDGKYLFGV	
	ADSKSKNIFD	FKKNPHLLDK	YSSKDIYKVV	KYKQVSGSAK	MLPKVVFAGS	
	NEKIFGHLIS	KRILEIREKK	LYTAAAGDRK	AVAEWIDFMK	SAIAIHPEWN	

Table 5 - Cas12a orthologs						
	EYFKFKFKNT	AEYDNANKFY	EDIDKQTYSL	EKVEIPTFYI	DEMVSQHKLY	
	LFQLYTKDFS	DKKKKKGTDN	LHTMYWHGVF	SDENLKAVTE	GTQPIIKLNG	
	EAEMFMRNPS	IEFQVTHEHN	KPIANKNPLN	TKKESVFNVD	LIKDKRYTER	
	KFYFHCPIITL	NFRADKPIKY	NEKINRFVEN	NPDVCIIGID	RGERHLLYYT	
	VINQTDGILE	QGSLNKISGS	YTNDKGEKVN	KETDYHDLDD	RKEKGKHVAQ	
	QAWETIENIK	ELKAGYLSQV	VYKLTQLMLQ	YNAVIVLENL	NVGFKRGRTK	
	VEKQVYQKFE	KAMIDKLNLY	VFKDRGYEMN	GSYAKGLQLT	DKFESFDKIG	
	KQTGCIYYVI	PSYTSIDDPK	TGFVNLLNAK	LRYENITKAQ	DTIRKFDNIS	
	YNAKADYFEE	AFDYRSFGVD	MARNEWVVCT	CGDLRWEYSA	KTRETAKYSV	
	TDRLKELFKA	HGIDYVGGEN	LVSHITEVAD	KHFLSTLLFY	LRLVLKMRYT	
	VSGTENENDF	ILSPVEYAPG	KFFDSREATS	TEPMNADANG	AYHIALKGLM	
	TIRGIEDGKL	HNYGKGGENA	AWFKFMQNGE	YKNGG		
WP_044910712.1 type V CRISPR- associated protein Cpf1 [Lachnospiraceae bacterium MC2017]	MDYGNQGFER	RAPLTKTITL	RLKPIGETRE	TIREQKLLEQ	DAAFRKLVET	(SEQ ID NO: 155)
	VTPIVDDCIR	KIADNALCHF	GTEYDFSCLG	NAISKNDKA	IKKETEKVEK	
	LLAKVLTENL	PDGLRKVNDI	NSAAFIQDTL	TSFVQDDADK	RVLIQELKKG	
	TVLMQRFLTT	RITALTVWLP	DRVFENFNIF	IENAEKMRIL	LDSPLNEKIM	
	KFDPDAEQYA	SLEFYGQCLS	QKDIDSYNLI	ISGIYADDEV	KNPGINEIVK	
	EYNQQIRGDK	DESPLPKLKK	LHKQILMPVE	KAFFVRVLSN	DSDARSILEK	
	ILKDTMELPS	KIIEAMKEAD	AGDIAVYGSR	LHELSHVIYG	DHGKLSQIIY	
	DKESKRISL	METLSPKERK	ESKKRLEGLE	EHIRKSTYTF	DELNRYAEKN	
	VMAAYIAAVE	ESCAEIMRKE	KDLRTLSSKE	DVKIRGNRHN	TLIVKNYFNA	
	WTVFRNLIRI	LRRKSEAEID	SDFYDVLDD	VEVLSLTYKG	ENLCRSYITK	
	KIGSDLKPEI	ATYGSALRPN	SRWWSPEKGF	NVKFHTIVRR	DGRLYYFILP	
	KGAKPVELED	MDGDIECLQM	RKIPNPTIFL	PKLVFKDPEA	FFRDNPDEAD	
	FVFLSGMKAP	VTITRETYEA	YRYKLYTVGK	LRDGEVSEEE	YKRALLQVLT	
	AYKEFLENRM	IYADLNFGFK	DLEEKYDSSE	FIKQVETHNT	FMCWAKVSSS	
	QLDDLKSGN	GLLFEIWSER	LESYYKYGNE	KVLRGYEGVL	LSILKDENLV	
	SMRTLNSRP	MLVYRPKESS	KPMVVHRDGS	RVVDRFDKDG	KYIPPEVHDE	
	LYRFFNNLLI	KEKLGEKARK	ILDNKKVKVK	VLESERVKWS	KFYDEQFAVT	
	FSVKKNADCL	DTTKDLNAEV	MEQYSESRL	ILIRNTTDL	YYLVLDKNGK	
	VLKQSRSLNI	NDGARDVDWK	ERFRQVTKDR	NEGYNEWDS	RTSNDLKEVY	
	LNALKEIAE	AVIEYNAILI	IEKMSNAFKD	KYSFLDDVTF	KGFETKLLAK	
	LSDLHFRGIK	DGEPCSFTNP	LQLCQNDNSK	ILQDGVIFMV	PNSMTRSLDP	
	DTGFIFAIND	HNIRTKKAKL	NFLSKFDQLK	VSEEGCLIMK	YSGDSLPTHN	
	TDNRVWNCCC	NHPITNYDRE	TKKVEFIEEP	VEELSRVLEE	NGIETDTELN	
	KLNERENVPG	KVVDIAISLV	LNYLRGTVSG	VAGQRAVYYS	PVTGKKYDIS	
	FIQAMNLNRK	CDYYRIGSKE	RGEWTFVVAQ	LIN		
WP_081834226 type V CRISPR- associated protein Cpf1 [Lachnospiraceae bacterium MC2017].	MTMDYGNQG	ERRAPLTKTI	TLRLKPIGET	RETIREQKLL	EQDAAFRKLV	(SEQ ID NO: 156)
	ETVTPIVDDC	IRKIADNALC	HFGTEYDFSC	LGNAISKND	KAIKKETEKV	
	EKLLAKVLTE	NLPDGLRKVN	DINSAAFIQD	TLTSFVQDDA	DKRVLIQELK	
	GKTVLMQRFL	TTRITALTVW	LPDRVFENFN	IFIENAEKMR	ILLDSPLNEK	
	IMKFDPAEQ	YASLEFYGQC	LSQKDIDSYN	LIISGIYADD	EVKNPGINEI	
	VKEYNQIRG	DKDESPLPKL	KKLHKQILMP	VEKAFFVRVL	SNDSDARSIL	
	EKILKDTMEL	PSKIIIEAMKE	ADAGDIAVYG	SRLHELSHVI	YGDHGKLSQI	
	IYDKESKRIS	ELMETLSPKE	RKESKKRLEG	LEEHIRKSTY	TFDELNRYAE	
	KNVMAAYIAA	VEESCAEIMR	KEKDLRTLSS	KEDVKIRGNR	HNTLIVKNYF	
	NAWTVFRNLI	RILRRKSEAE	IDSDFYDVLD	DSVEVLSLTY	KGENLCRSYI	
	TKKIGSDLKP	EIATYGSALR	PNSRWWSPE	KFNVKFHTIV	RRDGRLYYFI	
	LPKGAKPVEL	EDMDGDIECL	QMRKIPNPTI	FLPKLVFKDP	EAFFRDNPEA	
	DEFVFLSGMK	APVTITRETY	EAYRYKLYTV	GKLDRGEVSE	EYKRALLOV	
	LTAYKEFLEN	RMİYADLNFG	FKDLEEKYDS	SEFIKQVETH	NTFMCWAKVS	
	SSQLDDLKVS	GNGLLFEIWS	ERLESYYKYG	NEKVLRGYEG	VLLSILKDEN	
	LVSMTLLNS	RPMLVYRPKE	SSKPMVVHRD	GSRVVDRFDK	DGKYIPPEVH	
	DELYRFFNNL	LIKEKLGEKA	RKILDNKKVK	VKLESERVK	WSKFYDEQFA	
	VTFSVKKNAD	CLDTTKDLNA	EVMEQYSES	RLILIRNTTD	ILYYLVLDKN	
	GKVLKQSRSL	IINDGARDVD	WKEFRQVTK	DRNEGYNEW	YSRTSNDLKE	
	VYLNALKEI	AEAVIEYNAI	LIIEKMSNAF	KDKYSFLDDV	TFKGFETKLL	
	AKLSDLHFRG	IKDGEPCSFT	NPLQLCQND	NKILQDGVIF	MVPNSMTRSL	
	DPDTGFIFAI	NDHNIRTKKA	KLNFLSKFDQ	LKVSSEGCIL	MKYSGDSLPT	

Table 5 - Cas12a orthologs						
	HNTDNRVWNC	CCNHPITNYD	RETKKVEFIE	EPVEELSRVL	EENGIEDTDE	
	LNKLNREENV	PGKVVDIAYS	LVLNLYRGTV	SGVAGQRAVY	YSPVTGKKYD	
	ISFIQAMNLN	RKCDYYRIGS	KERGEWTDVF	AQLIN		
WP_027216152.1 type V CRISPR- associated protein Cpf1 [Butyrivibrio fibrisolvens]	MYYESLTCLY	PIKKTIRNEL	VPIGKTLENI	KKNNILEADE	DRKIAYIRVK	(SEQ ID NO: 157)
	AIMDDYHKRL	INEALSGFAL	IDLDKAANLY	LSRSKSADDI	BSFSRFQDKL	
	RKAIKRLRE	HENFGKIGNK	DIIPLLQKLS	ENEDDYNALE	SFKNFYTYFE	
	SYNDVRLNLY	SDKEKSSTVA	YRLINENLPR	FLDNIRAYDA	VQKAGITSEE	
	LSSEAQDGLF	LVNTFNNVLI	QDGINTYNED	IGKLNVAINL	YNQKNASVQG	
	FRKVPKMKVL	YKQILSDREE	SFIDEFESDT	ELLDSLESHY	ANLAKYFGSN	
	KVQLLFTALR	ESKGVNVVVK	NDIAKTSFSN	VVFGSWSRID	ELINGEYDDN	
	NNRKKDEKYY	DKRQKELKKN	KSYTIEKIIIT	LSTEDVDVIG	KYIEKLES DI	
	DDIRFKGKNF	YEAVLCGHDR	SKKLSKNKGA	VEAIKGYLDS	VKDFERDLKL	
	INGSGQELEK	NLVVYGEQEA	VLSLSGIDS	LYNMTRNYLT	KKPFSTEKIK	
	LNFNKPTFLD	GWDYGNEEAY	LGFFMIKEGN	YFLAVMDANW	NKEFRNIPSV	
	DKSDCYKKVI	YKQISSPEKS	IQNLMLVIDGK	TVKKNGRKEK	EGIHSGENLI	
	LEELKNTYLP	KKINDIRKRR	SYLNGDTFSK	KDLTEFIGYY	KQRVIEYYNG	
	YSFYFKSDDD	YASFKEFQED	VGRQAYQISY	VDPVPSFVDD	LINSGLYLFL	
	RVYNKDFSEY	SKGRNLNLHTL	YFKMLFDERN	LKNVVYKLNQ	QAEVFYRPSS	
	IKKEELIVHR	AGEEIKNKNP	KRAAQKPTRR	LDYDIVKDRR	YSQDKFMLHT	
	SIIMNFGAEE	NVSFNIDVNG	VLNRNEDKVV	IGIDRGERNL	LYVVVIDPEG	
	KILEQRSLNC	ITDSNLDIET	DYHRLLEKE	SDRKIARRDW	TTIENIKELK	
	AGYLSQVVHI	VAELVLKYNA	IICLEDLNFQ	FKRGRQKVEK	QVYQKFEKML	
	IDKLNLYVMD	KSREQLSPEK	ISGALNALQL	TPDFKSFKVL	GKQTGIIYYV	
	PAYLTSKIDP	MTGFANLFYV	KYENVDKAKE	FFSKFDSIKY	NKDGKNWNTK	
	GYFEFAFDYK	KFTDRAYGRV	SEWTVCTVGE	RIIKFKNKEK	NNSYDDKVID	
	LTNSLKELFD	SYKVTYESEV	DLKDAILAID	DPAFYRDLTR	RLQQTQLMRN	
	SSCDGSRDYI	ISPVKNSKGE	FFCSDNNDT	TPNDADANGA	FNIARKGLWV	
	LNEIRNSEEG	SKINLAMNSA	QWLEYAQDNT	I		
WP_016301126.1 type V CRISPR- associated protein Cpf1 [Lachnospiraceae bacterium COE1]	MHENNGKIAD	NFIGIYPVSK	TLRFELKPVG	KTQEYIEKHG	ILDEDLKRAG	(SEQ ID NO: 158)
	DYKSVKKIID	AYHKYFIDEA	LNGIQLDGLK	NYELEYEKKR	DNNEKEBFQK	
	IQMSLRKQIV	KRFSEHPQYK	YLPKKELIKN	VLPEFTKDNA	EEQTLVKSQFQ	
	EFTTYFEGFH	QNRKNMYSDE	EKSTAIAYRV	VHQNLPKYID	NMRIFSMILN	
	TDIRSDLTEL	FNNLKTMDI	TIVEEYFAID	GFNKVVNQKG	IDVYNTILGA	
	FSTDDNTKIK	GLNEYINLYN	QKNKAKLPKL	KPLFKQILSD	RDKISFIEPQ	
	FDSDETEVLEA	VDMFYNRLLQ	FVIENEGQIT	ISKLLTNFSA	YDLNKIYVKN	
	DTTISAISND	LFDDWSYISK	AVRENYDSEN	VDKNKRAAAY	EKKKEKALSK	
	IKMYSIEELN	FFVKKYSCNE	CHIEGYFERR	ILEILDKMRY	AYESCKILHD	
	KGLINNISLC	QDRQAISELK	DFLDSIKEVQ	WLLKPLMIGQ	EQADKEEAFY	
	TELLRIWEEL	EPITLLYNKV	RNYVTCKPYT	LEKVKLNFYK	STLLDGWDKN	
	KEKDNLGIIL	LKDQGYLGI	MNRRNNKIAD	DAPLAKTDNV	YRKMEYKLLT	
	KVSANLPRIF	LKDYNPSEE	MLEKYEKGTH	LKGENFCIDD	CRELIDFFKK	
	GKQYEDWGQ	FDKFSDES	YDDISAFYKE	VEHQGYKITF	RDIDETYIDS	
	LVNEGKLYLF	QIYNKDFSPY	SKGTKNLHTL	YWEMLFSQQN	LQNIVYKLNQ	
	NAEIFYRKAS	INQKDVVVK	ADLPIKNKDP	QNSKKESMFD	YDIKDKRFT	
	CDKYQFHVPI	TMNFKALGEN	HFNRKVNRLI	HDAENMHIIG	IDRGERNLIY	
	LCMIDMKGNI	VKQISLNEII	SYDKNKLEHK	RNYHQLLKTR	EDENKSARQS	
	WQTIHTIKEL	KEGYLSQVIH	VITDLMVEYN	AIVVLEDLNF	GFKQGRQKFE	
	RQVYQKFEKM	LIDKLNLYLD	KSKGMDDEDG	LLHAYQLTDE	FKSFKQLGKQ	
	SGFLYYIPAW	NTSKLDPTTG	FVNLFYTKYE	SVEKSKEFIN	NFTSILYNQE	
	REYFEFLFDY	SAFTSKAEGS	RLKWTVCSSK	ERVETYNPK	KNNEWDTQKI	
	DLTFELKKLF	NDYSISLLDG	DLREQMGKID	KADFYKKFMK	LFALIVQMRN	
	SDEREDKLIS	PVLNKYGAF	ETGKNRMPL	DADANGAYNI	ARKGLWIEK	
	IKNTDVEQLD	KVKLTISNKE	WLQYAEHIL			
WP_035635841.1 type V CRISPR- associated protein Cpf1	MSKLEKFTNC	YLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	(SEQ ID NO: 159)
	KKLLDRYYS	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	
	LRKEIAKAFK	GNEGYSKLFK	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	
	FTGFFDNREN	MFSEEAKSTS	IAFCINENL	TRYISNMDIF	EKVDAIFDKH	
	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	IGGFVTSGE	
	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRRESLSFY	GEGYTSDEEV	

Table 5 - Cas12a orthologs						
[Lachnospiraceae bacterium ND2006]	LEVFRNTLNK IFGEWNVIRD QEYADADLSV AVVAIMKDLL DHIYDAIRNY SKYYLAIMDK KWMAYYNPSE NAYDFNFSET MFQIYNKDFS LKKEELVVHP AINKCPKNIF VEQYSLNEII AGYISQVVHK IDKLYNMVVK TSKIDPSTGF NFSRTDADYI KYGINYQQGD ISPVKNSDGI AEDEKLDKVK	NSEIFSSIKK KWNAEYDDIH VEKLKEIIIIQ DSVKSFENYI VTQKPYSKDK KYAKCLQKID DIQKIYKNGT EKYKDIAGFY DKSHGTPNLH ANSPIANKNP KINTEVRVLL NNFNGIRIKT ICELVEKYDA KSNPCATGGA VNLLKTKYTS KKWKLYSYGN IRALLCEQSD FYDSRNYEAG IAISNKEWLE	LEKLFKNFDE LKKKAVVTEK KVDEIYKVYK KAFEGEGKET FKLYFQNPQF KDDVNGNYEK FKKGDMPNLN REVVEQGYKV TMYFKLLFDE DNPCKTTTSL KHDDNPYVIG DYHSLLDKKE VIALEDLNSG LKGQITNKF IADSKKFIS RIRIFRNPCK KAFYSSFMAL ENAILPKNAD YAQTSVKH	YSSAGIFVKN YEDDRRKSFK SSEKLFADDF NRDESFGYGF MGWWDKDKET INYKLLPGPN DCHKLIDFFK NNHGQIRLSG YDVYKDKRFS IDRGERNLLY KERFEARQNW FKNSRVKVEK ESFKSMSTQN FDRIMYVPEE NNVFDWEEVC MSLMLQMRNS ANGAYNIARK	GPAISTISKD KIGSFSLEQL VLEKSLKKND VLAYDILLKV DYRATILRYG KMLPKVFFSK DSISRYPKWS DKLVEEGKLY GAELFMRRAS EDQYELHIP IVVVDGKGN TSIENIKELK QVYQKFEKML GFIFYIPAWL DLFEFALDYK LTSAYKELFN ITGRTDVDFL VLWAIQGFKK	
WP_051666128.1 type V CRISPR- associated protein Cpf1 [Lachnospiraceae bacterium ND2006]	MLKNVGIDRL NKRLLEVEDEK KKTRTEKENK DDKDEIALVN YISNMDIFEK GIDVYNAIIG DRESLSFYGE SAGIFVKNGP DDRRKSFKKI EKLFDADFVL DESFYGDFVL GWDKDKETDY YKLLPGPNKM HKLIDFFKDS ESASKKEVDK HGQIRLSGGA VYKDKRFSED RGERNLLYIV RFEARQNWTS NSRVKVEKQV FKSMSTQNGF RIMYVPEEDL VFDWEEVCLT LMLQMRNSIT GAYNIARKVL	DVEKGRKNMS RAEDYKGVKK ELENLEINLR SFNGFTTAFT VDAIFDKHEV GFVTESGEKI GYTSDEEVLE AISTISKDIF GSFSLQLQE EKSLKKNDV AYDILLKVDH RATILRYGSK LPKVFFSKKW ISRYPKWSNA LVEEGKLYMF ELFMRRASLK QYELHIPAI VVDGKNIVE YISQVVKIC YQKFEKMLID IFYIPAWLTS FEFALDYKNF SAYKELFNKY GRTDVDFLIS WAIGQFKKAE	KLEKFTNCYS LLDRYLSFI KEIAKAFKGN GFFDNRENMF QEIKEKILNS KGLNEYINLY VFRNTLNKNS GEWNVIRDKW YADADLSVVE VAIMKDLLDS IYDAIRNYVT YYLAIMDKKY MAYYNPSEDI YDFNFSETEK QIYNKDFSDK KEELVVHPAN NKCPKNIFKI QYSLNEIINN YISQVVKIC KLYNMVDKKS KIDPSTGFVN SRTDADYIKK GINYQQGDIR PVKNSDGIFY DEKLDKVKIA	LSKTLRFKAI NDVLHSIKLK EGYKSLFKKD SEBAKSTSLA DYDVEDFFEG NQKTKQKLPK EIFSSIKKLE NAEYDDIHLK KLKEIIIIQKV VKSFENYIKA QKPYSKDKFK AKCLQKIDKD QKIYKNGTFK YKDIAGFYRE SHGTPNLHTM NTEVRVLLKH FNGIRIKTDY ELVEKYDAVI NPCATGGALK LLKTKYTSIA WKLVSYGNRI ALLCEQSDKA DSRNYEAQEN ISNKEWLEYA	PVGKTQENID NLNNYISLFR IETILPEFL FRCINENLTR EFFNFVLTQE FKPLYKQVLS KLKFNPEYS KKAVVTEKYE DEIYKVYGSS FFGEGKETNR LYFQNPQFMG DVNGNYEKIN KGDMFNLNDC VEEQGYKVSF YFKLLFDENN DDNPYVIGID HSLLDKKEKE ALEDLNSGFK GYQITNKFES DSKKFISFSD RIFRNPCKNN FYSSFMALMS AILPKNADAN QTSVKH	(SEQ ID NO: 160)
WP_015504779.1 type V CRISPR- associated protein Cpf1 [Candidatus Methanomethylophi lus alvus]	MDAKEFTGQY ELDDNHRAF KRRKEISAYL SVYFTGYHES DIISEVSGNL QAFNVVLNLR DCICDYVSKI DWDAIETALM NRAEDICRVI FHELEIFSVG TDKIKVNLKF RTSDEDESSF KGMHKSASF FNEDVAGAGY MHTMYWGLF	PLSKTLRFEL LNRVLPQIDM QDADGYKGLF RENIYSDEDM GVDDIGKYFD HQKDPGFEEK EKSETVERAL HSSSENDDKK SETAPFINDL DEFKCAAFY PTLADGWDLN EKMEYKLLPS DLGFCHLID YMSLRKIPCS SPQNLESPVF	RPIGRTWDNL DWHPIAEAF AKPALDEAMK VSVAYRITED VSNYNNFLSQ QFKQLYKQIL KLVRNISSFD SVYDSAEAF RAVDLDSLND SELEEVSEQL KERDNKAAIL PVKMLPKIFV YYKRCIAEYP EYVRLLEKES KLSGGAELFF	EASGYLAEDR KVHKNPGNKE IAKENGNESED NFPRFVSNAL AGIDDYNHII SVRTSKSYIP LRGIFVNKK LDDIFSSVKK DGEEAAVSKI IBIIPLFNKA RKDGKYLLAI KSKAAKEKYG GWDVDFDKFR IYLFQIYNKD RKSSIPNDK	HRAECYPRAK LAQDYNLQLS IEVLEAFNGF IFDKLNESH GGHTTEDGLI KQFDNSKEMV LRILSNKLIG FSDASAEDIG RESLEPYMDL RSFCTRKRYS LDMKKDLSSI LTDRLMECYD ETSDYGSMEKE YSENAHGNKN TVHPKGSVLV	(SEQ ID NO: 161)

Table 5 - Cas12a orthologs						
	PRNDVNGRRI	PDSIYRELTR	YFNRGDCRIS	DEAKSYLDKV	KTKKADHDIV	
	KDRRFTVDM	MFHVPIAMNF	KATSKPNLNK	KVIDGIIDDQ	DLKIIIGIDRG	
	ERNLIYVTMV	DRKGNILYQD	SLNILNGYDY	RKALDVREYD	NKEARNNWT	
	VEGIRKMEG	YLSLAVSKLA	DMIENNAII	VMEDLNHGFK	AGRSKIEKQV	
	YQKFESMLIN	KLGYMVLKDK	SIDQSGGALH	GYQLANHVTT	LASVGKQCGV	
	IFYIPAAFTS	KIDPTTGAFD	LFALSNVKNV	ASMREFFSKM	KSVIYDKAEG	
	KFAFTFDYLD	YNVNSECGRT	LWTVYTVGER	FTYSRVNREY	VRKVPDIIY	
	DALQKAGISV	EGDLRDRIAE	SDGDTLKSIF	YAFKYALDMR	VENREEDYIQ	
	SPVKNASGEF	FCSKNAGKSL	PQDSDANGAY	NIALKGILQL	RMLSEQYDPN	
	AESIRLPLIT	NKAWLTFMQS	GKMTWKN			
WP_044910713.1 type V CRISPR- associated protein Cpf1 [Lachnospiraceae bacterium MC2017]	MGLYDGFVNR	YSVSKTLRFE	LIPQGRTRREY	IETNGILSDD	EERAKDYKTI	(SEQ ID NO: 162)
	KRLIDEYHKD	YISRCLKNVN	ISCLEEYYHL	YNSSNRDKRH	EELDALSQDM	
	RGEIASFLTG	NDEYKEQKSR	DIIINERIIN	FASTDEELAA	VKRFRKFTSY	
	FTGFFTNREN	MYSAEKKSTA	IAHRIIDVNL	PKYVDNIKAF	NTAIEAGVFD	
	IAEFESNFKA	ITDEHEVSDL	LDITKYSRFI	RNEDIIIIYNT	LLGGISMKDE	
	KIQGLNELIN	LHNQKHPGKK	VPLLKVLYKQ	ILGDSQTHSF	VDDQFEDDQ	
	VINAVKAVTD	TFSETLLGSL	KIIINNIGHY	DLDRYIYKAG	QDITTLKRA	
	LNDWHIITEC	LESEYDDKFP	KNKKSPTYEE	MRNRYVKSFK	SFSIGRLNSL	
	VTTYTEQACF	LENYLGSGFG	DTDNCLTDF	TNSLMEVEHL	LNSEYPVTNR	
	LITDYESVRI	LKRLDSEME	VIHFLKPLL	NGNESDKDLV	FYGEFEAEYE	
	KLLPVIKVYN	RVRNYLTRKP	FSTEKIKLNF	NSPTLLCGWS	QSKEKEYMGV	
	ILRKDGQYYL	GIMTPSNKKI	FSEAPKPED	CYEKMLRYI	PHPYQMLPKV	
	FFSKSNIAFF	NPSDEILRIK	KQESFKKGKS	FNRDDCHKFI	DFYKDSINRH	
	EEWRKFNFKF	SDTDSYEDIS	RFYKEVENQA	FSMSFTKIPT	VYIDSLVDEG	
	KLYLFKLHNK	DFSEHSKGKP	NLHTVYWNAL	FSEYNLQNTV	YQLNGSABIF	
	FRKASIPENE	RVIHKKNVPI	TRKVAELNGK	KEVSVPFYDI	IKNRRYTVDK	
	QFQHVPLKMN	FKADEKKRIN	DDVIEAIRSN	KGIHVIDR	GERNLLYLSL	
	INEEGRIIEQ	RSLNIDSSE	GHTQNYRDL	DSREKREKA	RENWQEIQEI	
	KDLKTGYLSQ	AIHTITKWMK	EYNAIIVLED	LNDRTNGRK	KVEKQVYQKF	
	EKMLIDKLN	YVDKDEEFDR	MGGTHRALQL	TEKFESFQKL	GRQTGFIFYV	
	PAWNTSKLDP	TTGFDLLYP	KYKSVDATKD	FIKKFDFIRF	NSEKNYFEFG	
	LHYSNFTERA	IGCRDEWILC	SYGNRIVNFR	NAAKNNSWDY	KEIDITKQLL	
	DLFEKNGIDV	KQENLIDSIC	EMKDKPFFKS	LIANIKLILQ	IRNSASGTDI	
	DYMISPAMND	RGEFFDTRKG	LQQLPLDADA	NGAYNIAKKG	LWIVDQIRNT	
	TGNNVKMAMS	NREWMHFAQE	SRLA			
KKQ36153.1 hypothetical protein US52_C0007G0008 [candidate division WS6 bacterium GW2011_GWA2_3 7_6]	MKNVFGGFTN	LYSLTKTLRF	ELKPTSKTQK	LMKRNNVIQT	DEEIDKLYHD	(SEQ ID NO: 163)
	EMKPILDEIH	RRFINDALAQ	KIFISASLDN	FLKVVKNYKV	ESAKKNIKQN	
	QVKLLQKEIT	IKTLGLRREV	VSGFITVSCK	WKDKYVGLGI	KLKGDGYKVL	
	TEQAVLDILK	IEFPNKAKEY	DKFRGFWTYF	SGFNENRKNY	YSEEDKATSI	
	ANRIVENLS	RYIDNIIAFE	EILQKIPNLK	KFKQDLIDTS	YNYLNLQAGI	
	DKYNKIIGGY	IVDKDKKIQQ	INEKVNLYTQ	QTKKKLPKLG	FLFKQIGSER	
	KGFGIFEIKE	GKEWEQLGDL	FKLQRTKINS	NGREKGLFDS	LRTMYREFFD	
	EIKRDSNSQA	RYSLDKIYFN	KASVNTISNS	WFTNWNKFAE	LLNIKEDKKN	
	GEKKIPEQIS	IEDIKDSLIS	IPKENLEELF	KLTNREKHDR	TRFFGNSAWV	
	TFLNIWQNEI	EESFNKLEEK	EKDFKKNAAI	KFQKNNLVQK	NYIKEVCDRM	
	LAIERMAKYH	LPKDSNLSRE	EDFYWIIDNL	SEQREIYKYV	NAFRNYISKK	
	PYNKSKMKLN	FENGNNLGGW	SDGQERNKAG	VILRNGNKYY	LGVLINRGIF	
	RTDKINNEIY	RTGSSKWERL	ILSNLKFQTL	AGKGLGKHG	VSYGNMNPKE	
	SVPSLQKFIR	ENYLKKYPQL	TEVSNTKFLS	KKDFDAAIKE	ALKECFMTNF	
	INIAENKLL	AEDKGDLYLF	EITNKDFSGK	KSGKDNHITI	YWKYLFSES	
	CKSPIIGLNG	GAEIFFREGQ	KDKLHTKLDK	KGKKVFDAGR	YSEDKLFHFV	
	SITINYGKPK	NIKFRDIINQ	LITSMNVNII	GIDRGEKHL	YYSVIDSNGI	
	ILKQGSNLKI	RVGDKEVDFN	KKLTERANEM	KKARQSWEQI	GNIKNFKEGY	
	LSQAIHEIYQ	LMIKYNAIIV	LEDLNTEFKA	KRLSKVEKSV	YKKFELKLAR	
	KLNLHLKDR	NTNEIGGVLK	AYQLTPTIGG	GDVSKFEKAK	QWGMFVYRA	
	NYTSTTDPVT	GWRKHLYISN	FSNNSVIKSF	FDPTNRDTGI	EIFYSGKYRS	
	WGFYVQKET	GKKWELFATK	ELERFKYNQT	TKLCEKINLY	DKFEELFKGI	
	DKSADIYSQL	CNVLDPRWKS	LVYLVNLLNQ	IRNVDKNAEG	NKNDFIQSPV	
	YPFFDSRKT	GKTEPINGDA	NGALNIARKG	LMLVERIKNN	PEKYEQLIRD	

Table 5 - Cas12a orthologs						
	TEWDAWIQNF	NKVN				
WP_044919442.1 type V CRISPR- associated protein Cpf1 [Lachnospiraceae bacterium MA2020]	MYYESLTKQY GILDEYHKQL RKEVVEKLKA TSYNKVRENL GLGEEQDSL GFRKIPKMKM SKVSAFFDAL ANENKKDDK DIENIIINNE KLINSSGQEL VKLNFNSTL VAKTEKVFCK HKKGNMFSLE EVEKQGYKLT LYFMMLFDQR PNRARTKETS SAIRIDENVN TDYHALDER AIIICLEDLNF ELGGALNALQ MKCENVEKSK CTNGERIIKY IISNEAEFY LSDGSVPKDA AQTHLL	PVSKTIRNEL INEALDNCTL HENFTKIGKK YSDKEKSSTV FIVETFNKTL LYKQILSDRE RESKGNVYV YFEKRQKELK TFLRIVINEH EKDLIVYSAH LNGWDRNKET VDYKLLPVPN DCHNLIDFFK YTDIDETYIN NIDDVYKLN TFSYDIVKDK VIGIDRGERN EGGRDKARKD GFKRGRQKVE LTSKFKSFKE RFFDGFDFIR RNPDKNMFD RRLYRLLQQT DANGAYNIAR	IPIGKTLDNI PSLKIAAEIY DILDILLEKLP AYRLINENFP TQDGIDTYS ESFIDEFQSD KNDLAKTAMS KNKSYSLEHL DRSRKLAKNR EELLVELKQV DNLGVLLKLD QMLPKVFFAK ESISKHEDWS DLIERNELYL GEAEVFYRPA RYSKDKFTLH LLYVVVIDSK WNTVENIRDL KQVYQKFEKM LGKQSGVIYY FNALENVFEF EKVVVVTDEM LQMRNSTSDG KGLWVLEQIR	RQNNILES DV LKNQKEVSDR SISEDDYNAL KFLDNVKS YR QVGKINSSIN EVLIDNVESY NIVFENWRTF CNLS EDS CNL KAVKAIKDFL DSL YNMTRNY GKY YLGIMNT SNIDFYNPSS KFGFKFS DTA FQIYNKDFSM SISEDELIH IPITMNF GVD GNILEQISLN KAGYLSQV V LIDKLNLYVI VPAYLT SKID GFDYRSFTQR KNLFEQYKIP TRDYIISP VK QKSEGEKINL	KRKQNYEHVK EDFNKTQDLL ESFRNFYTYF FVKTAGILAD LYNQKNQKAN GSVLIESLKS DDL NQ EYDL IENYIHQISD DSIKVLEREL LTKKPFST EK SANKAFVNPP BIYSNYKKGT SYNDISEFYR YSKGKLNLHT KAGEEIKNKN EVRGRNVKDM SIINKEYDIE VVAKLVLKYN DKSREQTSPK PTTG FANLFY ACGINSKWTV YEDGRNVKDM NKREAYFNSE AMTNAEWLEY	(SEQ ID NO: 164)
WP_035798880.1 type V CRISPR- associated protein Cpf1 [Butyrivibrio sp. NC3005]	MYYQNLTKKY GIMDEYHKQL RREVTGRLKE TSYNKVRENL CIEEEEQDAL EFKKIPKMKV EKINIFFDAL ANENKKDDK DIEKICIYNG KLINGSGQEL VKLNFNKSTL AAKTENVFKK HKKGPSFSID EVEKQGYKLT LYFMMLFDQR PNRAKVKETS NALRTDDNVN TNYHALDER AIIICLEDLNF KMGGALNALQ IKYENIEKAK YTNGERIIKY IISKAADFY FSEGTMPKDA AQLHLL	PVSKTIRNEL INEALDNYML HENYTKIGKK YSDEEKSSTV FMVETFNMTL LYKQILSDRE RESEGNVYV YFEKRQKELK EFEKIVVNEH EKNLVVYVGQ LNGWDKNKET VDYKLLPGSN DCHNLIDFFK FTDIDESYIN NLDNVVYKLN TFSYDIVKDK VIGIDRGERN EDDRNKARKD GFKRGRQKVE LTSKFKSFAE QFFDGFDFIR PNPEKNLFD KRLFRLLHQT DANGAYNIAR	IPIGKTLENI PSLNQAAEY DILDILLEKLP AYRLINENLP TQEGIDMYNY EVFIGEFKDD KNDLSKTTMS KNKSYTLEQM DSSRKLSKNI EEALEQLRPV DNLGILFFKD KMLPKVFFAK ESIKKHEDWS DLIEKNELYL GEAEVFYRPA RYSKYKFTLH LLYVVVINSE WNTIENIKEL KQVYQKFEKM LGKQSGIYY FNKKDDMF EF EKVINVTDEI LQMRNSTSDG KGLWVLEQIR	RKNNILES DV LKKHVDVEDR SISEEDYNAL KFLDNIKSYA QIGKVNSAIN ETLLSSIGAY NIVFGSWSAF SNLSKEDISP KAVKVIKDYL DSL YNLTRNY GKY YLGIMNT SNIGYYPST KFGFEFS DTA FQIYNKDFSE SIAENELVIH IPITMNF GVD GKILEQISLN KTGYLSQV V LIEKLNLYVI VPAYLT SKID SFDYKSFTQK KGLFKQYRIP TRDYIISP VK QKDEGEKVNL	KRKQDYEHVK BEFKKTQDLL ESFRNFYTYF FVKAAGVLAD LYNQKNHKVE GNVLMTYLKS DELLNQEYDL IENYIERISE DSIKELEHDI LTKKPFST EK TANKAFVNPP ELYSNYKKGT DYRDISEFYR YSKGKLNLHT KAGEGINKNKN EVRFRNDVIN SIINKEYDIE VVAKLVLKYN DKSREQVSPE PTTG FVNLFY ACGIRSKWIV YENGEDIKEI NDRGEFFCSE SMTNAEWLKY	(SEQ ID NO: 165)
WP_027109509.1 type V CRISPR- associated protein Cpf1 [Lachnospiraceae bacterium NC2008]	MENYDLSLTR VKELMDEFHK YMRKQIADAL YFTDYKNRK NDNYDEETLD EENFVRVPKM LPINIDRYIK YDAKYTGKVN	QYPVTKTIRQ SIEKSLVGI KNSRDYGYVD NIYSDEEKHS KTFMIESFNE KVLFKQILSD TLMDSNNGTG DKYEEKREKA	ELKPVGKTLE KLDGLSEFEK NKDLIEKILP TVGYRCINEN CLTQSGVEAY RTSLFDGLII IYVKNDSSLT YKSNDSFELN	NIKNAEII EA LYKIKTKTDE ERVKDENS LN LLIFMSNIEV NSVVASIKTA ESDDEL LDKL TLSNYLTDSW YIQNLLGINV	DKQKKEAYVK DKNRISELFY ALSCFKGFTT YQIYKKANIK TNLYIQKNNK CSFSAEVDKF SSIRNAFNEN IDKYIERINF	(SEQ ID NO: 166)

