(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
09 November 2023 (09.11.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/215831 A1

(51) International Patent Classification:
C12N 15/113 (2010.01) *C12N 9/12* (2006.01)
C12N 15/90 (2006.01)

(21) International Application Number:
PCT/US2023/066609

(22) International Filing Date:
04 May 2023 (04.05.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/338,077 04 May 2022 (04.05.2022) US

(71) Applicant: **TOME BIOSCIENCES, INC.** [US/US]; 100 Talcott Avenue, Watertown, MA 02472 (US).

(72) Inventors: **FINN, Jonathan, Douglas**; Tome Biosciences, Inc., 100 Talcott Avenue, Watertown, MA 02472 (US). **DONG, Xiaolong**; Tome Biosciences, Inc., 100 Talcott Avenue, Watertown, MA 02472 (US).

(74) Agent: **BECKER, Daniel, M.** et al.; Fenwick & West LLP, 801 California Street, Mountain View, CA 94041 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: GUIDE RNA COMPOSITIONS FOR PROGRAMMABLE GENE INSERTION

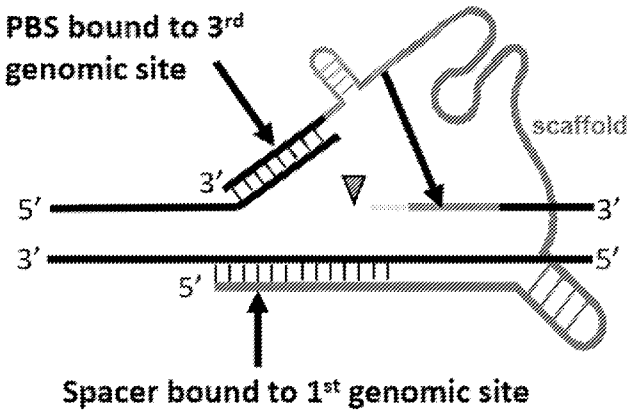

FIG. 4B

(57) **Abstract:** Provided herein are attachment site containing guide RNAs (atgRNAs) and compositions useful for site-specifically incorporating an integration recognition site into a cellular genome or intracellular nucleic acid sequence. For example, provided herein is an atgRNA comprising a spacer, a scaffold, a clamp, a reverse transcriptase template, and a primer binding site.

WO 2023/215831 A1

GUIDE RNA COMPOSITIONS FOR PROGRAMMABLE GENE INSERTION

1. CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 63/338,077, filed May 4, 2022; which is hereby incorporated in its entirety by reference.

2. SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing with 558 sequences, which has been submitted via Patent Center and is hereby incorporated by reference in its entirety. Said XML copy, created on May 4, 2023, is named 52322WO-SequenceListing.xml, and is 757,295 bytes in size.

3. BACKGROUND

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

[0004] There is a need in the art for techniques and gene editor compositions which address and overcome these shortcomings and enable the programmable genomic integration of a large gene(s).

4. SUMMARY

[0005] This disclosure features attachment site containing guide RNA (atgRNA) and guide RNA complex compositions capable of binding to a gene editor protein composition and mediating incorporation of an integration recognition site into a target DNA sequence. In some embodiments, the integration recognition stie (also referred to as a beacon) is placed in the genome by partial genomic replacement using single guide RNA or dual guide RNAs. In typical embodiments, the atgRNA comprises, from 5' to 3': a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence; a reverse transcriptase (RT) template comprising: an integration recognition site; and a clamp segment comprising a sequence where upon being reverse transcribed the sequence (reverse

complement clamp "rcClamp") is at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence; and a primer binding site complementary to a third genomic site of the double-stranded target DNA sequence, wherein the third genomic site is on a second strand of the target DNA sequence.

[0006] Inclusion of a clamp segment in the atgRNA allows for a reduction in the number of nucleotides in the primer binding site needed to bind the third genomic site and facilitate reverse transcription. By relaxing the requirement that the clamp binding site (i.e., second genomic site) be immediately adjacent to the primer binding site (i.e., third genomic site) or by allowing reduction in the number of nucleotides in the PBS, this disclosure provides for atgRNA compositions which are capable of being reduced in total sequence length as compared to standard prime editing guide RNAs or gene editing RNAs capable of integration recognition site insertion. The clamp segment binding site (i.e., second genomic site) can be selected from within a large genomic window, wherein for example, the genomic window frame can be shifted to screen for variable sequence distance from the third genomic site and/or constrained to select for efficient (e.g., melting temperature, rate of binding, low off target, etc.) hybridization to a defined length clamp segment. The present disclosure provides for single guide RNA, split guide RNA, and dual guide RNA methods for programmable gene insertion by partial genomic replacement.

[0007] 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. Non-limiting examples of PASTE are also described in U.S. Pat. 11,572,556, which is hereby incorporated by reference in its entirety.

[0008] In one aspect, this disclosure features an attachment site containing guide RNA (atgRNA) comprising from 5' to 3': a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence; a reverse transcriptase template comprising: an integration recognition site; and a clamp segment comprising a sequence where upon being reverse transcribed the sequence is at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence; and a primer binding site complementary to a third genomic site of the double-stranded target DNA sequence, wherein the third genomic site is on a second strand of the target DNA

sequence, wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site and the clamp segment.

[0009] In another aspect, this disclosure features an attachment site containing guide RNA (atgRNA) comprising from 5' to 3': a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence; a reverse transcriptase template comprising: an integration recognition site; and a clamp segment comprising a sequence where upon being reverse transcribed comprises a sequence at least partially complementary to a second genomic site of first strand of the double-stranded target DNA sequence; and a primer binding site comprising: a sequence complementary to a third genomic site of the double-stranded DNA target sequence, wherein the third genomic site is on a second strand of the target DNA sequence; and an anti-reverse transcriptase template (anti-RTT) sequence comprising a sequence at least partially complementary to all or a portion of the RT template, wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site and the clamp segment.

[0010] In another aspect, this disclosure features an attachment site containing guide RNA (atgRNA), comprising from 5' to 3': a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence; a clamp segment at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence; a reverse transcriptase (RT) template comprising an integration recognition site; and a primer binding site complementary to a third genomic site of the double-stranded target DNA sequence, wherein the third genomic site is on a second strand of the target DNA sequence; wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site.

[0011] In another aspect, this disclosure features an attachment site containing guide RNA (atgRNA) comprising from 5' to 3': a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence; a clamp segment at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence; a reverse transcriptase (RT) template comprising an integration recognition site; and a primer binding site comprising: a sequence complementary to a third genomic site of the double-stranded DNA target sequence, wherein the

third genomic site is on a second strand of the target DNA sequence; and an anti-reverse transcriptase template (anti-RTT) sequence comprising a sequence at least partially complementary to all or a portion of the RT template, wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site.

[0012] In some embodiments, upon introducing the atgRNA into a cell with a gene editor polypeptide, all or a portion of the single-stranded DNA sequence comprising the integration recognition site is site-specifically integrated into the double-stranded target DNA sequence.

[0013] In some embodiments, upon introducing the atgRNA into a cell with a gene editor polypeptide, the atgRNA comprising the clamp segment enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to the integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment.

[0014] In some embodiments, the atgRNA has a length of 150 to 190 nucleotides. In some embodiments, the atgRNA has a length of 160 to 180 nucleotides.

[0015] In some embodiments, the clamp segment has a length of 1 nucleotide to 40 nucleotides. In some embodiments, the clamp segment has a length of 1 to 10 nucleotides, 11 to 20 nucleotides, 21 to 30 nucleotides, or 31 to 40 nucleotides. In some embodiments, the clamp segment has a length of 10 nucleotides.

[0016] In some embodiments, the RT template has a length of 10 nucleotides to 200 nucleotides. In some embodiments, the RT template has a length of 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 or 50 nucleotides.

[0017] In some embodiments, the second genomic site and the third genomic site are separated by one or more genomic bases in the double-stranded target DNA sequence.

[0018] In some embodiments, the second genomic site and the third genomic site are separated by 1 to 10 genomic bases, 11 to 20 genomic bases, 21 to30 genomic bases, 31 to 40 genomic bases, or 41 to 45 genomic bases in the double-stranded target DNA sequence.

[0019] In some embodiments, the second genomic site and the third genomic site are separated by 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, or 45 genomic bases in the double-stranded target DNA sequence.

[0020] In some embodiments, the second genomic site and the third genomic site are adjacent in the double-stranded target DNA sequence.

[0021] In some embodiments, the clamp segment comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity to the second genomic site.

[0022] In some embodiments, the clamp segment comprises a sequence that is capable of binding to the second genomic site of the first strand of the double-stranded target DNA sequence with at least a first nucleotide mismatch.

[0023] In some embodiments, the clamp segment comprises a sequence that is capable of binding to the second genomic site of the first strand of the double-stranded target DNA sequence with a first nucleotide mismatch and a second nucleotide mismatch. In some embodiments, the first nucleotide mismatch and the second nucleotide mismatch are consecutive. In some embodiments, the first nucleotide mismatch and the second nucleotide mismatch are non-consecutive.

[0024] In some embodiments, the anti-RTT sequence comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity to all or a portion the RT template. In some embodiments, the anti-RTT sequence has a length between 1 nucleotide to 20 nucleotides, 5 nucleotides to 15 nucleotides, or 10 nucleotides to 20 nucleotides. In some embodiments, the anti-RTT sequence has a length of 10 nucleotides.

[0025] In some embodiments, upon introducing the atgRNA into a cell, the anti-RTT sequence increases stability of the atgRNA as compared to an atgRNA that does not comprise the anti-RTT sequence.

[0026] In another aspect, this disclosure features a system for site-specifically incorporating an integration recognition site into a mammalian cell genome at a target DNA sequence, comprising: any of the attachment site containing gRNAs (atgRNAs) described herein; and a gene editor polypeptide comprising a DNA binding nickase domain linked to a reverse transcriptase domain capable of incorporating the integration recognition site into the target DNA sequence, wherein, upon introducing into the cell, the system integrates the integration recognition site into the target DNA sequence.

[0027] In some embodiments, upon introducing the system into the cell, the atgRNA comprising the clamp segment enhances integration efficiency of the integration recognition site at the target

DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment.

[0028] In another aspect, this disclosure features a method for site-specifically incorporating an integration recognition site into a mammalian cell genome at a target DNA sequence, comprising: incorporating an integration target recognition site into the genome by delivering into the cell: any of the attachment site containing gRNAs (atgRNAs) described herein; and a gene editor polypeptide or polynucleotide encoding a gene editor polypeptide, wherein the gene editor polypeptide comprises a DNA binding nickase domain linked to a reverse transcriptase domain, wherein the integration recognition site is incorporated into the target DNA sequence by the gene editor polypeptide; and optionally, a nicking gRNA.

[0029] In some embodiments, the method using the atgRNA comprising the clamp segment enhances integration efficiency of the integration recognition site at the target DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment.

[0030] In another aspect, this disclosure features a system for site-specifically integrating a donor polynucleotide template into a mammalian cell genome at a target DNA sequence, comprising: any of the attachment site containing gRNAs (atgRNAs) described herein; a gene editor polypeptide comprising a DNA binding nickase domain linked to a reverse transcriptase domain capable of incorporating the integration recognition site into the target DNA sequence, an integration enzyme; and a donor polynucleotide template linked to a sequence that is an integration cognate of the integration recognition site present in the atgRNA, whereby the integration enzyme integrates the donor polynucleotide template into the target DNA sequence.

[0031] In some embodiments, upon introducing the system into the cell, the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site at the target DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment.

[0032] In another aspect, this disclosure features a method for site-specifically integrating a donor polynucleotide template into a mammalian cell genome at a target DNA sequence, comprising: incorporating an integration recognition site into the genome by delivering into the cell: any of the attachment site containing gRNAs (atgRNAs) described herein; and a gene editor polypeptide or polynucleotide encoding a gene editor polypeptide, wherein the gene editor polypeptide comprises

a DNA binding nickase domain linked to a reverse transcriptase domain and is capable of incorporating the integration recognition site into the target DNA sequence; and optionally, a nicking gRNA; and integrating the donor polynucleotide template into the genome by delivering into the cell: an integration enzyme; and a donor polynucleotide template, wherein the donor polynucleotide template is linked to a sequence that is an integration cognate of the integration recognition site present in the atgRNA, and wherein the donor polynucleotide template is integrated into the cellular genome at the incorporated genomic integration recognition site by the integration enzyme.

[0033] In some embodiments, the method enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment.

5. BRIEF DESCRIPTION OF THE DRAWINGS

[0034] These and other features, aspects, and advantages of the present disclosure will become better understood with regard to the following description, and accompanying drawings, where:

[0035] FIG. 1A-1E shows analysis of AttP variants. FIG. 1A shows a non-limiting schematic of AttP mutations tested for improving integration efficiency (SEQ ID NOS: 394 and 540-542, respectively, in order of appearance). FIG. 1B shows integration efficiencies of wildtype and mutant AttP sites across a panel of AttB lengths. FIG. 1C shows a non-limiting schematic of multiplexed integration of different cargo sets at specific genomic loci. Three fluorescent cargos (GFP, mCherry, and YFP) are inserted orthogonally at three different loci (ACTB, LMNB1, NOLC1) for in-frame gene tagging. FIG. 1D shows orthogonality of top 4 AttB/AttP dinucleotide pairs evaluated for GFP integration with PASTE at the ACTB locus. FIG. 1E shows efficiency of multiplexed PASTE insertion of combinations of fluorophores at ACTB, LMNB1, and NOLC1 loci. Data are mean (n= 3) ± s.e.m.

[0036] FIG. 2 illustrates a schematic of single atgRNA and dual atgRNA approaches for step 1: beacon placement ("integration recognition site") and step 2: gene insertion, thereby illustrating PASTE.

[0037] FIGs. 3A-3C illustrates a version 1 attachment site-containing guide RNA (atgRNAv1) (FIG. 3A) and a mechanism by which the atRNAv1 can be used for site-specifically integrating an integration recognition site into a mammalian cell genome at a desired target site (e.g., a target

DNA sequence) (FIGs. 3B-3C). FIG. 3A shows an atgRNA-v1 comprising from 5' to 3': a spacer sequence that binds to a first genomic site, a scaffold sequence, a clamp segment comprising a sequence that is complementary (binds) to a second genomic site or comprising a sequence (see rcClamp) where upon being reverse transcribed the sequence is at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence, a sequence encoding an integration recognition site ("AttB"), and a primer binding stie (PBS) that binds to a third genomic site. FIG. 3B shows atgRNA v1 with the spacer hybridized to the first genomic site and the PBS hybridized to the third genomic site, where the nick is indicated by the inverted triangle. FIG. 3C shows a mechanism by which the atgRNA-v1 as shown in FIGs. 3A-3B can be used for site-specifically integrating an integration recognition site into a mammalian cell genome at a desired target site (e.g., a target DNA sequence). For example, a nick (as indicated by the inverted triangle) on the non-targeting strand induced by a nickase (e.g., nCas9) facilitates binding of the PBS to a sequence 5' to the nick site on the non-targeting strand. Following "3' flap synthesis by a reverse transcriptase, the clamp (i.e., reverse complement of the clamp) binds the targeting strand immediately 3' to the site of the nick on the non-targeting strand. Following favored repair, the integration recognition site (e.g., AttB) is integrated into the genome at the target sequence.

[0038] FIGs. 4A-4C illustrate a version 2 attachment site-containing guide RNA (atgRNAv2) (FIG. 4A) and a mechanism by which the atRNAv2 can be used for site-specifically integrating an integration recognition site into a mammalian cell genome at a desired target site (e.g., a target DNA sequence) (FIGs. 4B-4C). FIG. 4A shows an atgRNA-v2 comprising 5' to 3': a spacer sequence that binds a first genomic site, a scaffold sequence, a clamp comprising a sequence that has complementarity (binds) to a second genomic site which is 3' to the nick on the targeting strand (i.e., a one or more separation between the second genomic site and a third genomic site) or comprising a sequence (see rcClamp) where upon being reverse transcribed the sequence is at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence, a sequence encoding an integration recognition site (e.g., AttB), and a primer binding site (PBS) that binds to a third genomic site. FIG. 4B shows atgRNA v2 with the spacer hybridized to the first genomic site and the PBS hybridized to the third genomic site, where the nick is indicated by the inverted triangle. FIG. 4C shows the mechanism by which the atgRNA-v2 as shown in FIGs. 4A-4B can be used for site-specifically integrating an integration recognition site into a genome of a cell. For example, a nick (as indicated by the inverted triangle) on the non-targeting strand induced by a nickase (e.g., nCas9) facilitates binding of the PBS to a sequence 5' to the nick site on the non-targeting strand. Following "3' flap synthesis by a RT," the clamp (i.e., reverse complement of the clamp) binds the targeting strand at a sequence that is 5' to the nick on

the corresponding non-targeting strand. In this case, the sequence to which the clamp (i.e., the reverse complement of the clamp) binds on the targeting strand is greater than 1 nucleotide 5' to the site of the nick on the non-targeting strand. Following favored repair, the integration recognition site (e.g., AttB) is integrated into the genome at the target sequence.

[0039] FIG. 5 shows ddPCR data for beacon placement (i.e., incorporation of integration recognition site into a genome of a cell) at the LMNB1 locus using atgRNAv1 or atgRNAv2.

[0040] FIGs. 6A-6B illustrate the use of an atgRNA comprising an anti-reverse transcription (RT) sequence that is at least partially complementary to all or a portion of the RT template sequence. FIG. 6A shows exonuclease-mediated degradation of the 3' end of an atgRNA with the result being an atgRNA where the 3' end including the PBS, AttB, and clamp has been degraded. FIG. 6B shows an atgRNA comprising the anti-RT sequence that binds (e.g., intramolecular binding) to the RT template, thereby preventing exonuclease-mediated degradation of the 3' end of the atgRNA. "X" through the exonuclease indicates blocking of the exonucleases ability to degrade the atgRNA.

[0041] FIGs. 7A-7C shows percent beacon placement data for atgRNA capable of site-specifically incorporating an integrase recognition site into a genomic locus. The AtgRNA were designed to target the *LMNB1* locus (FIG. 7A), the hF9 locus (FIG. 7B), and the ACTB locus (FIG. 7C) and each capable of site-specifically incorporating an integrase recognition site into the *LMBN1* locus hF9 locus, and ACTB locus, respectively. Each atgRNA included a 10 nucleotide clamp and the indicated separation between the second and third genomic sites. FIG. 7A shows percent beacon placement (i.e., efficiency of integration of the integration recognition site) for atgRNA targeting the *LMNB1* locus having a clamp segment at least partially complementary to a second genomic site and a primer binding site complementary to a third genomic site, where the second genomic site and the third genomic site are separated by 0 to 36 base pairs (see FIG. 7A "clamp+0" (i.e., 0 base pair separation) to "clamp+36" (i.e., 36 base pair separation). FIG. 7B shows percent beacon placement (i.e., efficiency of integration of the integration recognition site) for atgRNA targeting the *hF9* locus having a clamp segment at least partially complementary to a second genomic site and a primer binding site complementary to a third genomic site, where the second genomic site and the third genomic site are separated by 0 to 42 base pairs (see FIG. 7B "clamp+0" (i.e., 0 base pair separation) to "clamp+42" (i.e., 42 base pair separation). FIG. 7C shows percent beacon placement (i.e., efficiency of integration of the integration recognition site) for atgRNA targeting the *ACTB* locus having a clamp segment at least partially complementary to a second genomic site and a primer binding site complementary to a third genomic site, where the second genomic site

and the third genomic site are separated by 0 to 33 base pairs (see FIG. 7C "clamp+0" (i.e., 0 base pair separation) to "clamp+33" (i.e., 33 base pair separation).

6. DETAILED DESCRIPTION

[0042] Described herein are attachment site containing guide RNA (atgRNA) and guide RNA complex compositions capable of binding to a gene editor protein composition and mediating incorporation of an integration recognition site into a target DNA sequence. In some embodiments, the integration recognition stie (also referred to as a beacon) is placed in the genome by partial genomic replacement using a single guide RNA or dual guide RNAs. In typical embodiments, the guide RNA comprises, in 5' to 3' order, comprises a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence; a reverse transcriptase template comprising: an integration recognition site; and a clamp segment comprising a sequence where upon being reverse transcribed the sequence (reverse complement clamp "rcClamp") is at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence; and a primer binding site complementary to a third genomic site of the double-stranded target DNA sequence, wherein the third genomic site is on a second strand of the target DNA sequence.

[0043] Overall, this disclosure is based in part on the finding that by relaxing the requirement that the clamp binding site (i.e., second genomic site) be immediately adjacent to the primer binding site (i.e., third genomic site) the size limitations associated with atgRNAs can be reduced as compared to standard prime editing guide RNAs or gene editing RNAs capable of integration recognition site insertion. This finding enables the clamp segment binding site (i.e., second genomic site) to be selected from within a large genomic window, wherein for example, the genomic window frame can be shifted to screen for variable sequence distance from the third genomic site.

6.1. Terminology

[0044] 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 disclosure belongs. As used herein, the following terms have the meanings ascribed to them below.

[0045] "**Gene editor**" as used herein, is a protein that that can be used to perform gene editing, gene modification, gene insertion, gene deletion, or gene inversion. As used herein, the terms

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

[0046] "**Prime editor fusion protein"** as used herein, describes a protein that is used in prime editing. "Prime editor system" as used herein describes the components used in prime editing. Prime editing uses CRISPR enzyme that nicks or cuts only single strand of double stranded DNA, i.e., a nickase; the nickase can occur either naturally or by mutation or modification of a nuclease that makes double stranded cuts. The nickase is programmed (directed) with a prime-editing guide RNA (pegRNA). The skilled person in the art would appreciate that the pegRNA both specifies the target site and encodes the desired edit. Described herein are attachment site containing guide RNA (atgRNA) that both specifies the target and encodes for the desired integrase target recognition site. The nickase may be programmed (directed) with an atgRNA. Advantageously the nickase is a catalytically-impaired Cas9 endonuclease, a Cas9 nickase, that is fused to the reverse transcriptase. During genetic editing, the Cas9 nickase part of the protein is guided to the DNA target site by the atgRNA (pegRNA), whereby a nick or single stranded cut occurs. The reverse transcriptase domain then uses the atgRNA (pegRNA) to template reverse transcription of the desired edit, directly polymerizing DNA onto the nicked target DNA strand. The edited DNA strand replaces the original DNA strand, creating a heteroduplex containing one edited strand and one unedited strand. Afterward, optionally, the prime editor (PE) guides resolution of the heteroduplex to favor copying the edit onto the unedited strand, completing the process (typically achieved with a nickase gRNA). Other enzymes that can be used to nick or cut only a single strand of double stranded DNA includes a cleavase (e.g., cleavase I enzyme).

[0047] 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 that encodes any of the gene editor polypeptides 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.

[0048] 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 atgRNA or pegRNA that target opposite DNA strands, programming not only the sites that are nicked but also the outcome of the repair (i.e., PrimeDel by *Choi et al.* Nat. Biotechnology, October 14, 2021; *Choi et al.* is incorporated herein by reference and TwinPE by *Anzalone et al.*BioRxiv, November 2, 2021; *Anzalone et al.* is incorporated herein by reference). In some embodiments described herein, the DNA deletion is induced using a single atgRNA. In some embodiments, the DNA deletion and replacement is induced using a wild type Cas9 prime editor (PE-Cas9) system (i.e., PEDAR by *Jiang et al.* Nat. Biotechnology, October 14, 2021; *Jiang et al.* is incorporated herein by reference in its entirety) In some embodiments, the DNA replacement is an integrase target recognition site or recombinase target recognition site. In certain embodiments, the constructs and methods described herein may be utilized to incorporate the pair of pegRNAs used in PrimeDel, TwinPE (WO2021226558 incorporated by reference herein in its entirety), or PEDAR, the prime editor fusion protein or Gene Writer protein, optionally a nickase guide RNA (ngRNA), an integrase, a nucleic acid cargo, and optionally a recombinase into a 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.

[0049] 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 PEI. 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 Casl2a/b. Alternatively, the CRISPR enzyme nickase can be another mutation of Cas9, such as Cas9(Dl0A). A CRISPR enzyme, such as a CRISPR enzyme nickase, such as Cas9 (wild type), Cas9(H840A), Cas9(Dl0A) 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, Vl29P, L139P, Tl97A, 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	ttgagcgggCCCcaccgt
SEQ ID NO: 02	RT_mut_E562Q	cagcgggctCAGctgatagca
SEQ ID NO: 03	RT_mut_D653N	cggatggctAACcaagcggcc

[0050] 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]	**atgactcactatcag**gccttgcttttggacacggaccgggtccagttcggaccggtggtagccctgaacccggct acgctgctcccactgcctgaggaagggctgcaacacaactgccttgatGGGACAGGTGGCGGTG GTGTCACCGTCAAGTTCAAGTACAAGGGTGAGGAACTTGAAGTTGAT ATTAGCAAAATCAAGAAGGTTTGGCGCGTTGGTAAAATGATATCTTT TACTTATGACGACAACGGCAAGACAGGTAGAGGGGCAGTGTCTGAG AAAGACGCCCCCAAGGAGCTGTTGCAAATGTTGGAAAAGTCTGGGA AAAAGtctggcggctcaaaaagaaccgccgacggcagcgaattcgagcccaagaagaagaggaaagtc	4

[0051] PE3, PE3b, PE4, PE5, and/or PEmax, which a skilled person can incorporate into a single nucleic acid construct 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).

[0052] The skilled person can readily incorporate into a gene editor composition described herein a prime editing or CRISPR system. Examples of prime editors or gene 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 disclosure. Further, from the disclosures of the foregoing, the skilled person can incorporate the selected CRISPR enzyme, as part of the gene editor, gene editor fusion, prime editor, or prime editor fusion compositions described herein.

[0053] 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.

[0054] 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.

[0055] 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 in 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 some embodiments, the split construct can be reconstituted via nanobody binding ALFA-tagged proteins. In certain embodiments, the split construct can be adapted into one or more single nucleic acid polynucleotides.

[0056] 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	GGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGCGACGTGGAGGAGAACCCTGGACCT	5	GSGATNFSLLKQAGDVEENPGP	13
B - (GGGS)3	GGGGGAGGAGGTTCTGGAGGCGGAGGCTCCGGAGGCGGAGGGTCA	6	GGGGSGGGGSGGGGS	14
C - GGGGS	GGAGGTGGCGGGAGC	7	GGGGS	15
D - PAPAP	CCCGCACCAGCGCCT	8	PAPAP	16
E -(EAAAK)3	GAGGCAGCTGCCAAGGAAGCCGCTGCCAAGGAGGCGGCCGCAAAG	9	EAAAKEAAAKEAAAK	17
F - XTEN	AGTGGGAGCGAGACCCCTGGGACTAGCGAGTCAGCTACACCCGAAAGC	10	SGSETPGTSESATPES	18
G - (GGS)6	GGGGGGTCAGGTGGATCCGGCGGAAGTGGCGGATCCGGTGGATCTGGCGGCAGT	11	GGSGGSGGSGGSGGSGGS	19
H - EAAAK	GAAGCTGCTGCTAAG	12	EAAAK	20
(GGGGS)4	GGCGGCGGCGGCAGCGGCGGCGGCGGCAGCGGCGGCGGCGGCAGCGGCGGCGGCGGCAGC	543	GGGGSGGGGSGGGGSGGGGS	551
PAS8	GGCGGCGCGAGCCCGGCGGGCGGC	544	GGASPAGG	552
PAS12	GGCGGCGCGAGCCCGGCGGCGCCGGCGCCGGCGGGC	545	GGASPAAPAPAG	553
A(EAAK)4ALEA(EAAAK)4A	GCGGAAGCGGCGAAAGAAGCGGCGAAAGAAGCGGCGAAAGAAGCGGCGAAAGCGCTGGAAGCGGAAGCGGCGGCGAAAGAAGCGGCGGCGAAAGAAGCGGCGGCGAAAGAAGCGGCGGCGAAAGCG	546	AEAAKEAAKEAAKEAAKALEAEAAAKEAAAKEAAAKEAAAKA	554
Camel	GCGCATCATAGCGAAGATCCGGGCGGCGGCGGCAGCGG	547	AHHSEDPGGGGSGGGGSGGGGS	555

Table 3.				
	Sequence (5'-3')	**SEQ ID NO:**	**Amino acid sequence**	**SEQ ID NO:**
	CGGCGGCGGCAGCGGCGGC GGCGGCAGC			
FRF	GGCGGCGGCGGCAGCGAAG CGGCGGCGAAAGGCGGCGG CGGCAGC	548	GGGGSEAAAKGGGG S	556
RFR	GAAGCGGCGGCGAAAGGCG GCGGCGGCAGCGAAGCGGC GGCGAAA	549	EAAAKGGGGSEAAA K	557
Modified XTEN (mXTEN)	AGCGGCGGCAGCAGCGGCG GCAGCAGCGGCAGCGAAAC CCCGGGCACCAGCGAAAGC GCGACCCCGGAAAGCAGCG GCGGCAGCAGCGGCGGCAG CAGCACC	550	SGGSSGGSSGSETPG TSESATPESSGGSSG GSST	558

[0057] 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 in 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

[0058] The skilled person can incorporate a selected CRISPR enzyme, described below, as part of the prime editor fusion, as a component of the gene editor polypeptide 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 a helix referred to as the bridge helix (residues 60–93), the REC1 (residues 94–179 and 308–713) domain, and the REC2 (residues 180–307) domain. The NUC lobe consists of the RuvC (residues 1–59, 718–769, and 909–1098), HNH (residues 775–908), and PAM-interacting (PI) (residues 1099–1368) domains. The negatively charged sgRNA:target DNA heteroduplex is accommodated in a positively charged groove at the interface between the REC and NUC lobes.

In the NUC lobe, the RuvC domain is assembled from the three split RuvC motifs (RuvC I–III) and interfaces with the PI domain to form a positively charged surface that interacts with the 30 tail of the sgRNA. The HNH domain lies between the RuvC II–III motifs and forms only a few contacts with the rest of the protein. Structural aspects of SpCas9 are described by Nishimasu et al., Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA, Cell 156, 935-949, February 27, 2014.

[0059] *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.

[0060] *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 (Δ1099-1368) abolished the cleavage activity, indicating that the PI domain is critical for SpCas9 function and a major determinant for the PAM specificity.

[0061] *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 α42 and α43) and the PI domain/stem loop 3 (β hairpin formed by β3 and β4).

[0062] *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.

[0063] *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 20-hydroxyl groups of G1, C15, U16, and G19 hydrogen bond with Val1009, Tyr450, Arg447/Ile448, and Thr404, respectively.

[0064] 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.

[0065] *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 20-hydroxyl and phosphate groups of sgRNA G1 interact with Val1009 and Gln926, respectively.

[0066] *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.

[0067] 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.

[0068] *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 20-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.

[0069] 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.

[0070] 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.

[0071] 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.

[0072] Cas9 proteins smaller than SpCas9 allow more efficient packaging of nucleic acids encoding CRISPR systems, e.g., Cas9 and sgRNA into one rAAV ("all-in-one-AAV") particle. In addition, efficient packaging of CRISPR systems can be achieved in other viral vector systems (i.e., lentiviral, integration deficient lentiviral, hd-AAV, etc.) and non viral vector systems (i.e., lipid nanoparticle). Small Cas9 proteins can be advantageous for multidomain-Cas-nuclease-based systems for prime editing. Well characterized smaller Cas9 proteins include *Staphylococcus*

aureus (SauCas9, 1053 amino acid residues) and *Campylobacter jejuni* (CjCas9, 984 amino residues). However, both recognize longer PAMs, 5′-NNGRRT-3′ for SauCas9 (R = A or G) and 5′-NNNNRYAC-3′ for CjCas9 (Y = C or T), which reduces the number of uniquely addressable target sites in the genome, in comparison to the NGG SpCas9 PAM. Among smaller Cas9s, Schmidt et al. identified *Staphylococcus lugdunensis* (Slu) Cas9 as having genome-editing activity and provided homology mapping to SpCas9 and SauCas9 to facilitate generation of nickases and inactive ("dead") enzymes (Schmidt et al., 2021, Improved CRISPR genome editing using small highly active and specific engineered RNA-guided nucleases. *Nat Commun* 12, 4219. doi.org/10.1038/s41467-021-24454-5) and engineered nucleases with higher cleavage activity by fragmenting and shuffling Cas9 DNAs. The small Cas9s and nickases are useful in the instant disclosure.

[0073] 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, 21, 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 (SEQ ID NO: 18).

[0074] 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.

[0075] 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_010922251.1	MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFFHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVVDELVKV MGRHKPENIV IEMARENQTT QKGQKNSRER MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVDH IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS KLVSDFRKDF QFYKVREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPTVA YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDLIIKLPK YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA PAAFKYFDTT IDRKRYTSTK EVLDATLIHQ SITGLYETRI DLS	(SEQ ID NO: 21)
AJN60021.1 GI: 757015977 J7RUA5.1 WP_053019794.1 Staphylococcus aureus	MKRNYILGLD IGITSVGYGI IDYETRDVID AGVRLFKEAN VENNEGRRSK RGARRLKRRR RHRIQRVKKL LFDYNLLTDH SELSGINPYE ARVKGLSQKL SEEEFSAALL HLAKRRGVHN VNEVEEDTGN ELSTKEQISR NSKALEEKYV AELQLERLKK DGEVRGSINR FKTSDYVKEA KQLLKVQKAY HQLDQSFIDT YIDLLETRRT YYEGPGEGSP FGWKDIKEWY EMLMGHCTYF PEELRSVKYA YNADLYNALN DLNNLVITRD ENEKLEYYEK FQIIENVFKQ KKKPTLKQIA KEILVNEEDI KGYRVTSTGK PEFTNLKVYH DIKDITARKE IIENAELLDQ IAKILTIYQS SEDIQEELTN LNSELTQEEI EQISNLKGYT GTHNLSLKAI NLILDELWHT NDNQIAIFNR LKLVPKKVDL SQQKEIPTTL VDDFILSPVV KRSFIQSIKV INAIIKKYGL PNDIIIELAR EKNSKDAQKM INEMQKRNRQ TNERIEEIIR TTGKENAKYL IEKIKLHDMQ EGKCLYSLEA IPLEDLLNNP FNYEVDHIIP RSVSFDNSFN NKVLVKQEEN SKKGNRTPFQ YLSSSDSKIS YETFKKHILN LAKGKGRISK TKKEYLLEER DINRFSVQKD FINRNLVDTR YATRGLMNLL RSYFRVNNLD VKVKSINGGF TSFLRRKWKF KKERNKGYKH	(SEQ ID NO: 22)

Table 4- Cas9 orthologs

	HAEDALIIAN ADFIFKEWKK LDKAKKVMEN QMFEEKQAES MPEIETEQEY KEIFITPHQI KHIKDFKDYK YSHRVDKKPN RELINDTLYS TRKDDKGNTL IVNNLNGLYD KDNDKLKKLI NKSPEKLLMY HHDPQTYQKL KLIMEQYGDE KNPLYKYYEE TGNYLTKYSK KDNGPVIKKI KYYGNKLNAH LDITDDYPNS RNKVVKLSLK PYRFDVYLDN GVYKFVTVKN LDVIKKENYY EVNSKCYEEA KKLKKISNQA EFIASFYNND LIKINGELYR VIGVNNDLLN RIEVNMIDIT YREYLENMND KRPPRIIKTI ASKTQSIKKY STDILGNLYE VKSKKHPQII KKG	
AJN60008.1 GI: 757015964 WP_002864485.1 Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819	MARILAFDIG ISSIGWAFSE NDELKDCGVR IFTKVENPKT GESLALPRRL ARSARKRLAR RKARLNHLKH LIANEFKLNY EDYQSFDESL AKAYKGSLIS 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 LKLVTPLMLE GKKYDEACNE LNLKVAINED KKDFLPAFNE TYYKDEVTNP VVLRAIKEYR KVLNALLKKY GKVHKINIEL AREVGKNHSQ RAKIEKEQNE NYKAKKDAEL ECEKLGLKIN SKNILKLRLF KEQKEFCAYS GEKIKISDLQ DEKMLEIDHI YPYSRSFDDS YMNKVLVFTK QNQEKLNQTP FEAFGNDSAK WQKIEVLAKN LPTKKQKRIL DKNYKDKEQK NFKDRNLNDT RYIARLVLNY TKDYLDFLPL SDDENTKLND TQKGSKVHVE AKSGMLTSAL RHTWGFSAKD RNNHLHHAID AVIIAYANNS IVKAFSDFKK EQESNSAELY AKKISELDYK NKRKFFEPFS GFRQKVLDKI DEIFVSKPER KKPSGALHEE TFRKEEEFYQ SYGGKEGVLK ALELGKIRKV NGKIVKNGDM FRVDIFKHKK TNKFYAVPIY TMDFALKVLP NKAVARSKKG EIKDWILMDE NYEFCFSLYK DSLILIQTKD MQEPEFVYYN AFTSSTVSLI VSKHDNKFET LSKNQKILFK NANEKEVIAK SIGIQNLKVF EKYIVSALGE VTKAEFRQRE DFKK	(SEQ ID NO: 23)
Streptococcus thermophilus LMD-9 AJN60026.1 GI: 757015982 WP_011680957.1	MSDLVLGLDI GIGSVGVGIL NKVTGEIIHK NSRIFPAAQA ENNLVRRTNR QGRRLARRKK HRRVRLNRLF EESGLITDFT KISINLNPYQ LRVKGLTDEL SNEELFIALK NMVKHRGISY LDDASDDGNS SVGDYAQIVK ENSKQLETKT PGQIQLERYQ TYGQLRGDFT VEKDGKKHRL INVFPTSAYR SEALRILQTQ QEFNPQITDE FINRYLEILT GKRKYYHGPG NEKSRTDYGR YRTSGETLDN IFGILIGKCT FYPDEFRAAK ASYTAQEFNL LNDLNNLTVP TETKKLSKEQ KNQIINYVKN EKAMGPAKLF KYIAKLLSCD VADIKGYRID KSGKAEIHTF EAYRKMKTLE TLDIEQMDRE TLDKLAYVLT LNTEREGIQE ALEHEFADGS FSQKQVDELV QFRKANSSIF GKGWHNFSVK LMMELIPELY ETSEEQMTIL TRLGKQKTTS SSNKTKYIDE KLLTEEIYNP VVAKSVRQAI KIVNAAIKEY GDFDNIVIEM ARETNEDDEK KAIQKIQKAN KDEKDAAMLK AANQYNGKAE LPHSVFHGHK QLATKIRLWH QQGERCLYTG KTISIHDLIN NSNQFEVDHI LPLSITFDDS LANKVLVYAT ANQEKGQRTP YQALDSMDDA WSFRELKAFV RESKTLSNKK KEYLLTEEDI SKFDVRKKFI ERNLVDTRYA SRVVLNALQE HFRAHKIDTK VSVVRGQFTS QLRRHWGIEK TRDTYHHHAV DALIIAASSQ LNLWKKQKNT LVSYSEDQLL DIETGELISD DEYKESVFKA PYQHFVDTLK SKEFEDSILF SYQVDSKFNR KISDATIYAT RQAKVGKDKA DETYVLGKIK DIYTQDGYDA FMKIYKKDKS KFLMYRHDPQ TFEKVIEPIL ENYPNKQINE KGKEVPCNPF LKYKEEHGYI RKYSKKGNGP EIKSLKYYDS KLGNHIDITP KDSNNKVVLQ SVSPWRADVY FNKTTGKYEI LGLKYADLQF EKGTGTYKIS QEKYNDIKKK EGVDSDSEFK FTLYKNDLLL VKDTETKEQQ LFRFLSRTMP KQKHYVELKP YDKQKFEGGE ALIKVLGNVA NSGQCKKGLG KSNISIYKVR TDVLGNQHII KNEGDKPKLD F	(SEQ ID NO: 24)
Parvibaculum lavamentivorans DS-1 AJN60020.1	MERIFGFDIG TTSIGFSVID YSSTQSAGNI QRLGVRIFPE ARDPDGTPLN QQRRQKRMMR RQLRRRRIRR KALNETLHEA GFLPAYGSAD WPVVMADEPY ELRRRGLEEG LSAYEFGRAI YHLAQHRHFK GRELEESDTP DPDVDDEKEA ANERAATLKA LKNEQTTLGA WLARRPPSDR KRGIHAHRNV VAEEFERLWE VQSKFHPALK SEEMRARISD TIFAQRPVFW RKNTLGECRF MPGEPLCPKG SWLSQQRRML EKLNNLAIAG GNARPLDAEE RDAILSKLQQ QASMSWPGVR	(SEQ ID NO: 25)

Table 4- Cas9 orthologs		
GI: 757015976 WP_011995013.1	SALKALYKQR GEPGAEKSLK FNLELGGESK LLGNALEAKL ADMFGPDWPA HPRKQEIRHA VHERLWAADY GETPDKKRVI ILSEKDRKAH REAAANSFVA DFGITGEQAA QLQALKLPTG WEPYSIPALN LFLAELEKGE RFGALVNGPD WEGWRRTNFP HRNQPTGEIL DKLPSPASKE ERERISQLRN PTVVRTQNEL RKVVNNLIGL YGKPDRIRIE VGRDVGKSKR EREEIQSGIR RNEKQRKKAT EDLIKNGIAN PSRDDVEKWI LWKEGQERCP YTGDQIGFNA LFREGRYEVE HIWPRSRSFD NSPRNKTLCR KDVNIEKGNR MPFEAFGHDE DRWSAIQIRL QGMVSAKGGT GMSPGKVKRF LAKTMPEDFA ARQLNDTRYA AKQILAQLKR LWPDMGPEAP VKVEAVTGQV TAQLRKLWTL NNILADDGEK TRADHRHHAI DALTVACTHP GMTNKLSRYW QLRDDPRAEK PALTPPWDTI RADAEKAVSE IVVSHRVRKK VSGPLHKETT YGDTGTDIKT KSGTYRQFVT RKKIESLSKG ELDEIRDPRI KEIVAAHVAG RGGDPKKAFP PYPCVSPGGP EIRKVRLTSK QQLNLMAQTG NGYADLGSNH HIAIYRLPDG KADFEIVSLF DASRRLAQRN PIVQRTRADG ASFVMSLAAG EAIMIPEGSK KGIWIVQGVW ASGQVVLERD TDADHSTTTR PMPNPILKDD AKKVSIDPIG RVRPSND	
Corynebacterium diphtheriae NCTC 13129 AJN60012.1 GI: 757015968 WP_010933968.1	MKYHVGIDVG TFSVGLAAIE VDDAGMPIKT LSLVSHIHDS GLDPDEIKSA VTRLASSGIA RRTRRLYRRK RRRLQQLDKF IQRQGWPVIE LEDYSDPLYP WKVRAELAAS YIADEKERGE KLSVALRHIA RHRGWRNPYA KVSSLYLPDG PSDAFKAIRE EIKRASGQPV PETATVGQMV TLCELGTLKL RGEGGVLSAR LQQSDYAREI QEICRMQEIG QELYRKIIDV VFAAESPKGS ASSRVGKDPL QPGKNRALKA SDAFQRYRIA ALIGNLRVRV DGEKRILSVE EKNLVFDHLV NLTPKKEPEW VTIAEILGID RGQLIGTATM TDDGERAGAR PPTHDTNRSI VNSRIAPLVD WWKTASALEQ HAMVKALSNA EVDDFDSPEG AKVQAFFADL DDDVHAKLDS LHLPVGRAAY SEDTLVRLTR RMLSDGVDLY TARLQEFGIE PSWTPPTPRI GEPVGNPAVD RVLKTVSRWL ESATKTWGAP ERVIIEHVRE GFVTEKRARE MDGDMRRRAA RNAKLFQEMQ EKLNVQGKPS RADLWRYQSV QRQNCQCAYC GSPITFSNSE MDHIVPRAGQ GSTNTRENLV AVCHRCNQSK GNTPFAIWAK NTSIEGVSVK EAVERTRHWV TDTGMRSTDF KKFTKAVVER FQRATMDEEI DARSMESVAW MANELRSRVA QHFASHGTTV RVYRGSLTAE ARRASGISGK LKFFDGVGKS RLDRRHHAID AAVIAFTSDY VAETLAVRSN LKQSQAHRQE APQWREFTGK DAEHRAAWRV WCQKMEKLSA LLTEDLRDDR VVVMSNVRLR LGNGSAHKET IGKLSKVKLS SQLSVSDIDK ASSEALWCAL TREPGFDPKE GLPANPERHI RVNGTHVYAG DNIGLFPVSA GSIALRGGYA ELGSSFHHAR VYKITSGKKP AFAMLRVYTI DLLPYRNQDL FSVELKPQTM SMRQAEKKLR DALATGNAEY LGWLVVDDEL VVDTSKIATD QVKAVEAELG TIRRWRVDGF FSPSKLRLRP LQMSKEGIKK ESAPELSKII DRPGWLPAVN KLFSDGNVTV VRRDSLGRVR LESTAHLPVT WKVQ	(SEQ ID NO: 26)
Streptococcus pasteurianus WP_013852048.1	MTNGKILGLD IGIASVGVGI IEAKTGKVVH ANSRLFSAAN AENNAERRGF RGSRRLNRRK KHRVKRVRDL FEKYGIVTDF RNLNLNPYEL RVKGLTEQLK NEELFAALRT ISKRRGISYL DDAEDDSTGS TDYAKSIDEN RRLLKNKTPG QIQLERLEKY GQLRGNFTVY DENGEAHRLI NVFSTSDYEK EARKILETQA DYNKKITAEF IDDYVEILTQ KRKYYHGPGN EKSRTDYGRF RTDGTTLENI FGILIGKCNF YPDEYRASKA SYTAQEYNFL NDLNNLKVST ETGKLSTEQK ESLVEFAKNT ATLGPAKLLK EIAKILDCKV DEIKGYREDD KGKPDLHTFE PYRKLKFNLE SINIDDLSRE VIDKLADILT LNTEREGIED AIKRNLPNQF TEEQISEIIK VRKSQSTAFN KGWHSFSAKL MNELIPELYA TSDEQMTILT RLEKFKVNKK SSKNTKTIDE KEVTDEIYNP VVAKSVRQTI KIINAAVKKY GDFDKIVIEM PRDKNADDEK KFIDKRNKEN KKEKDDALKR AAYLYNSSDK LPDEVFHGNK QLETKIRLWY QQGERCLYSG KPISIQELVH NSNNFEIDHI LPLSLSFDDS LANKVLVYAW TNQEKGQKTP YQVIDSMDAA WSFREMKDYV LKQKGLGKKK RDYLLTTENI DKIEVKKKFI ERNLVDTRYA SRVVLNSLQS ALRELGKDTK VSVVRGQFTS QLRRKWKIDK SRETYHHHAV DALIIAASSQ LKLWEKQDNP MFVDYGKNQV VDKQTGEILS VSDDEYKELV FQPPYQGFVN TISSKGFEDE ILFSYQVDSK YNRKVSDATI YSTRKAKIGK DKKEETYVLG KIKDIYSQNG FDTFIKKYNK DKTQFLMYQK DSLTWENVIE VILRDYPTTK KSEDGKNDVK CNPFEEYRRE NGLICKYSKK GKGTPIKSLK YYDKKLGNCI DITPEESRNK VILQSINPWR ADVYFNPETL KYELMGLKYS DLSFEKGTGN YHISQEKYDA IKEKEGIGKK SEFKFTLYRN DLILIKDIAS GEQEIYRFLS	(SEQ ID NO: 27)

Table 4- Cas9 orthologs		
	RTMPNVNHYV ELKPYDKEKF DNVQELVEAL GEADKVGRCI KGLNKPNISI YKVRTDVLGN KYFVKKKGDK PKLDFKNNK K	
Neisseria cinerea ATCC 14685 AJN60019.1 GI: 757015975 WP_003676410.1	MAAFKPNPMN YILGLDIGIA SVGWAIVEID EEENPIRLID LGVRVFERAE VPKTGDSLAA ARRLARSVRR LTRRRAHRLL RARRLLKREG VLQAADFDEN GLIKSLPNTP WQLRAAALDR KLTPLEWSAV LLHLIKHRGY LSQRKNEGET ADKELGALLK GVADNTHALQ TGDFRTPAEL ALNKFEKESG HIRNQRGDYS HTFNRKDLQA ELNLLFEKQK EFGNPHVSDG LKEGIETLLM TQRPALSGDA VQKMLGHCTF EPTEPKAAKN TYTAERFVWL TKLNNLRILE QGSERPLTDT ERATLMDEPY RKSKLTYAQA RKLLDLDDTA FFKGLRYGKD NAEASTLMEM KAYHAISRAL EKEGLKDKKS PLNLSPELQD EIGTAFSLFK TDEDITGRLK DRVQPEILEA LLKHISFDKF VQISLKALRR IVPLMEQGNR YDEACTEIYG DHYGKKNTEE KIYLPPIPAD EIRNPVVLRA LSQARKVING VVRRYGSPAR IHIETAREVG KSFKDRKEIE KRQEENRKDR EKSAAKFREY FPNFVGEPKS KDILKLRLYE QQHGKCLYSG KEINLGRLNE KGYVEIDHAL PFSRTWDDSF NNKVLALGSE NQNKGNQTPY EYFNGKDNSR EWQEFKARVE TSRFPRSKKQ RILLQKFDED GFKERNLNDT RYINRFLCQF VADHMLLTGK GKRRVFASNG QITNLLRGFW GLRKVRAEND RHHALDAVVV ACSTIAMQQK ITRFVRYKEM NAFDGKTIDK ETGEVLHQKA HFPQPWEFFA QEVMIRVFGK PDGKPEFEEA DTPEKLRTLL AEKLSSRPEA VHKYVTPLFI SRAPNRKMSG QGHMETVKSA KRLDEGISVL RVPLTQLKLK DLEKMVNRER EPKLYEALKA RLEAHKDDPA KAFAEPFYKY DKAGNRTQQV KAVRVEQVQK TGVWVHNHNG IADNATIVRV DVFEKGGKYY LVPIYSWQVA KGILPDRAVV QGKDEEDWTV MDDSFEFKFV LYANDLIKLT AKKNEFLGYF VSLNRATGAI DIRTHDTDST KGKNGIFQSV GVKTALSFQK YQIDELGKEI RPCRLKKRPP VR	(SEQ ID NO: 28)
AJN60009.1 GI: 757015965 St1Cas9+SpCas9	MSDLVLGLDI GIGSVGVGIL NKVTGEIIHK NSRIFPAAQA ENNLVRRTNR QGRRLARRKK HRRVRLNRLF EESGLITDFT KISINLNPYQ LRVKGLTDEL SNEELFIALK NMVKHRGISY LDDASDDGNS SVGDYAQIVK ENSKQLETKT PGQIQLERYQ TYGQLRGDFT VEKDGKKHRL INVFPTSAYR SEALRILQTQ QEFNPQITDE FINRYLEILT GKRKYYHGPG NEKSRTDYGR YRTSGETLDN IFGILIGKCT FYPDEFRAAK ASYTAQEFNL LNDLNNLTVP TETKKLSKEQ KNQIINYVKN EKAMGPAKLF KYIAKLLSCD VADIKGYRID KSGKAEIHTF EAYRKMKTLE TLDIEQMDRE TLDKLAYVLT LNTEREGIQE ALEHEFADGS FSQKQVDELV QFRKANSSIF GKGWHNFSVK LMMELIPELY ETSEEQMTIL TRLGKQKTTS SSNKTKYIDE KLLTEEIYNP VVAKSVRQAI KIVNAAIKEY GDFDNIVIEM ARENQTTQKG QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT QRKFDNLTKA ERGGLSELDK AGFIKRQLVE TRQITKHVAQ ILDSRMNTKY DENDKLIREV KVITLKSKLV SDFRKDFQFY KVREINNYHH AHDAYLNAVV GTALIKKYPK LESEFVYGDY KVYDVRKMIA KSEQEIGKAT AKYFFYSNIM NFFKTEITLA NGEIRKRPLI ETNGETGEIV WDKGRDFATV RKVLSMPQVN IVKKTEVQTG GFSKESILPK RNSDKLIARK KDWDPKKYGG FDSPTVAYSV LVVAKVEKGK SKKLKSVKEL LGITIMERSS FEKNPIDFLE AKGYKEVKKD LIIKLPKYSL FELENGRKRM LASAGELQKG NELALPSKYV NFLYLASHYE KLKGSPEDNE QKQLFVEQHK HYLDEIIEQI SEFSKRVILA DANLDKVLSA YNKHRDKPIR EQAENIIHLF TLTNLGAPAA FKYFDTTIDR KRYTSTKEVL DATLIHQSIT GLYETRIDLS QLGGD	(SEQ ID NO: 29)
Campylobacter lari Cas9 BAK69486.1	MRILGFDIGI NSIGWAFVEN DELKDCGVRI FTKAENPKNK ESLALPRRNA RSSRRRLKRR KARLIAIKRI LAKELKLNYK DYVAADGELP KAYEGSLASV YELRYKALTQ NLETKDLARV ILHIAKHRGY MNKNEKKSND AKKGKILSAL KNNALKLENY QSVGEYFYKE FFQKYKKNTK NFIKIRNTKD NYNNCVLSSD LEKELKLILE KQKEFGYNYS EDFINEILKV AFFQRPLKDF SHLVGACTFF EEEKRACKNS YSAWEFVALT KIINEIKSLE KISGEIVPTQ TINEVLNLIL DKGSITYKKF RSCINLHESI SFKSLKYDKE NAENAKLIDF RKLVEFKKAL GVHSLSRQEL DQISTHITLI KDNVKLKTVL EKYNLSNEQI NNLLEIEFND YINLSFKALG MILPLMREGK RYDEACEIAN LKPKTVDEKK DFLPAFCDSI FAHELSNPVV NRAISEYRKV LNALLKKYGK VHKIHLELAR DVGLSKKARE	(SEQ ID NO: 30)

Table 4- Cas9 orthologs		
	KIEKEQKENQ AVNAWALKEC ENIGLKASAK NILKLKLWKE QKEICIYSGN KISIEHLKDE KALEVDHIYP YSRSFDDSFI NKVLVFTKEN QEKLNKTPFE AFGKNIEKWS KIQTLAQNLP YKKKNKILDE NFKDKQQEDF ISRNLNDTRY IATLIAKYTK EYLNFLLLSE NENANLKSGE KGSKIHVQTI SGMLTSVLRH TWGFDKKDRN NHLHHALDAI IVAYSTNSII KAFSDFRKNQ ELLKARFYAK ELTSDNYKHQ VKFFEPFKSF REKILSKIDE IFVSKPPRKR ARRALHKDTF HSENKIIDKC SYNSKEGLQI ALSCGRVRKI GTKYVENDTI VRVDIFKKQN KFYAIPIYAM DFALGILPNK IVITGKDKNN NPKQWQTIDE SYEFCFSLYK NDLILLQKKN MQEPEFAYYN DFSISTSSIC VEKHDNKFEN LTSNQKLLFS NAKEGSVKVE SLGIQNLKVF EKYIITPLGD KIKADFQPRE NISLKTSKKY GLR	
AJN60010.1 GI: 757015966 SpCas9+St1Cas9	MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFFHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVVDELVKV MGRHKPENIV IEMARETNED DEKKAIQKIQ KANKDEKDAA MLKAANQYNG KAELPHSVFH GHKQLATKIR LWHQQGERCL YTGKTISIHD LINNSNQFEV DHILPLSITF DDSLANKVLV YATANQEKGQ RTPYQALDSM DDAWSFRELK AFVRESKTLS NKKKEYLLTE EDISKFDVRK KFIERNLVDT RYASRVVLNA LQEHFRAHKI DTKVSVVRGQ FTSQLRRHWG IEKTRDTYHH HAVDALIIAA SSQLNLWKKQ KNTLVSYSED QLLDIETGEL ISDDEYKESV FKAPYQHFVD TLKSKEFEDS ILFSYQVDSK FNRKISDATI YATRQAKVGK DKADETYVLG KIKDIYTQDG YDAFMKIYKK DKSKFLMYRH DPQTFEKVIE PILENYPNKQ INEKGKEVPC NPFLKYKEEH GYIRKYSKKG NGPEIKSLKY YDSKLGNHID ITPKDSNNKV VLQSVSPWRA DVYFNKTTGK YEILGLKYAD LQFEKGTGTY KISQEKYNDI KKKEGVDSDS EFKFTLYKND LLLVKDTETK EQQLFRFLSR TMPKQKHYVE LKPYDKQKFE GGEALIKVLG NVANSGQCKK GLGKSNISIY KVRTDVLGNQ HIIKNEGDKP KLDF	(SEQ ID NO: 31)
SpCas9 inactive AJN60011.1 GI: 757015967	MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFFHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL HEHIANLAGS PAIKKGILQT VKVVDELVKV MGRHKPENIV IAMARENQTT QKGQKNSRER MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVDA IVPQSFLKDD SIDAKVLTRS DKARGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI	(SEQ ID NO: 32)

Table 4- Cas9 orthologs		
	REVKVITLKS KLVSDFRKDF QFYKVREINN YHHAHAAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPTVA YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDLIIKLPK YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA PAAFKYFDTT IDRKRYTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD	
AJN60013.1 GI: 757015969 WP_005430658.1 Sutterella wadsworthensis 3_1_45B	MTQSERRFSC SIGIDMGAKY TGVFYALFDR EELPTNLNSK AMTLVMPETG PRYVQAQRTA VRHRLRGQKR YTLARKLAFL VVDDMIKKQE KRLTDEEWKR GREALSGLLK RRGYSRPNAD GEDLTPLENV RADVFAAHPA FSTYFSEVRS LAEQWEEFTA NISNVEKFLG DPNIPADKEF IEFAVAEGLI DKTEKKAYQS ALSTLRANAN VLTGLRQMGH KPRSEYFKAI EADLKKDSRL AKINEAFGGA ERLARLLGNL SNLQLRAERW YFNAPDIMKD RGWEPDRFKK TLVRAFKFFH PAKDQNKQHL ELIKQIENSE DIIETLCTLD PNRTIPPYED QNNRRPPLDQ TLLLSPEKLT RQYGEIWKTW SARLTSAEPT LAPAAEILER STDRKSRVAV NGHEPLPTLA YQLSYALQRA FDRSKALDPY ALRALAAGSK SNKLTSARTA LENCIGGQNV KTFLDCARRY YREADDAKVG LWFDNADGLL ERSDLHPPMK KKILPLLVAN ILQTDETTGQ KFLDEIWRKQ IKGRETVASR CARIETVRKS FGGGFNIAYN TAQYREVNKL PRNAQDKELL TIRDRVAETA DFIAANLGLS DEQKRKFANP FSLAQFYTLI ETEVSGFSAT TLAVHLENAW RMTIKDAVIN GETVRAAQCS RLPAETARPF DGLVRRLVDR QAWEIAKRVS TDIQSKVDFS NGIVDVSIFV EENKFEFSAS VADLKKNKRV KDKMLSEAEK LETRWLIKNE RIKKASRGTC PYTGDRLAEG GEIDHILPRS LIKDARGIVF NAEPNLIYAS SRGNQLKKNQ RYSLSDLKAN YRNEIFKTSN IAAITAEIED VVTKLQQTHR LKFFDLLNEH EQDCVRHALF LDDGSEARDA VLELLATQRR TRVNGTQIWM IKNLANKIRE ELQNWCKTTN NRLHFQAAAT NVSDAKNLRL KLAQNQPDFE KPDIQPIASH SIDALCSFAV GSADAERDQN GFDYLDGKTV LGLYPQSCEV IHLQAKPQEE KSHFDSVAIF KEGIYAEQFL PIFTLNEKIW IGYETLNAKG ERCGAIEVSG KQPKELLEML APFFNKPVGD LSAHATYRIL KKPAYEFLAK AALQPLSAEE KRLAALLDAL RYCTSRKSLM SLFMAANGKS LKKREDVLKP KLFQLKVELK GEKSFKLNGS LTLPVKQDWL RICDSPELAD AFGKPCSADE LTSKLARIWK RPVMRDLAHA PVRREFSLPA IDNPSGGFRI RRTNLFGNEL YQVHAINAKK YRGFASAGSN VDWSKGILFN ELQHENLTEC GGRFITSADV TPMSEWRKVV AEDNLSIWIA PGTEGRRYVR VETTFIQASH WFEQSVENWA ITSPLSLPAS FKVDKPAEFQ KAVGTELSEL LGQPRSEIFI ENVGNAKHIR FWYIVVSSNK KMNESYNNVS KS	(SEQ ID NO: 33)
AJN60014.1 GI: 757015970 WP_011212792.1 Legionella pneumophila str. Paris	MESSQILSPI GIDLGGKFTG VCLSHLEAFA ELPNHANTKY SVILIDHNNF QLSQAQRRAT RHRVRNKKRN QFVKRVALQL FQHILSRDLN AKEETALCHY LNNRGYTYVD TDLDEYIKDE TTINLLKELL PSESEHNFID WFLQKMQSSE FRKILVSKVE EKKDDKELKN AVKNIKNFIT GFEKNSVEGH RHRKVYFENI KSDITKDNQL DSIKKKIPSV CLSNLLGHLS NLQWKNLHRY LAKNPKQFDE QTFGNEFLRM LKNFRHLKGS QESLAVRNLI QQLEQSQDYI SILEKTPPEI TIPPYEARTN TGMEKDQSLL LNPEKLNNLY PNWRNLIPGI IDAHPFLEKD LEHTKLRDRK RIISPSKQDE KRDSYILQRY LDLNKKIDKF KIKKQLSFLG QGKQLPANLI ETQKEMETHF NSSLVSVLIQ IASAYNKERE DAAQGIWFDN AFSLCELSNI NPPRKQKILP LLVGAILSED FINNKDKWAK FKIFWNTHKI GRTSLKSKCK EIEEARKNSG NAFKIDYEEA LNHPEHSNNK ALIKIIQTIP DIIQAIQSHL GHNDSQALIY HNPFSLSQLY TILETKRDGF HKNCVAVTCE NYWRSQKTEI DPEISYASRL PADSVRPFDG VLARMMQRLA YEIAMAKWEQ IKHIPDNSSL LIPIYLEQNR FEFEESFKKI KGSSSDKTLE QAIEKQNIQW EEKFQRIINA SMNICPYKGA SIGGQGEIDH IYPRSLSKKH FGVIFNSEVN LIYCSSQGNR EKKEEHYLLE HLSPLYLKHQ FGTDNVSDIK NFISQNVANI KKYISFHLLT PEQQKAARHA LFLDYDDEAF KTITKFLMSQ QKARVNGTQK FLGKQIMEFL STLADSKQLQ LEFSIKQITA EEVHDHRELL SKQEPKLVKS RQQSFPSHAI DATLTMSIGL KEFPQFSQEL DNSWFINHLM PDEVHLNPVR SKEKYNKPNI SSTPLFKDSL YAERFIPVWV KGETFAIGFS EKDLFEIKPS	(SEQ ID NO: 34)

Table 4- Cas9 orthologs		
	NKEKLFTLLK TYSTKNPGES LQELQAKSKA KWLYFPINKT LALEFLHHYF HKEIVTPDDT TVCHFINSLR YYTKKESITV KILKEPMPVL SVKFESSKKN VLGSFKHTIA LPATKDWERL FNHPNFLALK ANPAPNPKEF NEFIRKYFLS DNNPNSDIPN NGHNIKPQKH KAVRKVFSLP VIPGNAGTMM RIRRKDNKGQ PLYQLQTIDD TPSMGIQINE DRLVKQEVLM DAYKTRNLST IDGINNSEGQ AYATFDNWLT LPVSTFKPEI IKLEMKPHSK TRRYIRITQS LADFIKTIDE ALMIKPSDSI DDPLNMPNEI VCKNKLFGNE LKPRDGKMKI VSTGKIVTYE FESDSTPQWI QTLYVTQLKK QP	
AJN60015.1 GI: 757015971 WP_002681289.1 Treponema denticola ATCC 35405	MKKEIKDYFL GLDVGTGSVG WAVTDTDYKL LKANRKDLWG MRCFETAETA EVRRLHRGAR RRIERRKKRI KLLQELFSQE IAKTDEGFFQ RMKESPFYAE DKTILQENTL FNDKDFADKT YHKAYPTINH LIKAWIENKV KPDPRLLYLA CHNIIKKRGH FLFEGDFDSE NQFDTSIQAL FEYLREDMEV DIDADSQKVK EILKDSSLKN SEKQSRLNKI LGLKPSDKQK KAITNLISGN KINFADLYDN PDLKDAEKNS ISFSKDDFDA LSDDLASILG DSFELLLKAK AVYNCSVLSK VIGDEQYLSF AKVKIYEKHK TDLTKLKNVI KKHFPKDYKK VFGYNKNEKN NNNYSGYVGV CKTKSKKLII NNSVNQEDFY KFLKTILSAK SEIKEVNDIL TEIETGTFLP KQISKSNAEI PYQLRKMELE KILSNAEKHF SFLKQKDEKG LSHSEKIIML LTFKIPYYIG PINDNHKKFF PDRCWVVKKE KSPSGKTTPW NFFDHIDKEK TAEAFITSRT NFCTYLVGES VLPKSSLLYS EYTVLNEINN LQIIIDGKNI CDIKLKQKIY EDLFKKYKKI TQKQISTFIK HEGICNKTDE VIILGIDKEC TSSLKSYIEL KNIFGKQVDE ISTKNMLEEI IRWATIYDEG EGKTILKTKI KAEYGKYCSD EQIKKILNLK FSGWGRLSRK FLETVTSEMP GFSEPVNIIT AMRETQNNLM ELLSSEFTFT ENIKKINSGF EDAEKQFSYD GLVKPLFLSP SVKKMLWQTL KLVKEISHIT QAPPKKIFIE MAKGAELEPA RTKTRLKILQ DLYNNCKNDA DAFSSEIKDL SGKIENEDNL RLRSDKLYLY YTQLGKCMYC GKPIEIGHVF DTSNYDIDHI YPQSKIKDDS ISNRVLVCSS CNKNKEDKYP LKSEIQSKQR GFWNFLQRNN FISLEKLNRL TRATPISDDE TAKFIARQLV ETRQATKVAA KVLEKMFPET KIVYSKAETV SMFRNKFDIV KCREINDFHH AHDAYLNIVV GNVYNTKFTN NPWNFIKEKR DNPKIADTYN YYKVFDYDVK RNNITAWEKG KTIITVKDML KRNTPIYTRQ AACKKGELFN QTIMKKGLGQ HPLKKEGPFS NISKYGGYNK VSAAYYTLIE YEEKGNKIRS LETIPLYLVK DIQKDQDVLK SYLTDLLGKK EFKILVPKIK INSLLKINGF PCHITGKTND SFLLRPAVQF CCSNNEVLYF KKIIRFSEIR SQREKIGKTI SPYEDLSFRS YIKENLWKKT KNDEIGEKEF YDLLQKKNLE IYDMLLTKHK DTIYKKRPNS ATIDILVKGK EKFKSLIIEN QFEVILEILK LFSATRNVSD LQHIGGSKYS GVAKIGNKIS SLDNCILIYQ SITGIFEKRI DLLKV	(SEQ ID NO: 35)
AJN60016.1 GI: 757015972 EFE28295.1 Filifactor alocis ATCC 35896	MTKEYYLGLD VGTNSVGWAV TDSQYNLCKF KKKDMWGIRL FESANTAKDR RLQRGNRRRL ERKKQRIDLL QEIFSPEICK IDPTFFIRLN ESRLHLEDKS NDFKYPLFIE KDYSDIEYYK EFPTIFHLRK HLIESEEKQD IRLIYLALHN IIKTRGHFLI DGDLQSAKQL RPILDTFLLS LQEEQNLSVS LSENQKDEYE EILKNRSIAK SEKVKKLKNL FEISDELEKE EKKAQSAVIE NFCKFIVGNK GDVCKFLRVS KEELEIDSFS FSEGKYEDDI VKNLEEKVPE KVYLFEQMKA MYDWNILVDI LETEEYISFA KVKQYEKHKT NLRLLRDIIL KYCTKDEYNR MFNDEKEAGS YTAYVGKLKK NNKKYWIEKK RNPEEFYKSL GKLLDKIEPL KEDLEVLTMM IEECKNHTLL PIQKNKDNGV IPHQVHEVEL KKILENAKKY YSFLTETDKD GYSVVQKIES IFRFRIPYYV GPLSTRHQEK GSNVWMVRKP GREDRIYPWN MEEIIDFEKS NENFITRMTN KCTYLIGEDV LPKHSLLYSK YMVLNELNNV KVRGKKLPTS LKQKVFEDLF ENKSKVTGKN LLEYLQIQDK DIQIDDLSGF DKDFKTSLKS YLDFKKQIFG EEIEKESIQN MIEDIIKWIT IYGNDKEMLK RVIRANYSNQ LTEEQMKKIT GFQYSGWGNF SKMFLKGISG SDVSTGETFD IITAMWETDN NLMQILSKKF TFMDNVEDFN SGKVGKIDKI TYDSTVKEMF LSPENKRAVW QTIQVAEEIK KVMGCEPKKI FIEMARGGEK VKKRTKSRKA QLLELYAACE EDCRELIKEI EDRDERDFNS MKLFLYYTQF GKCMYSGDDI DINELIRGNS KWDRDHIYPQ SKIKDDSIDN LVLVNKTYNA KKSNELLSED IQKKMHSFWL SLLNKKLITK SKYDRLTRKG DFTDEELSGF IARQLVETRQ STKAIADIFK QIYSSEVVYV KSSLVSDFRK KPLNYLKSRR VNDYHHAKDA YLNIVVGNVY NKKFTSNPIQ WMKKNRDTNY SLNKVFEHDV VINGEVIWEK CTYHEDTNTY DGGTLDRIRK IVERDNILYT EYAYCEKGEL	(SEQ ID NO: 36)

Table 4- Cas9 orthologs		
	FNATIQNKNG NSTVSLKKGL DVKKYGGYFS ANTSYFSLIE FEDKKGDRAR HIIGVPIYIA NMLEHSPSAF LEYCEQKGYQ NVRILVEKIK KNSLLIINGY PLRIRGENEV DTSFKRAIQL KLDQKNYELV RNIEKFLEKY VEKKGNYPID ENRDHITHEK MNQLYEVLLS KMKKFNKKGM ADPSDRIEKS KPKFIKLEDL IDKINVINKM LNLLRCDNDT KADLSLIELP KNAGSFVVKK NTIGKSKIIL VNQSVTGLYE NRREL	
AJN60017.1 GI: 757015973 WP_014613259.1 Staphylococcus pseudintermedius ED99	MGRKPYILSL DIGTGSVGYA CMDKGFNVLK YHDKDALGVY LFDGALTAQE RRQFRTSRRR KNRRIKRLGL LQELLAPLVQ NPNFYQFQRQ FAWKNDNMDF KNKSLSEVLS FLGYESKKYP TIYHLQEALL LKDEKFDPEL IYMALYHLVK YRGHFLFDHL KIENLTNNDN MHDFVELIET YENLNNIKLN LDYEKTKVIY EILKDNEMTK NDRAKRVKNM EKKLEQFSIM LLGLKFNEGK LFNHADNAEE LKGANQSHTF ADNYEENLTP FLTVEQSEFI ERANKIYLSL TLQDILKGKK SMAMSKVAAY DKFRNELKQV KDIVYKADST RTQFKKIFVS SKKSLKQYDA TPNDQTFSSL CLFDQYLIRP KKQYSLLIKE LKKIIPQDSE LYFEAENDTL LKVLNTTDNA SIPMQINLYE AETILRNQQK YHAEITDEMI EKVLSLIQFR IPYYVGPLVN DHTASKFGWM ERKSNESIKP WNFDEVVDRS KSATQFIRRM TNKCSYLINE DVLPKNSLLY QEMEVLNELN ATQIRLQTDP KNRKYRMMPQ IKLFAVEHIF KKYKTVSHSK FLEIMLNSNH RENFMNHGEK LSIFGTQDDK KFASKLSSYQ DMTKIFGDIE GKRAQIEEII QWITIFEDKK ILVQKLKECY PELTSKQINQ LKKLNYSGWG RLSEKLLTHA YQGHSIIELL RHSDENFMEI LTNDVYGFQN FIKEENQVQS NKIQHQDIAN LTTSPALKKG IWSTIKLVRE LTSIFGEPEK IIMEFATEDQ QKGKKQKSRK QLWDDNIKKN KLKSVDEYKY IIDVANKLNN EQLQQEKLWL YLSQNGKCMY SGQSIDLDAL LSPNATKHYE VDHIFPRSFI KDDSIDNKVL VIKKMNQTKG DQVPLQFIQQ PYERIAYWKS LNKAGLISDS KLHKLMKPEF TAMDKEGFIQ RQLVETRQIS VHVRDFLKEE YPNTKVIPMK AKMVSEFRKK FDIPKIRQMN DAHHAIDAYL NGVVYHGAQL AYPNVDLFDF NFKWEKVREK WKALGEFNTK QKSRELFFFK KLEKMEVSQG ERLISKIKLD MNHFKINYSR KLANIPQQFY NQTAVSPKTA ELKYESNKSN EVVYKGLTPY QTYVVAIKSV NKKGKEKMEY QMIDHYVFDF YKFQNGNEKE LALYLAQREN KDEVLDAQIV YSLNKGDLLY INNHPCYFVS RKEVINAKQF ELTVEQQLSL YNVMNNKETN VEKLLIEYDF IAEKVINEYH HYLNSKLKEK RVRTFFSESN QTHEDFIKAL DELFKVVTAS ATRSDKIGSR KNSMTHRAFL GKGKDVKIAY TSISGLKTTK PKSLFKLAES RNEL	(SEQ ID NO: 37)
AJN60018.1 GI: 757015974 WP_014567561.1 Lactobacillus johnsonii DPC 6026	MTKIKDDYIV GLDIGTDSCG WVAMNSNNDI LKLQGKTAIG SRLFEGGKSA AERRLFRTTH RRIKRRRWRL KLLEEFFDPY MAEVDPYFFA RLKESGLSPL DKRKTVSSIV FPTSAEDKKF YDDYPTIYHL RYKLMTEDEK FDLREVYLAI HHIIKYRGNF LYNTSVKDFK ASKIDVKSSI EKLNELYENL GLDLNVEFNI SNTAEIEKVL KDKQIFKRDK VKKIAELFAI KTDNKEQSKR IKDISKQVAN AVLGYKTRFD TIALKEISKD ELSDWNFKLS DIDADSKFEA LMGNLDENEQ AILLTIKELF NEVTLNGIVE DGNTLSESMI NKYNDHRDDL KLLKEVIENH IDRKKAKELA LAYDLYVNNR HGQLLQAKKK LGKIKPRSKE DFYKVVNKNL DDSRASKEIK KKIELDSFMP KQRTNANGVI PYQLQQLELD KIIENQSKYY PFLKEINPVS SHLKEAPYKL DELIRFRVPY YVGPLISPNE STKDIQTKKN QNFAWMIRKE EGRITPWNFD QKVDRIESAN KFIKRMTTKD TYLFGEDVLP ANSLLYQKFT VLNELNNIRI NGKRISVDLK QEIYENLFKK HTTVTVKKLE NYLKENHNLV KVEIKGLADE KKFNSGLTTY NRFKNLNIFD NQIDDLKYRN DFEKIIEWST IFEDKSIYKE KLRSIDWLNE KQINALSNIR LQGWGRLSKK LLAQLHDHNG QTIIEQLWDS QNNFMQIVTQ ADFKDAIAKA NQNLLVATSV EDILNNAYTS PANKKAIRQV IKVVDDIVKA ASGKVPKQIA IEFTRDADEN PKRSQTRGSK LQKVYKDLST ELASKTIAEE LNEAIKDKKL VQDKYYLYFM QLGRDAYTGE PINIDEIQKY DIDHILPQSF IKDDALDNRV LVSRAVNNGK SDNVPVKLFG NEMAANLGMT IRKMWEEWKN IGLISKTKYN NLLTDPDHIN KYKSAGFIRR QLVETSQIIK LVSTILQSRY PNTEIITVKA KYNHYLREKF DLYKSREVND YHHAIDAYLS AICGNLLYQN YPNLRPFFVY GQYKKFSSDP DKEKAIFNKT RKFSFISQLL KNKSENSKEI AKKLKRAYQF KYMLVSRETE TRDQEMFKMT VYPRFSHDTV KAPRNLIPKK MGMSPDIYGG YTNNSDAYMV IVRIDKKKGT EYKILGIPTR ELVNLKKAEK EDHYKSYLKE ILTPRILYNK NGKRDKKITS FEIVKSKIPY KQVIQDGDKK FMLGSSTYVY NAKQLTLSTE	(SEQ ID NO: 38)

Table 4- Cas9 orthologs		
	SMKAITNNFD KDSDENDALI KAYDEILDKV DKYLPLFDIN KFREKLHSGR EKFIKLSLED KKDTILKVLE GLHDNAVMTK IPTIGLSTPL GFMQFPNGVI LSENAKLIYQ SPTGLFKKSV KISDL	
Mycoplasma gallisepticum str. F AJN60022.1 GI: 757015978 WP_014574789.1	MNNSIKSKPE VTIGLDLGVG SVGWAIVDNE TNIIHHLGSR LFSQAKTAED RRSFRGVRRL IRRRKYKLKR FVNLIWKYNS YFGFKNKEDI LNNYQEQQKL HNTVLNLKSE ALNAKIDPKA LSWILHDYLK NRGHFYEDNR DFNVYPTKEL AKYFDKYGYY KGIIDSKEDN DNKLEEELTK YKFSNKHWLE EVKKVLSNQT GLPEKFKEEY ESLFSYVRNY SEGPGSINSV SPYGIYHLDE KEGKVVQKYN NIWDKTIGKC NIFPDEYRAP KNSPIAMIFN EINELSTIRS YSIYLTGWFI NQEFKKAYLN KLLDLLIKTN GEKPIDARQF KKLREETIAE SIGKETLKDV ENEEKLEKED HKWKLKGLKL NTNGKIQYND LSSLAKFVHK LKQHLKLDFL LEDQYATLDK INFLQSLFVY LGKHLRYSNR VDSANLKEFS DSNKLFERIL QKQKDGLFKL FEQTDKDDEK ILAQTHSLST KAMLLAITRM TNLDNDEDNQ KNNDKGWNFE AIKNFDQKFI DITKKNNNLS LKQNKRYLDD RFINDAILSP GVKRILREAT KVFNAILKQF SEEYDVTKVV IELARELSEE KELENTKNYK KLIKKNGDKI SEGLKALGIS EDEIKDILKS PTKSYKFLLW LQQDHIDPYS LKEIAFDDIF TKTEKFEIDH IIPYSISFDD SSSNKLLVLA ESNQAKSNQT PYEFISSGNA GIKWEDYEAY CRKFKDGDSS LLDSTQRSKK FAKMMKTDTS SKYDIGFLAR NLNDTRYATI VFRDALEDYA NNHLVEDKPM FKVVCINGSV TSFLRKNFDD SSYAKKDRDK NIHHAVDASI ISIFSNETKT LFNQLTQFAD YKLFKNTDGS WKKIDPKTGV VTEVTDENWK QIRVRNQVSE IAKVIEKYIQ DSNIERKARY SRKIENKTNI SLFNDTVYSA KKVGYEDQIK RKNLKTLDIH ESAKENKNSK VKRQFVYRKL VNVSLLNNDK LADLFAEKED ILMYRANPWV INLAEQIFNE YTENKKIKSQ NVFEKYMLDL TKEFPEKFSE FLVKSMLRNK TAIIYDDKKN IVHRIKRLKM LSSELKENKL SNVIIRSKNQ SGTKLSYQDT INSLALMIMR SIDPTAKKQY IRVPLNTLNL HLGDHDFDLH NMDAYLKKPK FVKYLKANEI GDEYKPWRVL TSGTLLIHKK DKKLMYISSF QNLNDVIEIK NLIETEYKEN DDSDSKKKKK ANRFLMTLST ILNDYILLDA KDNFDILGLS KNRIDEILNS KLGLDKIVK	(SEQ ID NO: 39)
AJN60023.1 GI: 757015979	MRILGFDIGI NSIGWAFVEN DELKDCGVRI FTKAENPKNK ESLALPRRNA RSSRRRLKRR KARLIAIKRI LAKELKLNYK DYVAADGELP KAYEGSLASV YELRYKALTQ NLETKDLARV ILHIAKHRGY MNKNEKKSND AKKGKILSAL KNNALKLENY QSVGEYFYKE FFQKYKKNTK NFIKIRNTKD NYNNCVLSSD LEKELKLILE KQKEFGYNYS EDFINEILKV AFFQRPLKDF SHLVGACTFF EEEKRACKNS YSAWEFVALT KIINEIKSLE KISGEIVPTQ TINEVLNLIL DKGSITYKKF RSCINLHESI SFKSLKYDKE NAENAKLIDF RKLVEFKKAL GVHSLSRQEL DQISTHITLI KDNVKLKTVL EKYNLSNEQI NNLLEIEFND YINLSFKALG MILPLMREGK RYDEACEIAN LKPKTVDEKK DFLPAFCDSI FAHELSNPVV NRAISEYRKV LNALLKKYGK VHKIHLELAR DVGLSKKARE KIEKEQKENQ AVNAWALKEC ENIGLKASAK NILKLKLWKE QKEICIYSGN KISIEHLKDE KALEVDHIYP YSRSFDDSFI NKVLVFTKEN QEKLNKTPFE AFGKNIEKWS KIQTLAQNLP YKKKNKILDE NFKDKQQEDF ISRNLNDTRY IATLIAKYTK EYLNFLLLSE NENANLKSGE KGSKIHVQTI SGMLTSVLRH TWGFDKKDRN NHLHHALDAI IVAYSTNSII KAFSDFRKNQ ELLKARFYAK ELTSDNYKHQ VKFFEPFKSF REKILSKIDE IFVSKPPRKR ARRALHKDTF HSENKIIDKC SYNSKEGLQI ALSCGRVRKI GTKYVENDTI VRVDIFKKQN KFYAIPIYAM DFALGILPNK IVITGKDKNN NPKQWQTIDE SYEFCFSLYK NDLILLQKKN MQEPEFAYYN DFSISTSSIC VEKHDNKFEN LTSNQKLLFS NAKEGSVKVE SLGIQNLKVF EKYIITPLGD KIKADFQPRE NISLKTSKKY GLR	(SEQ ID NO: 30)
AJN60025.1 GI: 757015981	MSDLVLGLDI GIGSVGVGIL NKVTGEIIHK NSRIFPAAQA ENNLVRRTNR QGRRLARRKK HRRVRLNRLF EESGLITDFT KISINLNPYQ LRVKGLTDEL SNEELFIALK NMVKHRGISY LDDASDDGNS SVGDYAQIVK ENSKQLETKT PGQIQLERYQ TYGQLRGDFT VEKDGKKHRL INVFPTSAYR SEALRILQTQ QEFNPQITDE FINRYLEILT GKRKYYHGPG NEKSRTDYGR YRTSGETLDN IFGILIGKCT FYPDEFRAAK ASYTAQEFNL LNDLNNLTVP TETKKLSKEQ KNQIINYVKN EKAMGPAKLF KYIAKLLSCD VADIKGYRID KSGKAEIHTF	(SEQ ID NO: 41)

Table 4- Cas9 orthologs

	EAYRKMKTLE TLDIEQMDRE TLDKLAYVLT LNTEREGIQE ALEHEFADGS FSQKQVDELV QFRKANSSIF GKGWHNFSVK LMMELIPELY ETSEEQMTIL TRLGKQKTTS SSNKTKYIDE KLLTEEIYNP VVAKSVRQAI KIVNAAIKEY GDFDNIVIEM ARETNEDDEK KAIQKIQKAN KDEKDAAMLK AANQYNGKAE LPHSVFHGHK QLATKIRLWH QQGERCLYTG KTISIHDLIN NSNQFEVDHI LPLSITFDDS LANKVLVYAT ANQEKGQRTP YQALDSMDDA WSFRELKAFV RESKTLSNKK KEYLLTEEDI SKFDVRKKFI ERNLVDTRYA SRVVLNALQE HFRAHKIDTK VSVVRGQFTS QLRRHWGIEK TRDTYHHHAV DALIIAASSQ LNLWKKQKNT LVSYSEDQLL DIETGELISD DEYKESVFKA PYQHFVDTLK SKEFEDSILF SYQVDSKFNR KISDATIYAT RQAKVGKDKA DETYVLGKIK DIYTQDGYDA FMKIYKKDKS KFLMYRHDPQ TFEKVIEPIL ENYPNKQINE KGKEVPCNPF LKYKEEHGYI RKYSKKGNGP EIKSLKYYDS KLGNHIDITP KDSNNKVVLQ SVSPWRADVY FNKTTGKYEI LGLKYADLQF EKGTGTYKIS QEKYNDIKKK EGVDSDSEFK FTLYKNDLLL VKDTETKEQQ LFRFLSRTMP KQKHYVELKP YDKQKFEGGE ALIKVLGNVA NSGQCKKGLG KSNISIYKVR TDVLGNQHII KNEGDKPKLM	
WP_002664048.1 Bergeyella zoohelcum ATCC 43767	MKHILGLDLG TNSIGWALIE RNIEEKYGKI IGMGSRIVPM GAELSKFEQG QAQTKNADRR TNRGARRLNK RYKQRRNKLI YILQKLDMLP SQIKLKEDFS DPNKIDKITI LPISKKQEQL TAFDLVSLRV KALTEKVGLE DLGKIIYKYN QLRGYAGGSL EPEKEDIFDE EQSKDKKNKS FIAFSKIVFL GEPQEEIFKN KKLNRRAIIV ETEEGNFEGS TFLENIKVGD SLELLINISA SKSGDTITIK LPNKTNWRKK MENIENQLKE KSKEMGREFY ISEFLLELLK ENRWAKIRNN TILRARYESE FEAIWNEQVK HYPFLENLDK KTLIEIVSFI FPGEKESQKK YRELGLEKGL KYIIKNQVVF YQRELKDQSH LISDCRYEPN EKAIAKSHPV FQEYKVWEQI NKLIVNTKIE AGTNRKGEKK YKYIDRPIPT ALKEWIFEEL QNKKEITFSA IFKKLKAEFD LREGIDFLNG MSPKDKLKGN ETKLQLQKSL GELWDVLGLD SINRQIELWN ILYNEKGNEY DLTSDRTSKV LEFINKYGNN IVDDNAEETA IRISKIKFAR AYSSLSLKAV ERILPLVRAG KYFNNDFSQQ LQSKILKLLN ENVEDPFAKA AQTYLDNNQS VLSEGGVGNS IATILVYDKH TAKEYSHDEL YKSYKEINLL KQGDLRNPLV EQIINEALVL IRDIWKNYGI KPNEIRVELA RDLKNSAKER ATIHKRNKDN QTINNKIKET LVKNKKELSL ANIEKVKLWE AQRHLSPYTG QPIPLSDLFD KEKYDVDHII PISRYFDDSF TNKVISEKSV NQEKANRTAM EYFEVGSLKY SIFTKEQFIA HVNEYFSGVK RKNLLATSIP EDPVQRQIKD TQYIAIRVKE ELNKIVGNEN VKTTTGSITD YLRNHWGLTD KFKLLLKERY EALLESEKFL EAEYDNYKKD FDSRKKEYEE KEVLFEEQEL TREEFIKEYK ENYIRYKKNK LIIKGWSKRI DHRHHAIDAL IVACTEPAHI KRLNDLNKVL QDWLVEHKSE FMPNFEGSNS ELLEEILSLP ENERTEIFTQ IEKFRAIEMP WKGFPEQVEQ KLKEIIISHK PKDKLLLQYN KAGDRQIKLR GQLHEGTLYG ISQGKEAYRI PLTKFGGSKF ATEKNIQKIV SPFLSGFIAN HLKEYNNKKE EAFSAEGIMD LNNKLAQYRN EKGELKPHTP ISTVKIYYKD PSKNKKKKDE EDLSLQKLDR EKAFNEKLYV KTGDNYLFAV LEGEIKTKKT SQIKRLYDII SFFDATNFLK EEFRNAPDKK TFDKDLLFRQ YFEERNKAKL LFTLKQGDFV YLPNENEEVI LDKESPLYNQ YWGDLKERGK NIYVVQKFSK KQIYFIKHTI ADIIKKDVEF GSQNCYETVE GRSIKENCFK LEIDRLGNIV KVIKR	(SEQ ID NO: 42)
CBK78998.1 Coprococcus catus GD/7	MKQEYFLGLD MGTGSLGWAV TDSTYQVMRK HGKALWGTRL FESASTAEER RMFRTARRRL DRRNWRIQVL QEIFSEEISK VDPGFFLRMK ESKYYPEDKR DAEGNCPELP YALFVDDNYT DKNYHKDYPT IYHLRKMLME TTEIPDIRLV YLVLHHMMKH RGHFLLSGDI SQIKEFKSTF EQLIQNIQDE ELEWHISLDD AAIQFVEHVL KDRNLTRSTK KSRLIKQLNA KSACEKAILN LLSGGTVKLS DIFNNKELDE SERPKVSFAD SGYDDYIGIV EAELAEQYYI IASAKAVYDW SVLVEILGNS VSISEAKIKV YQKHQADLKT LKKIVRQYMT KEDYKRVFVD TEEKLNNYSA YIGMTKKNGK KVDLKSKQCT QADFYDFLKK NVIKVIDHKE ITQEIESEIE KENFLPKQVT KDNGVIPYQV HDYELKKILD NLGTRMPFIK ENAEKIQQLF EFRIPYYVGP LNRVDDGKDG KFTWSVRKSD ARIYPWNFTE VIDVEASAEK FIRRMTNKCT YLVGEDVLPK DSLVYSKFMV LNELNNLRLN GEKISVELKQ RIYEELFCKY RKVTRKKLER YLVIEGIAKK GVEITGIDGD FKASLTAYHD FKERLTDVQL SQRAKEAIVL NVVLFGDDKK LLKQRLSKMY	(SEQ ID NO: 43)

Table 4- Cas9 orthologs		
	PNLTTGQLKG ICSLSYQGWG RLSKTFLEEI TVPAPGTGEV WNIMTALWQT NDNLMQLLSR NYGFTNEVEE FNTLKKETDL SYKTVDELYV SPAVKRQIWQ TLKVVKEIQK VMGNAPKRVF VEMAREKQEG KRSDSRKKQL VELYRACKNE ERDWITELNA QSDQQLRSDK LFLYYIQKGR CMYSGETIQL DELWDNTKYD IDHIYPQSKT MDDSLNNRVL VKKNYNAIKS DTYPLSLDIQ KKMMSFWKML QQQGFITKEK YVRLVRSDEL SADELAGFIE RQIVETRQST KAVATILKEA LPDTEIVYVK AGNVSNFRQT YELLKVREMN DLHHAKDAYL NIVVGNAYFV KFTKNAAWFI RNNPGRSYNL KRMFEFDIER SGEIAWKAGN KGSIVTVKKV MQKNNILVTR KAYEVKGGLF DQQIMKKGKG QVPIKGNDER LADIEKYGGY NKAAGTYFML VKSLDKKGKE IRTIEFVPLY LKNQIEINHE SAIQYLAQER GLNSPEILLS KIKIDTLFKV DGFKMWLSGR TGNQLIFKGA NQLILSHQEA AILKGVVKYV NRKNENKDAK LSERDGMTEE KLLQLYDTFL DKLSNTVYSI RLSAQIKTLT EKRAKFIGLS NEDQCIVLNE ILHMFQCQSG SANLKLIGGP GSAGILVMNN NITACKQISV INQSPTGIYE KEIDLIKL	
WP_002235162.1 Neisseria meningitidis Z2491	MAAFKPNPIN YILGLDIGIA SVGWAMVEID EDENPICLID LGVRVFERAE VPKTGDSLAM ARRLARSVRR LTRRRAHRLL RARRLLKREG VLQAADFDEN GLIKSLPNTP WQLRAAALDR KLTPLEWSAV LLHLIKHRGY LSQRKNEGET ADKELGALLK GVADNAHALQ TGDFRTPAEL ALNKFEKESG HIRNQRGDYS HTFSRKDLQA ELILLFEKQK EFGNPHVSGG LKEGIETLLM TQRPALSGDA VQKMLGHCTF EPAEPKAAKN TYTAERFIWL TKLNNLRILE QGSERPLTDT ERATLMDEPY RKSKLTYAQA RKLLGLEDTA FFKGLRYGKD NAEASTLMEM KAYHAISRAL EKEGLKDKKS PLNLSPELQD EIGTAFSLFK TDEDITGRLK DRIQPEILEA LLKHISFDKF VQISLKALRR IVPLMEQGKR YDEACAEIYG DHYGKKNTEE KIYLPPIPAD EIRNPVVLRA LSQARKVING VVRRYGSPAR IHIETAREVG KSFKDRKEIE KRQEENRKDR EKAAAKFREY FPNFVGEPKS KDILKLRLYE QQHGKCLYSG KEINLGRLNE KGYVEIDHAL PFSRTWDDSF NNKVLVLGSE NQNKGNQTPY EYFNGKDNSR EWQEFKARVE TSRFPRSKKQ RILLQKFDED GFKERNLNDT RYVNRFLCQF VADRMRLTGK GKKRVFASNG QITNLLRGFW GLRKVRAEND RHHALDAVVV ACSTVAMQQK ITRFVRYKEM NAFDGKTIDK ETGEVLHQKT HFPQPWEFFA QEVMIRVFGK PDGKPEFEEA DTPEKLRTLL AEKLSSRPEA VHEYVTPLFV SRAPNRKMSG QGHMETVKSA KRLDEGVSVL RVPLTQLKLK DLEKMVNRER EPKLYEALKA RLEAHKDDPA KAFAEPFYKY DKAGNRTQQV KAVRVEQVQK TGVWVRNHNG IADNATMVRV DVFEKGDKYY LVPIYSWQVA KGILPDRAVV QGKDEEDWQL IDDSFNFKFS LHPNDLVEVI TKKARMFGYF ASCHRGTGNI NIRIHDLDHK IGKNGILEGI GVKTALSFQK YQIDELGKEI RPCRLKKRPP VR	(SEQ ID NO: 44)
WP_012414420.1 Elusimicrobium minutum Pei191	MQKNINTKQN HIYIKQAQKI KEKLGDKPYR IGLDLGVGSI GFAIVSMEEN DGNVLLPKEI IMVGSRIFKA SAGAADRKLS RGQRNNHRHT RERMRYLWKV LAEQKLALPV PADLDRKENS SEGETSAKRF LGDVLQKDIY ELRVKSLDER LSLQELGYVL YHIAGHRGSS AIRTFENDSE EAQKENTENK KIAGNIKRLM AKKNYRTYGE YLYKEFFENK EKHKREKISN AANNHKFSPT RDLVIKEAEA ILKKQAGKDG FHKELTEEYI EKLTKAIGYE SEKLIPESGF CPYLKDEKRL PASHKLNEER RLWETLNNAR YSDPIVDIVT GEITGYYEKQ FTKEQKQKLF DYLLTGSELT PAQTKKLLGL KNTNFEDIIL QGRDKKAQKI KGYKLIKLES MPFWARLSEA QQDSFLYDWN SCPDEKLLTE KLSNEYHLTE EEIDNAFNEI VLSSSYAPLG KSAMLIILEK IKNDLSYTEA VEEALKEGKL TKEKQAIKDR LPYYGAVLQE STQKIIAKGF SPQFKDKGYK TPHTNKYELE YGRIANPVVH QTLNELRKLV NEIIDILGKK PCEIGLETAR ELKKSAEDRS KLSREQNDNE SNRNRIYEIY IRPQQQVIIT RRENPRNYIL KFELLEEQKS QCPFCGGQIS PNDIINNQAD IEHLFPIAES EDNGRNNLVI SHSACNADKA KRSPWAAFAS AAKDSKYDYN RILSNVKENI PHKAWRFNQG AFEKFIENKP MAARFKTDNS YISKVAHKYL ACLFEKPNII CVKGSLTAQL RMAWGLQGLM IPFAKQLITE KESESFNKDV NSNKKIRLDN RHHALDAIVI AYASRGYGNL LNKMAGKDYK INYSERNWLS KILLPPNNIV WENIDADLES FESSVKTALK NAFISVKHDH SDNGELVKGT MYKIFYSERG YTLTTYKKLS ALKLTDPQKK KTPKDFLETA LLKFKGRESE MKNEKIKSAI ENNKRLFDVI QDNLEKAKKL LEEENEKSKA EGKKEKNIND ASIYQKAISL SGDKYVQLSK KEPGKFFAIS KPTPTTTGYG YDTGDSLCVD LYYDNKGKLC GEIIRKIDAQ QKNPLKYKEQ GFTLFERIYG	(SEQ ID NO: 45)

Table 4- Cas9 orthologs		
	GDILEVDFDI HSDKNSFRNN TGSAPENRVF IKVGTFTEIT NNNIQIWFGN IIKSTGGQDD SFTINSMQQY NPRKLILSSC GFIKYRSPIL KNKEG	
WP_009105777.1 Treponema sp. JC4	MIMKLEKWRL GLDLGTNSIG WSVFSLDKDN SVQDLIDMGV RIFSDGRDPK TKEPLAVARR TARSQRKLIY RRKLRRKQVF KFLQEQGLFP KTKEECMTLK SLNPYELRIK ALDEKLEPYE LGRALFNLAV RRGFKSNRKD GSREEVSEKK SPDEIKTQAD MQTHLEKAIK ENGCRTITEF LYKNQGENGG IRFAPGRMTY YPTRKMYEEE FNLIRSKQEK YYPQVDWDDI YKAIFYQRPL KPQQRGYCIY ENDKERTFKA MPCSQKLRIL QDIGNLAYYE GGSKKRVELN DNQDKVLYEL LNSKDKVTFD QMRKALCLAD SNSFNLEENR DFLIGNPTAV KMRSKNRFGK LWDEIPLEEQ DLIIETIITA DEDDAVYEVI KKYDLTQEQR DFIVKNTILQ SGTSMLCKEV SEKLVKRLEE IADLKYHEAV ESLGYKFADQ TVEKYDLLPY YGKVLPGSTM EIDLSAPETN PEKHYGKISN PTVHVALNQT RVVVNALIKE YGKPSQIAIE LSRDLKNNVE KKAEIARKQN QRAKENIAIN DTISALYHTA FPGKSFYPNR NDRMKYRLWS ELGLGNKCIY CGKGISGAEL FTKEIEIEHI LPFSRTLLDA ESNLTVAHSS CNAFKAERSP FEAFGTNPSG YSWQEIIQRA NQLKNTSKKN KFSPNAMDSF EKDSSFIARQ LSDNQYIAKA ALRYLKCLVE NPSDVWTTNG SMTKLLRDKW EMDSILCRKF TEKEVALLGL KPEQIGNYKK NRFDHRHHAI DAVVIGLTDR SMVQKLATKN SHKGNRIEIP EFPILRSDLI EKVKNIVVSF KPDHGAEGKL SKETLLGKIK LHGKETFVCR ENIVSLSEKN LDDIVDEKIK SKVKDYVAKH KGQKIEAVLS DFSKENGIKK VRCVNRVQTP IEITSGKISR YLSPEDYFAA VIWEIPGEKK TFKAQYIRRN EVEKNSKGLN VVKPAVLENG KPHPAAKQVC LLHKDDYLEF SDKGKMYFCR IAGYAATNNK LDIRPVYAVS YCADWINSTN ETMLTGYWKP TPTQNWVSVN VLFDKQKARL VTVSPIGRVF RK	(SEQ ID NO: 46)
WP_002460848.1 Staphylococcus lugdunensis M23590	MNQKFILGLD IGITSVGYGL IDYETKNIID AGVRLFPEAN VENNEGRRSK RGSRRLKRRR IHRLERVKKL LEDYNLLDQS QIPQSTNPYA IRVKGLSEAL SKDELVIALL HIAKRRGIHK IDVIDSNDDV GNELSTKEQL NKNSKLLKDK FVCQIQLERM NEGQVRGEKN RFKTADIIKE IIQLLNVQKN FHQLDENFIN KYIELVEMRR EYFEGPGKGS PYGWEGDPKA WYETLMGHCT YFPDELRSVK YAYSADLFNA LNDLNNLVIQ RDGLSKLEYH EKYHIIENVF KQKKKPTLKQ IANEINVNPE DIKGYRITKS GKPQFTEFKL YHDLKSVLFD QSILENEDVL DQIAEILTIY QDKDSIKSKL TELDILLNEE DKENIAQLTG YTGTHRLSLK CIRLVLEEQW YSSRNQMEIF THLNIKPKKI NLTAANKIPK AMIDEFILSP VVKRTFGQAI NLINKIIEKY GVPEDIIIEL ARENNSKDKQ KFINEMQKKN ENTRKRINEI IGKYGNQNAK RLVEKIRLHD EQEGKCLYSL ESIPLEDLLN NPNHYEVDHI IPRSVSFDNS YHNKVLVKQS ENSKKSNLTP YQYFNSGKSK LSYNQFKQHI LNLSKSQDRI SKKKKEYLLE ERDINKFEVQ KEFINRNLVD TRYATRELTN YLKAYFSANN MNVKVKTING SFTDYLRKVW KFKKERNHGY KHHAEDALII ANADFLFKEN KKLKAVNSVL EKPEIESKQL DIQVDSEDNY SEMFIIPKQV QDIKDFRNFK YSHRVDKKPN RQLINDTLYS TRKKDNSTYI VQTIKDIYAK DNTTLKKQFD KSPEKFLMYQ HDPRTFEKLE VIMKQYANEK NPLAKYHEET GEYLTKYSKK NNGPIVKSLK YIGNKLGSHL DVTHQFKSST KKLVKLSIKP YRFDVYLTDK GYKFITISYL DVLKKDNYYY IPEQKYDKLK LGKAIDKNAK FIASFYKNDL IKLDGEIYKI IGVNSDTRNM IELDLPDIRY KEYCELNNIK GEPRIKKTIG KKVNSIEKLT TDVLGNVFTN TQYTKPQLLF KRGN	(SEQ ID NO: 47)
WP_011681470.1 Streptococcus thermophilus LMD-9	MTKPYSIGLD IGTNSVGWAV TTDNYKVPSK KMKVLGNTSK KYIKKNLLGV LLFDSGITAE GRRLKRTARR RYTRRRNRIL YLQEIFSTEM ATLDDAFFQR LDDSFLVPDD KRDSKYPIFG NLVEEKAYHD EFPTIYHLRK YLADSTKKAD LRLVYLALAH MIKYRGHFLI EGEFNSKNND IQKNFQDFLD TYNAIFESDL SLENSKQLEE IVKDKISKLE KKDRILKLFP GEKNSGIFSE FLKLIVGNQA DFRKCFNLDE KASLHFSKES YDEDLETLLG YIGDDYSDVF LKAKKLYDAI LLSGFLTVTD NETEAPLSSA MIKRYNEHKE DLALLKEYIR NISLKTYNEV FKDDTKNGYA GYIDGKTNQE DFYVYLKKLL AEFEGADYFL EKIDREDFLR KQRTFDNGSI PYQIHLQEMR AILDKQAKFY PFLAKNKERI EKILTFRIPY YVGPLARGNS DFAWSIRKRN EKITPWNFED VIDKESSAEA FINRMTSFDL YLPEEKVLPK HSLLYETFNV YNELTKVRFI AESMRDYQFL DSKQKKDIVR	(SEQ ID NO: 48)