Table 5 - Cas12a orthologs						
	DIKEICEAYK	EMTKNCFEDH	DKTKKLQKNI	KAVASIKSYL	DSLKNIERDI	
	KLLNGTGLES	RNEFFYGEQS	TVLEBITKVD	ELYNITRNYL	TKKPFSTCKM	
	KLNFNPNQLL	GGWDVNKERD	CYGVILIKDN	NYYLIGIMDKS	ANKSFLNIKE	
	SKNENAYKKV	NCKLLPGPNK	MFPKVFFAKS	NIDYYDPTHE	IKKLYDKGTF	
	KKGNSFNLED	CHKLIDFYKE	SIKKNDWKN	FNFNFSDTKD	YEDISGFFRE	
	VEAQNYKITY	TNVSCDFIES	LVDEGKLYLF	QIYNKDFSEY	ATGNLNLHTL	
	YLMKLFDERN	LKDLCIKMNG	EAEVFYRPAS	ILDEDKVVHK	ANQKITNKNT	
	NSKKKESIFS	YDIVKDKRYT	VDKFFIHLPI	TLNYKEQNVS	RFNDYIREIL	
	KKSKNIRVIG	IDRGERNLLY	VVCDSDGSI	LYQRSINEIV	SGSHKTDYHK	
	LLDNKEKERL	SSRRDWKTIE	NIKDLKAGYM	SQVVNEIYNL	ILKYNAIVVL	
	EDLNIGFKNG	RKKVEKQVYQ	NFEKALIDKL	NYLCIDKTRE	QLSPSSPGGV	
	LNAYQLTAKF	ESFEKIGKQT	GCIFYVPAYL	TSQIDPTTGF	VNLFYQKDT	
	KQGLQLFFRK	FKKINFDKVA	SNFEFVFDYN	DFTNKAEGTK	TNWTISTQGT	
	RIAKYRSDDA	NGKWSRTVH	PTDIIKEALN	REKINYNDGH	DLIDEIVSIE	
	KSAVLKEIYY	GFKLTLQLRN	STLANEEQE	DYIISPVKNS	SGNYFDSRIT	
	SKELPCDADA	NGAYNIARKG	LWALEQIRNS	ENVSKVKLAI	SNKEWFYEQ	
	NNIPSL					
WP_049895985.1 type V CRISPR- associated protein Cpf1 [Oribacterium sp. NK2B42] WP_029202018	METEILKYDF	FEREGKMYMY	DGLTKQYALS	KTIRNELVPI	GKTLDNICKN	(SEQ ID NO: 167)
	RILEADIKRK	SDYEHVKKLM	DMYHKKIINE	ALDNFKLSVL	EDAADIYFNK	
	QNDERDIDAF	LKIQDKLRKE	IVEQLKGHTD	YSKVGNKDFL	GLLKAASTEE	
	DRILIESFDN	FYTYFTSYNK	VRNLYSAED	KSSTVAYRLI	NENLPKFFDN	
	IKAYRTVRNA	GVISGDMISV	EQDELFEVDI	FNHTLTQYGI	DTYNHMIGQL	
	NSAINLYNQK	MHGAGSFKKL	PKMKELYKQL	LTEREEEFIE	EYTDDEVILIT	
	SVHNYVSYLI	DYLSNDKVES	FFDTRLKSDG	KEVFIKNDVS	KTTMSNILFD	
	NWSTIDDLIN	HEYDSAPENV	KKTKDDKYFE	KRQKDLKKNK	SYSLSKIAAL	
	CRDTTILEKY	IRRLVDDIEK	IYTSNNVFS	IVLSKHDRSK	KLSKNTNAVQ	
	AIKNMLDSIK	DFEHDVMLIN	GSGQEIKKNL	NVYSEQEALA	GILRQVDHIY	
	NLTRNYLTKK	PFSTEKIKLN	FNRPTFLDGW	DNKKEEANLG	ILLIKDNRYK	
	LGIMNTSSNK	AFVNPPKAIS	NDIYKKVDYK	LLPGPNKMPL	KVFFATKNIA	
	YYAPSEELLS	KYRKGTTHKK	DSFSIDDCRN	LIDFFKSSIN	KNTDWSTFGF	
	NFSDTNSYND	ISDFYREVEK	QGYKLSFTDI	DACYIKDLVD	NNELYLFQIY	
	NKDFSPYSKG	KLNHLTYLYFK	MLFDQRNLDN	VVYKLNGEAE	VFYRPASIES	
	DEQIIHKSGQ	NIKNKNQKRS	NCKKTSTFDY	DIVKDRRYCK	DKFMLHLPI	
	VNFGTNEGSK	FNELVNNAIR	ADKDVNIGI	DRGERNLLYV	VVVDPCGKII	
	EQISLNTIVD	KEYDIETDYH	QLLDEKEGSR	DKARKDWNTI	ENIKELKEGY	
	LSQVVNIIAK	LVLKYDAIIC	LEDLNFQFKR	GRQKVEKQVY	QKFEKMLIDK	
	MNYLVLDKSR	KQESPQKPGG	ALNALQLTSA	FKSFKEKLGQ	TGIIYYVPAY	
	LTSKIDPTTG	FANLFYIKYE	SVDKARDFFS	KFDFIRYNQM	DNYFEFGFDY	
	KSFTERASGC	KSKWIACTNG	ERIVKYRNSD	KNNSFDDKT	ILTDEYRSLF	
	DKYLQNYIDE	DDLKDQILQI	DSADFYKNLI	KLFQLTLQMR	NSSSDGKRKY	
	IISPVKNYRE	EFFCSEFSDD	TFPRDADANG	AYNIARKGLW	VIKQIRETKS	
	GTKINLAMSN	SEWLEYAQCN	LL			
WP_028248456.1 type V CRISPR- associated protein Cpf1 [Pseudobutyrvibrio ruminis]	MYYQNLTKMY	PISKTLRNEL	IPVGKTLENI	RKNGILEADI	QRKADYEHVK	(SEQ ID NO: 168)
	KLMDNYHKQL	INEALQGVHL	SDLSDAYDLY	FNLSKEKNSV	DAFSKCQDKL	
	RKEIVSLLKN	HENFPKIGNK	EIIKLLQSLY	DNDTDYKALD	SFSNFYTYFS	
	SYNEVRKNLY	SDEEKSTVA	YRLINENLPK	FLDNIAKAYI	AKKAGVRAEG	
	LSEEDQDCLF	IIETTFERTLT	QDGIDNYNAA	IGKLNTAINL	FNQQNKKQEG	
	FRKVPQMKCL	YKQILSDREE	AFIDEFSDD	DLITNIESFA	ENMNVLNSE	
	IITDFKIALV	ESDGLSVYIK	NDVSKTSFSN	IVFGSWNAID	EKLSDEYDLA	
	NSKKKKDEKY	YEKRQKELKK	NKSYDLETII	GLFDDNSDVI	GKYIEKLES	
	ITAI AEAKND	FDEIVLRKHD	KNKSLRKNTN	AVEAIIKSYLD	TVKDFERDIK	
	LINGSGQEVE	KNLVVYAEQE	NILAEIKNVD	SLYNMSRNYL	TQKPFSTEKF	
	KLNFRNATLL	NGWDKNKETD	NLGILFEKDG	MYYLIGIMNTK	ANKIFVNIPK	
	ATSNDVYHKV	NYKLLPGPNK	MLPKVFFAQS	NLDYYKPSEE	LLAKYKAGTH	
	KKGDNFSLED	CHALIDFFKA	SIEKHPDWSS	FGFEFSETCT	YEDLSGFYRE	
	VEKQGYKITY	TDVDADYITS	LVERDELYLF	QIYNKDFSPY	SKGNLNLHTI	
	YLQMLFDQRN	LNNVVYKLNG	EAEVFYRPAS	INDEEVIHK	AGEEIKNKNS	
	KRAVDKPTSK	FGYDIKDRR	YSKDKFMLHI	PVTMNFVGD	TRRFNDVND	
	ALRNDEKVRV	IGIDRGERNL	LYVVVVDTDG	TILEQISLNS	IINNEYSIET	

Table 5 - Cas12a orthologs						
	DYHKLLDEKE IICLEDLNFQ FGGALNALQL KYENVEKAKE TYGERIIKFR MSVDEADFFR CDASKPKDAD QRNQI	GDRDRARKNW FKRGRQKVEK TSKFTSFKDM FFSRFDSISY NTEKNNSFDD SLTRLFQQTM ANGAFNIARK	TTIENIKELK QVYQKFEKML GKQTGIIYYV KTIVLSEEFK QMRNSSNDVT GLWVLEQIRN	EGYLSQVVNV IDKLNLYLVID PAYLTSKIDP FDYKKFTDRA ELFSIYGISY RDYIISPIMN TPSGDKLNLA	IAKLVLKYNA KSRKQDKPEE TTGFANLFYV CGARSQWTV EDGAELKNKI DRGEFFNSEA MSNAEWLEYA	
WP_028830240 type V CRISPR- associated protein Cpf1 [Proteocatella sphenisci]	MENFKNLYPI IDEKFKEETIE DEIFNNYLKP ESSGSMAYRI FITQSGITHY KMILSDRVSN KYKQLGNLPG SINYISELLT EKTNLIKDLL FNSVYNKSRN FNNDRGDYDY KQFLSKIWKA NHDEKYQDVF DGKLYVFQIW IFYRPASLNY INYKSEKLNS VEQKHLDEII KEGYISHVIN LIKKFNYIID IDPVTGTFVNL NNRYSISKTK GTLKSQDVET RENIKNLPAD NQQE	NKTLRFELRP ERLKYTEFSE DKNIDALFKN IDENLTTYLN NEIIGGISKS SFVLDTIEND ISYSSIVNAI DTDVSSNIK DILKSIQRFY YLTRKQYSDK FLGIWNKSTP KHPTTPEFDK GFNLRNTEDY SKDFSIDSKG CEDIIKKGHH KSLNNRTNEN NTDTKGVEHK EIQLQKEKYN KKDPETYIHG LYADDLKYKN WTLTSYGTRI YKKFMSLFKL ADANGAYNIA	YGKTLENFKK CDLGNMTSKD DPSNPVISTF NIEKIKKLPE ENVKIQGINE TELIEMISDL CSDYDNNFGD RYKELEQNYQ DLFDIVDEDK KIKLNFDSP ANEKIIPLED KYKEGRHKKG NSYTEFLEDV TKNLNTIYFE HAELKDKFDY LGQFTHIIGI THYLNKLEEK ALIVMENLNY YQLTNPITTL QEQAQSFQIK QTFRNPQKNN MLQLRNSVTG RKGIMAIENI	SGLLEKDAFK KKITDKAATN KGFTTYFVNF ELKSQLEGID GINLYCQKNK INKTEISQDV GKRKKSYEND VCKENFNATN NPSAEFYTWL LAKGWDANKE NGLFEKMQYK PDFEKEFLHE ERCNYNLSFN SLFSEENMIE PIIKDKRYSQ DRGERHLIYL SKTRDNERKS GFKNSRIKVE DKIGNQSGIV IDNIYFENG KWDSAEDYLT TDIDYMISPV MNGISDPLKI	ANSRRSMQAI LKKQVILSFD FEIRKHIFKG QIDKLNYYNE VKLPRLTPLY IMSDIQNIFI RKKHLETNVY WMNIKNIKQS SKNAEKLDPE IDNSTIIMRK LYPDPKMLP LIDCFKHGLV KIADTSNLIN KMFKLSGEAE DKFFFHVPMV TVVDVSTGEI WEAIETIKEL KQVYQKFETA LYIPAWNTSK EEFKLILNID SNEDYLKYIQ	(SEQ ID NO: 169)
WP_084502895.1 type V CRISPR- associated protein Cpf1 [Proteocatella sphenisci]	MIILYISTSN LEKDAFKANS TDKAATNLKK TTYFVNFFEI SQLEGIDQID LYCQKNKVKL TEISQDVIMS KKSyenDRKK ENFNATNWMN AEFYTWLSKN GWDANKEIDN FEKMQYKLYP EKEFLHELID NYNLSFNKIA SEENMIEKMF KDKRYSQDKF ERHLIYLTVV RDNERKSWEA NSRIKVEKQV GNQSGIVLYI IYFENGFEKF SAEYDLTEEF DYMISPVTDK ISDPLKISNE	MNMEGVFMEN RRSMQAIIDE QVILSFDDEI RKHIFKGESS KLNNYNEFIT PRLTPLYKMI DIQNIFIYK HLETNVYSIN IKNIKQSEKT AEKLDFFENS STIIMRKFN CFKHGLVNHD DTSNLINDGK KLSGEAEIFY FFHVPVINY DVSTGEIVEQ KLILNIDGTL TGTHFDSREN DYLKYIQNQQ	FKNLYPINKT KFKETIEERL FNNYLKPDKN GSMAYRIIDE QSGITHYNEI LSDRVSNFSV QLGNLPGISY YISELLTDTD NLIKDLLDIL VYNKSRNYLT DRGDYDYFLG LSKIWKAKHP EKYQDVFGFN LYVFIWSKD KSEKLSKSL YSISKTKWTL KSQDVETYKK IKNLPADADA E	LRFELRPYGG KYTEFSECDL IDALFKNDPS NLTTYLNNEI IGGISKSENV LDTIENDTEL SSIVNAICSD VSSNIKMRKY KSIQRFYDLF RKQYSDKKIK IWNKSTPANE TTPFDKDYK LRNTEDYNSY FSIDSKGTKN NNRTNENLQ TKGVEHKTHY KLQEKYNALI PETYIHGYQL DDLKYKNQEQ TSYGTRIQT FMSLFKLMLQ NGAYNIARKG	TLENFKKSGL GNMTSKDKKI NPVISTFKGF KIKKLPEELK KIQGINEGIN IEMISDLINK YDNNFGDGKR ELEQNYQVCK DIVDEKNPS LNFDSPTLAK KIIPLEDNGL EGRHKKGPDF TEFLEDVERC LNTIYFESLF LKD KFDYPII FTHIIGIDRG LNKLEEKSKT VMENLNYGFK TNPITTLDKI AKSFIQKIDN RNPQKNNKWD LRNSVTGTDI IMAIENIMNG	(SEQ ID NO: 170)
WP_055225123.1 Eubacterium rectale	MNNGTNNFQN RQILKDIDDD EQTEYRKAIH	FIGISSLOKT YYRGFISSETL KKFANDDRFK	LRNALIPTET SSIDDIDWTS NMFSAKLISD	TQQFIVKNGI LFEKMEIQLK ILPEFVIHNN	IKEDELRGEN NGDNKDTLIK NYSASEKEEK	(SEQ ID NO: 171)

Table 5 - Cas12a orthologs						
	TQVIKLF SRF	ATSF KDY FKN	RANCF SADDI	SSSSCH RIVN	DNAEIFF SNA	
	LVYRRIVKSL	SNDDINKISG	DMKDSL KEMS	LEBIYSY EKY	GEFITQEGIS	
	FYNDICGKVN	SFMNLYCQKN	KENKNLYK LQ	KLHKQILCIA	DTSYEVYPKF	
	ESDEEVYQSV	NGFLDNISSK	HIVERLRKIG	DNYNGYNLDK	IYIVSKFYES	
	VSQKTYRDWE	TINTALEIHY	NNILPGNGKS	KADKVKKAVK	NDLQKSITEI	
	NELVSNYKLC	SDDNIKAETY	IHEISHILNN	FEAQELKYNP	EIHLVESELK	
	ASELKNVLDV	IMNAFWHCSV	FMTEELVDKD	NNFYAELEEI	YDEIYPVISL	
	YNLVRNYVTQ	KPYSTKKIKL	NFGIPTLADG	WSKSKEYSNN	AAILMRDNLY	
	YLGIFNAKNK	PDKKIEGNT	SENKGDYKKM	IYNLLPGPNK	MIPKVF LSSK	
	TGVETYPKPSA	YILEGYKQNK	HIKSSKDFDI	TFCHDLIDYF	KNCIAIHPEW	
	KNFGFDFS DT	STYEDISGFY	REVELQGYKI	DWTYISEKDI	DLLQEK GQLY	
	LFQIYNKDFS	KKSTGNDNLH	TMYLKNLFSE	ENLKDIVLKL	NGEAEIFFRK	
	SSIKNPIIHK	KGSILVNRTY	EAEKDKQFGN	IQIVRKNPIE	NIYQELYKYF	
	NDKSDKELSD	EAAKLKNVVG	HHEAATNIVK	DYRYTYDKYF	LHMPITINFK	
	ANKTGFINDR	ILQYIAKEKD	LHVIGIDRGE	RNLIYVSVID	TCGNIVEQKS	
	FNIVNGYDYQ	IKLKQQEGAR	QIARKEWKEI	GKIKEIKEGY	LSLVIHEISK	
	MVIKYNALIA	MEDLSYGFKK	GRFKVERQVY	QKFETMLINK	LNLYLVFKDIS	
	ITENGGLLKG	YQLTYIPDKL	KNVGHQCGCI	FYVPAAYTSK	IDPTTG FVNI	
	FKFKDLTVDA	KREFIKKFDS	IRYDSEKNLF	CFTFDYNNFI	TQNTVMSKSS	
	WSVYTYGVRI	KRRFVNGRFS	NESDTIDITK	DMEKTLEMTD	INWRDGHDLR	
	QDIIDYEIVQ	HIFEIFRLTV	QMRNSLSELE	DRDYDRLISP	VLNENNIFYD	
	SAKAGDALPK	DADANGAYCI	ALKGLYEIKQ	ITENWKEDGK	FSRDKLKISN	
	KDWFD FIQNK	RYL				
WP_055237260.1 Eubacterium rectale	MNNGTNNFQN	FIGISSLOKT	LRNALIPTET	TQQFIVKNGI	IKEDELRGEN	(SEQ ID NO: 172)
	RQILKDIMDD	YYRGFIS ETL	SSIDDIDWTS	LF EKMEIQLK	NGDNKDTLIK	
	EQAEKRKA IY	KKFADDDR FK	NMFSAKLISD	ILPEFVIHNN	NYSASEKEEK	
	TQVIKLF SRF	ATSF KDY FKN	RANCF SADDI	SSSSCH RIVN	DNAEIFF SNA	
	LVYRRIVKSL	SNDDINKISG	DMKDSL KEMS	LDEIYSY EKY	GEFITQEGIS	
	FYNDICGKVN	SFMNLYCQKN	KENKNLYK LK	KLHKQILCIA	DTSYEVYPKF	
	ESDEEVYQSV	NGFLDNISSK	HIVERLRKIG	DNYNGYNLDK	IYIVSRFYES	
	VSQKTYRDWE	TINTALEIHY	NNILPGNGKS	KADKVKKAVK	NDLQKSITEI	
	NELVSNYKLC	PDDNIKAETY	IHEISHILNN	FEAQELKYNP	EIHLVESELK	
	ASELKNVLDV	IMNAFWHCSV	FMTEELVDKD	NNFYAELEEI	YDEIYPVISL	
	YNLVRNYVTQ	KPYSTKKIKL	NFGIPTLADG	WSKSKEYSNN	AAILMRDNLY	
	YLGIFNAKNK	PDKKIEGNT	SENKGDYKKM	IYNLLPGPNK	MIPKVF LSSK	
	TGVETYPKPSA	YILEGYKQNK	HLKSSKDFDI	TFCDRLIDYF	KNCIAIHPEW	
	KNFGFDFS DT	STYEDISGFY	REVELQGYKI	DWTYISEKDI	DLLQEK GQLY	
	LFQIYNKDFS	KKSTGNDNLH	TMYLKNLFSE	ENLKDIVLKL	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	LNLYLVFKDIS	
	ITENGGLLKG	YQLTYIPEKL	KNVGHQCGCI	FYVPAAYTSK	IDPTTG FANI	
	FKFKDLTVDA	KREFIKKFDS	IRYDSEKNLF	CFTFDYNNFI	TQNTVMSKSS	
	WSVYTYGVRI	KRRFVNGRFS	NESDTIDITK	DMEKTLEMTD	INWRDGHDLR	
	QDIIDYEIVQ	HIFEIFKLTV	QMRNSLSELE	DRDYDRLISP	VLNENNIFYD	
	SAKAGDALPK	DADANGAYCI	ALKGLYEIKQ	ITENWKEDGK	FSRDKLKISN	
	KDWFD FIQNK	RYL				
WP_055272206.1 Eubacterium rectale	MNNGTNNFQN	FIGISSLOKT	LRNALPTTET	TQQFIVKNGI	IKEDELRGEN	(SEQ ID NO: 173)
	RQILKDIMDD	YYRGFIS ETL	SSIDDIDWTS	LF EKMEIQLK	NGDNKDTLIK	
	EQAEKRKA IY	KKFADDDR FK	NMFSAKLISD	ILPEFVIHNN	NYSASEKEEK	
	TQVIKLF SRF	ATSF KDY FKN	RANCF SADDI	SSSSCH RIVN	DNAEIFF SNA	
	LVYRRIVKSL	SNDDINKISG	DMKDSL KEMS	LEKIYSY EKY	GEFITQEGIS	
	FYNDICGKVN	SFMNLYCQKN	KENKNLYK LK	KLHKQILCIA	DTSYEVYPKF	
	ESDEEVYQSV	NGFLDNISSK	HIVERLRKIG	DNYNGYNLDK	IYIVSKFYES	
	VSQKTYRDWE	TINTALEIHY	NNILPGNGKS	KADKVKKAVK	NDLQKSITEI	
	NELVSNYKLC	PDDNIKAETY	IHEISHILNN	FEAQELKYNP	EIHLVESELK	
	ASELKNVLDV	IMNAFWHCSV	FMTEELVDKD	NNFYAELEEI	YDEIYPVISL	

Table 5 - Cas12a orthologs						
	YNLVRNYVTQ	KPYSTKKIKL	NFGIPTLADG	WSKSKEYSNN	AIIILMRDNL	
	YLGIFNAKNK	PEKKIIEGNT	SENKGDYKKM	IYNLLPGPNK	MIPKVFLSSK	
	TGVETYKPSA	YILEGYKQNK	HLKSSKDFDI	TFCRDLIDYF	KNCIAIHPEW	
	KNFGFDFSDT	STYEDISGFY	REVELQGYKI	DWTYISEKDI	DLLQEKQGLY	
	LFQIYNKDFS	KKSTGNDNLH	TMYLKNLFSE	ENLKDVVLKL	NGEAEIFFRK	
	SSIKNPIIHK	KGSILVNRTY	EAEKDKQFGN	IQIVRKTIPE	NIYQELYKYF	
	NDKSKDELSD	EAALKNAVG	HHEAATNIVK	DYRYTYDKYF	LHMPITINFK	
	ANKTSFINDR	ILQYIAKEKD	LHVIGIDRGE	RNLIYVSVID	TCGNIVEQKS	
	FNIVNGYDYQ	IKLKQQEGAR	QIARKEWKEI	GKIKEIKEGY	LSLVIHEISK	
	MVIKYNIAIA	MEDLSYGFKK	GRFKVERQVY	QKFETMLINK	LNYLVFKDIS	
	ITENGGLLKG	YQLTYIPEKL	KNVGHQCGCI	FYVPAAYTSK	IDPTTGFWNI	
	FKFKDLTVDA	KREFIKKFDS	IRYDSDKNLF	CFTFDYNNFI	TQNTVMSKSS	
	WSVYTYGVRI	KRRFVNGRFS	NESDTIDITK	DMEKTLEMTD	INWRDGHDLR	
	QDIIDYEIVQ	HIFEIFKLTV	QMRNSLSELE	DRNYDRLISP	VLNENNIFYD	
	SAKAGDALPK	DADANGAYCI	ALKGLYEIKQ	ITENWKEDGK	FSRDCLKISN	
	KDWDFIQNK	RYL				
OLA16049.1 Eubacterium sp. 41_20	MNNGTNNFQN	FIGISSLOKT	LRNALIPTET	TQQFIVKNGI	IKEDELRGKN	(SEQ ID NO: 174)
	RQILKDIMDD	YYRGFISSETL	SSIDDIDWTS	LFEEKMEIQLK	NGDNKDTLIK	
	EQAEKRKAIIY	KKFADDDRFRK	NMFSAKLISD	ILPEFVIHNN	NYSASEKKEK	
	TQVIKLFSRF	ATSPKDYFKN	RANCFSAADDI	SSSSCHRIVN	DNAEIFFSNA	
	LVYRRIVKNL	SNDDINKISG	DMKDSLKEMS	LEEIYSYEKY	GEFITQEGIS	
	FYNDICGKVN	SFMNLYCQKN	KENKNLYKLR	KLHKQILCIA	DTSYEVYPKF	
	ESDEEVYQSV	NGFLDNISSE	HIVERLRKIG	DNYNDYNLDK	IYIVSKFYES	
	VSQKTYRDWE	TINTALEIHY	NNILPGNGKS	KADKVKKAVK	NDLQKSITEI	
	NELVSNYKLC	SDDNIKAETY	IHEISHILNN	FEAHELKYNP	BIHLVESELK	
	ASELKNVLDI	IMNAFWHCSV	FMTEELVDKD	NNFYAELEEI	YDEIYPVISL	
	YNLVRNYVTQ	KPYSTKKIKL	NFGIPTLADG	WSKSKEYSNN	AIIILMRDNL	
	YLGIFNAKNK	PDKKIEGNT	SENKGDYKKM	IYNLLPGPNK	MIPKVFLSSK	
	TGVETYKPSA	YILEGYKQNK	HLKSSKDFDI	TFCRDLIDYF	KNCIAIHPEW	
	KNFGFDFSDT	SAYEDISGFY	REVELQGYKI	DWTYISEKDI	DLLQEKQGLY	
	LFQIYNKDFS	KKSTGNDNLH	TMYLKNLFSE	ENLKDIVLKL	NGEAEIFFRK	
	SSIKNPIIHK	KGSILVNRTY	EAEKDKQFGN	IQIVRKTIPE	NIYQELYKYF	
	NDKSKDELSD	EAALKNVVG	HHEAATNIVK	DYRYTYDKYF	LHMPITINFK	
	ANKTSFINDR	ILQYIAKEKD	LHVIGIDRGE	RNLIYVSVID	TCGNIVEQKS	
	FNIVNGYDYQ	IKLKQQEGAR	QIARKEWKEI	GKIKEIKEGY	LSLVIHEISK	
	MVIKYNIAIA	MEDLSYGFKK	GRFKVERQVY	QKFETMLINK	LNYLVFKDIS	
	ITENGGLLKG	YQLTYIPDKL	KNVGHQCGCI	FYVPAAYTSK	IDPTTGFWNI	
	FKFKDLTVDA	KREFIKKFDS	IRYDSEKNLF	CFTFDYNNFI	TQNTVMSKSS	
	WSVYTYGVRI	KRRFVNGRFS	NESDTIDITK	DMEKTLEMTD	INWRDGHDLR	
	QDIIDYEIVQ	HIFEIFKLTV	QMRNSLSELE	DRDYDRLISP	VLNENNIFYD	
	SAKAGYALPK	DADANGAYCI	ALKGLYEIKQ	ITENWKEDGK	FSRDCLKISN	
	KDWDFIQNK	RYL				

Table 6 - Cas12b (C2c1) orthologs						
Alicyclobacillus macrosporangiidus strain DSM 17980	MVAVKSIKVK	LMLGHLPEIR	EGLWHLHEAV	NLGVRYYTEW	LALLRQGNLY	(SEQ ID NO: 175)
	RRGKDGAQEC	YMTAEQCRQE	LLVRLRDRQK	RNGHTGDPGT	DEELLGVARR	
	LYELLVPQSV	GKKGQAQMLA	SGFLSPLADP	KSEGKGKTSK	SGRKPAWMGM	
	KEAGDSRWVE	AKARYEANKA	KDPTKQVIAS	LEMYGLRPLF	DVFTETYKTI	
	RWMPLGKHQG	VRAWDRDMFQ	QSLERLMSWE	SWNERVGAEF	ARLVDRDRF	
	REKHFTGQEH	LVALAQRLEQ	EMKEASPGFE	SKSSQAHRT	KRALRGADGI	
	IDDWLKLSEG	EPVDRFDEIL	RKRQAQNPRR	FGSHDLFLKL	AEPVFQPLWR	
	EDPSFLSRWA	SYNEVLNKLK	DAKQFATFTL	PSPCSNPVWA	RFENAEGTNI	
	FKYDFLFDHF	GKGRHGVRFO	RMIVMRDGVF	TEVEGIVVPI	APSRQLDALA	
	PNDAAAPIDV	FVGDPAPAGA	FRGQFGGAKI	QYRRSALVRK	GRREEKAYLC	
	GFRLPSQRR	GTPADDAGEV	FLNLSLRVES	QSEQAGRRNP	PYAAVFHISD	
	QTRRVIVRYG	EIERYLAEHP	DTGIPGSRGL	TSGLRVMSVD	LGLRTSAAIS	

Table 6 - Cas12b (C2c1) orthologs						
	VFRVAHRDEL	TPDAHGRQPF	FFPIHGMDHL	VALHERSHLI	RLPGETESKK	
	VRSIREQRLD	RLNRLRSQMA	SLRLLVRTGV	LDEQKRDRNW	ERLQSSMERG	
	GERMPDWD	LFQAQVRYLA	QHRDASGEAW	GRMVQAAVRT	LWRQLAKQVR	
	DWRKEVRRNA	DKVKIRGIAR	DVPGGHSQAQ	LDYLERQYRF	LRSWSAFSVQ	
	AGQVVRAERD	SRFAVALREH	IDNGKKDRLK	KLADRILMEA	LGYVYVTDGR	
	RAGQWQAVYP	PCQLVLEEL	SEYRFSNDRP	PSENSQLMVW	SHRGVLEELI	
	HQAQVHDLV	GTIPAAFSSR	FDARTGAPGI	RCRRVPSIPL	KDAPSIPIWL	
	SHYLKQTERD	AAALRPGELI	PTGDGEFLVT	PAGRGASGVR	VVHADINAAH	
	NLQRRLEWENF	DLSDIRVRCD	RREGKDGTVV	LIPRLTNQRV	KERYSGVIFT	
	SEDGVSFTVG	DAKTRRRSSA	SQEGGDDLSD	BEQELLAEAD	DARERSVVLV	
	RDPGSGFVNGG	RWTAQRAFWG	MVHNRIETLL	AERFSVSGAA	EKVRG	
Bacillus hisashii strain C4 WP_095142515.1	MATRSFILKI	EPNEEVKKGL	WKTHEVLNHG	IAYYMNILKL	IRQEAIYEH	(SEQ ID NO: 176)
	EQDPKNPKKV	SKAEIQAELEW	DFVLKMQKCN	SFTHEVDKDE	VFNILRELYE	
	ELVPSSVEKK	GEANQLSNKF	LYPLVDPNSQ	SGKGTASSGR	KPRWYNLKIA	
	GDPSWEEEEK	KWEEDKKKDP	LAKILGKLA	YGLIPLFIPY	TDSNEPIVKE	
	IKWMEKSRNQ	SVRRLDKDMF	IQALERFLSW	ESWNLKVKKE	YEKVEKEYKT	
	LEERIKEDIQ	ALKALEQYKE	ERQEQLLRDT	LNTNEYRLSK	RGLRGRWREI	
	QKWLKMDENE	PSEKYLEVFK	DYQRKHPREA	GDYSVYEFSL	KKENHFIWRN	
	HPEYPYLYAT	FCEIDKKKKD	AKQATFTLA	DPINHPLWVR	FEERSGSNLN	
	KYRILTEQLH	TEKLKKKLT	QLDRLIYPT	SGGWEEKGKV	DIVLLPSRQF	
	YNQIFLDIEE	KGKHAFTYKD	ESIKFPLKGT	LGGARVQFDR	DHLRRYPHKV	
	ESGNGVRIYF	NMTVNIEPTE	SPVSKSLKIH	RDDFPKVVNF	KPKELTEWIK	
	DSKGKLLKSG	IESLEIGLRV	MSIDLQGRQA	AAASIFEVVD	QKPDIEGKLF	
	FPIKGTETYA	VHRASFNIK	PGETLVKSRE	VLRKAREDN	KLMNQKLNFL	
	RNVLFHFQFE	DITEREKVRT	KWISRQENS	VPLVYQDELI	QIRELMYKPY	
	KDWVAFKQL	HKRLVEBIG	EVKHWRKSLS	DGRKGGLYGS	LKNIDEIDRT	
	RKFLRLWSLR	PTEPGEVRR	EPGQRFADQ	LNHLNALKED	RLKKMANTII	
	MHALGYCYDV	RKKKWAQKNP	ACQIILFEDL	SNYNPYEERS	RFENSKLMKW	
	SRREIPRQVA	LQGEIYGLQV	GEVGAQFSSR	FHAKTGSPGI	RCSVVTKEKL	
	QDNRFFKNLQ	REGRLTLDKI	AVLKEGDLYP	DKGGEKFISL	SKDRKCVTTH	
	ADINAAQNLQ	KRFWTRTHGF	YKVYCKAYQV	DGQTVYIPES	KDQKQKIEE	
	FGEGYFILKD	GVYEWVNAK	LKIKKGSSKQ	SSSELVDSI	LKDSFDLASE	
	LKGEKMLYR	DPSGNVFPD	KWMAAGVFFG	KLERILISK	TNQYSISTIE	
	DDSSKQSM					
Candidatus Lindowbacteria bacterium RIFCSPLOWO2 OGH55994.1	MPRDDLLLT	NLNSTAKGIR	ERGKTEGTD	KKKSGRKSSW	PMDKAAWETA	(SEQ ID NO: 177)
	KTSDDSAHFL	EKLQHPDLK	DAFGNLSSGG	SKKLEYKKL	AGSAPWKESQ	
	SVILEKAARW	KEAQEREER	EQDSSEHGSK	AAYRRLFDAG	CLPMPEFAKY	
	IDENQIEFGD	LKLSDCGAEW	KRGMWNQAGQ	RVRSHMGWQR	RREKENAVYS	
	LRKELFEKGG	AIRRKSEEL	TPEDILPGKA	APDQNDWQER	PAYGNQMWFI	
	GLRSYEENEM	AKYABEAGMG	SRAPRIRRG	TIKGWSKLRE	RWLQILKRN	
	QATRDDLIGE	LNALRSQDPR	AYGDARLFDW	LSKTDQRFLW	DGFDADGKIL	
	CGRDDRDVCS	AFVAYNEEFA	DEPSSITLTE	TDERLHPVWP	FFGESSAVPY	
	EIEYDLETAC	PTAIRLPLLV	GKENGGAER	QGTRLPLAEY	ADLASSFQLP	
	TPVRLDVLVE	IREVTRAGRK	VTCPFSYFKQ	NGVWYVREGE	IPSGESIQIK	
	QTDRKIENGK	IFISSKLKMA	YRDDLMVSPA	TGDFGSIKIL	WERIELASHV	
	DQKKLPETAP	ARSRVFVSFS	CNVVERAPRK	QLTRKPDVAV	VTIPSGVDQ	
	LVVVSTDVRT	GKSKSSSAPP	LPPGSRLWPA	DAVHGDPLR	ILSVDLGRHR	
	SAYAVWELGL	QKSWRAGVL	KGSTQTPVYA	DCTGTGLLCL	PGDGEDTPAE	
	EESLRLRSRQ	IRRLNLQNS	ILRVSRLLSL	DKFEKTIFEQ	SDVRDRPNKK	
	GLRIRRRCRT	EKTPLSEAEV	RKNCDKAAEI	LIRWADTDAM	AKSLAATGNA	
	DISFWKYMAV	KNPPLSAVVD	VAPSTIVPDD	GPDRTELKKK	RQEEEEKFAS	
	SIYENRVKLA	GALCSGYDAD	HRRPATGGLW	HDLDRTLIRE	ISYGDGQKQ	
	NPRKLNNEGI	LRLRLRPRA	RPDWREFHRT	LNDANRIPKG	RTLRRGLSMG	
	RLNFLKEVGD	FVKKWSRPR	WPGDRRHIPP	GQLFDRQDAE	HLEHLRDDRI	
	KRLAHLIVAQ	ALGFEPDIRR	GLWKYVDGST	GEILWQHPET	RRFFAEGAAG	
	ELREVSRPAE	IDDDAAARPH	TVSAPAHIVV	FENLIRYRFQ	SDRPKTENAG	
	LMQWAHRQIV	HFTKQVASLV	GLKVAMVYAA	FSSKFCSCRG	SPGARVSRFD	
	PAWRNQEWFK	RRTSNPRSKV	DHSLKRASED	PTADETRPWW	LIEGGKEFVC	
	ANAKCSAHDE	PLNADENAAA	NIGLRFLRGV	EDFRTKVNPA	GALKGKLRF	