Table 4- Cas9 orthologs		
	LYFKDKRKVT DKDIIEYLHA IYGYDGIELK GIEKQFNSSL STYHDLLNII NDKEFLDDSS NEAIIEEIIH TLTIFEDREM IKQRLSKFEN IFDKSVLKKL SRRHYTGWGK LSAKLINGIR DEKSGNTILD YLIDDGISNR NFMQLIHDDA LSFKKKIQKA QIIGDEDKGN IKEVVKSLPG SPAIKKGILQ SIKIVDELVK VMGGRKPESI VVEMARENQY TNQGKSNSQQ RLKRLEKSLK ELGSKILKEN IPAKLSKIDN NALQNDRLYL YYLQNGKDMY TGDDLDIDRL SNYDIDHIIP QAFLKDNSID NKVLVSSASN RGKSDDVPSL EVVKKRKTFW YQLLKSKLIS QRKFDNLTKA ERGGLSPEDK AGFIQRQLVE TRQITKHVAR LLDEKFNNKK DENNRAVRTV KIITLKSTLV SQFRKDFELY KVREINDFHH AHDAYLNAVV ASALLKKYPK LEPEFVYGDY PKYNSFRERK SATEKVYFYS NIMNIFKKSI SLADGRVIER PLIEVNEETG ESVWNKESDL ATVRRVLSYP QVNVVKKVEE QNHGLDRGKP KGLFNANLSS KPKPNSNENL VGAKEYLDPK KYGGYAGISN SFTVLVKGTI EKGAKKKITN VLEFQGISIL DRINYRKDKL NFLLEKGYKD IELIIELPKY SLFELSDGSR RMLASILSTN NKRGEIHKGN QIFLSQKFVK LLYHAKRISN TINENHRKYV ENHKKEFEEL FYYILEFNEN YVGAKKNGKL LNSAFQSWQN HSIDELCSSF IGPTGSERKG LFELTSRGSA ADFEFLGVKI PRYRDYTPSS LLKDATLIHQ SVTGLYETRI DLAKLGEG	
WP_009293010.1 Bacteroides fragilis NCTC 9343 Cas9	MKRILGLDLG TNSIGWALVN EAENKDERSS IVKLGVRVNP LTVDELTNFE KGKSITTNAD RTLKRGMRRN LQRYKLRRET LTEVLKEHKL ITEDTILSEN GNRTTFETYR LRAKAVTEEI SLEEFARVLL MINKKRGYKS SRKAKGVEEG TLIDGMDIAR ELYNNNLTPG ELCLQLLDAG KKFLPDFYRS DLQNELDRIW EKQKEYYPEI LTDVLKEELR GKKRDAVWAI CAKYFVWKEN YTEWNKEKGK TEQQEREHKL EGIYSKRKRD EAKRENLQWR VNGLKEKLSL EQLVIVFQEM NTQINNSSGY LGAISDRSKE LYFNKQTVGQ YQMEMLDKNP NASLRNMVFY RQDYLDEFNM LWEKQAVYHK ELTEELKKEI RDIIIFYQRR LKSQKGLIGF CEFESRQIEV DIDGKKKIKT VGNRVISRSS PLFQEFKIWQ ILNNIEVTVV GKKRKRRKLK ENYSALFEEL NDAEQLELNG SRRLCQEEKE LLAQELFIRD KMTKSEVLKL LFDNPQELDL NFKTIDGNKT GYALFQAYSK MIEMSGHEPV DFKKPVEKVV EYIKAVFDLL NWNTDILGFN SNEELDNQPY YKLWHLLYSF EGDNTPTGNG RLIQKMTELY GFEKEYATIL ANVSFQDDYG SLSAKAIHKI LPHLKEGNRY DVACVYAGYR HSESSLTREE IANKVLKDRL MLLPKNSLHN PVVEKILNQM VNVINVIIDI YGKPDEIRVE LARELKKNAK EREELTKSIA QTTKAHEEYK TLLQTEFGLT NVSRTDILRY KLYKELESCG YKTLYSNTYI SREKLFSKEF DIEHIIPQAR LFDDSFSNKT LEARSVNIEK GNKTAYDFVK EKFGESGADN SLEHYLNNIE DLFKSGKISK TKYNKLKMAE QDIPDGFIER DLRNTQYIAK KALSMLNEIS HRVVATSGSV TDKLREDWQL IDVMKELNWE KYKALGLVEY FEDRDGRQIG RIKDWTKRND HRHHAMDALT VAFTKDVFIQ YFNNKNASLD PNANEHAIKN KYFQNGRAIA PMPLREFRAE AKKHLENTLI SIKAKNKVIT GNINKTRKKG GVNKNMQQTP RGQLHLETIY GSGKQYLTKE EKVNASFDMR KIGTVSKSAY RDALLKRLYE NDNDPKKAFA GKNSLDKQPI WLDKEQMRKV PEKVKIVTLE AIYTIRKEIS PDLKVDKVID VGVRKILIDR LNEYGNDAKK AFSNLDKNPI WLNKEKGISI KRVTISGISN AQSLHVKKDK DGKPILDENG RNIPVDFVNT GNNHHVAVYY RPVIDKRGQL VVDEAGNPKY ELEEVVVSFF EAVTRANLGL PIIDKDYKTT EGWQFLFSMK QNEYFVFPNE KTGFNPKEID LLDVENYGLI SPNLFRVQKF SLKNYVFRHH LETTIKDTSS ILRGITWIDF RSSKGLDTIV KVRVNHIGQI VSVGEY	(SEQ ID NO: 49)
AOL40912.1 Veillonella atypica ACS-134-V-Col7a	METQTSNQLI TSHLKDYPKQ DYFVGLDIGT NSVGWAVTNT SYELLKFHSH KMWGSRLFEE GESAVTRRGF RSMRRRLERR KLRLKLLEEL FADAMAQVDS TFFIRLHESK YHYEDKTTGH SSKHILFIDE DYTDQDYFTE YPTIYHLRKD LMENGTDDIR KLFLAVHHIL KYRGNFLYEG ATFNSNAFTF EDVLKQALVN ITFNCFDTNS AISSISNILM ESGKTKSDKA KAIERLVDTY TVFDEVNTPD KPQKEQVKED KKTLKAFANL VLGLSANLID LFGSVEDIDD DLKKLQIVGD TYDEKRDELA KVWGDEIHII DDCKSVYDAI ILMSIKEPGL TISQSKVKAF DKHKEDLVIL KSLLKLDRNV YNEMFKSDKK GLHNYVHYIK QGRTEETSCS REDFYKYTKK IVEGLADSKD KEYILNEIEL QTLLPLQRIK DNGVIPYQLH LEELKVILDK CGPKFPFLHT VSDGFSVTEK LIKMLEFRIP YYVGPLNTHH NIDNGGFSWA VRKQAGRVTP WNFEEKIDRE KSAAAFIKNL TNKCTYLFGE DVLPKSSLLY SEFMLLNELN NVRIDGKALA QGVKQHLIDS IFKQDHKKMT	(SEQ ID NO: 50)

Table 4- Cas9 orthologs

	KNRIELFLKD NNYITKKHKP EITGLDGEIK NDLTSYRDMV RILGNNFDVS MAEDIITDIT IFGESKKMLR QTLRNKFGSQ LNDETIKKLS KLRYRDWGRL SKKLLKGIDG CDKAGNGAPK TIIELMRNDS YNLMEILGDK FSFMECIEEE NAKLAQGQVV NPHDIIDELA LSPAVKRAVW QALRIVDEVA HIKKALPSRI FVEVARTNKS EKKKKDSRQK RLSDLYSAIK KDDVLQSGLQ DKEFGALKSG LANYDDAALR SKKLYLYYTQ MGRCAYTGNI IDLNQLNTDN YDIDHIYPRS LTKDDSFDNL VLCERTANAK KSDIYPIDNR IQTKQKPFWA FLKHQGLISE RKYERLTRIA PLTADDLSGF IARQLVETNQ SVKATTTLLR RLYPDIDVVF VKAENVSDFR HNNNFIKVRS LNHHHHAKDA YLNIVVGNVY HEKFTRNFRL FFKKNGANRT YNLAKMFNYD VICTNAQDGK AWDVKTSMNT VKKMMASNDV RVTRRLLEQS GALADATIYK ASVAAKAKDG AYIGMKTKYS VFADVTKYGG MTKIKNAYSI IVQYTGKKGE EIKEIVPLPI YLINRNATDI ELIDYVKSVI PKAKDISIKY RKLCINQLVK VNGFYYYLGG KTNDKIYIDN AIELVVPHDI ATYIKLLDKY DLLRKENKTL KASSITTSIY NINTSTVVSL SNKVGIDVFD YFMSKLRTPL YMKMKGNKVD ELSSTGRSKF IKMTLEEQSI YLLEVLNLLT NSKTTFDVKP LGITGSRSTI GVKIHNLDEF KIINESITGL YSNEVTIV	
WP_013389026.1 Ilyobacter polytropus DSM 2926	MKYSIGLDIG IASVGWSVIN KDKERIEDMG VRIFQKAENP KDGSSLASSR REKRGSRRRN RRKKHRLDRI KNILCESGLV KKNEIEKIYK NAYLKSPWEL RAKSLEAKIS NKEIAQILLH IAKRRGFKSF RKTDRNADDT GKLLSGIQEN KKIMEEKGYL TIGDMVAKDP KFNTHVRNKA GSYLFSFSRK LLEDEVRKIQ AKQKELGNTH FTDDVLEKYI EVFNSQRNFD EGPSKPSPYY SEIGQIAKMI GNCTFESSEK RTAKNTWSGE RFVFLQKLNN FRIVGLSGKR PLTEEERDIV EKEVYLKKEV RYEKLRKILY LKEEERFGDL NYSKDEKQDK KTEKTKFISL IGNYTIKKLN LSEKLKSEIE EDKSKLDKII EILTFNKSDK TIESNLKKLE LSREDIEILL SEEFSGTLNL SLKAIKKILP YLEKGLSYNE ACEKADYDYK NNGIKFKRGE LLPVVDKDLI ANPVVLRAIS QTRKVVNAII RKYGTPHTIH VEVARDLAKS YDDRQTIIKE NKKRELENEK TKKFISEEFG IKNVKGKLLL KYRLYQEQEG RCAYSRKELS LSEVILDESM TDIDHIIPYS RSMDDSYSNK VLVLSGENRK KSNLLPKEYF DRQGRDWDTF VLNVKAMKIH PRKKSNLLKE KFTREDNKDW KSRALNDTRY ISRFVANYLE NALEYRDDSP KKRVFMIPGQ LTAQLRARWR LNKVRENGDL HHALDAAVVA VTDQKAINNI SNISRYKELK NCKDVIPSIE YHADEETGEV YFEEVKDTRF PMPWSGFDLE LQKRLESENP REEFYNLLSD KRYLGWFNYE EGFIEKLRPV FVSRMPNRGV KGQAHQETIR SSKKISNQIA VSKKPLNSIK LKDLEKMQGR DTDRKLYEAL KNRLEEYDDK PEKAFAEPFY KPTNSGKRGP LVRGIKVEEK QNVGVYVNGG QASNGSMVRI DVFRKNGKFY TVPIYVHQTL LKELPNRAIN GKPYKDWDLI DGSFEFLYSF YPNDLIEIEF GKSKSIKNDN KLTKTEIPEV NLSEVLGYYR GMDTSTGAAT IDTQDGKIQM RIGIKTVKNI KKYQVDVLGN VYKVKREKRQ TF	(SEQ ID NO: 51)
WP_005864263.1 Parabacteroides sp. 20_3	MKKIVGLDLG TNSIGWALIN AYINKEHLYG IEACGSRIIP MDAAILGNFD KGNSISQTAD RTSYRGIRRL RERHLLRRER LHRILDLLGF LPKHYSDSLN RYGKFLNDIE CKLPWVKDET GSYKFIFQES FKEMLANFTE HHPILIANNK KVPYDWTIYY LRKKALTQKI SKEELAWILL NFNQKRGYYQ LRGEEEETPN KLVEYYSLKV EKVEDSGERK GKDTWYNVHL ENGMIYRRTS NIPLDWEGKT KEFIVTTDLE ADGSPKKDKE GNIKRSFRAP KDDDWTLIKK KTEADIDKIK MTVGAYIYDT LLQKPDQKIR GKLVRTIERK YYKNELYQIL KTQSEFHEEL RDKQLYIACL NELYPNNEPR RNSISTRDFC HLFIEDIIFY QRPLKSKKSL IDNCPYEENR YIDKESGEIK HASIKCIAKS HPLYQEFRLW QFIVNLRIYR KETDVDVTQE LLPTEADYVT LFEWLNEKKE IDQKAFFKYP PFGFKKTTSN YRWNYVEDKP YPCNETHAQI IARLGKAHIP KAFLSKEKEE TLWHILYSIE DKQEIEKALH SFANKNNLSE EFIEQFKNFP PFKKEYGSYS AKAIKKLLPL MRMGKYWSIE NIDNGTRIRI NKIIDGEYDE NIRERVRQKA INLTDITHFR ALPLWLACYL VYDRHSEVKD IVKWKTPKDI DLYLKSFKQH SLRNPIVEQV ITETLRTVRD IWQQVGHIDE IHIELGREMK NPADKRARMS QQMIKNENTN LRIKALLTEF LNPEFGIENV RPYSPSQQDL LRIYEEGVLN SILELPEDIG IILGKFNQTD TLKRPTRSEI LRYKLWLEQK YRSPYTGEMI PLSKLFTPAY EIEHIIPQSR YFDDSLSNKV ICESEINKLK DRSLGYEFIK NHHGEKVELA FDKPVEVLSV EAYEKLVHES YSHNRSKMKK LLMEDIPDQF IERQLNDSRY ISKVVKSLLS NIVREENEQE AISKNVIPCT GGITDRLKKD WGINDVWNKI	(SEQ ID NO: 52)

Table 4- Cas9 orthologs

	VLPRFIRLNE LTESTRFTSI NTNNTMIPSM PLELQKGFNK KRIDHRHHAM DAIIIACANR NIVNYLNNVS ASKNTKITRR DLQTLLCHKD KTDNNGNYKW VIDKPWETFT QDTLTALQKI TVSFKQNLRV INKTTNHYQH YENGKKIVSN QSKGDSWAIR KSMHKETVHG EVNLRMIKTV SFNEALKKPQ AIVEMDLKKK ILAMLELGYD TKRIKNYFEE NKDTWQDINP SKIKVYYFTK ETKDRYFAVR KPIDTSFDKK KIKESITDTG IQQIMLRHLE TKDNDPTLAF SPDGIDEMNR NILILNKGKK HQPIYKVRVY EKAEKFTVGQ KGNKRTKFVE AAKGTNLFFA IYETEEIDKD TKKVIRKRSY STIPLNVVIE RQKQGLSSAP EDENGNLPKY ILSPNDLVYV PTQEEINKGE VVMPIDRDRI YKMVDSSGIT ANFIPASTAN LIFALPKATA EIYCNGENCI QNEYGIGSPQ SKNQKAITGE MVKEICFPIK VDRLGNIIQV GSCILTN	
GAP01010.1 Fructobacillus fructosus KCTC 3544	MVYDVGLDIG TGSVGWVALD ENGKLARAKG KNLVGVRLFD TAQTAADRRG FRTTRRRLSR RKWRLRLLDE LFSAEINEID SSFFQRLKYS YVHPKDEENK AHYYGGYLFP TEEETKKFHR SYPTIYHLRQ ELMAQPNKRF DIREIYLAIH HLVKYRGHFL SSQEKITIGS TYNPEDLANA IEVYADEKGL SWELNNPEQL TEIISGEAGY GLNKSMKADE ALKLFEFDNN QDKVAIKTLL AGLTGNQIDF AKLFGKDISD KDEAKLWKLK LDDEALEEKS QTILSQLTDE EIELFHAVVQ AYDGFVLIGL LNGADSVSAA MVQLYDQHRE DRKLLKSLAQ KAGLKHKRFS EIYEQLALAT DEATIKNGIS TARELVEESN LSKEVKEDTL RRLDENEFLP KQRTKANSVI PHQLHLAELQ KILQNQGQYY PFLLDTFEKE DGQDNKIEEL LRFRIPYYVG PLVTKKDVEH AGGDADNHWV ERNEGFEKSR VTPWNFDKVF NRDKAARDFI ERLTGNDTYL IGEKTLPQNS LRYQLFTVLN ELNNVRVNGK KFDSKTKADL INDLFKARKT VSLSALKDYL KAQGKGDVTI TGLADESKFN SSLSSYNDLK KTFDAEYLEN EDNQETLEKI IEIQTVFEDS KIASRELSKL PLDDDQVKKL SQTHYTGWGR LSEKLLDSKI IDERGQKVSI LDKLKSTSQN FMSIINNDKY GVQAWITEQN TGSSKLTFDE KVNELTTSPA NKRGIKQSFA VLNDIKKAMK EEPRRVYLEF AREDQTSVRS VPRYNQLKEK YQSKSLSEEA KVLKKTLDGN KNKMSDDRYF LYFQQQGKDM YTGRPINFER LSQDYDIDHI IPQAFTKDDS LDNRVLVSRP ENARKSDSFA YTDEVQKQDG SLWTSLLKSG FINRKKYERL TKAGKYLDGQ KTGFIARQLV ETRQIIKNVA SLIEGEYENS KAVAIRSEIT ADMRLLVGIK KHREINSFHH AFDALLITAA GQYMQNRYPD RDSTNVYNEF DRYTNDYLKN LRQLSSRDEV RRLKSFGFVV GTMRKGNEDW SEENTSYLRK VMMFKNILTT KKTEKDRGPL NKETIFSPKS GKKLIPLNSK RSDTALYGGY SNVYSAYMTL VRANGKNLLI KIPISIANQI EVGNLKINDY IVNNPAIKKF EKILISKLPL GQLVNEDGNL IYLASNEYRH NAKQLWLSTT DADKIASISE NSSDEELLEA YDILTSENVK NRFPFFKKDI DKLSQVRDEF LDSDKRIAVI QTILRGLQID AAYQAPVKII SKKVSDWHKL QQSGGIKLSD NSEMIYQSAT GIFETRVKIS DLL	(SEQ ID NO: 53)
Bacillus smithii WP_003354196.1	MNYKMGLDIG IASVGWAVIN LDLKRIEDLG VRIFDKAEHP QNGESLALPR RIARSARRRL RRRKHRLERI RRLLVSENVL TKEEMNLLFK QKKQIDVWQL RVDALERKLN NDELARVLLH LAKRRGFKSN RKSERNSKES SEFLKNIEEN QSILAQYRSV GEMIVKDSKF AYHKRNKLDS YSNMIARDDL EREIKLIFEK QREFNNPVCT ERLEEKYLNI WSSQRPFASK EDIEKKVGFC TFEPKEKRAP KATYTFQSFI VWEHINKLRL VSPDETRALT EIERNLLYKQ AFSKNKMTYY DIRKLLNLSD DIHFKGLLYD PKSSLKQIEN IRFLELDSYH KIRKCIENVY GKDGIRMFNE TDIDTFGYAL TIFKDDEDIV AYLQNEYITK NGKRVSNLAN KVYDKSLIDE LLNLSFSKFA HLSMKAIRNI LPYMEQGEIY SKACELAGYN FTGPKKKEKA LLLPVIPNIA NPVVMRALTQ SRKVVNAIIK KYGSPVSIHI ELARDLSHSF DERKKIQKDQ TENRKKNETA IKQLIEYELT KNPTGLDIVK FKLWSEQQGR CMYSLKPIEL ERLLEPGYVE VDHILPYSRS LDDSYANKVL VLTKENREKG NHTPVEYLGL GSERWKKFEK FVLANKQFSK KKKQNLLRLR YEETEEKEFK ERNLNDTRYI SKFFANFIKE HLKFADGDGG QKVYTINGKI TAHLRSRWDF NKNREESDLH HAVDAVIVAC ATQGMIKKIT EFYKAREQNK ESAKKKEPIF PQPWPHFADE LKARLSKFPQ ESIEAFALGN YDRKKLESLR PVFVSRMPKR SVTGAAHQET LRRCVGIDEQ SGKIQTAVKT KLSDIKLDKD GHFPMYQKES DPRTYEAIRQ RLLEHNNDPK KAFQEPLYKP KKNGEPGPVI RTVKIIDTKN KVVHLDGSKT VAYNSNIVRT DVFEKDGKYY CVPVYTMDIM KGTLPNKAIE ANKPYSEWKE MTEEYTFQFS LFPNDLVRIV LPREKTIKTS	(SEQ ID NO: 54)

Table 4- Cas9 orthologs		
	TNEEIIIKDI FAYYKTIDSA TGGLELISHD RNFSLRGVGS KTLKRFEKYQ VDVLGNIHKV KGEKRVGLAA PTNQKKGKTV DSLQSVSD	
Mycoplasma canis PG 14 EIE39736.1 WP_004794730.1	MEKKRKVTLG FDLGIASVGW AIVDSETNQV YKLGSRLFDA PDTNLERRTQ RGTRRLLRRR KYRNQKFYNL VKRTEVFGLS SREAIENRFR ELSIKYPNII ELKTKALSQE VCPDEIAWIL HDYLKNRGYF YDEKETKEDF DQQTVESMPS YKLNEFYKKY GYFKGALSQP TESEMKDNKD LKEAFFFDFS NKEWLKEINY FFNVQKNILS ETFIEEFKKI FSFTRDISKG PGSDNMPSPY GIFGEFGDNG QGGRYEHIWD KNIGKCSIFT NEQRAPKYLP SALIFNFLNE LANIRLYSTD KKNIQPLWKL SSVDKLNILL NLFNLPISEK KKKLTSTNIN DIVKKESIKS IMISVEDIDM IKDEWAGKEP NVYGVGLSGL NIEESAKENK FKFQDLKILN VLINLLDNVG IKFEFKDRND IIKNLELLDN LYLFLIYQKE SNNKDSSIDL FIAKNESLNI ENLKLKLKEF LLGAGNEFEN HNSKTHSLSK KAIDEILPKL LDNNEGWNLE AIKNYDEEIK SQIEDNSSLM AKQDKKYLND NFLKDAILPP NVKVTFQQAI LIFNKIIQKF SKDFEIDKVV IELAREMTQD QENDALKGIA KAQKSKKSLV EERLEANNID KSVFNDKYEK LIYKIFLWIS QDFKDPYTGA QISVNEIVNN KVEIDHIIPY SLCFDDSSAN KVLVHKQSNQ EKSNSLPYEY IKQGHSGWNW DEFTKYVKRV FVNNVDSILS KKERLKKSEN LLTASYDGYD KLGFLARNLN DTRYATILFR DQLNNYAEHH LIDNKKMFKV IAMNGAVTSF IRKNMSYDNK LRLKDRSDFS HHAYDAAIIA LFSNKTKTLY NLIDPSLNGI ISKRSEGYWV IEDRYTGEIK ELKKEDWTSI KNNVQARKIA KEIEEYLIDL DDEVFFSRKT KRKTNRQLYN ETIYGIATKT DEDGITNYYK KEKFSILDDK DIYLRLLRER EKFVINQSNP EVIDQIIEII ESYGKENNIP SRDEAINIKY TKNKINYNLY LKQYMRSLTK SLDQFSEEFI NQMIANKTFV LYNPTKNTTR KIKFLRLVND VKINDIRKNQ VINKFNGKNN EPKAFYENIN SLGAIVFKNS ANNFKTLSIN TQIAIFGDKN WDIEDFKTYN MEKIEKYKEI YGIDKTYNFH SFIFPGTILL DKQNKEFYYI SSIQTVRDII EIKFLNKIEF KDENKNQDTS KTPKRLMFGI KSIMNNYEQV DISPFGINKK IFE	(SEQ ID NO: 55)
Odoribacter laneus YIT EHP49880.1	METTLGIDLG TNSIGLALVD QEEHQILYSG VRIFPEGINK DTIGLGEKEE SRNATRRAKR QMRRQYFRKK LRKAKLLELL IAYDMCPLKP EDVRRWKNWD KQQKSTVRQF PDTPAFREWL KQNPYELRKQ AVTEDVTRPE LGRILYQMIQ RRGFLSSRKG KEEGKIFTGK DRMVGIDETR KNLQKQTLGA YLYDIAPKNG EKYRFRTERV RARYTLRDMY IREFEIIWQR QAGHLGLAHE QATRKKNIFL EGSATNVRNS KLITHLQAKY GRGHVLIEDT RITVTFQLPL KEVLGGKIEI EEEQLKFKSN ESVLFWQRPL RSQKSLLSKC VFEGRNFYDP VHQKWIIAGP TPAPLSHPEF EEFRAYQFIN NIIYGKNEHL TAIQREAVFE LMCTESKDFN FEKIPKHLKL FEKFNFDDTT KVPACTTISQ LRKLFPHPVW EEKREEIWHC FYFYDDNTLL FEKLQKDYAL QTNDLEKIKK IRLSESYGNV SLKAIRRINP YLKKGYAYST AVLLGGIRNS FGKRFEYFKE YEPEIEKAVC RILKEKNAEG EVIRKIKDYL VHNRFGFAKN DRAFQKLYHH SQAITTQAQK ERLPETGNLR NPIVQQGLNE LRRTVNKLLA TCREKYGPSF KFDHIHVEMG RELRSSKTER EKQSRQIREN EKKNEAAKVK LAEYGLKAYR DNIQKYLLYK EIEEKGGTVC CPYTGKTLNI SHTLGSDNSV QIEHIIPYSI SLDDSLANKT LCDATFNREK GELTPYDFYQ KDPSPEKWGA SSWEEIEDRA FRLLPYAKAQ RFIRRKPQES NEFISRQLND TRYISKKAVE YLSAICSDVK AFPGQLTAEL RHLWGLNNIL QSAPDITFPL PVSATENHRE YYVITNEQNE VIRLFPKQGE TPRTEKGELL LTGEVERKVF RCKGMQEFQT DVSDGKYWRR IKLSSSVTWS PLFAPKPISA DGQIVLKGRI EKGVFVCNQL KQKLKTGLPD GSYWISLPVI SQTFKEGESV NNSKLTSQQV QLFGRVREGI FRCHNYQCPA SGADGNFWCT LDTDTAQPAF TPIKNAPPGV GGGQIILTGD VDDKGIFHAD DDLHYELPAS LPKGKYYGIF TVESCDPTLI PIELSAPKTS KGENLIEGNI WVDEHTGEVR FDPKKNREDQ RHHAIDAIVI ALSSQSLFQR LSTYNARREN KKRGLDSTEH FPSPWPGFAQ DVRQSVVPLL VSYKQNPKTL CKISKTLYKD GKKIHSCGNA VRGQLHKETV YGQRTAPGAT EKSYHIRKDI RELKTSKHIG KVVDITIRQM LLKHLQENYH IDITQEFNIP SNAFFKEGVY RIFLPNKHGE PVPIKKIRMK EELGNAERLK DNINQYVNPR NNHHVMIYQD ADGNLKEEIV SFWSVIERQN QGQPIYQLPR EGRNIVSILQ INDTFLIGLK EEEPEVYRND LSTLSKHLYR VQKLSGMYYT FRHHLASTLN NEREEFRIQS LEAWKRANPV KVQIDEIGRI TFLNGPLC	(SEQ ID NO: 56)

Table 4- Cas9 orthologs		
Akkermansia muciniphila ATCC BAA-835 WP_012421034.1	MSRSLTFSFD IGYASIGWAV IASASHDDAD PSVCGCGTVL FPKDDCQAFK RREYRRLRRN IRSRRVRIER IGRLLVQAQI ITPEMKETSG HPAPFYLASE ALKGHRTLAP IELWHVLRWY AHNRGYDNNA SWSNSLSEDG GNGEDTERVK HAQDLMDKHG TATMAETICR ELKLEEGKAD APMEVSTPAY KNLNTAFPRL IVEKEVRRIL ELSAPLIPGL TAEIIELIAQ HHPLTTEQRG VLLQHGIKLA RRYRGSLLFG QLIPRFDNRI ISRCPVTWAQ VYEAELKKGN SEQSARERAE KLSKVPTANC PEFYEYRMAR ILCNIRADGE PLSAEIRREL MNQARQEGKL TKASLEKAIS SRLGKETETN VSNYFTLHPD SEEALYLNPA VEVLQRSGIG QILSPSVYRI AANRLRRGKS VTPNYLLNLL KSRGESGEAL EKKIEKESKK KEADYADTPL KPKYATGRAP YARTVLKKVV EEILDGEDPT RPARGEAHPD GELKAHDGCL YCLLDTDSSV NQHQKERRLD TMTNNHLVRH RMLILDRLLK DLIQDFADGQ KDRISRVCVE VGKELTTFSA MDSKKIQREL TLRQKSHTDA VNRLKRKLPG KALSANLIRK CRIAMDMNWT CPFTGATYGD HELENLELEH IVPHSFRQSN ALSSLVLTWP GVNRMKGQRT GYDFVEQEQE NPVPDKPNLH ICSLNNYREL VEKLDDKKGH EDDRRRKKKR KALLMVRGLS HKHQSQNHEA MKEIGMTEGM MTQSSHLMKL ACKSIKTSLP DAHIDMIPGA VTAEVRKAWD VFGVFKELCP EAADPDSGKI LKENLRSLTH LHHALDACVL GLIPYIIPAH HNGLLRRVLA MRRIPEKLIP QVRPVANQRH YVLNDDGRMM LRDLSASLKE NIREQLMEQR VIQHVPADMG GALLKETMQR VLSVDGSGED AMVSLSKKKD GKKEKNQVKA SKLVGVFPEG PSKLKALKAA IEIDGNYGVA LDPKPVVIRH IKVFKRIMAL KEQNGGKPVR ILKKGMLIHL TSSKDPKHAG VWRIESIQDS KGGVKLDLQR AHCAVPKNKT HECNWREVDL ISLLKKYQMK RYPTSYTGTP R	(SEQ ID NO: 57)
Dinoroseobacter shibae DFL 12 = DSM 16493 WP_012177079.1	MRLGLDIGTS SIGWWLYETD GAGSDARITG VVDGGVRIFS DGRDPKSGAS LAVDRRAARA MRRRRDRYLR RRATLMKVLA ETGLMPADPA EAKALEALDP FALRAAGLDE PLPLPHLGRA LFHLNQRRGF KSNRKTDRGD NESGKIKDAT ARLDMEMMAN GARTYGEFLH KRRQKATDPR HVPSVRTRLS IANRGGPDGK EEAGYDFYPD RRHLEEEFHK LWAAQGAHHP ELTETLRDLL FEKIFFQRPL KEPEVGLCLF SGHHGVPPKD PRLPKAHPLT QRRVLYETVN QLRVTADGRE ARPLTREERD QVIHALDNKK PTKSLSSMVL KLPALAKVLK LRDGERFTLE TGVRDAIACD PLRASPAHPD RFGPRWSILD ADAQWEVISR IRRVQSDAEH AALVDWLTEA HGLDRAHAEA TAHAPLPDGY GRLGLTATTR ILYQLTADVV TYADAVKACG WHHSDGRTGE CFDRLPYYGE VLERHVIPGS YHPDDDDITR FGRITNPTVH IGLNQLRRLV NRIIETHGKP HQIVVELARD LKKSEEQKRA DIKRIRDTTE AAKKRSEKLE ELEIEDNGRN RMLLRLWEDL NPDDAMRRFC PYTGTRISAA MIFDGSCDVD HILPYSRTLD DSFPNRTLCL REANRQKRNQ TPWQAWGDTP HWHAIAANLK NLPENKRWRF APDAMTRFEG ENGFLDRALK DTQYLARISR SYLDTLFTKG GHVWVVPGRF TEMLRRHWGL NSLLSDAGRG AVKAKNRTDH RHHAIDAAVI AATDPGLLNR ISRAAGQGEA AGQSAELIAR DTPPPWEGFR DDLRVRLDRI IVSHRADHGR IDHAARKQGR DSTAGQLHQE TAYSIVDDIH VASRTDLLSL KPAQLLDEPG RSGQVRDPQL RKALRVATGG KTGKDFENAL RYFASKPGPY QAIRRVRIIK PLQAQARVPV PAQDPIKAYQ GGSNHLFEIW RLPDGEIEAQ VITSFEAHTL EGEKRPHPAA KRLLRVHKGD MVALERDGRR VVGHVQKMDI ANGLFIVPHN EANADTRNND KSDPFKWIQI GARPAIASGI RRVSVDEIGR LRDGGTRPI	(SEQ ID NO: 58)
Wolinella succinogenes DSM 1740 WP_011139289.1	MIERILGVDL GISSLGWAIV EYDKDDEAAN RIIDCGVRLF TAAETPKKKE SPNKARREAR GIRRVLNRRR VRMNMIKKLF LRAGLIQDVD LDGEGGMFYS KANRADVWEL RHDGLYRLLK GDELARVLIH IAKHRGYKFI GDDEADEESG KVKKAGVVLR QNFEAAGCRT VGEWLWRERG ANGKKRNKHG DYEISIHRDL LVEEVEAIFV AQQEMRSTIA TDALKAAYRE IAFFVRPMQR IEKMVGHCTY FPEERRAPKS APTAEKFIAI SKFFSTVIID NEGWEQKIIE RKTLEELLDF AVSREKVEFR HLRKFLDLSD NEIFKGLHYK GKPKTAKKRE ATLFDPNEPT ELEFDKVEAE KKAWISLRGA AKLREALGNE FYGRFVALGK HADEATKILT YYKDEGQKRR ELTKLPLEAE MVERLVKIGF SDFLKLSLKA IRDILPAMES GARYDEAVLM LGVPHKEKSA ILPPLNKTDI DILNPTVIRA FAQFRKVANA LVRKYGAFDR VHFELAREIN TKGEIEDIKE SQRKNEKERK EAADWIAETS FQVPLTRKNI LKKRLYIQQD GRCAYTGDVI ELERLFDEGY CEIDHILPRS RSADDSFANK VLCLARANQQ KTDRTPYEWF GHDAARWNAF ETRTSAPSNR	(SEQ ID NO: 59)

Table 4- Cas9 orthologs		
	VRTGKGKIDR LLKKNFDENS EMAFKDRNLN DTRYMARAIK TYCEQYWVFK NSHTKAPVQV RSGKLTSVLR YQWGLESKDR ESHTHHAVDA IIIAFSTQGM VQKLSEYYRF KETHREKERP KLAVPLANFR DAVEEATRIE NTETVKEGVE VKRLLISRPP RARVTGQAHE QTAKPYPRIK QVKNKKKWRL APIDEEKFES FKADRVASAN QKNFYETSTI PRVDVYHKKG KFHLVPIYLH EMVLNELPNL SLGTNPEAMD ENFFKFSIFK DDLISIQTQG TPKKPAKIIM GYFKNMHGAN MVLSSINNSP CEGFTCTPVS MDKKHKDKCK LCPEENRIAG RCLQGFLDYW SQEGLRPPRK EFECDQGVKF ALDVKKYQID PLGYYYEVKQ EKRLGTIPQM RSAKKLVKK	
Parasutterella excrementihominis YIT 11859 WP_008864843.1	MGKTHIIGVG LDLGGTYTGT FITSHPSDEA EHRDHSSAFT VVNSEKLSFS SKSRTAVRHR VRSYKGFDLR RRLLLLVAEY QLLQKKQTLA PEERENLRIA LSGYLKRRGY ARTEAETDTS VLESLDPSVF SSAPSFTNFF NDSEPLNIQW EAIANSPETT KALNKELSGQ KEADFKKYIK TSFPEYSAKE ILANYVEGRR AILDASKYIA NLQSLGHKHR SKYLSDILQD MKRDSRITRL SEAFGSTDNL WRIIGNISNL QERAVRWYFN DAKFEQGQEQ LDAVKLKNVL VRALKYLRSD DKEWSASQKQ IIQSLEQSGD VLDVLAGLDP DRTIPPYEDQ NNRRPPEDQT LYLNPKALSS EYGEKWKSWA NKFAGAYPLL TEDLTEILKN TDRKSRIKIR SDVLPDSDYR LAYILQRAFD RSIALDECSI RRTAEDFENG VVIKNEKLED VLSGHQLEEF LEFANRYYQE TAKAKNGLWF PENALLERAD LHPPMKNKIL NVIVGQALGV SPAEGTDFIE EIWNSKVKGR STVRSICNAI ENERKTYGPY FSEDYKFVKT ALKEGKTEKE LSKKFAAVIK VLKMVSEVVP FIGKELRLSD EAQSKFDNLY SLAQLYNLIE TERNGFSKVS LAAHLENAWR MTMTDGSAQC CRLPADCVRP FDGFIRKAID RNSWEVAKRI AEEVKKSVDF TNGTVKIPVA IEANSFNFTA SLTDLKYIQL KEQKLKKKLE DIQRNEENQE KRWLSKEERI RADSHGICAY TGRPLDDVGE IDHIIPRSLT LKKSESIYNS EVNLIFVSAQ GNQEKKNNIY LLSNLAKNYL AAVFGTSDLS QITNEIESTV LQLKAAGRLG YFDLLSEKER ACARHALFLN SDSEARRAVI DVLGSRRKAS VNGTQAWFVR SIFSKVRQAL AAWTQETGNE LIFDAISVPA ADSSEMRKRF AEYRPEFRKP KVQPVASHSI DAMCIYLAAC SDPFKTKRMG SQLAIYEPIN FDNLFTGSCQ VIQNTPRNFS DKTNIANSPI FKETIYAERF LDIIVSRGEI FIGYPSNMPF EEKPNRISIG GKDPFSILSV LGAYLDKAPS SEKEKLTIYR VVKNKAFELF SKVAGSKFTA EEDKAAKILE ALHFVTVKQD VAATVSDLIK SKKELSKDSI ENLAKQKGCL KKVEYSSKEF KFKGSLIIPA AVEWGKVLWN VFKENTAEEL KDENALRKAL EAAWPSSFGT RNLHSKAKRV FSLPVVATQS GAVRIRRKTA FGDFVYQSQD TNNLYSSFPV KNGKLDWSSP IIHPALQNRN LTAYGYRFVD HDRSISMSEF REVYNKDDLM RIELAQGTSS RRYLRVEMPG EKFLAWFGEN SISLGSSFKF SVSEVFDNKI YTENAEFTKF LPKPREDNKH NGTIFFELVG PRVIFNYIVG GAASSLKEIF SEAGKERS	(SEQ ID NO: 60)
Streptococcus sanguinis SK49 WP_002933589.1	MTKFNKNYSI GLDIGVSSVG YAVVTEDYRV PAFKFKVLGN TEKEKIKKNL IGSTTFVSAQ PAKGTRVFRV NRRRIDRRNH RITYLRDIFQ KEIEKVDKNF YRRLDESFRV LGDKSEDLQI KQPFFGDKEL ETAYHKKYPT IYHLRKHLAD ADKNSPVADI REVYMAISHI LKYRGHFLTL DKINPNNINM QNSWIDFIES CQEVFDLEIS DESKNIADIF KSSENRQEKV KKILPYFQQE LLKKDKSIFK QLLQLLFGLK TKFKDCFELE EEPDLNFSKE NYDENLENFL GSLEEDFSDV FAKLKVLRDT ILLSGMLTYT GATHARFSAT MVERYEEHRK DLQRFKFFIK QNLSEQDYLD IFGRKTQNGF DVDKETKGYV GYITNKMVLT NPQKQKTIQQ NFYDYISGKI TGIEGAEYFL NKISDGTFLR KLRTSDNGAI PNQIHAYELE KIIERQGKDY PFLLENKDKL LSILTFKIPY YVGPLAKGSN SRFAWIKRAT SSDILDDNDE DTRNGKIRPW NYQKLINMDE TRDAFITNLI GNDIILLNEK VLPKRSLIYE EVMLQNELTR VKYKDKYGKA HFFDSELRQN IINGLFKNNS KRVNAKSLIK YLSDNHKDLN AIEIVSGVEK GKSFNSTLKT YNDLKTIFSE ELLDSEIYQK ELEEIIKVIT VFDDKKSIKN YLTKFFGHLE ILDEEKINQL SKLRYSGWGR YSAKLLLDIR DEDTGFNLLQ FLRNDEENRN LTKLISDNTL SFEPKIKDIQ SKSTIEDDIF DEIKKLAGSP AIKRGILNSI KIVDELVQII GYPPHNIVIE MARENMTTEE GQKKAKTRKT KLESALKNIE NSLLENGKVP HSDEQLQSEK LYLYYLQNGK DMYTLDKTGS PAPLYLDQLD QYEVDHIIPY SFLPIDSIDN KVLTHRENNQ QKLNNIPDKE TVANMKPFWE KLYNAKLISQ TKYQRLTTSE RTPDGVLTES MKAGFIERQL VETRQIIKHV ARILDNRFSD	(SEQ ID NO: 61)

Table 4- Cas9 orthologs		
	TKIITLKSQL ITNFRNTFHI AKIRELNDYH HAHDAYLAVV VGQTLLKVYP KLAPELIYGH HAHFNRHEEN KATLRKHLYS NIMRFFNNPD SKVSKDIWDC NRDLPIIKDV IYNSQINFVK RTMIKKGAFY NQNPVGKFNK QLAANNRYPL KTKALCLDTS IYGGYGPMNS ALSIIIIAER FNEKKGKIET VKEFHDIFII DYEKFNNNPF QFLNDTSENG FLKKNNINRV LGFYRIPKYS LMQKIDGTRM LFESKSNLHK ATQFKLTKTQ NELFFHMKRL LTKSNLMDLK SKSAIKESQN FILKHKEEFD NISNQLSAFS QKMLGNTTSL KNLIKGYNER KIKEIDIRDE TIKYFYDNFI KMFSFVKSGA PKDINDFFDN KCTVARMRPK PDKKLLNATL IHQSITGLYE TRIDLSKLGE D	
Actinomyces sp. oral taxon 180 str. F0310 AOL41039.1	MLHCIAVIRV PPSEEPGFFE THADSCALCH HGCMTYAAND KAIRYRVGID VGLRSIGFCA VEVDDEDHPI RILNSVVHVH DAGTGGPGET ESLRKRSGVA ARARRRGRAE KQRLKKLDVL LEELGWGVSS NELLDSHAPW HIRKRLVSEY IEDETERRQC LSVAMAHIAR HRGWRNSFSK VDTLLLEQAP SDRMQGLKER VEDRTGLQFS EEVTQGELVA TLLEHDGDVT IRGFVRKGGK ATKVHGVLEG KYMQSDLVAE LRQICRTQRV SETTFEKLVL SIFHSKEPAP SAARQRERVG LDELQLALDP AAKQPRAERA HPAFQKFKVV ATLANMRIRE QSAGERSLTS EELNRVARYL LNHTESESPT WDDVARKLEV PRHRLRGSSR ASLETGGGLT YPPVDDTTVR VMSAEVDWLA DWWDCANDES RGHMIDAISN GCGSEPDDVE DEEVNELISS ATAEDMLKLE LLAKKLPSGR VAYSLKTLRE VTAAILETGD DLSQAITRLY GVDPGWVPTP APIEAPVGNP SVDRVLKQVA RWLKFASKRW GVPQTVNIEH TREGLKSASL LEEERERWER FEARREIRQK EMYKRLGISG PFRRSDQVRY EILDLQDCAC LYCGNEINFQ TFEVDHIIPR VDASSDSRRT NLAAVCHSCN SAKGGLAFGQ WVKRGDCPSG VSLENAIKRV RSWSKDRLGL TEKAMGKRKS EVISRLKTEM PYEEFDGRSM ESVAWMAIEL KKRIEGYFNS DRPEGCAAVQ VNAYSGRLTA CARRAAHVDK RVRLIRLKGD DGHHKNRFDR RNHAMDALVI ALMTPAIART IAVREDRREA QQLTRAFESW KNFLGSEERM QDRWESWIGD VEYACDRLNE LIDADKIPVT ENLRLRNSGK LHADQPESLK KARRGSKRPR PQRYVLGDAL PADVINRVTD PGLWTALVRA PGFDSQLGLP ADLNRGLKLR GKRISADFPI DYFPTDSPAL AVQGGYVGLE FHHARLYRII GPKEKVKYAL LRVCAIDLCG IDCDDLFEVE LKPSSISMRT ADAKLKEAMG NGSAKQIGWL VLGDEIQIDP TKFPKQSIGK FLKECGPVSS WRVSALDTPS KITLKPRLLS NEPLLKTSRV GGHESDLVVA ECVEKIMKKT GWVVEINALC QSGLIRVIRR NALGEVRTSP KSGLPISLNL R	(SEQ ID NO: 62)
Rhodovulum sp. PH10 WP_008386983.1	MGIRFAFDLG TNSIGWAVWR TGPGVFGEDT AASLDGSGVL IFKDGRNPKD GQSLATMRRV PRQSRKRRDR FVLRRRDLLA ALRKAGLFPV DVEEGRRLAA TDPYHLRAKA LDESLTPHEM GRVIFHLNQR RGFRSNRKAD RQDREKGKIA EGSKRLAETL AATNCRTLGE FLWSRHRGTP RTRSPTRIRM EGEGAKALYA FYPTREMVRA EFERLWTAQS RFAPDLLTPE RHEEIAGILF RQRDLAPPKI GCCTFEPSER RLPRALPSVE ARGIYERLAH LRITTGPVSD RGLTRPERDV LASALLAGKS LTFKAVRKTL KILPHALVNF EEAGEKGLDG ALTAKLLSKP DHYGAAWHGL SFAEKDTFVG KLLDEADEER LIRRLVTENR LSEDAARRCA SIPLADGYGR LGRTANTEIL AALVEETDET GTVVTYAEAV RRAGERTGRN WHHSDERDGV ILDRLPYYGE ILQRHVVPGS GEPEEKNEAA RWGRLANPTV HIGLNQLRKV VNRLIAAHGR PDQIVVELAR ELKLNREQKE RLDRENRKNR EENERRTAIL AEHGQRDTAE NKIRLRLFEE QARANAGIAL CPYTGRAIGI AELFTSEVEI DHILPVSLTL DDSLANRVLC RREANREKRR QTPFQAFGAT PAWNDIVARA AKLPPNKRWR FDPAALERFE REGGFLGRQL NETKYLSRLA KIYLGKICDP DRVYVTPGTL TGLLRARWGL NSILSDSNFK NRSDHRHHAV DAVVIGVLTR GMIQRIAHDA ARAEDQDLDR VFRDVPVPFE DFRDHVRERV STITVAVKPE HGKGGALHED TSYGLVPDTD PNAALGNLVV RKPIRSLTAG EVDRVRDRAL RARLGALAAP FRDESGRVRD AKGLAQALEA FGAENGIRRV RILKPDASVV TIADRRTGVP YRAVAPGENH HVDIVQMRDG SWRGFAASVF EVNRPGWRPE WEVKKLGGKL VMRLHKGDMV ELSDKDGQRR VKVVQQIEIS ANRVRLSPHN DGGKLQDRHA DADDPFRWDL ATIPLLKDRG CVAVRVDPIG VVTLRRSNV	(SEQ ID NO: 63)

Table 4- Cas9 orthologs		
Bifidobacterium bifidum S17 WP_013362995.1	MSRKNYVDDY AISLDIGNAS VGWSAFTPNY RLVRAKGHEL IGVRLFDPAD TAESRRMART TRRRYSRRRW RLRLLDALFD QALSEIDPSF LARRKYSWVH PDDENNADCW YGSVLFDSNE QDKRFYEKYP TIYHLRKALM EDDSQHDIRE IYLAIHHMVK YRGNFLVEGT LESSNAFKED ELLKLLGRIT RYEMSEGEQN SDIEQDDENK LVAPANGQLA DALCATRGSR SMRVDNALEA LSAVNDLSRE QRAIVKAIFA GLEGNKLDLA KIFVSKEFSS ENKKILGIYF NKSDYEEKCV QIVDSGLLDD EEREFLDRMQ GQYNAIALKQ LLGRSTSVSD SKCASYDAHR ANWNLIKLQL RTKENEKDIN ENYGILVGWK IDSGQRKSVR GESAYENMRK KANVFFKKMI ETSDLSETDK NRLIHDIEED KLFPIQRDSD NGVIPHQLHQ NELKQIIKKQ GKYYPFLLDA FEKDGKQINK IEGLLTFRVP YFVGPLVVPE DLQKSDNSEN HWMVRKKKGE ITPWNFDEMV DKDASGRKFI ERLVGTDSYL LGEPTLPKNS LLYQEYEVLN ELNNVRLSVR TGNHWNDKRR MRLGREEKTL LCQRLFMKGQ TVTKRTAENL LRKEYGRTYE LSGLSDESKF TSSLSTYGKM CRIFGEKYVN EHRDLMEKIV ELQTVFEDKE TLLHQLRQLE GISEADCALL VNTHYTGWGR LSRKLLTTKA GECKISDDFA PRKHSIIEIM RAEDRNLMEI ITDKQLGFSD WIEQENLGAE NGSSLMEVVD DLRVSPKVKR GIIQSIRLID DISKAVGKRP SRIFLELADD IQPSGRTISR KSRLQDLYRN ANLGKEFKGI ADELNACSDK DLQDDRLFLY YTQLGKDMYT GEELDLDRLS SAYDIDHIIP QAVTQNDSID NRVLVARAEN ARKTDSFTYM PQIADRMRNF WQILLDNGLI SRVKFERLTR QNEFSEREKE RFVQRSLVET RQIMKNVATL MRQRYGNSAA VIGLNAELTK EMHRYLGFSH KNRDINDYHH AQDALCVGIA GQFAANRGFF ADGEVSDGAQ NSYNQYLRDY LRGYREKLSA EDRKQGRAFG FIVGSMRSQD EQKRVNPRTG EVVWSEEDKD YLRKVMNYRK MLVTQKVGDD FGALYDETRY AATDPKGIKG IPFDGAKQDT SLYGGFSSAK PAYAVLIESK GKTRLVNVTM QEYSLLGDRP SDDELRKVLA KKKSEYAKAN ILLRHVPKMQ LIRYGGGLMV IKSAGELNNA QQLWLPYEEY CYFDDLSQGK GSLEKDDLKK LLDSILGSVQ CLYPWHRFTE EELADLHVAF DKLPEDEKKN VITGIVSALH ADAKTANLSI VGMTGSWRRM NNKSGYTFSD EDEFIFQSPS GLFEKRVTVG ELKRKAKKEV NSKYRTNEKR LPTLSGASQP	(SEQ ID NO: 64)
Barnesiella intestinihominis YIT 11860 WP_008863245.1	MKNILGLDLG LSSIGWSVIR ENSEEQELVA MGSRVVSLTA AELSSFTQGN 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 FDFLNTHEKL KGSDLLKLLG LSKTYGYRLG EQFKTGIQGN KTRVEIERAL GNYPDKKRLL QFNLQEESSS MVNTETGEII PMISLSFEQE PLYRLWHVLY SIDDREQLQS VLRQKFGIDD DEVLERLSAI DLVKAGFGNK SSKAIRRILP FLQLGMNYAE ACEAAGYNHS NNYTKAENEA RALLDRLPAI KKNELRQPVV EKILNQMVNV VNALMEKYGR FDEIRVELAR ELKQSKEERS NTYKSINKNQ RENEQIAKRI VEYGVPTRSR IQKYKMWEES KHCCIYCGQP VDVGDFLRGF DVEVEHIIPK SLYFDDSFAN KVCSCRSCNK EKNNRTAYDY MKSKGEKALS DYVERVNTMY TNNQISKTKW QNLLTPVDKI SIDFIDRQLR ESQYIARKAK EILTSICYNV TATSGSVTSF LRHVWGWDTV LHDLNFDRYK KVGLTEVIEV NHRGSVIRRE QIKDWSKRFD HRHHAIDALT IACTKQAYIQ RLNNLRAEEG PDFNKMSLER YIQSQPHFSV AQVREAVDRI LVSFRAGKRA VTPGKRYIRK NRKRISVQSV LIPRGALSEE SVYGVIHVWE KDEQGHVIQK QRAVMKYPIT SINREMLDKE KVVDKRIHRI LSGRLAQYND NPKEAFAKPV YIDKECRIPI RTVRCFAKPA INTLVPLKKD DKGNPVAWVN PGNNHHVAIY RDEDGKYKER TVTFWEAVDR CRVGIPAIVT QPDTIWDNIL QRNDISENVL ESLPDVKWQF VLSLQQNEMF ILGMNEEDYR YAMDQQDYAL LNKYLYRVQK LSKSDYSFRY HTETSVEDKY DGKPNLKLSM QMGKLKRVSI KSLLGLNPHK VHISVLGEIK EIS	(SEQ ID NO: 65)

Table 4- Cas9 orthologs		
Aminomonas paucivorans DSM 12260 WP_006299850.1	MIGEHVRGGC LFDDHWTPNW GAFRLPNTVR TFTKAENPKD GSSLAEPRRQ ARGLRRRLRR KTQRLEDLRR LLAKEGVLSL SDLETLFRET PAKDPYQLRA EGLDRPLSFP EWVRVLYHIT KHRGFQSNRR NPVEDGQERS RQEEEGKLLS GVGENERLLR EGGYRTAGEM LARDPKFQDH RRNRAGDYSH TLSRSLLLEE ARRLFQSQRT LGNPHASSNL EEAFLHLVAF QNPFASGEDI RNKAGHCSLE PDQIRAPRRS ASAETFMLLQ KTGNLRLIHR RTGEERPLTD KEREQIHLLA WKQEKVTHKT LRRHLEIPEE WLFTGLPYHR SGDKAEEKLF VHLAGIHEIR KALDKGPDPA VWDTLRSRRD LLDSIADTLT FYKNEDEILP RLESLGLSPE NARALAPLSF SGTAHLSLSA LGKLLPHLEE GKSYTQARAD AGYAAPPPDR HPKLPPLEEA DWRNPVVFRA LTQTRKVVNA LVRRYGPPWC IHLETARELS QPAKVRRRIE TEQQANEKKK QQAEREFLDI VGTAPGPGDL LKMRLWREQG GFCPYCEEYL NPTRLAEPGY AEMDHILPYS RSLDNGWHNR VLVHGKDNRD KGNRTPFEAF GGDTARWDRL VAWVQASHLS APKKRNLLRE DFGEEAEREL KDRNLTDTRF ITKTAATLLR DRLTFHPEAP KDPVMTLNGR LTAFLRKQWG LHKNRKNGDL HHALDAAVLA VASRSFVYRL SSHNAAWGEL PRGREAENGF SLPYPAFRSE VLARLCPTRE EILLRLDQGG VGYDEAFRNG LRPVFVSRAP SRRLRGKAHM ETLRSPKWKD HPEGPRTASR IPLKDLNLEK LERMVGKDRD RKLYEALRER LAAFGGNGKK AFVAPFRKPC RSGEGPLVRS LRIFDSGYSG VELRDGGEVY AVADHESMVR VDVYAKKNRF YLVPVYVADV ARGIVKNRAI VAHKSEEEWD LVDGSFDFRF SLFPGDLVEI EKKDGAYLGY YKSCHRGDGR LLLDRHDRMP RESDCGTFYV STRKDVLSMS KYQVDPLGEI RLVGSEKPPF VL	(SEQ ID NO: 66)
Ralstonia syzygii R24 CCA84553.1	MAEKQHRWGL DIGTNSIGWA VIALIEGRPA GLVATGSRIF SDGRNPKDGS SLAVERRGPR QMRRRRDRYL RRRDRFMQAL INVGLMPGDA AARKALVTEN PYVLRQRGLD QALTLPEFGR ALFHLNQRRG FQSNRKTDRA TAKESGKVKN AIAAFRAGMG NARTVGEALA RRLEDGRPVR ARMVGQGKDE HYELYIAREW IAQEFDALWA SQQRFHAEVL ADAARDRLRA ILLFQRKLLP VPVGKCFLEP NQPRVAAALP SAQRFRLMQE LNHLRVMTLA DKRERPLSFQ ERNDLLAQLV ARPKCGFDML RKIVFGANKE AYRFTIESER RKELKGCDTA AKLAKVNALG TRWQALSLDE QDRLVCLLLD GENDAVLADA LREHYGLTDA QIDTLLGLSF EDGHMRLGRS ALLRVLDALE SGRDEQGLPL SYDKAVVAAG YPAHTADLEN GERDALPYYG ELLWRYTQDA PTAKNDAERK FGKIANPTVH IGLNQLRKLV NALIQRYGKP AQIVVELARN LKAGLEEKER IKKQQTANLE RNERIRQKLQ DAGVPDNREN RLRMRLFEEL GQGNGLGTPC IYSGRQISLQ RLFSNDVQVD HILPFSKTLD DSFANKVLAQ HDANRYKGNR GPFEAFGANR DGYAWDDIRA RAAVLPRNKR NRFAETAMQD WLHNETDFLA RQLTDTAYLS RVARQYLTAI CSKDDVYVSP GRLTAMLRAK WGLNRVLDGV MEEQGRPAVK NRDDHRHHAI DAVVIGATDR AMLQQVATLA ARAREQDAER LIGDMPTPWP NFLEDVRAAV ARCVVSHKPD HGPEGGLHND TAYGIVAGPF EDGRYRVRHR VSLFDLKPGD LSNVRCDAPL QAELEPIFEQ DDARAREVAL TALAERYRQR KVWLEELMSV LPIRPRGEDG KTLPDSAPYK AYKGDSNYCY ELFINERGRW DGELISTFRA NQAAYRRFRN DPARFRRYTA GGRPLLMRLC INDYIAVGTA AERTIFRVVK MSENKITLAE HFEGGTLKQR DADKDDPFKY LTKSPGALRD LGARRIFVDL IGRVLDPGIK GD	(SEQ ID NO: 67)
Catenibacterium mitsuokai DSM 15897 WP_006506696.1	IVDYCIGLDL GTGSVGWAVV DMNHRLMKRN GKHLWGSRLF SNAETAANRR ASRSIRRRYN KRRERIRLLR AILQDMVLEK DPTFFIRLEH TSFLDEEDKA KYLGTDYKDN YNLFIDEDFN DYTYYHKYPT IYHLRKALCE STEKADPRLI YLALHHIVKY RGNFLYEGQK FNMDASNIED KLSDIFTQFT SFNNIPYEDD EKKNLEILEI LKKPLSKKAK VDEVMTLIAP EKDYKSAFKE LVTGIAGNKM NVTKMILCEP IKQGDSEIKL KFSDSNYDDQ FSEVEKDLGE YVEFVDALHN VYSWVELQTI MGATHTDNAS ISEAMVSRYN KHHDDLKLLK DCIKNNVPNK YFDMFRNDSE KSKGYYNYIN RPSKAPVDEF YKYVKKCIEK VDTPEAKQIL NDIELENFLL KQNSRTNGSV PYQMQLDEMI KIIDNQAEYY PILKEKREQL LSILTFRIPY YFGPLNETSE HAWIKRLEGK ENQRILPWNY QDIVDVDATA EGFIKRMRSY CTYFPDEEVL PKNSLIVSKY EVYNELNKIR VDDKLLEVDV KNDIYNELFM KNKTVTEKKL KNWLVNNQCC SKDAEIKGFQ KENQFSTSLT PWIDFTNIFG KIDQSNFDLI ENIIYDLTVF EDKKIMKRRL KKKYALPDDK VKQILKLKYK DWSRLSKKLL DGIVADNRFG SSVTVLDVLE MSRLNLMEII	(SEQ ID NO: 68)

Table 4- Cas9 orthologs		
	NDKDLGYAQM IEEATSCPED GKFTYEEVER LAGSPALKRG IWQSLQIVEE ITKVMKCRPK YIYIEFERSE EAKERTESKI KKLENVYKDL DEQTKKEYKS VLEELKGFDN TKKISSDSLF LYFTQLGKCM YSGKKLDIDS LDKYQIDHIV PQSLVKDDSF DNRVLVVPSE NQRKLDDLVV PFDIRDKMYR FWKLLFDHEL ISPKKFYSLI KTEYTERDEE RFINRQLVET RQITKNVTQI IEDHYSTTKV AAIRANLSHE FRVKNHIYKN RDINDYHHAH DAYIVALIGG FMRDRYPNMH DSKAVYSEYM KMFRKNKNDQ KRWKDGFVIN SMNYPYEVDG KLIWNPDLIN EIKKCFYYKD CYCTTKLDQK SGQLFNLTVL SNDAHADKGV TKAVVPVNKN RSDVHKYGGF SGLQYTIVAI EGQKKKGKKT ELVKKISGVP LHLKAASINE KINYIEEKEG LSDVRIIKDN IPVNQMIEMD GGEYLLTSPT EYVNARQLVL NEKQCALIAD IYNAIYKQDY DNLDDILMIQ LYIELTNKMK VLYPAYRGIA EKFESMNENY VVISKEEKAN IIKQMLIVMH RGPQNGNIVY DDFKISDRIG RLKTKNHNLN NIVFISQSPT GIYTKKYKL	
Mycoplasma synoviae 53 AOL40776.1	MLRLYCANNL VLNNVQNLWK YLLLLIFDKK IIFLFKIKVI LIRRYMENNN KEKIVIGFDL GVASVGWSIV NAETKEVIDL GVRLFSEPEK ADYRRAKRTT RRLLRRKKFK REKFHKLILK NAEIFGLQSR NEILNVYKDQ SSKYRNILKL KINALKEEIK PSELVWILRD YLQNRGYFYK NEKLTDEFVS NSFPSKKLHE HYEKYGFFRG SVKLDNKLDN KKDKAKEKDE EEESDAKKES EELIFSNKQW INEIVKVFEN QSYLTESFKE EYLKLFNYVR PFNKGPGSKN SRTAYGVFST DIDPETNKFK DYSNIWDKTI GKCSLFEEEI RAPKNLPSAL IFNLQNEICT IKNEFTEFKN WWLNAEQKSE ILKFVFTELF NWKDKKYSDK KFNKNLQDKI KKYLLNFALE NFNLNEEILK NRDLENDTVL GLKGVKYYEK SNATADAALE FSSLKPLYVF IKFLKEKKLD LNYLLGLENT EILYFLDSIY LAISYSSDLK ERNEWFKKLL KELYPKIKNN NLEIIENVED IFEITDQEKF ESFSKTHSLS REAFNHIIPL LLSNNEGKNY ESLKHSNEEL KKRTEKAELK AQQNQKYLKD NFLKEALVPL SVKTSVLQAI KIFNQIIKNF GKKYEISQVV IEMARELTKP NLEKLLNNAT NSNIKILKEK LDQTEKFDDF TKKKFIDKIE NSVVFRNKLF LWFEQDRKDP YTQLDIKINE IEDETEIDHV IPYSKSADDS WFNKLLVKKS TNQLKKNKTV WEYYQNESDP EAKWNKFVAW AKRIYLVQKS DKESKDNSEK NSIFKNKKPN LKFKNITKKL FDPYKDLGFL ARNLNDTRYA TKVFRDQLNN YSKHHSKDDE NKLFKVVCMN GSITSFLRKS MWRKNEEQVY RFNFWKKDRD QFFHHAVDAS IIAIFSLLTK TLYNKLRVYE SYDVQRREDG VYLINKETGE VKKADKDYWK DQHNFLKIRE NAIEIKNVLN NVDFQNQVRY SRKANTKLNT QLFNETLYGV KEFENNFYKL EKVNLFSRKD LRKFILEDLN EESEKNKKNE NGSRKRILTE KYIVDEILQI LENEEFKDSK SDINALNKYM DSLPSKFSEF FSQDFINKCK KENSLILTFD AIKHNDPKKV IKIKNLKFFR EDATLKNKQA VHKDSKNQIK SFYESYKCVG FIWLKNKNDL EESIFVPINS RVIHFGDKDK DIFDFDSYNK EKLLNEINLK RPENKKFNSI NEIEFVKFVK PGALLLNFEN QQIYYISTLE SSSLRAKIKLLNKMDKGKAVS MKKITNPDEY KIIEHVNPL GINLNWTKKL ENNN	(SEQ ID NO: 69)
Flavobacterium branchiophilum FL-15 WP_014084151.1	MAKILGLDLG TNSIGWAVVE RENIDFSLID KGVRIFSEGV KSEKGIESSR AAERTGYRSA RKIKYRRKLR KYETLKVLSL NRMCPLSIEE VEEWKKSGFK DYPLNPEFLK WLSTDEESNV NPYFFRDRAS KHKVSLFELG RAFYHIAQRR GFLSNRLDQS AEGILEEHCP KIEAIVEDLI SIDEISTNIT DYFFETGILD SNEKNGYAKD LDEGDKKLVS LYKSLLAILK KNESDFENCK SEIIERLNKK DVLGKVKGKI KDISQAMLDG NYKTLGQYFY SLYSKEKIRN QYTSREEHYL SEFITICKVQ GIDQINEEEK INEKKFDGLA KDLYKAIFFQ RPLKSQKGLI GKCSFEKSKS RCAISHPDFE EYRMWTYLNT IKIGTQSDKK LRFLTQDEKL KLVPKFYRKN DFNFDVLAKE LIEKGSSFGF YKSSKKNDFF YWFNYKPTDT VAACQVAASL KNAIGEDWKT KSFKYQTINS NKEQVSRTVD YKDLWHLLTV ATSDVYLYEF AIDKLGLDEK NAKAFSKTKL KKDFASLSLS AINKILPYLK EGLLYSHAVF VANIENIVDE NIWKDEKQRD YIKTQISEII ENYTLEKSRF EIINGLLKEY KSENEDGKRV YYSKEAEQSF ENDLKKKLVL FYKSNEIENK EQQETIFNEL LPIFIQQLKD YEFIKIQRLD QKVLIFLKGK NETGQIFCTE EKGTAEEKEK KIKNRLKKLY HPSDIEKFKK KIIKDEFGNE KIVLGSPLTP SIKNPMAMRA LHQLRKVLNA LILEGQIDEK TIIHIEMARE LNDANKRKGI QDYQNDNKKF REDAIKEIKK LYFEDCKKEV EPTEDDILRY QLWMEQNRSE IYEEGKNISI CDIIGSNPAY DIEHTIPRSR SQDNSQMNKT LCSQRFNREV	(SEQ ID NO: 70)

Table 4- Cas9 orthologs		
	KKQSMPIELN NHLEILPRIA HWKEEADNLT REIEIISRSI KAAATKEIKD KKIRRRHYLT LKRDYLQGKY DRFIWEEPKV GFKNSQIPDT GIITKYAQAY LKSYFKKVES VKGGMVAEFR KIWGIQESFI DENGMKHYKV KDRSKHTHHT IDAITIACMT KEKYDVLAHA WTLEDQQNKK EARSIIEASK PWKTFKEDLL KIEEEILVSH YTPDNVKKQA KKIVRVRGKK QFVAEVERDV NGKAVPKKAA SGKTIYKLDG EGKKLPRLQQ GDTIRGSLHQ DSIYGAIKNP LNTDEIKYVI RKDLESIKGS DVESIVDEVV KEKIKEAIAN KVLLLSSNAQ QKNKLVGTVW MNEEKRIAIN KVRIYANSVK NPLHIKEHSL LSKSKHVHKQ KVYGQNDENY AMAIYELDGK RDFELINIFN LAKLIKQGQG FYPLHKKKEI KGKIVFVPIE KRNKRDVVLK RGQQVVFYDK EVENPKDISE IVDFKGRIYI IEGLSIQRIV RPSGKVDEYG VIMLRYFKEA RKADDIKQDN FKPDGVFKLG ENKPTRKMNH NQFTAFVEGI DFKVLPSGKF EKI	
Eubacterium yurii subsp. margaretiae ATCC 43715 EFM38267.1	MENKQYYIGL DVGTNSVGWA VTDTSYNLLR AKGKDMWGAR LFEKANTAAE RRTKRTSRRR SEREKARKAM LKELFADEIN RVDPSFFIRL EESKFFLDDR SENNRQRYTL FNDATFTDKD YYEKYKTIFH LRSALINSDE KFDVRLVFLA ILNLFSHRGH FLNASLKGDG DIQGMDVFYN DLVESCEYFE IELPRITNID NFEKILSQKG KSRTKILEEL SEELSISKKD KSKYNLIKLI SGLEASVVEL YNIEDIQDEN KKIKIGFRES DYEESSLKVK EIIGDEYFDL VERAKSVHDM GLLSNIIGNS KYLCEARVEA YENHHKDLLK IKELLKKYDK KAYNDMFRKM TDKNYSAYVG SVNSNIAKER RSVDKRKIED LYKYIEDTAL KNIPDDNKDK IEILEKIKLG EFLKKQLTAS NGVIPNQLQS RELRAILKKA ENYLPFLKEK GEKNLTVSEM IIQLFEFQIP YYVGPLDKNP KKDNKANSWA KIKQGGRILP WNFEDKVDVK GSRKEFIEKM VRKCTYISDE HTLPKQSLLY EKFMVLNEIN NIKIDGEKIS VEAKQKIYND LFVKGKKVSQ KDIKKELISL NIMDKDSVLS GTDTVCNAYL SSIGKFTGVF KEEINKQSIV DMIEDIIFLK TVYGDEKRFV KEEIVEKYGD EIDKDKIKRI LGFKFSNWGN LSKSFLELEG ADVGTGEVRS IIQSLWETNF NLMELLSSRF TYMDELEKRV KKLEKPLSEW TIEDLDDMYL SSPVKRMIWQ SMKIVDEIQT VIGYAPKRIF VEMTRSEGEK VRTKSRKDRL KELYNGIKED SKQWVKELDS KDESYFRSKK MYLYYLQKGR CMYSGEVIEL DKLMDDNLYD IDHIYPRSFV KDDSLDNLVL VKKEINNRKQ NDPITPQIQA SCQGFWKILH DQGFMSNEKY SRLTRKTQEF SDEEKLSFIN RQIVETGQAT KCMAQILQKS MGEDVDVVFS KARLVSEFRH KFELFKSRLI NDFHHANDAY LNIVVGNSYF VKFTRNPANF IKDARKNPDN PVYKYHMDRF FERDVKSKSE VAWIGQSEGN SGTIVIVKKT MAKNSPLITK KVEEGHGSIT KETIVGVKEI KFGRNKVEKA DKTPKKPNLQ AYRPIKTSDE RLCNILRYGG RTSISISGYC LVEYVKKRKT IRSLEAIPVY LGRKDSLSEE KLLNYFRYNL NDGGKDSVSD IRLCLPFIST NSLVKIDGYL YYLGGKNDDR IQLYNAYQLK MKKEEVEYIR KIEKAVSMSK FDEIDREKNP VLTEEKNIEL YNKIQDKFEN TVFSKRMSLV KYNKKDLSFG DFLKNKKSKF EEIDLEKQCK VLYNIIFNLS NLKEVDLSDI GGSKSTGKCR CKKNITNYKE FKLIQQSITG LYSCEKDLMT I	(SEQ ID NO: 71)
Acidovorax ebreus WP_012655176.1	MAQHVFGLDI GIASVGWAIL GEQRIIDLGV RCFDKAETAK EGDPLNLTRR QARLLRRRLY RRAWRLTQLS RLLKRKGLIA DAKLFAKAPS YGDSAWELRR QGLDRLLTPL EWARVIYHQC KHRGFHWTSK AEEAKADSDA EGGRVKQGLA HTKALMQAKN YRSAAEMVLA EFPDAQRNKR GQYDKALSRV LLGEELALLF ATQRRLGNPH ASDFFEKLIL GDGDRKSGLF WQQKPALSGA DLLKMLGKCT FEKGEYRAPK ASFSVERHVW LTRLNNLRIV VDGRSRPLNE AERQAALLLP YQTETSKYKT LKNAFIKAGL WGDGVRFGGL AYPSQAQIDA EKTKDPEDQF LVKLPAWHEL RKAFKAAGHE ALWQQISTPA LDGDPTLLDQ IATVLSVYKD GAEVVQQLRQ LALPEPAASI AVLEKISFDK FSSLSLKALR RIVPLMQSGL RYDEAVAQIP EYGHHSQRIE PGAAKHLYLP PFYEAQRKYA GKGDHIGSMQ FRDDADIPRN PVVLRALNQA RKVVNALIRE YGSPIAVNIE MARDLSRPLD ERNKVKRAQE EFRDRNDRAR SEFERDFGYK PKAAAFEKWM LYREQLGQCA YSQQPLDIQR VLDDHNYAQV DHALPYSRSY DDSKNNKVLV LTHENQNKGN RTAFEYLTSF PDGEDGERWR TFVAWVQGNK AYRMAKRNRL LRKNYGVDES KGFIDRNLND TRYICKFFKN YVEEHLQLAA RADGDTARRC VVVNGQLTAF LRARWGLTKV RGDSDRHHAL DAAVVAACTH GMVKALADYS RRKEISFLQE GFPDPETGEI LNPAAFDRAR QHFPEPWTHF AHELKARLFT DDLAALREDM QRLGSYTTED LGRLRTLFVS RAPQRRSGGA VHKETIYAQP ESLKQQGGVI	(SEQ ID NO: 72)

Table 4- Cas9 orthologs		
	EKILLTSLKL QDFDKLLNPE SNDHFVEPHR NERLYAAIRQ RLEQFGGRAD KAFGPDNLFH KPDKNNQPTG PVVRSIKLVR GKQTGIPIRG GLAKNDSMLR VDIFTKAGKF HLVPVYVHHR VTGLPNRAIV AFKDEDEWTL IDESFAFLFS VYPNDYVKVT LKKEQQSGYY SGADRSTGAM NLWAHDRAAS VGKDGLIRGI GVKTALSVEK FNVDVLGRIY LAPPETRSGL A	
Porphyromonas sp. oral taxon 279 str. F0450 WP_009433518.1	MLMSKHVLGL DLGVGSIGWC LIALDAQGDP AEILGMGSRV VPLNNATKAI EAFNAGAAFT ASQERTARRT MRRGFARYQL RRYRLRRELE KVGMLPDAAL IQLPLLELWE LRERAATAGR RLTLPELGRV LCHINQKRGY RHVKSDAAAI VGDEGEKKKD SNSAYLAGIR ANDEKLQAEH KTVGQYFAEQ LRQNQSESPT GGISYRIKDQ IFSRQCYIDE YDQIMAVQRV HYPDILTDEF IRMLRDEVIF MQRPLKSCKH LVSLCEFEKQ ERVMRVQQDD GKGGWQLVER RVKFGPKVAP KSSPLFQLCC IYEAVNNIRL TRPNGSPCDI TPEERAKIVA HLQSSASLSF AALKKLLKEK ALIADQLTSK SGLKGNSTRV ALASALQPYP QYHHLLDMEL ETRMMTVQLT DEETGEVTER EVAVVTDSYV RKPLYRLWHI LYSIEEREAM RRALITQLGM KEEDLDGGLL DQLYRLDFVK PGYGNKSAKF ICKLLPQLQQ GLGYSEACAA VGYRHSNSPT SEEITERTLL EKIPLLQRNE LRQPLVEKIL NQMINLVNAL KAEYGIDEVR VELARELKMS REERERMARN NKDREERNKG VAAKIRECGL YPTKPRIQKY MLWKEAGRQC LYCGRSIEEE QCLREGGMEV EHIIPKSVLY DDSYGNKTCA CRRCNKEKGN RTALEYIRAK GREAEYMKRI NDLLKEKKIS YSKHQRLRWL KEDIPSDFLE RQLRLTQYIS RQAMAILQQG IRRVSASEGG VTARLRSLWG YGKILHTLNL DRYDSMGETE RVSREGEATE ELHITNWSKR MDHRHHAIDA LVVACTRQSY IQRLNRLSSE FGREDKKKED QEAQEQQATE TGRLSNLERW LTQRPHFSVR TVSDKVAEIL ISYRPGQRVV TRGRNIYRKK MADGREVSCV QRGVLVPRGE LMEASFYGKI LSQGRVRIVK RYPLHDLKGE VVDPHLRELI TTYNQELKSR EKGAPIPPLC LDKDKKQEVR SVRCYAKTLS LDKAIPMCFD EKGEPTAFVK SASNHHLALY RTPKGKLVES IVTFWDAVDR ARYGIPLVIT HPREVMEQVL QRGDIPEQVL SLLPPSDWVF VDSLQQDEMV VIGLSDEELQ RALEAQNYRK ISEHLYRVQK MSSSYYVFRY HLETSVADDK NTSGRIPKFH RVQSLKAYEE RNIRKVRVDL LGRISLL	(SEQ ID NO: 73)
Mycoplasma ovipneumoniae SC01 WP_010320922.1	MHNKKNITIG FDLGIASIGW AIIDSTTSKI LDWGTRTFEE RKTANERRAF RSTRRNIRRK AYRNQRFINL ILKYKDLFEL KNISDIQRAN KKDTENYEKI ISFFTEIYKK CAAKHSNILE VKVKALDSKI EKLDLIWILH DYLENRGFFY DLEEENVADK YEGIEHPSIL LYDFFKKNGF FKSNSSIPKD LGGYSFSNLQ WVNEIKKLFE VQEINPEFSE KFLNLFTSVR DYAKGPGSEH SASEYGIFQK DEKGKVFKKY DNIWDKTIGK CSFFVEENRS PVNYPSYEIF NLLNQLINLS TDLKTTNKKI WQLSSNDRNE LLDELLKVKE KAKIISISLK KNEIKKIILK DFGFEKSDID DQDTIEGRKI IKEEPTTKLE VTKHLLATIY SHSSDSNWIN INNILEFLPY LDAICIILDR EKSRGQDEVL KKLTEKNIFE VLKIDREKQL DFVKSIFSNT KFNFKKIGNF SLKAIREFLP KMFEQNKNSE YLKWKDEEIR RKWEEQKSKL GKTDKKTKYL NPRIFQDEII SPGTKNTFEQ AVLVLNQIIK KYSKENIIDA IIIESPREKN DKKTIEEIKK RNKKGKGKTL EKLFQILNLE NKGYKLSDLE TKPAKLLDRL RFYHQQDGID LYTLDKINID QLINGSQKYE IEHIIPYSMS YDNSQANKIL TEKAENLKKG KLIASEYIKR NGDEFYNKYY EKAKELFINK YKKNKKLDSY VDLLDEDSAKN RFRFLTLQDY DEFQVEFLAR NLNDTRYSTK LFYHALVEHF ENNEFFTYID ENSSKHKVKI STIKGHVTKY FRAKPVQKNN GPNENLNNNK PEKIEKNREN NEHHAVDAAI VAIIGNKNPQ IANLLTLADN KTDKKFLLHD ENYKENIETG ELVKIPKFEV DKLAKVEDLK KIIQEKYEEA KKHTAIKFSR KTRTILNGGL SDETLYGFKY DEKEDKYFKI IKKKLVTSKN EELKKYFENP FGKKADGKSE YTVLMAQSHL SEFNKLKEIF EKYNGFSNKT GNAFVEYMND LALKEPTLKA EIESAKSVEK LLYYNFKPSD QFTYHDNINN KSFKRFYKNI RIIEYKSIPI KFKILSKHDG GKSFKDTLFS LYSLVYKVYE NGKESYKSIP VTSQMRNFGI DEFDFLDENL YNKEKLDIYK SDFAKPIPVN CKPVFVLKKG SILKKKSLDI DDFKETKETE EGNYYFISTI SKRFNRDTAY GLKPLKLSVV KPVAEPSTNP IFKEYIPIHL DELGNEYPVK IKEHTDDEKL MCTIK	(SEQ ID NO: 74)

Table 4- Cas9 orthologs		
Wolinella succinogenes WP_011139431.1	MLVSPISVDL GGKNTGFFSF TDSLDNSQSG TVIYDESFVL SQVGRRSKRH SKRNNLRNKL VKRLFLLILQ EHHGLSIDVL PDEIRGLFNK RGYTYAGFEL DEKKKDALES DTLKEFLSEK LQSIDRDSDV EDFLNQIASN AESFKDYKKG FEAVFASATH SPNKKLELKD ELKSEYGENA KELLAGLRVT KEILDEFDKQ ENQGNLPRAK YFEELGEYIA TNEKVKSFFD SNSLKLTDMT KLIGNISNYQ LKELRRYFND KEMEKGDIWI PNKLHKITER FVRSWHPKND ADRQRRAELM KDLKSKEIME LLTTTEPVMT IPPYDDMNNR GAVKCQTLRL NEEYLDKHLP NWRDIAKRLN HGKFNDDLAD STVKGYSEDS TLLHRLLDTS KEIDIYELRG KKPNELLVKT LGQSDANRLY GFAQNYYELI RQKVRAGIWV PVKNKDDSLN LEDNSNMLKR CNHNPPHKKN QIHNLVAGIL GVKLDEAKFA EFEKELWSAK VGNKKLSAYC KNIEELRKTH GNTFKIDIEE LRKKDPAELS KEEKAKLRLT DDVILNEWSQ KIANFFDIDD KHRQRFNNLF SMAQLHTVID TPRSGFSSTC KRCTAENRFR SETAFYNDET GEFHKKATAT CQRLPADTQR PFSGKIERYI DKLGYELAKI KAKELEGMEA KEIKVPIILE QNAFEYEESL RKSKTGSNDR VINSKKDRDG KKLAKAKENA EDRLKDKDKR IKAFSSGICP YCGDTIGDDG EIDHILPRSH TLKIYGTVFN PEGNLIYVHQ KCNQAKADSI YKLSDIKAGV SAQWIEEQVA NIKGYKTFSV LSAEQQKAFR YALFLQNDNE AYKKVVDWLR TDQSARVNGT QKYLAKKIQE KLTKMLPNKH LSFEFILADA TEVSELRRQY ARQNPLLAKA EKQAPSSHAI DAVMAFVARY QKVFKDGTPP NADEVAKLAM LDSWNPASNE PLTKGLSTNQ KIEKMIKSGD YGQKNMREVF GKSIFGENAI GERYKPIVVQ EGGYYIGYPA TVKKGYELKN CKVVTSKNDI AKLEKIIKNQ DLISLKENQY IKIFSINKQT ISELSNRYFN MNYKNLVERD KEIVGLLEFI VENCRYYTKK VDVKFAPKYI HETKYPFYDD WRRFDEAWRY LQENQNKTSS KDRFVIDKSS LNEYYQPDKN EYKLDVDTQP IWDDFCRWYF LDRYKTANDK KSIRIKARKT FSLLAESGVQ GKVFRAKRKI PTGYAYQALP MDNNVIAGDY ANILLEANSK TLSLVPKSGI SIEKQLDKKL DVIKKTDVRG LAIDNNSFFN ADFDTHGIRL IVENTSVKVG NFPISAIDKS AKRMIFRALF EKEKGKRKKK TTISFKESGP VQDYLKVFLK KIVKIQLRTD GSISNIVVRK NAADFTLSFR SEHIQKLLK	(SEQ ID NO: 75)
Streptococcus mutans UA159 WP_002263549.1	MKKPYSIGLD IGTNSVGWAV VTDDYKVPAK KMKVLGNTDK SHIEKNLLGA LLFDSGNTAE DRRLKRTARR RYTRRRNRIL YLQEIFSEEM GKVDDSFFHR LEDSFLVTED KRGERHPIFG NLEEEVKYHE NFPTIYHLRQ YLADNPEKVD LRLVYLALAH IIKFRGHFLI EGKFDTRNND VQRLFQEFLA VYDNTFENSS LQEQNVQVEE ILTDKISKSA KKDRVLKLFP NEKSNGRFAE FLKLIVGNQA DFKKHFELEE KAPLQFSKDT YEEELEVLLA QIGDNYAELF LSAKKLYDSI LLSGILTVTD VGTKAPLSAS MIQRYNEHQM DLAQLKQFIR QKLSDKYNEV FSDVSKDGYA GYIDGKTNQE AFYKYLKGLL NKIEGSGYFL DKIEREDFLR KQRTFDNGSI PHQIHLQEMR AIIRRQAEFY PFLADNQDRI EKLLTFRIPY YVGPLARGKS DFAWLSRKSA DKITPWNFDE IVDKESSAEA FINRMTNYDL YLPNQKVLPK HSLLYEKFTV YNELTKVKYK TEQGKTAFFD ANMKQEIFDG VFKVYRKVTK DKLMDFLEKE FDEFRIVDLT GLDKENKVFN ASYGTYHDLC KILDKDFLDN SKNEKILEDI VLTLTLFEDR EMIRKRLENY SDLLTKEQVK KLERRHYTGW GRLSAELIHG IRNKESRKTI LDYLIDDGNS NRNFMQLIND DALSFKEEIA KAQVIGETDN LNQVVSDIAG SPAIKKGILQ SLKIVDELVK IMGHQPENIV VEMARENQFT NQGRRNSQQR LKGLTDSIKE FGSQILKEHP VENSQLQNDR LFLYYLQNGR DMYTGEELDI DYLSQYDIDH IIPQAFIKDN SIDNRVLTSS KENRGKSDDV PSKDVVRKMK SYWSKLLSAK LITQRKFDNL TKAERGGLTD DDKAGFIKRQ LVETRQITKH VARILDERFN TETDENNKKI RQVKIVTLKS NLVSNFRKEF ELYKVREIND YHHAHDAYLN AVIGKALLGV YPQLEPEFVY GDYPHFHGHK ENKATAKKFF YSNIMNFFKK DDVRTDKNGE IIWKKDEHIS NIKKVLSYPQ VNIVKKVEEQ TGGFSKESIL PKGNSDKLIP RKTKKFYWDT KKYGGFDSPI VAYSILVIAD IEKGKSKKLK TVKALVGVTI MEKMTFERDP VAFLERKGYR NVQEENIIKL PKYSLFKLEN GRKRLLASAR ELQKGNEIVL PNHLGTLLYH AKNIHKVDEP KHLDYVDKHK DEFKELLDVV SNFSKKYTLA EGNLEKIKEL YAQNNGEDLK ELASSFINLL TFTAIGAPAT FKFFDKNIDR KRYTSTTEIL NATLIHQSIT GLYETRIDLN KLGGD	(SEQ ID NO: 76)

Table 4- Cas9 orthologs		
Prevotella timonensis CRIS 5C-B1 WP_008122718.1	MNKRILGLDT GTNSLGWAVV DWDEHAQSYE LIKYGDVIFQ EGVKIEKGIE SSKAAERSGY KAIRKQYFRR RLRKIQVLKV LVKYHLCPYL SDDDLRQWHL QKQYPKSDEL MLWQRTSDEE GKNPYYDRHR CLHEKLDLTV EADRYTLGRA LYHLTQRRGF LSNRLDTSAD NKEDGVVKSG ISQLSTEMEE AGCEYLGDYF YKLYDAQGNK VRIRQRYTDR NKHYQHEFDA ICEKQELSSE LIEDLQRAIF FQLPLKSQRH GVGRCTFERG KPRCADSHPD YEEFRMLCFV NNIQVKGPHD LELRPLTYEE REKIEPLFFR KSKPNFDFED IAKALAGKKN YAWIHDKEER AYKFNYRMTQ GVPGCPTIAQ LKSIFGDDWK TGIAETYTLI QKKNGSKSLQ EMVDDVWNVL YSFSSVEKLK EFAHHKLQLD EESAEKFAKI KLSHSFAALS LKAIRKFLPF LRKGMYYTHA SFFANIPTIV GKEIWNKEQN RKYIMENVGE LVFNYQPKHR EVQGTIEMLI KDFLANNFEL PAGATDKLYH PSMIETYPNA QRNEFGILQL GSPRTNAIRN PMAMRSLHIL RRVVNQLLKE SIIDENTEVH VEYARELNDA NKRRAIADRQ KEQDKQHKKY GDEIRKLYKE ETGKDIEPTQ TDVLKFQLWE EQNHHCLYTG EQIGITDFIG SNPKFDIEHT IPQSVGGDST QMNLTLCDNR FNREVKKAKL PTELANHEEI LTRIEPWKNK YEQLVKERDK QRTFAGMDKA VKDIRIQKRH KLQMEIDYWR GKYERFTMTE VPEGFSRRQG TGIGLISRYA GLYLKSLFHQ ADSRNKSNVY VVKGVATAEF RKMWGLQSEY EKKCRDNHSH HCMDAITIAC IGKREYDLMA EYYRMEETFK QGRGSKPKFS KPWATFTEDV LNIYKNLLVV HDTPNNMPKH TKKYVQTSIG KVLAQGDTAR GSLHLDTYYG AIERDGEIRY VVRRPLSSFT KPEELENIVD ETVKRTIKEA IADKNFKQAI AEPIYMNEEK GILIKKVRCF AKSVKQPINI RQHRDLSKKE YKQQYHVMNE NNYLLAIYEG LVKNKVVREF EIVSYIEAAK YYKRSQDRNI FSSIVPTHST KYGLPLKTKL LMGQLVLMFE ENPDEIQVDN TKDLVKRLYK VVGIEKDGRI KFKYHQEARK EGLPIFSTPY KNNDDYAPIF RQSINNINIL VDGIDFTIDI LGKVTLKE	(SEQ ID NO: 77)
Clostridium cellulolyticum H10 ACL77411.1	MKYTLGLDVG IASVGWAVID KDNNKIIDLG VRCFDKAEES KTGESLATAR RIARGMRRRI SRRSQRLRLV KKLFVQYEII KDSSEFNRIF DTSRDGWKDP WELRYNALSR ILKPYELVQV LTHITKRRGF KSNRKEDLST TKEGVVITSI KNNSEMLRTK NYRTIGEMIF METPENSNKR NKVDEYIHTI AREDLLNEIK YIFSIQRKLG SPFVTEKLEH DFLNIWEFQR PFASGDSILS KVGKCTLLKE ELRAPTSCYT SEYFGLLQSI NNLVLVEDNN TLTLNNDQRA KIIEYAHFKN EIKYSEIRKL LDIEPEILFK AHNLTHKNPS GNNESKKFYE MKSYHKLKST LPTDIWGKLH SNKESLDNLF YCLTVYKNDN EIKDYLQANN LDYLIEYIAK LPTFNKFKHL SLVAMKRIIP FMEKGYKYSD ACNMAELDFT GSSKLEKCNK LTVEPIIENV TNPVVIRALT QARKVINAII QKYGLPYMVN IELAREAGMT RQDRDNLKKE HENNRKAREK ISDLIRQNGR VASGLDILKW RLWEDQGGRC AYSGKPIPVC DLLNDSLTQI DHIYPYSRSM DDSYMNKVLV LTDENQNKRS YTPYEVWGST EKWEDFEARI YSMHLPQSKE KRLLNRNFIT KDLDSFISRN LNDTRYISRF LKNYIESYLQ FSNDSPKSCV VCVNGQCTAQ LRSRWGLNKN REESDLHHAL DAAVIACADR KIIKEITNYY NERENHNYKV KYPLPWHSFR QDLMETLAGV FISRAPRRKI TGPAHDETIR SPKHFNKGLT SVKIPLTTVT LEKLETMVKN TKGGISDKAV YNVLKNRLIE HNNKPLKAFA EKIYKPLKNG TNGAIIRSIR VETPSYTGVF RNEGKGISDN SLMVRVDVFK KKDKYYLVPI YVAHMIKKEL PSKAIVPLKP ESQWELIDST HEFLFSLYQN DYLVIKTKKG ITEGYYRSCH RGTGSLSLMP HFANNKNVKI DIGVRTAISI EKYNVDILGN KSIVKGEPRR GMEKYNSFKS N	(SEQ ID NO: 78)
Francisella tularensis subsp. novicida U112 WP_003038941.1	MNFKILPIAI DLGVKNTGVF SAFYQKGTSL ERLDNKNGKV YELSKDSYTL LMNNRTARRH QRRGIDRKQL VKRLFKLIWT EQLNLEWDKD TQQAISFLFN RRGFSFITDG YSPEYLNIVP EQVKAILMDI FDDYNGEDDL DSYLKLATEQ ESKISEIYNK LMQKILEFKL MKLCTDIKDD KVSTKTLKEI TSYEFELLAD YLANYSESLK TQKFSYTDKQ GNLKELSYYH HDKYNIQEFL KRHATINDRI LDTLLTDDLD IWNFNFEKFD FDKNEEKLQN QEDKDHIQAH LHHFVFAVNK IKSEMASGGR HRSQYFQEIT NVLDENNHQE GYLKNFCENL HNKKYSNLSV KNLVNLIGNL SNLELKPLRK YFNDKIHAKA DHWDEQKFTE TYCHWILGEW RVGVKDQDKK DGAKYSYKDL CNELKQKVTK AGLVDFLLEL DPCRTIPPYL DNNNRKPPKC QSLILNPKFL DNQYPNWQQY LQELKKLQSI QNYLDSFETD LKVLKSSKDQ PYFVEYKSSN QQIASGQRDY KDLDARILQF IFDRVKASDE LLLNEIYFQA KKLKQKASSE LEKLESSKKL DEVIANSQLS QILKSQHTNG	(SEQ ID NO: 79)

Table 4- Cas9 orthologs		
	IFEQGTFLHL VCKYYKQRQR ARDSRLYIMP EYRYDKKLHK YNNTGRFDDD NQLLTYCNHK PRQKRYQLLN DLAGVLQVSP NFLKDKIGSD DDLFISKWLV EHIRGFKKAC EDSLKIQKDN RGLLNHKINI ARNTKGKCEK EIFNLICKIE GSEDKKGNYK HGLAYELGVL LFGEPNEASK PEFDRKIKKF NSIYSFAQIQ QIAFAERKGN ANTCAVCSAD NAHRMQQIKI TEPVEDNKDK IILSAKAQRL PAIPTRIVDG AVKKMATILA KNIVDDNWQN IKQVLSAKHQ LHIPIITESN AFEFEPALAD VKGKSLKDRR KKALERISPE NIFKDKNNRI KEFAKGISAY SGANLTDGDF DGAKEELDHI IPRSHKKYGT LNDEANLICV TRGDNKNKGN RIFCLRDLAD NYKLKQFETT DDLEIEKKIA DTIWDANKKD FKFGNYRSFI NLTPQEQKAF RHALFLADEN PIKQAVIRAI NNRNRTFVNG TQRYFAEVLA NNIYLRAKKE NLNTDKISFD YFGIPTIGNG RGIAEIRQLY EKVDSDIQAY AKGDKPQASY SHLIDAMLAF CIAADEHRND GSIGLEIDKN YSLYPLDKNT GEVFTKDIFS QIKITDNEFS DKKLVRKKAI EGFNTHRQMT RDGIYAENYL PILIHKELNE VRKGYTWKNS EEIKIFKGKK YDIQQLNNLV YCLKFVDKPI SIDIQISTLE ELRNILTTNN IAATAEYYYI NLKTQKLHEY YIENYNTALG YKKYSKEMEF LRSLAYRSER VKIKSIDDVK QVLDKDSNFI IGKITLPFKK EWQRLYREWQ NTTIKDDYEF LKSFFNVKSI TKLHKKVRKD FSLPISTNEG KFLVKRKTWD NNFIYQILND SDSRADGTKP FIPAFDISKN EIVEAIIDSF TSKNIFWLPK NIELQKVDNK NIFAIDTSKW FEVETPSDLR DIGIATIQYK IDNNSRPKVR VKLDYVIDDD SKINYFMNHS LLKSRYPDKV LEILKQSTII EFESSGFNKT IKEMLGMKLA GIYNETSNN	
Azospirillum sp. B510 AOL40891.1	MARPAFRAPR REHVNGWTPD PHRISKPFFI LVSWHLLSRV VIDSSSGCFP GTSRDHTDKF AEWECAVQPY RLSFDLGTNS IGWGLLNLDR QGKPREIRAL GSRIFSDGRD PQDKASLAVA RRLARQMRRR RDRYLTRRTR LMGALVRFGL MPADPAARKR LEVAVDPYLA RERATRERLE PFEIGRALFH LNQRRGYKPV RTATKPDEEA GKVKEAVERL EAAIAAAGAP TLGAWFAWRK TRGETLRARL AGKGKEAAYP FYPARRMLEA EFDTLWAEQA RHHPDLLTAE AREILRHRIF HQRPLKPPPV GRCTLYPDDG RAPRALPSAQ RLRLFQELAS LRVIHLDLSE RPLTPAERDR IVAFVQGRPP KAGRKPGKVQ KSVPFEKLRG LLELPPGTGF SLESDKRPEL LGDETGARIA PAFGPGWTAL PLEEQDALVE LLLTEAEPER AIAALTARWA LDEATAAKLA GATLPDFHGR YGRRAVAELL PVLERETRGD PDGRVRPIRL DEAVKLLRGG KDHSDFSREG ALLDALPYYG AVLERHVAFG TGNPADPEEK RVGRVANPTV HIALNQLRHL VNAILARHGR PEEIVIELAR DLKRSAEDRR REDKRQADNQ KRNEERKRLI LSLGERPTPR NLLKLRLWEE QGPVENRRCP YSGETISMRM LLSEQVDIDH ILPFSVSLDD SAANKVVCLR EANRIKRNRS PWEAFGHDSE RWAGILARAE ALPKNKRWRF APDALEKLEG EGGLRARHLN DTRHLSRLAV EYLRCVCPKV RVSPGRLTAL LRRRWGIDAI LAEADGPPPE VPAETLDPSP AEKNRADHRH HALDAVVIGC IDRSMVQRVQ LAAASAEREA AAREDNIRRV LEGFKEEPWD GFRAELERRA RTIVVSHRPE HGIGGALHKE TAYGPVDPPE EGFNLVVRKP IDGLSKDEIN SVRDPRLRRA LIDRLAIRRR DANDPATALA KAAEDLAAQP ASRGIRRVRV LKKESNPIRV EHGGNPSGPR SGGPFHKLLL AGEVHHVDVA LRADGRRWVG HWVTLFEAHG GRGADGAAAP PRLGDGERFL MRLHKGDCLK LEHKGRVRVM QVVKLEPSSN SVVVVEPHQV KTDRSKHVKI SCDQLRARGA RRVTVDPLGR VRVHAPGARV GIGGDAGRTA MEPAEDIS	(SEQ ID NO: 80)
Peptoniphilus duerdenii ATCC BAA-1640 WP_008901059.1	MKNLKEYYIG LDIGTASVGW AVTDESYNIP KFNGKKMWGV RLFDDAKTAE ERRTQRGSRR RLNRRKERIN LLQDLFATEI SKVDPNFFLR LDNSDLYRED KDEKLKSKYT LFNDKDFKDR DYHKKYPTIH HLIMDLIEDE GKKDIRLLYL ACHYLLKNRG HFIFEGQKFD TKNSFDKSIN DLKIHLRDEY NIDLEFNNED LIEIITDTTL NKTNKKKELK NIVGDTKFLK AISAIMIGSS QKLVDLFEDG EFEETTVKSV DFSTTAFDDK YSEYEEALGD TISLLNILKS IYDSSILENL LKDADKSKDG NKYISKAFVK KFNKHGKDLK TLKRIIKKYL PSEYANIFRN KSINDNYVAY TKSNITSNKR TKASKFTKQE DFYKFIKKHL DTIKETKLNS SENEDLKLID EMLTDIEFKT FIPKLKSSDN GVIPYQLKLM ELKKILDNQS KYYDFLNESD EYGTVKDKVE SIMEFRIPYY VGPLNPDSKY AWIKRENTKI TPWNFKDIVD LDSSREEFID RLIGRCTYLK EEKVLPKASL IYNEFMVLNE LNNLKLNEFL ITEEMKKAIF EELFKTKKKV TLKAVSNLLK KEFNLTGDIL LSGTDGDFKQ GLNSYIDFKN IIGDKVDRDD YRIKIEEIIK LIVLYEDDKT	(SEQ ID NO: 81)

Table 4- Cas9 orthologs		
	YLKKKIKSAY KNDFTDDEIK KIAALNYKDW GRLSKRFLTG IEGVDKTTGE KGSIIYFMRE YNLNLMELMS GHYTFTEEVE KLNPVENREL CYEMVDELYL SPSVKRMLWQ SLRVVDEIKR IIGKDPKKIF IEMARAKEAK NSRKESRKNK LLEFYKFGKK AFINEIGEER YNYLLNEINS EEESKFRWDN LYLYYTQLGR CMYSLEPIDL ADLKSNNIYD QDHIYPKSKI YDDSLENRVL VKKNLNHEKG NQYPIPEKVL NKNAYGFWKI LFDKGLIGQK KYTRLTRRTP FEERELAEFI ERQIVETRQA TKETANLLKN ICQDSEIVYS KAENASRFRQ EFDIIKCRTV NDLHHMHDAY LNIVVGNVYN TKFTKNPLNF IKDKDNVRSY NLENMFKYDV VRGSYTAWIA DDSEGNVKAA TIKKVKRELE GKNYRFTRMS YIGTGGLYDQ NLMRKGKGQI PQKENTNKSN IEKYGGYNKA SSAYFALIES DGKAGRERTL ETIPIMVYNQ EKYGNTEAVD KYLKDNLELQ DPKILKDKIK INSLIKLDGF LYNIKGKTGD SLSIAGSVQL IVNKEEQKLI KKMDKFLVKK KDNKDIKVTS FDNIKEEELI KLYKTLSDKL NNGIYSNKRN NQAKNISEAL DKFKEISIEE KIDVLNQIIL LFQSYNNGCN LKSIGLSAKT GVVFIPKKLN YKECKLINQS ITGLFENEVD LLNL	
Lactobacillus coryniformis subsp. torquens KCTC 3535 WP_010014406.1	MGYRIGLDVG ITSTGYAVLK TDKNGLPYKI LTLDSVIYPR AENPQTGASL AEPRRIKRGL RRRTRRTKFR KQRTQQLFIH SGLLSKPEIE QILATPQAKY SVYELRVAGL DRRLTNSELF RVLYFFIGHR GFKSNRKAEL NPENEADKKQ MGQLLNSIEE IRKAIAEKGY RTVGELYLKD PKYNDHKRNK GYIDGYLSTP NRQMLVDEIK QILDKQRELG NEKLTDEFYA TYLLGDENRA GIFQAQRDFD EGPGAGPYAG DQIKKMVGKD IFEPTEDRAA KATYTFQYFN LLQKMTSLNY QNTTGDTWHT LNGLDRQAII DAVFAKAEKP TKTYKPTDFG ELRKLLKLPD DARFNLVNYG SLQTQKEIET VEKKTRFVDF KAYHDLVKVL PEEMWQSRQL LDHIGTALTL YSSDKRRRRY FAEELNLPAE LIEKLLPLNF SKFGHLSIKS MQNIIPYLEM GQVYSEATTN TGYDFRKKQI SKDTIREEIT NPVVRRAVTK TIKIVEQIIR RYGKPDGINI ELARELGRNF KERGDIQKRQ DKNRQTNDKI AAELTELGIP VNGQNIIRYK LHKEQNGVDP YTGDQIPFER AFSEGYEVDH IIPYSISWDD SYTNKVLTSA KCNREKGNRI PMVYLANNEQ RLNALTNIAD NIIRNSRKRQ KLLKQKLSDE ELKDWKQRNI NDTRFITRVL YNYFRQAIEF NPELEKKQRV LPLNGEVTSK IRSRWGFLKV REDGDLHHAI DATVIAAITP KFIQQVTKYS QHQEVKNNQA LWHDAEIKDA EYAAEAQRMD ADLFNKIFNG FPLPWPEFLD ELLARISDNP VEMMKSRSWN TYTPIEIAKL KPVFVVRLAN HKISGPAHLD TIRSAKLFDE KGIVLSRVSI TKLKINKKGQ VATGDGIYDP ENSNNGDKVV YSAIRQALEA HNGSGELAFP DGYLEYVDHG TKKLVRKVRV AKKVSLPVRL KNKAAADNGS MVRIDVFNTG KKFVFVPIYI KDTVEQVLPN KAIARGKSLW YQITESDQFC FSLYPGDMVH IESKTGIKPK YSNKENNTSV VPIKNFYGYF DGADIATASI LVRAHDSSYT ARSIGIAGLL KFEKYQVDYF GRYHKVHEKK RQLFVKRDE	(SEQ ID NO: 82)
Ignavibacterium album JCM 16511 WP_014561873.1	MEFKKVLGLD IGTNSIGCAL LSLPKSIQDY GKGGRLEWLT SRVIPLDADY MKAFIDGKNG LPQVITPAGK RRQKRGSRRL KHRYKLRRSR LIRVFKTLNW LPEDFPLDNP KRIKETISTE GKFSFRISDY VPISDESYRE FYREFGYPEN EIEQVIEEIN FRRKTKGKNK NPMIKLLPED WVVYYLRKKA LIKPTTKEEL IRIIYLFNQR RGFKSSRKDL TETAILDYDE FAKRLAEKEK YSAENYETKF VSITKVKEVV ELKTDGRKGK KRFKVILEDS RIEPYEIERK EKPDWEGKEY TFLVTQKLEK GKFKQNKPDL PKEEDWALCT TALDNRMGSK HPGEFFFDEL LKAFKEKRGY KIRQYPVNRW RYKKELEFIW TKQCQLNPEL NNLNINKEIL RKLATVLYPS QSKFFGPKIK EFENSDVLHI ISEDIIYYQR DLKSQKSLIS ECRYEKRKGI DGEIYGLKCI PKSSPLYQEF RIWQDIHNIK VIRKESEVNG KKKINIDETQ LYINENIKEK LFELFNSKDS LSEKDILELI SLNIINSGIK ISKKEEETTH RINLFANRKE LKGNETKSRY RKVFKKLGFD GEYILNHPSK LNRLWHSDYS NDYADKEKTE KSILSSLGWK NRNGKWEKSK NYDVFNLPLE VAKAIANLPP LKKEYGSYSA LAIRKMLVVM RDGKYWQHPD QIAKDQENTS LMLFDKNLIQ LTNNQRKVLN KYLLTLAEVQ KRSTLIKQKL NEIEHNPYKL ELVSDQDLEK QVLKSFLEKK NESDYLKGLK TYQAGYLIYG KHSEKDVPIV NSPDELGEYI RKKLPNNSLR NPIVEQVIRE TIFIVRDVWK SFGIIDEIHI ELGRELKNNS EERKKTSESQ EKNFQEKERA RKLLKELLNS SNFEHYDENG NKIFSSFTVN PNPDSPLDIE KFRIWKNQSG LTDEELNKKL KDEKIPTEIE VKKYILWLTQ KCRSPYTGKI IPLSKLFDSN VYEIEHIIPR SKMKNDSTNN	(SEQ ID NO: 83)