Table 6 - Cas12b (C2c1) orthologs						
	TGIHSFRPPV LFRDPSGAFR QPQSGPGLAK	SGSPFWSPMA NKQYWYEGKI PGGD	EPAQKKKIGA FWSNVMMAVE	AAPGADVDEA AKIAGASVGA	GDADESGVVV KPVAASWGQA	
Elusimicrobia bacterium RIFOXYA12 OGS02326.1	MNRIYQGRVT TLALAAISGS SILGLEQNAS ERLPFFCPRN TLDISLKFSS IDKHGSSSLKV DREVTNDPIA IMSPHRFYQD FEGDTRIDLL QKQWSELAEK WLKSGTEEW FPPKSETEEK ARFSYSAPRL TRITLMAKGL LRWPHEKKPK KSRFIGEACD LFGEKGRSAT IVARRAQAAV EKVKETIISL LDQTGKAMPN AKRDDSIPDC ELNETCDMHG AVRDKLKEMC GFRAVEVIPG KPEKPRTLV ISDPRIWEIH DRKPNYFADF NKRHIEDWSN	KVEVPDGGKDE AVGSDEKSI FDIAAKHILR IQSKRSPTSK PKIVFLEDPK PAPGSKPSGL DARVNDKPHF TQKRKEAARK RKLVTDTLW GETTEGKLE AENPLKAWLN EVSKNTHSLT RRDCLRSENN DDIQLGFPVE NPPEQPWWGI KKWYAKVSRM PLEAEETAEL SRLHRWAWFF LKVQKQELLPT VLIRGQRGLS CPDLLEKLDH AYAKVDNPVS EVFFPLCERR FELKYPWSWL PIAGGPPIFVP PRLRTEKRDG SGKVDWGFNG KQKQ	KGNIKWKKLE LREWAVQVQN TSEAKPEQRA KARAECLKQF YPSAIVFKYF DYFTNIALIR LEEKIKAIEG LGESETPDNN VLKHEQTEHQ YKELQYELTD TSLASEHIKN ENLYKAPWLQ ANSQELQKEV DSFSCLAVDL GLLRLPGEDV IKLFGANEKD DEAKRSDDAI LLTRLANRVL MDRIEQITEL MKEQRVNQTA FIVIEDLSRY KAGSAWVSLP KDKKDKAGNL ISEVGLSSFG RLFAREKRKY IKNESGLTLV	NWSDILWQHH IWEKAKKKAT SALIRLLEEI QEEVRRFHN DNACKKHKEL EKEKNRAAWF KGGQFKESDS EGKKTEYSIS SDFGSATLYQ KKRPIHFTPA SLQFTAGLIR PMMRALGIDE SNGISWKGQF GQRYAGAFAR KVWRDASKID VMPDNWSKEL REILESDDTD PLRGRSWEWK RKRFAQALNQS HMLAEALGL RSSQGRSPRE PILLETAAYS AKEALNIRTV LKPQVVQADI GEEKVEVQPS SGKALWWTIN	MLFQDAVNYY VFEGPQKRLT DKKNHNVCV QPEEVKNNAV VGIKKAPTES KATEKLMGK BFDLAAFI EDDDVDSLPG ERTLRIFPDI HLAKPEFHPI HPVYSPRYFD KTNVGGKAIK EKADRQNFAN NWGGIASLSA LDVSTIEKKG KENGFAFRKE SFPEQNDKLL LKQKVNKNEI KHHQKNDGFI LRRQIGKKAP KLAEPKDKK NSRLMKWCHR SRFCSRSGVA SEQLKAFNQD NAAINLGLRA KNEKAKKVKD QLQWERCDFI	(SEQ ID NO: 178)
Omnitrophica WOR_2 bacterium RIFCSPHIGHO2 OGX36711.1	MNRIYQGRVT ALAAMAEGLP SESASFGDAL FCDDLKKKPN SLVKCVPDQG LRQHASDVVN FPCPFTFDYV LARKNNKPIF EEREKRRSEA VKLFGDLDEE QDLQKAVTGF RWVKSUVYAA GAKHIREGLV LQPMMAALGL PVDLDISKLE PGQFWWDNPT AGGQSWFTKL AIALAKQLLH RYKTWLRSW QLQKLIADAA HLLVSDGQSP KPEFGSSTCG HSLFTEEREK LMKWCHRKIN SEDRPFWRKT VKEDQPLPAI KWLPRKGEA LNIACYGFAT	KVEKLKNGKS DKHPINVLRK KKILPPAPEN WDYSEELAR INPEGARNLV FIHKNKNQFS RAAVPEPKVK KAFTALEKWS EAEKYMMDE GSEHATGKIY QREHKLDMGY ADTRELAEE EVSLAVKDQS SSDARGCFTR IQKEGMYCLA CAAGSLRLPG NENSAELES RLTEQHKELW VDLRNFLEKA AEIPWLRGQR ESLPDPCPDL ADMHGEHEVI EKVKLLAEPF IEKQSVAREV RQADINAAVN NKRENAQFGN IQNLQHPFLA	PDDRBEELKDW RMEEAWEFEF KEVRALAVAL KTGSGDWVAG KEALDHLEGV SLPNGRLITE KVDQEKKSEQ DQNSKSCWSD NPEWKPKKET GPSRASLRGF TAFFLKLCER ERLQEPVRYT GKYGTCRVRL DSKGNVKEPA QOMNTAYEKN IDLSQRRAAD EDTEVIREGK ARDWLGDNAK DKTLDEIRKV LLHIAAYRALP GLSIARIEQL TDKIVRLKEE PGRSPVDFV GIPVIEVFAS FDCLDNLVVK IGLRAIAGPS VDLDLEVKN HASAVFSRQK	QTALWRHHEL RKTVTPAKNL LAEKARTLKP LWSEDALNKI SNGTKKEKND RARGGGININ SATWTELEKR FDKCAFEEAL EGDDVREVP GKLRNEWVDL DYWDIWRDDT PAEPQFSRRL HYSAPRLIRD VALMSDFVGR KLKQRFHLIW YALLHAGVNR RQIELSGKKG RFSFPEQNDK PYFASWGELA LRENTWRWIE ENFRRAVLSL RVNQTAHLII LEDLSRYTTD YSSKFDARTG GLNGIHLVLP CYHAHPKVRL RLDIDSDVLK GAVARLQWEV	FQDAVSYYTL RDSVPRWLGL QKTSASYWGR DELAOKSLKS PGPAKKTNNW KTYAGVLFKA ILRIGDDPIE KTLNQFNQKT LKGDPREYKL FTKANDNP EVEVKKIREK FMFSDIKGKQ HLSDGSSSMW KRELRLMLNF PGMELKETQE DSKTFFVELGQ RNATQSEYDQ LIDLYYGALS GNGTNEATVQ NGKDGKGKPL NRLLRHEIGT AQSLGVRLKG KSRSRSENSR APGFRAVEVT QNGGPLFIAA IKGESGTDKG GDNTNLFHDP CRAINSRRLE	(SEQ ID NO: 179)

Table 6 - Cas12b (C2c1) orthologs						
	AWQKKAEEKAA VKR					
Phycisphaerae bacterium ST- NAGAB-D1 (transposase) AQT69685.1	MATKSYRARI KKLAQIMLSR LQAERFATVK RELMQQWRKD QLAAEQQPTA FKTVTDDVRK FDSFKRPPTL GNWFFNWMPA TDAERKAGIA AFSSDGKTRK SPSIPPLRQV KGASAIIVSLA LVSEQAPEFG EKCTDFGYPV LKDAETKTF	LTD SRLAAAL SNTFAHGVMD NRAIHTKSRG RAAWLKDKNE SKNNRPRKGT QIITKYPQON TLPSPYRHPY SLKPDPRLPV GAKLMIKNNR PLRILSIDLG RRHDYDLKRK REHSADLILF ITVKDDINPW WEPGGHLFRC TVHTDKKTPP	DRTHVVFVES QITRDQPTST KVIPSPELTA WQOKHPEFMQ KRFARWHLWY KYITRLLDWL WFTMELDQFY SWRAETFETE SELLFTVFEQ IRHIGAFALT RRRHGKPVKG ENLHSLKFSA MTSRICSNCN PHCDHRVNAD LIFKPRPQFD	LKQMINTYLR LDEEWTDLAR VPAKFWHQVC FYNGPYQNFL KWLSENPEII EDNNPELCTL KKADFENGTI GRFPPYLGK DCPPRVKWK QGTRNDSAWQ QRSNANLQAH FDERWMNRQL LPGFRFSMKK INAAANLANK LWADSVKTRK	MQNGKFGPDH RIHKTGTPLF DSASAYIRS KLCCCCRITS EWRNKASASD ENLRRTYVKK QLLLIDEDDD IGKLSRPAP TKNRKCPADN TESLKKGIIN RTNMAQDRFK RDMNRRHIVE KNPYREKLPR FFGLGYWNG QLGPDPF	(SEQ ID NO: 180)
Planctomycetes bacterium RBG_13_46_10 OHB62175.1	MSVRSFQARV KQKSLYKSIS DAAKLSSQGI EHNEWLKEKE KWLSDNPDA PEMKALDNLH NPKPTNPEGY DPRLSLIRPV LIFPDKTPKA LPHGLVSCAV HPYRWTEGPD DRFKKAARTI INRALAGWNR RYSIIRENNR FKSYDWKRL	ECDKQTMHLL QSILEANAQN HVDYDKQVLG KWESEDEHKK AWRGNKAVIN GYERNFVRR RKLILPKKAG KGRRVVRKGG AYLYFTCDIP DLSMRRGTTG LGHIAXHKKR VNFALNTENA RHLVERVIEM REIRFGYVEK SEKKQKEEIE	WRTHKVFNER ADYLLNSVSI DLPGMMSQMV YLDLREKFEQ PLSEKAQIRI RKTCKNPDPG DLGSLEMRL EQGQTKETDS DEPLTETAKK FATLCRYENG IRILRSKRK ASKNGFYPR AKDAGFKRRV LFACPNCGYC TIESKLMCKL	LPBIIKILFK KGWKPGTAKK CRQSVEAISC FEQSIGGKIT NKAKPNKKN DHKPTFTLPH TGEKNKGNYP YEFFDKHLKK IQWLETGDTV KIHLRSRNL PVKGEESHID DVLLLENLEG FEIPPYGTSQ ANADHNASVN CAMHKISRG	MKRGECCQND YRNASFTWAD HIELTKKWEK KRRGRWHLYL VERDEFFKAN PTIHPRFVVF DDWISVKFKA WRPAKLSGVK KKGKKRKKKV WVGKKEGKGC LQKHIDYMG LIPDAEKERG VCSKCGALGR LNRRFLIEDS ISK	(SEQ ID NO: 181)
Spirochaetes bacterium GWB1_27_13 OHD16008.1	MSFTISYPFK GEDETGKRIY ICLVSSSMEN DYGKILKRL SYESWCKLMI FEQDSGFAKK DNSKLVGDRN RFTMPDFKKH FEENYTFNLA YIKYCDTAKN KSLFAKEKLQ DWMDEKTFQN LNDQNHKIIIR ERKNKIGIFT GIIDWNNVDN GGLSMQRIDF DNRVKQGVNY LTEYTFAEEDR KCLSDNNAPG KSIKINDILP VVLKKDKDET KRIKMDKHEK KNKMTKLLRE	LIINKNDEAK IEETBYKKQI KKGFPNAQQA RDNNLIPIFT KERFDKEVKI GKFYITGRML LFEFTIKLEN PLWYRYEDPS KLKKLSENIT TSYGKSGGIR PQIFNTVDNK VKELNEGIRV INDKNYAKH RQINKLNQLL KILSYENKED LEKLNKLKRL ILMTALGYIK PREENSKLRM IRCHQVTKKD LRGGELFAST FHLKNPNYL ESTDEGDDYS RNNKKNKSG	ALLDTHQYMN SLNIFSPLEF KFTDIKKITA KELENKCENK KGFDEIKEKW QCLWNGDIDY NSNFRNYKIE FIPKSKNKEF LYFNRNELEN ITSLKIQKKF LSVDLGQRFF IYSKTIKLSG KIIRNDEIDK VINLHKELDK WVARTQKSAE DNEIKNDSKK WSHRQIFNFL LIDNSWFLKI KDGKLHIVQA KNYYSILNFV KNQIALFRDD AAHFFIKVAE	EGVKYYLEKL AESKGYILKE KLSPTALDRM QERDKIFEIL LKEKDRSEQN LKIKRDINKN VVKDENYITI EFIDSNEEEE YKDGKKMDSY GNEEQTNFLS AAVSCFEIMS EDDDLYKERK EKFKELIETT KLEIDFKEFI SIVLTPKFGY KQKEDWVKK QQKASLWGIL VVKDDAFCDL DINASRNIAK PTNEELTFFK SGIFFDKSLW	LMFRQEKIFI EFKTLMLRELY ENNNISLIHK IFAQAIKLL EKYEERQKT LINILNKYQT QIWLDPRPMP PLITERNNEY DKKDQKSKQ TVFTLSRIDY YFTQNQITKK EIDNNKLFFN INKNYKLSYQ KRYVKNTYND RECRKPIFRS KLKEHINELK RACQIILMEK VGDFVAPYTS IEINKENVKN RFLSQINPFR VEENKDIKPT VDGKIFWSVV	(SEQ ID NO: 182)
Verrucomicrobiacea e bacterium UBA2429 GCA_002343505.1	MPLSRIYQGR YQLCLVALAS AASLFEEAALA DPKANPTYDY VKLQPDKNFV	TNSLIILTPT SDGTRPLSKL KVLEGNEAPE SATARASASG GAEARALLLE	PQEPWDHKAL HEQMKAASWE RARELAGELL LTKLAAVIHA AAHFFIKVAE	ARFDSPLWRH AKTDTEDSWR LDKIEGDIQQ ENVTEEALKQ SPPTKLAEVL	HALFQDAVNY VRLARRLGIP AGRGYWPFRF VAAEMDLSTW ARFPDGLALW	(SEQ ID NO: 183)

Table 6 - Cas12b (C2c1) orthologs						
	QALPEKIAAL	PEETQVPRNR	KASPDLTTFAT	LLFQHFPSLF	TAAVLGLSVG	
	KPKSVKAPKV	VEKVSARRKA	NAVTOAVVIE	EPFIDFAELG	DDPIKLARGE	
	RGFVFPAFTS	LSFWAVPGPH	VPVWKEFDIA	AFKEALKTVN	QFKLKTSENR	
	ALLAEAQRRL	DYMDEKTHDW	KTGDSDEPGH	IPPRLKSDPN	FTLIQALTQD	
	EGVSNKATGD	QHIPKGVYTG	GLRGFYAIKK	DWCELWERKA	DKSQGTPTEE	
	ELISIVTDYQ	RDHVDVVDV	GLFRALCEPR	FWPLWQPLTD	EQEAEIRKAG	
	RAKDMISAYR	VWLELQEDVV	RLAQPIRFTP	AHAENSRLRF	MFSDISGSHG	
	AEFGSDGKSL	EVSIAVDVDG	KLQPVRAKLE	FSAPRAARDE	LEGLSGGSES	
	MRWFQPMKA	LDCPEVEMPA	LEKCAVSLMP	DVVKKGGGKW	VRLLLNFPAT	
	LEPEGLIRHI	GKQAMWYKQF	NGTYKPRTQQ	LDTGLHLYWP	GLEKAPBAED	
	AAAWNREEI	RAKGFSVLVS	DLGQRDAGAW	ALLESRSKKA	FSRNRQPFIE	
	LGEAGGLWS	TALLGLGLMR	LPGEDARTGA	LDDQGKRAVE	FHGKAGRNAL	
	EAEWQEAEM	ALLFGGEEAK	SRLGPGFDHL	SHSKQNEELL	RILSRAQSRL	
	ARFHRWSCRI	HEKPEATGDD	VIDYQGVDL	LTKTAEAMLE	NLKALYTNAG	
	GILDSKSKQP	LTLVGLRKKL	EAQKVEPEKI	AAVLKPHAEI	IFQRLGTLIP	
	ELKQHLRVSL	ERLANRELPL	RHREWVWNEA	FEKLEQGNFK	KEENPKWIRG	
	QRGLSMARIE	QIENLRKRFM	SLRRQMSLIP	GEQVKQGVED	KGQRKPWICE	
	DILNKLDRMK	QQRVNQTAHL	ILAQALGLRL	RPHLANDAER	EEKDIHGEYE	
	LIPGRKPVDF	IVMEDLSRYL	SSQGRAPSEN	GRLMKWCHRA	VLAQLKQMCCE	
	PFGIPVLEVP	AAYSRRFCAL	TGVPGFRAVE	VHDGNAEDFR	WKRLIKKAEK	
	DKSSKDAEAA	AMLFQDLHDL	NIEAREARKQ	DKKLPLRRLF	APVAGGPLFI	
	PMVGGGPRQA	DMNAAINLGL	RAIASPTCLR	ARPKIRAEK	DGKHQAMLGN	
	KLEKAAALTL	EPPKEPTKEL	AAQKRTNFFL	DEKFVGKFDT	AHVTTSGKKL	
	RLSGGMSLWK	AIKDGAQORV	KKINDARIAK	WKNNPPPEPD	PDDEIQF	
Alicyclobacillus kakegawensis WP_067936067.1	MAVKSIVKVL	RLSECPDILA	GMWQLHRATN	AGVRYYTEWV	SLMRQEILYS	(SEQ ID NO: 184)
	RGPDGGQCY	MTAEDCQREL	LRLRLNRQLH	NGRQDQPGTD	ADLLAISRRRL	
	YEILVLQSIG	KRGDAQQIAS	SFLSPLVDPN	SKGGRGEAKS	GRKPAWQKMR	
	DQGDPRWVAA	REKYEQRKAV	DPSKEILNSL	DALGLRPLFA	VFTETYSRGV	
	DWKPLGKSQG	VRTWDRDMFQ	QALERLMSWE	SWNRRVGEEY	ARLFQQKMKF	
	EQEHFAEQSH	LVKLARALEA	DMRAASQGFE	AKRGTAHQIT	RRALRGADRV	
	FEIWKSIPEE	ALFSQYDEVI	RQVQAEKRRD	FGSHDLFAKL	AEPKYQPLWR	
	ADETFLTRYA	LYNGVLRDLE	KARQFATFTL	PDACVNPWT	RFESSQGSNL	
	HKYEFLLDHL	GPGRHAVRFQ	RLLVVESEGA	KERDSVVVPV	APSGQLDKLV	
	LREEEKSSVA	LHLHDTARPD	GFMAEWAGAK	LQYERSTLAR	KARRDKQGM	
	SWRRQPSMLM	SAAQMLEDAK	QAGDVYLNIS	VRVKSPSEVR	GQRRPPYAAL	
	FRIDDKQRRV	TVNYNKLSAY	LEEHPDKQIP	GAPGLLSGLR	VMSVDLGLRT	
	SASISVFRVA	KKEEVEALGD	GRPPHYPIH	GTDDLAVVHE	RSHLIQMPGE	
	TETKQLRKLR	EERQAVLRPL	FAQLALLRLL	VRCGAADERI	RTRSWQRLTK	
	QGREFTKRLT	PSWREALELE	LTRLEAYCGR	VPDDEWSRIV	DRTVIALWRR	
	MGKQVRDWRK	QVKSAGVKV	KGYQLDVVGG	NSLAQIDYLE	QQYKFLRRWS	
	FFARASGLVV	RADRESHFAV	ALRQHIENAK	RDRLKKLADR	ILMEALGYVY	
	EASGPREGQW	TAQHPPCQLI	ILEELSAYRF	SDDRPPSENS	KLMAWGHRI	
	LEELVNQAQV	HDVLVGTVYA	AFSSRFDART	GAPGVRCRRV	PARFVGATVD	
	DSLPLWLTEF	LDKHLRLDKNL	LRPDDVIPTG	EGEFLVSPCG	EEAARVRQVH	
	ADINAAQNLQ	RLRWQNFDT	ELRLRCDVKM	GGEGLTVLPR	VNNARAKQLF	
	GKKVLVSQDG	VTFFERSQTG	GKPHSEKQTD	LTDKELELIA	EADDEARAKSV	
	VLFRDPSGHI	GKGHWIRQRE	FWSLVKQRIE	SHTAERIRVR	GVGSSLD	
Bacillus sp._V3-13 WP_101661451.1	MAIRSIKLM	KTNSGTDISIY	LRKALWRTHQ	LINEGIAYYM	NLLTLYRQEA	(SEQ ID NO: 185)
	IGDKTKEAYQ	AELINIIRNQ	QRNNGSSEEH	GSDQEILALL	RQLYELIIPS	
	SIGESGDANQ	LGNKFLYPLV	DPNSQSGKGT	SNAGRKPWRK	RLKEEGNPDW	
	ELEKKKDEER	KAKDPTVKIF	DNLNKYGLLP	LFPLFTNIQK	DIEWLPLGKR	
	QSVRKWDKDM	FIQAIERLLS	WESWNRVRAD	EYKQLKEKTE	SYKHELTGG	
	EEWIEKIRKF	EKERMELEK	NAFAPNDGYF	ITSRQIRGWD	RVYEKWSKLP	
	ESASPEELWK	VVAEQQNKMS	EGFGDPKVFS	FLANRENNDI	WRGHSEIRYH	
	IAAYNGLQKK	LSRTKEQATF	TLPDIAIEHPL	WIRYESPGGT	NLNLFKLEEK	
	QKKNYVYVTL	KIIPWSEEKW	IEKENIEIPL	APSIQFNQRI	KLKQHVKGKQ	
	EISFSDYSSR	ISLDGVLGGS	RIQFNRYKIK	NHKELLGEGD	IGPVFFNLV	
	DVAPLQETRN	GRLQSPIGKA	LKVISSDFSK	VIDYKPKELM	DWMNTGSASN	
	SFGVASLLEG	MRVMSIDMGQ	RTSASVSIFE	VVKELPKDQE	QKLFYSINDT	

Table 6 - Cas12b (C2c1) orthologs						
	ELFAIHKRSF ETKKTDPDERK IWKESLVELH RLLSISWSKRS TALGFKYDKE MKWAHRSIPR EDLKAGSNTL DSDNNELTVI SQTETIKKYF DISKTIELAQ LKKKILSNKV	LLNLPGEVVT KAIHKLMEIV HRIEPYVGQI RTPGEANRIE EKDRYKRWKE TVSMQGEMFG KRLIEDGFIN HADINAAQNL GKGSFVKNN EQKKYLTMF EL	KNNKQQRQER QSYDSWTASQ VSKWRKGLSE TDEPFGSSLL TYPACQIILF LQVGDVRSY ESELAYLKKG QKRFWQQNSE EQEVYKWEKS RDPSTGYFFNN	RKKRQFVRSQ KEVWEKELNL GRKNLAGISM QHIQNVKDDR ENLNRYLFNL SSRFHAKTGA DIIPSQGGEL VYRVPCQLAR EKMKIKTDTT ETWRPQKEYW	IRMLANVLRL LTNMAAFNDE WNIDELEDTR LKQMANLIIM DRSRRENSRL PGIRCHALTE FVTLSKRYKK MGEDKLYIPK FDLQDLGFE SIVNNIIKSC	
Desulfatirhabdium butyrivorans WP_028326052.1	MPLSNPPVT LTLRGGLDHT GAPSSYIVAS LSAAIRDDAV PMKDPEDKSS IAAWADNASQ TPNPVHELLR LNDVESVCGF ARRISLAHKW GAVEPYRIRR GDIQLFEALA VPAYRHPDEL IKPVPLCWQS NITGLFEQAD RWLVTFVSKL RLPGLRVLSV HIKVKKQGIG QGEEKDAREA DALKELGWIP RHGNRARIAY GWRDEAAKQL AKTSLGSDVIT TIRNVGGLSL ALELLREQRV HAVVIENLAN VSAWYTSRQD DARERFLCEL IHADLNAAAN GQPLRQPAQT IGEWKETSAY	QRAYTLRLRG LADTKVKGGK GDEPAKDRND WVNRSKVFDE ETEQEDKAKD GGSTVDDLV QNTSLNKS HL TYRVDKDGQP IKRAEAERHK RAVDGWKEVV EDDALCVWHK LHPVFCDFGK KRLTRDLALG WNGRLQAPRQ QPQGPWCAYA DLGHRYAAAC KQTEVDKTTI SNEEIWALHQ APDSSENLSR YLISEVKIRP WDSRIATLAG CKELSDAWKE SRLATITEFR KQLASRIAEA YRPDETRTRL SRTGAPGIRC NKTWKAKTPA IGLRALTDPD NADGAASKIR QNDVQYRVIR	ADPSDLWRE GKPDRTPTPE NVVSALEEIL AVKSVGSSLT LVQKAGQWLS SELRQHFDTK DDLKKKANTR VSVADYSKYD FEEDAKRIAN AAWSKSDCKS DGEATNEPDF SRWKINYDVH NNHRNDAASA QLEAIAVVRD EQHGLNTNPQ AVWEAVNTET YRRIGADTLP MECKLDRTKP EDGEAKDYRE GGIQEKL DEN YKAPEENGDN RWEDEDQRWK RKQVGVFFTR ALGIGSEG GK ENRRMTWSA QDVSVREFMQ EWKAGFVRI WPGKWWYVPC KGKKNRTAGT MLKEHIKSLD	ALWHTHEAVN ERKARRILLA QSRKVAKSEI REEAWDMLER SRYGTSEGAD ESKATNGLDW AESCKSKIGS VDYKWGTARH VPARAREWLD TEDRIAAARA QPLIDYSLAI KNVQAPFYRG VTRADRLGRA NPRLSEQERN YWPHADTNRD VKEACQNVGR DGRPHAPAWA LIDRLIASGW SLAVDDLMFS GRIDLLQDAL VSDVAYRKKQ KLLRWFKDWV LRPDGTRHEI GWDGGRPRQ SKVHKYLS EA SPFWRKQVKQ PLRGGEIFVS DPVSFESKMD SKEKVYLWRD NRTGDNVEG	KGAKVFGDWL LSWLSVESKL DDWKRDCSAS FFGSRDAYLT FCRMSDIY GK IIGLSSYTG H KGQRPYSDAI YIFAVMLDHA SFCKERSVTS LQDDSEIDKF EAEFFKRQFK LCLTLWTGSE ASNVTKSDMV LRMCGMIEHI RKVHARILIP DMPKEHDLYL RLDRQFLIKL GLLKRQMARL AVRTLRLALQ ALWHELFSSP QVYREQLRNV LPSGTQANNA GEQFGQKTL D RINDSRFAPC CQLNGLYLCT AEAKHDENKG ADSKSPSAKG YVKGCAAVKV ISAFPLESNE	(SEQ ID NO: 186)
Desulfonatronum thiodismutans WP_031386437.1	MVLGRKDDTA PVHVPEQVA LLDDLKPLK YLGALPEWAT PANQDALKNK LWLELGLLPL ALARAKRDEL GRALRSWTRV DGQEALWKER APGGSNLRTY SNVSFDQIQK TDLASKPGHV KFADQVRPGL DPEKLWAKHE LRLSVLQEDD LWKQHCHFFH WAVTYLEAVR DRIKTGADMI	ELRRALWTT EDALAMAREA GDAQKIGTNY PISKQEFDGK AYNKDDWKKE FIPVFDKTMV AALFLGMEDG REEWLRHGD DCVTSFSLN QIRKTENGLW GSKMVGRCRY WFKLTLDVRP RVLSVDLGVR RSFKITLPG PRTEHLRLFM DKAEKVVAER GLILRWNM RG IQAARGFVPR	EHVNLAAVEV QRRNGWPVVG AGPLFSDSTC DASHLRFKAT KDKGISSWAV GNLWNRLAVR FAGLREYELR QESRKNICNR DADGLLEK RK ADVLLSPRN QSANQQFEG L QAPQGWLDGK SFAACSVFEL NPSRKEEIAR EAIVDDPAKS FSRWRTE TRP RTYGEVNRQD KNGAGWVQVH	ERVLLRCRGR EDEEILLALR RRDEGKDVAC GGDDAFFRVS KYIQKQLQLG LALAHLLSWE RNESIKQHAF LQDRLRGKFG GYALMTFADA ESAAVEEKT F LGGAEILFDR GRPALPPEAK VRGGPDQGT Y RAAMEELRSL ALNAELFKGF KSSSWQDWRE KKQFGTVASA EPCRLILFED	SYWTLDRRGD YLYEQIVPSC CGPFHEVAGK IEKANAWYED QDPRTEVRRK SWNHRVQDQ EPVDRPYVVS DPDVFHWLAE RLHPRWAMY E NVRLAPSGQL KRIANEQHGA HFKTALSNKS FPAADGRTV D NGDIRRLKAI GDDRFRSTPD RRGYAGGKSY LLHHINQLKE LARYRFR TDR	(SEQ ID NO: 187)

Table 6 - Cas12b (C2c1) orthologs						
	SRRENSRLMR	WSHREIVNEV	GMQGEYGLH	VDTTEAGFSS	RYLASSGAPG	
	VRCRHLVEED	FHDGLPGMHL	VGELDWLLPK	DKDRTANEAR	RLLGGMVRPG	
	MLVPWDGGEL	FATLNAASQL	HVIHADINAA	QNLQRRFWGR	CGEAIRIVCN	
	QLSVDGSTRY	EMAKAPKARL	LGALQQLKNG	DAPFHLTSIP	NSQKPENSIV	
	MTPTNAGKKY	RAGPGEKSSG	EEDELALDIV	EQABELAQGR	KTFFRDPSPGV	
	FFAPDRWLPS	EIYWSRIRRR	IWQVTLERN	SGRQERAEMD	EMPY	
Lentisphaeria bacterium DCFZ01000012.1	MAVELNRIYQ	GRVNHVYIFD	ENQNQVSVDN	GDDLLFVHHE	LYQDAINYYL	(SEQ ID NO: 188)
	VALAAMALDS	KDSLFGKFKM	QIRAVWPDFY	RNGQLRPLGK	HSLIRSLGHA	
	AELNTSNGAD	IAMNLILEDG	GIPSEILNAA	LEHLAEKCTG	DVSQGLKTFE	
	PRFCDTAYHG	NWDVDAKSFS	EKKGRQRLVD	ALYSLHPVQA	VQELAPEIEI	
	GWGGVKTQTG	KFFTGDDEAKA	SLKKAISYFL	QDTGKNSPEL	QEYFSVAGKQ	
	PLEQYLKID	TFPEISFGRI	SSHQNNINISN	AMWILKFFPD	QYSVDLIKLN	
	IPNKKYEIGI	APQWGDPPVK	LSRGKRGYTF	RAFTDLAMWE	KNWKFVDRAA	
	FSDALKTINQ	FRNKTQERN	QLKRYCAALN	WMDGESSDKK	PPVEPADADA	
	VDEAATSVLP	ILAGDKRWNA	LLQLQKELGI	CNDFTENELM	DYGLSLRTIR	
	GYQKLRSMM	EKEEKMRAKT	ADDEEISQAL	QEI I I K F Q S S	HRDTIGSVSL	
	FLKLAEPKYF	CVVHDAKDNQ	NFASVDMVAD	AVRYYSYQEE	KARLEEPIQI	
	TPADARYSRR	VSDLYALVYK	NAKECKTGYG	LRPDGNFVFE	IAQKNAKGYA	
	PAKVVLAFSA	PRLKRDGLID	KEFSAYYPPV	LQAFRLREEEA	PKQSFKTTAV	
	ILMPDWDKNG	KRRILLNFPI	KLDVSAIHQK	TDHRFENQFY	FANNTNTCLL	
	WPSYQYKKPV	TWYQGGKPPD	VVAVDLGQRS	AGAVSRITVS	TEKREHSAI	
	GEAGGTQWYA	YRKFSGLLRL	PGEDATVIRD	GQRTTEELSGN	AGRLSTEEET	
	VQACVLCMKL	IGDATLLGGS	DEKTIRSFPK	QNDKLLIAFR	RATGRMKQLQ	
	RWLWMLNENG	LCDKAKTEIS	NSDWLVNKN	DNVLKEEKQH	REMLPAILLQ	
	IADRVLPRLG	RKWDWVLPQ	SNSFVLQQT	HGSGDPHKKI	CGQRGLSFAR	
	IEQLESRLMR	CQALNRILMR	KTGEKPATLA	EMRNNPIPD	CPDILMRDLA	
	MKEQRINQTA	NLILAQALGL	RHCLHSESAT	KRKENGMMHGE	YEKIPGVEPA	
	AFVVLLEDSR	YRFSQDRSSY	ENSRLMKWSH	RKILEKLALL	CEVFNVPILQ	
	VGAAYSSKFS	ANAI PGFRAE	ECSIDQLSFY	PWRELKDSRE	KALVEQIRKI	
	GHRLLTFAK	ATIIMPRNGG	PVFIPFVPSD	SKDTLIQADI	NASFNIGLRG	
	VADATNLLCN	NRVSCDRKKD	CWQVKRSSNF	SKMVYPEKLS	LSFDPIKKQE	
	GAGGNFFVLG	CSERILTGTS	EKSPVFTSSE	MAKKYPNLMF	GSALWRNEIL	
	KLERCKKINQ	SRLDKFIACK	EVQNEL			
Laceyella sediminis WP_106341859.1	MSIRSFKLKI	KTSGVNAEE	LRRGLWRTHQ	LINDGIAYYM	NWLVLRLQED	(SEQ ID NO: 189)
	LFIRNEETNE	IEKRSKEEIQ	GELLERVHKQ	QQRNQWSGEV	DDQTLQLTLR	
	HLVEEIVPSV	IGKSGNASLK	ARFFLGPLVD	PNNKTTKDVS	KSGPTPKWKK	
	MKDAGDPNWV	QEYKYMAER	QTLVRLEEMG	LIPLFPMTYD	EVGDIHWLPQ	
	ASGYTRTWDR	DMFQQAIRL	LSWESWNRV	RERRAQFEKK	THDFASRFSE	
	SDVQWMNKL	EYEAQQEKS	EENAFAPNEP	YALTKKALRG	WERVYHSWMR	
	LDSAAEEAY	WQEVATCQTA	MRGEFGDPAI	YQFLAQKENH	DIWRGYPERV	
	IDFAELNHLQ	RELRAKEDA	TFTLPDSVDH	PLWVRYEAPG	GTNIHGVDLV	
	QDTRKRLTLI	LDKFIPLDEN	GSWHEVKKVP	FSLAKSKQFH	RQVWLQEEQK	
	QKKREVVFYD	YSTNLPHLGT	LAGAKLQWDR	NFLNKRTQQQ	IEETGEIGKV	
	FFNISVDVRP	AVEVKNGRLQ	NGLGKALTVL	THPDGTHIVT	GWKAEQLEKW	
	VGESGRVSSL	GLDSLSEGLR	VMSIDLQORT	SATVSVFEIT	KEAPDNPYKF	
	FYQLEGTELF	AVHQRSFLLA	LPGENPPQKI	KQMRERIRWE	RNRKQQVDQ	
	LSAILRLHKK	VNERERIQAI	DKLLQKVASW	QLNEEIATAW	NQALSQLYSK	
	AKENDLQWNQ	AIKNAHHQLE	PVVGKQISLW	RKDLSTGRQG	IAGLSLWSIE	
	ELEATKKLLT	RWSKRSREPG	VVKRIERFET	FAKQIQHHIN	QVKENRLKQL	
	ANLIVMTALG	YKYDQEQKKW	IEVYPACQVV	LFENLRSYRF	SYERSRRENK	
	KLMEWSHRSI	PKLVQMGGEL	FGLQVADVYA	AYSSRYHGRT	GAPGIRCHAL	
	TEADLRNETN	I I H E L I E A G F	IKEEHRPYLQ	QGDLPVWSSG	ELFATLQKPY	
	DNPRILTLHA	DINAAQNIQK	RFWHPMSWFR	VNCESVMEGE	IVTYVPKNKT	
	VHKKQGKTFR	FVKVEGSDVY	EWAKWSKNRN	KNTFSSITER	KPPSSMILFR	
	DPSGTFEKEQ	EWVEQKTFWG	KVQSMIQAYM	KKTIVQRMEE		
Methylobacterium nodulans (long form)	MYEAIVLADD	ANAQLANAF	GPLTDPNSAG	FLEAFNKVDR	PAPSWLDQVP	(SEQ ID NO: 190)
	ASDPIDPAVL	AEANAWLDTD	AGRAWLVDTG	APPRWRSLAA	KQDPIWPREF	
	ARKLGELRKE	AASGTSIIK	ALKRDFGVLP	LFQPSLAPRI	LGSRSSLTWP	
	DRLAFRLAVG	HLLSWESWCT	RARDEHTARV	QRLEQFSSAH	LKGDLATKVS	