Table 4- Cas9 orthologs		
	LVICELGVNK AKGDRLAANF ISESNGKCKF GEVEYTLLKY GDYLQYCKDT FKYQKAKYKN LLATEPPEDF IERQINDTRY IGRKLAELLT PVVKDSKNII FTIGSITSEL KITWGLNGVW KDILRPRFKR LESIINKKLI FQDEDDPNKY HFDLSINPQL DKEGLKRLDH RHHALDATII AATTREHVRY LNSLNAADND EEKREYFLSL CNHKIRDFKL PWENFTSEVK SKLLSCVVSY KESKPILSDP FNKYLKWEYK NGKWQKVFAI QIKNDRWKAV RRSMFKEPIG TVWIKKIKEV SLKEAIKIQA IWEEVKNDPV RKKKEKYIYD DYAQKVIAKI VQELGLSSSM RKQDDEKLNK FINEAKVSAG VNKNLNTTNK TIYNLEGRFY EKIKVAEYVL YKAKRMPLNK KEYIEKLSLQ KMFNDLPNFI LEKSILDNYP EILKELESDN KYIIEPHKKN NPVNRLLLEH ILEYHNNPKE AFSTEGLEKL NKKAINKIGK PIKYITRLDG DINEEEIFRG AVFETDKGSN VYFVMYENNQ TKDREFLKPN PSISVLKAIE HKNKIDFFAP NRLGFSRIIL SPGDLVYVPT NDQYVLIKDN SSNETIINWD DNEFISNRIY QVKKFTGNSC YFLKNDIASL ILSYSASNGV GEFGSQNISE YSVDDPPIRI KDVCIKIRVD RLGNVRPL	
uncultured delta proteobacterium HF0070_07E19 ADI19058.1	MSSKAIDSLE QLDLFKPQEY TLGLDLGIKS IGWAILSGER IANAGVYLFE TAEELNSTGN KLISKAAERG RKRRIRRMLD RKARRGRHIR YLLEREGLPT DELEEVVVHQ SNRTLWDVRA EAVERKLTKQ ELAAVLFHLV RHRGYFPNTK KLPPDDESDS ADEEQGKINR ATSRLREELK ASDCKTIGQF LAQNRDRQRN REGDYSNLMA RKLVFEEALQ ILAFQRKQGH ELSKDFEKTY LDVLMGQRSG RSPKLGNCSL IPSELRAPSS APSTEWFKFL QNLGNLQISN AYREEWSIDA PRRAQIIDAC SQRSTSSYWQ IRRDFQIPDE YRFNLVNYER RDPDVDLQEY LQQQERKTLA NFRNWKQLEK IIGTGHPIQT LDEAARLITL IKDDEKLSDQ LADLLPEASD KAITQLCELD FTTAAKISLE AMYRILPHMN QGMGFFDACQ QESLPEIGVP PAGDRVPPFD EMYNPVVNRV LSQSRKLINA VIDEYGMPAK IRVELARDLG KGRELRERIK LDQLDKSKQN DQRAEDFRAE FQQAPRGDQS LRYRLWKEQN CTCPYSGRMI PVNSVLSEDT QIDHILPISQ SFDNSLSNKV LCFTEENAQK SNRTPFEYLD AADFQRLEAI SGNWPEAKRN KLLHKSFGKV AEEWKSRALN DTRYLTSALA DHLRHHLPDS KIQTVNGRIT GYLRKQWGLE KDRDKHTHHA VDAIVVACTT PAIVQQVTLY HQDIRRYKKL GEKRPTPWPE TFRQDVLDVE EEIFITRQPK KVSGGIQTKD TLRKHRSKPD RQRVALTKVK LADLERLVEK DASNRNLYEH LKQCLEESGD QPTKAFKAPF YMPSGPEAKQ RPILSKVTLL REKPEPPKQL TELSGGRRYD SMAQGRLDIY RYKPGGKRKD EYRVVLQRMI DLMRGEENVH VFQKGVPYDQ GPEIEQNYTF LFSLYFDDLV EFQRSADSEV IRGYYRTFNI ANGQLKISTY LEGRQDFDFF GANRLAHFAK VQVNLLGKVI K	(SEQ ID NO: 84)
Ruminococcus albus 8 WP_002846926.1	MGNYYLGLDV GIGSIGWAVI NIEKKRIEDF NVRIFKSGEI QEKNRNSRAS QQCRRSRGLR RLYRRKSHRK LRLKNYLSII GLTTSEKIDY YYETADNNVI QLRNKGLSEK LTPEEIAACL IHICNNRGYK DFYEVNVEDI EDPDERNEYK EEHDSIVLIS NLMNEGGYCT PAEMICNCRE FDEPNSVYRK FHNSAASKNH YLITRHMLVK EVDLILENQS KYYGILDDKT IAKIKDIIFA QRDFEIGPGK NERFRRFTGY LDSIGKCQFF KDQERGSRFT VIADIYAFVN VLSQYTYTNN RGESVFDTSF ANDLINSALK NGSMDKRELK AIAKSYHIDI SDKNSDTSLT KCFKYIKVVK PLFEKYGYDW DKLIENYTDT DNNVLNRIGI VLSQAQTPKR RREKLKALNI GLDDGLINEL TKLKLSGTAN VSYKYMQGSI EAFCEGDLYG KYQAKFNKEI PDIDENAKPQ KLPPFKNEDD CEFFKNPVVF RSINETRKLI NAIIDKYGYP AAVNIETADE LNKTFEDRAI DTKRNNDNQK ENDRIVKEII ECIKCDEVHA RHLIEKYKLW EAQEGKCLYS GETITKEDML RDKDKLFEVD HIVPYSLILD NTINNKALVY AEENQKKGQR TPLMYMNEAQ AADYRVRVNT MFKSKKCSKK KYQYLMLPDL NDQELLGGWR SRNLNDTRYI CKYLVNYLRK NLRFDRSYES SDEDDLKIRD HYRVFPVKSR FTSMFRRWWL NEKTWGRYDK AELKKLTYLD HAADAIIIAN CRPEYVVLAG EKLKLNKMYH QAGKRITPEY EQSKKACIDN LYKLFRMDRR TAEKLLSGHG RLTPIIPNLS EEVDKRLWDK NIYEQFWKDD KDKKSCEELY RENVASLYKG DPKFASSLSM PVISLKPDHK YRGTITGEEA IRVKEIDGKL IKLKRKSISE ITAESINSIY TDDKILIDSL KTIFEQADYK DVGDYLKKTN QHFFTTSSGK RVNKVTVIEK VPSRWLRKEI DDNNFSLLND SSYYCIELYK DSKGDNNLQG IAMSDIVHDR KTKKLYLKPD FNYPDDYYTH VMYIFPGDYL RIKSTSKKSG EQLKFEGYFI SVKNVNENSF RFISDNKPCA KDKRVSITKK DIVIKLAVDL MGKVQGENNG KGISCGEPLS	(SEQ ID NO: 85)

Table 4- Cas9 orthologs		
	LLKEKN	
Lactobacillus farciminis KCTC 3681 WP_010018949.1	MTKKEQPYNI GLDIGTSSVG WAVTNDNYDL LNIKKKNLWG VRLFEEAQTA KETRLNRSTR RRYRRRKNRI NWLNEIFSEE LAKTDPSFLI RLQNSWVSKK DPDRKRDKYN LFIDGPYTDK EYYREFPTIF HLRKELILNK DKADIRLIYL ALHNILKYRG NFTYEHQKFN ISNLNNNLSK ELIELNQQLI KYDISFPDDC DWNHISDILI GRGNATQKSS NILKDFTLDK ETKKLLKEVI NLILGNVAHL NTIFKTSLTK DEEKLNFSGK DIESKLDDLD SILDDDQFTV LDAANRIYST ITLNEILNGE SYFSMAKVNQ YENHAIDLCK LRDMWHTTKN EEAVEQSRQA YDDYINKPKY GTKELYTSLK KFLKVALPTN LAKEAEEKIS KGTYLVKPRN SENGVVPYQL NKIEMEKIID NQSQYYPFLK ENKEKLLSIL SFRIPYYVGP LQSAEKNPFA WMERKSNGHA RPWNFDEIVD REKSSNKFIR RMTVTDSYLV GEPVLPKNSL IYQRYEVLNE LNNIRITENL KTNPIGSRLT VETKQRIYNE LFKKYKKVTV KKLTKWLIAQ GYYKNPILIG LSQKDEFNST LTTYLDMKKI FGSSFMEDNK NYDQIEELIE WLTIFEDKQI LNEKLHSSKY SYTPDQIKKI SNMRYKGWGR LSKKILMDIT TETNTPQLLQ LSNYSILDLM WATNNNFISI MSNDKYDFKN YIENHNLNKN EDQNISDLVN DIHVSPALKR GITQSIKIVQ EIVKFMGHAP KHIFIEVTRE TKKSEITTSR EKRIKRLQSK LLNKANDFKP QLREYLVPNK KIQEELKKHK NDLSSERIML YFLQNGKSLY SEESLNINKL SDYQVDHILP RTYIPDDSLE NKALVLAKEN QRKADDLLLN SNVIDRNLER WTYMLNNNMI GLKKFKNLTR RVITDKDKLG FIHRQLVQTS QMVKGVANIL DNMYKNQGTT CIQARANLST AFRKALSGQD DTYHFKHPEL VKNRNVNDFH HAQDAYLASF LGTYRLRRFP TNEMLLMNGE YNKFYGQVKE LYSKKKKLPD SRKNGFIISP LVNGTTQYDR NTGEIIWNVG FRDKILKIFN YHQCNVTRKT EIKTGQFYDQ TIYSPKNPKY KKLIAQKKDM DPNIYGGFSG DNKSSITIVK IDNNKIKPVA IPIRLINDLK DKKTLQNWLE ENVKHKKSIQ IIKNNVPIGQ IIYSKKVGLL SLNSDREVAN RQQLILPPEH SALLRLLQIP DEDLDQILAF YDKNILVEIL QELITKMKKF YPFYKGEREF LIANIENFNQ ATTSEKVNSL EELITLLHAN STSAHLIFNN IEKKAFGRKT HGLTLNNTDF IYQSVTGLYE TRIHIE	(SEQ ID NO: 86)
Eubacterium dolichum DSM 3991 WP_004800457.1	MMEVFMGRLV LGLDIGITSV GFGIIDLDES EIVDYGVRLF KEGTAAENET RRTKRGGRRL KRRRVTRRED MLHLLKQAGI ISTSFHPLNN PYDVRVKGLN ERLNGEELAT ALLHLCKHRG SSVETIEDDE AKAKEAGETK KVLSMNDQLL KSGKYVCEIQ KERLRTNGHI RGHENNFKTR AYVDEAFQIL SHQDLSNELK SAIITIISRK RMYYDGPGGP LSPTPYGRYT YFGQKEPIDL IEKMRGKCSL FPNEPRAPKL AYSAELFNLL NDLNNLSIEG EKLTSEQKAM ILKIVHEKGK ITPKQLAKEV GVSLEQIRGF RIDTKGSPLL SELTGYKMIR EVLEKSNDEH LEDHVFYDEI AEILTKTKDI EGRKKQISEL SSDLNEESVH QLAGLTKFTA YHSLSFKALR LINEEMLKTE LNQMQSITLF GLKQNNELSV KGMKNIQADD TAILSPVAKR AQRETFKVVN RLREIYGEFD SIVVEMAREK NSEEQRKAIR ERQKFFEMRN KQVADIIGDD RKINAKLREK LVLYQEQDGK TAYSLEPIDL KLLIDDPNAY EVDHIIPISI SLDDSITNKV LVTHRENQEK GNLTPISAFV KGRFTKGSLA QYKAYCLKLK EKNIKTNKGY RKKVEQYLLN ENDIYKYDIQ KEFINRNLVD TSYASRVVLN TLTTYFKQNE IPTKVFTVKG SLTNAFRRKI NLKKDRDEDY GHHAIDALII ASMPKMRLLS TIFSRYKIED IYDESTGEVF SSGDDSMYYD DRYFAFIASL KAIKVRKFSH KIDTKPNRSV ADETIYSTRV IDGKEKVVKK YKDIYDPKFT ALAEDILNNA YQEKYLMALH DPQTFDQIVK VVNYYFEEMS KSEKYFTKDK KGRIKISGMN PLSLYRDEHG MLKKYSKKGD GPAITQMKYF DGVLGNHIDI SAHYQVRDKK VVLQQISPYR TDFYYSKENG YKFVTIRYKD VRWSEKKKKY VIDQQDYAMK KAEKKIDDTY EFQFSMHRDE LIGITKAEGE ALIYPDETWH NFNFFFHAGE TPEILKFTAT NNDKSNKIEV KPIHCYCKMR LMPTISKKIV RIDKYATDVV GNLYKVKKNT LKFEFD	(SEQ ID NO: 87)
Nitratifractor salsuginis DSM 16511 ADV46720.1	MKKILGVDLG ITSFGYAILQ ETGKDLYRCL DNSVVMRNNP YDEKSGESSQ SIRSTQKSMR RLIEKRKKRI RCVAQTMERY GILDYSETMK INDPKNNPIK NRWQLRAVDA WKRPLSPQEL FAIFAHMAKH RGYKSIATED LIYELELELG LNDPEKESEK KADERRQVYN ALRHLEELRK KYGGETIAQT IHRAVEAGDL RSYRNHDDYE KMIRREDIEE EIEKVLLRQA ELGALGLPEE QVSELIDELK ACITDQEMPT IDESLFGKCT FYKDELAAPA YSYLYDLYRL YKKLADLNID	(SEQ ID NO: 88)

Table 4- Cas9 orthologs		
	GYEVTQEDRE KVIEWVEKKI AQGKNLKKIT HKDLRKILGL APEQKIFGVE DERIVKGKKE PRTFVPFFFL ADIAKFKELF ASIQKHPDAL QIFRELAEIL QRSKTPQEAL DRLRALMAGK GIDTDDRELL ELFKNKRSGT RELSHRYILE ALPLFLEGYD EKEVQRILGF DDREDYSRYP KSLRHLHLRE GNLFEKEENP INNHAVKSLA SWALGLIADL SWRYGPFDEI ILETTRDALP EKIRKEIDKA MREREKALDK IIGKYKKEFP SIDKRLARKI QLWERQKGLD LYSGKVINLS QLLDGSADIE HIVPQSLGGL STDYNTIVTL KSVNAAKGNR LPGDWLAGNP DYRERIGMLS EKGLIDWKKR KNLLAQSLDE IYTENTHSKG IRATSYLEAL VAQVLKRYYP FPDPELRKNG IGVRMIPGKV TSKTRSLLGI KSKSRETNFH HAEDALILST LTRGWQNRLH RMLRDNYGKS EAELKELWKK YMPHIEGLTL ADYIDEAFRR FMSKGEESLF YRDMFDTIRS ISYWVDKKPL SASSHKETVY SSRHEVPTLR KNILEAFDSL NVIKDRHKLT TEEFMKRYDK EIRQKLWLHR IGNTNDESYR AVEERATQIA QILTRYQLMD AQNDKEIDEK FQQALKELIT SPIEVTGKLL RKMRFVYDKL NAMQIDRGLV ETDKNMLGIH ISKGPNEKLI FRRMDVNNAH ELQKERSGIL CYLNEMLFIF NKKGLIHYGC LRSYLEKGQG SKYIALFNPR FPANPKAQPS KFTSDSKIKQ VGIGSATGII KAHLDLDGHV RSYEVFGTLP EGSIEWFKEE SGYGRVEDDP HH	
Rhodospirillum rubrum ATCC 11170 WP_011388212.1	MRPIEPWILG LDIGTDSLGW AVFSCEEKGP PTAKELLGGG VRLFDSGRDA KDHTSRQAER GAFRRARRQT RTWPWRRDRL IALFQAAGLT PPAAETRQIA LALRREAVSR PLAPDALWAA LLHLAHHRGF RSNRIDKRER AAAKALAKAK PAKATAKATA PAKEADDEAG FWEGAEAALR QRMAASGAPT VGALLADDLD RGQPVRMRYN QSDRDGVVAP TRALIAEELA EIVARQSSAY PGLDWPAVTR LVLDQRPLRS KGAGPCAFLP GEDRALRALP TVQDFIIRQT LANLRLPSTS ADEPRPLTDE EHAKALALLS TARFVEWPAL RRALGLKRGV KFTAETERNG AKQAARGTAG NLTEAILAPL IPGWSGWDLD RKDRVFSDLW AARQDRSALL ALIGDPRGPT RVTEDETAEA VADAIQIVLP TGRASLSAKA ARAIAQAMAP GIGYDEAVTL ALGLHHSHRP RQERLARLPY YAAALPDVGL DGDPVGPPPA EDDGAAAEAY YGRIGNISVH IALNETRKIV NALLHRHGPI LRLVMVETTR ELKAGADERK RMIAEQAERE RENAEIDVEL RKSDRWMANA RERRQRVRLA RRQNNLCPYT STPIGHADLL GDAYDIDHVI PLARGGRDSL DNMVLCQSDA NKTKGDKTPW EAFHDKPGWI AQRDDFLARL DPQTAKALAW RFADDAGERV ARKSAEDEDQ GFLPRQLTDT GYIARVALRY LSLVTNEPNA VVATNGRLTG LLRLAWDITP GPAPRDLLPT PRDALRDDTA ARRFLDGLTP PPLAKAVEGA VQARLAALGR SRVADAGLAD ALGLTLASLG GGGKNRADHR HHFIDAAMIA VTTRGLINQI NQASGAGRIL DLRKWPRTNF EPPYPTFRAE VMKQWDHIHP SIRPAHRDGG SLHAATVFGV RNRPDARVLV QRKPVEKLFL DANAKPLPAD KIAEIIDGFA SPRMAKRFKA LLARYQAAHP EVPPALAALA VARDPAFGPR GMTANTVIAG RSDGDGEDAG LITPFRANPK AAVRTMGNAV YEVWEIQVKG RPRWTHRVLT RFDRTQPAPP PPPENARLVM RLRRGDLVYW PLESGDRLFL VKKMAVDGRL ALWPARLATG KATALYAQLS CPNINLNGDQ GYCVQSAEGI RKEKIRTTSC TALGRLRLSK KAT	(SEQ ID NO: 89)
Finegoldia magna ATCC 29328 WP_012290141.1	MKSEKKYYIG LDVGTNSVGW AVTDEFYNIL RAKGKDLWGV RLFEKADTAA NTRIFRSGRR RNDRKGMRLQ ILREIFEDEI KKVDKDFYDR LDESKFWAED KKVSGKYSLF NDKNFSDKQY FEKFPTIFHL RKYLMEEHGK VDIRYYFLAI NQMMKRRGHF LIDGQISHVT DDKPLKEQLI LLINDLLKIE LEEELMDSIF EILADVNEKR TDKKNNLKEL IKGQDFNKQE GNILNSIFES IVTGKAKIKN IISDEDILEK IKEDNKEDFV LTGDSYEENL QYFEEVLQEN ITLFNTLKST YDFLILQSIL KGKSTLSDAQ VERYDEHKKD LEILKKVIKK YDEDGKLFKQ VFKEDNGNGY VSYIGYYLNK NKKITAKKKI SNIEFTKYVK GILEKQCDCE DEDVKYLLGK IEQENFLLKQ ISSINSVIPH QIHLFELDKI LENLAKNYPS FNNKKEEFTK IEKIRKTFTF RIPYYVGPLN DYHKNNGGNA WIFRNKGEKI RPWNFEKIVD LHKSEEEFIK RMLNQCTYLP EETVLPKSSI LYSEYMVLNE LNNLRINGKP LDTDVKLKLI EELFKKKTKV TLKSIRDYMV RNNFADKEDF DNSEKNLEIA SNMKSYIDFN NILEDKFDVE MVEDLIEKIT IHTGNKKLLK KYIEETYPDL SSSQIQKIIN LKYKDWGRLS RKLLDGIKGT KKETEKTDTV INFLRNSSDN LMQIIGSQNY SFNEYIDKLR KKYIPQEISY EVVENLYVSP SVKKMIWQVI RVTEEITKVM GYDPDKIFIE MAKSEEEKKT TISRKNKLLD LYKAIKKDER DSQYEKLLTG LNKLDDSDLR SRKLYLYYTQ MGRDMYTGEK	(SEQ ID NO: 90)

Table 4- Cas9 orthologs		
	IDLDKLFDST HYDKDHIIPQ SMKKDDSIIN NLVLVNKNAN QTTKGNIYPV PSSIRNNPKI YNYWKYLMEK EFISKEKYNR LIRNTPLTNE ELGGFINRQL VETRQSTKAI KELFEKFYQK SKIIPVKASL ASDLRKDMNT LKSREVNDLH HAHDAFLNIV AGDVWNREFT SNPINYVKEN REGDKVKYSL SKDFTRPRKS KGKVIWTPEK GRKLIVDTLN KPSVLISNES HVKKGELFNA TIAGKKDYKK GKIYLPLKKD DRLQDVSKYG GYKAINGAFF FLVEHTKSKK RIRSIELFPL HLLSKFYEDK NTVLDYAINV LQLQDPKIII DKINYRTEII IDNFSYLIST KSNDGSITVK PNEQMYWRVD EISNLKKIEN KYKKDAILTE EDRKIMESYI DKIYQQFKAG KYKNRRTTDT IIEKYEIIDL DTLDNKQLYQ LLVAFISLSY KTSNNAVDFT VIGLGTECGK PRITNLPDNT YLVYKSITGI YEKRIRIK	
Eubacterium rectale ATCC 33656 WP_012742555.1	MNYTEKEKLF MKYILALDIG IASVGWAILD KESETVIEAG SNIFPEASAA DNQLRRDMRG AKRNNRRLKT RINDFIKLWE NNNLSIPQFK STEIVGLKVR AITEEITLDE LYLILYSYLK HRGISYLEDA LDDTVSGSSA YANGLKLNAK ELETHYPCEI QQERLNTIGK YRGQSQIINE NGEVLDLSNV FTIGAYRKEI QRVFEIQKKY HPELTDEFCD GYMLIFNRKR KYYEGPGNEK SRTDYGRFTT KLDANGNYIT EDNIFEKLIG KCSVYPDELR AAAASYTAQE YNVLNDLNNL TINGRKLEEN EKHEIVERIK SSNTINMRKI ISDCMGENID DFAGARIDKS GKEIFHKFEV YNKMRKALLE IGIDISNYSR EELDEIGYIM TINTDKEAMM EAFQKSWIDL SDDVKQCLIN MRKTNGALFN KWQSFSLKIM NELIPEMYAQ PKEQMTLLTE MGVTKGTQEE FAGLKYIPVD VVSEDIFNPV VRRSVRISFK ILNAVLKKYK ALDTIVIEMP RDRNSEEQKK RINDSQKLNE KEMEYIEKKL AVTYGIKLSP SDFSSQKQLS LKLKLWNEQD GICLYSGKTI DPNDIINNPQ LFEIDHIIPR SISFDDARSN KVLVYRSENQ KKGNQTPYYY LTHSHSEWSF EQYKATVMNL SKKKEYAISR KKIQNLLYSE DITKMDVLKG FINRNINDTS YASRLVLNTI QNFFMANEAD TKVKVIKGSY THQMRCNLKL DKNRDESYSH HAVDAMLIGY SELGYEAYHK LQGEFIDFET GEILRKDMWD ENMSDEVYAD YLYGKKWANI RNEVVKAEKN VKYWHYVMRK SNRGLCNQTI RGTREYDGKQ YKINKLDIRT KEGIKVFAKL AFSKKDSDRE RLLVYLNDRR TFDDLCKIYE DYSDAANPFV QYEKETGDII RKYSKKHNGP RIDKLKYKDG EVGACIDISH KYGFEKGSKK VILESLVPYR MDVYYKEENH SYYLVGVKQS DIKFEKGRNV IDEEAYARIL VNEKMIQPGQ SRADLENLGF KFKLSFYKND IIEYEKDGKI YTERLVSRTM PKQRNYIETK PIDKAKFEKQ NLVGLGKTKF IKKYRYDILG NKYSCSEEKF TSFC	(SEQ ID NO: 91)
Corynebacterium diphtheriae C7 (beta) AEX66236.1 WP_014318431.1	MKYHVGIDVG TFSVGLAAIE VDDAGMPIKT LSLVSHIHDS GLDPDKIKSA VTRLASSGIA RRTRRLYRRK RRRLQQLDKF IQRQGWPVIE LEDYSDPLYP WKVRAELAAS YIADEKERGE KLSVALRHIA RHRGWRNPYA KVSSLYLPDE PSDAFKAIRE EIKRASGQPV PETATVGQMV TLCELGTLKL RGEGGVLSAR LQQSDHAREI QEICRMQEIG QELYRKIIDV VFAAESPKGS ASSRVGKDPL QPGKNRALKA SDAFQRYRIA ALIGNLRVRV DGEKRILSVE EKNLVFDHLV NLAPKKEPEW VTIAEILGID RGQLIGTATM TDDGERAGAR PPTHDTNRSI VNSRIAPLVD WWKTASALEQ HAMVKALSNA EVDDFDSPEG AKVQAFFADL DDDVHAKLDS LHLPVGRAAY SEDTLVRLTR RMLADGVDLY TARLQEFGIE PSWTPPAPRI GEPVGNPAVD RVLKTVSRWL ESATKTWGAP ERVIIEHVRE GFVTEKRARE MDGDMRRRAA RNAKLFQEMQ EKLNVQGKPS RADLWRYQSV QRQNCQCAYC GSPITFSNSE MDHIVPRAGQ GSTNTRENLV AVCHRCNQSK GNTPFAIWAK NTSIEGVSVK EAVERTRHWV TDTGMRSTDF KKFTKAVVER FQRATMDEEI DARSMESVAW MANELRSRVA QHFASHGTTV RVYRGSLTAE ARRASGISGK LEFLDGVGKS RLDRRHHAID AAVIAFTSDY VAETLAVRSN LKQSQAHRQE APQWREFTGK DAEHRAAWRV WCQKMEKLSA LLTEDLRDDR VVVMSNVRLR LGNGSAHEET IGKLSKVKLG SQLSVSDIDK ASSEALWCAL TREPDFDPKD GLPANPERHI RVNGTHVYAG DNIGLFPVSA GSIALRGGYA ELGSSFHHAR VYKITSGKKP AFAMLRVYTI DLLPYRNQDL FSVELKPQTM SMRQAEKKLR DALATGNAEY LGWLVVDDEL VVDTSKIATD QVKAVEAELG TIRRWRVDGF FGDTRLRLRP LQMSKEGIKK ESAPELSKII DRPGWLPAVN KLFSEGNVTV VRRDSLGRVR LESTAHLPVT WKVQ	(SEQ ID NO: 92)

Table 4- Cas9 orthologs		
Roseburia inulinivorans DSM 16841 WP_007889305.1	MNAEHGKEGL LIMEENFQYR IGLDIGITSV GWAVLQNNSQ DEPVRITDLG VRIFDVAENP KNGDALAAPR RDARTTRRRL RRRRHRLERI KFLLQENGLI EMDSFMERYY KGNLPDVYQL RYEGLDRKLK DEELAQVLIH IAKHRGFRST RKAETKEKEG GAVLKATTEN QKIMQEKGYR TVGEMLYLDE AFHTECLWNE KGYVLTPRNR PDDYKHTILR SMLVEEVHAI FAAQRAHGNQ KATEGLEEAY VEIMTSQRSF DMGPGLQPDG KPSPYAMEGF GDRVGKCTFE KDEYRAPKAT YTAELFVALQ KINHTKLIDE FGTGRFFSEE ERKTIIGLLL SSKELKYGTI RKKLNIDPSL KFNSLNYSAK KEGETEEERV LDTEKAKFAS MFWTYEYSKC LKDRTEEMPV GEKADLFDRI GEILTAYKND DSRSSRLKEL GLSGEEIDGL LDLSPAKYQR VSLKAMRKMQ PYLEDGLIYD KACEAAGYDF RALNDGNKKH LLKGEEINAI VNDITNPVVK RSVSQTIKVI NAIIQKYGSP QAVNIELARE MSKNFQDRTN LEKEMKKRQQ ENERAKQQII ELGKQNPTGQ DILKYRLWND QGGYCLYSGK KIPLEELFDG GYDIDHILPY SITFDDSYRN KVLVTAQENR QKGNRTPYEY FGADEKRWED YEASVRLLVR DYKKQQKLLK KNFTEEERKE FKERNLNDTK YITRVVYNMI RQNLELEPFN HPEKKKQVWA VNGAVTSYLR KRWGLMQKDR STDRHHAMDA VVIACCTDGM IHKISRYMQG RELAYSRNFK FPDEETGEIL NRDNFTREQW DEKFGVKVPL PWNSFRDELD IRLLNEDPKN FLLTHADVQR ELDYPGWMYG EEESPIEEGR YINYIRPLFV SRMPNHKVTG SAHDATIRSA RDYETRGVVI TKVPLTDLKL NKDNEIEGYY DKDSDRLLYQ ALVRQLLLHG NDGKKAFAED FHKPKADGTE GPVVRKVKIE KKQTSGVMVR GGTGIAANGE MVRIDVFREN GKYYFVPVYT ADVVRKVLPN RAATHTKPYS EWRVMDDANF VFSLYSRDLI HVKSKKDIKT NLVNGGLLLQ KEIFAYYTGA DIATASIAGF ANDSNFKFRG LGIQSLEIFE KCQVDILGNI SVVRHENRQE FH	(SEQ ID NO: 93)
Alicycliphilus denitrificans K601 WP_013517127.1	MRSLRYRLAL DLGSTSLGWA LFRLDACNRP TAVIKAGVRI FSDGRNPKDG SSLAVTRRAA RAMRRRRDRL LKRKTRMQAK LVEHGFFPAD AGKRKALEQL NPYALRAKGL QEALLPGEFA RALFHINQRR GFKSNRKTDK KDNDSGVLKK AIGQLRQQMA EQGSRTVGEY LWTRLQQGQG VRARYREKPY TTEEGKKRID KSYDLYIDRA MIEQEFDALW AAQAAFNPTL FHEAARADLK DTLLHQRPLR PVKPGRCTLL PEEERAPLAL PSTQRFRIHQ EVNHLRLLDE NLREVALTLA QRDAVVTALE TKAKLSFEQI RKLLKLSGSV QFNLEDAKRT ELKGNATSAA LARKELFGAA WSGFDEALQD EIVWQLVTEE GEGALIAWLQ THTGVDEARA QAIVDVSLPE GYGNLSRKAL ARIVPALRAA VITYDKAVQA AGFDHHSQLG FEYDASEVED LVHPETGEIR SVFKQLPYYG KALQRHVAFG SGKPEDPDEK RYGKIANPTV HIGLNQVRMV VNALIRRYGR PTEVVIELAR DLKQSREQKV EAQRRQADNQ RRNARIRRSI AEVLGIGEER VRGSDIQKWI CWEELSFDAA DRRCPYSGVQ ISAAMLLSDE VEVEHILPFS KTLDDSLNNR TVAMRQANRI KRNRTPWDAR AEFEAQGWSY EDILQRAERM PLRKRYRFAP DGYERWLGDD KDFLARALND TRYLSRVAAE YLRLVCPGTR VIPGQLTALL RGKFGLNDVL GLDGEKNRND HRHHAVDACV IGVTDQGLMQ RFATASAQAR GDGLTRLVDG MPMPWPTYRD HVERAVRHIW VSHRPDHGFE GAMMEETSYG IRKDGSIKQR RKADGSAGRE ISNLIRIHEA TQPLRHGVSA DGQPLAYKGY VGGSNYCIEI TVNDKGKWEG EVISTFRAYG VVRAGGMGRL RNPHEGQNGR KLIMRLVIGD SVRLEVDGAE RTMRIVKISG SNGQIFMAPI HEANVDARNT DKQDAFTYTS KYAGSLQKAK TRRVTISPIGEVRDPGFKG	(SEQ ID NO: 94)
Sphaerochaeta globosa str. Buddy WP_013607849.1	MSKKVSRRYE EQAQEICQRL GSRPYSIGLD LGVGSIGVAV AAYDPIKKQP SDLVFVSSRI FIPSTGAAER RQKRGQRNSL RHRANRLKFL WKLLAERNLM LSYSEQDVPD PARLRFEDAV VRANPYELRL KGLNEQLTLS ELGYALYHIA NHRGSSSVRT FLDEEKSSDD KKLEEQQAMT EQLAKEKGIS TFIEVLTAFN TNGLIGYRNS ESVKSKGVPV PTRDIISNEI DVLLQTQKQF YQEILSDEYC DRIVSAILFE NEKIVPEAGC CPYFPDEKKL PRCHFLNEER RLWEAINNAR IKMPMQEGAA KRYQSASFSD EQRHILFHIA RSGTDITPKL VQKEFPALKT SIIVLQGKEK AIQKIAGFRF RRLEEKSFWK RLSEEQKDDF FSAWTNTPDD KRLSKYLMKH LLLTENEVVD ALKTVSLIGD YGPIGKTATQ LLMKHLEDGL TYTEALERGM ETGEFQELSV WEQQSLLPYY GQILTGSTQA LMGKYWHSAF KEKRDSEGFF KPNTNSDEEK YGRIANPVVH QTLNELRKLM NELITILGAK PQEITVELAR ELKVGAEKRE DIIKQQTKQE KEAVLAYSKY CEPNNLDKRY IERFRLLEDQ AFVCPYCLEH ISVADIAAGR ADVDHIFPRD DTADNSYGNK	(SEQ ID NO: 95)

Table 4- Cas9 orthologs		
	VVAHRQCNDI KGKRTPYAAF SNTSAWGPIM HYLDETPGMW RKRRKFETNE EEYAKYLQSK GFVSRFESDN SYIAKAAKEY LRCLFNPNNV TAVGSLKGME TSILRKAWNL QGIDDLLGSR HWSKDADTSP TMRKNRDDNR HHGLDAIVAL YCSRSLVQMI NTMSEQGKRA VEIEAMIPIP GYASEPNLSF EAQRELFRKK ILEFMDLHAF VSMKTDNDAN GALLKDTVYS ILGADTQGED LVFVVKKKIK DIGVKIGDYE EVASAIRGRI TDKQPKWYPM EMKDKIEQLQ SKNEAALQKY KESLVQAAAV LEESNRKLIE SGKKPIQLSE KTISKKALEL VGGYYYLISN NKRTKTFVVK EPSNEVKGFA FDTGSNLCLD FYHDAQGKLC GEIIRKIQAM NPSYKPAYMK QGYSLYVRLY QGDVCELRAS DLTEAESNLA KTTHVRLPNA KPGRTFVIII TFTEMGSGYQ IYFSNLAKSK KGQDTSFTLT TIKNYDVRKV QLSSAGLVRY VSPLLVDKIE KDEVALCGE	
Fusobacterium nucleatum subsp. vincentii ATCC 49256 WP_005888649.1	MKKQKFSDYY LGFDIGTNSV GWCVTDLDYN VLRFNKKDMW GSRLFDEAKT AAERRVQRNS RRRLKRRKWR LNLLEEIFSD EIMKIDSNFF RRLKESSLWL EDKNSKEKFT LFNDDNYKDY DFYKQYPTIF HLRDELIKNP EKKDIRLIYL ALHSIFKSRG HFLFEGQNLK EIKNFETLYN NLISFLEDNG INKSIDKDNI EKLEKIICDS GKGLKDKEKE FKGIFNSDKQ LVAIFKLSVG SSVSLNDLFD TDEYKKEEVE KEKISFREQI YEDDKPIYYS ILGEKIELLD IAKSFYDFMV LNNILSDSNY ISEAKVKLYE EHKKDLKNLK YIIRKYNKEN YDKLFKDKNE NNYPAYIGLN KEKDKKEVVE KSRLKIDDLI KVIKGYLPKP ERIEEKDKTI FNEILNKIEL KTILPKQRIS DNGTLPYQIH EVELEKILEN QSKYYDFLNY EENGVSTKDK LLKTFKFRIP YYVGPLNSYH KDKGGNSWIV RKEEGKILPW NFEQKVDIEK SAEEFIKRMT NKCTYLNGED VIPKDSFLYS EYIILNELNK VQVNDEFLNE ENKRKIIDEL FKENKKVSEK KFKEYLLVNQ IANRTVELKG IKDSFNSNYV SYIKFKDIFG EKLNLDIYKE ISEKSILWKC LYGDDKKIFE KKIKNEYGDI LNKDEIKKIN SFKFNTWGRL SEKLLTGIEF INLETGECYS SVMEALRRTN YNLMELLSSK FTLQESIDNE NKEMNEVSYR DLIEESYVSP SLKRAILQTL KIYEEIKKIT GRVPKKVFIE MARGGDESMK NKKIPARQEQ LKKLYDSCGN DIANFSIDIK EMKNSLSSYD NNSLRQKKLY LYYLQFGKCM YTGREIDLDR LLQNNDTYDI DHIYPRSKVI KDDSFDNLVL VLKNENAEKS NEYPVKKEIQ EKMKSFWRFL KEKNFISDEK YKRLTGKDDF ELRGFMARQL VNVRQTTKEV GKILQQIEPE IKIVYSKAEI ASSFREMFDF IKVRELNDTH HAKDAYLNIV AGNVYNTKFT EKPYRYLQEI KENYDVKKIY NYDIKNAWDK ENSLEIVKKN MEKNTVNITR FIKEEKGELF NLNPIKKGET SNEIISIKPK LYDGKDNKLN EKYGYYTSLK AAYFIYVEHE KKNKKVKTFE RITRIDSTLI KNEKNLIKYL VSQKKLLNPK IIKKIYKEQT LIIDSYPYTF TGVDSNKKVE LKNKKQLYLE KKYEQILKNA LKFVEDNQGE TEENYKFIYL KKRNNNEKNE TIDAVKERYN IEFNEMYDKF LEKLSSKDYK NYINNKLYTN FLNSKEKFKK LKLWEKSLIL REFLKIFNKN TYGKYEIKDS QTKEKLFSFP EDTGRIRLGQ SSLGNNKELL EESVTGLFVK KIKL	(SEQ ID NO: 96)
Pasteurella multocida subsp. multocida str. Pm70 WP_010907033.1	MQTTNLSYIL GLDLGIASVG WAVVEINENE DPIGLIDVGV RIFERAEVPK TGESLALSRR LARSTRRLIR RRAHRLLLAK RFLKREGILS TIDLEKGLPN QAWELRVAGL ERRLSAIEWG AVLLHLIKHR GYLSKRKNES QTNNKELGAL LSGVAQNHQL LQSDDYRTPA ELALKKFAKE EGHIRNQRGA YTHTFNRLDL LAELNLLFAQ QHQFGNPHCK EHIQQYMTEL LMWQKPALSG EAILKMLGKC THEKNEFKAA KHTYSAERFV WLTKLNNLRI LEDGAERALN EEERQLLINH PYEKSKLTYA QVRKLLGLSE QAIFKHLRYS KENAESATFM ELKAWHAIRK ALENQGLKDT WQDLAKKPDL LDEIGTAFSL YKTDEDIQQY LTNKVPNSVI NALLVSLNFD KFIELSLKSL RKILPLMEQG KRYDQACREI YGHHYGEANQ KTSQLLPAIP AQEIRNPVVL RTLSQARKVI NAIIRQYGSP ARVHIETGRE LGKSFKERRE IQKQQEDNRT KRESAVQKFK ELFSDFSSEP KSKDILKFRL YEQQHGKCLY SGKEINIHRL NEKGYVEIDH ALPFSRTWDD SFNNKVLVLA SENQNKGNQT PYEWLQGKIN SERWKNFVAL VLGSQCSAAK KQRLLTQVID DNKFIDRNLN DTRYIARFLS NYIQENLLLV GKNKKNVFTP NGQITALLRS RWGLIKAREN NNRHHALDAI VVACATPSMQ QKITRFIRFK EVHPYKIENR YEMVDQESGE IISPHFPEPW AYFRQEVNIR VFDNHPDTVL KEMLPDRPQA NHQFVQPLFV SRAPTRKMSG QGHMETIKSA KRLAEGISVL RIPLTQLKPN LLENMVNKER EPALYAGLKA RLAEFNQDPA KAFATPFYKQ GGQQVKAIRV EQVQKSGVLV RENNGVADNA SIVRTDVFIK NNKFFLVPIY TWQVAKGILP	(SEQ ID NO: 97)

Table 4- Cas9 orthologs		
	NKAIVAHKNE DEWEEMDEGA KFKFSLFPND LVELKTKKEY FFGYYIGLDR ATGNISLKEH DGEISKGKDG VYRVGVKLAL SFEKYQVDEL GKNRQICRPQ QRQPVR	
Alcanivorax pacificus W11-5 WP_008738269.1	MRYRVGLDLG TASVGAAVFS MDEQGNPMEL IWHYERLFSE PLVPDMGQLK PKKAARRLAR QQRRQIDRRA SRLRRIAIVS RRLGIAPGRN DSGVHGNDVP TLRAMAVNER IELGQLRAVL LRMGKKRGYG GTFKAVRKVG EAGEVASGAS RLEEEMVALA SVQNKDSVTV GEYLAARVEH GLPSKLKVAA NNEYYAPEYA LFRQYLGLPA IKGRPDCLPN MYALRHQIEH EFERIWATQS QFHDVMKDHG VKEEIRNAIF FQRPLKSPAD KVGRCSLQTN LPRAPRAQIA AQNFRIEKQM ADLRWGMGRR AEMLNDHQKA VIRELLNQQK ELSFRKIYKE LERAGCPGPE GKGLNMDRAA LGGRDDLSGN TTLAAWRKLG LEDRWQELDE VTQIQVINFL ADLGSPEQLD TDDWSCRFMG KNGRPRNFSD EFVAFMNELR MTDGFDRLSK MGFEGGRSSY SIKALKALTE WMIAPHWRET PETHRVDEEA AIRECYPESL ATPAQGGRQS KLEPPPLTGN EVVDVALRQV RHTINMMIDD LGSVPAQIVV EMAREMKGGV TRRNDIEKQN KRFASERKKA AQSIEENGKT PTPARILRYQ LWIEQGHQCP YCESNISLEQ ALSGAYTNFE HILPRTLTQI GRKRSELVLA HRECNDEKGN RTPYQAFGHD DRRWRIVEQR ANALPKKSSR KTRLLLLKDF EGEALTDESI DEFADRQLHE SSWLAKVTTQ WLSSLGSDVY VSRGSLTAEL RRRWGLDTVI PQVRFESGMP VVDEEGAEIT PEEFEKFRLQ WEGHRVTREM RTDRRPDKRI DHRHHLVDAI VTALTSRSLY QQYAKAWKVA DEKQRHGRVD VKVELPMPIL TIRDIALEAV RSVRISHKPD RYPDGRFFEA TAYGIAQRLD ERSGEKVDWL VSRKSLTDLA PEKKSIDVDK VRANISRIVG EAIRLHISNI FEKRVSKGMT PQQALREPIE FQGNILRKVR CFYSKADDCV RIEHSSRRGH HYKMLLNDGF AYMEVPCKEG ILYGVPNLVR PSEAVGIKRA PESGDFIRFY KGDTVKNIKT GRVYTIKQIL GDGGGKLILT PVTETKPADL LSAKWGRLKV GGRNIHLLRL CAE	(SEQ ID NO: 98)
Mycoplasma mobile 163K AAT27519.1	MYFYKNKENK LNKKVVLGLD LGIASVGWCL TDISQKEDNK FPIILHGVRL FETVDDSDDK LLNETRRKKR GQRRRNRRLF TRKRDFIKYL IDNNIIELEF DKNPKILVRN FIEKYINPFS KNLELKYKSV TNLPIGFHNL RKAAINEKYK LDKSELIVLL YFYLSLRGAF FDNPEDTKSK EMNKNEIEIF DKNESIKNAE FPIDKIIEFY KISGKIRSTI NLKFGHQDYL KEIKQVFEKQ NIDFMNYEKF AMEEKSFFSR IRNYSEGPGN EKSFSKYGLY ANENGNPELI INEKGQKIYT KIFKTLWESK IGKCSYDKKL YRAPKNSFSA KVFDITNKLT DWKHKNEYIS ERLKRKILLS RFLNKDSKSA VEKILKEENI KFENLSEIAY NKDDNKINLP IINAYHSLTT IFKKHLINFE NYLISNENDL SKLMSFYKQQ SEKLFVPNEK GSYEINQNNN VLHIFDAISN ILNKFSTIQD RIRILEGYFE FSNLKKDVKS SEIYSEIAKL REFSGTSSLS FGAYYKFIPN LISEGSKNYS TISYEEKALQ NQKNNFSHSN LFEKTWVEDL IASPTVKRSL RQTMNLLKEI FKYSEKNNLE IEKIVVEVTR SSNNKHERKK IEGINKYRKE KYEELKKVYD LPNENTTLLK KLWLLRQQQG YDAYSLRKIE ANDVINKPWN YDIDHIVPRS ISFDDSFSNL VIVNKLDNAK KSNDLSAKQF IEKIYGIEKL KEAKENWGNW YLRNANGKAF NDKGKFIKLY TIDNLDEFDN SDFINRNLSD TSYITNALVN HLTFSNSKYK YSVVSVNGKQ TSNLRNQIAF VGIKNNKETE REWKRPEGFK SINSNDFLIR EEGKNDVKDD VLIKDRSFNG HHAEDAYFIT IISQYFRSFK RIERLNVNYR KETRELDDLE KNNIKFKEKA SFDNFLLINA LDELNEKLNQ MRFSRMVITK KNTQLFNETL YSGKYDKGKN TIKKVEKLNL LDNRTDKIKK IEEFFDEDKL KENELTKLHI FNHDKNLYET LKIIWNEVKI EIKNKNLNEK NYFKYFVNKK LQEGKISFNE WVPILDNDFK IIRKIRYIKF SSEEKETDEI IFSQSNFLKI DQRQNFSFHN TLYWVQIWVY KNQKDQYCFI SIDARNSKFE KDEIKINYEK LKTQKEKLQI INEEPILKIN KGDLFENEEK ELFYIVGRDE KPQKLEIKYI LGKKIKDQKQ IQKPVKKYFP NWKKVNLTYM GEIFKK	(SEQ ID NO: 99)
gamma proteobacterium HTCC5015	MTKNYISPIA IDLGAKFTGV ALYQYLEGAD CTQEVAKGLL VDDRGNVTWS QEGRRGKRHQ VRGYKRRKMA KRLLWLILDS EYGIKREEVT EPLLKFINGL LNRRGYTYIS EEVDEESMNV SPLPFSEMMP DYFNSSAPLL EQLAKLLSDK NKLVRFRAEG KIPSNKNEFK KLLDTALDGK YKDEKKELSE AWGNILIASE	(SEQ ID NO: 100)

Table 4- Cas9 orthologs		
WP_008284239.1	NVLKSTVDGH KSRSEYLANI KEDIKSNEEL EKQISSKEID GFYNLVGHLS NFQLRLLRKY FNDPNMSGVS YWDEKRLEKY FYQWVQGWHT KGGTDEAEKK NIILKTKGAP LLKTLKSLSA DLTIPPYEDQ NNRRPPKCQS VLLSDEKLTM HYPKWKEWVG QLVKQNDNAY LNENVTLANA LHRIVERSRS IDPYQLRLLI SITDAEKRND LAGYKRLKLS LGSEVDEFLL LVKNIVDETK EAREGLWFET ENKLFFKCGK TPPRKEKLKS TLLSAVLGKN LSDDEQSSFI EEFWKSGTPK IERRNVRGWC RLASQVQKTY GVYLKEYGLQ QLHKLEAGKK LDDKPLALLY KNSGLIASKI GEALNIEPDE VSRFASPHSL AQIFNIIEGD VAGFNKTCRA CTYENIWRMQ EEKVESLLTN QLLSEIHGER KVPLKSAMCT RLSADSTRPF DGQMASIIEH IARKIAQHKI AQINDVPKEF SIDIPIIIES NQFSFTAELE EIKRGRGSAK AKKAKELGEK SKAGWVSKTE RIKTSSEGIC PYTGAPLGGS GEIDHIIPRS LTGRTKKTVF NSEANLIYCS SKGNHDKGNR VYVIEQLNDK YLKKQFSTSD VNLIKKKIKT TIQRFTEGGE KLRSFSELSR EDQKAFRHAL FVPELKSEVT SLLAVKNITR VNGTQAWLAK KIASLLAEHL DKQGRDYTLS AHQIDPWSVS KQRKMLASAE PIWAKKDPQP AASHVVDAVC TFLEALEQPH TASRLKTISS TSFEKTGWRS ALIPDLIKVD ALDRRPKYRR YNIGSTSLFK DGIYAERFLP ILIDENGLMA GYDIDNSLKA KGADVVFESL SPFLLFKGEE VGAQSLSDWQ ERIDGRYLYM SIDKVKAFDY LQEKVGEKDI AAELLNSIHF TQRKTELRAK FSDDSGKKMK TLDAIRKSLK LTVTVNEIGK RKEKCGFSGT IGIPAKSAWE NLLDEPLLET YWGTKMPPQE IWEKVYRKHF PRNIPNQAHR KVRKDFSLPV VDSVSGGFRV KRKTPNGYNY QLLAIDGYSA VGFKKEGDNV DFKSPALVPQ IAESKSVTPI SSELVHLDKN EIVYFDEWRK IDISDSDLKQ FVSSLELAPG SQNRFYIRFT VDEDQFERHF KSALRVNGIQ DLDTVNKTFD WNREIPSLLI PPRSNLFLLE TGQKITFEYI ANGANAEVKK AYSLRRA	
Planococcus antarcticus DSM 14505 ANU10858.1	MKNYTIGLDI GVASVGWVCI DENYKILNYN NRHAFGVHEF ESAESAAGRR LKRGMRRRYN RRKKRLQLLQ SLFDSYITDS GFFSKTDSQH FWKNNNEFEN RSLTEVLSSL RISSRKYPTI YHLRSDLIES NKKMDLRLVY LALHNLVKYR GHFLQEGNWS EAASAEGMDD QLLELVTRYA ELENLSPLDL SESQWKAAET LLLNRNLTKT DQSKELTAMF GKEYEPFCKL VAGLGVSLHQ LFPSSEQALA YKETKTKVQL SNENVEEVME LLLEEESALL EAVQPFYQQV VLYELLKGET YVAKAKVSAF KQYQKDMASL KNLLDKTFGE KVYRSYFISD KNSQREYQKS HKVEVLCKLD QFNKEAKFAE TFYKDLKKLL EDKSKTSIGT TEKDEMLRII KAIDSNQFLQ KQKGIQNAAI PHQNSLYEAE KILRNQQAHY PFITTEWIEK VKQILAFRIP YYIGPLVKDT TQSPFSWVER KGDAPITPWN FDEQIDKAAS AEAFISRMRK TCTYLKGQEV LPKSSLTYER FEVLNELNGI QLRTTGAESD FRHRLSYEMK CWIIDNVFKQ YKTVSTKRLL QELKKSPYAD ELYDEHTGEI KEVFGTQKEN AFATSLSGYI SMKSILGAVV DDNPAMTEEL IYWIAVFEDR EILHLKIQEK YPSITDVQRQ KLALVKLPGW GRFSRLLIDG LPLDEQGQSV LDHMEQYSSV FMEVLKNKGF GLEKKIQKMN QHQVDGTKKI RYEDIEELAG SPALKRGIWR SVKIVEELVS IFGEPANIVL EVAREDGEKK RTKSRKDQWE ELTKTTLKND PDLKSFIGEI KSQGDQRFNE QRFWLYVTQQ GKCLYTGKAL DIQNLSMYEV DHILPQNFVK DDSLDNLALV MPEANQRKNQ VGQNKMPLEI IEANQQYAMR TLWERLHELK LISSGKLGRL KKPSFDEVDK DKFIARQLVE TRQIIKHVRD LLDERFSKSD IHLVKAGIVS KFRRFSEIPK IRDYNNKHHA MDALFAAALI QSILGKYGKN FLAFDLSKKD RQKQWRSVKG SNKEFFLFKN FGNLRLQSPV TGEEVSGVEY MKHVYFELPW QTTKMTQTGD GMFYKESIFS PKVKQAKYVS PKTEKFVHDE VKNHSICLVE FTFMKKEKEV QETKFIDLKV IEHHQFLKEP ESQLAKFLAE KETNSPIIHA RIIRTIPKYQ KIWIEHFPYY FISTRELHNA RQFEISYELM EKVKQLSERS SVEELKIVFG LLIDQMNDNY PIYTKSSIQD RVQKFVDTQL YDFKSFEIGF EELKKAVAAN AQRSDTFGSR ISKKPKPEEV AIGYESITGL KYRKPRSVVG TKR	(SEQ ID NO: 101)
Prevotella sp. C561 WP_009013303.1	MTQKVLGLDL GTNSIGSAVR NLDLSDDLQW QLEFFSSDIF RSSVNKESNG REYSLAAQRS AHRRSRGLNE VRRRRLWATL NLLIKHGFCP MSSESLMRWC TYDKRKGLFR EYPIDDKDFN AWILLDFNGD GRPDYSSPYQ LRRELVTRQF DFEQPIERYK LGRALYHIAQ HRGFKSSKGE TLSQQETNSK PSSTDEIPDV AGAMKASEEK LSKGLSTYMK EHNLLTVGAA FAQLEDEGVR VRNNNDYRAI RSQFQHEIET IFKFQQGLSV ESELYERLIS EKKNVGTIFY KRPLRSQRGN VGKCTLERSK PRCAIGHPLF EKFRAWTLIN NIKVRMSVDT LDEQLPMKLR	(SEQ ID NO: 102)

Table 4- Cas9 orthologs		
	LDLYNECFLA FVRTEFKFED IRKYLEKRLG IHFSYNDKTI NYKDSTSVAG CPITARFRKM LGEEWESFRV EGQKERQAHS KNNISFHRVS YSIEDIWHFC YDAEEPEAVL AFAQETLRLE RKKAEELVRI WSAMPQGYAM LSQKAIRNIN KILMLGLKYS DAVILAKVPE LVDVSDEELL SIAKDYYLVE AQVNYDKRIN SIVNGLIAKY KSVSEEYRFA DHNYEYLLDE SDEKDIIRQI ENSLGARRWS LMDANEQTDI LQKVRDRYQD FFRSHERKFV ESPKLGESFE NYLTKKFPMV EREQWKKLYH PSQITIYRPV SVGKDRSVLR LGNPDIGAIK NPTVLRVLNT LRRRVNQLLD DGVISPDETR VVVETARELN DANRKWALDT YNRIRHDENE KIKKILEEFY PKRDGISTDD IDKARYVIDQ REVDYFTGSK TYNKDIKKYK FWLEQGGQCM YTGRTINLSN LFDPNAFDIE HTIPESLSFD SSDMNLTLCD AHYNRFIKKN HIPTDMPNYD KAITIDGKEY PAITSQLQRW VERVERLNRN VEYWKGQARR AQNKDRKDQC MREMHLWKME LEYWKKKLER FTVTEVTDGF KNSQLVDTRV ITRHAVLYLK SIFPHVDVQR GDVTAKFRKI LGIQSVDEKK DRSLHSHHAI DATTLTIIPV SAKRDRMLEL FAKIEEINKM LSFSGSEDRT GLIQELEGLK NKLQMEVKVC RIGHNVSEIG TFINDNIIVN HHIKNQALTP VRRRLRKKGY IVGGVDNPRW QTGDALRGEI HKASYYGAIT QFAKDDEGKV LMKEGRPQVN PTIKFVIRRE LKYKKSAADS GFASWDDLGK AIVDKELFAL MKGQFPAETS FKDACEQGIY MIKKGKNGMP DIKLHHIRHV RCEAPQSGLK IKEQTYKSEK EYKRYFYAAV GDLYAMCCYT NGKIREFRIY SLYDVSCHRK SDIEDIPEFI TDKKGNRLML DYKLRTGDMI LLYKDNPAEL YDLDNVNLSR RLYKINRFES QSNLVLMTHH LSTSKERGRS LGKTVDYQNL PESIRSSVKS LNFLIMGENR DFVIKNGKII FNHR	
Alicyclobacillus hesperidum URH17-3-68 WP_006446566.1	MAYRLGLDIG ITSVGWAVVA LEKDESGLKP VRIQDLGVRI FDKAEDSKTG ASLALPRREA RSARRRTRRR RHRLWRVKRL LEQHGILSME QIEALYAQRT SSPDVYALRV AGLDRCLIAE EIARVLIHIA HRRGFQSNRK SEIKDSDAGK LLKAVQENEN LMQSKGYRTV AEMLVSEATK TDAEGKLVHG KKHGYVSNVR NKAGEYRHTV SRQAIVDEVR KIFAAQRALG NDVMSEELED SYLKILCSQR NFDDGPGGDS PYGHGSVSPD GVRQSIYERM VGSCTFETGE KRAPRSSYSF ERFQLLTKVV NLRIYRQQED GGRYPCELTQ TERARVIDCA YEQTKITYGK LRKLLDMKDT ESFAGLTYGL NRSRNKTEDT VFVEMKFYHE VRKALQRAGV FIQDLSIETL DQIGWILSVW KSDDNRRKKL STLGLSDNVI EELLPLNGSK FGHLSLKAIR KILPFLEDGY SYDVACELAG YQFQGKTEYV KQRLLPPLGE GEVTNPVVRR ALSQAIKVVN AVIRKHGSPE SIHIELAREL SKNLDERRKI EKAQKENQKN NEQIKDEIRE ILGSAHVTGR DIVKYKLFKQ QQEFCMYSGE KLDVTRLFEP GYAEVDHIIP YGISFDDSYD NKVLVKTEQN RQKGNRTPLE YLRDKPEQKA KFIALVESIP LSQKKKNHLL MDKRAIDLEQ EGFRERNLSD TRYITRALMN HIQAWLLFDE TASTRSKRVV CVNGAVTAYM RARWGLTKDR DAGDKHHAAD AVVVACIGDS LIQRVTKYDK FKRNALADRN RYVQQVSKSE GITQYVDKET GEVFTWESFD ERKFLPNEPL EPWPFFRDEL LARLSDDPSK NIRAIGLLTY SETEQIDPIF VSRMPTRKVT GAAHKETIRS PRIVKVDDNK GTEIQVVVSK VALTELKLTK DGEIKDYFRP EDDPRLYNTL RERLVQFGGD AKAAFKEPVY KISKDGSVRT PVRKVKIQEK LTLGVPVHGG RGIAENGGMV RIDVFAKGGK YYFVPIYVAD VLKRELPNRL ATAHKPYSEW RVVDDSYQFK FSLYPNDAVM IKPSREVDIT YKDRKEPVGC RIMYFVSANI ASASISLRTH DNSGELEGLG IQGLEVFEKY VVGPLGDTHP VYKERRMPFR VERKMN	(SEQ ID NO: 103)
Lactobacillus rhamnosus GG WP_014569977.1	MTKLNQPYGI GLDIGSNSIG FAVVDANSHL LRLKGETAIG ARLFREGQSA ADRRGSRTTR RRLSRTRWRL SFLRDFFAPH ITKIDPDFFL RQKYSEISPK DKDRFKYEKR LFNDRTDAEF YEDYPSMYHL RLHLMTHTHK ADPREIFLAI HHILKSRGHF LTPGAAKDFN TDKVDLEDIF PALTEAYAQV YPDLELTFDL AKADDFKAKL LDEQATPSDT QKALVNLLLS SDGEKEIVKK RKQVLTEFAK AITGLKTKFN LALGTEVDEA DASNWQFSMG QLDDKWSNIE TSMTDQGTEI FEQIQELYRA RLLNGIVPAG MSLSQAKVAD YGQHKEDLEL FKTYLKKLND HELAKTIRGL YDRYINGDDA KPFLREDFVK ALTKEVTAHP NEVSEQLLNR MGQANFMLKQ RTKANGAIPI QLQQRELDQI IANQSKYYDW LAAPNPVEAH RWKMPYQLDE LLNFHIPYYV GPLITPKQQA ESGENVFAWM VRKDPSGNIT PYNFDEKVDR EASANTFIQR MKTTDTYLIG EDVLPKQSLL YQKYEVLNEL NNVRINNECL GTDQKQRLIR EVFERHSSVT IKQVADNLVA HGDFARRPEI RGLADEKRFL SSLSTYHQLK EILHEAIDDP TKLLDIENII TWSTVFEDHT	(SEQ ID NO: 104)

Table 4- Cas9 orthologs

	IFETKLAEIE WLDPKKINEL SGIRYRGWGQ FSRKLLDGLK LGNGHTVIQE LMLSNHNLMQ ILADETLKET MTELNQDKLK TDDIEDVIND AYTSPSNKKA LRQVLRVVED IKHAANGQDP SWLFIETADG TGTAGKRTQS RQKQIQTVYA NAAQELIDSA VRGELEDKIA DKASFTDRLV LYFMQGGRDI YTGAPLNIDQ LSHYDIDHIL PQSLIKDDSL DNRVLVNATI NREKNNVFAS TLFAGKMKAT WRKWHEAGLI SGRKLRNLML RPDEIDKFAK GFVARQLVET RQIIKLTEQI AAAQYPNTKI IAVKAGLSHQ LREELDFPKN RDVNHYHHAF DAFLAARIGT YLLKRYPKLA PFFTYGEFAK VDVKKFREFN FIGALTHAKK NIIAKDTGEI VWDKERDIRE LDRIYNFKRM LITHEVYFET ADLFKQTIYA AKDSKERGGS KQLIPKKQGY PTQVYGGYTQ ESGSYNALVR VAEADTTAYQ VIKISAQNAS KIASANLKSR EKGKQLLNEI VVKQLAKRRK NWKPSANSFK IVIPRFGMGT LFQNAKYGLF MVNSDTYYRN YQELWLSREN QKLLKKLFSI KYEKTQMNHD ALQVYKAIID QVEKFFKLYD INQFRAKLSD AIERFEKLPI NTDGNKIGKT ETLRQILIGL QANGTRSNVK NLGIKTDLGL LQVGSGIKLD KDTQIVYQSP SGLFKRRIPL ADL	
Enterococcus faecalis TX0012 WP_002408901.1 EFT93846.1	MYSIGLDLGI SSVGWSVIDE RTGNVIDLGV RLFSAKNSEK NLERRTNRGG RRLIRRKTNR LKDAKKILAA VGFYEDKSLK NSCPYQLRVK GLTEPLSRGE IYKVTLHILK KRGISYLDEV DTEAAKESQD YKEQVRKNAQ LLTKYTPGQI QLQRLKENNR VKTGINAQGN YQLNVFKVSA YANELATILK TQQAFYPNEL TDDWIALFVQ PGIAEEAGLI YRKRPYYHGP GNEANNSPYG RWSDFQKTGE PATNIFDKLI GKDFQGELRA SGLSLSAQQY NLLNDLTNLK IDGEVPLSSE QKEYILTELM TKEFTRFGVN DVVKLLGVKK ERLSGWRLDK KGKPEIHTLK GYRNWRKIFA EAGIDLATLP TETIDCLAKV LTLNTEREGI ENTLAFELPE LSESVKLLVL DRYKELSQSI STQSWHRFSL KTLHLLIPEL MNATSEQNTL LEQFQLKSDV RKRYSEYKKL PTKDVLAEIY NPTVNKTVSQ AFKVIDALLV KYGKEQIRYI TIEMPRDDNE EDEKKRIKEL HAKNSQRKND SQSYFMQKSG WSQEKFQTTI QKNRRFLAKL LYYYEQDGIC AYTGLPISPE LLVSDSTEID HIIPISISLD DSINNKVLVL SKANQVKGQQ TPYDAWMDGS FKKINGKFSN WDDYQKWVES RHFSHKKENN LLETRNIFDS EQVEKFLARN LNDTRYASRL VLNTLQSFFT NQETKVRVVN GSFTHTLRKK WGADLDKTRE THHHHAVDAT LCAVTSFVKV SRYHYAVKEE TGEKVMREID FETGEIVNEM SYWEFKKSKK YERKTYQVKW PNFREQLKPV NLHPRIKFSH QVDRKANRKL SDATIYSVRE KTEVKTLKSG KQKITTDEYT IGKIKDIYTL DGWEAFKKKQ DKLLMKDLDE KTYERLLSIA ETTPDFQEVE EKNGKVKRVK RSPFAVYCEE NDIPAIQKYA KKNNGPLIRS LKYYDGKLNK HINITKDSQG RPVEKTKNGR KVTLQSLKPY RYDIYQDLET KAYYTVQLYY SDLRFVEGKY GITEKEYMKK VAEQTKGQVV RFCFSLQKND GLEIEWKDSQ RYDVRFYNFQ SANSINFKGL EQEMMPAENQ FKQKPYNNGA INLNIAKYGK EGKKLRKFNT DILGKKHYLF YEKEPKNIIK	(SEQ ID NO: 105)
Candidatus Puniceispirillum marinum IMCC1322 WP_013047413.1	MRRLGLDLGT NSIGWCLLDL GDDGEPVSIF RTGARIFSDG RDPKSLGSLK ATRREARLTR RRRDRFIQRQ KNLINALVKY GLMPADEIQR QALAYKDPYP IRKKALDEAI DPYEMGRAIF HINQRRGFKS NRKSADNEAG VVKQSIADLE MKLGEAGART IGEFLADRQA TNDTVRARRL SGTNALYEFY PDRYMLEQEF DTLWAKQAAF NPSLYIEAAR ERLKEIVFFQ RKLKPQEVGR CIFLSDEDRI SKALPSFQRF RIYQELSNLA WIDHDGVAHR ITASLALRDH LFDELEHKKK LTFKAMRAIL RKQGVVDYPV GFNLESDNRD HLIGNLTSCI MRDAKKMIGS AWDRLDEEEQ DSFILMLQDD QKGDDEVRSI LTQQYGLSDD VAEDCLDVRL PDGHGSLSKK AIDRILPVLR DQGLIYYDAV KEAGLGEANL YDPYAALSDK LDYYGKALAG HVMGASGKFE DSDEKRYGTI SNPTVHIALN QVRAVVNELI RLHGKPDEVV IEIGRDLPMG ADGKRELERF QKEGRAKNER ARDELKKLGH IDSRESRQKF QLWEQLAKEP VDRCCPFTGK MMSISDLFSD KVEIEHLLPF SLTLDDSMAN KTVCFRQANR DKGNRAPFDA FGNSPAGYDW QEILGRSQNL PYAKRWRFLP DAMKRFEADG GFLERQLNDT RYISRYTTEY ISTIIPKNKI WVVTGRLTSL LRGFWGLNSI LRGHNTDDGT PAKKSRDDHR HHAIDAIVVG MTSRGLLQKV SKAARRSEDL DLTRLFEGRI DPWDGFRDEV KKHIDAIIVS HRPRKKSQGA LHNDTAYGIV EHAENGASTV VHRVPITSLG KQSDIEKVRD PLIKSALLNE TAGLSGKSFE NAVQKWCADN SIKSLRIVET VSIIPITDKE GVAYKGYKGD GNAYMDIYQD PTSSKWKGEI VSRFDANQKG FIPSWQSQFP TARLIMRLRI NDLLKLQDGE IEEIYRVQRL SGSKILMAPH TEANVDARDR	(SEQ ID NO: 106)

Table 4- Cas9 orthologs

	DKNDTFKLTS KSPGKLQSAS ARKVHISPTG LIREG	
Oenococcus kitaharae DSM 17330 EHN59352.1	MARDYSVGLD IGTSSVGWAA IDNKYHLIRA KSKNLIGVRL FDSAVTAEKR RGYRTTRRRL SRRHWRLRLL NDIFAGPLTD FGDENFLARL KYSWVHPQDQ SNQAHFAAGL LFDSKEQDKD FYRKYPTIYH LRLALMNDDQ KHDLREVYLA IHHLVKYRGH FLIEGDVKAD SAFDVHTFAD AIQRYAESNN SDENLLGKID EKKLSAALTD KHGSKSQRAE TAETAFDILD LQSKKQIQAI LKSVVGNQAN LMAIFGLDSS AISKDEQKNY KFSFDDADID EKIADSEALL SDTEFEFLCD LKAAFDGLTL KMLLGDDKTV SAAMVRRFNE HQKDWEYIKS HIRNAKNAGN GLYEKSKKFD GINAAYLALQ SDNEDDRKKA KKIFQDEISS ADIPDDVKAD FLKKIDDDQF LPIQRTKNNG TIPHQLHRNE LEQIIEKQGI YYPFLKDTYQ ENSHELNKIT ALINFRVPYY VGPLVEEEQK IADDGKNIPD PTNHWMVRKS NDTITPWNLS QVVDLDKSGR RFIERLTGTD TYLIGEPTLP KNSLLYQKFD VLQELNNIRV SGRRLDIRAK QDAFEHLFKV QKTVSATNLK DFLVQAGYIS EDTQIEGLAD VNGKNFNNAL TTYNYLVSVL GREFVENPSN EELLEEITEL QTVFEDKKVL RRQLDQLDGL SDHNREKLSR KHYTGWGRIS KKLLTTKIVQ NADKIDNQTF DVPRMNQSII DTLYNTKMNL MEIINNAEDD FGVRAWIDKQ NTTDGDEQDV YSLIDELAGP KEIKRGIVQS FRILDDITKA VGYAPKRVYL EFARKTQESH LTNSRKNQLS TLLKNAGLSE LVTQVSQYDA AALQNDRLYL YFLQQGKDMY SGEKLNLDNL SNYDIDHIIP QAYTKDNSLD NRVLVSNITN RRKSDSSNYL PALIDKMRPF WSVLSKQGLL SKHKFANLTR TRDFDDMEKE RFIARSLVET RQIIKNVASL IDSHFGGETK AVAIRSSLTA DMRRYVDIPK NRDINDYHHA FDALLFSTVG QYTENSGLMK KGQLSDSAGN QYNRYIKEWI HAARLNAQSQ RVNPFGFVVG SMRNAAPGKL NPETGEITPE ENADWSIADL DYLHKVMNFR KITVTRRLKD QKGQLYDESR YPSVLHDAKS KASINFDKHK PVDLYGGFSS AKPAYAALIK FKNKFRLVNV LRQWTYSDKN SEDYILEQIR GKYPKAEMVL SHIPYGQLVK KDGALVTISS ATELHNFEQL WLPLADYKLI NTLLKTKEDN LVDILHNRLD LPEMTIESAF YKAFDSILSF AFNRYALHQN ALVKLQAHRD DFNALNYEDK QQTLERILDA LHASPASSDL KKINLSSGFG RLFSPSHFTL ADTDEFIFQS VTGLFSTQKT VAQLYQETK	(SEQ ID NO: 107)
Helicobacter mustelae 12198 WP_013022389.1	MIRTLGIDIG IASIGWAVIE GEYTDKGLEN KEIVASGVRV FTKAENPKNK ESLALPRTLA RSARRRNARK KGRIQQVKHY LSKALGLDLE CFVQGEKLAT LFQTSKDFLS PWELRERALY RVLDKEELAR VILHIAKRRG YDDITYGVED NDSGKIKKAI AENSKRIKEE QCKTIGEMMY KLYFQKSLNV RNKKESYNRC VGRSELREEL KTIFQIQQEL KSPWVNEELI YKLLGNPDAQ SKQEREGLIF YQRPLKGFGD KIGKCSHIKK GENSPYRACK HAPSAEEFVA LTKSINFLKN LTNRHGLCFS QEDMCVYLGK ILQEAQKNEK GLTYSKLKLL LDLPSDFEFL GLDYSGKNPE KAVFLSLPST FKLNKITQDR KTQDKIANIL GANKDWEAIL KELESLQLSK EQIQTIKDAK LNFSKHINLS LEALYHLLPL MREGKRYDEG VEILQERGIF SKPQPKNRQL LPPLSELAKE ESYFDIPNPV LRRALSEFRK VVNALLEKYG GFHYFHIELT RDVCKAKSAR MQLEKINKKN KSENDAASQL LEVLGLPNTY NNRLKCKLWK QQEEYCLYSG EKITIDHLKD QRALQIDHAF PLSRSLDDSQ SNKVLCLTSS NQEKSNKTPY EWLGSDEKKW DMYVGRVYSS NFSPSKKRKL TQKNFKERNE EDFLARNLVD TGYIGRVTKE YIKHSLSFLP LPDGKKEHIR IISGSMTSTM RSFWGVQEKN RDHHLHHAQD AIIIACIEPS MIQKYTTYLK DKETHRLKSH QKAQILREGD HKLSLRWPMS NFKDKIQESI QNIIPSHHVS HKVTGELHQE TVRTKEFYYQ AFGGEEGVKK ALKFGKIREI NQGIVDNGAM VRVDIFKSKD KGKFYAVPIY TYDFAIGKLP NKAIVQGKKN GIIKDWLEMD ENYEFCFSLF KNDCIKIQTK EMQEAVLAIY KSTNSAKATI ELEHLSKYAL KNEDEEKMFT DTDKEKNKTM TRESCGIQGL KVFQKVKLSV LGEVLEHKPR NRQNIALKTT PKHV	(SEQ ID NO: 108)
Bradyrhizobium sp. BTAi1 WP_012044026.1	MKRTSLRAYR LGVDLGANSL GWFVVWLDDH GQPEGLGPGG VRIFPDGRNP QSKQSNAAGR RLARSARRRR DRYLQRRGKL MGLLVKHGLM PADEPARKRL ECLDPYGLRA KALDEVLPLH HVGRALFHLN QRRGLFANRA IEQGDKDASA IKAAAGRLQT SMQACGARTL GEFLNRRHQL RATVRARSPV GGDVQARYEF YPTRAMVDAE FEAIWAAQAP HHPTMTAEAH DTIREAIFSQ RAMKRPSIGK CSLDPATSQD DVDGFRCAWS HPLAQRFRIW QDVRNLAVVE TGPTSSRLGK EDQDKVARAL LQTDQLSFDE IRGLLGLPSD ARFNLESDRR DHLKGDATGA	(SEQ ID NO: 109)

Table 4- Cas9 orthologs		
	ILSARRHFGP AWHDRSLDRQ IDIVALLESA LDEAAIIASL GTTHSLDEAA AQRALSALLP DGYCRLGLRA IKRVLPLMEA GRTYAEAASA AGYDHALLPG GKLSPTGYLP YYGQWLQNDV VGSDDERDTN ERRWGRLPNP TVHIGIGQLR RVVNELIRWH GPPAEITVEL TRDLKLSPRR LAELEREQAE NQRKNDKRTS LLRKLGLPAS THNLLKLRLW DEQGDVASEC PYTGEAIGLE RLVSDDVDID HLIPFSISWD DSAANKVVCM RYANREKGNR TPFEAFGHRQ GRPYDWADIA ERAARLPRGK RWRFGPGARA QFEELGDFQA RLLNETSWLA RVAKQYLAAV THPHRIHVLP GRLTALLRAT WELNDLLPGS DDRAAKSRKD HRHHAIDALV AALTDQALLR RMANAHDDTR RKIEVLLPWP TFRIDLETRL KAMLVSHKPD HGLQARLHED TAYGTVEHPE TEDGANLVYR KTFVDISEKE IDRIRDRRLR DLVRAHVAGE RQQGKTLKAA VLSFAQRRDI AGHPNGIRHV RLTKSIKPDY LVPIRDKAGR IYKSYNAGEN AFVDILQAES GRWIARATTV FQANQANESH DAPAAQPIMR VFKGDMLRID HAGAEKFVKI VRLSPSNNLL YLVEHHQAGV FQTRHDDPED SFRWLFASFD KLREWNAELV RIDTLGQPWR RKRGLETGSE DATRIGWTRP KKWP	
Acidaminococcus sp. D21 WP_009016219.1	MGKMYYLGLD IGTNSVGYAV TDPSYHLLKF KGEPMWGAHV FAAGNQSAER RSFRTSRRRL DRRQQRVKLV QEIFAPVISP IDPRFFIRLH ESALWRDDVA ETDKHIFFND PTYTDKEYYS DYPTIHHLIV DLMESSEKHD PRLVYLAVAW LVAHRGHFLN EVDKDNIGDV LSFDAFYPEF LAFLSDNGVS PWVCESKALQ ATLLSRNSVN DKYKALKSLI FGSQKPEDNF DANISEDGLI QLLAGKKVKV NKLFPQESND ASFTLNDKED AIEEILGTLT PDECEWIAHI RRLFDWAIMK HALKDGRTIS ESKVKLYEQH HHDLTQLKYF VKTYLAKEYD DIFRNVDSET TKNYVAYSYH VKEVKGTLPK NKATQEEFCK YVLGKVKNIE CSEADKVDFD EMIQRLTDNS FMPKQVSGEN RVIPYQLYYY ELKTILNKAA SYLPFLTQCG KDAISNQDKL LSIMTFRIPY FVGPLRKDNS EHAWLERKAG KIYPWNFNDK VDLDKSEEAF IRRMTNTCTY YPGEDVLPLD SLIYEKFMIL NEINNIRIDG YPISVDVKQQ VFGLFEKKRR VTVKDIQNLL LSLGALDKHG KLTGIDTTIH SNYNTYHHFK SLMERGVLTR DDVERIVERM TYSDDTKRVR LWLNNNYGTL TADDVKHISR LRKHDFGRLS KMFLTGLKGV HKETGERASI LDFMWNTNDN LMQLLSECYT FSDEITKLQE AYYAKAQLSL NDFLDSMYIS NAVKRPIYRT LAVVNDIRKA CGTAPKRIFI EMARDGESKK KRSVTRREQI KNLYRSIRKD FQQEVDFLEK ILENKSDGQL QSDALYLYFA QLGRDMYTGD PIKLEHIKDQ SFYNIDHIYP QSMVKDDSLD NKVLVQSEIN GEKSSRYPLD AAIRNKMKPL WDAYYNHGLI SLKKYQRLTR STPFTDDEKW DFINRQLVET RQSTKALAIL LKRKFPDTEI VYSKAGLSSD FRHEFGLVKS RNINDLHHAK DAFLAIVTGN VYHERFNRRW FMVNQPYSVK TKTLFTHSIK NGNFVAWNGE EDLGRIVKML KQNKNTIHFT RFSFDRKEGL FDIQPLKAST GLVPRKAGLD VVKYGGYDKS TAAYYLLVRF TLEDKKTQHK LMMIPVEGLY KARIDHDKEF LTDYAQTTIS EILQKDKQKV INIMFPMGTR HIKLNSMISI DGFYLSIGGK SSKGKSVLCH AMVPLIVPHK IECYIKAMES FARKFKENNK LRIVEKFDKI TVEDNLNLYE LFLQKLQHNP YNKFFSTQFD VLTNGRSTFT KLSPEEQVQT LLNILSIFKT CRSSGCDLKS INGSAQAARI MISADLTGLS KKYSDIRLVE QSASGLFVSK SQNLLEYL	(SEQ ID NO: 110)
Methylosinus trichosporium OB3b WP_003611034.1	MRVLGLDAGI ASLGWALIEI EESNRGELSQ GTIIGAGTWM FDAPEEKTQA GAKLKSEQRR TFRGQRRVVR RRRQRMNEVR RILHSHGLLP SSDRDALKQP GLDPWRIRAE ALDRLLGPVE LAVALGHIAR HRGFKSNSKG AKTNDPADDT SKMKRAVNET REKLARFGSA AKMLVEDESF VLRQTPTKNG ASEIVRRFRN REGDYSRSLL RDDLAAEMRA LFTAQARFQS AIATADLQTA FTKAAFFQRP LQDSEKLVGP CPFEVDEKRA PKRGYSFELF RFLSRLNHVT LRDGKQERTL TRDELALAAA DFGAAAKVSF TALRKKLKLP ETTVFVGVKA DEESKLDVVA RSGKAAEGTA RLRSVIVDAL GELAWGALLC SPEKLDKIAE VISFRSDIGR ISEGLAQAGC NAPLVDALTA AASDGRFDPF TGAGHISSKA ARNILSGLRQ GMTYDKACCA ADYDHTASRE RGAFDVGGHG REALKRILQE ERISRELVGS PTARKALIES IKQVKAIVER YGVPDRIHVE LARDVGKSIE EREEITRGIE KRNRQKDKLR GLFEKEVGRP PQDGARGKEE LLRFELWSEQ MGRCLYTDDY ISPSQLVATD DAVQVDHILP WSRFADDSYA NKTLCMAKAN QDKKGRTPYE WFKAEKTDTE WDAFIVRVEA LADMKGFKKR NYKLRNAEEA AAKFRNRNLN DTRWACRLLA EALKQLYPKG EKDKDGKERR RVFSRPGALT DRLRRAWGLQ	(SEQ ID NO: 111)

Table 4- Cas9 orthologs

	WMKKSTKGDR IPDDRHHALD AIVIAATTES LLQRATREVQ EIEDKGLHYD LVKNVTPPWP GFREQAVEAV EKVFVARAER RRARGKAHDA TIRHIAVREG EQRVYERRKV AELKLADLDR VKDAERNARL IEKLRNWIEA GSPKDDPPLS PKGDPIFKVR LVTKSKVNIA LDTGNPKRPG TVDRGEMARV DVFRKASKKG KYEYYLVPIY PHDIATMKTP PIRAVQAYKP EDEWPEMDSS YEFCWSLVPM TYLQVISSKG EIFEGYYRGM NRSVGAIQLS AHSNSSDVVQ GIGARTLTEF KKFNVDRFGR KHEVERELRT WRGETWRGKA YI	
Actinomyces coleocanis DSM 15436 WP_006546479.1	MDNKNYRIGI DVGLNSIGFC AVEVDQHDTP LGFLNLSVYR HDAGIDPNGK KTNTTRLAMS GVARRTRRLF RKRKRRLAAL DRFIEAQGWT LPDHADYKDP YTPWLVRAEL AQTPIRDEND LHEKLAIAVR HIARHRGWRS PWVPVRSLHV EQPPSDQYLA LKERVEAKTL LQMPEGATPA EMVVALDLSV DVNLRPKNRE KTDTRPENKK PGFLGGKLMQ SDNANELRKI AKIQGLDDAL LRELIELVFA ADSPKGASGE LVGYDVLPGQ HGKRRAEKAH PAFQRYRIAS IVSNLRIRHL GSGADERLDV ETQKRVFEYL LNAKPTADIT WSDVAEEIGV ERNLLMGTAT QTADGERASA KPPVDVTNVA FATCKIKPLK EWWLNADYEA RCVMVSALSH AEKLTEGTAA EVEVAEFLQN LSDEDNEKLD SFSLPIGRAA YSVDSLERLT KRMIENGEDL FEARVNEFGV SEDWRPPAEP IGARVGNPAV DRVLKAVNRY LMAAEAEWGA PLSVNIEHVR EGFISKRQAV EIDRENQKRY QRNQAVRSQI ADHINATSGV RGSDVTRYLA IQRQNGECLY CGTAITFVNS EMDHIVPRAG LGSTNTRDNL VATCERCNKS KSNKPFAVWA AECGIPGVSV AEALKRVDFW IADGFASSKE HRELQKGVKD RLKRKVSDPE IDNRSMESVA WMARELAHRV QYYFDEKHTG TKVRVFRGSL TSAARKASGF ESRVNFIGGN GKTRLDRRHH AMDAATVAML RNSVAKTLVL RGNIRASERA IGAAETWKSF RGENVADRQI FESWSENMRV LVEKFNLALY NDEVSIFSSL RLQLGNGKAH DDTITKLQMH KVGDAWSLTE IDRASTPALW CALTRQPDFT WKDGLPANED RTIIVNGTHY GPLDKVGIFG KAAASLLVRG GSVDIGSAIH HARIYRIAGK KPTYGMVRVF APDLLRYRNE DLFNVELPPQ SVSMRYAEPK VREAIREGKA EYLGWLVVGD ELLLDLSSET SGQIAELQQD FPGTTHWTVA GFFSPSRLRL RPVYLAQEGL GEDVSEGSKS IIAGQGWRPA VNKVFGSAMP EVIRRDGLGR KRRFSYSGLP VSWQG	(SEQ ID NO: 112)
Caenispirillum salinarum AK4 WP_009541330.1	MPVLSPLSPN AAQGRRRWSL ALDIGEGSIG WAVAEVDAEG RVLQLTGTGV TLFPSAWSNE NGTYVAHGAA DRAVRGQQQR HDSRRRRLAG LARLCAPVLE RSPEDLKDLT RTPPKADPRA IFFLRADAAR RPLDGPELFR VLHHMAAHRG IRLAELQEVD PPPESDADDA APAATEDEDG TRRAAADERA FRRLMAEHMH RHGTQPTCGE IMAGRLRETP AGAQPVTRAR DGLRVGGGVA VPTRALIEQE FDAIRAIQAP RHPDLPWDSL RRLVLDQAPI AVPPATPCLF LEELRRRGET FQGRTITREA IDRGLTVDPL IQALRIRETV GNLRLHERIT EPDGRQRYVP RAMPELGLSH GELTAPERDT LVRALMHDPD GLAAKDGRIP YTRLRKLIGY DNSPVCFAQE RDTSGGGITV NPTDPLMARW IDGWVDLPLK ARSLYVRDVV ARGADSAALA RLLAEGAHGV PPVAAAAVPA ATAAILESDI MQPGRYSVCP WAAEAILDAW ANAPTEGFYD VTRGLFGFAP GEIVLEDLRR ARGALLAHLP RTMAAARTPN RAAQQRGPLP AYESVIPSQL ITSLRRAHKG RAADWSAADP EERNPFLRTW TGNAATDHIL NQVRKTANEV ITKYGNRRGW DPLPSRITVE LAREAKHGVI RRNEIAKENR ENEGRRKKES AALDTFCQDN TVSWQAGGLP KERAALRLRL AQRQEFFCPY CAERPKLRAT DLFSPAETEI DHVIERRMGG DGPDNLVLAH KDCNNAKGKK TPHEHAGDLL DSPALAALWQ GWRKENADRL KGKGHKARTP REDKDFMDRV GWRFEEDARA KAEENQERRG RRMLHDTARA TRLARLYLAA AVMPEDPAEI GAPPVETPPS PEDPTGYTAI YRTISRVQPV NGSVTHMLRQ RLLQRDKNRD YQTHHAEDAC LLLLAGPAVV QAFNTEAAQH GADAPDDRPV DLMPTSDAYH QQRRARALGR 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 ARLNGLDEGA GVTGNNAPPR PLRQDIDRLT PLWRDHGTAP GGYLGTAVGE LEDKARSALR GKAMRQTLTD AGITAEAGWR LDSEGAVCDL EVAKGDTVKK DGKTYKVGVI	(SEQ ID NO: 113)

Table 4- Cas9 orthologs		
	TQGIFGMPVD AAGSAPRTPE DCEKFEEQYG IKPWKAKGIP LA	
Coriobacteriu m glomerans PW2 WP_01370957 5.1	MKLRGIEDDY SIGLDMGTSS VGWAVTDERG TLAHFKRKPT WGSRLFREAQ TAAVARMPRG QRRRYVRRRW RLDLLQKLFE QQMEQADPDF FIRLRQSRLL RDDRAEEHAD YRWPLFNDCK FTERDYYQRF PTIYHVRSWL METDEQADIR LIYLALHNIV KHRGNFLREG QSLSAKSARP DEALNHLRET LRVWSSERGF ECSIADNGSI LAMLTHPDLS PSDRRKKIAP LFDVKSDDAA ADKKLGIALA GAVIGLKTEF KNIFGDFPCE DSSIYLSNDE AVDAVRSACP DDCAELFDRL CEVYSAYVLQ GLLSYAPGQT ISANMVEKYR RYGEDLALLK KLVKIYAPDQ YRMFFSGATY PGTGIYDAAQ ARGYTKYNLG PKKSEYKPSE SMQYDDFRKA VEKLFAKTDA RADERYRMMM DRFDKQQFLR RLKTSDNGSI YHQLHLEELK AIVENQGRFY PFLKRDADKL VSLVSFRIPY YVGPLSTRNA RTDQHGENRF AWSERKPGMQ DEPIFPWNWE SIIDRSKSAE KFILRMTGMC TYLQQEPVLP KSSLLYEEFC VLNELNGAHW SIDGDDEHRF DAADREGIIE ELFRRKRTVS YGDVAGWMER ERNQIGAHVC GGQGEKGFES KLGSYIFFCK DVFKVERLEQ SDYPMIERII LWNTLFEDRK ILSQRLKEEY GSRLSAEQIK TICKKRFTGW GRLSEKFLTG ITVQVDEDSV SIMDVLREGC PVSGKRGRAM VMMEILRDEE LGFQKKVDDF NRAFFAENAQ ALGVNELPGS PAVRRSLNQS IRIVDEIASI AGKAPANIFI EVTRDEDPKK KGRRTKRRYN DLKDALEAFK KEDPELWREL CETAPNDMDE RLSLYFMQRG KCLYSGRAID IHQLSNAGIY EVDHIIPRTY VKDDSLENKA LVYREENQRK TDMLLIDPEI RRRMSGYWRM LHEAKLIGDK KFRNLLRSRI DDKALKGFIA RQLVETGQMV KLVRSLLEAR YPETNIISVK ASISHDLRTA AELVKCREAN DFHHAHDAFL ACRVGLFIQK RHPCVYENPI GLSQVVRNYV RQQADIFKRC RTIPGSSGFI VNSFMTSGFD KETGEIFKDD WDAEAEVEGI RRSLNFRQCF ISRMPFEDHG VFWDATIYSP RAKKTAALPL KQGLNPSRYG SFSREQFAYF FIYKARNPRK EQTLFEFAQV PVRLSAQIRQ DENALERYAR ELAKDQGLEF IRIERSKILK NQLIEIDGDR LCITGKEEVR NACELAFAQD EMRVIRMLVS EKPVSRECVI SLFNRILLHG DQASRRLSKQ LKLALLSEAF SEASDNVQRN VVLGLIAIFN GSTNMVNLSD IGGSKFAGNV RIKYKKELAS PKVNVHLIDQ SVTGMFERRT KIGL	(SEQ ID NO: 114)

[0076] 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.

[0077] 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

[0078] 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 a-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.

[0079] *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.

[0080] *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.

[0081] 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.

[0082] *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.

[0083] 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)

[0084] 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).

[0085] 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_33_10]	MSNFFKNFTN LYELSKTLRF ELKPVGDTLT NMKDHLEYDE KLQTFLKDQN IDDAYQALKP QFDEIHEEFI TDSLESKKAK EIDFSEYLDL FQEKKELNDS EKKLRNKIGE TFNKAGEKWK KEKYPQYEWK KGSKIANGAD ILSCQDMLQF IKYKNPEDEK IKNYIDDTLK GFFTYFGGFN QNRANYYETK KEASTAVATR IVHENLPKFC DNVIQFKHII KRKKDGTVEK TERKTEYLNA YQYLKNNNKI TQIKDAETEK MIESTPIAEK IFDVYYFSSC LSQKQIEEYN RIIGHYNLLI NLYNQAKRSE GKHLSANEKK YKDLPKFKTL YKQIGCGKKK DLFYTIKCDT EEEANKSRNE GKESHSVEEI INKAQEAINK YFKSNNDCEN INTVPDFINY ILTKENYEGV YWSKAAMNTI SDKYFANYHD LQDRLKEAKV FQKADKKSED DIKIPEAIEL SGLFGVLDSL ADWQTTLFKS SILSNEDKLK IITDSQTPSE ALLKMIFNDI EKNMESFLKE TNDIITLKKY KGNKEGTEKI KQWFDYTLAI NRMLKYFLVK ENKIKGNSLD TNISEALKTL IYSDDAEWFK WYDALRNYLT QKPQDEAKEN KLKLNFDNPS LAGGWDVNKE CSNFCVILKD KNEKKYLAIM KKGENTLFQK EWTEGRGKNL TKKSNPLFEI NNCEILSKME YDFWADVSKM IPKCSTQLKA VVNHFKQSDN EFIFPIGYKV TSGEKFREEC KISKQDFELN NKVFNKNELS VTAMRYDLSS TQEKQYIKAF QKEYWELLFK QEKRDTKLTN NEIFNEWINF CNKKYSELLS WERKYKDALT NWINFCKYFL SKYPKTTLFN YSFKESENYN SLDEFYRDVD ICSYKLNINT TINKSILDRL VEEGKLYLFE IKNQDSNDGK SIGHKNNLHT IYWNAIFENF DNRPKLNGEA EIFYRKAISK DKLGIVKGKK TKNGTEIIKN YRFSKEKFIL HVPITLNFCS NNEYVNDIVN TKFYNFSNLH FLGIDRGEKH LAYYSLVNKN GEIVDQGTLN LPFTDKDGNQ RSIKKEKYFY NKQEDKWEAK EVDCWNYNDL LDAMASNRDM ARKNWQRIGT IKEAKNGYVS LVIRKIADLA VNNERPAFIV LEDLNTGFKR SRQKIDKSVY QKFELALAKK LNFLVDKNAK RDEIGSPTKA LQLTPPVNNY GDIENKKQAG IMLYTRANYT SQTDPATGWR KTIYLKAGPE ETTYKKDGKI KNKSVKDQII ETFTDIGFDG KDYYFEYDKG EFVDEKTGEI KPKKWRLYSG ENGKSLDRFR GEREKDKYEW KIDKIDIVKI LDDLFVNFDK NISLLKQLKE GVELTRNNEH GTGESLRFAI NLIQQIRNTG NNERDNDFIL SPVRDENGKH FDSREYWDKE TKGEKISMPS SGDANGAFNI ARKGIIMNAH ILANSDSKDL SLFVSDEEWD LHLNNKTEWK KQLNIFSSRK AMAKRKK	(SEQ ID NO: 116)
KKR91555_(modified) hypothetical protein UU43_C0004G0003 [Parcubacteria (Falkowbacteria) bacterium GW2011_GWA2_41_14]	MLFFMSTDIT NKPREKGVFD NFTNLYEFSK TLTFGLIPLK WDDNKKMIVE DEDFSVLRKY GVIEEDKRIA ESIKIAKFYL NILHRELIGK VLGSLKFEKK NLENYDRLLG EIEKNNKNEN ISEDKKKEIR KNFKKELSIA QDILLKKVGE VFESNGSGIL SSKNCLDELT KRFTRQEVDK LRRENKDIGV EYPDVAYREK DGKEETKSFF AMDVGYLDDF HKNRKQLYSV KGKKNSLGRR 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 TVTLGGKKES ITSLDEKKNR LKEKLGWFEG KVRENEKMKD EEEGEFCSTV LAYSQAVLNI TKRAEIFWLN EKQDAKVGED NKDMIFYKKF DEFADDGFAP FFYFDKFGNY LKRRSRNTTK EIKLHFGNDD LLEGWDMNKE PEYWSFILRD RNQYYLGIGK KDGEIFHKKL GNSVEAVKEA YELENEADFY EKIDYKQLNI DRFEGIAFPK KTKTEEAFRQ VCKKRADEFL GGDTYEFKIL LAIKKEYDDF KARRQKEKDW DSKFSKEKMS KLIEYYITCL GKRDDWKRFN LNFRQPKEYE DRSDFVRHIQ RQAYWIDPRK VSKDYVDKKV AEGEMFLFKV HNKDFYDFER KSEDKKNHTA NLFTQYLLEL FSCENIKNIK SKDLIESIFE LDGKAEIRFR PKTDDVKLKI YQKKGKDVTY ADKRDGNKEK EVIQHRRFAK DALTLHLKIR LNFGKHVNLF DFNKLVNTEL FAKVPVKILG MDRGENNLIY YCFLDEHGEI ENGKCGSLNR VGEQIITLED DKKVKEPVDY FQLLVDREGQ RDWEQKNWQK MTRIKDLKKA YLGNVVSWIS KEMLSGIKEG VVTIGVLEDL NSNFKRTRFF RERQVYQGFE KALVNKLGYL VDKKYDNYRN VYQFAPIVDS VEEMEKNKQI GTLVYVPASY TSKICPHPKC GWRERLYMKN SASKEKIVGL LKSDGIKISY DQKNDRFYFE YQWEQEHKSD GKKKKYSGVD KVFSNVSRMR WDVEQKKSID FVDGTDGSIT NKLKSLLKGK GIELDNINQQ IVNQQKELGV EFFQSIIFYF NLIMQIRNYD KEKSGSEADY IQCPSCLFDS RKPEMNGKLS AITNGDANGA YNIARKGFMQ LCRIRENPQE PMKLITNREW DEAVREWDIY SAAQKIPVLS EEN	(SEQ ID NO: 117)

Table 5 - Cas12a orthologs		
KDN25524_(modified) hypothetical protein MBO_03467 [Moraxella bovoculi 237] > WP_052585281.1 type V CRISPR-associated protein Cpf1 [Moraxella bovoculi]	MLFQDFTHLY PLSKTVRFEL KPIDRTLEHI HAKNFLSQDE TMADMHQKVK VILDDYHRDF IADMMGEVKL TKLAEFYDVY LKFRKNPKDD ELQKQLKDLQ AVLRKEIVKP IGNGGKYKAG YDRLFGAKLF KDGKELGDLA KFVIAQEGES SPKLAHLAHF EKFSTYFTGF HDNRKNMYSD EDKHTAIAYR LIHENLPRFI DNLQILTTIK QKHSALYDQI INELTASGLD VSLASHLDGY HKLLTQEGIT AYNTLLGGIS GEAGSPKIQG INELINSHHN QHCHKSERIA KLRPLHKQIL SDGMSVSFLP SKFADDSEMC QAVNEFYRHY ADVFAKVQSL FDGFDDHQKD GIYVEHKNLN ELSKQAFGDF ALLGRVLDGY YVDVVNPEFN ERFAKAKTDN AKAKLTKEKD KFIKGVHSLA SLEQAIEHYT ARHDDESVQA GKLGQYFKHG LAGVDNPIQK IHNNHSTIKG FLERERPAGE RALPKIKSGK NPEMTQLRQL KELLDNALNV AHFAKLLTTK TTLDNQDGNF YGEFGVLYDE LAKIPTLYNK VRDYLSQKPF STEKYKLNFG NPTLLNGWDL NKEKDNFGVI LQKDGCYYLA LLDKAHKKVF DNAPNTGKSI YQKMIYKYLE VRKQFPKVFF SKEAIAINYH PSKELVEIKD KGRQRSDDER LKLYRFILEC LKIHPKYDKK FEGAIGDIQL FKKDKKGREV PISEKDLFDK INGIFSSKPK LEMEDFFIGE FKRYNPSQDL VDQYNIYKKI DSNDNRKKEN FYNNHPKFKK DLVRYYYESM CKHEEWEESF EFSKKLQDIG CYVDVNELFT EIETRRLNYK ISFCNINADY IDELVEQGQL YLFQIYNKDF SPKAHGKPNL HTLYFKALFS EDNLADPIYK LNGEAQIFYR KASLDMNETT IHRAGEVLEN KNPDNPKKRQ FVYDIIKDKR YTQDKFMLHV PITMNFGVQG MTIKEFNKKV NQSIQQYDEV NVIGIDRGER HLLYLTVINS KGEILEQCSL NDITTASANG TQMTTPYHKI LDKREIERLN ARVGWGEIET IKELKSGYLS HVVHQISQLM LKYNAIVVLE DLNFGFKRGR FKVEKQIYQN FENALIKKLN HLVLKDKADD EIGSYKNALQ LTNNFTDLKS IGKQTGFLFY VPAWNTSKID PETGFVDLLK PRYENIAQSQ AFFGKFDKIC YNADKDYFEF HIDYAKFTDK AKNSRQIWTI CSHGDKRYVY DKTANQNKGA AKGINVNDEL KSLFARHHIN EKQPNLVMDI CQNNDKEFHK SLMYLLKTLL ALRYSNASSD EDFILSPVAN DEGVFFNSAL ADDTQPQNAD ANGAYHIALK GLWLLNELKN SDDLNKVKLA IDNQTWLNFA QNR	(SEQ ID NO: 118)
KKT48220_(modified) hypothetical protein UW39_C0001G0044 [Parcubacteria bacterium GW2011_GWC2_44_17]	MENIFDQFIG KYSLSKTLRF ELKPVGKTED FLKINKVFEK DQTIDDSYNQ AKFYFDSLHQ KFIDAALASD KTSELSFQNF ADVLEKQNKI ILDKKREMGA LRKRDKNAVG IDRLQKEIND AEDIIQKEKE KIYKDVRTLF DNEAESWKTY YQEREVDGKK ITFSKADLKQ KGADFLTAAG ILKVLKYEFP EEKEKEFQAK NQPSLFVEEK ENPGQKRYIF DSFDKFAGYL TKFQQTKKNL YAADGTSTAV ATRIADNFII FHQNTKVFRD KYKNNHTDLG FDEENIFEIE RYKNCLLQRE IEHIKNENSY NKIIGRINKK IKEYRDQKAK DTKLTKSDFP FFKNLDKQIL GEVEKEKQLI EKTREKTEED VLIERFKEFI ENNEERFTAA KKLMNAFCNG EFESEYEGIY LKNKAINTIS RRWFVSDRDF ELKLPQQKSK NKSEKNEPKV KKFISIAEIK NAVEELDGDI FKAVFYDKKI IAQGGSKLEQ FLVIWKYEFE YLFRDIEREN GEKLLGYDSC LKIAKQLGIF PQEKEAREKA TAVIKNYADA GLGIFQMMKY FSLDDKDRKN TPGQLSTNFY AEYDGYYKDF EFIKYYNEFR NFITKKPFDE DKIKLNFENG ALLKGWDENK EYDFMGVILK KEGRLYLGIM HKNHRKLFQS MGNAKGDNAN RYQKMIYKQI ADASKDVPRL LLTSKKAMEK FKPSQEILRI KKEKTFKRES KNFSLRDLHA LIEYYRNCIP QYSNWSFYDF QFQDTGKYQN IKEFTDDVQK YGYKISFRDI DDEYINQALN EGKMYLFEVV NKDIYNTKNG SKNLHTLYFE HILSAENLND PVFKLSGMAE IFQRQPSVNE REKITTQKNQ CILDKGDRAY KYRRYTEKKI MFHMSLVLNT GKGEIKQVQF NKIINQRISS SDNEMRVNVI GIDRGEKNLL YYSVVKQNGE IIEQASLNEI NGVNYRDKLI EREKERLKNR QSWKPVVKIK DLKKGYISHV IHKICQLIEK YSAIVVLEDL NMRFKQIRGG IERSVYQQFE KALIDKLGYL VFKDNRDLRA PGGVLNGYQL SAPFVSFEKM RKQTGILFYT QAEYTSKTDP ITGFRKNVYI SNSASLDKIK EAVKKFDAIG WDGKEQSYFF KYNPYNLADE KYKNSTVSKE WAIFASAPRI RRQKGEDGYW KYDRVKVNEE FEKLLKVWNF VNPKATDIKQ EIIKKEKAGD LQGEKELDGR LRNFWHSFIY LFNLVLELRN SFSLQIKIKA GEVIAVDEGV DFIASPVKPF FTTPNPYIPS NLCWLAVENA DANGAYNIAR KGVMILKKIR EHAKKDPEFK KLPNLFISNA EWDEAARDWG KYAGTTALNL DH	(SEQ ID NO: 119)
WP_031492824_(modified) hypothetical protein	MSSLTKFTNK YSKQLTIKNE LIPVGKTLEN IKENGLIDGD EQLNENYQKA KIIVDDFLRD FINKALNNTQ IGNWRELADA LNKEDEDNIE KLQDKIRGII VSKFETFDLF SSYSIKKDEK IIDDDNDVEE EELDLGKKTS SFKYIFKKNL FKLVLPSYLK TTNQDKLKII SSFDNFSTYF RGFFENRKNI FTKKPISTSI	(SEQ ID NO: 120)