Table 6 - Cas12b (C2c1) orthologs						
	TLREYERARK EALHAI IATE NAMERLVERS QITLPLLKAA NEVFEAKLGS GVDDQLARAA LKDTAPTGTG LRQLRGGLNR ELERCSTVAD SMWSVQHLTD KDDRLLKTGAD RPRRENSQLM CLDTPAAFSS GYKPGDLVPL LPCGKSAVQG AEWRRLSGAQ PTDLWFPSAA	EQIAQLGLPM QTRKRGRFGD RDTALMTLPD DDGRCIDTPL ADLLLNWDHL FHFQSAKGAK AFPLAEFRLW HRQLLRAATV PLWQDTCCKRA VRRFLQSWSL LIVQAARGFQ QWAHRGV PDM RYHASTMTPG PGGEVVFCLN QIRWAPLSMG KDRDEAAAAE FWSIVRAKTV	GERDFLITVR PDLFRWLARP PVAHPRSAQW SFSLAPSDQL RGRIRDRVDA SKHADSVQAG AVHERSFTLE QKGERDAYLT ARLYRTEFGA AGRASGDIRR RNEFGYVWQK VGMQGEIYGI IRCHPLRKRE ANGLSRIHAD KRQAGALGGF DEELQGLEEE GRLRSHLDAQ	MTRGWDDLRE ENHHVWADGH EAEGGSNLRN QGVVLTQKQK GDIGSAFLKL LRVLSIDLGV LPGENVGAAG DLREAWSAKE VVSEWRSRTR LDRERGGVFA HAPCHVILFE QDRRDPSAR FEDQGFLELL INAAQNLRQR GYLEPTGHDS LLERSGERV AEASYAVAAG	KWRRSGDKGQ ADAVGVLARV YQLEAVGGEL QQKITYCTNM ALDVAHVLPD RSFATCSVFE QQWRAQADAE LWPFEASLLS SREDRKYAGK KDLLDHIDAL DLSRYRMRTD KHARQPLAAF KRENEGLDLN FWTQHGDAFR GSCQWRKTTE FFRDPGSGVVL L	
Opitutaceae bacterium WP_009513281.1	MSLNRIYQGR YLVALLALAD APYITPGNNA EGRSYWPKFC GSFDTYSIAT SLKKIPDDDS SATPPPKNAE GDNIHPTWKS LAKIPVKYTT LQRRITIRGFR SVQLFNELIQ DIDALKAPVK NTFTTEIAAR VPWLQPMMEA AALVEQLGNA DHFTVLGVDL ARMIRLPGED DELLGADPRR LDEIHAERAG LPLRGRREW EELRRRCQSL RVDQTAHAIL NLSRYLSSQD SRFSSRDGSA EARAVRGLFD LNAAINIALR PAVTLSPTNN TGRGLWASVK	VAAVETGTAL KNNPVLGLPI PTLDEVFRSI DPDSTANFAG PDTRTPQLTG RLNLQGYVGS TTPPAGVPLP FDIAAFKYAL GEAEPPPIILA ELRRIWRGHA NPKYWPIWQA LTPADPEYSR NAADGNRWRA LAPLPTLPQD GRWQNQFFGS GTRDAGALAL ARLFVRGKLV HSYPEINDKL EKPSPLPPLA RPHVEVPDCH NRALRHKPGE AAALGVRLRA RARSENTRLM GFRAVHLTPD RLDRFNAGHV GIAAPDRHDI GASPEDSDAL QRAWNRVARL	AKGNVEWMPA SQMDNPQSPY LAGNPTDRAT DPAMLRREQH PKARARLEQA SAKGEVQARL AASAADPVRI TVINQIEEKT NDLRIPLLRE PAGTVFSSEL PDVETARQWA RQYDFNAVSK THVRIHYSAP LTGMPVFLMP REDPFALRWP LNVTAQKPAK QEPYGERGRN LVALRRAQAR RDDAIKSTDE ILAQSDPGTD QWCHRQIVQK HRHRMPWSRI PGKPWRTLLA HHRLRAENKK PERVSNLFVD NETVTDNNRN	AGGDEVWQW HVWGSFRRQG LDAALMQLLK RLLLPQVLHD ITLWRVRLPE FALLLFRHLE ARGKRSFVFR KERQKECAEL LLQNIKVDTA KEKLAGELRQ DAGFADDPLA FGAGSRSANR RLLRDGLRRP DVTLSGERRI ADGAVKTAKG PVHRIIGED ASLLEWEDAR LARLQNRSWR ALLSQRDIIR DTKRLVAGQR EEIADPCPAL RDIHGEYERF LRLQLCETYGI LARLKAHEED PLPGGPVFVP RILSLRLGTQ IAGVANFERV EEEDDIPM	HELFAAAINY RQRTGLSQAV ACDGAGATQQ PAITHDSPAL SAADFDRLAS RSSFTLGLLR AFTSLPCWHG ETDFDYMGR LTDGEAVSYG FQTDNSTTIG ALVQEAELOE HEPGQTERGH DTDGNEALEA LLNLPVTLPE KTHIPWHQDR GRTWYASLAD NIIILRLQNP LRDLAESDKA RSFVQIANLI GISHERIEQI LEKINRLRDQ RAPADFVVIE PVLAVPAAYS GKRLEKTVLD LGDATPMQAD REKARWPGGA TIEGVSQKFA	(SEQ ID NO: 191)
Thermomonas hydrothermalis WP_072754838.1	MSEKTTQRAY GDWLLTLRGG EDEHGAPKEF CGPSLKAHIR AGIGDGKDKD AYEKIAKWS SATREYLKTL VLKDVENSCE RQFESDAQKL EGWSYVQAW AICVWRDQEG DFGNSRWSIQ NLHWSKRLT EKEWNGRLQA	TLRLNRASGE LCHTLVEMEV IVATGRDSAD EDAVWVNRRA DAEGPARQGE QAQNGDNGKA DKKSTVTQED LTYLQDNSPA KNLQERAPSA AEASCDTEDE TQNPSILIDY FAIHKEIRDR ADLALDQNP PRAELDRIAK	CAVCQNNSCD PAKGNPPQR DRAKKVEEKL LFDAAVERIK KAKDLVQKAG TIEKLACALR LNQLRKLAD RHREFSVMLD VEWLDRFCES RIAAARKVQA VTGKTAHNQ DKGAKQDTRQ PNPTEVTRAD LEEQKTEQA	CWHDALWATH PTDQERRDRR REILEKRDFQ TLTWEEAWDF QWLSARFGIG PSEPPTLDTV DARNCRKKVG HAARRVSMHA RSMTTGANTG DPEIEKFGDI KRFKVPAYRH LQNRHGLKMR RLGRAASSAF EKLRKRLRWY	KAVNRGAKAF VLLALSWLSV EHEIDAWLQD LEPFFGTQYF TGADFMSMAE LKCISGPGHK KKGKKPWAD SWIKKAEQRR SGYRIRKRAI QLFEALAADE PDELHHPVFC LWNGRSMTDV DHVKIKNVFN VSFSPCLSPS	(SEQ ID NO: 192)

Table 6 - Cas12b (C2c1) orthologs						
	GPFIYVAGQH LGHRFAAACA KRRTVVYRRI LWTVHKLEVE EPSAETDEKE DYKPMPPGGQK DWEDNEAKKL AKALAENDQL SIRHVGGLSI NRRLLAARDR TPCHAVVIES FLEVPANYTS DAKDRFLVDL RRAIQADLNA FNDVRSLEPTG RVIELLRRHA	NIQPKRSGQY VWETLSSDAF GPDQLLDNTP VGRTPVLIDR GEIRSISRVS YYFHEAKEAS WQNHIATLPN RQHLDHTWKE TRINTISGLY LREQRVKQLA LKTYPDDLRL RQCSRTGLPG YDHLNNLQSK AANIGLRALL DNSSRRAPRE GLPTS	APHAQANKGR RREIQGLNVL HPAPWARLDR MVRSGFGKTE DELMSSALGT KNDDETKRRD YQTPPEISAE RWESDDQQWK QILKAFKMRP SRIIEAALGV TRRENRLMQ IRCDDVPTGD GEALPATVRV DPDWRGRWWY IENLWRDPSG	ARLAQLILSR AGGSGEGDLF QFLIKLQGED KQKERLKKLR LRLALKRHGN NQIEFLQDAL LKRVERNKKR ERLRSCLKWI EPDDLKKNIP GRIKIPKNGK WSSAKVRKYL FLKAPWRRRA PRQGGNLFIA VPCKDGTSEP DSLESGTWSP	LPDLRILSVD LHVEMTGDDG EGVREASNEE ELGWISAMPN RARIAFAMTA SLWHDLFSSP KENRDKLRTA FPRGKAEDNP QKGDDELENF LPKPRPTTVD KEGCELYGLH INTAREKNNG GAQLDDTNKE ALDRIEGSTA TRAYWDTVQS	
Methylobacterium nodulans WP_043747912.1	MYEAIVLADD ASDPIDPAVL ARKLGE LRKE DRLAFRLAVG TLREYERARK EALHAI IATE NAMERLVERS QITLPLLKAA	ANAQLANAFI AEANAWLDTD AASGTSATIK HLLSWESWCT EQIAQLGLPM QTRKRGRFGD RDTALMTLPD DDGRCIDTPL	GPLTDPNSAG AGRAWLVDTG ALKRDFGVLP RARDEHTARV GERDFLITVR PDLFRWLARP PVAHPRSAQW	FLEAFNKVDR APPRWRSLAA LFQPSLAPRI QRLEQFSSAH MTRGWDDLRE ENHHVWADGH EAEGGSNLRN	PAPSWLDQVP KQDPIWPREF LGSRSSLTFW LKGDLATKVS KWRRSRGDKGQ ADAVGLVARV YQLEAVGGEL	(SEQ ID NO: 193)
Chloracidobacteriu m thermophilum WP_058868187.1	MPQQA KPPVT LTLRGGLDHT GAPAGLI IAF LSAAIRDDAV RDAYLAPAKG ATVYEAIKAW PGYKSATRNF YSNSILEKVE EAERRKFEEF GWKEVVAEWS VCVWHKDGDA FCDFGKSRWD RDLALGNNTG WNGRLQAPRQ WCAYAKIHGL YAAACAVWEA IYRRIGADTL QLEAALGRPV ELMSSAVRTM LQDALVLWHD RKKKQRENRE FKDWVFPRGN GGNGRRGNHA AETKEQFGQS RPDVRVDEPC CQLHGLHLRE AEAKQAANKG ADANS PAAKG KVDGSAVVNP SPLECIEFGE L	GRAYTLRLRG LADTPVKGGK GTEAAEERNR WVNRSAFDE SEDESSEAKQ DGKASLEMA LNQLAAQTIV SVCGFYTLQD AKKIDQVPEA KSDCKTVEDR AKAPDPQPLI ICFDVHKNMQ DAGASEVTRA QLEATARYVE KEDPQYWPHA LSTEAFQREI PDGKPHPAPW SLIDRLVASG RLALKRHGDR LFSSRGWRDD KLYDVAKALA HASDPTIRKR SDPTIRKVG ALDALEHLRE HAIVIEDLTH VSASYTSRQD DARERLLCDL IQADLNAAAN DQPLRQSAQA WKEYAAYQNE	ADSNDPSWRD GKPD PDPTDE KVVAAL EEL AVESIGSSGS EDQAKDLVQK DKAIADLATA TQQDFVSLKD GGPARHSEFA AKDWLDRFCL IAAARALQDD DYALAAEAEF TPFPRALCLT DRLGRAASRA KHDWDQKAEK DTNKNRKGHA KGRITILRGRT ARLDRQFLIK WGGSDKQKAR ARIAHYLITD TAKQLWNAHV QDVTLREALH QLINPSGGNG LSLPLRLATLT QRVKQLASRI YRPEETRTRR SRTGAPGVRC NARWKDRTAA IGLRALTDPD QSGDAAKDN VQCRVIRILK	ALWQTHEAVN ERKARRILLA KSRGVDQNEI SGSSLTREEP AGQWLSSRF LSEFN PASND KANND AQECK VILDHAARRV ERSGVSGALE PEIDKFGDIQ KKRHFKVPAY LWTGSEMKRI ASNVTKSDVV MRNAIQWLVT RLILSRLPGL DGNALYCHTR LQGE EGVRE LEGLKQLGWD EKTTPGGIKE AKLHGYKAP DAWKRWEND RRGNHASDPT EFRRKVQVGF AEAALGVGRV ENRQLMTWSS QDVPVKEFMR DWEKAGAVRI WAGKWWYVPC GNKGAGKSKE EQIKGRDKQP	RGAQAFGDWL LSWLSVESKL NAWKKDCSAS WDMLEFFGS TGKGADFRM LQGV LGLISG QNTGSKGQRP SLAHTWIKLA PYRIRRAVD LFEALAEDDA RHPDALLHPI PLCWQSKRLA NIAGLFEQAD FSARLQPPGP RVLAVDLGH HKANGKEVRT ASNEE I WAVH PADKPSLSVD TLDEKGRIDL EPGEDSSGAE DERWKKQLRW IRKRLINPS FTRLKPDGTR RRPVEGKDPK SKVKKYLAEA SPFWRKQVKQ PLQGGEIFVS DPASFPRVDR VVNLWRDISS HEGSKEDDIP	(SEQ ID NO: 194)
Desulfovibrio inopinatus WP_027186183.1	MPTRTINLKL YRTVDNEGKE LYERLVPSVN MKLKEEKAPG	VLGKNPENAT AEIPRHAVQE ENNEAGDAQA WEAASQIWIQ	LRRALFSTHR EALAFAKAAQ ANAWVSPLMS SDEGQSLLNK	LVNQATKRIE RHNGCISTYE AESEGGLSVY PGSPPRWIRK	EFLLLCRGEA DQEILDVLRQ DKVLDPPPWW LRSGQPWQDD	(SEQ ID NO: 195)

Table 6 - Cas12b (C2c1) orthologs						
	FVSDQKKKQD	ELTKGNAPLI	KQLKEMGLLP	LVNPFRRHLL	DPEGKGVSPW	
	DRLAVRAAVA	HFISWESWNH	RTRAEYNSLK	LRRDEFEAAS	DEFKDDFTLL	
	RQYEAKRHST	LKSIALADDS	NPYRIGVRS	RAWNRVREEW	IDKGATEEQR	
	VTILSKLQTQ	LRGKFGDPL	FNWLAQDRHV	HLWSPRDSVT	PLVRINAVDK	
	VLRRRKPYAL	MTFAHPRFHP	RWILYEAPGG	SNLRQYALDC	TENALHITLP	
	LLVDDAHGTW	IEKKIRVPLA	PSGQIQDLTL	EKLEKKKNRL	YYRSGFQQFA	
	GLAGGAEVLF	HRPYMEHDER	SEESLLERPG	AVWFKLTLDV	ATQAPPNWLD	
	GKGRVRTPPE	VHHFKTALSN	KSKHTRTLQP	GLRVLSVDLG	MRTFASCVSF	
	ELIEGKPETG	RAFPVADERS	MDSPNKLWAK	HERSFKLTLP	GETPSRKEEE	
	ERSIARABIY	ALKRDIQRLK	SLLRLGEEDN	DNRRDALLEQ	FFKGWGEEDV	
	VPGQAFPRSL	FQGLGAAPFR	STPELWRQHC	QTYYDKAEAC	LAKHISDWRK	
	RTRPRPTSRE	MWYKTRSYHG	GKSIWMLEYL	DAVRKLLLSW	SLRGRTYGAI	
	NRQDTARFGS	LASRLHHIN	SLKEDRIKTG	ADSIVQAARG	YIPLPHGKGW	
	EQRYEPCQLI	LFEDLARYRF	RVDRPRRENS	QLMQWNHRAI	VAETTMQAEI	
	YGQIVENTAA	GFSSRFHAAT	GAPGVRCRFL	LERDFDNDLP	KPYLLRELSW	
	MLGNTKVESE	EEKLRLLEK	IRPGSLVPWD	GGEQFATLHP	KRQTLCVIHA	
	DMNAAQNLQR	RFFGRCGEAF	RLVCQPHGDD	VLRLASTPGA	RLLGALQQLE	
	NGQGAFELVR	DMGSTSQMNR	FVMKSLGKKK	IKPLQDNNGD	DELEDVLSVL	
	PEEDDTGRIT	VFRDSSGIF	PCNVWIPAKQ	FWPAVRAMIW	KVMASHSLG	
Desulfonatronum thiodismutans WP_031386437.1	MVLGRKDDTA	ELRRALWTTT	EHVNLAVAEE	ERVLLRCRGR	SYWTLDRRGD	(SEQ ID NO: 187)
	PVHVPEQVA	EDALAMAREA	QRRNGWPVVG	EDEEILLALR	YLYEQIVPSC	
	LLDDLKPLK	GDAQKIGTNY	AGPLFDSBTC	RRDEGKDVAC	CGPFHEVAGK	
	YLGALPEWAT	PISKQEFDGK	DASHLRFKAT	GGDDAFFRVS	IEKANAWYED	
	PANQDALKNK	AYNKDDWKKE	KDKGISSWAV	KYIQKQLQLG	QDPRTEVRRK	
	LWLELGLLPL	FIPVFDKTMV	GNLWNRLAVR	LALAHLLSWE	SWNHRVQDQ	
	ALARAKRDEL	AALFLGMDG	FAGLREYELR	RNESIKQHAF	EPVDRPVVVS	
	GRALRSWTRV	REEWLRHGD	QESRKNICNR	LQDRLRGKFG	DPDVFWHLAE	
	DGQEALWKER	DCVTSFSLN	DADGLLEKRR	GYALMTFADA	RLHPRWAME	
	APGGSNLRTY	QIRKTENGLW	ADVLLSPRN	ESAAVEEKT	NVRLAPSGQL	
	SNVSFDQIQK	GSKMVGRCRY	QSANQQFEG	LGGAEILFDR	KRIANEQHGA	
	TDLASKPGHV	WFKLTLDVRP	QAPQGWLDGK	GRPALPPEAK	HFKTALSNNK	
	KFADQVRPGL	RVLSVDLGVR	SFAACSVFEL	VRGGPDQGT	FPAADGRTV	
	DPEKLWAKHE	RSFKITLPG	NPSRKEEIR	RAAMEELRSL	NGDIRRLKAI	
	LRLSVLQEDD	PRTEHLRLFM	EATVDDPAKS	ALNAELFKGF	GDDRFRSTPD	
	LWKQHCHFFH	DKAEKVVAER	FSRWRTETRP	KSSSWQDWRE	RRGYAGGKSY	
	WAVTYLEAVR	GLILRWNMGR	RTYGEVNRQD	KKQFGTVASA	LLHHINQLKE	
	DRIKTGADMI	IQAARGFVPR	KNGAGWVQVH	EPCRLILFED	LARYRFRFDR	
	SRRENSRLMR	WSHREIVNEV	GMQGEYGLH	VDTTEAGFSS	RYLASSGAPG	
	VRCRHLVEED	FHDGLPGMHL	VGELDWLLPK	DKDRTANEAR	RLLGGMVRPG	
	MLVPWDGDEL	FATLNAASQL	HVIHADINAA	QNLQRRFWGR	CGEAIRVCN	
	QLSVDGSTRY	EMAKAPKARL	LGALQQLKNG	DAPFHLTSIP	NSQKPENSIV	
	MTPTNAGKKY	RAGPGEKSSG	EEDELALDIV	EQAEELAQGR	KTFFRDPSPG	
	FFAPDRWLPS	EIYWSRIRRR	IWQVTLEARN	SGRQERAEMD	EMPY	
Tuberibacillus calidus WP_027726362.1	MATKSFILKM	KTKNNPQLRL	SLWKTHELFN	FGVAYYMDLL	SLFRQKDLYM	(SEQ ID NO: 196)
	HNDEDPDHPV	VLKKEEIQR	LWMKVRETQQ	KNGFHGEVSK	DEVLETLRAL	
	YEELVPSAVG	KSGEANQISN	KYLYPLTDPA	SQSGKGTANS	GRKPRWKKLK	
	EAGDPSWKDA	YEKWEKERQE	DPKLKILAAL	QSFGLIPLFR	PFTENDHKAV	
	ISVKWMPKSK	NQSVRKFDKD	MFNQAIERFL	SWESWNEKVA	EDYEKTVSIY	
	ESLQKELKGI	STKAFEIMER	VEKAYEAHLR	EITFSNSTYR	IGNRAIRGWT	
	EIVKKWMKLD	PSAPQGNLYD	VVKDYQRRHP	RESGDFKLFE	LLSRPENQAA	
	WREYPEFLPL	YVKYRHAEQR	MKTAKKQATF	TLCDPIRHPL	WVRYEERSGT	
	NLNKYRLIMN	EKEKVQFDR	LICLNADGHY	EEQEDVTVP	APSQQFDDQI	
	KFSSDGTGKG	KHNFSYYHKG	INVELKGTG	GARIQFDREH	LLRRQGVKAG	
	NVGRIFLNV	LNIEPMQPF	RSGNLQTSVG	KALKVYVDGY	PKVVNFKPKE	
	LTEHIKESEK	NLTGLGVESL	PTGLRVMSVD	LGQRQAAAT	IFEVVSSEKPD	
	DNKLFYPVKD	TDLFAVHR	FNKLPGEKR	TERRMLEQQK	RDQAIRDL	
	KLKFLKNVLN	MQKLEKTD	EKRVRNRWIK	REEREENPVY	VQEFEMISKV	
	LYSPHSVWVD	QLKSIHRKLE	EQLGKEISKW	RQSISQGRQG	VYGISLKNIE	
	DIEKTRRLLF	RWSMRPENPG	EVKQLQPGER	FAIDQQNHLN	HLKDDRICKL	

Table 6 - Cas12b (C2c1) orthologs						
	ANQIVMTALG	YRYDGKRKKW	IAKHPACQLV	LFEDLSRYAF	YDERSRIENR	
	NLMRWSRREI	PKQVAQIGGL	YGLLVGEVGA	QYSSRFHAKS	GAPGIRCRVV	
	KEHELYITEG	GQKVRNQKFL	DSLVENNIIE	PDDARRLEPG	DLIRDQGGDK	
	FATLDERGEL	VITHADINAA	QNLQKRFWTR	THGLYRIRCE	SREIKDAVVL	
	VPSDKDQKEK	MENLFGIGYL	QPFKQENDVY	KWVKGEKIKG	KKTSSQSDDK	
	ELVSEILQEA	SVMADDELKGN	RKTLFRDPSG	YVFPKDRWYT	GGRYFGTLEH	
	LLKRKLAERR	LFDGSSRRG	LFNGTDSNTN	VE		
Bacillus thermoamylovorans WP_041902512.1	MATRSFILKI	EPNEEVKKGL	WKTHEVLNHG	IAYYMNILKL	IRQEAIYEH	(SEQ ID NO: 197)
	EQDPKNPKKV	SKAEIQAELEW	DFVLKMQKCN	SFTHEVDKDV	VFNILRELYE	
	ELVPSSVEKK	GEANQLSNKF	LYPLVDPNSQ	SGKGTASSGR	KPRWYNLKIA	
	GDPSWEEKK	KWEEDKKKDP	LAKILGKLA	YGLIPLFIPF	TDSNEPIVKE	
	IKWMEKSRNQ	SVRRLDKDMF	IQALERFLSW	ESWNLKVKEE	YEKVEKEHKT	
	LEERIKEDIQ	AFKSLEQYEK	ERQEQLLRDT	LNTNEYRLSK	RGLRGWREII	
	QKWLKMDENE	PSEKYLEVFK	DYQRKHPREA	GDYSVYEFLS	KKENHFIWRN	
	HPEYPYLYAT	FCEIDKKKKD	AKQQATFTLA	DPINHPLWVR	FEERSGSNLN	
	KYRILTEQLH	TEKLKKKLT	QLDRLIYPT	SGGWEEKGKV	DIVLLPSRQF	
	YNQIFLDIEE	KGKHAFETYK	ESIKFPLKGT	LGGARVQFDR	DHLRRYPHKV	
	ESGNVGRIYF	NMTVNIEPTE	SPVSKSLKIH	RDDFPKFVNF	KPKELTEWIK	
	DSKGKKLKSG	IESLEIGLRV	MSIDLQQRQA	AAASIFEVVD	QKPDIEGKLF	
	FPIKGTLEYA	VHRASFNIKL	PGETLVKSRE	VLRKAREDNL	KLMNQKLNFL	
	RNVLFHQFQFE	DITEREKVRT	KWISRQENS	VPLVYQDELI	QIRELMYKPY	
	KDWVAFLLKQL	HKRLEVEIGK	EVKHWKSL	DGRKGLYGIS	LKNIDEIDRT	
	RKFLLRWSLR	PTEPGEVRR	EPGQRFADQ	LNHLNALKED	RLKKMANTII	
	MHALGYCYDV	RKKKWQAKNP	ACQIILFEDL	SNYNPYEERS	RFENSKLMKW	
	SRREIPRQVA	LQGEIYGLQV	GEVGAQFSSR	FHAKTGSPGI	RCSVVTKEKL	
	QDNRFKKNLQ	REGRLTLDKI	AVLKEGDLYP	DKGGEKFISL	SKDRKLVTH	
	ADINAAQNLO	KRFWTRTHGF	YKVYCKAYQV	DGQTVYIPES	KDQKQKIEE	
	FGEGYFILKD	GVYEWGNAGK	LKIKKGSSKQ	SSSELVDSDI	LKDSFDLASE	
	LKGEKLMLYR	DPSGNVFPD	KWMAAGVFFG	KLERILISKL	TNQYSISTIE	
	DDSSKQSM					
Bacillus sp. NSP2.1 WP_026557978.1	MAIRSIKLLK	KTHTGPEAQN	LRKGIWRTHR	LLNEGVAYYM	KMLLLFRQES	(SEQ ID NO: 198)
	TGERPKEELQ	EELICHIREQ	QQRNQADKNT	QALPLDKALE	ALRQLYELLV	
	PSSVGQSGDA	QIISRKFLSP	LVDPNSEGGK	GTSKAGAKPT	WQKKKEANDP	
	TSSQDYEKWK	KRREEDTPAS	VITLLEEYGI	RPFIPLYTNT	VTDIWLPLQ	
	SNQFVRTWDR	DMLQQAIRL	LSWESWNKRV	QEEYAKLKEK	MAQLNEQLEG	
	GQEWISLLEQ	YEENRERELR	ENMTAANDKY	RITKRQMKGW	NELYELWSTF	
	PASASHEQYK	EALKRVQQR	RGRFGDAHFF	QYLMEEKNRL	IWKGNPQRH	
	YFVARNELTK	RLEEAKQSAT	MTLPNARKHP	LWVRFDARGG	NLQDYILTAE	
	ADKPRSRFV	TFSQLIWPSE	SGWMEKKDVE	VELALSRQFY	QQVKLLKNDK	
	GKQKIEFKDK	SGSGTFNGHL	GGAKLQLERG	DLEKEEKNFE	DGEIGSVVLN	
	VVIDFEPLQE	VKNRQVQAPY	GQVLQLIRRP	NEFPKVTTYK	SEQLVEWIK	
	SPQHSAGVES	LASGFRVMSI	DLGLRAAAAT	SIFSVEESSD	KNAADFSYWI	
	EGTPLVAVHQ	RSYMLRLPGE	QVEKQVMEKR	DERFQLHQRV	KFQIRVLAQI	
	MRMANKQYGD	RWDELDSLKQ	AVEQKKSPLD	QTDRTFWEGI	VCDLTKVLP	
	NEADWEQAVV	QIHRKAEEYV	GKAVQAWKR	FAADERKGIA	GLSMWNIEEL	
	EGLRKLISW	SRRTNPQEV	NRFERGHTSH	QRLLTTHIQNV	KEDRLKQLSH	
	AIVMTALGYV	YDERKQEWCA	EYPACQVILF	ENLSQYRSNL	DRSTKENSTL	
	MKWAHRSIPK	YVHMQAEPYG	IQIGDVRAEY	SSRFYAKTGT	PGIRCKKVRG	
	QDLQGRRFEN	LQKRLVNEQF	LTEEQVKQLR	PGDIVPDDSG	ELFMTLTDGS	
	GSKEVVFLLQ	DINAHNLLQK	RFWQRYNELF	KVSCRIVIRD	EEEYLVPKTK	
	SVQAKLGKGL	FVKKSDTAWK	DVYVWDSQAK	LKGKTTFTTE	SESPEQLEDF	
	QEIIIEEAEA	KGTYRTLFRD	PSGVFFPESV	WYPQKDFWGE	VKRKLYGKLR	
	ERFLTAKAR					
Alicyclobacillus acidoterrestris WP_021296342.1	MAVKSIVKVL	RLDDMPEIRA	GLWKLHKEVN	AGVRYYTEWL	SLLRQENLYR	(SEQ ID NO: 199)
	RSPNGDGEQE	CDKTAECKA	ELLERLRARQ	VENGHRGPAG	SDDELLQLAR	
	QLYELLVPQA	IGAKGDAQOI	ARKFLSPLAD	KDAVGGLGIA	KAGNKPRWVR	
	MREAGEPGWE	EEKEKAETRK	SADRTADVLR	ALADFGKPL	MRVYTDSEMS	
	SVEWKPLRKG	QAVRTWDRDM	FQQAIERMMS	WESWNQVRGQ	EYAKLVBQKN	
	RFEQKNFVGQ	EHLVHLVNQL	QQDMKEASPG	LESKEQTAHY	VTGRALRGSD	

Table 6 - Cas12b (C2c1) orthologs						
	KVFEKWGKLA WREDASFLTR LHQYTFLFNE LPRDPNEPIA YLNVSVRVQS DDGKLGSEGL FFPIKGNNDNL YLRLLVRCGS SLHGICSDKE DVVGNSIEQ IDHAKEDRLK SEYQFNNDRP FDARTGAPGI LIPTGEGEIF CDWGEVDGEL FAQEKLESEE NQRIEGLVK	PDAPFDLYDA YAVYNSILRK FGERRHAIRF LYFRDYGAEQ QSEARGERRP LSGLRVMSVD VAVHERSQLL EDVGRRRERSW WMDAVYESVR IEYLERQYKF KLADRIIMEA PSENNQLMQW RCRRVPARCT VSPFSAEEDG VLIPRLTGKR AELLVEADEA QIRSRVPLQD	EIKNVQRRNT LNHAKMFATF HKLLKVENG HFTGEFGGAK PYAAVFRLVG LGLRTSASIS KLPGETESKD AKLIEQPVDA RVWRHMGKQV LKSWSFFGKV LGYVYALDER SHRGVFQELI QEHNPEPFPW FHQIHADLNA TADSYSNKFV REKSVVLMRD SACENTGDI	RRFGSHDLFA TLPDATAHPI AREVDDVTVP IQCRDQLAH DNHRAVHFD VFRVARKDEL LRAIREERQR ANHMTPDWRE RDWRKDVRSG SGQVIRAEKG GKGKWKVAKYP NQAQVHDLV WLNKFVVEHT AQNLQQRLWS YTNTGVTYYE PSGIINRGNW	KLAPEPYQAL WTRFDKLGGN ISMSEQLDNL MHRRRGARDV KLSDYLAEHP KPNSEGRVPF TLRQLRTQLA AFENELQKLL ERP KIRGYAK SRFAITLREH PCQLILLEEL GTMYAAFSSR LDACPLRADD DFDISQRLR RERGGKRRKV TRQKEFWSMV	
Alicyclobacillus hesperidum WP_074693942.1	MTVRSIRVKL YDEDEHGLTV QLYEQLVPSS LLKKLSGDPT LEVSWAPKKA KKFYDENFAT DRVADEWCKL LWRSDDPTKIL NIYSYRLLIP RLHFSIMEDG GECGPVVLNV WKSQGDGTL AGTDEFVAVQ LLRLCDTKTP NWKQKVISVH LEQERTLVA DLLIMTALGY NSQLMAWSHR SVKKQDIENA CDRQKGKIV RISNAFADSF AEAETWIDEA LLSRMSEREA	AVGSPQYRDV VQRTREDIQA VDKNKSGDAR WEEDYKKAME RQWVRKWDYD DPEFIKLAET ADDAPIEYE RFARLRLALQ EKRSKRYVE SLCVQYRDPG SVDVNPAPHR LRVMSVDLGV ERAFLLRLPG DERQBALHGL RQMERILGHA WSNHAREPGQ VYDETKPSGH RLEILKLQA AQKGWLARE HADLNAAQNL YKVFQGQVVF REQQKVIAL AAHKE	RRGLWKTHEI ELLSRLRTLQ MIARNFFNPL QKQESSVSFL LFQQSIERML LEGELQASSQ AAIKRVQARL AVARAKRHAR FSSLIMPDGE VDEPLRAELG DVQVLQSAKL RSSAAVVICR EGKGTNEDAP AQSLPSGAW VREWKVARKN VVRICKRGTV KWDKRYPPCQ DLHKLIVGTV LQRLNWTLEW QRRFWGGHAE STDEEDVYRW FRDASGQIHG	MNQGVRYICE SAHQHSGDMG TNPNSQGGG LLELRRFGLH SWESWTRRVK GFVAVDEHAF GRNFGSYVLF LRLPDIAHHP NRWAEHRNIR GAKIQFDRRY VSVSRDTNRI LEHRDSVVSS LRDVYAQLGT KDELHPLHVM AGKPPIRRGA AQWLVERVNH IILMEDLSRY FPAFSSRFDA LQPNDLIPTG SLCRVTCVV MVGEKISSRG GDWLVAKVFW	WLVLMRQEP TDEELLSLMR ISNAGRPKW PIFLPYTDTV ERFEKLVESE QIRPRSMRGF AHLAKPEYWS IWIRYDAK GK VPLAFSRQWE LIRRSSTLSA YLRPENLSAY GRRTATYRI IRQGIQILRS LQGVVHDSVD GGLSLRRIRQ LKEDRLKKLA RFQSDRPPSE QSGAPGVR CR DGELFVTPAC ERDGRRYAVP RSRGRTSDEE GWVERLVTAR	(SEQ ID NO: 200)
Alicyclobacillus acidiphilus WP_067623834.1	MAVSKMKVKL RSPNGDGEQE QLYELLVPQA MREAGEPGWE SVQWKPLRKG RFEQKNFVGQ KVFEKWEKLD WREDASFLTR LHQYTFLFNE LPRDPHELVA YLNLSVRVQS DDGKLGSEGL CFPIEGNENL YLRLLVRCGS LYGICGDREW VVGNSIEQI DHAKEDRLKK SEYQFNNDRP FDARTGAPGI	RLDNMPEIRA CYKTAECKA IGAKGDAQI EEKAKAEARK QAVRTWDRDM EHLVQLVNQL PDAPFDLYDT YAVYNSIVRK FGGRHAIRF LYFQDYGAEQ QSEARGERRP LSGLRVMSVD VAVHERSQLL EDVGRRRERSW TEAVYESVRR EYLERQYKFL LADRIIMEAL PSENNQLMQW RCRRVPARCA	GLWKLHTEVN ELLERLRARQ ARKFLSPLAD STDRTADVLR FQQAIERMMS QQDMKEASHG EIKNVQRRNT LNHAKMFATF QKLLTVEDGV HLAGEFGGAK LGLRTSASIS KLPGETESKD AKLIEQPMDA VWRHMGKQVR KSWSFSGKVS GYVYALDDER SHRGVFQELL REQNPEPFPW	AGVRYYTEWL VENGHCGPAG KDAVGGLGIA ALADFGKPL WESWNQVRGE LESKEQTAHY RRFGSHDLFA TLPDATAHPI AKEVDDVTVP IQYRRDQLNH DNHRAVHFD VFRVARKDEL LRAIREERQR NQMTPDWREA DWRKDVRSGE GQVIRAEKGS GKGKWKVAKYP NQAQVHDLV WLNKFVAEHK	SLLRQENLYR SDDELLQLAR KAGNKPWRV MRVYTDSDMS AYAKLVEQKS LTGRALRGSD KLAEPKYQAL WTRFDKLGGN ISMSAQLDDL LHARRGARDV KLSDYLAEHP KPNSEGRVPF TLRQLRTQLA FEDELQKLKS RPFKIRGYQKD RFAITLREH PCQLILLEEL GTMYAAFSSR LDGCPLRADD	(SEQ ID NO: 201)