Table 5 - Cas12a orthologs		
[Succinivibrio dextrinosolvens]	AYRIVHDNFP KFLDNIRCFN VWQTECPQLI VKADNYLKSK NVIAKDKSLA NYFTVGAYDY FLSQNGIDFY NNIIGGLPAF AGHEKIQGLN EFINQECQKD SELKSKLKNR HAFKMAVLFK QILSDREKSF VIDEFESDAQ VIDAVKNFYA EQCKDNNVIF NLLNLIKNIA FLSDDELDGI FIEGKYLSSV SQKLYSDWSK LRNDIEDSAN SKQGNKELAK KIKTNKGDVE KAISKYEFSL SELNSIVHDN TKFSDLLSCT LHKVASEKLV KVNEGDWPKH LKNNEEKQKI KEPLDALLEI YNTLLIFNCK SFNKNGNFYV DYDRCINELS SVVYLYNKTR NYCTKKPYNT DKFKLNFNSP QLGEGFSKSK ENDCLTLLFK KDDNYYVGII RKGAKINFDD TQAIADNTDN CIFKMNYFLL KDAKKFIPKC SIQLKEVKAH FKKSEDDYIL SDKEKFASPL VIKKSTFLLA TAHVKGKKGN IKKFQKEYSK ENPTEYRNSL NEWIAFCKEF LKTYKAATIF DITTLKKAEE YADIVEFYKD VDNLCYKLEF CPIKTSFIEN LIDNGDLYLF RINNKDFSSK STGTKNLHTL YLQAIFDERN LNNPTIMLNG GAELFYRKES IEQKNRITHK AGSILVNKVC KDGTSLDDKI RNEIYQYENK FIDTLSDEAK KVLPNVIKKE ATHDITKDKR FTSDKFFFHC PLTINYKEGD TKQFNNEVLS FLRGNPDINI IGIDRGERNL IYVTVINQKG EILDSVSFNT VTNKSSKIEQ TVDYEEKLAV REKERIEAKR SWDSISKIAT LKEGYLSAIV HEICLLMIKH NAIVVLENLN AGFKRIRGGL SEKSVYQKFE KMLINKLNYF VSKKESDWNK PSGLLNGLQL SDQFESFEKL GIQSGFIFYV PAAYTSKIDP TTGFANVLNL SKVRNVDAIK SFFSNFNEIS YSKKEALFKF SFDLDSLSKK GFSSFVKFSK SKWNVYTFGE RIIKPKNKQG YREDKRINLT FEMKKLLNEY KVSFDLENNL IPNLTSANLK DTFWKELFFI FKTTLQLRNS VTNGKEDVLI SPVKNAKGEF FVSGTHNKTL PQDCDANGAY HIALKGLMIL ERNNLVREEK DTKKIMAISN VDWFEYVQKR RGVL	
KKT50231_(modified) hypothetical protein UW40_C0007G0006 [Parcubacteria bacterium GW2011_GWF2_44_17]	MKPVGKTEDF LKINKVFEKD QTIDDSYNQA KFYFDSLHQK FIDAALASDK TSELSFQNFA DVLEKQNKII LDKKREMGAL RKRDKNAVGI DRLQKEINDA EDIIQKEKEK IYKDVRTLFD NEAESWKTYY QEREVDGKKI TFSKADLKQK GADFLTAAGI LKVLKYEFPE EKEKEFQAKN QPSLFVEEKE NPGQKRYIFD SFDKFAGYLT KFQQTKKNLY AADGTSTAVA TRIADNFIIF HQNTKVFRDK YKNNHTDLGF DEENIFEIER YKNCLLQREI EHIKNENSYN KIIGRINKKI KEYRDQKAKD TKLTKSDFPF FKNLDKQILG EVEKEKQLIE KTREKTEEDV LIERFKEFIE NNEERFTAAK KLMNAFCNGE FESEYEGIYL KNKAINTISR RWFVSDRDFE LKLPQQKSKN KSEKNEPKVK KFISIAEIKN AVEELDGDIF KAVFYDKKII AQGGSKLEQF LVIWKYEFEY LFRDIERENG EKLLGYDSCL KIAKQLGIFP QEKEAREKAT AVIKNYADAG LGIFQMMKYF SLDDKDRKNT PGQLSTNFYA EYDGYYKDFE FIKYYNEFRN FITKKPFDED KIKLNFENGA LLKGWDENKE YDFMGVILKK EGRLYLGIMH KNHRKLFQSM GNAKGDNANR YQKMIYKQIA DASKDVPRLL LTSKKAMEKF KPSQEILRIK KEKTFKRESK NFSLRDLHAL IEYYRNCIPQ YSNWSFYDFQ FQDTGKYQNI KEFTDDVQKY GYKISFRDID DEYINQALNE GKMYLFEVVN KDIYNTKNGS KNLHTLYFEH ILSAENLNDP VFKLSGMAEI FQRQPSVNER EKITTQKNQC ILDKGDRAYK YRRYTEKKIM FHMSLVLNTG KGEIKQVQFN KIINQRISSS DNEMRVNVIG IDRGEKNLLY YSVVKQNGEI IEQASLNEIN GVNYRDKLIE REKERLKNRQ SWKPVVKIKD LKKGYISHVI HKICQLIEKY SAIVVLEDLN MRFKQIRGGI ERSVYQQFEK ALIDKLGYLV FKDNRDLRAP GGVLNGYQLS APFVSFEKMR KQTGILFYTQ AEYTSKTDPI TGFRKNVYIS NSASLDKIKE AVKKFDAIGW DGKEQSYFFK YNPYNLADEK YKNSTVSKEW AIFASAPRIR RQKGEDGYWK YDRVKVNEEF EKLLKVWNFV NPKATDIKQE IIKKEKAGDL QGEKELDGRL RNFWHSFIYL FNLVLELRNS FSLQIKIKAG EVIAVDEGVD FIASPVKPFF TTPNPYIPSN LCWLAVENAD ANGAYNIARK GVMILKKIRE HAKKDPEFKK LPNLFISNAE WDEAARDWGK YAGTTALNLD H	(SEQ ID NO: 121)
WP_004356401_(modified) hypothetical protein [Prevotella disiens]	MKVMENYQEF TNLFQLNKTL RFELKPIGKT CELLEEGKIF ASGSFLEKDK VRADNVSYVK KEIDKKHKIF IEETLSSFSI SNDLLKQYFD CYNELKAFKK DCKSDEEEVK KTALRNKCTS IQRAMREAIS QAFLKSPQKK LLAIKNLIEN VFKADENVQH FSEFTSYFSG FETNRENFYS DEEKSTSIAY RLVHDNLPIF IKNIYIFEKL KEQFDAKTLS EIFENYKLYV AGSSLDEVFS LEYFNNTLTQ KGIDNYNAVI GKIVKEDKQE IQGLNEHINL YNQKHKDRRL PFFISLKKQI LSDREALSWL PDMFKNDSEV IKALKGFYIE DGFENNVLTP LATLLSSLDK YNLNGIFIRN NEALSSLSQN VYRNFSIDEA IDANAELQTF NNYELIANAL RAKIKKETKQ GRKSFEKYEE YIDKKVKAID SLSIQEINEL VENYVSEFNS NSGNMPRKVE DYFSLMRKGD FGSNDLIENI KTKLSAAEKL LGTKYQETAK	(SEQ ID NO: 122)

Table 5 - Cas12a orthologs		
	DIFKKDENSK LIKELLDATK QFQHFIKPLL GTGEEADRDL VFYGDFLPLY EKFEELTLLY NKVRNRLTQK PYSKDKIRLC FNKPKLMTGW VDSKTEKSDN GTQYGGYLFR KKNEIGEYDY FLGISSKAQL FRKNEAVIGD YERLDYYQPK ANTIYGSAYE GENSYKEDKK RLNKVIIAYI EQIKQTNIKK SIIESISKYP NISDDDKVTP SSLLEKIKKV SIDSYNGILS FKSFQSVNKE VIDNLLKTIS PLKNKAEFLD LINKDYQIFT EVQAVIDEIC KQKTFIYFPI SNVELEKEMG DKDKPLCLFQ ISNKDLSFAK TFSANLRKKR GAENLHTMLF KALMEGNQDN LDLGSGAIFY RAKSLDGNKP THPANEAIKC RNVANKDKVS LFTYDIYKNR RYMENKFLFH LSIVQNYKAA NDSAQLNSSA TEYIRKADDL HIIGIDRGER NLLYYSVIDM KGNIVEQDSL NIIRNNDLET DYHDLLDKRE KERKANRQNW EAVEGIKDLK KGYLSQAVHQ IAQLMLKYNA IIALEDLGQM FVTRGQKIEK AVYQQFEKSL VDKLSYLVDK KRPYNELGGI LKAYQLASSI TKNNSDKQNG FLFYVPAWNT SKIDPVTGFT DLLRPKAMTI KEAQDFFGAF DNISYNDKGY FEFETNYDKF KIRMKSAQTR WTICTFGNRI KRKKDKNYWN YEEVELTEEF KKLFKDSNID YENCNLKEEI QNKDNRKFFD DLIKLLQLTL QMRNSDDKGN DYIISPVANA EGQFFDSRNG DKKLPLDADA NGAYNIARKG LWNIRQIKQT KNDKKLNLSI SSTEWLDFVR EKPYLK	
CCB70584_(modified) Protein of unknown function [Flavobacterium branchiophilum FL-15]	MTNKFTNQYS LSKTLRFELI PQGKTLEFIQ EKGLLSQDKQ RAESYQEMKK TIDKFHKYFI DLALSNAKLT HLETYLELYN KSAETKKEQK FKDDLKKVQD NLRKEIVKSF SDGDAKSIFA ILDKKELITV ELEKWFENNE QKDIYFDEKF KTFTTYFTGF HQNRKNMYSV EPNSTAIAYR LIHENLPKFL ENAKAFEKIK QVESLQVNFR ELMGEFGDEG LIFVNELEEM FQINYYNDVL SQNGITIYNS IISGFTKNDI KYKGLNEYIN NYNQTKDKKD RLPKLKQLYK QILSDRISLS FLPDAFTDGK QVLKAIFDFY KINLLSYTIE GQEESQNLLL LIRQTIENLS SFDTQKIYLK NDTHLTTISQ QVFGDFSVFS TALNYWYETK VNPKFETEYS KANEKKREIL DKAKAVFTKQ DYFSIAFLQE VLSEYILTLD HTSDIVKKHS SNCIADYFKN HFVAKKENET DKTFDFIANI TAKYQCIQGI LENADQYEDE LKQDQKLIDN LKFFLDAILE LLHFIKPLHL KSESITEKDT AFYDVFENYY EALSLLTPLY NMVRNYVTQK PYSTEKIKLN FENAQLLNGW DANKEGDYLT TILKKDGNYF LAIMDKKHNK AFQKFPEGKE NYEKMVYKLL PGVNKMLPKV FFSNKNIAYF NPSKELLENY KKETHKKGDT FNLEHCHTLI DFFKDSLNKH EDWKYFDFQF SETKSYQDLS GFYREVEHQG YKINFKNIDS EYIDGLVNEG KLFLFQIYSK DFSPFSKGKP NMHTLYWKAL FEEQNLQNVI YKLNGQAEIF FRKASIKPKN IILHKKKIKI AKKHFIDKKT KTSEIVPVQT IKNLNMYYQG KISEKELTQD DLRYIDNFSI FNEKNKTIDI IKDKRFTVDK FQFHVPITMN FKATGGSYIN QTVLEYLQNN PEVKIIGLDR GERHLVYLTL IDQQGNILKQ ESLNTITDSK ISTPYHKLLD NKENERDLAR KNWGTVENIK ELKEGYISQV VHKIATLMLE ENAIVVMEDL NFGFKRGRFK VEKQIYQKLE KMLIDKLNYL VLKDKQPQEL GGLYNALQLT NKFESFQKMG KQSGFLFYVP AWNTSKIDPT TGFVNYFYTK YENVDKAKAF FEKFEAIRFN AEKKYFEFEV KKYSDFNPKA EGTQQAWTIC TYGERIETKR QKDQNNKFVS TPINLTEKIE DFLGKNQIVY GDGNCIKSQI ASKDDKAFFE TLLYWFKMTL QMRNSETRTD IDYLISPVMN DNGTFYNSRD YEKLENPTLP KDADANGAYH IAKKGLMLLN KIDQADLTKK VDLSISNRDW LQFVQKNK	(SEQ ID NO: 123)
WP_005398606_(modified) hypothetical protein [Helcococcus kunzii]	MFEKLSNIVS ISKTIRFKLI PVGKTLENIE KLGKLEKDFE RSDFYPILKN ISDDYYRQYI KEKLSDLNLD WQKLYDAHEL LDSSKKESQK NLEMIQAQYR KVLFNILSGE LDKSGEKNSK DLIKNNKALY GKLFKKQFIL EVLPDFVNNN DSYSEEDLEG LNLYSKFTTR LKNFWETRKN VFTDKDIVTA IPFRAVNENF GFYYDNIKIF NKNIEYLENK IPNLENELKE ADILDDNRSV KDYFTPNGFN YVITQDGIDV YQAIRGGFTK ENGEKVQGIN EILNLTQQQL RRKPETKNVK LGVLTKLRKQ ILEYSESTSF LIDQIEDDND LVDRINKFNV SFFESTEVSP SLFEQIERLY NALKSIKKEE VYIDARNTQK FSQMLFGQWD VIRRGYTVKI TEGSKEEKKK YKEYLELDET SKAKRYLNIR EIEELVNLVE GFEEVDVFSV LLEKFKMNNI ERSEFEAPIY GSPIKLEAIK EYLEKHLEEY HKWKLLLIGN DDLDTDETFY PLLNEVISDY YIIPLYNLTR NYLTRKHSDK DKIKVNFDFP TLADGWSESK ISDNRSIILR KGGYYYLGIL IDNKLLINKK NKSKKIYEIL IYNQIPEFSK SIPNYPFTKK VKEHFKNNVS DFQLIDGYVS PLIITKEIYD IKKEKKYKKD FYKDNNTNKN YLYTIYKWIE FCKQFLYKYK GPNKESYKEM YDFSTLKDTS LYVNLNDFYA DVNSCAYRVL FNKIDENTID NAVEDGKLLL FQIYNKDFSP ESKGKKNLHT LYWLSMFSEE NLRTRKLKLN GQAEIFYRKK	(SEQ ID NO: 124)

Table 5 - Cas12a orthologs		
	LEKKPIIHKE GSILLNKIDK EGNTIPENIY HECYRYLNKK IGREDLSDEA IALFNKDVLK YKEARFDIIK DRRYSESQFF FHVPITFNWD IKTNKNVNQI VQGMIKDGEI KHIIGIDRGE RHLLYYSVID LEGNIVEQGS LNTLEQNRFD NSTVKVDYQN KLRTREEDRD RARKNWTNIN KIKELKDGYL SHVVHKLSRL IIKYEAIVIM ENLNQGFKRG RFKVERQVYQ KFELALMNKL SALSFKEKYD ERKNLEPSGI LNPIQACYPV DAYQELQGQN GIVFYLPAAY TSVIDPVTGF TNLFRLKSIN SSKYEEFIKK FKNIYFDNEE EDFKFIFNYK DFAKANLVIL NNIKSKDWKI STRGERISYN SKKKEYFYVQ PTEFLINKLK ELNIDYENID 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 KARNDHYKEL KPIIDRIYKT YADQCLQLVQ LDWENLSAAI DSYRKEKTEE TRNALIEEQA TYRNAIHDYF IGRTDNLTDA INKRHAEIYK GLFKAELFNG KVLKQLGTVT TTEHENALLR SFDKFTTYFS GFYENRKNVF SAEDISTAIP HRIVQDNFPK FKENCHIFTR LITAVPSLRE HFENVKKAIG IFVSTSIEEV FSFPFYNQLL TQTQIDLYNQ LLGGISREAG TEKIKGLNEV LNLAIQKNDE TAHIIASLPH RFIPLFKQIL SDRNTLSFIL EEFKSDEEVI QSFCKYKTLL RNENVLETAE ALFNELNSID LTHIFISHKK LETISSALCD HWDTLRNALY ERRISELTGK ITKSAKEKVQ RSLKHEDINL QEIISAAGKE LSEAFKQKTS EILSHAHAAL DQPLPTTLKK QEEKEILKSQ LDSLLGLYHL LDWFAVDESN EVDPEFSARL TGIKLEMEPS LSFYNKARNY ATKKPYSVEK FKLNFQMPTL ASGWDVNKEK NNGAILFVKN GLYYLGIMPK QKGRYKALSF EPTEKTSEGF DKMYYDYFPD AAKMIPKCST QLKAVTAHFQ THTTPILLSN NFIEPLEITK EIYDLNNPEK EPKKFQTAYA KKTGDQKGYR EALCKWIDFT RDFLSKYTKT TSIDLSSLRP SSQYKDLGEY YAELNPLLYH ISFQRIAEKE IMDAVETGKL YLFQIYNKDF AKGHHGKPNL HTLYWTGLFS PENLAKTSIK LNGQAELFYR PKSRMKRMAH RLGEKMLNKK LKDQKTPIPD TLYQELYDYV NHRLSHDLSD EARALLPNVI TKEVSHEIIK DRRFTSDKFF FHVPITLNYQ AANSPSKFNQ RVNAYLKEHP ETPIIGIDRG ERNLIYITVI DSTGKILEQR SLNTIQQFDY QKKLDNREKE RVAARQAWSV VGTIKDLKQG YLSQVIHEIV DLMIHYQAVV VLENLNFGFK SKRTGIAEKA VYQQFEKMLI DKLNCLVLKD YPAEKVGGVL NPYQLTDQFT SFAKMGTQSG FLFYVPAPYT SKIDPLTGFV DPFVWKTIKN HESRKHFLEG FDFLHYDVKT GDFILHFKMN RNLSFQRGLP GFMPAWDIVF EKNETQFDAK GTPFIAGKRI VPVIENHRFT GRYRDLYPAN ELIALLEEKG IVFRDGSNIL PKLLENDDSH AIDTMVALIR SVLQMRNSNA ATGEDYINSP VRDLNGVCFD SRFQNPEWPM DADANGAYHI ALKGQLLLNH LKESKDLKLQ NGISNQDWLA YIQELRN	(SEQ ID NO: 125)
WP_004339290_(modified) hypothetical protein [Francisella tularensis]	MSIYQEFVNK YSLSKTLRFE LIPQGKTLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF FIEEILSSVC ISEDLLQNYS DVYFKLKKSD DDNLQKDFKS AKDTIKKQIS KYINDSEKFK NLFNQNLIDA KKGQESDLIL WLKQSKDNGI ELFKANSDIT DIDEALEIIK SFKGWTTYFK GFHENRKNVY SSNDIPTSII YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE ELTFDIDYKT SEVNQRVFSL DEVFEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRKGI NEYINLYSQQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA AFKTVEEKSI KETLSLLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY ITQQVAPKNL DNPSKKEQDL IAKKTEKAKY LSLETIKLAL EEFNKHRDID KQCRFEEILS NFAAIPMIFD EIAQNKDNLA QISIKYQNQG KKDLLQASAE EDVKAIKDLL DQTNNLLHRL KIFHISQSED KANILDKDEH FYLVFEECYF ELANIVPLYN KIRNYITQKP YSDEKFKLNF ENSTLASGWD KNKESANTAI LFIKDDKYYL GIMDKKHNKI FSDKAIEENK GEGYKKIVYK QIADASKDIQ NLMIIDGKTV CKKGRKDRNG VNRQLLSLKR KHLPENIYRI KETKSYLKNE ARFSRKDLYD FIDYYKDRLD YYDFEFELKP SNEYSDFNDF TNHIGSQGYK LTFENISQDY INSLVNEGKL YLFQIYSKDF SAYSKGRPNL HTLYWKALFD ERNLQDVVYK LNGEAELFYR KQSIPKKITH PAKETIANKN KDNPKKESVF EYDLIKDKRF TEDKFFFHCP ITINFKSSGA NKFNDEINLL LKEKANDVHI LSIDRGERHL AYYTLVDGKG NIIKQDNFNI IGNDRMKTNY HDKLAAIEKD RDSARKDWKK INNIKEMKEG YLSQVVHEIA KLVIEYNAIV VFEDLNFGFK RGRFKVEKQV YQKLEKMLIE KLNYLVFKDN EFDKTGGVLR AYQLTAPFET FKKMGKQTGI IYYVPAGFTS KICPVTGFVN	(SEQ ID NO: 126)

Table 5 - Cas12a orthologs		
	QLYPKYESVS KSQEFFSKFD KICYNLDKGY FEFSFDYKNF GDKAAKGKWT IASFGSRLIN FRNSDKNHNW DTREVYPTKE LEKLLKDYSI EYGHGECIKA AICGESDKKF FAKLTSVLNT ILQMRNSKTG TELDYLISPV ADVNGNFFDS RQAPKNMPQD ADANGAYHIG LKGLMLLDRI KNNQEGKKLN LVIKNEEYFE FVQNRNN	
WP_022501477 type V CRISPR-associated protein Cpf1 [Eubacterium sp. CAG:76]	MNKAADNYTG GNYDEFIALS KVQKTLRNEL KPTPFTAEHI KQRGIISEDE YRAQQSLELK KIADEYYRNY ITHKLNDINN LDFYNLFDAI EEKYKKNDKD NRDKLDLVEK SKRGEIAKML SADDNFKSMF EAKLITKLLP DYVERNYTGE DKEKALETLA LFKGFTTYFK GYFKTRKNMF SGEGGASSIC HRIVNVNASI FYDNLKTFMR IQEKAGDEIA LIEEELTEKL DGWRLEHIFS RDYYNEVLAQ KGIDYYNQIC GDINKHMNLY CQQNKFKANI FKMMKIQKQI MGISEKAFEI PPMYQNDEEV YASFNEFISR LEEVKLTDRL INILQNINIY NTAKIYINAR YYTNVSSYVY GGWGVIDSAI ERYLYNTIAG KGQSKVKKIE NAKKDNKFMS VKELDSIVAE YEPDYFNAPY IDDDDNAVKA FGGQGVLGYF NKMSELLADV SLYTIDYNSD DSLIENKESA LRIKKQLDDI MSLYHWLQTF IIDEVVEKDN AFYAELEDIC CELENVVTLY DRIRNYVTKK PYSTQKFKLN FASPTLAAGW SRSKEFDNNA IILLRNNKYY IAIFNVNNKP DKQIIKGSEE QRLSTDYKKM VYNLLPGPNK MLPKVFIKSD TGKRDYNPSS YILEGYEKNR HIKSSGNFDI NYCHDLIDYY KACINKHPEW KNYGFKFKET NQYNDIGQFY KDVEKQGYSI SWAYISEEDI NKLDEEGKIY LFEIYNKDLS AHSTGRDNLH TMYLKNIFSE DNLKNICIEL NGEAELFYRK SSMKSNITHK KDTILVNKTY INETGVRVSL SDEDYMKVYN YYNNNYVIDT ENDKNLIDII EKIGHRKSKI DIVKDKRYTE DKYFLYLPIT INYGIEDENV NSKIIEYIAK QDNMNVIGID RGERNLIYIS VIDNKGNIIE QKSFNLVNNY DYKNKLKNME KTRDNARKNW QEIGKIKDVK SGYLSGVISK IARMVIDYNA IIVMEDLNKG FKRGRFKVER QVYQKFENML ISKLNYLVFK ERKADENGGI LRGYQLTYIP KSIKNVGKQC GCIFYVPAAY TSKIDPATGF INIFDFKKYS GSGINAKVKD KKEFLMSMNS IRYINECSEE YEKIGHRELF AFSFDYNNFK TYNVSSPVNE WTAYTYGERI KKLYKDGRWL RSEVLNLTEN LIKLMEQYNI EYKDGHDIRE DISHMDETRN ADFICSLFEE LKYTVQLRNS KSEAEDENYD RLVSPILNSS NGFYDSSDYM ENENNTTHTM PKDADANGAY CIALKGLYEI NKIKQNWSDD KKFKENELYI NVTEWLDYIQ NRRFE	(SEQ ID NO: 127)
WP_014550095 type V CRISPR-associated protein Cpf1 [Francisella tularensis]	MSIYQEFVNK YSLSKTLRFE LIPQGKTLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF FIEEILSSVC ISEDLLQNYS DVYFKLKKSD DDNLQKDFKS AKDTIKKQIS EYIKDSEKFK NLFNQNLIDA KKGQESDLIL WLKQSKDNGI ELFKANSDIT DIDEALEIIK SFKGWTTYFK GFHENRKNVY SSNDIPTSII YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE ELTFDIDYKT SEVNQRVFSL DEVFEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRKGI NEYINLYSQQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA AFKTVEEKSI KETLSLLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY ITQQVAPKNL DNPSKKEQDL IAKKTEKAKY LSLETIKLAL EEFNKHRDID KQCRFEEILA NFAAIPMIFD EIAQNKDNLA QISIKYQNQG KKDLLQASAE DDVKAIKDLL DQTNNLLHRL KIFHISQSED KANILDKDEH FYLVFEECYF ELANIVPLYN KIRNYITQKP YSDEKFKLNF ENSTLANGWD KNKEPDNTAI LFIKDDKYYL GVMNKKNNKI FDDKAIKENK GEGYKKIVYK LLPGANKMLP KVFFSAKSIK FYNPSEDILR IRNHSTHTKN GNPQKGYEKF EFNIEDCRKF IDFYKESISK HPEWKDFGFR FSDTQRYNSI DEFYREVENQ GYKLTFENIS ESYIDSVVNQ GKLYLFQIYN KDFSAYSKGR PNLHTLYWKA LFDERNLQDV VYKLNGEAEL FYRKKSIPKK ITHPAKEAIA NKNKDNPKKE SFFEYDLIKD KRFTEDKFFF HCPITINFKS SGANKFNDEI NLLLKEKAND VHILSIDRGE RHLAYYTLVD GKGNIIKQDT FNIIGNDRMK TNYHDKLAAI EKDRDSARKD WKKINNIKEM KEGYLSQVVH EIAKLVIEHN AIVVFEDLNF GFKRGRFKVE KQVYQKLEKM LIEKLNYLVF KDNEFDKTGG VLRAYQLTAP FETFKKMGKQ TGIIYYVPAG FTSKICPVTG FVNQLYPKYE SVSKSQEFFS KFDKICYNLD KGYFEFSFDY KNFGDKAAKG KWTIASFGSR LINFRNSDKN HNWDTREVYP TKELEKLLKD YSIEYGHGEC IKAAICGESD KKFFAKLTSI LNTILQMRNS KTGTELDYLI SPVADVNGNF FDSRQAPKNM PQDADANGAY HIGLKGLMLL DRIKNNQEGK KLNLVIKNEE YFEFVQNRNN	(SEQ ID NO: 128)
WP_003034647	MSIYQEFVNK YSLSKTLRFE LIPQGKTLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF FIEEILSSVC ISEDLLQNYS DVYFKLKKSD DDNLQKDFKS	(SEQ ID NO: 129)

Table 5 - Cas12a orthologs		
type V CRISPR-associated protein Cpf1 [Francisella tularensis]	AKDTIKKQIS EYIKDSEKFK NLFNQNLIDA KKGQESDLIL WLKQSKDNGI ELFKANSDIT DIDEALEIIK SFKGWTTYFK GFHENRKNVY SSDDIPTSII YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE ELTFDIDYKT SEVNQRVFSL DEVFEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRKGI NEYINLYSQQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA AFKTVEEKSI KETLSLLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY ITQQVAPKNL DNPSKKEQDL IAKKTEKAKY LSLETIKLAL EEFNKHRDID KQCRFEEILA NFAAIPMIFD EIAQNKDNLA QISLKYQNQG KKDLLQASAE EDVKAIKDLL DQTNNLLHRL KIFHISQSED KANILDKDEH FYLVFEECYF ELANIVPLYN KIRNYITQKP YSDEKFKLNF ENSTLANGWD KNKEPDNTAI LFIKDDKYYL GVMNKKNNKI FDDKAIKENK GEGYKKIVYK LLPGANKMLP KVFFSAKSIK FYNPSEDILR IRNHSTHTKN GNPQKGYEKF EFNIEDCRKF IDFYKESISK HPEWKDFGFR FSDTQRYNSI DEFYREVENQ GYKLTFENIS ESYIDSVVNQ GKLYLFQIYN KDFSAYSKGR PNLHTLYWKA LFDERNLQDV VYKLNGEAEL FYRKQSIPKK ITHPAKEAIA NKNKDNPKKE SVFEYDLIKD KRFTEDKFFF HCPITINFKS SGANKFNDEI NLLLKEKAND VHILSIDRGE RHLAYYTLVD GKGNIIKQDT FNIIGNDRMK TNYHDKLAAI EKDRDSARKD WKKINNIKEM KEGYLSQVVH EIAKLVIEHN AIVVFEDLNF GFKRGRFKVE KQVYQKLEKM LIEKLNYLVF KDNEFDKTGG VLRAYQLTAP FETFKKMGKQ TGIIYYVPAG FTSKICPVTG FVNQLYPKYE SVSKSQEFFS KFDKICYNLD KGYFEFSFDY KNFGDKAAKG KWTIASFGSR LINFRNSDKN HNWDTREVYP TKELEKLLKD YSIEYGHGEC IKAAICGESD KKFFAKLTSV LNTILQMRNS KTGTELDYLI SPVADVNGNF FDSRQAPKNM PQDADANGAY HIGLKGLMLL DRIKNNQEGK KLNLVIKNEE YFEFVQNRNN	
WP_003040289.1 type V CRISPR-associated protein Cpf1 [Francisella tularensis subsp. novicida U112]	MSIYQEFVNK YSLSKTLRFE LIPQGKTLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF FIEEILSSVC ISEDLLQNYS DVYFKLKKSD DDNLQKDFKS AKDTIKKQIS EYIKDSEKFK NLFNQNLIDA KKGQESDLIL WLKQSKDNGI ELFKANSDIT DIDEALEIIK SFKGWTTYFK GFHENRKNVY SSNDIPTSII YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE ELTFDIDYKT SEVNQRVFSL DEVFEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRKGI NEYINLYSQQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA AFKTVEEKSI KETLSLLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY ITQQIAPKNL DNPSKKEQEL IAKKTEKAKY LSLETIKLAL EEFNKHRDID KQCRFEEILA NFAAIPMIFD EIAQNKDNLA QISIKYQNQG KKDLLQASAE DDVKAIKDLL DQTNNLLHKL KIFHISQSED KANILDKDEH FYLVFEECYF ELANIVPLYN KIRNYITQKP YSDEKFKLNF ENSTLANGWD KNKEPDNTAI LFIKDDKYYL GVMNKKNNKI FDDKAIKENK GEGYKKIVYK LLPGANKMLP KVFFSAKSIK FYNPSEDILR IRNHSTHTKN GSPQKGYEKF EFNIEDCRKF IDFYKQSISK HPEWKDFGFR FSDTQRYNSI DEFYREVENQ GYKLTFENIS ESYIDSVVNQ GKLYLFQIYN KDFSAYSKGR PNLHTLYWKA LFDERNLQDV VYKLNGEAEL FYRKQSIPKK ITHPAKEAIA NKNKDNPKKE SVFEYDLIKD KRFTEDKFFF HCPITINFKS SGANKFNDEI NLLLKEKAND VHILSIDRGE RHLAYYTLVD GKGNIIKQDT FNIIGNDRMK TNYHDKLAAI EKDRDSARKD WKKINNIKEM KEGYLSQVVH EIAKLVIEYN AIVVFEDLNF GFKRGRFKVE KQVYQKLEKM LIEKLNYLVF KDNEFDKTGG VLRAYQLTAP FETFKKMGKQ TGIIYYVPAG FTSKICPVTG FVNQLYPKYE SVSKSQEFFS KFDKICYNLD KGYFEFSFDY KNFGDKAAKG KWTIASFGSR LINFRNSDKN HNWDTREVYP TKELEKLLKD YSIEYGHGEC IKAAICGESD KKFFAKLTSV LNTILQMRNS KTGTELDYLI SPVADVNGNF FDSRQAPKNM PQDADANGAY HIGLKGLMLL GRIKNNQEGK KLNLVIKNEE YFEFVQNRNN	(SEQ ID NO: 130)
KKQ38174 hypothetical protein US54_C0016G0015 [Candidatus Roizmanbacteria bacterium GW2011_GWA2_37_7]	MKSFDSFTNL YSLSKTLKFE MRPVGNTQKM LDNAGVFEKD KLIQKKYGKT KPYFDRLHRE FIEEALTGVE LIGLDENFRT LVDWQKDKKN 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 GHFWKEKYYK ISKFQEKTNW EQFLAIFLYE FNSLFSDKIN TKDGETKQVG YYLFAKDLHN LILSEQIDIP	(SEQ ID NO: 131)

Table 5 - Cas12a orthologs		
	KDSKVTIKDF ADSVLTIYQM AKYFAVEKKR AWLAEYELDS FYTQPDTGYL QFYDNAYEDI VQVYNKLRNY LTKKPYSEEK WKLNFENSTL ANGWDKNKES DNSAVILQKG GKYYLGLITK GHNKIFDDRF QEKFIVGIEG GKYEKIVYKF FPDQAKMFPK VCFSAKGLEF FRPSEEILRI YNNAEFKKGE TYSIDSMQKL IDFYKDCLTK YEGWACYTFR HLKPTEEYQN NIGEFFRDVA EDGYRIDFQG ISDQYIHEKN EKGELHLFEI HNKDWNLDKA RDGKSKTTQK NLHTLYFESL FSNDNVVQNF PIKLNGQAEI FYRPKTEKDK LESKKDKKGN KVIDHKRYSE NKIFFHVPLT LNRTKNDSYR FNAQINNFLA NNKDINIIGV DRGEKHLVYY SVITQASDIL ESGSLNELNG VNYAEKLGKK AENREQARRD WQDVQGIKDL KKGYISQVVR KLADLAIKHN AIIILEDLNM RFKQVRGGIE KSIYQQLEKA LIDKLSFLVD KGEKNPEQAG HLLKAYQLSA PFETFQKMGK QTGIIFYTQA SYTSKSDPVT GWRPHLYLKY FSAKKAKDDI AKFTKIEFVN DRFELTYDIK DFQQAKEYPN KTVWKVCSNV ERFRWDKNLN QNKGGYTHYT NITENIQELF TKYGIDITKD LLTQISTIDE KQNTSFFRDF IFYFNLICQI RNTDDSEIAK KNGKDDFILS PVEPFFDSRK DNGNKLPENG DDNGAYNIAR KGIVILNKIS QYSEKNENCE KMKWGDLYVS NIDWDNFVTQ ANARH	
WP_022097749 type V CRISPR-associated protein Cpf1 [Eubacterium eligens CAG:72]	MNGNRSIVYR EFVGVTPVAK TLRNELRPVG HTQEHIIQNG LIQEDELRQE KSTELKNIMD DYYREYIDKS LSGLTDLDFT LLFELMNSVQ SSLSKDNKKA LEKEHNKMRE QICTHLQSDS DYKNMFNAKL FKEILPDFIK NYNQYDVKDK AGKLETLALF NGFSTYFTDF FEKRKNVFTK EAVSTSIAYR IVHENSLIFL ANMTSYKKIS EKALDEIEVI EKNNQDKMGD WELNQIFNPD FYNMVLIQSG IDFYNEICGV VNAHMNLYCQ QTKNNYNLFK MRKLHKQILA YTSTSFEVPK MFEDDMSVYN AVNAFIDETE KGNIIGKLKD IVNKYDELDE KRIYISKDFY ETLSCFMSGN WNLITGCVEN FYDENIHAKG KSKEEKVKKA VKEDKYKSIN DVNDLVEKYI DEKERNEFKN SNAKQYIREI SNIITDTETA HLEYDEHISL IESEEKADEI KKRLDMYMNM YHWVKAFIVD EVLDRDEMFY SDIDDIYNIL ENIVPLYNRV RNYVTQKPYT SKKIKLNFQS PTLANGWSQS KEFDNNAIIL IRDNKYYLAI FNAKNKPDKK IIQGNSDKKN DNDYKKMVYN LLPGANKMLP KVFLSKKGIE TFKPSDYIIS GYNAHKHIKT SENFDISFCR DLIDYFKNSI EKHAEWRKYE FKFSATDSYN DISEFYREVE MQGYRIDWTY ISEADINKLD EEGKIYLFQI YNKDFAENST GKENLHTMYF KNIFSEENLK NIVIKLNGQA ELFYRKASVK NPVKHKKDSV LVNKTYKNQL DNGDVVRIPI PDDIYNEIYK MYNGYIKESD LSEAAKEYLD KVEVRTAQKD IVKDYRYTVD KYFIHTPITI NYKVTARNNV NDMAVKYIAQ NDDIHVIGID RGERNLIYIS VIDSHGNIVK QKSYNILNNY DYKKKLVEKE KTREYARKNW KSIGNIKELK EGYISGVVHE IAMLMVEYNA IIAMEDLNYG FKRGRFKVER QVYQKFESML INKLNYFASK GKSVDEPGGL LKGYQLTYVP DNIKNLGKQC GVIFYVPAAF TSKIDPSTGF ISAFNFKSIS TNASRKQFFM QFDEIRYCAE KDMFSFGFDY NNFDTYNITM GKTQWTVYTN GERLQSEFNN ARRTGKTKSI NLTETIKLLL EDNEINYADG HDVRIDMEKM YEDKNSEFFA QLLSLYKLTV QMRNSYTEAE EQEKGISYDK IISPVINDEG EFFDSDNYKE SDDKECKMPK DADANGAYCI ALKGLYEVLK IKSEWTEDGF DRNCLKLPHA EWLDFIQNKR YE	(SEQ ID NO: 132)
WP_021739647 hypothetical protein [Eubacterium ramulus]	MIKKTIDTVL NVRPIFVGIQ HLYFYEGPCR FGEGDELMPE YDAMMNQEMN AAYVNEVVQH ETEGVHIMDP IYVERDDWFR SPEAMYEKMA EDIDKVDFYL FHFGIGRGDI YLEFAERYKK PVGAAPGLCC DGIGNTAAVK NRGLEAYAFM SWDEFDTWMR VLRVRKCLKN TRVLLAVRWD SNRSYSSYDN FINQSDVTNK WGIQFRHVNV HELLDQTHPV DPTTNPSTPG RKALNINDED MKEIEKITDE LIANAEACTM EPDMVKKTIQ AYYTVQKLLD AYDCNAFTAP CPDLCSTRRF SEEKFTLCMT HSLNDENGIS SACEYDINSV IGKVIMTNLS GKAPYMGNTN AIVFDKEGHM IPFHKFNDNT IEDIADKTNL YMTFHSTPNR NLKGLKAEKE RYRLAPFAYS GFGATIRYDF AQDIGQVITM IRISPDATKI FIAKGTISGG AGYEMKNCDQ GVFFNVADKV DFYHKQQYFG NHTVLAYGDY VEELKMLAEA LGIEAVIA	(SEQ ID NO: 133)
gi|800943167 WP_045971446.1 type V CRISPR-associated protein Cpf1	MKNFSNLYQV SKTVRFELKP IGNTLENIKN KSLLKNDSIR AESYQKMKKT IDEFHKYFID LALNNKKLSY LNEYIALYTQ SAEAKKEDKF KADFKKVQDN LRKEIVSSFT EGEAKAIFSV LDKKELITIE LEKWKNENNL AVYLDESFKS FTTYFTGFHQ NRKNMYSAEA NSTAIAYRLI HENLPKFIEN SKAFEKSSQI AELQPKIEKL YKEFEAYLNV NSISELFEID YFNEVLTQKG ITVYNNIIGG RTATEGKQKI QGLNEIINLY NQTKPKNERL PKLKQLYKQI LSDRISLSFL PDAFTEGKQV LKAVFEFYKI NLLSYKQDGV EESQNLLELI QQVVKNLGNQ	(SEQ ID NO: 134)

Table 5 - Cas12a orthologs		
[Flavobacterium sp. 316]	DVNKIYLKND TSLTTIAQQL FGDFSVFSAA LQYRYETVVN PKYTAEYQKA NEAKQEKLDK EKIKFVKQDY FSIAFLQEVV ADYVKTLDEN LDWKQKYTPS CIADYFTTHF IAKKENEADK TFNFIANIKA KYQCIQGILE QADDYEDELK QDQKLIDNIK FFLDAILEVV HFIKPLHLKS ESITEKDNAF YDVFENYYEA LNVVTPLYNM VRNYVTQKPY STEKIKLNFE NAQLLNGWDA NKEKDYLTTI LKRDGNYFLA IMDKKHNKTF QQFTEDDENY EKIVYKLLPG VNKMLPKVFF SNKNIAFFNP SKEILDNYKN NTHKKGATFN LKDCHALIDF FKDSLNKHED WKYFDFQFSE TKTYQDLSGF YKEVEHQGYK INFKKVSVSQ IDTLIEEGKM YLFQIYNKDF SPYAKGKPNM HTLYWKALFE TQNLENVIYK LNGQAEIFFR KASIKKKNII THKAHQPIAA KNPLTPTAKN TFAYDLIKDK RYTVDKFQFH VPITMNFKAT GNSYINQDVL AYLKDNPEVN IIGLDRGERH LVYLTLIDQK GTILLQESLN VIQDEKTHTP YHTLLDNKEI ARDKARKNWG SIESIKELKE GYISQVVHKI TKMMIEHNAI VVMEDLNFGF KRGRFKVEKQ IYQKLEKMLI DKLNYLVLKD KQPHELGGLY NALQLTNKFE SFQKMGKQSG FLFYVPAWNT SKIDPTTGFV NYFYTKYENV EKAKTFFSKF DSILYNKTKG YFEFVVKNYS DFNPKAADTR QEWTICTHGE RIETKRQKEQ NNNFVSTTIQ LTEQFVNFFE KVGLDLSKEL KTQLIAQNEK SFFEELFHLL KLTLQMRNSE SHTEIDYLIS PVANEKGIFY DSRKATASLP IDADANGAYH IAKKGLWIME QINKTNSEDD LKKVKLAISN REWLQYVQQV QKK	
WP_044110123.1 type V CRISPR-associated protein Cpf1 [Prevotella brevis]	MKQFTNLYQL SKTLRFELKP IGKTLEHINA NGFIDNDAHR AESYKKVKKL IDDYHKDYIE NVLNNFKLNG EYLQAYFDLY SQDTKDKQFK DIQDKLRKSI ASALKGDDRY KTIDKKELIR QDMKTFLKKD TDKALLDEFY EFTTYFTGYH ENRKNMYSDE AKSTAIAYRL IHDNLPKFID NIAVFKKIAN TSVADNFSTI YKNFEEYLNV NSIDEIFSLD YYNIVLTQTQ IEVYNSIIGG RTLEDDTKIQ GINEFVNLYN QQLANKKDRL PKLKPLFKQI LSDRVQLSWL QEEFNTGADV LNAVKEYCTS YFDNVEESVK VLLTGISDYD LSKIYITNDL ALTDVSQRMF GEWSIIPNAI EQRLRSDNPK KTNEKEEKYS DRISKLKKLP KSYSLGYINE CISELNGIDI ADYYATLGAI NTESKQEPSI PTSIQVHYNA LKPILDTDYP REKNLSQDKL TVMQLKDLLD DFKALQHFIK PLLGNGDEAE KDEKFYGELM QLWEVIDSIT PLYNKVRNYC TRKPFSTEKI KVNFENAQLL DGWDENKEST NASIILRKNG MYYLGIMKKE YRNILTKPMP SDGDCYDKVV YKFFKDITTM VPKCTTQMKS VKEHFSNSND DYTLFEKDKF IAPVVITKEI FDLNNVLYNG VKKFQIGYLN NTGDSFGYNH AVEIWKSFCL KFLKAYKSTS IYDFSSIEKN IGCYNDLNSF YGAVNLLLYN LTYRKVSVDY IHQLVDEDKM YLFMIYNKDF STYSKGTPNM HTLYWKMLFD ESNLNDVVYK LNGQAEVFYR KKSITYQHPT HPANKPIDNK NVNNPKKQSN FEYDLIKDKR YTVDKFMFHV PITLNFKGMG NGDINMQVRE YIKTTDDLHF IGIDRGERHL LYICVINGKG EIVEQYSLNE IVNNYKGTEY KTDYHTLLSE RDKKRKEERS SWQTIEGIKE LKSGYLSQVI HKITQLMIKY NAIVLLEDLN MGFKRGRQKV ESSVYQQFEK ALIDKLNYLV DKNKDANEIG GLLHAYQLTN DPKLPNKNSK QSGFLFYVPA WNTSKIDPVT GFVNLLDTRY ENVAKAQAFF KKFDSIRYNK EYDRFEFKFD YSNFTAKAED TRTQWTLCTY GTRIETFRNA EKNSNWDSRE IDLTTEWKTL FTQHNIPLNA NLKEAILLQA NKNFYTDILH LMKLTLQMRN SVTGTDIDYM VSPVANECGE FFDSRKVKEG LPVNADANGA YNIARKGLWL AQQIKNANDL SDVKLAITNK EWLQFAQKKQ YLKD	(SEQ ID NO: 135)
WP_036388671.1 type V CRISPR-associated protein Cpf1 [Moraxella caprae]	MLFQDFTHLY PLSKTMRFEL KPIGKTLEHI HAKNFLSQDE TMADMYQKVK AILDDYHRDF IADMMGEVKL TKLAEFYDVY LKFRKNPKDD GLQKQLKDLQ AVLRKEIVKP IGNGGKYKAG YDRLFGAKLF KDGKELGDLA KFVIAQEGES SPKLAHLAHF EKFSTYFTGF HDNRKNMYSD EDKHTAITYR LIHENLPRFI DNLQILATIK QKHSALYDQI INELTASGLD VSLASHLDGY HKLLTQEGIT AYNTLLGGIS GEAGSRKIQG INELINSHHN QHCHKSERIA KLRPLHKQIL SDGMGVSFLP SKFADDSEMC QAVNEFYRHY ADVFAKVQSL FDGFDDHQKD GIYVEHKNLN ELSKQAFGDF ALLGRVLDGY YVDVVNPEFN ERFAKAKTDN AKAKLTKEKD KFIKGVHSLA SLEQAIEHYT ARHDDESVQA GKLGQYFKHG LAGVDNPIQK IHNNHSTIKG FLERERPAGE RALPKIKSGK NPEMTQLRQL KELLDNALNV AHFAKLLTTK TTLDNQDGNF YGEFGALYDE LAKIPTLYNK VRDYLSQKPF STEKYKLNFG NPTLLNGWDL NKEKDNFGII LQKDGCYYLA LLDKAHKKVF DNAPNTGKNV YQKMIYKLLP GPNKMLPKVF FAKSNLDYYN PSAELLDKYA QGTHKKGNNF NLKDCHALID FFKAGINKHP EWQHFGFKFS PTSSYQDLSD FYREVEPQGY QVKFVDINAD YINELVEQGQ LYLFQIYNKD	(SEQ ID NO: 136)

Table 5 - Cas12a orthologs		
	FSPKAHGKPN LHTLYFKALF SKDNLANPIY KLNGEAQIFY RKASLDMNET TIHRAGEVLE NKNPDNPKKR QFVYDIIKDK RYTQDKFMLH VPITMNFGVQ GMTIKEFNKK VNQSIQQYDE VNVIGIDRGE RHLLYLTVIN SKGEILEQRS LNDITTASAN GTQMTTPYHK ILDKREIERL NARVGWGEIE TIKELKSGYL SHVVHQISQL MLKYNAIVVL EDLNFGFKRG RFKVEKQIYQ NFENALIKKL NHLVLKDEAD DEIGSYKNAL QLTNNFTDLK SIGKQTGFLF YVPAWNTSKI DPETGFVDLL KPRYENIAQS QAFFGKFDKI CYNADKDYFE FHIDYAKFTD KAKNSRQIWK ICSHGDKRYV YDKTANQNKG ATKGINVNDE LKSLFARHHI NDKQPNLVMD ICQNNDKEFH KSLIYLLKTL LALRYSNASS DEDFILSPVA NDEGMFFNSA LADDTQPQNA DANGAYHIAL KGLWVLEQIK NSDDLNKVKL AIDNQTWLNF AQNR	
WP_020988726.1 type V CRISPR-associated protein Cpf1 [Leptospira inadai]	MEDYSGFVNI YSIQKTLRFE LKPVGKTLEH IEKKGFLKKD KIRAEDYKAV KKIIDKYHRA YIEEVFDSVL HQKKKKDKTR FSTQFIKEIK EFSELYYKTE KNIPDKERLE ALSEKLRKML VGAFKGEFSE EVAEKYKNLF SKELIRNEIE KFCETDEERK QVSNFKSFTT YFTGFHSNRQ NIYSDEKKST AIGYRIIHQN LPKFLDNLKI IESIQRRFKD FPWSDLKKNL KKIDKNIKLT EYFSIDGFVN VLNQKGIDAY NTILGGKSEE SGEKIQGLNE YINLYRQKNN IDRKNLPNVK ILFKQILGDR ETKSFIPEAF PDDQSVLNSI TEFAKYLKLD KKKKSIIAEL KKFLSSFNRY ELDGIYLAND NSLASISTFL FDDWSFIKKS VSFKYDESVG DPKKKIKSPL KYEKEKEKWL KQKYYTISFL NDAIESYSKS QDEKRVKIRL EAYFAEFKSK DDAKKQFDLL ERIEEAYAIV EPLLGAEYPR DRNLKADKKE VGKIKDFLDS IKSLQFFLKP LLSAEIFDEK DLGFYNQLEG YYEEIDSIGH LYNKVRNYLT GKIYSKEKFK LNFENSTLLK GWDENREVAN LCVIFREDQK YYLGVMDKEN NTILSDIPKV KPNELFYEKM VYKLIPTPHM QLPRIIFSSD NLSIYNPSKS ILKIREAKSF KEGKNFKLKD CHKFIDFYKE SISKNEDWSR FDFKFSKTSS YENISEFYRE VERQGYNLDF KKVSKFYIDS LVEDGKLYLF QIYNKDFSIF SKGKPNLHTI YFRSLFSKEN LKDVCLKLNG EAEMFFRKKS INYDEKKKRE GHHPELFEKL KYPILKDKRY SEDKFQFHLP ISLNFKSKER LNFNLKVNEF LKRNKDINII GIDRGERNLL YLVMINQKGE ILKQTLLDSM QSGKGRPEIN YKEKLQEKEI ERDKARKSWG TVENIKELKE GYLSIVIHQI SKLMVENNAI VVLEDLNIGF KRGRQKVERQ VYQKFEKMLI DKLNFLVFKE NKPTEPGGVL KAYQLTDEFQ SFEKLSKQTG FLFYVPSWNT SKIDPRTGFI DFLHPAYENI EKAKQWINKF DSIRFNSKMD WFEFTADTRK FSENLMLGKN RVWVICTTNV ERYFTSKTAN SSIQYNSIQI TEKLKELFVD IPFSNGQDLK PEILRKNDAV FFKSLLFYIK TTLSLRQNNG KKGEEEKDFI LSPVVDSKGR FFNSLEASDD EPKDADANGA YHIALKGLMN LLVLNETKEE NLSRPKWKIK NKDWLEFVWE RNR	(SEQ ID NO: 137)
WP_023936172.1 type V CRISPR-associated protein Cpf1 [Porphyromonas crevioricanis]	MPWIDLKDFT NLYPVSKTLR FELKPVGKTL ENIEKAGILK EDEHRAESYR RVKKIIDTYH KVFIDSSLEN MAKMGIENEI KAMLQSFCEL YKKDHRTEGE DKALDKIRAV LRGLIVGAFT GVCGRRENTV QNEKYESLFK EKLIKEILPD FVLSTEAESL PFSVEEATRS LKEFDSFTSY FAGFYENRKN IYSTKPQSTA IAYRLIHENL PKFIDNILVF QKIKEPIAKE LEHIRADFSA GGYIKKDERL EDIFSLNYYI HVLSQAGIEK YNALIGKIVT EGDGEMKGLN EHINLYNQQR GREDRLPLFR PLYKQILSDR EQLSYLPESF EKDEELLRAL KEFYDHIAED ILGRTQQLMT SISEYDLSRI YVRNDSQLTD ISKKMLGDWN AIYMARERAY DHEQAPKRIT AKYERDRIKA LKGEESISLA NLNSCIAFLD NVRDCRVDTY LSTLGQKEGP HGLSNLVENV FASYHEAEQL LSFPYPEENN LIQDKDNVVL IKNLLDNISD LQRFLKPLWG MGDEPDKDER FYGEYNYIRG ALDQVIPLYN KVRNYLTRKP YSTRKVKLNF GNSQLLSGWD RNKEKDNSCV ILRKGQNFYL AIMNNRHKRS FENKVLPEYK EGEPYFEKMD YKFLPDPNKM LPKVFLSKKG IEIYEPSPKL LEQYGHGTHK KGDTFSMDDL HELIDFFKHS IEAHEDWKQF GFKFSDTATY ENVSSFYREV EDQGYKLSFR KVSESYVYSL IDQGKLYLFQ IYNKDFSPCS KGTPNLHTLY WRMLFDERNL ADVIYKLDGK AEIFFREKSL KNDHPTHPAG KPIKKKSRQK KGEESLFEYD LVKDRRYTMD KFQFHVPITM NFKCSAGSKV NDMVNAHIRE AKDMHVIGID RGERNLLYIC VIDSRGTILD QISLNTINDI DYHDLLESRD KDRQQERRNW QTIEGIKELK QGYLSQAVHR IAELMVAYKA VVALEDLNMG FKRGRQKVES SVYQQFEKQL IDKLNYLVDK KKRPEDIGGL LRAYQFTAPF KSFKEMGKQN GFLFYIPAWN TSNIDPTTGF VNLFHAQYEN VDKAKSFFQK FDSISYNPKK DWFEFAFDYK NFTKKAEGSR SMWILCTHGS RIKNFRNSQK NGQWDSEEFA LTEAFKSLFV RYEIDYTADL	(SEQ ID NO: 138)