Table 6 - Cas12b (C2c1) orthologs						
	LIPTGEGEFF	VSPFSAEEGD	FHQIHADLNA	AQNLQRRLWS	DFDISQIRLR	
	CDWGEVDGEP	VLIPTTGTGR	TADSYGNKVF	YTKTGVTTYE	RERGKKRRKV	
	FAQEELSEEE	AELLVEADEA	REKSVVLMRD	PSGIINRGDW	TRQKEFWSMV	
	NQRIEGYLVK	QIRSRVRLQE	SACENTGDI			
Alicyclobacillus macrosporangiidus SFU30094.1	MAVKSIKVKL	MLGHLPEIRE	GLWHLHEAVN	LGVRYYTEWL	ALLRQGNLYR	(SEQ ID NO: 202)
	RGKDGAQECY	MTAEQCRQEL	LVRRLDRQKR	NGHTGDPGTD	EELLGVARRL	
	YELLVPQSVG	KKGQAQMLAS	GFLSPLADPK	SEGGKGTSKS	GRKPAWMGMK	
	EAGDSRWVEA	KARYEANKAK	DPTKQVIASL	EMYGLRPLFD	VFTETYKTIR	
	WMPLGKHQGV	RAWDRDMFQQ	SLERLMSWES	WNERVGAEFA	RLVDRDRDRF	
	EKHFTGQEHL	VALAQRLEQE	MKEASPGFES	KSSQAHRIK	RALRGADGII	
	DDWLKLEGE	PVDRFDEILR	KRQAQNPRRF	GSHDLFLKLA	BPVFQPLWRE	
	DPSFLSRWAS	YNEVLNKLED	AKQFATFTLP	SPCSNPVWAR	FENAEGTNIF	
	KYDFLFDHFG	KGRHGVRFQR	MIVMRDGVPT	EVEGIVVPIA	PSRQLDALAP	
	NDAASPIDVF	VGDPAAPGAF	RGQFGGAKIQ	YRRSALVRKG	RREEKAYLCG	
	FRLPSQRRTG	TPADDAGEVF	LNLSLRVESQ	SEQAGRRNPP	YAAVFHISDQ	
	TRRVIVRYGE	IERYLAEHPD	TGIPGSRGLT	SGLRVMSVDL	GLRTSAAISV	
	FRVAHRDEL	PDAHGRQPPF	FPIHGMDHLV	ALHERSHLIR	LPGETESKKV	
	RSIREQRLDR	LNRLRSQMAS	LRLLVRTGVL	DEQKRDRNWE	RLQSSMERGG	
	ERMPDWWDL	FQAQVRYLAQ	HRDASGEAWG	RMVQAAVRTL	WRQLAKQVRD	
	WRKEVRRNAD	KVKIRGIARD	VPGGHSLAQL	DYLERQYRFL	RSWSAFSVQA	
	GQVVRAERDS	RFAVALREHI	DNGKKDRLKK	LADRILMEAL	GYVYVTDGRR	
	AGQWQAVYPP	CQLVLEELS	EYRFSNDRPP	SENSQLMVWS	HRGVLEELIH	
	QAQVHDVLVG	TIPAAFSSRF	DARTGAPGIR	CRRVPSIPLK	DAPSIPIWLS	
	HYLKQTERDA	AALRPGEIIP	TGDGEFLVTP	AGRGASGVRV	VHADINAAHN	
	LQRRWLWENFD	LSDIRVRCDR	REGKDGTVVL	IPRLTNQVRK	ERYSGVIFTS	
	EDGVSFVTVD	AKTRRRSSAS	QGEGDDLSD	EQELLAEADD	ARERSVVLFR	
	DPSGFVNGGR	WTAQRAFWGM	VHNRIETLLA	ERFSVSGAAE	KVRG	
Sulfobacillus thermosulfidooxidans PSR34340.1	RQSREDASPO	IIISASDLKA	DLLYHARQQQ	KEHVPRTIGS	DAEVLGALRQ	(SEQ ID NO: 203)
	VYELIVPSSV	GKSGDSKTIA	RKFLSPLTDP	DSAGGRDQSA	SGRKPTWTM	
	KAEGNPLWEE	KFRQWKDRKD	NDPTPFVLNQ	LADYGLLPLI	RLFTDVGENI	
	FDPKPKPGQFV	RPWDRSMFQQ	AIERLMSWES	WNQVRVQWE	ALTQKHSIFY	
	REQFTAEPDA	ALYRVAQSLE	EEMRKEHQGF	ATDAPEAFRI	RRVALKGFDR	
	LLERWQKTLG	KNGQSATLLD	DIRRVQSDLG	DKFGSAPLYQ	KLVDERWQRL	
	WTVDPTFLQR	YAAFNDLTQR	LQRAKRVANL	TLPDVAHAHPI	WSRYEGPNAS	
	SGNRYHIHLR	TTGQPSSVTF	DRILWPDGDG	GWYERKRVTV	FLRPSHQVDR	
	IREAPTDSVV	DNFPLVVEDQ	SARTILRASW	GGAKLEYDRN	RLPRQLKKGV	
	PDSIYLSLTL	NLDTTKPSGL	FHMQQNGRVW	IRKDVVMQYY	NEIPGDNVQF	
	KPLYVMSVDL	GIRSAVAISI	FSVQLKTGIE	EHRLTYPVAD	CPGLVAVHER	
	SVLLTMPGER	REQDRRYEQ	QRQGLRELRT	DMRGMNDLLR	GAYVDGDRRE	
	EFLARLSKLE	ETSPELWEPV	YRSLNDSKMA	PAAEWERLVV	YCHRVQEQSL	
	SSRIQNLRSR	RSAYRMSGGL	SLDHVQDLER	IRGIIASWTN	HPRIPGSVVR	
	WQQGRSHTVA	LGRHILELKR	DRVKKVANYL	IMTALGYAYD	SKRARGEKWV	
	RRYPSCHLMV	FEDLTRYRFR	TDRPRSENRO	LMRWTHQELI	AVTGIIQAEPH	
	GILVGTMYAG	FSSRFDAVTK	APGVRGATVR	QILRTRGMVR	LKEIAADVGV	
	DINTLRPHDV	LPTGDGEYLL	SVVRHRDSYR	LKQVHADINA	AHNLQRRLWT	
	QDEVFRVSCR	LALNSERVVA	TPPPSYNKRY	GKGFFEKGDN	GVYIWKTGK	
	IKISDMLEED	MDIPEDTAEI	LRGNSVTLFR	DPSGTIAGGN	WLEAKEFWGR	
	VNSLVNKGVR	DKILGGIPVD	NSSAHAE			
Spirochaeta sp. LUC14_002_19_P3 OQX29950.1	MGLLLPSSLR	TVNVTIHLIL	HPRKKGSRHR	EYAVMLDHAV	RKIFLAHNWI	(SEQ ID NO: 204)
	KRAEAERQKF	EADLYKIDRV	PQEARDWLDE	FCRERTESTG	SIDGYHIRRK	
	AVLGWEALVE	AWDQKDCLSV	EDRIAAARDL	QDNPGMDKFG	DIWLIEALAS	
	APCVWQKDGE	PNAQILLDYV	DAGEAEYKRS	HYKVPAYRHP	DPLLHPIPCD	
	FGQSRWSISF	DIHEFKKNGE	KNPVNIHALT	MGLVSKKRIV	KTELKWSSKR	
	LNSNLALSLE	SPEDAIEVSR	ATRLGRAAVG	ASQDRAVNIA	GLFESAGWNG	
	RLQAPRKQLE	ALAKLEEDKS	AEALAKALRN	RIKWFITFSP	KLQPHGPWME	
	YAERFSGEAP	SRAAVIKGKY	TVIHQDKTRR	RPLAKLHLCR	MPGLRVLSVD	
	LGHRHAAACA	VWETLSSES	EKKCREAGCL	PPAPEDLYLH	LKKKNKTAVY	
	RRIGGNFLPD	GNEHPAPWAK	LDRQFIIDLQ	GEEGCTRMAL	AGEIHWQVHC	
	EKVFGRSIPL	VDRLVRAGWG	EKNKQPEILQ	ELKQKGWVPL	EVSKTNTGYH	

Table 6 - Cas12b (C2c1) orthologs						
	YSLCVDSLMT DFIVEALMLW AKIIKQKERK ELRSLRDWIL PEGEPMDGTM KPENNKTPKR KVKKYLFESC PFRQREIAHA EIFVSALTGN PHSKCSGAAC KSTINLWRDI	LAVNTVRFAL YELATDSRWN ELLKKEFIPL PKGTRGKKRT AVGEGFGQKI PFTAVDEPCH QLHGLYLFV EENMAQENPC TLQADLNAAA LDNWRVGLKN SVLPLTEGQW	RRHACRARIA GSWEANFWDE AEGLENSRR IRHVGGLSLS LDDLETMKEQ AVVIENLTHY QASYTSRQDS NRYLIALHNK NIGLQALLDP NLYTGVRTPL	YYMEGGAIPE NFDKKLAEIQ ISIASQWRMV RLAVIKSLYR RVKQLASRVV RPENKRTRRE RTGAPGVRC WKNREYDKTA DWPGRWWYVP PGKNKGSTSG	GGLPENSGNK DAVNEREGDK WNEEDAIWQS VQKSFYTRMK EAALGTGRIK NRQLMTWSSS ELSVKKFLES PPLRI PHWGG AVKGC DGRR EDVHKSNAVE	
Bacillus hisashii strain C4 v4 mutant of WP_095142515.1 K846R S893R E837G	MATRSFILKI EQDPKNPKKV ELVPSSVEKK GDPSWEEEEK IKWMEKSRNQ LEERIKEDIQ QKWLKMDENE HPEYPYLYAT KYRILTEQLH YNQIFLDIEE ESGNVGRIYF DSKGGKLGSG FPIKGTETYA RNVLHFQQFE KDWVAFKQL RKFLLRWSLR MHALGYCYDV SRREIPRQVA QDNRFFKNLQ ADINAAQNLQ FGEGYFILKD LKGEKMLYR DDSSKQSM	EPNEEVKKGL SKAEIQAELEW GEANQLSNKF KWEEDKKKDP SVRRLDKDMF ALKALEQYEQ PSEKYLEVFK FCEIDKKKKD TEKLKKKLT KGKHAFTYKD NMTVNIEPTE IESLEIGLRV VHRASFNIKL DITEREKVRT HKRLEVEIGK PTEPGEVRR RKKKWAQAKNP LQGEIYGLQV REGRLTLDKI KRFWTRTHGF GVYEWVNAGK DPSGNVFPD KWMAAGVFFG	WKTHEVLNHG DFVLKMQKCN LYPLVDPNSQ LAKILGKLAE IQALERFLSW ERQEQLLRDT DYQRKHPREA AKQQATFTLA QLDRLIYPT ESIKFPLKGT SPVSKSLKIH MSIDLQORQA PGETLVKSRE KWISRQENS EVKHWRKSLS EPGQRFAIDQ ACQIILFEDL GEVGAQFSSR AVLKEGDLYP YKVYCKAYQV LKIKKGSSKQ KLERILISKL	IAYYMNILKL SFTHEVDKDE SGKGTASSGR YGLIPLFIPY ESWNLKVKEE LNTNEYRLSK GDYSVYEFLS DPINHPLWVR SGGWEEKGKV LGGARVQFDR RDDFPKVVNF AAASIFEVVD VLRKAREDNL VPLVYQDELI DGRKGLYGIS LNHLNALKED SNYNPYGERS FHAKTGSPGI DKGGEKFISL DGQTVYIPES SSSELVDSI FLRELPHRTF	IRQEAIYEH VFNILRELYE KPRWYNLKIA TDSNEPIVKE YEKVEKEYKT RGLRGWREII KKENHFIWRN FEERSGSNLN DIVLLPSRQF DHLRRYPHKV KPKELTEWIK QKPDIEGKLF KLMNQKLNFL QIRELMYKPY LKNIDEIDRT RLKKMANTII RFENSRLMKW RCRVVTEKEL SKDRKCVTTH KDQKQKIEE LKDSFDLASE TNQYSISTIE	(SEQ ID NO: 205)

Table 7- Cas12c (C2c3) orthologs						
OspCas12c AWU30132.1 KZX85786.1	MTKLRRHQKK SAYARATKGE GSLAHFLLSA IQNLTIISTLF RDLAEVISRK VLIKMNDLQG KNQNVGNVYK PSCHALIELI NSLSMDSEEL ADILLGSTQY DGEKIHLPA ALQKNFDLNF CLYRGSLVLR KLLRLTDSRP SSFINLPIT NKQSFVVTRT KDAGRLCVID	LTHDWAGSKK MTDGRKNMFT HALGFRIFSK NALTTSVRGK IASSFGTWKE DLKNRTIVFD NAITTTNANG AQLEAPELFE ERLIKSFQIH MLTNSLVEES WSELISLPFI NKALLNRTQH RAGIEVLKKH DLLHVIDEIL QKGDRLIDL LSPYLGSKLV TAKHLSNIKK	REVLGSNGKL HSFEPFKTKP SGEATAFQAS GEETSADPLI MTANPLQSLQ PDAPVFEYNA LGWLLNKGLS KNVFS DTRSE TPHCSLFIGA IATYQRTLNR FEAMCRSTKK KIFESNSELR QHDNLENKDR IQYDQINRDA YVPKDKDWLV SVFSSEEVL	QNPLLMPIVKK SLHQCELADK SKIEAYESKL ARFYTLTGTGK FFEEELHALD EDPADII IKL LLPVSTDEL VQGMIDSAVS QSLSQQLESL INYLSGVAGQ LAHLKNQYQT NALSKEPIVS EHVHERKHV ESLWLVRSY FVMLVTSFAK PSQMFEGRFA FLRELPHRTF	GQVTEFRKAF AYQSLHSYLP ASELACVDLS PLSRDTQGPE ANVSLSPA TARYAKEAVI LEFIGVERSH NHIARLSSSR PEALQSGVNS INGAIKRKAI LSNEFDLTLS YRDLLARLTS FVSPDRKAK LLAGLPDQLS SNLSGLQYRA DILQSDYMWV IQTEVRGLGV	(SEQ ID NO: 206)

Table 7- Cas12c (C2c3) orthologs						
	NVDGIAFNNG IQFERAYYKK PGEYGMGLSL SPYANYLEQS TKVLSFYTWG PGSQVSSYGN IKLFPNDPST SIRHSPEFIA EL	DIPSLKTFSN DDQIHEDA VSINNGEVL KDSAAGDIAH DNDAQNSIRK SQRSCCCGRN VIERRRHNLG KKRGIGSEYF	CVQVKVSRTN RKIRFQMPAT SGFIHINSLI ILDRLIYKLN QHWFGASHWD PIEQLREMAK PSRIPVADRT CAYSDCNSSL	TSLVQTLNRW ELVHASDDAG NFASKKSNHQ ALPVFEALSG IKGMLRQPPT DTSIKELKIR FKNISPSSLE NSEANAAANV	FEGGKVSPPS WTPSYLLGID TKVVRPQQYK NSQSAADQVW EKKPKPYIAF NSEIQLFDGT FKELITIVSR AQKFQKQLFF	
QFN42172.1	MRSNYHGGRN TYGIAEGVGH KLSAIMGIQL GRGPANKPDE QGDFFPSIASL LGRERPDLQ IDEYKVPADQ ANYWKRLDL LPARLVSAEE KQKLENLQDA EEKLQRLDLA DPEEYALRL LGSLYRSPYS LGDLQRLWIN EEDDIARDVC PKDKAWEYPD GAGVSEYLVQ MVGASSFKTH RLQVEAAVPV TGEVKPIHDN ALQNYRENV	ARQWRKQISG GSLIGLVCAF PTLSPEGLDL HLEIAGEIA PAAIMLSPAN NKGPFVSSLQ HGAVTQVKSF KSLLATTEFT AIDRLMGVGI DDEEFLKGLK RSEHFQTISE LQRIGRLANR TSRHQPFSID LAGFAISQRL LKAFNLVYSA RLNTAKGPIN APHDWYTPLD LDSVLLSDKI VDTRDRTVPE NGNPVVGTV	LARRTKETVF HLSGFRLFSK IFQSPPRS KQVFPKFGGW STVDFEGDYI DALVSSQNNG VDAIPLNPLF LPESISDPKA PTAADIAQVE IELPSGDKEP WAEENAVTLD VSPVSAGSIR VGKAKAIDWI RGLPDTVPNA INGCLFGALR AAVSSDWIEK LRDVAHLVTG KLGDFTHIID PDTLFDHIVA VPSIRRLMKA MERYNAFPVL	TYKFPLETDA AGEAMAFRNR GIAPVWSENE DDLASDPDKA AIDPAAETLL LSWLFVGVGFQ DTTHYGEFRA VSLFSGLLVD RVADEIGAFI PAINRISGGA PIAAMVELER ELLKPVFMEE AGLDQISSDI LAQVRCDDV EGFIVRTRFQ DGAVIKPVET LPVEKNITKL QHYRQSVTYG IDLGERSVGF VRSHRRRRQP EFQIKNFQAG	AEIDFDKAVQ SRYPTDAFAE VRNRLYTNT LAAADKYFQS HQAVERCAAR HWKEKSPKEL SVAGKVRSWV PQGLKKVADS GQVQQFNNQV PDAAAEISEL LRLAERGATG REFNLFFHNR EKALSGAGEA RIPPLLAMLL RIGTDQIHVY VRNLSSTGFA KRLTNRTAFR GKVKISYEPE AVFDIKSCLR NQKVNQTYST AKQLEIVYGS	(SEQ ID NO: 207)
QFN42158.1	MKKFELKQNF GKLPKYENLI KDIFKDNDYF VIVAEYRKEI TGCEIIDEIL KSDELLNPYF GLTLFQTEST QDFRSSLGGH EYSGCNRDEI DFSEIVNSLH LKKIPKLNKL GTIDIFESRF CEKIKNWFKE LEKRYEVTNT KEQHIPTKIA LLSGLSIRLS EYKDKQILIV WMFKLPFGAS DEIMMNDESE KKGNTESFPF IRRLIHSVKK YNAFPILKQ GNSWNTPILR CHVCGRNALE VKNLAKKGNR NAAINIGRRF	RNNYSGKTLR SLYDTIEDIK NQFKIQSHLD GKLKNKDESS NDSFGTLP AVATILKSMP EKLCAFMNVS LDSWTTNYLK QQMIDFVVNE SFVQQIDNSL SGGVPAWKE KYDELLKKS QNIFDSSKDF VGALLEQCES QPKLDDSTYQ RNSFYLRTKF NEEYDVVVK NAEKCKVLKL LSEMTELLADE KNIVSIDQGE YRGKKQRIQN VGNLESGSKQ ISNPNSNNK LLKNDDSTGK ERTYASINER LKDCILDDNK	NFRQTLAQIA KGTLSYYLFT LPDFVPSKIY EHQCEELFKK KKMVLASTQ PEIQQDKKSA DKRVIEQVQD RLDELNDLLL QNRIKLQESL EQSSNEANSI IREIEQKFHE KNNLQSADEL NRYFINQKGF DPAIVNDPFS ESVSPTLK SWIGNNSLIY TFESVYKIVK EKNKKFKPKL PVRQQMSNGK AGFAYAVFKL FNQKFDSTMF LMLVYKAVNS NIVKNINGKK VKKYQINQDG APMSKDTTQS EKD	SSDGEQD YVKANLTPT YAKAVKLPAE NLPKNLSLPD NALLGKGNQ FSELKKKIEK AFRSVLNKL IFKHPSSKKD MRSLVEFRAL LEKEQITSSE CPKETTWKIP SKDNNEKNRI SVSKDSLARL IEIIPDDYVM SDCGNERAEP TLRENTGDI KFLAAKVD YEELKIYPGY EVTIGGEVIK RYFCVFNCP	IKFKLDCSKT SASQAKAFST NGKDNFKAS SWQDLINNC IAYDPDSTFI HNALSWIFGK LDLNHCTLKF IFMIDGKDFI ICSDDISTVK NEKWDIWKN EVMWIDAGN RFRQGNLDV NSPYNLSANL WFSINISGIS LNSIFTVYKS AAYFKSDLWN LPLLKQLPHD SGPSTYFNRI SLAIPDQTSR IATGLIPIPS CGLIVALMCK NDQRRSWWYQ SVSAYMTSCI LYRKPDRLTP CHNKEQHADV	(SEQ ID NO: 208)
QFN42173.1	MNARDWRKHV GYGSLGLAC TITLTLPETL	GVLAQQHKET AVHLSGFRLF DRLFSSRPKR	TRTYTFPLDT STGKEAATFR RNGVPLPWNQ	TGSAIDFDA NRARYPNAAF DSIRDRLYN	LQAYNAVEGV QAALRKELGT WVKPRPGDTP	(SEQ ID NO: 209)

Table 7- Cas12c (C2c3) orthologs					
	DAVLFIQIATG	IAQEITEDVS	SWTDLAKNSD	RGLKAAHRYF	ARVGGFPAFD
	NLTPPATVQP	TDTTIDYDPN	APFHLVSHAD	QTLIHQSISL	CAHRIRQEDP
	ALDPNKSQFI	KQLQNNFLSQ	TFYGLSWLFG	AGYVHFRECT	ANDLAIQYGI
	PNNCRDGIHQ	IKSFADAILP	NTFFEKKHYR	KDSRSVGKKA	KSWISNYWQR
	LLQLQTWVDD	HTWVTLPOEL	TEAQFKPLFR	GLLVDAVELM	AIAERLPQRL
	ADCRDSDLCL	MGKGPQAATK	NDVEIVEKVR	EEIESFVGQI	EQLGNQLRHQ
	LENENNDQVH	RDNLHQLKNR	LPLDLRRPQA	LNKISGGVPD	VAKSIRGLET
	QLDQVLKERR	SHFGRLTKWA	KECGITLDPL	QPLIESEKQR	VAERGSAHDA
	KELAIRLLLO	RIGRLGHRLS	PTNATAIQEL	LRPVFAVKRE	FNLFFHNHMG
	ALYRSPYSTS	RHQPFQINVD	VAHGTDWIGT	IETLIQNLFT	QIQDDALLRD
	LVQLEGFVFS	HKLRLALPGVI	PSELARPNNL	QQMGLPALLL	VLLQADQVHR
	ETVLRVFNLY	GSAINGYLFQ	ALRPGFIVRA	GFQRLETKKL	RYVPKAQSWQ
	YPDRLHHAKS	AIKNSLSAGW	IKKNHQGAIL	PQKTLTALVK	QKSLKDTGVP
	EYLVQAPHDW	YVPIDLRGPA	IPIEGLTVGT	EGPELTQLGP	MKDDCAFRAI
	GPSSFKSKID	AGLLPQDVKY	GDMTLIFDQH	YQQSISFANG	TFSIQYQPTS
	LQVKAAPV	DKRPRDTRNN	SHLYDRIVAI	DLGERKIGYA	IFDLKQVLKS
	EQLEPMREDG	KPLIGSISIR	SIRGLMKAVQ	THRNRRQPNY	RIDQTYSKAL
	MHYRESVIGD	VCNAIDTLCA	RYGGFPVLES	SVRNFEVGSA	QLKTVYGSVS
	RRYTWSAVDA	HKNQRQQYWL	GGTKDKIPIW	THPYLMTREW	DEKNSKWSNR
	SKPLKMHPGV	EVHPAGTSQI	CHQCKRNPIG	ALWNVADTVV	LDDQGGQLDLD
	DGTIRLNSGY	IDTTEIKRAR	RKKIRLPENK	PLTGSHKTS	VRVARRNLR
	QPPKSTRAKD	TTQSRYTCLY	VDCGHECHAD	ENAAINIGRK	YLQERIHIEA
	SRQALSTR				
QFN42174.1	MVAGLKKIKR	DGVTMKSNIH	GGVKARAWRK	RIGGLARRQK	ETVFTYKFPL
	ETEEAGIDFD	KAVQTYGIAE	GISQGSILGL	VCAFHLSGFR	LFSKADETKA
	FCNQGRYPNQ	AFABKLRNEL	SVTLPKLSPQ	SLDVLFQSSP	KSKNGVAPBW
	SKNAIRNRLY	TNWTGKGAGT	NPDEHLLLEIA	EDIAAEIDSD	LDGWKDLLEH
	PEKGLSAAAR	YFQAQGD FPS	LTGLPPSVPL	TPQNSTVAFE	GDPVCLNPSD
	NTLLHQAVAR	CAGRILQEQP	NLSPDKNRFI	NQLQDELVSS	QNNGLSWLFG
	VGFKYWKEMS	VDQLADDYKV	KSTDLDALKQ	VKSFDIAIPL	NPLFDTPHYG
	EFRAVAGKM	RSWVKNYWKR	LLDLKSQGLT	ANINLPEGLD	EQRAENLFSG
	LLIDSKGLRQ	VTDKLP SRLK	KAEDTIDRLM	GDGNPTSDDI	EQVETVA AEI
	SAFIGQVEQF	NNQLEQRLN	PLEGDDDETFL	KQLKIDLP AE	FKKPPAINRI
	SGGSPDP TAE	IAELEEK LDR	LMSARKEHYE	TIAEWASANK	VTLDPM EAMT
	TLEAQR LTER	GAEGDQEEFA	LRLLLQRIGR	LANRLSPQGA	TAIRDLLRPV
	FTEKREFNLF	FHNRMGSLYR	SPYSTSRHQF	FTIDVAVAKN	TDWMDALDGI
	AETIMKGLSQ	AGDELSLRQL	EEDEVSREVC	LKAFNLVYSA	INGCLFRALR
	EGFIVRTKFQ	RLERDVLSYV	PKTKLWNYPQ	RLDTARGPIH	SALAAAWINK
	EGSVDPVET	VTALSDTGFS	DDGIPEYLVQ	APHDWYLRDW	INISGFSLSQ
	RLRGLPDTVP	GELALVRSAD	DVRIPPMLAL	TPIDLRDISK	PVSGLPVKKN
	ITGLKRQKKQ	TAFRMVGPSS	FKSHLDSTLL	SEEVKLG DFT	LIFDQYYKQR
	VSYNGRVKIT	FEPDR LHVEA	AVPVIDKRVR	PSTEEDALFD	HLLAIDLGEK
	RVGYAVYDIK	ACLRTGDIKP	LEDGDGKPIV	GSVAVPSIRR	LMKAVRSHRQ
	QRQPNQKVNQ	TYSTALMNYR	ENVIGDVCNR	IDTLM EK YNA	FPVLESSVMN
	FEAGSRQLEM	VYGSVLHRYT	YSKIDAHTAK	RKEYWYTGEY	WDHPYLM AHK
	WNERTRSYSG	SLSALTLYPG	VMVHPAGTSQ	RCHQCKRNPM	VEIKQLTGQV
	EINADGSLEL	DDGTICLYEG	YDYSPEEYK	AKREKRRLDP	NVPLSGRHQA
	KHVSAAVAKRN	LRRPTVSMMS	GDTTQARYVC	LYTDCDFTGH	ADENAAINIG
	WKYLTERIAL	SESKDKAGV			
					(SEQ ID NO: 210)

Table 8 - Cas12e (CasY) orthologs					
APG80656.1	MSKRHPRISG	VKGYRLHAQR	LEYTGKSGAM	RTIKYPLYSS	PSGGRTVPRE
GI: 1110962136	IVSAINDDYV	GLYGLSNFDD	LYNAEKRNEE	KVYSVLDFWY	DCVQYGA VFS
	YTAPGLLKNV	AEVRGGSYEL	TKTLKGSHLY	DELQIDKVIK	FLNKKEISRA
QFN42175.1	NGSLDKLKKD	IIDCFKA EYR	ERHKDQCNKL	ADDIKNAK KD	AGASLGERQK
	KLFRDFFGIS	EQSENDKPSF	TNPLNLTCCCL	LPFDTVNNNR	NRGEVLFNKL
	KEYAQKLDKN	EGSLEMWEYI	GIGNSGTAFS	NFLGEGFLGR	LRENKITELK
					(SEQ ID NO: 211)

KAMMDITDAW	RGQEQUEELE	KRLRILAALT	IKLREPKFDN	HWGGYRSDIN
GKLSSWLQNY	INQTVKIKED	LKGHKKDLKK	AKEMINRFGE	SDTKEEAVVS
SLLESIEKIV	PDDSDADDEKP	DIPATAIYRR	FLSDGRLTLN	RFVQREDVQE
ALIKERLEAE	KKKKPKKRKK	KSDAEDEKET	IDFKELFPHL	AKPLKLVPNF
YGDSKRELYK	KYKNAAIYTD	ALWKAWEKIY	KSAFSSSLKN	SFFDITDFDKD
FFIKRLQKIF	SVYRRFNTDK	WKPIVKNSFA	PYCDIVSLAE	NEVLYKPKQS
RSRKSAAIDK	NRVRLPSTEN	IAKAGIALAR	ELSVAGFDWK	DLLKKEEHEE
YIDLIELHKT	ALALLLAVTE	TQLDISALDF	VENGTVKDFM	KTRDGNLVLE
GRFLEMFQS	IVFSELRLGLA	GLMSRKEFIT	RSAIQTMNGK	QAELLYIPHE
FQSAKITTPK	EMSR AFLDLA	PAEFATSLEP	ESLSEKSLK	LKQMRYPHY
FGYELTRTGQ	GIDGGVAENA	LRLEKSPVKK	REIKCKQYKT	LGRGQNKIVL
YVRSSYYQTQ	FLEWFLHRPK	NVQTDVAVSG	SFLIDEKKVK	TRWNYDALTV
ALEPVSGSER	VFVSQPFTIF	PEKSABEEGQ	RYLGIDIGEY	GIAYTALEIT
GDSAKILDQN	FISDPQLKTL	REEVKGLKLD	QRRGTFAMPS	TKIARIRESL
VHSLRNRIHH	LALKHKAKIV	YELEVSRFEE	GKQIKKKVYA	TLKKADVYSE
IDADKNLQTT	VWGKLAVASE	ISASYTSQFC	GACKKLWRAE	MQVDETITTO
ELIGTVRVIK	GGTLIDAIDK	FMRPPIFDEN	DTPFPKYRDF	CDKHHISKKM
RGNSCLFICP	FCRANADADI	QASQTIALLR	YVKEEKKVED	YFERFRKLKN
IKVLGQMKKI				

6.4. Protospacer Adjacent Motif

[0116] 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.

[0117] 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.

[0118] It will also be appreciated that Cas9 enzymes from different bacterial species (i.e., Cas9 orthologs) can have varying PAM specificities and in 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 thermophilus* (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.

[0119] 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

[0120] “**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).

[0121] 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 PEgRNA. In various embodiments, the prime editors disclosed herein is comprised of PE1.

[0122] 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 PEgRNA. In various embodiments, the prime editors disclosed herein is comprised of PE2.

[0123] In various embodiments, the prime editors disclosed herein is comprised of PE2 and co-expression of MMR protein MLH1dn, that is PE4.

[0124] 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 is comprised of PE3.

[0125] In various embodiments, the prime editors disclosed herein is comprised of PE3 and co-expression of MMR protein MLH1dn, that is PE5.

[0126] 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

[0127] 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.

[0128] 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).

[0129] In typical embodiments, the guide RNA (e.g., atgRNA) or guide RNA complex is capable of binding a DNA binding nickase selected from the group consisting of: Cas9-D10A, Cas9-H840A, Cas12a/b/c/d/e nickase, CasX nickase, SaCas9 nickase, and CasY nickase. In certain embodiments, the nickase is linked or fused to one or more of a reverse transcriptase. In certain embodiments, the nickase is linked or fused to one or more of a reverse transcriptase

and integrase. In certain embodiments, the nickase is linked or fused to one or more of an integrase.

6.7. Attachment Site-Containing Guide RNA (atgRNA)

[0130] 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).

[0131] As used herein, the term “cognate integrase recognition site” or “integration cognate” or “cognate pair” refers to a first integrase recognition site (e.g., any of the integrase recognition sites described herein) and a second integrase recognition site (e.g., any of the integrase recognition sites described herein) that can be recombined. Recombination between a first integrase recognition site (e.g., any of the integrase recognition sites described herein) and a second recognition site (e.g., any of the integrase recognition sites described herein) is mediated by functional symmetry between the two integrase recognition sites and the central dinucleotide of each of the two integrase recognition sites. In some cases, a first integrase recognition site (e.g., any of the integrase recognition sites described herein) that can be recombined with a second integrase recognition site (e.g., any of the integrase 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 BxB1 integrase mediates recombination between the attB site and the attP site.

[0132] In some cases, a single nucleic acid construct includes a first cognate pair (e.g., a first integrase recognition site and a second integrase recognition site) and a second cognate pair (e.g., a third integrase recognition site and a fourth recognition site). In such cases, the first cognate pair and the second cognate pair have different central dinucleotides that enable recombination only with the other integrase recognition site within the cognate pair.

[0133] 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, 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).

[0134] 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.

[0135] In some embodiments, the single nucleic acid construct (i.e., “installer”) contains a nucleotide sequence encoding an attachment site-containing guide RNA (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 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.

[0136] In some embodiments, the single nucleic acid construct (i.e., “installer”) contains a nucleotide sequence encoding a first attachment site-containing guide RNA (atgRNA) and a nucleotide sequence encoding a second attachment site-containing guide RNA (atgRNA). In some embodiments, where the single nucleic acid construct (i.e., “installer”) 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.

[0137] 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 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).

[0138] In some embodiments, upon introducing the nucleic acid construct into a cell, the first atgRNA incorporates the first integrase recognition site into the cell's genome at the target sequence.

[0139] In some embodiments, upon introducing the nucleic acid construct into a cell, the first pair of atgRNAs incorporate the first integrase recognition site into the cell's genome at the target sequence.

[0140] 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	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatc aacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCATCATC CATGGccgatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGC <i>GAGAA</i>	212
ACTB N-term PBS 13 RT 29 AttB 46 atgRNA with v2 scaffold	GCTATTCTCGCAGCTCACCAgttttagagctatgctggaacagcatagcaagttcaaat aaggc tagtccgttatcaacttgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATAT CATCATCCATGGccgatgatcctgacgacggagaccgccgtcgtcgacaagccggcc <i>TGAGCTGCGA GAA</i>	213
ACTB N-term PBS_13_RT_29_with TP901-1 minimal AttB f atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatc aacttgaaaaagtggcaccgagtcggtgc GAGTCGGTGCGACGAGCGCGGC GATATCATCATCCATGGcacaattaacatctcaatcaaggtaaaTGCTTGAGC <i>TGCGAGAA</i>	214
ACTB N-term PBS_13_RT_29_with TP901-1 minimal AttB rc atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatc aacttgaaaaagtggcaccgagtcggtgc GAGTCGGTGCGACGAGCGCGGC GATATCATCATCCATGGagcatttaccttgattgagatgttaattgtgTGAGCTG <i>CGAGAA</i>	215
ACTB N-term PBS_13_RT_29_with PhiBT1 minimal AttB f atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatc aacttgaaaaagtggcaccgagtcggtgc GAGTCGGTGCGACGAGCGCGGC GATATCATCATCCATGGcagggttttgacgaaagtgatccagatgatccagTGAG <i>CTGCGAGAA</i>	216
ACTB N-term PBS_13_RT_29_with PhiBT1 minimal AttB f atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatc aacttgaaaaagtggcaccgagtcggtgc GAGTCGGTGCGACGAGCGCGGC GATATCATCATCCATGGcagggttttgacgaaagtgatccagatgatccagTGAG <i>CTGCGAGAA</i>	217

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
T_29_with PhiBT1 minimal AttB rc atgRNA	GATATCATCATCCATGGctggatcatctggatcactttcgtcaaaacctgTGAGCTGCGAGAA	
ACTB N-term Nicking guide 1 +48 guide	GAAGCCGGCCTTGACATGCgttttagagctagaatagcaagttaaaataaggctagtc cggtat caactgaaaaagtggcaccgagtcggtgc	218
ACTB N-term PBS_18_R T_16_with Lo x71_Cre atgRNA	GAAGCCGGCCTTGACATGCgttttagagctagaatagcaagttaaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc ATATCATCATCCATGGtaccgttc gtatagcatatacattatacgaagtattTGAGCTGCGAGAATAGCC	219
ACTB N-term PBS_13_R T_29_with Lo x71_Cre atgRNA	GAAGCCGGCCTTGACATGCgttttagagctagaatagcaagttaaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGGtaccgttcgtatagcatatacattatacgaagtattTGAGCTGCGAGAA	220
ACTB N-term PBS 13 RT 34 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc TCGACGACGAGCGCGGCGAT ATCATCATCCATGGcggatgatcctgacgacggagaccgccgtcgtcgacaagccg gccTGAGCTGCGAGAA	221
ACTB N-term PBS 13 RT 26 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GAGCGCGGCGATATCATCAT CCATGGcggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTG CGAGAA	222
ACTB N-term PBS 13 RT 23 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc CGCGGCGATATCATCATCCA TGGcggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGA GAA	223
ACTB N-term PBS 13 RT 20 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GGCGATATCATCATCCATGGc cggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGAGAA	224
ACTB N-term PBS 13 RT 16 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc ATATCATCATCCATGGcggatg atcctgacgacggagaccgccgtcgtcgacaagccggccTGAGCTGCGAGAA	225

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 18 RT 34 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcTCGACGACGAGCGCGGCAT ATCATCATCCATGGccggatgatcctgacgacggagaccgccgctcgtcgacaagccg gccTGAGCTGCGAGAATAGCC	226
ACTB N-term PBS 18 RT 29 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCATATCAT CATCCATGGccggatgatcctgacgacggagaccgccgctcgtcgacaagccggccTGA GCTGCGAGAATAGCC	227
ACTB N-term PBS 18 RT 16 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcATATCATCATCCATGGccggatg atcctgacgacggagaccgccgctcgtcgacaagccggccTGAGCTGCGAGAATAGCC	228
LMNB1 N-term PBS 13 RT 39 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcCTGCCCATCCGCGGCGGCAC GGGGGTTCGAGTCGCCATGccggatgatcctgacgacggagaccgccgctcgtcg acaagccggccCGGGCGGCGGAGA	229
LMNB1 N-term PBS 13 RT 34 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcCATCCGCGGCGGCACGGGGG TCGAGTCGCCATGccggatgatcctgacgacggagaccgccgctcgtcgacaagccg gccCGGGCGGCGGAGA	230
LMNB1 N-term PBS 13 RT 29 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGCGGCGGCACGGGGGTTCGCA GTTCGCCATGccggatgatcctgacgacggagaccgccgctcgtcgacaagccggccCGG GCGGCGGAGA	231
LMNB1 N-term PBS 13 RT 24 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGGCACGGGGGTTCGAGTCGC CATGccggatgatcctgacgacggagaccgccgctcgtcgacaagccggccCGGGCGGC GGAGA	232
LMNB1 N-term PBS 13 RT 19 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGGGGGTTCGAGTCGCCATGcc ggatgatcctgacgacggagaccgccgctcgtcgacaagccggccCGGGCGGCGGAGA	233
LMNB1 N-term PBS 18 RT 39 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcCTGCCCATCCGCGGCGGCAC GGGGGTTCGAGTCGCCATGccggatgatcctgacgacggagaccgccgctcgtcg acaagccggccCGGGCGGCGGAGACAGCG	234
LMNB1 N-term PBS 18 RT 34 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcCATCCGCGGCGGCACGGGGG TCGAGTCGCCATGccggatgatcctgacgacggagaccgccgctcgtcgacaagccg gccCGGGCGGCGGAGACAGCG	235

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
LMNB1 N-term PBS 18 RT 29 atgRNA	GCTGTCTCCGCCGCCGCCAGtttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGGGTTCGCA GTCCG CATG cgggatgatcctgacgacggagaccgccgctcgtcgacaagccggccCGG GCGGCGGAGACAGCG	236
LMNB1 N-term PBS 18 RT 24 atgRNA	GCTGTCTCCGCCGCCGCCAGtttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GGCACGGGGGTTCGAGTCCG CATG cgggatgatcctgacgacggagaccgccgctcgtcgacaagccggcc CGGGCGGC GGAGACAGCG	237
LMNB1 N-term PBS 18 RT 19 atgRNA	GCTGTCTCCGCCGCCGCCAGtttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GGGGGTTCGAGTCCGCATGcc gggatgatcctgacgacggagaccgccgctcgtcgacaagccggccCGGGCGGCGGAGAC AGCG	238
LMNB1 N-term Nicking guide 1 +46	GCGTGGTGGGGCCGCCAGCGtttttagagctagaatagcaagttaaataaggctagt ccgtatcaactgaaaaagtggcaccgagtcggtgc	239
ACTB N-term PBS 13 RT 29 AttB 42 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG gatgatcctgacgacggagaccgccgctcgtcgacaagccg TGAGCT GCGAGAA	240
ACTB N-term PBS 13 RT 29 AttB 40 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG gatgatcctgacgacggagaccgccgctcgtcgacaagccg TGAGCTGC GAGAA	241
ACTB N-term PBS 13 RT 29 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG atgatcctgacgacggagaccgccgctcgtcgacaagcc TGAGCTGCG AGAA	242
ACTB N-term PBS 13 RT 29 AttB 36 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG atgatcctgacgacggagaccgccgctcgtcgacaagc TGAGCTGCGAG AA	243
LMNB1 N-term PBS 13 RT 29 AttB 44 atgRNA v2	GCTGTCTCCGCCGCCGCCAGtttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGGGTTCGCA GTCCG CATG cgggatgatcctgacgacggagaccgccgctcgtcgacaagccggcCGGG CGGCGGAGA	244

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
LMNB1 N-term PBS 13 RT 29 AttB 42 atgRNA v2	GCTGTCTCCGCCGCCGCCAgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGGGTTCGA GTCCG CCATGggatgatcctgacgacggagaccgccgctcgtcgacaagccgCGGGCGG GCGGAGA	245
LMNB1 N-term PBS 13 RT 29 AttB 40 atgRNA v2	GCTGTCTCCGCCGCCGCCAgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGGGTTCGA GTCCG CCATGgatgatcctgacgacggagaccgccgctcgtcgacaagccgCGGGCGG CGGAGA	246
LMNB1 N-term PBS 13 RT 29 AttB 38 atgRNA v2	GCTGTCTCCGCCGCCGCCAgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGGGTTCGA GTCCG CCATGatgatcctgacgacggagaccgccgctcgtcgacaagccCGGGCGGC GGAGA	247
NOLC1 N-term PBS 18 RT 29 AttB 46 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCG GCGTCCGCC cggatgatcctgacgacggagaccgccgctcgtcgacaagccggccTCCT CCAGGCAATACGCG	248
NOLC1 N-term PBS 13 RT 29 AttB 46 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCG GCGTCCGCC CcggatgatcctgacgacggagaccgccgctcgtcgacaagccggccTCCT CCAGGCAAT	249
NOLC1 N-term PBS 13 RT 29 AttB 44 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCG GCGTCCGCC CcggatgatcctgacgacggagaccgccgctcgtcgacaagccggcTCCTC CAGGCAAT	250
NOLC1 N-term PBS 13 RT 29 AttB 42 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCG GCGTCCGCC CggatgatcctgacgacggagaccgccgctcgtcgacaagccggTCCTCC AGGCAAT	251
NOLC1 N-term PBS 13 RT 29 AttB 40 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCG GCGTCCGCC CgatgatcctgacgacggagaccgccgctcgtcgacaagccgTCCTCCAG GCAAT	252

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
NOLC1 N-term PBS 13 RT 29 AttB 38 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCG GCGTCCGCCatgatcctgacgacggagaccgccgtcgtcgacaagccTCCTCCAGG CAAT	253
NOLC1 nicking guide -43	GAGCCGAGCACGAGGGGATACgttttagagctagaatagcaagttaaataaggcta gtcgttatcaactgaaaaagtggcaccgagtcggtgc	254
ACTB N-term PBS 13 RT 20 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GCGGATATCATCATCCATGGA tgatcctgacgacggagaccgccgtcgtcgacaagccTGAGCTGCGAGAA	255
ACTB N-term PBS 13 RT 15 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc TATCATCATCCATGGatgatcctga cgacggagaccgccgtcgtcgacaagccTGAGCTGCGAGAA	256
ACTB N-term PBS 13 RT 10 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc TCATCCATGGatgatcctgacgacgga gaccgccgtcgtcgacaagccTGAGCTGCGAGAA	257
ACTB N-term PBS 9 RT 20 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GCGGATATCATCATCCATGGA tgatcctgacgacggagaccgccgtcgtcgacaagccTGAGCTGCG	258
ACTB N-term PBS 9 RT 15 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc TATCATCATCCATGGatgatcctga cgacggagaccgccgtcgtcgacaagccTGAGCTGCG	259
ACTB N-term PBS 9 RT 10 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc TCATCCATGGatgatcctgacgacgga gaccgccgtcgtcgacaagccTGAGCTGCG	260
LMNB1 N-term PBS 13 RT 20 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCAGgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc CGGGGGTTCGAGTCGCCATG atgatcctgacgacggagaccgccgtcgtcgacaagccCGGGCGGCGGAGA	261

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
LMNB1 N-term PBS 13 RT 15 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCAGtttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GTCGCAGTCGCCATGatgatcctg acgacggagaccgccgtcgtcgacaagccCGGGCGGCGGAGA	262
LMNB1 N-term PBS 13 RT 10 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCAGtttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc AGTCGCCATGatgatcctgacgacgga gaccgccgtcgtcgacaagccCGGGCGGCGGAGA	263
LMNB1 N-term PBS 9 RT 20 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCAGtttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc CGGGGGTTCGCAGTCGCCATG atgatcctgacgacggagaccgccgtcgtcgacaagccCGGGCGGCG	264
LMNB1 N-term PBS 9 RT 15 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCAGtttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GTCGCAGTCGCCATGatgatcctg acgacggagaccgccgtcgtcgacaagccCGGGCGGCG	265
LMNB1 N-term PBS 9 RT 10 AttB 38 atgRNA	GCTGTCTCCGCCGCCGCCAGtttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc AGTCGCCATGatgatcctgacgacgga gaccgccgtcgtcgacaagccCGGGCGGCG	266
SUPT16H N-term PBS 13 RT 24 Bxb1-GT_Initial length	GAGAAGCGGCGTCCGGGGCTAgttttagagctagaatagcaagttaaataaggcta gtccgttatcaactgaaaaagtggcaccgagtcggtgc TCTTTGTCCAGAGTCACAG CCATAccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccCCCCGGAC GCCGC	267
SRRM2 N-term PBS 13 RT 24 Bxb1 Initial length	GGGCACGGGGCCATGTACAAgttttagagctagaatagcaagttaaataaggctagt ccgttatcaactgaaaaagtggcaccgagtcggtgc GGCGTCGGCAGCCCGATCC CGTTGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTACATGGC CCCGT	268
DEPDC4 N-term PBS 18 RT 24 Bxb1 Initial length	GTGTCAGGTGGGGCGGGGCTAgttttagagctagaatagcaagttaaataaggctag tccgttatcaactgaaaaagtggcaccgagtcggtgc GCTGGCTCCTCCCCTGGCA CCATAccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccCCCCGCC CACCTGACAC	269
NES N-term PBS 13 RT	GAGTGGGTCAGACGAGCAGGAgtttttagagctagaatagcaagttaaataaggcta gtccgttatcaactgaaaaagtggcaccgagtcggtgc CGACTCCTCCCCCATGCA CCCTCCATCccggatgatcctgacgacggagaccgccgtcgtcgacaagccggccTGCT CGTCTGACC	270

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
29 Bxb1 Initial length		
SUPT16H nicking guide - 53	GCAGCCACCCGCTCTCGGCCCgtttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	271
SRRM2 N-term nicking guide 1 +87	GTGTAGTCAGGCCGCTCACCCgtttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	272
DEPDC4 N-term Nicking guide 1 +59	GCTGACAAGTCTACGGAACCTgtttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	273
NES N-term Nicking guide 2 +79	GCTCCTCCAGCGCCTTGACCgtttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	274
HITI_ACT B_guide	GCTATTCTCGCAGCTCACCAgtttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	275
HITI_SUP TH16_guide	AGAAGCGGCGTCCGGGGCTAgtttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	276
HITI_SRR M2_guide	GGGCACGGGGCCATGTACAAgtttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	277
HITI_NOL C1_guide	GCGTATTGCCTGGAGGATGGgtttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	278
HITI_DEP DC4_guide	TGTCAGGTGGGGCGGGGCTAgtttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	279
HITI_NES_guide	AGTGGGTCAGACGAGCAGGAgtttagagctagaaatagcaagttaaataaggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc	280
HITI_LMN B1_guide	GCTGTCTCCGCCGCCGCCAgtttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	281
HDR Cas9 ACTB guide	GCTATTCTCGCAGCTCACCAgtttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	275
HDR Cas9 LMNB1 guide	GGGGTCGCAGTCGCCATGGCgtttagagctagaaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	282

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB original length atgRNAs for dinucleotides	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATG GccggatgatcctgacgacggagXX cgccgtcgtcgacaagccggccTGAG <i>GCTGCGAGAA</i> 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	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATG GccggatgatcctgacgacggagAC cgccgtcgtcgacaagccggccTGAG <i>CTGCGAGAA</i>	284
ACTB N-term PBS 13 RT 29 atgRNA with AttB 46 CT for multiplexing	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATG GccggatgatcctgacgacggagAG cgccgtcgtcgacaagccggccTGAG <i>CTGCGAGAA</i>	285
NOLC1 N-term PBS 18 RT 29 atgRNA with AttB 46 GA for multiplexing	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCG GCGTCCGCC CccggatgatcctgacgacggagTC cgccgtcgtcgacaagccggccTCC <i>TCCAGGCAATACGCG</i>	286
LMNB1 N-term PBS 18 RT 29 atgRNA with AttB 46 AG for multiplexing	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGGGTGCA GTCGCCATG GccggatgatcctgacgacggagCT cgccgtcgtcgacaagccggccCG <i>GGCGGCGGAGACAGCG</i>	287
EMX1 Cas9 guide 1	GTCACCTCCAATGACTAGGGgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc	288

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
EMX1 Cas9 guide 2	GGGCAACCACAAACCCACGAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	289
ACTB N-term PBS 13 RT 29 AttB 56 GA atgRNA	GCTATTCTCGCAGCTCACCAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGctatgccggtatgctgacgacggagtcgccgtcgtcgacaagccggccctagcTGAGCTGCGAGAA	290
ACTB N-term PBS 13 RT 29 AttB 51 GA atgRNA	GCTATTCTCGCAGCTCACCAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGtgccggtatgctgacgacggagtcgccgtcgtcgacaagccggccctatGAGCTGCGAGAA	291
ACTB N-term PBS 13 RT 29 AttB 46 GA atgRNA	GCTATTCTCGCAGCTCACCAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGcggatgctgacgacggagtcgccgtcgtcgacaagccggccTGAGCTGCGAGAA	292
ACTB N-term PBS 13 RT 29 AttB 41 GA atgRNA	GCTATTCTCGCAGCTCACCAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGggatgctgacgacggagtcgccgtcgtcgacaagccgTGAGCTGCGAGAA	293
ACTB N-term PBS 13 RT 29 AttB 36 GA atgRNA	GCTATTCTCGCAGCTCACCAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGtgatcctgacgacggagtcgccgtcgtcgacaagcTGAGCTGCGAGAA	294
ACTB N-term PBS 13 RT 29 AttB 31 GA atgRNA	GCTATTCTCGCAGCTCACCAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGatcctgacgacggagtcgccgtcgtcgacaTGAGCTGCGAGAA	295
ACTB N-term PBS 13 RT 29 AttB 26 GA atgRNA	GCTATTCTCGCAGCTCACCAgtttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGcctgacgacggagtcgccgtcgtcgTGAGCTGCGAGAA	296

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 21 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATC ATCCATGG gacgacggagtccgccgtc <i>TGAGCTGCGAGAA</i>	297
ACTB N-term PBS 13 RT 29 AttB 16 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGG gacgacggagtccgccg <i>TGAGCTGCGAGAA</i>	298
ACTB N-term PBS 13 RT 29 AttB 11 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGG gacgacggagtccg <i>TGAGCTGCGAGAA</i>	299
ACTB N-term PBS 13 RT 29 AttB 6 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGG cgaggt <i>TGAGCTGCGAGAA</i>	300
ACTB N-term PBS_18_RT_34_with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGACATGCGgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcTCGACGACGAGCGCGGCGAT ATCATCATCCATGG taccgttcgtatagcatacattatacgaagttat <i>TGAGCTGC GAGAATAGCC</i>	301
ACTB N-term PBS_18_RT_29_with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGACATGCGgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGG taccgttcgtatagcatacattatacgaagttat <i>TGAGCTGCGAGAAT AGCC</i>	302
ACTB N-term PBS_13_RT_34_with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGACATGCGgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcTCGACGACGAGCGCGGCGAT ATCATCATCCATGG taccgttcgtatagcatacattatacgaagttat <i>TGAGCTGC GAGAA</i>	303
ACTB N-term PBS_13_RT_16_with_	GAAGCCGGCCTTGACATGCGgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcATATCATCATCCATGG taccgttc gtatagcatacattatacgaagttat <i>TGAGCTGCGAGAA</i>	304

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
Lo x71_CreatgRNA		
ACTB N-term Nicking guide 2 +93 guide	CCCCACGATGGAGGGGAAGAgtttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	305
LMNB1 N-term Nicking guide 2 +87 guide	CCTTCTCCTGGAGCCGCGACgtttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc	306
ACTB N-term PBS 13 RT 29 AttB 46 N191352_1 43_72 integrase	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGcattatatgttcttacagtatggcgccggattgtaaaaacatataatgTGA GCTGCGAGAA	307
ACTB N-term PBS 13 RT 29 AttB 46 N684346_90_69 integrase	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGcgttatagggtattacagtatggcggtcggtactgcaataccctataacgTG AGCTGCGAGAA	308
ACTB N-term PBS 13 RT 29 AttB 46 N675015_95_5 integrase	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGgtatcattttcatatagttagcacctgcacactatagaaatgatacaTGA GCTGCGAGAA	309
ACTB N-term PBS 13 RT 29 AttB 46 N189929_49_54 integrase	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGgttctactatctgtatatgcacacatgtggcataaagacatagtagacaTG AGCTGCGAGAA	310
ACTB N-term PBS 13 RT 29 AttB 46	GCTATTCTCGCAGCTCACCAgtttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCATGGcatcgaccctgacgcagtcggaggcggtccatgcgtctgacctcattTG AGCTGCGAGAA	311

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
N203911_4 5186_6 integrase		
ACTB N-term PBS 13 RT 29 AttB 46 N687663_5 3_29 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGgttagtaccctaaatgacaaaaggctcatcctttatcatttgggtactaacTGA GCTGCGAGAA	312
ACTB N-term PBS 13 RT 29 AttB 46 N687611_9 0_68 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGcttattaaaaccggttcgcttctgtcaaagcggcatcggtttataaacTGA GCTGCGAGAA	313
ACTB N-term PBS 13 RT 29 AttB 46 N190156_2 34_12 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGggcgtgatggctgtgaacctcaacatgacgacgaacacgacctcgggccT GAGCTGCGAGAA	314
ACTB N-term PBS 13 RT 29 AttB 46 N191533_2 24_76 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGtctacatcttgaatatcaagttataactttgaattatatcagtttataTGAG CTGCGAGAA	315
ACTB N-term PBS 13 RT 29 AttB 46 N208621_9 _15 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGaattatatctaaaagcactaagctccgccatactgcttttagatataataTGA GCTGCGAGAA	316
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_ce reus_AH18	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGgatatggggaagtgaatcagtacaaccgccacagtaccTGAGCTGCG AGAA	317

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
7_38 bp_Att		
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH18 7_38 bp_Att_rc	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGggtactgtggcgggtgtactgattcacttccccatattGAGCTGCGAG AA	318
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugduensis_N920143_38bp_Att	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGtgggtgtacaggtgccacattagttgtaccatttatGAGCTGCGAG AA	319
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugduensis_N920143_38bp_Att_rc	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGcataaatgttacaactaatgtggcacctgtaccaccaTGAGCTGCGA GAA	320
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_38bp_Att	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGgttgttttccagatccagttggtcctgtaataataagTGAGCTGCGAG AA	321
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_N	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGcttatatttacaggaccaactggatctggaataaacTGAGCTGCGA GAA	322

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
VH_391-98_38bp_Att_rc		
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH18 7_Att 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGtactgtggcggtgtactgattcacttccccatatTGAGCTGCGAGAA	323
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH18 7_Att 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGtactgtggcggtgtactgattcacttccccataTGAGCTGCGAGAA	324
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH18 7_Att 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGactgtggcggtgtactgattcacttccccatTGAGCTGCGAGAA	325
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH18 7_Att_rc 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGatattgggaagtgaatcagtacaaccgccacagtacTGAGCTGCGAGAA	326
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH18 7_Att_rc 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGtatgggaagtgaatcagtacaaccgccacagtaTGAGCTGCGAGAA	327
ACTB N-term PBS 13 RT 29 AttB 46	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgcGACGAGCGCGGCGATATCAT CATCCATGGatgggaagtgaatcagtacaaccgccacagtTGAGCTGCGAGAA	328

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
Bacillus_cereus_AH187_Att_rc 32bp		
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugduensis_N920143_Att 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG <i>GataaatggtacaactaatgtggcacctgtaccaccTGAGCTGCGAGAA</i>	329
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugduensis_N920143_Att 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG <i>GtaaatggtacaactaatgtggcacctgtaccaccTGAGCTGCGAGAA</i>	330
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugduensis_N920143_Att 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG <i>GaaatggtacaactaatgtggcacctgtaccacTGAGCTGCGAGAA</i>	331
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugduensis_N920143_Att_rc 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG <i>gggtggtacaggtgccacattagttgtaccatttatTGAGCTGCGAGAA</i> <i>A</i>	332
ACTB N-term PBS 13 RT 29 AttB 46	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG <i>gggtggtacaggtgccacattagttgtaccatttaTGAGCTGCGAGAA</i>	333

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
Staphylococcus_lugduensis_N920143_Att_rc 34bp		
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugduensis_N920143_Att_rc 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGGgtgttacaggtgccacattagtgtaccattTGAGCTGCGAGAA	334
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_N VH_391-98_Att 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGGTtatatttacaggaccaactggatctggaaaaacaTGAGCTGCGAGAA	335
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_N VH_391-98_Att 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGGTtatatttacaggaccaactggatctggaaaaacaTGAGCTGCGAGAA	336
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_N VH_391-98_Att 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGGatatttacaggaccaactggatctggaaaaacTGAGCTGCGAGAA	337
ACTB N-term PBS 13 RT 29 AttB 46	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGGTgttttccagatccagttggtcctgtaaatataaTGAGCTGCGAGAA	338

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
Bacillus_cytotoxicus_N VH_391-98_Att_rc 36bp		
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_N VH_391-98_Att_rc 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG gtttttccagatccagttggctctgtaaata TGAGCTGCGAGAA	339
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_N VH_391-98_Att_rc 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GACGAGCGCGGCGATATCAT CATCCATGG gtttttccagatccagttggctctgtaaata TGAGCTGCGAGAA	340
Bacillus_cereus_AH18 7_Att_rc_36 LMNB1 PBS 9 RT 10 AttB 36 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc AGTCGCCATG ataggggaagtgaatc agtacaaccgccacgtacCGGGCGGCG	341
Bacillus_cereus_AH18 7_Att_rc_36 NOLC1 PBS 18 RT 29 AttB 36 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCG GCGTCCGCC ataggggaagtgaatcagtacaaccgccacgtac TCTCCAGGCA ATACGCG	342
Bacillus_cereus_AH18 7_Att_rc_36 SUPT16H PBS 13	GAGAAGCGGCGTCCGGGGCTAgttttagagctagaatagcaagttaaataaggcta gtccgtatcaactgaaaaagtggcaccgagtcggtgc TCTTTGTCCAGAGTCACAG CCATA ataggggaagtgaatcagtacaaccgccacgtac CCCCGGACGCCGC	343

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
RT 24 AttB 36 atgRNA		
Bacillus_cereus_AH187_Att_rc_36 SRRM2 PBS 13 RT 24 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GGCGTCGGCAGCCCCGATCCCGTTGatatggggaagtgaatcagtacaaccgccacagtacTACATGGCCCCGT	344
Bacillus_cereus_AH187_Att_rc_36 DEPDC4 PBS 18 RT 24 AttB 36 atgRNA	GTGTCAGGTGGGGCGGGGCTAgttttagagctagaatagcaagttaaataaggctagtccggttatcaactgaaaaagtggcaccgagtcggtgc GCTGGCTCCTCCCCCTGGCA CCATAatatggggaagtgaatcagtacaaccgccacagtacCCCCGCCCCACCTGACAC	345
Bacillus_cereus_AH187_Att_rc_36 NES PBS 13 RT 28 AttB 36 atgRNA	GAGTGGGTCAGACGAGCAGGAgtttagagctagaatagcaagttaaataaggctagtccggttatcaactgaaaaagtggcaccgagtcggtgc CGACTCCTCCCCCATGCA CCCTCCATCatatggggaagtgaatcagtacaaccgccacagtacTGCTCGTCTGAC	346
B. cereus LMNB1_P BS 9 RT 20 AttB 36 atgRNA	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc CGGGGGTTCGCAGTCGCCATGatatggggaagtgaatcagtacaaccgccacagtacCGGGCGGCG	347
B. cereus LMNB1_P BS 13 RT 20 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc CGGGGGTTCGCAGTCGCCATGat atggggaagtgaatcagtacaaccgccacagtacCGGGCGGCGGAGA	348
B. cereus LMNB1_P BS 13 RT 29 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GCGGCGGCACGGGGGTTCGC AGTCGCCATGatatggggaagtgaatcagtacaaccgccacagtacCGGGCGGCGGAGA	349
B. cereus NOLC1_P BS 13 RT	GCGTATTGCCTGGAGGATGGgttttagagctagaatagcaagttaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcggtgc GAACCACGCGGCGAATGCCCG CGTCCGCCatatggggaagtgaatcagtacaaccgccacagtacTCCTCCAGGCAAT	350

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
29 AttB 36 atgRNA		
B. cereus NOLC1_P BS 13 RT 20 AttB 36 atgRNA	GGGCACGGGGCCATGTACAgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgc GGCGAATGCCGGCGTCCGC CatatggggaagtgaatcagtacaaccgccacagtacTCCTCCAGGCAAT	351
B. cereus NOLC1_P BS 18 RT 20 AttB 36 atgRNA	GGGCACGGGGCCATGTACAgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgc GGCGAATGCCGGCGTCCGC CatatggggaagtgaatcagtacaaccgccacagtacTCCTCCAGGCAATACGCG	352
B. cereus SRRM2_P BS 9 RT 24 AttB 36 atgRNA	GGGCACGGGGCCATGTACAgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgc GGCGTCGGCAGCCCGATCC CGTTGatatggggaagtgaatcagtacaaccgccacagtacTACATGGCC	353
B. cereus SRRM2_P BS 9 RT 10 AttB 36 atgRNA	GGGCACGGGGCCATGTACAgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgc GATCCCGTTGatatggggaagtgaat cagtacaaccgccacagtacTACATGGCC	354
B. cereus SRRM2_P BS 13 RT 10 AttB 36 atgRNA	GGGCACGGGGCCATGTACAgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgc GATCCCGTTGatatggggaagtgaat cagtacaaccgccacagtacTACATGGCCCCGT	355
Screen validation guides ACTB_1_1 1_24_38	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgcgcgcgggcgatatcatcatccatggatgatcctgacgacggagaccgccgtcgtcgacaagcctgagctgcgag	356
Screen validation guides ACTB_1_1 6_18_43	GCTATTCTCGCAGCTCACCAgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgcgcatatcatcatccatggcgatgatcctgacgacggagaccgccgtcgtcgacaagccggtgagctgcgagaatag	357
Screen validation guides LMNB1_1_8_26_38	GCTGTCTCCGCCGCCGCCAgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgcgcgccacgggggtcgagtcgcatgatgatcctgacgacggagaccgccgtcgtcgacaagcccggcgccg	358

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
Screen validation guides NOLC1_1_15_16_43	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgcaatgccggcgccgcccggatgacgacg gagaccgcccgtcgtcgacaagccggctcctccaggcaatac	359
Screen validation guides NOLC1_1_14_10_38	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgcccgcgcccgcctgacgacggagaccgcc gtcgtcgacaagcctcctccaggcaata	360
Screen validation guides SERPIN_1_3_32_38	GGGAAATGCATCTTGCACAAGgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgcagcccctccatgctctctagctgttgccattgggctt gtcgacgacggcggtcctcgtcgcaggatcattgaagatgcatt	361
Screen validation guides DEPDC4_8_10_38	GTGTCAGGTGGGGCGGGGCTAGgttttagagctagaaatagcaagttaaataaggctag tccgttatcaactgaaaaagtgccaccgagtcggtgctggcaccataatgatcctgacgacggagaccgc cgtcgtcgacaagcccccgccc	362
SERPIN Nicking guide -107 guide	GTGGGGACAGCCCCGTCTCTgttttagagctagaaatagcaagttaaataaggctagtc cggtatcaactgaaaaagtgccaccgagtcggtgc	363
SERPIN Nicking guide -91 guide	GCTCTTGGGAAAAAAACCCTAGgttttagagctagaaatagcaagttaaataaggctag tccgttatcaactgaaaaagtgccaccgagtcggtgc	364
SERPIN Nicking guide -90 guide	GTCTTGGGAAAAAAACCCTAAGgttttagagctagaaatagcaagttaaataaggctag tccgttatcaactgaaaaagtgccaccgagtcggtgc	365
SERPIN Nicking guide -84 guide	GAAAAAAACCCTAAGGGCTGgttttagagctagaaatagcaagttaaataaggctagt ccgttatcaactgaaaaagtgccaccgagtcggtgc	366
SERPIN Nicking guide -67 guide	GCTGAGGATCCTTGTGAGTGTgttttagagctagaaatagcaagttaaataaggctagt ccgttatcaactgaaaaagtgccaccgagtcggtgc	367

Table 9		
Description	Sequence (5'-3')	SEQ ID NO:
SERPIN Nicking guide -66 guide	GTGAGGATCCTTGTGAGTGTGtttttagagctagaatagcaagttaaataaggctagtccggtatcaactgaaaaagtggcaccgagtcggtgc	368
SERPIN Nicking guide -63 guide	GGATCCTTGTGAGTGTGGGgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgc	369
SERPIN Nicking guide -62 guide	GATCCTTGTGAGTGTGGGTgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgc	370
SERPIN Nicking guide -49 guide	GTTGGGTGGGAACAGCTCCCgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgc	371
SERPIN Nicking guide -46 guide	GGGTGGGAACAGCTCCCAGGgttttagagctagaatagcaagttaaataaggctagtcggtatcaactgaaaaagtggcaccgagtcggtgc	372
SERPIN Nicking guide +34 guide	GCTTCTGTGCAGCAGTTTCCGgttttagagctagaatagcaagttaaataaggctagtcggtatc aactgaaaaagtggcaccgagtcggtgc	373
SERPIN Nicking guide +48 guide	GTTTCCCTGGCCACTAAATAGgttttagagctagaatagcaagttaaataaggctagtcggtatc aactgaaaaagtggcaccgagtcggtgc	374
SERPIN Nicking guide +49 guide	GTTCCCTGGCCACTAAATAGTgttttagagctagaatagcaagttaaataaggctagtcggtatc aactgaaaaagtggcaccgagtcggtgc	375
SERPIN Nicking guide +71 guide	GATTAGATAGAAGCCCTCCAgtttagagctagaatagcaagttaaataaggctagtcggtatca actgaaaaagtggcaccgagtcggtgc	376
SERPIN Nicking guide +72 guide	GATTAGATAGAAGCCCTCCAAgttttagagctagaatagcaagttaaataaggctagtcggtat caactgaaaaagtggcaccgagtcggtgc	377

6.8. Integrases/Recombinases and Integration/Recombination Sites

[0141] In typical embodiments, the single nucleic acid construct (i.e., “installer”) contains an integrase or recombinase. In some embodiments, the single nucleic acid construct (i.e., “installer”) contains an integrase and a recombinase. In some embodiments, the single nucleic acid construct (i.e., “installer”) contains at least one integrase (e.g., at least two integrases) and at least one recombinase (e.g., at least two recombinases). In some embodiments, an integration enzyme (e.g., an integrase or a recombinase) is selected from the group consisting of Cre, Dre, Vika, Bxb1, ϕ C31, RDF, FLP, ϕ BT1, R1, R2, R3, R4, R5, TP901-1, A118, ϕ FC1, 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.

[0142] In one embodiment, the single nucleic acid construct (i.e., “installer”) contains an integrase (e.g., any of the integrases described herein (e.g., any of the large serine integrases described herein). In one embodiment, the single nucleic acid construct (i.e., “installer”)

contains a recombinase (e.g., any of the recombinases described herein). In some embodiments, the single nucleic acid construct (i.e., “installer”) contains a large serine integrase (e.g., any of the large serine integrases described herein) and a recombinase. In some embodiments, the single nucleic acid construct (i.e., “installer”) contains a BxB1 integrase and a flippase (e.g., FLP).

[0143] 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.

[0144] In typical embodiments, the integrase used herein is selected from below.

Table 10 - Integrases

Data base	SRA accession	bioproject accession	nucleotide accession or ORF ID	protein ID	Internal protein ID	Proposed names	Alternative names	organism/source	description	Sequence	SEQ ID NO:	Length	Group
EN A	SRS1205298	PRJEB26277	NA	NA	N189929_49_54	SsuINT	NA	human gut metagenome	stool sample from male in USA	MEKNRAVLRLSKEDVDKVN KGDDSSSIKSQRLLLTDFALERG FKIVGVYSDDDESGLYDDRPDF ERMMTDAKLDEFDIII AKTQSRF SRNMEHIEKYLHHDLPNLGIRFI GAVDGVDTESDENKKSQRQINGL VNEWYCEDLSKNIRSAFKAKM KDGQFLGSSCPYGYKKDPQNH NHLVVDDYAAKVVQKIFNLYL EGYGKAKIGSILSSEGILPTLYK KDILKQNYHNSKALDTTQNWS YQTIHTLNNEVYLGHLIQNKV NTMSYKDKNKRLPKKWIIVR NTHEPIITEEMFQDVQKLQKNR TRSVENIEPNGLFSGLIFCADCK HAMSRYARRGEKGFVGYVCK TYKTQGNFCESHSIDYDELEE AVLFSIKNEARSILQEEIDELR KVQAYDETKSYEMQLENIKS MEKIEKYKKKTYDNYMDDLIS RDDYKKYVTEYDKEIGGLKQQ QELINSKTDLKEISTQYDEWVE AFINYVDIDKL TREIVIELIEKIEV NKDGSINIYYKFKNPYIS	378	527	INTc
EN A	ERS396461	PRJEB26280	NA	NA	N190156_234_12	SssINT	NA	human gut metagenome	stool sample from Spain	MNTVIYARYSAGPRQTDQSIDG QLRVCTEFCKQRGLTVVDTYC DRHISGR TDERPEFQRLIADAKA HKFEAVVVYKTDRFARNKYDS AIYKRELRRNGIQIFYAAEAPE GPEGIILESLMEGLAEYYSAELA QKIKRGLNESALKCQSLGSGRP LGYTVDEQKHQIDPESSQAVK TIFEMYIKGESNAICDYL NARG LRTSQGNLFNKNSINRIKRNKY IGEYRYNDIVVEGGMPAII SKET FCMAQAEMERRRTHRAPVSPK AEYLLAGKLF CGHCKGPMQGV SGTGKSGNKWYYYCANTRGK ERTCDKKQVSRDRLEKAVVDF TVRYILQENVLEELSKKVYAAQ ERQNTASEIAFYEKKLAENKK AANILRAIESGAMTQALPARLQ ELENEQTVIQGELSYLK GARLA FTEDQILFALLQHLDPRPGESER DYHRRITDFVSEVYL YDDRMLI YFNISSADGKLKHADLSAIESGV FDAGLISSSRASSFSTRCALI	379	510	INTd

EN A	ERS 1015 837	PRJ EB2 6832	NA	NA	N1913 52_14 3_72	SscINT	NA	human gutmetage nome	stool sample from China	MNEKNLEIGAAYIRVSTDDQTE LSPDAQLRVILEAAKKDGIIPQE FVFMEDRGRSGRRADNRPEFQR MISTARQNPSPFYLYLWKFSR FARNQEESAFYKGILRKKCGVTI KSVSEPIMEGMFGRLEVMIIEWS DEFYSVNLSGEVLRGMTQKALE HGYQLTPCLGYDAVGHGRPYVI NEEQYQIVFIHRSFFDGKDMT WIAREANRRGYHTRRGPNFDTR AVRIILTNSFYVGLVKWNDVTF QGTHECRESVTSVFSANQERLN RIHRPRGRRQASSCKHWLSGLL KCSICGASLGYNQTKDLTKRGH AFQCWKYTKGIHPGSCSVSLK AEAAVLESQMILETGEVEYTY EQREKHLDDNKLTLIQKSLERL DTKELRIREAYESGIDTLDEFKT NKARLQRERDQLMEELEELHSQ EPEPDPGKEILIERIQNVYDLL QSPDNDNDKGNVRSIIKKIV YIKESKTFCFYYYYV	380	482	INTd
EN A	ERS 1289 677	PRJ EB2 6924	NA	NA	N1915 33_22 4_76	Ssc2INT	NA	human gutmetage nome	stool sample from China	MERTIKVIQPGTVKIPTKKRVAA YARVSSGKDAMLHLSAQVSY YSNMIQKQNEWSYVGIYADEAI TGTKDRRVEFNRLIQDCTDGKI DMIITKSISRFAFNRLTMLEVVR KLKNINVDVYFEKENIHSISGDG ELMLTILASFQEESSVSSENCK WRIRKGFEQGEINLRFLYGYRI NKGKIEIYEKEAEIVRMIFDDYL NGEGCTRIGNKLRKMKVNLKR GGMWNSERVVDIINKNEYTGN ALLQKKYVKDHLKSKLVRNKG ILTQYYAEGTHPAIIDIKTFEIAQ KIMEANR TKFQKCGSNRYLFT SKIECGICGKNYRHKDREGKST WVCANHLKYGNSRCIAKPLNE EKLKLLINEALELKYFDEEIFIR NIKRIKVTGNQTIEFILKDGK VIE EGMI	381	406	INTc
EN A	ERS 265 5827	PRJ EB2 8245	NA	NA	N2039 11_45 186_6	SsdINT	NA	human gut metageno me	stool sample from Denmark	MKKIKIDRAIQERPATRKQTRN EKIRQSLTEHVDVQVIPAITDRE GYEKPCLRVCAYCRVSTDMDT QALSYELQVQNYTDYIRGNDE WRFAGIYADRGISGTSKLRDE FNRMIEDCKAGKIDLIITKAVTR FARNVLDICISTIRMLKQLEHPV AVYFETERINTLDTTSETYGLI SLFAQGESESKSES LKWSYIRR WKRGTGIYPAWSLLGYEMGED GKWQIVEAEAEELVRIYDMYLN GYSSPQIAEILTRSGVPTATNQT VWSSGGVLGILRNEKYCGNVL CQKTMTVDVFSHKAIKNTGOK TQYFIEGHDPILRSDWDRVQQ MIDEKYYRKRGRRTKPRIVLK GCLAGFTQIDLDWDEDDIARIF YSTTPAAEVATPAMADHIEIKV KGEN	382	401	INTc

EN A	SRS 2949 42	PRJ EB3 0046	NA	NA	N2086 21_9_15	SmcIN T	NA	human gut metageno me	sample from 72- year-old male from China	MKTAAAYIRVSTDDQVEYSPDS QIKLIRDYAKRNDYILPDEFIFR DDGISGKSAKHRPEFTKMIALA KSPEHPFDAILVWKFSRFARNQ EESIVFKNILRKIGVEVRSVSEPI SEDPFGSLVERIIEWTDEYYIINL SGEVKRGMLEKISRGGQPVVPPP VGYKMENGQYIPDENAHFIKEI FEAYAAGEGARHIAQRLAAQG CLTKRGNPIDNRFVDYVLHNPV YIGKLRSVNSHAASSRHYDSA DIIVFDGTHEPLISSELWESVQK RLHEVKTLYPKYQRREQPVFSM LKGLVRCSSCGSTLCYCRTSEPS LQCHSYARGSCRQSHSINIATAN EAVIKGLQLAVDKLDFAIAPAK PHYSADAPGTNKLAAEYKKM ERIKAAANGTDTLEEY AANK KKISAEIARLEAELQQESNVKPI NKKAFKRVSEIHKYISDPHNSE AAKNQALRTVISYIIFDRAATTF NIIFHF	383	476	INTd
Met aSU B	NA	NA	NA	NA	N6750 15_95_5	UhmIN T	NA	urban human microbio me	NA	MKIAIYARKSKYSPTGESVENQI QLCKEYLQAKYKSETLEIDEYK DEGYSGGNTNRPDFKKLIAQIE DYDMLICYRLDRISRNVAADFSS TLTLLQNNKCDVFSIKEQFDITS PMGRAMIYISSVFAQLERE'TIAE RIRDNMELAKMGRWLGGTIP MGFDSEPIITFIDENMKERSMTK LIPNVEELKVIELIYEKYLQLGS MGKVVTYLLQNNIKTKKGKDF TLGSIK VILTNPIYVKANQEVVN HLKTQGITICGDVDGKK ALLTY NKTGTISNDVGITIVKDKSEW IAAVANHKGII PADKWLQAQNI KDKNKDSFPALGRSNTTIASRV LRCDKCESTMGVTHGHINPVTG KKHYYNCTLKKRSKGVRCDN KPAKAAEVDEAILITLENMFKA KSSIIDLKAKNKKARRIEMISSN RVDVINKIIEDKTKQIDNLVNKL SLDDDLTDILFKKIKGLKAEIKE LEDELLTLTSDNIKLNEDEVVLD FTEKLLEKCSIIRTLDILEQQQIV DALIPLVTWNGDTEVLNIYPLG SPELELKEAESKKK	384	550	INTd
Sega ta- Paso lli	NA	PRJ NA4 2243 4	NA	NA	N6843 46_90_69	SacINT	NA	human gut metageno me	stool sample from adult in China	MKEKVSEKRTGAIYIRVSTDKQ EELSPDAQLRLLLDYAKKDSID VPKEYIFQDNGISGRKANKRPA FQNMIALAKSKEHPIDTII VWKF SRFARNQEEISIVYKSLKKNNV DVVSVSEPLIDGPFGLIERIIEW MDEYYISIRLSGEVMRGMTQNA MRGHYQSDAPIGYTSPGDKKPP VINPDTVQIPLMIKDMFLSGSTQ LQIARKLNDSGYRTRKGNLWD ARGVRYVLENPFYIGKSRWNYT ERGRRLKPADEVIIYADGNWEA LWDEDTFKEIQKRLALNMRKS KSRDISAAKHWSGLLICSSCGG TLAFGGAHNMRGFQCWKYSKG FCSESHYISTGPIEKMVLEYLEA VMHSPALSYTVISSSSVDASSKL SDLERQLQKIDAKEKRIKAAAYL NEIDTLEEYKANKTALEEEERT VEKEIEELTSDVKYSKEDLDK KMKQNISDLLRVLRDESADYIQ KGNMMRNVDHIVFNKNTSL DVFLKLVV	385	493	INTd

Sega ta- Paso Li	ERR 1136 864	PRJ EB1 1532	NA	NA	N6876 11_90 _68	RsaINT	NA	human gut metageno me	rectal swab from adult in Isreal	MKITKKQPLRPRGRSEDKRQST KNVIRDAYINGPQKEVQIIPAKR DMEAEETKKKLRVCAYCRVST DEDTQASSYELQVQNYTRMIRE NPEWEFAGIFADEGISGTSVLHR EHFLEMIEKCKAGEIDLITKQV SRFARNVLDLSNYIFMLRKLD PVG VYFETEKLTLDKSSDMVI TVLSLVAQSESEQKSNLKWSE KRRRAQGLGIYPSWALLGYRLD DEKNWEIVEDEADIVRTIYSLYL DGYSTQIAELLTKSGIPTVKGL SVWSSGSLGILKNEKFCGDAL CQKTVTIDFFTHKSVKNNGIEPQ YFVEGHHIPIIEKNDWLLAQQIR KERRYRKRSTHRKPRIVVKGA LSGFMIVDTSWDEEYVDSLLISA TQKPEPAPVIAEEDENFIVIEKE	386	404	INTc
Sega ta- Paso lli	ERR 1136 737	PRJ EB1 1532	NA	NA	N6876 63_53 _29	Rsa2INT	NA	human gut metageno me	rectal swab from adult in Isreal	MADIQPVKNGALYIRVSTHLQE ELSPDAQKRLLMYAEAHNIIV LKEHIYIDSGISGRSARQRPQFN NMIAEAKSKEHPFDVILVWKYS RFARNQEEISIVYKSMLKREND VISVSEPISSDDPFGSLIERIIEWM DEYYSIRLSGEVSRGMAENAMR GNYQARPLGYRIPGYRQTPVI VPEEAELIQLIFDLYTEKKMGIF EIVRYLNEHGYQTGHKKPFQRR SVTYILKNPTYIGKTIWNQHDQ DHKLRLDKSEWIIADGKHEPIISK EQFDKAQKRIESTYKPAYRKPT SVCHHWLSSLLKCSSCGRTLVV KRTASKKKDRMYVNFQCYGYQ KGICNTNQSISAIKLEPVMHAL EDAMTSGKIHFDVLPNTTLDSS QKQQLTRLNEIEKKEERIKRAY RDGIDTLEEYKENKSIQTEKEM LLKKIEHIEEPALSPPEAKPIMM DRIKNVYEITNPDIGMEENKA ARSIEKIVFDRATGSVNIFFYLA HCP	387	498	INTd
NC BI	NA	NA	NC_0026 56.1	NP_07530 2.1	NA	BxbINT	Bxb1 integrase	Mycobact erium phage Bxb1	NA	MRALVVIRLSRVTDATTSPERQ LESCQQLCAQRGWDVVGVAED LDVSGAVDPFDRKRRPNLARW LAFEEQPFQVIVAYRVDRLTRSI RHLQQLVHWAEDHKLVVSAT EAHFDTTTPFAAVVIALMGTV QMELEAIKERNRSAAHFNIRAG KYRGSLPPWGYLPTRVDGEWR LVPDPVQREIRILEVYHRVVDNH EPLHLVAHDLNRRGVLSPKDYF AQLQGREPQGREWSATALKRS MISEAMLGYATLNGKTVRDDD GAPLVRAEPILTREQLEALRAEL VKTSRAKPAVSTPSLLRLVLC AVCGEPAYKFAGGGRKHPYR CRSMGFPHKCGNGTVMAAEW DAFCEEQVLDLLGDAERLEKV WVAGSDSAVELAEVNAELVDL TSLIGSPAYRAGSPQREALDARI AALAAEQEELEGLEARPSGWE WRETGQRFQDWWREQDTAAK NTWLRSMNVRLTFDVRGGLTR TIDFGDLQEYEQHLRLGSVVER LHTGMS*	388	501	INTa

NC BI	NA	NA	NC_002747.1	NP_12664.1	NA	Tp9INT	TP901-lintegrase	Lactococcus phage TP901-1	NA	MTKKVAIYTRVSTTNQAEEGFS IDEQIDRLTKYAEAMGWQVSDT YTDAGFSGAKLERPAMQRLIND IENKAFDITVLVYKLDRLSRSVR DTLYLVKDVFTKNKIDFISLNE IDTSSAMGSLFLTILSAINEFERE NIKERMTMGKLGRAKSGKSMM WTKTAFGYHHNRKTGILEIVPL QATIVEQIFTDYLSGISLTKLRD KLNESGHIGKDIPWSYRTLROT LDNPVYCGYIKFKDSLFEFMMHK PIIPYETYLKVQKELEERQQQTY ERNNNPRPFQAKYMLSGMARCY GYCGAPLKIVLGHKRKDGSR MKYHCANRFPRTKGITVYND NKKCDSGTYDLSNLENTVIDNL IGFQENNDLSLLKIINGNNQPI LDTSSFKKQISQIDKKIQKNSD LYLNDFTITMDELKDRDLSLQAE KKLLKAKISENKFNDSTDVFELV KTQLGSIPINELSYDNKKKIVN NLVSKVDVTADNVDIIFKFQLA*	389	486	INTd
NC BI	NA	NA	NC_004664.2	NP_81374.2	NA	Bt1INT	PhiBTintegrase	Streptomyces virus phiBT1	NA	MSPFIAPDVPEHLLDTVRVFLY ARQSKGRSDGSDVSTEAQLAA GRALVASRNAQGGARVWVAG EFVDVGRSGWDPNVTRADFER MMGEVRAGEGDVVVNLSRL TRKGAHDALEIDNELKKHGVRF MSVLEPFLDTSPIGVAIFALIAA LAKQDSDLKAERLKGAKDEIAA LGGVHSSAPFGMRAVRKKVD NLVISVLEPDEDNPDHVELVER MAKMSFEGVSDNAIATTFEKEK IPSPGMAERRATEKRLASIKARR LNGAEKPIMWRAQTVRWILNH PAIGGFATERVKHGKAHINVIRR DPGGKPLTPHTGILSGSKWLEL QEKRSKGNLSDRKPGAEVEPTL LSGWRFLGCRICGSGMGQSQG GRKRNGDLAEGNYMCANPKG HGGLSVKRSEDEFVASKVWA RLRTADMEDEHDQAWIAAAAE RFALQHDLAGVADERREQQA HLDNVRRSIKDLQADRKAGLYV GREELETWRSTVLQYRSYAEAC TTRLAELDEKMNGSTRVPSEWF SGEDPTAEGGIWASWDVYERR EFLSFFLDSVMVDRGRHPETKK YIPLKDRVTLKWAELLKEEDEA SEATERELAAL*	390	595	INTa

NC BI	NA	NA	NC 0116 58.1	WP 0002 8620 6.1	NA	BceINT	NA	Bacillus reus AH187	NA	MYPYDVPDYAGSYRPESLDVCI YLKRSRKDVEEERRAIEEGSSY NALERHRKRLFAIAKAENHNID IFEEVASGESIQERPQMQLLRK LEGNEIDGVLVIDLDRGRGDM LDAGMIDRAFRYSSTKIITPTDV YDPDESDELWELVFGIKSLSRQEL KSITKRLQNGRIDSVEGKHIGK KPPYGYLKDENLRLYPDEKA WIVKKIFELMCDGKGROMIAAE LDRLGIDPPVTKRGAWDSSTITS IINKNEVYTGIVWGFKHKHRRN GKYTRHKNPOEKWIMYENAE PIISKELFDAANEAHSSRHKPAV ITSKLTNPLAGILKCKLCG YTMLIQTRKDRPHNYLRNNPA CKGKQKQSVFNLVEEKLLYSLO QIVDEY QAQKVEEVEIDSKLISFKEKAIH SKE KELKELQAQKGNLHDLLEQGIY TVE IFLERQKNLVERITSIENDIEVLQ KEIE TEQIKEHNKTEFIPALKTVIESY HKTT NIELKNQLLTILSTVYYRHPD WKT NEFEIQVYFKIS*	391	529	INTc
NC BI	NA	NA	NC 00 9674 .1	WP 0120 9542 9.1	NA	BcyINT	NA	Bacill us cytotoxic us NVH391 -98	NA	MYPYDVPDYAGSAVGIIYRVST QEQAEGHSIESQKKKLASYCEI QGWDYRFRYIEEGISGKNTNRP KLKLLMEHIEKGKINILLVYRLD RLTRSVIDLHKLLNFLQEHGCA FKSATETYDTTANGRMSMGIV SLLAQWETENMSERIKLNLEHK VLVEGERVGAIPYGFDSLDEK LVKNEKSAILDMVERVENGW SVNRIVNYLNTNDRNWSNP GVLRLLRNPALYGATRWNDKI AENTHEGIIISKERFNLQOILAD RSIHHRDVKGTIFYQGVLRCP VCDQTLNVNRFIKKRKGTEYC GVLRYRCQPCIKQNKYNLAIGE RFLKALNEYMSTVEFQTVDEVE IPKKSEREMLESQOQIARKREK YQKAWASDLMSDDEFKLMVE TRETYDECKQKLESCEDPIKIDE TYLKEIVYMFHQTFNDLESEKQ KEFISKFIRTYTVKEQQPIRPD KSKTGKGGKQVIITEVE FYQS*	392	487	INTd
NC BI	NA	NA	NC 0173 53.1	WP 0145 3323 8.1	NA	StuINT	NA	Staphyloc occus lugdunensis N920143	NA	MYPYDVPDYAGSKVAIYTRVSS AEQANEGYSIHEQKKKLISYCEI HDWNEYKVFTDAGISGGSMKR PALQKLMKHLSSFDLVLVYKLD RLTRNVRLDMLLEFEQYNVS FKSATEVFDTTSAIGKLFITMVG AMAEWERETIRERSLFGSRAAV REGNYIREAPFCYDNIEGKLHPN EYAKVIDLIVSMFKKGISANEIA RRLNSSKVHVPNKSWNRNSLI RLMRSPVLRGHTKYGDMIENT HEPVLSEHDYNAINNAISSKTHK SKVKHHAIFRGALVCPQCNRRRL HLYAGTVKDRKGYKYDVRRY KCETCSKNKDVKNVSFNESEVE NKFVNLLKSYELNKFHIRKVEP VKKIEYDIDKINKQKINYTRWS LGYIEDDEYFELMEEINATKKMI EEQTENKQSVSKEQIQSINNFIL KGWEELTIKDEELILSTVDKIE FNFIKDKKHK TNTLDINNIHFKFS*	393	473	INTd

[0145] Sequences of insertion sites (i.e., recognition target sites) suitable for use in embodiments of the disclosure are presented below.

Table 11				
Bxb1_AttP_ GT_original_ _site	GTGGTTTGTCTGGTCAACCA CCGCGGTCTCAGTGGTGTAC GGTACAAACCCA	SEQ ID NO: 394	TGGGTTTGTACCGTACACC ACTGAGACCGCGGTGGTTG ACCAGACAAACCAC	SEQ ID NO: 473
Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bxb1_AttP_ CG_site	GTGGTTTGTCTGGTCAACCA CCGCGcgCTCAGTGGTGTAC GGTACAAACCCA	395	TGGGTTTGTACCGTACACC ACTGAGCGCGCGGTGGTTG ACCAGACAAACCAC	474
Bxb1_AttP_ GC_site	GTGGTTTGTCTGGTCAACCA CCGCGgcCTCAGTGGTGTAC GGTACAAACCCA	396	TGGGTTTGTACCGTACACC ACTGAGGCCGCGGTGGTTG ACCAGACAAACCAC	475
Bxb1_AttP_ AT_site	GTGGTTTGTCTGGTCAACCA CCGCGatCTCAGTGGTGTAC GGTACAAACCCA	397	TGGGTTTGTACCGTACACC ACTGAGATCGCGGTGGTTG ACCAGACAAACCAC	476
Bxb1_AttP_ TA_site	GTGGTTTGTCTGGTCAACCA CCGCGtaCTCAGTGGTGTAC GGTACAAACCCA	398	TGGGTTTGTACCGTACACC ACTGAGTACGCGGTGGTTG ACCAGACAAACCAC	477
Bxb1_AttP_ GG_site	GTGGTTTGTCTGGTCAACCA CCGCGggCTCAGTGGTGTAC GGTACAAACCCA	399	TGGGTTTGTACCGTACACC ACTGAGCCC GCGGTGGTTG ACCAGACAAACCAC	478
Bxb1_AttP_ TT_site	GTGGTTTGTCTGGTCAACCA CCGCGttCTCAGTGGTGTACG GTACAAACCCA	400	TGGGTTTGTACCGTACACC ACTGAGAACGCGGTGGTTG ACCAGACAAACCAC	479
Bxb1_AttP_ GA_site	GTGGTTTGTCTGGTCAACCA CCGCGgaCTCAGTGGTGTAC GGTACAAACCCA	401	TGGGTTTGTACCGTACACC ACTGAGTCCGCGGTGGTTG ACCAGACAAACCAC	480
Bxb1_AttP_ AG_site	GTGGTTTGTCTGGTCAACCA CCGCGagCTCAGTGGTGTAC GGTACAAACCCA	402	TGGGTTTGTACCGTACACC ACTGAGCTCGCGGTGGTTG ACCAGACAAACCAC	481
Bxb1_AttP_ CC_site	GTGGTTTGTCTGGTCAACCA CCGCGccCTCAGTGGTGTAC GGTACAAACCCA	403	TGGGTTTGTACCGTACACC ACTGAGGGCGCGGTGGTTG ACCAGACAAACCAC	482
Bxb1_AttP_ TC_site	GTGGTTTGTCTGGTCAACCA CCGCGtcCTCAGTGGTGTAC GGTACAAACCCA	404	TGGGTTTGTACCGTACACC ACTGAGGACGCGGTGGTTG ACCAGACAAACCAC	483
Bxb1_AttP_ CT_site	GTGGTTTGTCTGGTCAACCA CCGCGctCTCAGTGGTGTAC GGTACAAACCCA	405	TGGGTTTGTACCGTACACC ACTGAGAGCGCGGTGGTTG ACCAGACAAACCAC	484
Bxb1_AttP_ AA_site	GTGGTTTGTCTGGTCAACCA CCGCGGaaCTCAGTGGTGTAC GGTACAAACCCA	406	TGGGTTTGTACCGTACACC ACTGAGTTTCGCGGTGGTTG ACCAGACAAACCAC	485
Bxb1_AttP_ CA_site	GTGGTTTGTCTGGTCAACCA CCGCGcaCTCAGTGGTGTAC GGTACAAACCCA	407	TGGGTTTGTACCGTACACC ACTGAGTGCGCGGTGGTTG ACCAGACAAACCAC	486
Bxb1_AttP_ AC_site	GTGGTTTGTCTGGTCAACCA CCGCGacCTCAGTGGTGTAC GGTACAAACCCA	408	TGGGTTTGTACCGTACACC ACTGAGGTCGCGGTGGTTG ACCAGACAAACCAC	487

Table 11				
Bxb1_AttP_TG_site	GTGGTTTGTCTGGTCAACCA CCGCGtgCTCAGTGGTGTAC GGTACAAACCCA	409	TGGGTTTGTACCGTACACC ACTGAGCACGCGGTGGTTG ACCAGACAAACCAC	488
Bxb1_AttB_46_GT_original_site	GGCCGGCTTGTCGACGACG GCGGTCTCCGTCGTCAGGAT CATCCGG	410	CCGGATGATCCTGACGACG GAGACCGCCGTCGTCGACA AGCCGGCC	489
Bxb1_AttB_46_AA_site	GGCCGGCTTGTCGACGACG GCGGaaCTCCGTCGTCAGGAT CATCCGG	411	CCGGATGATCCTGACGACG GAGTTCGCCGTCGTCGACA AGCCGGCC	490
Bxb1_AttB_46_GA_site	GGCCGGCTTGTCGACGACG GCGgaCTCCGTCGTCAGGAT CATCCGG	412	CCGGATGATCCTGACGACG GAGTCCGCCGTCGTCGACA AGCCGGCC	491
Bxb1_AttB_46_CA_site	GGCCGGCTTGTCGACGACG GCGcaCTCCGTCGTCAGGAT CATCCGG	413	CCGGATGATCCTGACGACG GAGTGCGCCGTCGTCGACA AGCCGGCC	492
Bxb1_AttB_46_TA_site	GGCCGGCTTGTCGACGACG GCGtaCTCCGTCGTCAGGATC ATCCGG	414	CCGGATGATCCTGACGACG GAGTACGCCGTCGTCGACA AGCCGGCC	493
Bxb1_AttB_46_AG_site	GGCCGGCTTGTCGACGACG GCGagCTCCGTCGTCAGGAT CATCCGG	415	CCGGATGATCCTGACGACG GAGCTCGCCGTCGTCGACA AGCCGGCC	494
Bxb1_AttB_46_GG_site	GGCCGGCTTGTCGACGACG GCGggCTCCGTCGTCAGGAT CATCCGG	416	CCGGATGATCCTGACGACG GAGCCCGCCGTCGTCGACA AGCCGGCC	495
Bxb1_AttB_46_CG_site	GGCCGGCTTGTCGACGACG GCGcgCTCCGTCGTCAGGAT CATCCGG	417	CCGGATGATCCTGACGACG GAGCGCGCCGTCGTCGACA AGCCGGCC	496
Bxb1_AttB_46_TG_site	GGCCGGCTTGTCGACGACG GCGtgCTCCGTCGTCAGGAT CATCCGG	418	CCGGATGATCCTGACGACG GAGCACGCCGTCGTCGACA AGCCGGCC	497
Bxb1_AttB_46_AC_site	GGCCGGCTTGTCGACGACG GCGacCTCCGTCGTCAGGAT CATCCGG	419	CCGGATGATCCTGACGACG GAGGTGCGCCGTCGTCGACA AGCCGGCC	498
Bxb1_AttB_46_GC_site	GGCCGGCTTGTCGACGACG GCGgcCTCCGTCGTCAGGAT CATCCGG	420	CCGGATGATCCTGACGACG GAGGCCGCCGTCGTCGACA AGCCGGCC	499
Bxb1_AttB_46_CC_site	GGCCGGCTTGTCGACGACG GCGccCTCCGTCGTCAGGAT CATCCGG	421	CCGGATGATCCTGACGACG GAGGGCGCCGTCGTCGACA AGCCGGCC	500
Bxb1_AttB_46_TC_site	GGCCGGCTTGTCGACGACG GCGtcCTCCGTCGTCAGGATC ATCCGG	422	CCGGATGATCCTGACGACG GAGGACGCCGTCGTCGACA AGCCGGCC	501
Bxb1_AttB_46_AT_site	GGCCGGCTTGTCGACGACG GCGatCTCCGTCGTCAGGATC ATCCGG	423	CCGGATGATCCTGACGACG GAGATCGCCGTCGTCGACA AGCCGGCC	502
Bxb1_AttB_46_CT_site	GGCCGGCTTGTCGACGACG GCGctCTCCGTCGTCAGGATC ATCCGG	424	CCGGATGATCCTGACGACG GAGAGCGCCGTCGTCGACA AGCCGGCC	503
Bxb1_AttB_46_TT_site	GGCCGGCTTGTCGACGACG GCGttCTCCGTCGTCAGGATC ATCCGG	425	CCGGATGATCCTGACGACG GAGAACGCCGTCGTCGACA AGCCGGCC	504

Table 11				
Bxb1_AttB _38_GT_sit e	GGCTTGTCGACGACGGCGG TCTCCGTCGTCAGGATCAT	426	ATGATCCTGACGACGGAGA CCGCCGTCGTCGACAAGCC	505
Bxb1_AttB _38_AA_sit e	GGCTTGTCGACGACGGCGaa CTCCGTCGTCAGGATCAT	427	ATGATCCTGACGACGGAGT TCGCCGTCGTCGACAAGCC	506
Bxb1_AttB _38_GA_sit e	GGCTTGTCGACGACGGCGga CTCCGTCGTCAGGATCAT	428	ATGATCCTGACGACGGAGT CCGCCGTCGTCGACAAGCC	507
Bxb1_AttB _38_CA_sit e	GGCTTGTCGACGACGGCGca CTCCGTCGTCAGGATCAT	429	ATGATCCTGACGACGGAGT GCGCCGTCGTCGACAAGCC	508
Bxb1_AttB _38_TA_sit e	GGCTTGTCGACGACGGCGta CTCCGTCGTCAGGATCAT	430	ATGATCCTGACGACGGAGT ACGCCGTCGTCGACAAGCC	509
Bxb1_AttB _38_AG_sit e	GGCTTGTCGACGACGGCGag CTCCGTCGTCAGGATCAT	431	ATGATCCTGACGACGGAGC TCGCCGTCGTCGACAAGCC	510
Bxb1_AttB _38_GG_sit e	GGCTTGTCGACGACGGCGgg CTCCGTCGTCAGGATCAT	432	ATGATCCTGACGACGGAGC CCGCCGTCGTCGACAAGCC	511
Bxb1_AttB _38_CG_sit e	GGCTTGTCGACGACGGCGcg CTCCGTCGTCAGGATCAT	433	ATGATCCTGACGACGGAGC GCGCCGTCGTCGACAAGCC	512
Bxb1_AttB _38_TG_sit e	GGCTTGTCGACGACGGCGtg CTCCGTCGTCAGGATCAT	434	ATGATCCTGACGACGGAGC ACGCCGTCGTCGACAAGCC	513
Bxb1_AttB _38_AC_sit e	GGCTTGTCGACGACGGCGac CTCCGTCGTCAGGATCAT	435	ATGATCCTGACGACGGAGG TCGCCGTCGTCGACAAGCC	514
Bxb1_AttB _38_GC_sit e	GGCTTGTCGACGACGGCGgc CTCCGTCGTCAGGATCAT	436	ATGATCCTGACGACGGAGG CCGCCGTCGTCGACAAGCC	515
Bxb1_AttB _38_CC_sit e	GGCTTGTCGACGACGGCGcc CTCCGTCGTCAGGATCAT	437	ATGATCCTGACGACGGAGG GCGCCGTCGTCGACAAGCC	516
Bxb1_AttB _38_TC_sit e	GGCTTGTCGACGACGGCGtc CTCCGTCGTCAGGATCAT	438	ATGATCCTGACGACGGAGG ACGCCGTCGTCGACAAGCC	517
Bxb1_AttB _38_AT_sit e	GGCTTGTCGACGACGGCGat CTCCGTCGTCAGGATCAT	439	ATGATCCTGACGACGGAGA TCGCCGTCGTCGACAAGCC	518
Bxb1_AttB _38_CT_sit e	GGCTTGTCGACGACGGCGct CTCCGTCGTCAGGATCAT	440	ATGATCCTGACGACGGAGA GCGCCGTCGTCGACAAGCC	519
Bxb1_AttB _38_TT_sit e	GGCTTGTCGACGACGGCGttc TCCGTCGTCAGGATCAT	441	ATGATCCTGACGACGGAGA ACGCCGTCGTCGACAAGCC	520
Cre Lox 66 site	TACCGTTTCGTATAATGTATG CTATACGAAGTTAT	442	ATAACTTCGTATAGCATAC ATTATACGAACGGTA	521

Table 11				
Cre Lox 71 site	ATAACTTCGTATAATGTATG CTATACGAACGGTA	443	TACCGTTCGTATAGCATAC ATTATACGAAGTTAT	522
TP901-1 minimal AttB site	TTTACCTTGATTGAGATGTT AATTGTG	444	CACAATTAACATCTCAATC AAGGTAAA	523
TP901-1 minimal AttP site	GCGAGTTTTTATTTTCGTTTA TTTCAATTAAGGTAATAAA AAACTCCTTT	445	AAAGGAGTTTTTTAGTTAC CTTAATTGAAATAAACGAA ATAAAAACTCGC	524
PhiBT1 minimal AttB site	CTGGATCATCTGGATCACTT TCGTCAAAAACCTG	446	CAGGTTTTTGACGAAAGTG ATCCAGATGATCCAG	525
PhiBT1 minimal AttP site	TTCGGGTGCTGGGTTGTTGT CTCTGGACAGTGATCCATGG GAAACTACTCAGCACCA	447	TGGTGCTGAGTAGTTTCCC ATGGATCACTGTCCAGAGA CAACAACCCAGACCCGAA	526
Bacillus_cereus_AH187_Int30_38bp_Att	gatatggggaagtgaatcagtacaaccgcca cagtacc	448	ggtactgtggcggttgactgattcacttccc catatc	527
Staphylococcus_lugdunensis_N920143_Int12_38bp_Att	tgggtgggtacaggtgccacattagtgtacat ttatg	449	cataaatggtacaactaatgtggcacctgtac caccca	528
Bacillus_cytotoxicus_NVH_391-98_Int13_38bp_Att	gttgttttccagatccagttggtcctgtaaata aag	450	cttatattacaggaccaactggatctggaaa aacaac	529
Bacillus_cereus_AH187_Int30_Att_30	tggggaagtgaatcagtacaaccgccacag	451	ctgtggcggttgactgattcacttcccca	454
Bacillus_cereus_AH187_Int30_Att_28	ggggaagtgaatcagtacaaccgccaca	452	tgtggcggttgactgattcacttcccc	455
Bacillus_cereus_AH187_Int30_Att_26	gggaagtgaatcagtacaaccgccac	453	gtggcggttgactgattcacttccc	456
Bacillus_cereus_AH187_Int30_Att_rc_30	ctgtggcggttgactgattcacttcccca	454	tggggaagtgaatcagtacaaccgccacag	451
Bacillus_cereus_AH187_Int30_Att_rc_28	tgtggcggttgactgattcacttcccc	455	ggggaagtgaatcagtacaaccgccaca	452
Bacillus_cereus_AH187_Int30_Att_rc_26	gtggcggttgactgattcacttccc	456	gggaagtgaatcagtacaaccgccac	453

Table 11				
Bacillus_cyt otoxicus_ NVH_391- 98_Int13_A tt_30	ttttccagatccagttggctcgttaaata	457	tatttacaggaccaactggatctggaaaaa	460
Bacillus_cyt otoxicus_ NVH_391- 98_Int13_A tt_28	ttttccagatccagttggctcgttaaat	458	atttacaggaccaactggatctggaaaaa	461
Bacillus_cyt otoxicus_ NVH_391- 98_Int13_A tt_26	ttttccagatccagttggctcgttaaa	459	tttacaggaccaactggatctggaaa	462
Bacillus_cyt otoxicus_ NVH_391- 98_Int13_A tt_rc_30	tatttacaggaccaactggatctggaaaaa	460	ttttccagatccagttggctcgttaaata	457
Bacillus_cyt otoxicus_ NVH_391- 98_Int13_A tt_rc_28	atttacaggaccaactggatctggaaaaa	461	ttttccagatccagttggctcgttaaat	458
Bacillus_cyt otoxicus_ NVH_391- 98_Int13_A tt_rc_26	tttacaggaccaactggatctggaaa	462	ttttccagatccagttggctcgttaaa	459
N680429_5 60_31_50bp	CATTATATGTTTTTACAATC CGGGCCGCCATACTGTAAG AACATATAATG	463	cattatatgttcttacagtatggcgcccgga ttgtaaaaacatataatg	530
N191607_8 _101_50bp	CGTTATAGGGTATTGCAGTA CCGACCGCCATACTGTAATA CCCTATAACG	464	cgttatagggtattacagtatggcgctcggt ctgcaataacctataacg	531
N674992_1 _1308_50bp	TGTATCATTTTCATATAGTG TGCAGGTGCTAACTATATGA AAATGATACA	465	tgtatcattttcatatagttagcacctgcacact atatgaaaatgataca	532
N684613_5 4_96_50bp	TGTCTACTATGTCTTTATGC CACATGTGTGTCATATACAG ATAGTAGACA	466	tgtctactatctgtatatgcgacacatgtggca taaagacatagtagaca	533
N252616_1 21_74_50bp	AATGAGGTCAGACGCATGG AGCGCCGCCTCCGCATGCGT CAGGGTCGATG	467	catcgaccctgacgcatcgaggcgcgcg ctccatgcgtctgacctcatt	534
N683040_2 22_19_50bp	GTTAGTACCCAAATGATAA AAGGATGACCTTTTGTCATT TGGGTACTAAC	468	gtagtagacccaaatgacaaaaggtcatccttt tatcatttgggtactaac	535
N687537_1 73_59_50bp	GTTTATAAAACCGATGCCGC TTTGACAGAAGCGGAACGG GTTTTAATAAG	469	cttattaaaacccgttccgcttctgtcaaagc ggcatcggtttataaac	536

Table 11				
N183629_4 7_40_50bp	GGCCGCGAGGTCGTGTTTCGT CGTCATGTTGAGGTTACGA CCATCACGCC	470	ggcgtgatggctcgtgaacctcaacatgacg acgaacacgacctcgcgcc	537
N191533_2 24_76_50bp	TATAAACTGATATAATTCAA AGTTATAACTTGATATATTC AAGATGTAGA	471	tctacatcttgaatatacaagtataactttga attatatcagttata	538
N682356_1 88_20_50 bp	TATTATATCTAAAAGCAGTA TGGCGGAGCTTAGTGCTTTT AGATATAATT	472	aattatatctaaaagcactaagctccgccata ctgcttttagatataata	539

6.9. Nucleic acid construct design

[0146] A single nucleic acid construct is described herein that allows for programmable gene insertion (PGI) (e.g., incorporation of any template into any DNA locus using DNA delivery of a single component DNA).

[0147] In various embodiments, the nucleic acid construct contains a nucleotide sequence encoding an integrase, a nucleotide sequence encoding a prime editor fusion protein or a gene writer protein, a nucleotide sequence encoding at least a first attachment site-containing guide RNA (atgRNA), a DNA donor template (i.e., “cargo”), optionally a nucleotide sequence encoding a nickase guide RNA (ngRNA), and optionally a nucleotide sequence encoding a recombinase.

[0148] In various embodiments, the nucleic acid construct contains a nucleotide sequence encoding an integrase, a nucleotide sequence encoding a prime editor fusion protein or a gene writer protein, a nucleotide sequence encoding at least a first attachment site-containing guide RNA (atgRNA), a DNA donor template (i.e., “cargo”), a nucleotide sequence encoding a nickase guide RNA (ngRNA), and optionally a nucleotide sequence encoding a recombinase. In various embodiments, the nucleic acid construct contains a nucleotide sequence encoding an integrase, a nucleotide sequence encoding a prime editor fusion protein or a gene writer protein, a nucleotide sequence encoding at least a first attachment site-containing guide RNA (atgRNA), a DNA donor template (i.e., “cargo”), a nucleotide sequence encoding a nickase guide RNA (ngRNA), and a nucleotide sequence encoding a recombinase. In various embodiments, the nucleic acid construct contains a nucleotide sequence encoding an integrase, a nucleotide sequence encoding a prime editor fusion protein or a gene writer protein, a nucleotide sequence encoding a first attachment site-containing guide RNA (atgRNA), a second attachment site-containing guide RNA (atgRNA), a DNA donor template (i.e.,

“cargo”), and a nucleotide sequence encoding a recombinase, where the first atgRNA and the second atgRNA are an at least first pair of atgRNAs. In various embodiments, the nucleic acid construct contains a nucleotide sequence encoding an integrase, a nucleotide sequence encoding a prime editor fusion protein or a gene writer protein, a nucleotide sequence encoding a first attachment site-containing guide RNA (atgRNA), a nucleotide sequence encoding a second attachment site-containing guide RNA (atgRNA), and a DNA donor template (i.e., “cargo”), where the first atgRNA and the second atgRNA are an at least first pair of atgRNAs.

[0149] In various embodiments, the nucleic acid construct comprises: a nucleotide sequence encoding a prime editor fusion protein; a nucleotide sequence encoding at least a first attachment site-containing guide RNA (atgRNA); a nucleotide sequence encoding a recombinase; a nucleic acid cargo; and a nucleotide sequence encoding a nickase guide RNA (ngRNA).

[0150] In some embodiments, the nucleic acid construct comprises: a nucleotide sequence encoding a prime editor fusion protein, a nucleotide sequence encoding a first attachment site-containing guide RNA (atgRNA), a nucleotide sequence encoding a second attachment site-containing guide RNA (atgRNA), and a nucleotide sequence encoding a recombinase; a nucleic acid cargo; where the first atgRNA and the second atgRNA are an at least first pair of atgRNAs.

[0151] In some embodiments, a single promoter drives expression of all the different nucleotide sequences on the single nucleic acid construct. In some embodiments, two or more promoters drive expression of the different nucleotide sequences on the single nucleic acid construct. In typical embodiments, at least one promoter drives the expression of the prime editor fusion protein or the gene writer protein, atgRNA, optionally ngRNA, integrase (e.g., serine integrase), and optionally recombinase. In some embodiments, the promoter is an immediate early promoter such as a CMV promoter or a type III RNA polymerase III promoter such as a U6 promoter. In some embodiments, the promoter is any Pol II promoter. In some embodiments, the atgRNA and ngRNA are driven by any Pol III promoter. In some embodiments, the respective promoters used to drive the expression of the protein components, the atgRNA, and the ngRNA have different promoter expression strength, fidelity, selectivity, and/or tissue-specificity.

[0152] In various embodiments, the integrase that is encoded in the nucleic acid construct is fused to the prime editor fusion protein or the Gene Writer protein optionally by a linker. In

various embodiments, the recombinase that is encoded in the nucleic acid construct is fused to the prime editor fusion protein or the Gene Writer protein optionally by a linker.

[0153] In some embodiments, the nucleic acid construct contains a 5' inverted terminal repeat (ITR). In some embodiments, the nucleic acid construct contains a 3' inverted terminal repeat (ITR). In some embodiments, the nucleic acid construct contains a 5' and a 3' inverted terminal repeat. 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, AAV5, or both.

[0154] In typical embodiments, the nucleic acid construct further comprises at least one integrase recognition target site (e.g., an integrase recognition site in the nucleic acid construct used to facilitate integration of all or part of the nucleic acid construct into an integrase recognition site incorporated into a cell genome). In such cases, the at least one integrase recognition site is separate from the integration sequences encoded by the first atgRNA, second atgRNA, or both. In some embodiments, the at least one integrase recognition site is a cognate pair with the integration sequences encoded by the first atgRNA, second atgRNA, or by a combination of the first atgRNA and second atgRNA. In some embodiments, the at least one integrase recognition site is specific for a BxB1, *B. cereus* (BceINTc or Bcec), N191352_143_72 stool sample from China (SscINTd or Sscd), N684346_90_69 stool sample from adult in China (SacINTd or Sacd).

[0155] In certain embodiments, the nucleic acid construct further comprises at least one recombinase recognition target site (e.g., one recombinase recognition site, two recombination recognition sites, three recombinase recognition sites, or four recombinase recognitions site, or more). In some embodiments, the at least one recombinase recognition site is specific for a FLP, a FLP mutant, Cre, or a Cre mutant. In some embodiments, the nucleic acid construct comprises two recombinase recognition sites where the two sites flank the nucleic acid cargo. In such cases, the two recombinase recognition sites are capable of self-circularizing to form a circular construct when contacted with a recombinase.

[0156] In certain embodiments, the nucleic acid construct further comprises at least one recombinase recognition target site and at least one integrase recognition target site.

[0157] In typical embodiments, the nucleic acid construct contains a nucleic acid cargo (i.e., "integration" cargo) of interest. In some embodiments, the nucleic acid cargo is one or

more genes or gene fragments. In some embodiments, the nucleic acid cargo is at least one intron, at least one exon sequence, or a combination thereof. In some embodiments, the nucleic acid cargo is at least one intron fragment, at least exon fragment sequence, or a combination thereof. In some embodiments the nucleic acid cargo is an expression cassette. In some embodiments, the nucleic acid 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 DNA or RNA. In some embodiments, the nucleic acid cargo is a genetic, protein, or peptide tag and/or barcode.

[0158] In certain embodiments, the constructs and methods described herein may be utilized for monitoring a biological or biochemical cellular condition or circuits, such as pH via a marker. In some embodiments, the constructs and methods described herein may be utilized for recording, via writing directly to a genome or intracellular DNA element, cellular, environmental, chemical, or other cellular temporal or spatial related events. In some embodiments, the constructs and methods described herein may be utilized for recording, via writing directly to a genome or intracellular DNA element, cellular lineage information.

[0159] In certain embodiments, the genome to be programmably inserted into is eukaryotic or prokaryotic. In certain embodiments, the genome is mammalian, nonmammalian, human, murine, or NHP.

[0160] In additional embodiments, constructs and methods described herein may be utilized in agricultural settings for production of crops with improved properties or traits as well as to produce livestock, such as cattle, avian, or other species with improved or desirable features.

6.10. Integrase- or recombinase-mediated self-circularization of a subsequence of the single nucleic acid construct

[0161] In some embodiments, the single nucleic acid construct comprises a sub-sequence of the nucleic acid construct that is capable of self-circularizing to form a self-circular nucleic acid. In some embodiments, the single nucleic acid construct comprises a physical portion or region of the nucleic acid construct that is capable of self-circularizing to form a circular construct. As used herein, the term “sub-sequence” refers to a portion of the single nucleic acid construct that is capable of self-circularizing, where the subsequence is flanked by integrase 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 single nucleic acid construct. Recombination occurs when the single nucleic acid construct is contacted with an integrase or a recombinase under conditions that allow for recombination of the cognate pair or integrase or recombinase recognition sites.

[0162] In some embodiments, the sub-sequence of the single nucleic acid construct 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 single nucleic acid includes a first recombinase recognition site, a second recombinase recognition site, and an integrase recognition site (e.g., a second integrase recognition site), where the first and second recombinase recognition sites flank the integrase 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 (see, e.g., **FIG. 1**).

[0163] In some embodiments, the sub-sequence of the single nucleic acid construct includes a third integrase recognition site and a fourth integrase recognition site, wherein the third and fourth integrase recognition sites are a cognate pair. In some embodiments, the subsequence of the single nucleic acid construct includes the second integrase recognition site, the third integrase recognition site, the fourth integrase recognition site, where the third and fourth integrase recognition sites flank the second integrase. In such cases, the third integrase recognition site, the fourth integrase recognition site, and an integrase enable self - circularization and formation of the circular construct. In such cases, the third integrase recognition site and/or the fourth integrase recognition sites cannot recombine due, in part, to having different central dinucleotides with the first integrase recognition site and/or the second integrase recognition site.

[0164] In some embodiments where the subsequence includes three or more integrase recognition sites, each integrase recognition site or each pair of integrase recognition is capable of being recognized by a different integrase. In some embodiments where the subsequence includes three or more integrase recognition sites, each integrase recognition site or each pair of integrase recognition comprises a different central dinucleotide.

[0165] In some embodiments, self-circularizing is mediated at the integrase recognition sites or recombinase recognition sites. In some embodiments, the self-circularizing is mediated by an integrase or a recombinase.

[0166] In some embodiments, upon introducing the nucleic acid construct into a cell and after self-circularizing to form the self-circular nucleic acid, the self-circular nucleic acid comprising the second integrase recognition site is capable of being integrated into the cell's genome at the target sequence that contains the first integrase recognition site.

[0167] In some embodiments, following self-circularization, the self-circular nucleic acid comprises one or more additional integrase 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 integrase 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 integrase recognition sites into the genome along with the nucleic acid cargo. The integrated one or more integrase recognition sites serve as an integrase 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 integrase recognition sites with an integrase and the second additional nucleic that includes a sequence that is an integration cognate to the one or more integrase recognition sites, thereby integrating the additional nucleic acid cargo.

[0168] In some embodiments, the self-circular nucleic acid includes a second integrase recognition site that is capable of being integrated into a genomic locus that contains the first integrase recognition site (i.e., the first and second integrase recognition sites are a cognate pair). See, **FIGs. 1-2**.

[0169] In some embodiments, the single nucleic acid construct comprises two recombinase recognition sites where the two sites flank the nucleic acid cargo. In such cases, the two recombinase recognition sites are capable of self-circularizing to form a self-circular nucleic acid when contacted with a recombinase. **FIG. 1** illustrates a non-limiting example of a single nucleic acid construct that includes two recombinase recognition sites capable of self-circularizing to form a circular construct (e.g., a self-circular nucleic acid) when contacted with a recombinase. In **FIG. 1**, **101** and **102** are recombinase recognition sites present in the single nucleic acid construct. The single nucleic acid construct also includes a sequence encoding a

recombinase **103**. The recombinase **103** is expressed **104** and contacts **105** the recombinase recognition sites (**101** and **102**), thereby mediating self-circularization of a portion of the single nucleic acid construct and producing a self-circular nucleic acid **106**.

[0170] In some embodiments, the self-circular nucleic acid **106** includes a sequence **107** that is an integration cognate (e.g., a cognate pair) to the first integrase recognition sequence **108**. In such cases, the self-circular nucleic acid is integrated into a genome at the incorporation site of the first integrase recognition site. In some embodiments, integration of the self-circular nucleic acid into the genome is mediated by an integrase. For example, **FIG. 1** illustrates a non-limiting example where the single nucleic acid construct also includes a sequence encoding an integrase **109**. The integrase **109** is expressed and integrates **110** the circular construct **106** into the first integrase recognition site **108** site-specifically incorporated into the genome.

[0171] In some embodiments, the nucleic acid construct comprises two integrase recognition sites where the two sites flank the nucleic acid cargo. In such cases, the two integrase recognition sites are capable of self-circularizing to form a self-circular nucleic acid when contacted with an integrase. **FIG. 2** illustrates a non-limiting example of a single nucleic acid construct that includes two integration sequences capable of self-circularizing to form a circular construct (e.g., a self-circular nucleic acid) when contacted with a recombinase. In **FIG. 2**, **201** and **202** are integrase recognition sites (e.g., the third and fourth integrase recognition sites) present in the single nucleic acid construct. The single nucleic acid construct also includes a sequence encoding an integrase **203**. The integrase **203** is expressed **204** and contacts **205** the integrase recognition sites (**201** and **202**), thereby mediating self-circularization of a portion of the single nucleic acid construct and producing a self-circular nucleic acid **206**.

[0172] In some embodiments, the self-circular nucleic construct **206** includes a sequence **207** that is a cognate pair to the site-specifically incorporated integration sequence **208**. As shown in **FIG. 2**, one embodiment uses the same integrase for both self-circularizing and integration of the self-circular nucleic acid. The integrase **203** is expressed **204** and integrates **210** the self-circular nucleic acid **206** into the first integrase recognition site **208** site-specifically incorporated into the genome.

[0173] High efficiency and/or fast integrase recognition target sites allow for integrase-mediated template circularization to happen prior to integrase-mediated genomic integration at an integrase recognition target site within the genome (i.e. “beacon” or “landing pad”). In some embodiments, the integration rate can be altered by changing the dinucleotide used within the integrase recognition target site. In some embodiments, the integration rate can be altered by changing the integrase recognition target site sequence length. In some embodiments, the integration rate can be altered by changing the dinucleotide used within the integrase recognition target site and by changing the integrase recognition target site sequence length. For example, the attB/attP integrase recognition target site sequence length can be about 32-46 bp in length. In some embodiments, high efficiency and/or fast integrase target recognition is mediated by orthogonal integrases or recombinases.

[0174] In some embodiments where a single nucleic acid construct includes a first cognate pair (e.g., a first integrase recognition site and a second integrase recognition site) and a second cognate pair (e.g., a third integrase recognition site and a fourth recognition site), the first cognate pair and the second cognate pair are designed such that each cognate pair has a different integration rate. In such embodiments, the cognate pair with the faster integration rate recombines prior to the cognate pair with the slower integration rate. For example, as shown in **FIG. 2**, the first cognate pair is represented by **207** and **208** and the second cognate pair is represented by **201** and **202**. In one embodiment of the illustration in **FIG. 2**, the second cognate pair (i.e., **201** and **202**) has a faster integration rate whereby self-circularization occurs prior to integration into the genome.

[0175] In some embodiments, the self-circularizing is effected at an integrase or recombinase recognition target sequence. In typical embodiments, the self-circularizing is mediated by an integrase or a recombinase.

[0176] 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.

[0177] In some embodiments, the DNA cargo contains one or more orthogonal recombinase recognition target site(s). In some embodiments, the DNA cargo contains one or more orthogonal integrase recognition target site(s). The region that contains one or more orthogonal recombinase or integrase recognition target site(s) may be referred to as a multiple access site. Further, after DNA cargo integration into a genomic locus, the additional one or more orthogonal recombinase or integrase target recognition site(s) contained within the inserted DNA cargo may be subsequently targeted via a recombinase or integrase to incorporate additional DNA cargo. The DNA cargo may contain one or one or more orthogonal recombinase or integrase target recognition site(s). Hence, because each newly genomically incorporated DNA template, insert, or DNA cargo, may contain at least one “embedded” or “nested” orthogonal recombinase or integrase target recognition site(s) it becomes possible to programmatically (spatially and temporally) access, introduce, delete, and modify a genomic- or DNA-locus of interest at the orthogonal recombinase or integrase target recognition site(s).

[0178] In typical embodiments, the self-circular nucleic acid is capable of being integrated into a genomic locus that contains an integrase or recombinase recognition site (i.e., “beacon” or “landing pad” site). In typical embodiments, the self-circular nucleic acid contains the DNA cargo of interest. In some embodiments, the integrase or recombinase that mediates self-circularization is fused or linked to the prime editor protein fusion.

[0179] In typical embodiments, the nucleic acid construct that contains a nucleotide sequence encoding an integrase, encoding a prime editor fusion protein or a gene writer protein, a nucleotide sequence encoding one or more attachment site-containing guide RNA (atgRNA), optionally a nucleotide sequence encoding a nickase guide RNA (ngRNA), a nucleotide sequence encoding an integrase, a DNA cargo, and optionally a nucleotide sequence encoding a recombinase is vectorized.

[0180] In some embodiments, an integration target recognition site is incorporated (i.e., beacon placement) into a human primary cell genome using a single atgRNA and a single nicking guide RNA (ngRNA). In some embodiments, an integration target recognition site is incorporated into a human primary cell genome using two atgRNAs (dual or paired or twin atgRNAs). In certain embodiments, the nucleic acid construct comprises two atgRNAs.

[0181] In some embodiments, the atgRNA 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 comprised of a reverse transcriptase template. 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).

6.11. Genes and Targets

[0182] This disclosure provides compositions and 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.

[0183] 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. 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). 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)**

[0184] Over 2500 mutations have been identified associated with various diseases and defects.

[0185] 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 certain embodiments, appropriate cells include epithelial cells which may be derived from iPSCs. Proposed validation is detection of persistent CFTR mRNA and protein expression in transduced cells.

[0186] **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. Appropriate test cells include erythroid cells which may be derived from iPSCs. In some embodiments, validation is detection of persistent HBB mRNA and protein expression in transduced cells.

[0187] **DMD - Duchenne Muscular Dystrophy**

[0188] 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).

[0189] 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.

[0190] 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. Appropriate test target cells include cardiomyocytes derived from iPSCs. Proposed validation is detection of persistent DMD mRNA and protein expression in transduced cells.

[0191] F8 (Factor VIII)

[0192] 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.

[0193] In some embodiments, correcting factor VIII deficiency by recombining the FVIII gene into an SHS is a corrective path. Appropriate test target cells include liver cells and endothelial cells which may be derived from iPSCs. Proposed validation is detection of persistent FVIII mRNA and protein expression in transduced cells.

6.12. Methods of treatment

[0194] 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.

[0195] 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.

[0196] Methods of non-viral delivery of the single nucleic acid construct 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).

6.12.1.1 Lipid Nanoparticle Delivery

[0197] In some embodiments, the single nucleic acid construct is packaged in a LNP and administered intravenously. In some embodiments, the single nucleic acid construct is packaged in a LNP and administered intrathecally. In some embodiments, the single nucleic acid construct is packaged in a LNP and administered by intracerebral ventricular injection. In some embodiments, the single nucleic acid construct is packaged in a LNP and administered by intracisternal magna administration. In some embodiments, the single nucleic acid construct is packaged in a LNP and administered by intravitreal injection.

[0198] 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).

[0199] 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.

[0200] 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-dilinoeoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoeoyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoeoyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoeoyl-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.

[0201] 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-dilinoeoyloxy-N,N-dimethylaminopropane (DLin-DMA), 2,2-dilinoeoyl-4-dimethylaminomethyl-[1,3]-dioxolane (DLin-K-DMA), heptatriaconta-6,9,28,31-tetraen-19-yl 4-(dimethylamino)butanoate (DLin-MC3-DMA), 2,2-dilinoeoyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLin-KC2-DMA), 1,2-dioeoyloxy-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.

[0202] 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.

[0203] 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).

[0204] 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.

[0205] 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.

[0206] 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.

[0207] 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.

[0208] 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.

[0209] 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.

[0210] 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.

[0211] 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).

[0212] 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.

[0213] 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-dilinoeoyl-sn-glycero-3-phosphocholine (DLPC), 1,2-dimyristoyl-sn-glycero-phosphocholine (DMPC), 1,2-dioeoyl-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 (OChemsPC), 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-didocosaheptaenoyl-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-didocosaheptaenoyl--sn-glycero-3-phosphoethanolamine, 1,2-dioleoyl-sn-glycero-3-phospho-rac-(1-glycerol) sodium salt (DOPG), sphingomyelin, and any mixtures thereof.

[0214] 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.

[0215] 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 disclosed 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.

[0216] 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.

[0217] 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-diacylglyceramide (PEG-DAG), PEG-dipalmitoyl phosphatidylethanolamine (PEG-DPPE), or PEG-1,2-dimyristyloxypropyl-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.

[0218] 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.

[0219] 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.

[0220] 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; Klibanov 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).

[0221] 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.

[0222] 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.

[0223] 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., *Bioconjug 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).

[0224] 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.

[0225] 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.

[0226] 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.

[0227] 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.

[0228] LNPs 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.

[0229] 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.

[0230] 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.

[0231] 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.

6.12.2. Viral Vector Delivery

[0232] 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). As 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.

[0233] 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.

[0234] 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.

[0235] 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.

[0236] 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.

[0237] 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^{10} 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, 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.

[0238] 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.

[0239] 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)

[0240] 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.

[0241] 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.

[0242] 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.

[0243] 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.

[0244] 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.

[0245] 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.

[0246] 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.

[0247] 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).

[0248] 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.

[0249] 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.

[0250] 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.

[0251] 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, HEKn, HEKa, 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.

[0252] 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.

[0253] 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).

[0254] 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 races 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.

[0255] 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.

7. EXAMPLES

7.1. Example 1: **Single nucleic acid construct comprising PASTE components and a nucleic acid cargo of interest that is capable of recombinase-mediated subsequence circularization effects targeted integration of the cargo into a genomic locus**

[0256] A single construct “installer” that contains a prime editor fusion protein, an attachment site-containing guide RNA (atgRNA), a nickase guide RNA (ngRNA), an integrase, a recombinase, recombination target sites, integration target site, a DNA of interest, and flanking ITRs is designed (**FIG. 1**). Following delivery of the single nucleic acid construct “installer”, recombinase expression and binding at recombinase recognition sites leads to self-circularization of a subsequence of the single nucleic acid construct. A DNA of interest (e.g. gene) contained within the self-circularized nucleic acid integrates into a genomic locus of interest via an integrase. Genomic integration occurs at an integrase recognition target site (i.e., “beacon”) placed via prime editing or gene writing. For additional disclosure regarding the nucleic acid construct, self-circularization and integration see, for example, Section 6.9 and 6.10.

7.2. Example 2: **Single nucleic acid construct comprising PASTE components and a nucleic acid cargo of interest that is capable of integrase-mediated subsequence circularization effects targeted integration of the cargo into a genomic locus**

[0257] A single construct “installer” that contains a prime editor fusion protein, an attachment site-containing guide RNA (atgRNA), a nickase guide RNA (ngRNA), an integrase, integration target sites, a DNA of interest, and flanking ITRs is designed (**FIG. 2**). Following delivery of the single nucleic acid construct “installer”, integrase expression and binding at integrase recognition sites (attP2/attB2) leads to self-circularization of a subsequence of the single nucleic acid construct.

[0258] Stepwise control of self-circularization followed by genomic integration is achieved by use of central dinucleotide matched orthogonal integrase target recognition sites (i.e., attB/attP pairs) (**FIG. 3D** and **FIG. 4D**). Additionally, use of a kinetically fast attB/attP pair integrated into the single nucleic acid construct allows self-circularization prior to genomic integration. Screening of attB/attP pairs is achieved through a pooled attB/attP dinucleotide orthogonality assay (**FIG. 4C**) and relative insertion preferences for all attB/attP dinucleotide

pairs results shown in **FIG. 4E**. Improved genomic integration occurs via the selection of attP/attB mutant pairs (**FIG. 3A**) that demonstrate improved integration efficiency (**FIGs. 3B-C** and **FIGs. 4A-4B**).

[0259] A DNA of interest (e.g., gene) contained within the self-circularized nucleic acid integrates into a genomic locus of interest via the integrase via the attP1/attB1 sites. Genomic integration occurs at an attB1 integrase recognition target site (i.e., “beacon”) placed via prime editing or gene writing.

7.3. Example 3: Single nucleic acid construct comprising PASTE components wherein an integrase is linked to a prime editor and a nucleic acid cargo of interest that is capable of integrase-mediated subsequence circularization effects targeted integration of the cargo into a genomic locus

[0260] A single construct “installer” that contains a prime editor fusion protein linked to an integrase (**FIG. 6**), an attachment site-containing guide RNA (atgRNA), a nickase guide RNA (ngRNA), an integrase, integration target sites, a DNA of interest, and flanking ITRs is designed. Following delivery of the single nucleic acid construct “installer”, prime editor-integrase fusion (Cas9-RT-Integrase) expression and binding at integrase recognition sites (attP2/attB2) leads to self-circularization of a subsequence of the single nucleic acid construct.

[0261] Stepwise control of self-circularization followed by genomic integration is achieved by use of central dinucleotide matched orthogonal integrase target recognition sites (i.e., attB/attP pairs) (**FIG. 3D** and **FIG. 4D**). Additionally, use of a kinetically fast attB/attP pair integrated into the single nucleic acid construct allows self-circularization prior to genomic integration. Screening of attB/attP pairs is achieved through a pooled attB/attP dinucleotide orthogonality assay (**FIG. 4C**) and relative insertion preferences for all attB/attP dinucleotide pairs results shown in **FIG. 4E**. Improved genomic integration occurs via the selection of attP/attB mutant pairs (**FIG. 3A**) that demonstrate improved integration efficiency (**FIG. 3B** and **FIG. 4B**).

[0262] A DNA of interest (e.g., gene) contained within the self-circularized nucleic acid integrates into a genomic locus of interest via the integrase via the attP1/attB1 sites. Genomic integration occurs at an attB1 integrase recognition target site (i.e., “beacon”) placed via prime editing mediated by the prime editor-integrase fusion.

[0263] **FIG. 5** illustrates a schematic of single atgRNA and dual atgRNA approaches for beacon placement. The single construct “installer” that contains a prime editor fusion protein linked to an integrase (**FIG. 6**), a first attachment site-containing guide RNA (atgRNA), a second attachment site-containing guide (atgRNA), an integrase, integration target sites, a DNA of interest, and flanking ITRs is designed. In this version of the single construct “installer” the first atgRNA and the second atgRNAs collectively encode the entirety of the integration recognition site.

7.4. Example 4: Extrachromosomal circular DNA (EccDNA) sensor to evaluate template integrase-mediated circularization and programmable gene insertion within a *ACTB* beacon locus

[0264] A dual reporter (Nanoluc and GFP) extrachromosomal circular DNA (EccDNA) sensor capable of detecting Bxb1-mediated self-circularization was designed (**FIG. 7**). Bxb1-mediated circularization of the EccDNA sensor, which occurs at a attP'/attB' target recognition site within the EccDNA sensor, orients the EF1 α promoter upstream of nanoluc and GFP, thereby allowing for dual reporter expression. EccDNA circularization can also be confirmed by PCR amplification of the post-circularization attR' scar using primers P1 and P2 as shown in **FIG. 7**. Total EccDNA (linear and circularized) is quantified by primers P3 and P4 as shown in **FIG. 7**. The EccDNA construct contains an orthogonal attP (GT central dinucleotide, see **FIGS. 4A** and **4D**) to facilitate genomic insertion at a placed attB beacon site. Genomic integration of the EccDNA is verified using primers P5 and P6 (**FIG. 7**).

[0265] A transfection screen was performed to confirm Bxb1-mediated EccDNA circularization (**FIG. 8**). Plasmid expressed EccDNA sensor, prime editor protein, Bxb1, *ACTB* targeting atgRNA, and nicking guide RNA were transfected using Lipo3000 into HEK293T cells (200K cells in a 12-well plate). Cell samples were harvested 72 hours post transfection for circularization, beacon placement, and insertion analysis.

[0266] As confirmed by ddPCR, transfection of both EccDNA sensor and Bxb1 resulted in confirmed intracellular circularization (**FIG. 9**). Circularization efficiency was >50% for Bxb1-containing samples tested at a 25,000-fold dilution, whereas equivalent samples that lacked Bxb1 demonstrated <1% circularization. In addition to Bxb1 transfection, circularization occurred with plasmid-form transfection of PE2 prime editor, *ACTB* targeting atgRNA, and nicking guide RNA (**FIG. 10**), albeit at <4% circularization efficiency. It is hypothesized that the drop in circularization efficiency is due an interaction between the

plasmid-form atgRNA attB and the EccDNA AttP in the presence of BxB1. Unwanted cross talk is mitigated by use of synthetic RNAs that contain stabilizing chemical modifications.

[0267] Beacon placement facilitated by the plasmid-form transfection of PE2 prime editor, *ACTB* targeting atgRNA, and nicking guide RNA was verified by ddPCR (FIG. 11). Beacon placement efficiency was >40% for samples containing the requisite beacon placement PE2/atgRNA/ngRNA components, however samples that also included Bxb1 demonstrated <20% beacon placement. It is hypothesized that the drop in beacon placement efficiency is due an interaction between the plasmid-form atgRNA attB and the EccDNA AttP in the presence of BxB1. FIG. 12 demonstrates programmable gene insertion of the EccDNA at the *ACTB* beacon locus was confirmed by ddPCR.

7.5. Example 5: Extrachromosomal circular DNA (EccDNA) sensor to evaluate template integrase-mediated circularization and programmable gene insertion within a *LMNB* placed beacon

[0268] A transfection screen was performed to confirm Bxb1-mediated EccDNA circularization and subsequent programmable gene insertion at a *LMNB* placed attB beacon site. To mimic linear viral genomic DNA and to eliminate the potential for unwanted genome insertion of a transfected plasmid directly, a linearized EccDNA sensor was tested in cell transfections (FIG. 13). An EccDNA sensor called EccDNA-NC1 which lacks the attP'/B' cognate pair was developed as a non-circularizing negative control. *LMNB* targeting atgRNA and nicking guide RNA were transfected as synthetic RNAs (containing standard IDT chemical modifications). Prime editor protein and Bxb1 effectors were transfected in plasmid form. Transfection was conducted across 300,000 HEK293T cells in a 24-well plate format using Lipo3000 for plasmid delivery (PE2, BxB1, and EccDNA sensors) in conjunction with Lipo mRNAMAX for synthetic RNA delivery (atgRNA, ngRNA). Cell samples were harvested 72 hours post transfection for circularization, beacon placement, and insertion analysis.

[0269] Intracellular circularization of the EccDNA sensor in the presence of BxB1 was confirmed via GFP expression (FIG. 14). In a ddPCR format, co-delivery of EccDNA with BxB1 also demonstrated circularization (FIG. 15), whereas no circularization was observed in either the no BxB1 control or across any of the EccDNA-NC1 control replicates. EccDNA circularization was observed in the presence of BxB1 and PE2/atgRNA/ngRNA (FIG. 15).

[0270] Transfection of PE2 (plasmid form) with atgRNA/ngRNA (synthetic RNA form) did result in *LMNB* beacon placement, however at <5% beacon placement efficiency, with a further drop in efficiency observed when Bxb1 is co transfected (**FIG. 16**). Low (~1-2%) PGI of the linear EccDNA was observed Co- at the *LMNB* placed beacon (**FIG. 17**).

7.6. Example 6: Programmable gene insertion with a single nucleic acid construct (HDA) in mouse cells

[0271] In this example, a single nucleic acid construct having PGI components “all-in-one” (i.e., nucleotide sequence encoding the prime editor fusion protein, nucleotide sequence encoding a first atgRNA, a nucleotide sequence encoding a second atgRNA, a nucleotide sequence encoding an integrase, and a nucleic acid cargo) was compared with a four plasmid system to see which resulted in greater beacon placement, PGI, and PGI conversion rate.

[0272] An “all-in-one” construct as shown in **FIG. 18** was cloned in an adenoviral backbone (a helper dependent Adenoviral backbone) (SEQ ID NO: 559) using multistep Gibson assembly. Two clones (i.e., C5 and C8) were selected and used for further analysis. For the four plasmid system, the same components as shown in **FIG. 18** were cloned into four separate plasmids (e.g., a plasmid with a nucleotide sequence encoding a prime editor fusion protein and a nucleotide sequence encoding an integrase, a second plasmid encoding a first atgRNA, a third plasmid encoding a second atgRNA, and a fourth plasmid having the nucleic acid cargo).

[0273] Mouse Hepa 1-6 cells were transfected in a 48 well format with 50,000 cells per well seeded 1 day prior to transfection. Total of 200 ng plasmid DNA was transfected in each well using Lipfectamine 3000 (ThermoFisher) using 3:1 (Lipo3000 : DNA). As shown in **FIG. 18**, RFP driven by an EF1alpha promoter was used a marker for transduction. **FIGs. 19A-19J** shows successful transduction for both clones with RFP positive cells at day 2 post transfection. 72 hours after transfection RNA was collected and subjected to ddPCR and NGS analysis to assess beacon placement and PGI. Data for ddPCR is shown in **FIGs. 20A-20B**, **FIGs. 21A-21B**, and **FIG. 22**. NGS data is shown in in **FIGs. 23A-23B** and **FIG. 24**.

[0274] Beacon placement at the *NolcI* site in mouse Hepa 1-6 cells was detected using ddPCR (**FIG. 20A** and **FIG. 20B**). In particular, transfection of both single nucleic acid constructs (both clones) resulted in beacon placement at the *NolcI* site but was lower than when PGI components were delivered using a four plasmid system.

[0275] Once expressed BxB1 mediated PGI at the *Nolc1* site. In particular, PGI was detected at the *Nolc1* site in mouse Hepa 1-6 cells using ddPCR for both single nucleic acid constructs (both clones) but exhibited lower PGI than when PGI components were delivered using a four plasmid system (**FIG. 21A** and **FIG. 21B**).

[0276] Analysis of PGI conversion rate, calculated as $\text{PGI}\% / (\text{PGI}\% + \text{BP}\%)$, for the data in **FIGs. 20A-20B** and **FIGs. 21A-21B** show a higher PGI conversion rate when using the single nucleic acid construct as compared to the four plasmid system (**FIG. 22**). PGI conversion rate identifies the percentage of beacons where PGI occurred (i.e., integration of the nucleic acid cargo), thereby serving as a proxy for PGI efficiency.

[0277] Beacon placement and PGI were confirmed using next generation sequencing (NGS). As shown in **FIGs. 23A-23B** beacon placement (**FIG. 23A**) and PGI (**FIG. 23B**) were higher with the four plasmid system. However, the PGI conversion rate for the data in **FIG. 23A** and **FIG. 23B** showed a higher PGI conversion rate for both of the single nucleic acid constructs (both clones) as compared to the four plasmid system (**FIG. 24**).

[0278] Overall, this data shows successful PGI using a single nucleic acid construct in mouse cells. Additionally, this data shows that delivering all of the PGI components in a single nucleic acid construct results in more efficient PGI (i.e., higher PGI conversion rate) than when the delivering the components in separate plasmids.

7.7. Example 7: Programmable gene insertion with a single nucleic acid construct (HDAd) in human cells

[0279] In this example, a single nucleic acid construct having PGI components “all-in-one” (i.e., nucleotide sequence encoding the prime editor fusion protein, nucleotide sequence encoding a first atgRNA, a nucleotide sequence encoding a second atgRNA, a nucleotide sequence encoding an integrase, and a nucleic acid cargo) was compared with a four plasmid system to see which resulted in greater beacon placement and PGI.

[0280] The same construct shown in **FIG. 18** and used in Example 6 was also used for these experiments. Similarly, the same four plasmid system used in Example 6 was also used for these experiments.

[0281] human hHepG2 cells were transfected in a 48 well format with 50,000 cells per well seeded 1 day prior to transfection. Total of 300 ng plasmid DNA was transfected in each well

using Lipofectamine 3000 (ThermoFisher) using 3:1 (Lipo3000:DNA) with further experimental details provided in **Table 12**.

Table 12									
#	Cells	Plasmid1	Plasmid2	Plasmid3	Plasmid4	Total (ng)	Lipo3000 (uL)	P3000	Opti-MEM uL
13	hHepG2	AdVG012-1				300	0.9	0.6	10+10
14	hHepG2	duplicate							
15	hHepG2	AdVG012-2				300	0.9	0.6	10+10
16	hHepG2	duplicate							
17	hHepG2	PL216 (50 ng)	hF9 atgF (100ng)	hF9 atgR (70ng)	CNGNC (80ng)	300	0.9	0.6	10+10
18	hHepG2	duplicate							
19	hHepG2	NC							
20	hHepG2	duplicate							

[0282] **FIGs. 25A-25L** show the results at day 2 post transfection. **FIGs. 25E** and **25F** show successful adenovirus transduction for both all-in-one clones (RFP is a marker for all-in-one systems (“AIO-012-1” and “AIO-012-2”)) at day 2 post transfection. **FIGs. 25K** and **25L** show GFP expression (marker for four plasmid system (“4plasmids-hF9”)) at day 2 post transfection. 72 hours after transfection RNA was collected and subjected to ddPCR and NGS to assess beacon placement and PGI. ddPCR data for beacon placement is shown in **FIGs. 26A-26B**. ddPCR data for PGI is shown in **FIGs. 27A-27B**.

[0283] Beacon placement at the *human Factor IX site* in human HepG2 cells was detected using ddPCR (**FIG. 26A** and **FIG. 26B**). In particular, transfection of both single nucleic acid constructs (both clones) resulted in beacon placement at the *human Factor IX* site but was lower than when PGI components were delivered using the four plasmid system.

[0284] Once expressed BxB1 mediated PGI at the *human Factor IX I* site. In particular, PGI was detected at the *human Factor IX* site using ddPCR for both single nucleic acid constructs (both clones) but exhibited lower PGI than when PGI components were delivered using a four plasmid system (**FIG. 27A** and **FIG. 27B**).

[0285] Overall, this data shows successful PGI using a single nucleic acid construct in human cells.

8. EQUIVALENTS AND INCORPORATION BY REFERENCE

[0286] 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.

[0287] 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.

[0288] 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 nucleic acid construct comprising:
 - a) a nucleotide sequence encoding a prime editor system;
 - b) a nucleotide sequence encoding at least a first attachment site-containing guide RNA (atgRNA);
 - c) a nucleotide sequence encoding at least a first integrase;
 - d) a nucleic acid cargo;
 - e) optionally, a nucleotide sequence encoding a nickase guide RNA (ngRNA);
and
 - f) optionally a nucleotide sequence encoding a recombinase.
2. The nucleic acid construct of claim 1, wherein the prime editor system comprises a nucleotide sequence encoding a nickase and a nucleotide sequence encoding a reverse transcriptase.
3. The nucleic acid construct of claim 2, wherein the nucleotide sequence encoding the nickase and the nucleotide sequence encoding the reverse transcriptase are positioned in the construct such that when expressed the gene editor system comprises a fusion protein comprising the nickase and the reverse transcriptase.
4. The nucleic acid construct of any one of claims 1-3, wherein the first integrase that is encoded by a nucleotide sequence in the nucleic acid construct is fused to the prime editor system, the nickase, or the reverse transcriptase by a linker.
5. The nucleic acid construct of any one of claims 1-4, 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 a first integration recognition site.
6. The nucleic acid construct of claim 5, wherein the RT template comprises the entirety of the first integration recognition site.

7. The nucleic acid construct of any one of claims 1-6, wherein, upon introducing the nucleic acid construct into a cell, the first atgRNA incorporates the first integrase recognition site into the cell's genome at the target sequence.
8. The nucleic acid construct of any one of claims 1-7, further comprising a second atgRNA.
9. The nucleic acid construct of claim 8, 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; 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.
10. The nucleic acid construct of claim 9, wherein, upon introducing the nucleic acid construct into a cell, the first pair of atgRNAs incorporate the first integrase recognition site into the cell's genome at the target sequence.
11. The nucleic acid construct of any one of claims 1-10, further comprising a second integrase recognition site.
12. The nucleic acid construct of claim 11, wherein the second integrase recognition site and the first integrase recognition site are a first cognate pair.
13. The nucleic acid construct of claim 11 or 12, further comprising a third integrase recognition site.
14. The nucleic acid construct of any one of claims 11-13, further comprising a fourth integrase recognition site.
15. The nucleic acid construct of claim 14, wherein the third integrase recognition site and the fourth integrase recognition site are a second cognate pair.

16. The nucleic acid construct of any one of claims 10-15, wherein the second cognate pair has a faster integration rate than the first cognate pair, whereby in the presence of the first integrase the second cognate pair recombines prior to recombination of the first cognate pair.
17. The nucleic acid construct of any one of claims 1-16, further comprising a nucleotide sequence encoding a second integrase.
18. The nucleic acid construct of any one of claims 1-17, wherein the first integrase, the second integrase, or both, are selected from BxB1, Bcec, Sscd, Sacd, Int10, or Pa01.
19. The nucleic acid construct of claim 17 or 18, wherein the first integrase and the second integrase recognize different integration recognition sites.
20. The nucleic acid construct of any one of claims 1-19, further comprising at least a first recombinase recognition site.
21. The nucleic acid construct of claim 20, further comprising a second recombinase recognition site.
22. The nucleic acid construct of any one of claims 1-21, wherein the recombinase is FLP or Cre.
23. The nucleic acid construct of any one of claims 1-22, wherein the nucleic acid cargo comprises at least one of the following: a gene, an expression cassette, a logic gate system, or any combination thereof.
24. The nucleic acid construct of any one of claims 1-23, further comprising a sub-sequence of the nucleic acid construct that is capable of self-circularizing to form a self-circular nucleic acid.
25. The nucleic acid construct of claim 24, wherein the sub-sequence of the nucleic acid construct that is capable of self-circularizing includes the nucleic acid cargo, whereby upon self-circularizing the self-circular nucleic acid comprises the nucleic acid cargo.
26. The nucleic acid construct of claim 24 or 25, wherein the sub-sequence is flanked by the third integrase recognition site and the fourth integrase recognition site.

27. The nucleic acid construct of claim 26, wherein the sub-sequence includes the second integrase recognition site.
28. The nucleic acid construct of any one of claims 25-27, wherein self-circularizing is mediated by recombination of the third integrase recognition site and the fourth integration recognition site by the first integrase.
29. The nucleic acid construct of claim 28, wherein the sub-sequence is flanked by the first recombinase recognition site and the second recombinase recognition site.
30. The nucleic acid construct of claim 29, wherein self-circularizing is mediated by recombination of the first recombinase recognition site and a second recombinase recognition site by the recombinase.
31. The nucleic acid construct of any one of claims 24-30, wherein the self-circular nucleic acid comprises one or more additional integration recognition sites that enable integration of additional nucleic acid cargo.
32. The nucleic acid construct of any of claims 24-31, wherein, upon introducing the nucleic acid construct into a cell and after self-circularizing to form the self-circular nucleic acid, the self-circular nucleic acid comprising the second integrase recognition site is capable of being integrated into the cell's genome at the target sequence that contains the first integrase recognition site.
33. The nucleic acid construct of claim 32, wherein self-circularization to form the self-circular nucleic acid is effected by the first integrase and integration of the self-circular nucleic acid is effected by the second integrase.
34. The nucleic acid construct of any one of claims 1-33, further comprising a 5' inverted terminal repeat (ITR).
35. The nucleic acid construct of any one of claims 1-34, further comprising a 3' inverted terminal repeat (ITR).
36. A vector comprising any of the nucleic acid constructs of claims 1-35.

37. The vector of claim 36, wherein the vector is recombinant adenovirus, helper dependent adenovirus, AAV, lentivirus, HSV, anneovirus, retrovirus, Doggybone DNA (dbDNA), minicircle, plasmid, miniDNA, or nanoplasmid.
38. A pharmaceutical composition comprising any of the nucleic acid constructs or vectors of claims 1-37.
39. A method comprising administering an effective amount of a pharmaceutical composition of claim 38 to a patient in need thereof.