Table 5 - Cas12a orthologs		
	KTAIVDEKQK DFFVDLLKLF KLTVQMRNSW KEKDLDYLIS PVAGADGRFF DTREGNKSLP KDADANGAYN IALKGLWALR QIRQTSEGGK LKLAISNKEW LQFVQERSYE KD	
WP_009217842.1 type V CRISPR-associated protein Cpf1 [Bacteroidetes oral taxon 274]	MRKFNEFVGL YPISKTLRFE LKPIGKTLEH IQRNKLLEHD AVRADDYVKV KKIIDKYHKC LIDEALSGFT FDTEADGRSN NSLSEYYLYY NLKKRNEQEQ KTFKTIQNNL RKQIVNKLTQ SEKYKRIDKK ELITTDLPDF LTNESEKELV EKFKNFTTYF TEFHKNRKNM YSKEEKSTAI AFRLINENLP KFVDNIAAFE KVVSSPLAEK INALYEDFKE YLNVEEISRV FRLDYYDELL TQKQIDLYNA IVGGRTEEDN KIQIKGLNQY INEYNQQQTD RSNRLPKLKP LYKQILSDRE SVSWLPPKFD SDKNLLIKIK ECYDALSEKE KVFDKLESIL KSLSTYDLSK IYISNDSQLS YISQKMFGRW DIISKAIRED CAKRNPQKSR ESLEKFAERI DKKLKTIDSI SIGDVDECLA QLGETYVKRV EDYFVAMGES EIDDEQTDTT SFKKNIEGAY ESVKELLNNA DNITDNNLMQ DKGNVEKIKT LLDAIKDLQR FIKPLLGKGD EADKDGVFYG EFTSLWTKLD QVTPLYNMVR NYLTSKPYST KKIKLNFENS TLMDGWDLNK EPDNTTVIFC KDGLYYLGIM GKKYNRVFVD REDLPHDGEC YDKMEYKLLP GANKMLPKVF FSETGIQRFL PSEELLGKYE RGTHKKGAGF DLGDCRALID FFKKSIERHD DWKKFDFKFS DTSTYQDISE FYREVEQQGY KMSFRKVSVD YIKSLVEEGK LYLFQIYNKD FSAHSKGTPN MHTLYWKMLF DEENLKDVVY KLNGEAEVFF RKSSITVQSP THPANSPIKN KNKDNQKKES KFEYDLIKDR RYTVDKFLFH VPITMNFKSV GGSNINQLVK RHIRSATDLH IIGIDRGERH LLYLTVIDSR GNIKEQFSLN EIVNEYNGNT YRTDYHELLD TREGERTEAR RNWQTIQNIR ELKEGYLSQV IHKISELAIK YNAVIVLEDL NFGFMRSRQK VEKQVYQKFE KMLIDKLNYL VDKKKPVAET GGLLRAYQLT GEFESFKTLG KQSGILFYVP AWNTSKIDPV TGFVNLFDTH YENIEKAKVF FDKFKSIRYN SDKDWFEFVV DDYTRFSPKA EGTRRDWTIC TQGKRIQICR NHQRNNEWEG QEIDLTKAFK EHFEAYGVDI SKDLREQINT QNKKEFFEEL LRLLRLTLQM RNSMPSSDID YLISPVANDT GCFFDSRKQA ELKENAVLPM NADANGAYNI ARKGLLAIRK MKQEENDSAK ISLAISNKEW LKFAQTKPYL ED	(SEQ ID NO: 139)
WP_036890108.1 type V CRISPR-associated protein Cpf1 [Porphyromonas crevioricanis]	MDSLKDFTNL YPVSKTLRFE LKPVGKTLEN IEKAGILKED EHRAESYRRV KKIIDTYHKV FIDSSLENMA KMGIENEIKA MLQSFCELYK KDHRTEGEDK ALDKIRAVLR GLIVGAFTGV CGRRENTVQN EKYESLFKEK LIKEILPDFV LSTEAESLPF SVEEATRSLK EFDSFTSYFA GFYENRKNIY STKPQSTAIA YRLIHENLPK FIDNILVFQK IKEPIAKELE HIRADFSAGG YIKKDERLED IFSLNYYIHV LSQAGIEKYN ALIGKIVTEG DGEMKGLNEH INLYNQQRGR EDRLPLFRPL YKQILSDREQ LSYLPESFEK DEELLRALKE FYDHIAEDIL GRTQQLMTSI SEYDLSRIYV RNDSQLTDIS KKMLGDWNAI YMARERAYDH EQAPKRITAK YERDRIKALK GEESISLANL NSCIAFLDNV RDCRVDTYLS TLGQKEGPHG LSNLVENVFA SYHEAEQLLS FPYPEENNLI QDKDNVVLIK NLLDNISDLQ RFLKPLWGMG DEPDKDERFY GEYNYIRGAL DQVIPLYNKV RNYLTRKPYS TRKVKLNFGN SQLLSGWDRN KEKDNSCVIL RKGQNFYLAI MNNRHKRSFE NKMLPEYKEG EPYFEKMDYK FLPDPNKMLP KVFLSKKGIE IYKPSPKLLE QYGHGTHKKG DTFSMDDLHE LIDFFKHSIE AHEDWKQFGF KFSDTATYEN VSSFYREVED QGYKLSFRKV SESYVYSLID QGKLYLFQIY NKDFSPCSKG TPNLHTLYWR MLFDERNLAD VIYKLDGKAE IFFREKSLKN DHPTHPAGKP IKKKSRQKKG EESLFEYDLV KDRRYTMDKF QFHVPITMNF KCSAGSKVND MVNAHIREAK DMHVIGIDRG ERNLLYICVI DSRGTILDQI SLNTINDIDY HDLLESRDKD RQQEHRNWQT IEGIKELKQG YLSQAVHRIA ELMVAYKAVV ALEDLNMGFK RGRQKVESSV YQQFEKQLID KLNYLVDKKK RPEDIGGLLR AYQFTAPFKS FKEMGKQNGF LFYIPAWNTS NIDPTTGFVN LFHVQYENVD KAKSFFQKFD SISYNPKKDW FEFAFDYKNF TKKAEGSRSM WILCTHGSRI KNFRNSQKNG QWDSEEFALT EAFKSLFVRY EIDYTADLKT AIVDEKQKDF FVDLLKLFKL TVQMRNSWKE KDLDYLISPV AGADGRFFDT REGNKSLPKD ADANGAYNIA LKGLWALRQI RQTSEGGKLK LAISNKEWLQ FVQERSYEKD	(SEQ ID NO: 140)
WP_036887416.1 type V CRISPR-associated protein Cpf1	MDSLKDFTNL YPVSKTLRFE LKPVGKTLEN IEKAGILKED EHRAESYRRV KKIIDTYHKV FIDSSLENMA KMGIENEIKA MLQSFCELYK KDHRTEGEDK ALDKIRAVLR GLIVGAFTGV CGRRENTVQN EKYESLFKEK LIKEILPDFV LSTEAESLPF SVEEATRSLK EFDSFTSYFA GFYENRKNIY STKPQSTAIA YRLIHENLPK FIDNILVFQK IKEPIAKELE HIRADFSAGG YIKKDERLED	(SEQ ID NO: 141)

Table 5 - Cas12a orthologs		
[Porphyromonas crevioricanis]	IFSLNYYIHV LSQAGIEKYN ALIGKIVTEG DGEMKGLNEH INLYNQQRGR EDRLPLFRPL YKQILSDREQ LSYLPESFEK DEELLRALKE FYDHIAEDIL GRTQQLMTSI SEYDLSRIYV RNDSQLTDIS KKMLGDWNAI YMARERAYDH EQAPKRITAK YERDRIKALK GEESISLANL NSCIAFLDNV RDCRVDTYLS TLGQKEGPHG LSNLVENVFA SYHEAEQLLS FPYPEENNLI QDKDNVVLIK NLLDNISDLQ RFLKPLWGMG DEPDKDERFY GEYNYIRGAL DQVIPLYNKV RNYLTRKPYS TRKVKLNFGN SQLLSGWDRN KEKDNSCVIL RKGQNFYLAI MNNRHKRSFE NKVLPEYKEG EPYFEKMDYK FLPDPNKMLP KVFLSKKGIE IYKPSPKLLE QYGHGTHKKG DTFSMDDLHE LIDFFKHSIE AHEDWKQFGF KFSDTATYEN VSSFYREVED QGYKLSFRKV SESYVYSLID QGKLYLFQIY NKDFSPCSKG TPNLHTLYWR MLFDERNLAD VIYKLDGKAE IFFREKSLKN DHPTHPAGKP IKKKSRQKKG EESLFEYDLV KDRHYTMDKF QFHVPITMNF KCSAGSKVND MVNAHIREAK DMHVIGIDRG ERNLLYICVI DSRGTILDQI SLNTINDIDY HDLLESRDKD RQQERRNWQT IEGIKELKQG YLSQAVHRIA ELMVAYKAVV ALEDLNMGFK RGRQKVESSV YQQFEKQLID KLNYLVDKKK RPEDIGGLLR AYQFTAPFKS FKEMGKQNGF LFYIPAWNTS NIDPTTGFVN LFHAQYENVD KAKSFFQKFD SISYNPKKDW FEFAFDYKNF TKKAEGSRSM WILCTHGSRI KNFRNSQKNG QWDSEEFALT EAFKSLFVRY EIDYTADLKT AIVDEKQKDF FVDLLKLFKL TVQMRNSWKE KDLDYLISPV AGADGRFFDT REGNKSLPKD ADANGAYNIA LKGLWALRQI RQTSEGGKLK LAISNKEWLQ FVQERSYEKD	
WP_023941260.1 type V CRISPR-associated protein Cpf1 [Porphyromonas crevioricanis]	MDSLKDFTNL YPVSKTLRFE LKPVGKTLEN IEKAGILKED EHRAESYRRV KKIIDTYHKV FIDSSLENMA KMGIENEIKA MLQSFCELYK KDHRTEGEDK ALDKIRAVLR GLIVGAFTGV CGRRENTVQN EKYESLFKEK LIKEILPDFV LSTEAESLPF SVEEATRSLK EFDSFTSYFA GFYENRKNIY STKPQSTAIA YRLIHENLPK FIDNILVFQK IKEPIAKELE HIRADFSAGG YIKKDERLED IFSLNYYIHV LSQAGIEKYN ALIGKIVTEG DGEMKGLNEH INLYNQQRGR EDRLPLFRPL YKQILSDREQ LSYLPESFEK DEELLRALKE FYDHIAEDIL GRTQQLMTSI SEYDLSRIYV RNDSQLTDIS KKMLGDWNAI YMARERAYDH EQAPKRITAK YERDRIKALK GEESISLANL NSCIAFLDNV RDCRVDTYLS TLGQKEGPHG LSNLVENVFA SYHEAEQLLS FPYPEENNLI QDKDNVVLIK NLLDNISDLQ RFLKPLWGMG DEPDKDERFY GEYNYIRGAL DQVIPLYNKV RNYLTRKPYS TRKVKLNFGN SQLLSGWDRN KEKDNSCVIL RKGQNFYLAI MNNRHKRSFE NKVLPEYKEG EPYFEKMDYK FLPDPNKMLP KVFLSKKGIE IYKPSPKLLE QYGHGTHKKG DTFSMDDLHE LIDFFKHSIE AHEDWKQFGF KFSDTATYEN VSSFYREVED QGYKLSFRKV SESYVYSLID QGKLYLFQIY NKDFSPCSKG TPNLHTLYWR MLFDERNLAD VIYKLDGKAE IFFREKSLKN DHPTHPAGKP IKKKSRQKKG EESLFEYDLV KDRRYTMDKF QFHVPITMNF KCSAGSKVND MVNAHIREAK DMHVIGIDRG ERNLLYICVI DSRGTILDQI SLNTINDIDY HDLLESRDKD RQQERRNWQT IEGIKELKQG YLSQAVHRIA ELMVAYKAVV ALEDLNMGFK RGRQKVESSV YQQFEKQLID KLNYLVDKKK RPEDIGGLLR AYQFTAPFKS FKEMGKQNGF LFYIPAWNTS NIDPTTGFVN LFHAQYENVD KAKSFFQKFD SISYNPKKDW FEFAFDYKNF TKKAEGSRSM WILCTHGSRI KNFRNSQKNG QWDSEEFALT EAFKSLFVRY EIDYTADLKT AIVDEKQKDF FVDLLKLFKL TVQMRNSWKE KDLDYLISPV AGADGRFFDT REGNKSLPKD ADANGAYNIA LKGLWALRQI RQTSEGGKLK LAISNKEWLQ FVQERSYEKD	(SEQ ID NO: 142)
WP_037975888.1 type V CRISPR-associated protein Cpf1 [Synergistes jonesii]	MANSLKDFTN IYQLSKTLRF ELKPIGKTEE HINRKLIIMH DEKRGEDYKS VTKLIDDYHR KFIHETLDPA HFDWNPLAEA LIQSGSKNNK ALPAEQKEMR EKIISMFTSQ AVYKKLFKKE LFSELLPEMI KSELVSDLEK QAQLDAVKSF DKFSTYFTGF HENRKNIYSK KDTSTSIAFR IVHQNFPKFL ANVRAYTLIK ERAPEVIDKA QKELSGILGG KTLDDIFSIE SFNNVLTQDK IDYYNQIIGG VSGKAGDKKL RGVNEFSNLY RQQHPEVASL RIKMVPLYKQ ILSDRTTLSF VPEALKDDEQ AINAVDGLRS ELERNDIFNR IKRLFGKNNL YSLDKIWIKN SSISAFSNEL FKNWSFIEDA LKEFKENEFN GARSAGKKAE KWLKSKYFSF ADIDAAVKSY SEQVSADISS APSASYFAKF TNLIETAAEN GRKFSYFAAE SKAFRGDDGK TEIIKAYLDS LNDILHCLKP FETEDISDID TEFYSAFAEI YDSVKDVIPV YNAVRNYTTQ KPFSTEKFKL NFENPALAKG WDKNKEQNNT AIILMKDGKY YLGVIDKNNK LRADDLADDG SAYGYMKMNY KFIPTPHMEL PKVFLPKRAP KRYNPSREIL LIKENKTFIK DKNFNRTDCH KLIDFFKDSI	(SEQ ID NO: 143)

Table 5 - Cas12a orthologs		
	NKHKDWRTFG FDFSDTDSYE DISDFYMEVQ DQGYKLTFTR LSAEKIDKWV EEGRLFLFQI YNKDFADGAQ GSPNLHTLYW KAIFSEENLK DVVLKLNGEA ELFFRRKSID KPAVHAKGSM KVNRRDIDGN PIDEGTYVEI CGYANGKRDM ASLNAGARGL IESGLVRITE VKHELVKDKR YTIDKYFFHV PFTINFKAQG QGNINSDVNL FLRNNKDVNI IGIDRGERNL VYVSLIDRDG HIKLQKDFNI IGGMDYHAKL NQKEKERDTA RKSWKTIGTI KELKEGYLSQ VVHEIVRLAV DNNAVIVMED LNIGFKRGRF KVEKQVYQKF EKMLIDKLNY LVFKDAGYDA PCGILKGLQL TEKFESFTKL GKQCGIIFYI PAGYTSKIDP TTGFVNLFNI NDVSSKEKQK DFIGKLDSIR FDAKRDMFTF EFDYDKFRTY QTSYRKKWAV WTNGKRIVRE KDKDGKFRMN DRLLTEDMKN ILNKYALAYK AGEDILPDVI SRDKSLASEI FYVFKNTLQM RNSKRDTGED FIISPVLNAK GRFFDSRKTD AALPIDADAN GAYHIALKGS LVLDAIDEKL KEDGRIDYKD MAVSNPKWFE FMQTRKFDF	
WP_081839471.1 type V CRISPR-associated protein Cpf1 [Synergistes jonesii]	MENMANSLKD FTNIYQLSKT LRFELKPIGK TEEHINRKLI IMHDEKRGED YKSVTKLIDD YHRKFIHETL DPAHFDWNPL AEALIQSGSK NNKALPAEQK EMREKIISMF TSQAVYKKLF KKELFSELLP EMIKSELVSD LEKQAQLDAV KSFDKFSTYF TGFHENRKNI YSKKDTSTSI AFRIVHQNFP KFLANVRAYT LIKERAPEVI DKAQKELSGI LGGKTLDDIF SIESFNNVLT QDKIDYYNQI IGGVSGKAGD KKLRGVNEFS NLYRQQHPEV ASLRIKMVPL YKQILSDRTT LSFVPEALKD DEQAINAVDG LRSELERNDI FNRIKRLFGK NNLYSLDKIW IKNSSISAFS NELFKNWSFI EDALKEFKEN EFNGARSAGK KAEKWLKSKY FSFADIDAAV KSYSEQVSAD ISSAPSASYF AKFTNLIETA AENGRKFSYF AAESKAFRGD DGKTEIIKAY LDSLNDILHC LKPFETEDIS DIDTEFYSAF AEIYDSVKDV IPVYNAVRNY TTQKPFSTEK FKLNFENPAL AKGWDKNKEQ NNTAIILMKD GKYYLGVIDK NNKLRADDLA DDGSAYGYMK MNYKFIPTPH MELPKVFLPK RAPKRYNPSR EILLIKENKT FIKDKNFNRT DCHKLIDFFK DSINKHKDWR TFGFDFSDTD SYEDISDFYM EVQDQGYKLT FTRLSAEKID KWVEEGRLFL FQIYNKDFAD GAQGSPNLHT LYWKAIFSEE NLKDVVLKLN GEAELFFRRK SIDKPAVHAK GSMKVNRRDI DGNPIDEGTY VEICGYANGK RDMASLNAGA RGLIESGLVR ITEVKHELVK DKRYTIDKYF FHVPFTINFK AQGQGNINSD VNLFLRNNKD VNIIGIDRGE RNLVYVSLID RDGHIKLQKD FNIIGGMDYH AKLNQKEKER DTARKSWKTI GTIKELKEGY LSQVVHEIVR LAVDNNAVIV MEDLNIGFKR GRFKVEKQVY QKFEKMLIDK LNYLVFKDAG YDAPCGILKG LQLTEKFESF TKLGKQCGII FYIPAGYTSK IDPTTGFVNL FNINDVSSKE KQKDFIGKLD SIRFDAKRDM FTFEFDYDKF RTYQTSYRKK WAVWTNGKRI VREKDKDGKF RMNDRLLTED MKNILNKYAL AYKAGEDILP DVISRDKSLA SEIFYVFKNT LQMRNSKRDT GEDFIISPVL NAKGRFFDSR KTDAALPIDA DANGAYHIAL KGSLVLDAID EKLKEDGRID YKDMAVSNPK WFEFMQTRKF DF	(SEQ ID NO: 144)
WP_006283774.1 type V CRISPR-associated protein Cpf1 [Prevotella bryantii B14]	MQINNLKIIY MKFTDFTGLY SLSKTLRFEL KPIGKTLENI KKAGLLEQDQ HRADSYKKVK KIIDEYHKAF IEKSLSNFEL KYQSEDKLDS LEEYLMYYSM KRIEKTEKDK FAKIQDNLRK QIADHLKGDE SYKTIFSKDL IRKNLPDFVK SDEERTLIKE FKDFTTYFKG FYENRENMYS AEDKSTAISH RIIHENLPKF VDNINAFSKI ILIPELREKL NQIYQDFEEY LNVESIDEIF HLDYFSMVMT QKQIEVYNAI IGGKSTNDKK IQGLNEYINL YNQKHKDCKL PKLKLLFKQI LSDRIAISWL PDNFKDDQEA LDSIDTCYKN LLNDGNVLGE GNLKLLLENI DTYNLKGIFI RNDLQLTDIS QKMYASWNVI QDAVILDLKK QVSRKKKESA EDYNDRLKKL YTSQESFSIQ YLNDCLRAYG KTENIQDYFA KLGAVNNEHE QTINLFAQVR NAYTSVQAIL TTPYPENANL AQDKETVALI KNLLDSLKRL QRFIKPLLGK GDESDKDERF YGDFTPLWET LNQITPLYNM VRNYMTRKPY SQEKIKLNFE NSTLLGGWDL NKEHDNTAII LRKNGLYYLA IMKKSANKIF DKDKLDNSGD CYEKMVYKLL PGANKMLPKV FFSKSRIDEF KPSENIIENY KKGTHKKGAN FNLADCHNLI DFFKSSISKH EDWSKFNFHF SDTSSYEDLS DFYREVEQQG YSISFCDVSV EYINKMVEKG DLYLFQIYNK DFSEFSKGTP NMHTLYWNSL FSKENLNNII YKLNGQAEIF FRKKSLNYKR PTHPAHQAIK NKNKCNEKKE SIFDYDLVKD KRYTVDKFQF HVPITMNFKS TGNTNINQQV IDYLRTEDDT HIIGIDRGER HLLYLVVIDS HGKIVEQFTL NEIVNEYGGN IYRTNYHDLL DTREQNREKA RESWQTIENI KELKEGYISQ VIHKITDLMQ KYHAVVVLED LNMGFMRGRQ KVEKQVYQKF EEMLINKLNY LVNKKADQNS AGGLLHAYQL TSKFESFQKL GKQSGFLFYI PAWNTSKIDP VTGFVNLFDT	(SEQ ID NO: 145)

Table 5 - Cas12a orthologs		
	RYESIDKAKA FFGKFDSIRY NADKDWFEFA FDYNNFTTKA EGTRTNWTIC TYGSRIRTFR NQAKNSQWDN EEIDLTKAYK AFFAKHGINI YDNIKEAIAM ETEKSFFEDL LHLLKLTLQM RNSITGTTTD YLISPVHDSK GNFYDSRICD NSLPANADAN GAYNIARKGL MLIQQIKDST SSNRFKFSPI TNKDWLIFAQ EKPYLND	
WP_024988992 type V CRISPR-associated protein Cpf1 [Prevotella albensis]	MNIKNFTGLY PLSKTLRFEL KPIGKTKENI EKNGILTKDE QRAKDYLIVK GFIDEYHKQF IKDRLWDFKL PLESEGEKNS LEEYQELYEL TKRNDAQEAD FTEIKDNLRS SITEQLTKSG SAYDRIFKKE FIREDLVNFL EDEKDKNIVK QFEDFTTYFT GFYENRKNMY SSEEKSTAIA YRLIHQNLPK FMDNMRSFAK IANSSVSEHF SDIYESWKEY LNVNSIEEIF QLDYFSETLT QPHIEVYNYI IGKKVLEDGT EIKGINEYVN LYNQQQKDKS KRLPFLVPLY KQILSDREKL SWIAEEFDSD KKMLSAITES YNHLHNVLMG NENESLRNLL LNIKDYNLEK INITNDLSLT EISQNLFGRY DVFTNGIKNK LRVLTPRKKK ETDENFEDRI NKIFKTQKSF SIAFLNKLPQ PEMEDGKPRN IEDYFITQGA INTKSIQKED IFAQIENAYE DAQVFLQIKD TDNKLSQNKT AVEKIKTLLD ALKELQHFIK PLLGSGEENE KDELFYGSFL AIWDELDTIT PLYNKVRNWL TRKPYSTEKI KLNFDNAQLL GGWDVNKEHD CAGILLRKND SYYLGIINKK TNHIFDTDIT PSDGECYDKI DYKLLPGANK MLPKVFFSKS RIKEFEPSEA IINCYKKGTH KKGKNFNLTD CHRLINFFKT SIEKHEDWSK FGFKFSDTET YEDISGFYRE VEQQGYRLTS HPVSASYIHS LVKEGKLYLF QIWNKDFSQF SKGTPNLHTL YWKMLFDKRN LSDVVYKLNG QAEVFYRKSS IEHQNRIIHP AQHPITNKNE LNKKHTSTFK YDIIKDRRYT VDKFQFHVPI TINFKATGQN NINPIVQEVI RQNGITHIIG IDRGERHLLY LSLIDLKGNI IKQMTLNEII NEYKGVTYKT NYHNLLEKRE KERTEARHSW SSIESIKELK DGYMSQVIHK ITDMMVKYNA IVVLEDLNGG FMRGRQKVEK QVYQKFEKKL IDKLNYLVDK KLDANEVGGV LNAYQLTNKF ESFKKIGKQS GFLFYIPAWN TSKIDPITGF VNLFNTRYES IKETKVFWSK FDIIRYNKEK NWFEFVFDYN TFTTKAEGTR TKWTLCTHGT RIQTFRNPEK NAQWDNKEIN LTESFKALFE KYKIDITSNL KESIMQETEK KFFQELHNLL HLTLQMRNSV TGTDIDYLIS PVADEDGNFY DSRINGKNFP ENADANGAYN IARKGLMLIR QIKQADPQKK FKFETITNKD WLKFAQDKPY LKD	(SEQ ID NO: 146)
WP_039658684.1 type V CRISPR-associated protein Cpf1 [Smithella sp. SC_K08D17]	MQTLFENFTN QYPVSKTLRF ELIPQGKTKD FIEQKGLLKK DEDRAEKYKK VKNIIDEYHK DFIEKSLNGL KLDGLEKYKT LYLKQEKDDK DKKAFDKEKE NLRKQIANAF RNNEKFKTLF AKELIKNDLM SFACEEDKKN VKEFEAFTTY FTGFHQNRAN MYVADEKRTA IASRLIHENL PKFIDNIKIF EKMKKEAPEL LSPFNQTLKD MKDVIKGTTL EEIFSLDYFN KTLTQSGIDI YNSVIGGRTP EEGKTKIKGL NEYINTDFNQ KQTDKKKRQP KFKQLYKQIL SDRQSLSFIA EAFKNDTEIL EAIEKFYVNE LLHFSNEGKS TNVLDAIKNA VSNLESFNLT KMYFRSGASL TDVSRKVFGE WSIINRALDN YYATTYPIKP REKSEKYEER KEKWLKQDFN VSLIQTAIDE YDNETVKGKN SGKVIADYFA KFCDDKETDL IQKVNEGYIA VKDLLNTPCP ENEKLGSNKD QVKQIKAFMD SIMDIMHFVR PLSLKDTDKE KDETFYSLFT PLYDHLTQTI ALYNKVRNYL TQKPYSTEKI KLNFENSTLL GGWDLNKETD NTAIILRKDN LYYLGIMDKR HNRIFRNVPK ADKKDFCYEK MVYKLLPGAN KMLPKVFFSQ SRIQEFTPSA KLLENYANET HKKGDNFNLN HCHKLIDFFK DSINKHEDWK NFDFRFSATS TYADLSGFYH EVEHQGYKIS FQSVADSFID DLVNEGKLYL FQIYNKDFSP FSKGKPNLHT LYWKMLFDEN NLKDVVYKLN GEAEVFYRKK SIAEKNTTIH KANESIINKN PDNPKATSTF NYDIVKDKRY TIDKFQFHIP ITMNFKAEGI FNMNQRVNQF LKANPDINII GIDRGERHLL YYALINQKGK ILKQDTLNVI ANEKQKVDYH NLLDKKEGDR ATARQEWGVI ETIKELKEGY LSQVIHKLTD LMIENNAIIV MEDLNFGFKR GRQKVEKQVY QKFEKMLIDK LNYLVDKNKK ANELGGLLNA FQLANKFESF QKMGKQNGFI FYVPAWNTSK TDPATGFIDF LKPRYENLNQ AKDFFEKFDS IRLNSKADYF EFAFDFKNFT EKADGGRTKW TVCTTNEDRY AWNRALNNNR GSQEKYDITA ELKSLFDGKV DYKSGKDLKQ QIASQESADF FKALMKNLSI TLSLRHNNGE KGDNEQDYIL SPVADSKGRF FDSRKADDDM PKNADANGAY HIALKGLWCL EQISKTDDLK KVKLAISNKE WLEFVQTLKG	(SEQ ID NO: 147)
WP_037385181 type V CRISPR-associated protein	MQTLFENFTN QYPVSKTLRF ELIPQGKTKD FIEQKGLLKK DEDRAEKYKK VKNIIDEYHK DFIEKSLNGL KLDGLEEYKT LYLKQEKDDK DKKAFDKEKE NLRKQIANAF RNNEKFKTLF AKELIKNDLM SFACEEDKKN VKEFEAFTTY FTGFHQNRAN MYVADEKRTA IASRLIHENL PKFIDNIKIF EKMKKEAPEL	(SEQ ID NO: 148)

Table 5 - Cas12a orthologs		
Cpf1 [Smithella sp. SCADC]	LSPFNQTLKD MKDVIKGTTL EEIFSLDYFN KTLTQSGIDI YNSVIGGRTP EEGKTKIKGL NEYINTDFNQ KQTDKKKRQP KFKQLYKQIL SDRQSLSFIA EAFKNDTEIL EAIEKFYVNE LLHFSNEGKS TNVLDAIKNA VSNLESFNLT KIYFRSGTSL TDVSRKVFGE WSIINRALDN YYATTYPIKP REKSEKYEER KEKWLKQDFN VSLIQTAIDE YDNETVKGKN SGKVIVDYFA KFCDDKETDL IQKVNEGYIA VKDLLNTPYP ENEKLGSNKD QVKQIKAFMD SIMDIMHFVR PLSLKDTDKE KDETFYSLFT PLYDHLTQTI ALYNKVRNYL TQKPYSTEKI KLNFENSTLL GGWDLNKETD NTAIILRKEN LYYLGIMDKR HNRIFRNVPK ADKKDSCYEK MVYKLLPGAN KMLPKVFFSQ SRIQEFTPSA KLLENYENET HKKGDNFNLN HCHQLIDFFK DSINKHEDWK NFDFRFSATS TYADLSGFYH EVEHQGYKIS FQSIADSFID DLVNEGKLYL FQIYNKDFSP FSKGKPNLHT LYWKMLFDEN NLKDVVYKLN GEAEVFYRKK SIAEKNTTIH KANESIINKN PDNPKATSTF NYDIVKDKRY TIDKFQFHVP ITMNFKAEGI FNMNQRVNQF LKANPDINII GIDRGERHLL YYTLINQKGK ILKQDTLNVI ANEKQKVDYH NLLDKKEGDR ATARQEWGVI ETIKELKEGY LSQVIHKLTD LMIENNAIIV MEDLNFGFKR GRQKVEKQVY QKFEKMLIDK LNYLVDKNKK ANELGGLLNA FQLANKFESF QKMGKQNGFI FYVPAWNTSK TDPATGFIDF LKPRYENLKQ AKDFFEKFDS IRLNSKADYF EFAFDFKNFT GKADGGRTKW TVCTTNEDRY AWNRALNNNR GSQEKYDITA ELKSLFDGKV DYKSGKDLKQ QIASQELADF FRTLMKYLSV TLSLRHNNGE KGETEQDYIL SPVADSMGKF FDSRKAGDDM PKNADANGAY HIALKGLWCL EQISKTDDLK KVKLAISNKE WLEFMQTLKG	
WP_039871282.1 type V CRISPR-associated protein Cpf1 [Prevotella bryantii B14]	MKFTDFTGLY SLSKTLRFEL KPIGKTLENI KKAGLLEQDQ HRADSYKKVK KIIDEYHKAF IEKSLSNFEL KYQSEDKLDS LEEYLMYYSM KRIEKTEKDK FAKIQDNLRK QIADHLKGDE SYKTIFSKDL IRKNLPDFVK SDEERTLIKE FKDFTTYFKG FYENRENMYS AEDKSTAISH RIIHENLPKF VDNINAFSKI ILIPELREKL NQIYQDFEEY LNVESIDEIF HLDYFSMVMT QKQIEVYNAI IGGKSTNDKK IQGLNEYINL YNQKHKDCKL PKLKLLFKQI LSDRIAISWL PDNFKDDQEA LDSIDTCYKN LLNDGNVLGE GNLKLLLENI DTYNLKGIFI RNDLQLTDIS QKMYASWNVI QDAVILDLKK QVSRKKKESA EDYNDRLKKL YTSQESFSIQ YLNDCLRAYG KTENIQDYFA KLGAVNNEHE QTINLFAQVR NAYTSVQAIL TTPYPENANL AQDKETVALI KNLLDSLKRL QRFIKPLLGK GDESDKDERF YGDFTPLWET LNQITPLYNM VRNYMTRKPY SQEKIKLNFE NSTLLGGWDL NKEHDNTAII LRKNGLYYLA IMKKSANKIF DKDKLDNSGD CYEKMVYKLL PGANKMLPKV FFSKSRIDEF KPSENIIENY KKGTHKKGAN FNLADCHNLI DFFKSSISKH EDWSKFNFHF SDTSSYEDLS DFYREVEQQG YSISFCDVSV EYINKMVEKG DLYLFQIYNK DFSEFSKGTP NMHTLYWNSL FSKENLNNII YKLNGQAEIF FRKKSLNYKR PTHPAHQAIK NKNKCNEKKE SIFDYDLVKD KRYTVDKFQF HVPITMNFKS TGNTNINQQV IDYLRTEDDT HIIGIDRGER HLLYLVVIDS HGKIVEQFTL NEIVNEYGGN IYRTNYHDLL DTREQNREKA RESWQTIENI KELKEGYISQ VIHKITDLMQ KYHAVVVLED LNMGFMRGRQ KVEKQVYQKF EEMLINKLNY LVNKKADQNS AGGLLHAYQL TSKFESFQKL GKQSGFLFYI PAWNTSKIDP VTGFVNLFDT RYESIDKAKA FFGKFDSIRY NADKDWFEFA FDYNNFTTKA EGTRTNWTIC TYGSRIRTFR NQAKNSQWDN EEIDLTKAYK AFFAKHGINI YDNIKEAIAM ETEKSFFEDL LHLLKLTLQM RNSITGTTTD YLISPVHDSK GNFYDSRICD NSLPANADAN GAYNIARKGL MLIQQIKDST SSNRFKFSPI TNKDWLIFAQ EKPYLND	(SEQ ID NO: 149)
EKE28449.1 hypothetical protein ACD_3C00058G0015 [uncultured bacterium (gcode 4)]	MFKGDAFTGL YEVQKTLRFE LVPIGLTQSY LENDWVIQKD KEVEENYGKI KAYFDLIHKE FVRQSLENAW LCQLDDFYEK YIELHNSLET RKDKNLAKQF EKVMKSLKKE FVSFFDAKWN EWKQKFSFLK KWWIDVLNEK EVLDLMAEFY PDEKELFDKF DKFFTYFSNF KESRKNFYAD DGRAWAIATR AIDENLITFI KNIEDFKKLN SSFREFVNDN FSEEDKQIFE IDFYNNCLLQ PWIDKYNKIV WWYSLENWEK VQWLNEKINN FKQNQNKSNS KDLKFPRMKL LYKQILGDKE KKVYIDEIRD DKNLIDLIDN SKRRNQIKID NANDIINDFI NNNAKFELDK IYLTRQSINT ISSKYFSSWD YIRWYFWTGE LQEFVSFYDL KETFWKIEYE TLENIFKDCY VKGINTESQN NIVFETQGIY ENFLNIFKFE FNQNISQISL LEWELDKIQN EDIKKNEKQV EVIKNYFDSV MSVYKMTKYF SLEKWKKRVE LDTDNNFYND FNEYLEGFEI WKDYNLVRNY ITKKQVNTDK IKLNFDNSQF LTWWDKDKEN ERLGIILRRE WKYYLWILKK WNTLNFGDYL QKEWEIFYEK MNYKQLNNVY RQLPRLLFPL TKKLNELKWD ELKKYLSKYI QNFWYNEEIA QIKIEFDIFQ ESKEKWEKFD IDKLRKLIEY YKKWVLALYS DLYDLEFIKY	(SEQ ID NO: 150)

Table 5 - Cas12a orthologs		
	KNYDDLSIFY SDVEKKMYNL NFTKIDKSLI DGKVKSWELY LFQIYNKDFS ESKKEWSTEN IHTKYFKLLF NEKNLQNLVV KLSWWADIFF RDKTENLKFK KDKNGQEILD HRRFSQDKIM FHISITLNAN CWDKYWFNQY VNEYMNKERD IKIIWIDRWE KHLAYYCVID KSWKIFNNEI WTLNELNWVN YLEKLEKIES SRKDSRISWW EIENIKELKN GYISQVINKL TELIVKYNAI IVFEDLNIWF KRWRQKIEKQ IYQKLELALA KKLNYLTQKD KKDDEILWNL KALQLVPKVN DYQDIWNYKQ SWIMFYVRAN YTSVTCPNCW LRKNLYISNS ATKENQKKSL NSIAIKYNDW KFSFSYEIDD KSWKQKQSLN KKKFIVYSDI ERFVYSPLEK LTKVIDVNKK LLELFRDFNL SLDINKQIQE KDLDSVFFKS LTHLFNLILQ LRNSDSKDNK DYISCPSCYY HSNNWLQWFE FNWDANWAYN IARKGIILLD RIRKNQEKPD LYVSDIDWDN FVQSNQFPNT IIPIQNIEKQ VPLNIKI	
WP_018359861.1 type V CRISPR-associated protein Cpf1 [Porphyromonas macacae]	MKTQHFFEDF TSLYSLSKTI RFELKPIGKT LENIKKNGLI RRDEQRLDDY EKLKKVIDEY HEDFIANILS SFSFSEEILQ SYIQNLSESE ARAKIEKTMR DTLAKAFSED ERYKSIFKKE LVKKDIPVWC PAYKSLCKKF DNFTTSLVPF HENRKNLYTS NEITASIPYR IVHVNLPKFI QNIEALCELQ KKMGADLYLE MMENLRNVWP SFVKTPDDLC NLKTYNHLMV QSSISEYNRF VGGYSTEDGT KHQGINEWIN IYRQRNKEMR LPGLVFLHKQ ILAKVDSSSF ISDTLENDDQ VFCVLRQFRK LFWNTVSSKE DDAASLKDLF CGLSGYDPEA IYVSDAHLAT ISKNIFDRWN YISDAIRRKT EVLMPRKKES VERYAEKISK QIKKRQSYSL AELDDLLAHY SEESLPAGFS LLSYFTSLGG QKYLVSDGEV ILYEEGSNIW DEVLIAFRDL QVILDKDFTE KKLGKDEEAV SVIKKALDSA LRLRKFFDLL SGTGAEIRRD SSFYALYTDR MDKLKGLLKM YDKVRNYLTK KPYSIEKFKL HFDNPSLLSG WDKNKELNNL SVIFRQNGYY YLGIMTPKGK NLFKTLPKLG AEEMFYEKME YKQIAEPMLM LPKVFFPKKT KPAFAPDQSV VDIYNKKTFK TGQKGFNKKD LYRLIDFYKE ALTVHEWKLF NFSFSPTEQY RNIGEFFDEV REQAYKVSMV NVPASYIDEA VENGKLYLFQ IYNKDFSPYS KGIPNLHTLY WKALFSEQNQ SRVYKLCGGG ELFYRKASLH MQDTTVHPKG ISIHKKNLNK KGETSLFNYD LVKDKRFTED KFFFHVPISI NYKNKKITNV NQMVRDYIAQ NDDLQIIGID RGERNLLYIS RIDTRGNLLE QFSLNVIESD KGDLRTDYQK ILGDREQERL RRRQEWKSIE SIKDLKDGYM SQVVHKICNM VVEHKAIVVL ENLNLSFMKG RKKVEKSVYE KFERMLVDKL NYLVVDKKNL SNEPGGLYAA YQLTNPLFSF EELHRYPQSG ILFFVDPWNT SLTDPSTGFV NLLGRINYTN VGDARKFFDR FNAIRYDGKG NILFDLDLSR FDVRVETQRK LWTLTTFGSR IAKSKKSGKW MVERIENLSL CFLELFEQFN IGYRVEKDLK KAILSQDRKE FYVRLIYLFN LMMQIRNSDG EEDYILSPAL NEKNLQFDSR LIEAKDLPVD ADANGAYNVA RKGLMVVQRI KRGDHESIHR IGRAQWLRYV QEGIVE	(SEQ ID NO: 151)
WP_013282991 type V CRISPR-associated protein Cpf1 [Butyrivibrio proteoclasticus]	MLLYENYTKR NQITKSLRLE LRPQGKTLRN IKELNLLEQD KAIYALLERL KPVIDEGIKD IARDTLKNCE LSFEKLYEHF LSGDKKAYAK ESERLKKEIV KTLIKNLPEG IGKISEINSA KYLNGVLYDF IDKTHKDSEE KQNILSDILE TKGYLALFSK FLTSRITTLE QSMPKRVIEN FEIYAANIPK MQDALERGAV SFAIEYESIC SVDYYNQILS QEDIDSYNRL ISGIMDEDGA KEKGINQTIS EKNIKIKSEH LEEKPFRILK QLHKQILEER EKAFTIDHID SDEEVVQVTK EAFEQTKEQW ENIKKINGFY AKDPGDITLF IVVGPNQTHV LSQLIYGEHD RIRLLLEEYE KNTLEVLPRR TKSEKARYDK FVNAVPKKVA KESHTFDGLQ KMTGDDRLFI LYRDELARNY MRIKEAYGTF ERDILKSRRG IKGNRDVQES LVSFYDELTK FRSALRIINS GNDEKADPIF YNTFDGIFEK ANRTYKAENL CRNYVTKSPA DDARIMASCL GTPARLRTHW WNGEENFAIN DVAMIRRGDE YYYFVLTPDV KPVDLKTKDE TDAQIFVQRK GAKSFLGLPK ALFKCILEPY FESPEHKNDK NCVIEEYVSK PLTIDRRAYD IFKNGTFKKT NIGIDGLTEE KFKDDCRYLI DVYKEFIAVY TRYSCFNMSG LKRADEYNDI GEFFSDVDTR LCTMEWIPVS FERINDMVDK KEGLLFLVRS MFLYNRPRKP YERTFIQLFS DSNMEHTSML LNSRAMIQYR AASLPRRVTH KKGSILVALR DSNGEHIPMH IREAIYKMKN NFDISSEDFI MAKAYLAEHD VAIKKANEDI IRNRRYTEDK FFLSLSYTKN ADISARTLDY INDKVEEDTQ DSRMAVIVTR NLKDLTYVAV VDEKNNVLEE KSLNEIDGVN YRELLKERTK IKYHDKTRLW QYDVSSKGLK EAYVELAVTQ ISKLATKYNA VVVVESMSST FKDKFSFLDE QIFKAFEARL CARMSDLSFN TIKEGEAGSI SNPIQVSNNN GNSYQDGVIY FLNNAYTRTL CPDTGFVDVF DKTRLITMQS KRQFFAKMKD IRIDDGEMLF TFNLEEYPTK RLLDRKEWTV KIAGDGSYFD KDKGEYVYVN DIVREQIIPA LLEDKAVFDG NMAEKFLDKT AISGKSVELI YKWFANALYG IITKKDGEKI YRSPITGTEI	(SEQ ID NO: 152)

Table 5 - Cas12a orthologs

	DVSKNTTYNF GKKFMFKQEY RGDGDFLDAF LNYMQAQDIA V	
WP_048112740.1 type V CRISPR-associated protein Cpf1 [Candidatus Methanoplasma termitum]	MNNYDEFTKL YPIQKTIRFE LKPQGRTMEH LETFNFFEED RDRAEKYKIL KEAIDEYHKK FIDEHLTNMS LDWNSLKQIS EKYYKSREEK DKKVFLSEQK RMRQEIVSEF KKDDRFKDLF SKKLFSELLK EEIYKKGNHQ EIDALKSFDK FSGYFIGLHE NRKNMYSDGD EITAISNRIV NENFPKFLDN LQKYQEARKK YPEWIIKAES ALVAHNIKMD EVFSLEYFNK VLNQEGIQRY NLALGGYVTK SGEKMMGLND ALNLAHQSEK SSKGRIHMTP LFKQILSEKE SFSYIPDVFT EDSQLLPSIG GFFAQIENDK DGNIFDRALE LISSYAEYDT ERIYIRQADI NRVSNVIFGE WGTLGGLMRE YKADSINDIN LERTCKKVDK WLDSKEFALS DVLEAIKRTG NNDAFNEYIS KMRTAREKID AARKEMKFIS EKISGDEESI HIIKTLLDSV QQFLHFFNLF KARQDIPLDG AFYAEFDEVH SKLFAIVPLY NKVRNYLTKN NLNTKKIKLN FKNPTLANGW DQNKVYDYAS LIFLRDGNYY LGIINPKRKK NIKFEQGSGN GPFYRKMVYK QIPGPNKNLP RVFLTSTKGK KEYKPSKEII EGYEADKHIR GDKFDLDFCH KLIDFFKESI EKHKDWSKFN FYFSPTESYG DISEFYLDVE KQGYRMHFEN ISAETIDEYV EKGDLFLFQI YNKDFVKAAT GKKDMHTIYW NAAFSPENLQ DVVVKLNGEA ELFYRDKSDI KEIVHREGEI LVNRTYNGRT PVPDKIHKKL TDYHNGRTKD LGEAKEYLDK VRYFKAHYDI TKDRRYLNDK IYFHVPLTLN FKANGKKNLN KMVIEKFLSD EKAHIIGIDR GERNLLYYSI IDRSGKIIDQ QSLNVIDGFD YREKLNQREI EMKDARQSWN AIGKIKDLKE GYLSKAVHEI TKMAIQYNAI VVMEELNYGF KRGRFKVEKQ IYQKFENMLI DKMNYLVFKD APDESPGGVL NAYQLTNPLE SFAKLGKQTG ILFYVPAAYT SKIDPTTGFV NLFNTSSKTN AQERKEFLQK FESISYSAKD GGIFAFAFDY RKFGTSKTDH KNVWTAYTNG ERMRYIKEKK RNELFDPSKE IKEALTSSGI KYDGGQNILP DILRSNNNGL IYTMYSSFIA AIQMRVYDGK EDYIISPIKN SKGEFFRTDP KRRELPIDAD ANGAYNIALR GELTMRAIAE KFDPDSEKMA KLELKHKDWF EFMQTRGD	(SEQ ID NO: 153)
WP_027407524.1 type V CRISPR-associated protein Cpf1 [Anaerovibrio sp. RM50]	MVAFIDEFVG QYPVSKTLRF EARPVPETKK WLESDQCSVL FNDQKRNEYY GVLKELLDDY YRAYIEDALT SFTLDKALLE NAYDLYCNRD TNAFSSCCEK LRKDLVKAFG NLKDYLLGSD QLKDLVKLKA KVDAPAGKGK KKIEVDSRLI NWLNNNAKYS AEDREKYIKA IESFEGFVTY LTNYKQAREN MFSSEDKSTA IAFRVIDQNM VTYFGNIRIY EKIKAKYPEL YSALKGFEKF FSPTAYSEIL SQSKIDEYNY QCIGRPIDDA DFKGVNSLIN EYRQKNGIKA RELPVMSMLY KQILSDRDNS FMSEVINRNE EAIECAKNGY KVSYALFNEL LQLYKKIFTE DNYGNIYVKT QPLTELSQAL FGDWSILRNA LDNGKYDKDI INLAELEKYF SEYCKVLDAD DAAKIQDKFN LKDYFIQKNA LDATLPDLDK ITQYKPHLDA MLQAIRKYKL FSMYNGRKKM DVPENGIDFS NEFNAIYDKL SEFSILYDRI RNFATKKPYS DEKMKLSFNM PTMLAGWDYN NETANGCFLF IKDGKYFLGV ADSKSKNIFD FKKNPHLLDK YSSKDIYYKV KYKQVSGSAK MLPKVVFAGS NEKIFGHLIS KRILEIREKK LYTAAAGDRK AVAEWIDFMK SAIAIHPEWN EYFKFKFKNT AEYDNANKFY EDIDKQTYSL EKVEIPTEYI DEMVSQHKLY LFQLYTKDFS DKKKKKGTDN LHTMYWHGVF SDENLKAVTE GTQPIIKLNG EAEMFMRNPS IEFQVTHEHN KPIANKNPLN TKKESVFNYD LIKDKRYTER KFYFHCPITL NFRADKPIKY NEKINRFVEN NPDVCIIGID RGERHLLYYT VINQTGDILE QGSLNKISGS YTNDKGEKVN KETDYHDLLD RKEKGKHVAQ QAWETIENIK ELKAGYLSQV VYKLTQLMLQ YNAVIVLENL NVGFKRGRTK VEKQVYQKFE KAMIDKLNYL VFKDRGYEMN GSYAKGLQLT DKFESFDKIG KQTGCIYYVI PSYTSHIDPK TGFVNLLNAK LRYENITKAQ DTIRKFDSIS YNAKADYFEF AFDYRSFGVD MARNEWVVCT CGDLRWEYSA KTRETKAYSV TDRLKELFKA HGIDYVGGEN LVSHITEVAD KHFLSTLLFY LRLVLKMRYT VSGTENENDF ILSPVEYAPG KFFDSREATS TEPMNADANG AYHIALKGLM TIRGIEDGKL HNYGKGGENA AWFKFMQNQE YKNNG	(SEQ ID NO: 154)
WP_044910712.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium MC2017]	MDYGNGQFER RAPLTKTITL RLKPIGETRE TIREQKLLEQ DAAFRKLVET VTPIVDDCIR KIADNALCHF GTEYDFSCLG NAISKNDSKA IKKETEKVEK LLAKVLTENL PDGLRKVNDI NSAAFIQDTL TSFVQDDADK RVLIQELKGK TVLMQRFLTT RITALTVWLP DRVFENFNIF IENAEKMRIL LDSPLNEKIM KFDPDAEQYA SLEFYGQCLS QKDIDSYNLI ISGIYADDEV KNPGINEIVK EYNQQIRGDK DESPLPKLKK LHKQILMPVE KAFFVRVLSN DSDARSILEK ILKDTEMLPS KIIEAMKEAD AGDIAVYGSR LHELSHVIYG DHGKLSQIIY DKESKRISEL METLSPKERK ESKKRLEGLE EHIRKSTYTF DELNRYAEKN VMAAYIAAVE ESCAEIMRKE KDLRTLLSKE DVKIRGNRHN TLIVKNYFNA	(SEQ ID NO: 155)

Table 5 - Cas12a orthologs		
	WTVFRNLIRI LRRKSEAEID SDFYDVLDDS VEVLSLTYKG ENLCRSYITK KIGSDLKPEI ATYGSALRPN SRWWSPGEKF NVKFHTIVRR DGRLYYFILP KGAKPVELED MDGDIECLQM RKIPNPTIFL PKLVFKDPEA FFRDNPEADE FVFLSGMKAP VTITRETYEA YRYKLYTVGK LRDGEVSEEE YKRALLQVLT AYKEFLENRM IYADLNFGFK DLEEYKDSSE FIKQVETHNT FMCWAKVSSS QLDDLVKSGN GLLFEIWSER LESYYKYGNE KVLRGYEGVL LSILKDENLV SMRTLLNSRP MLVYRPKESS KPMVVHRDGS RVVDRFDKDG KYIPPEVHDE LYRFFNNLLI KEKLGEKARK ILDNKKVKVK VLESERVKWS KFYDEQFAVT FSVKKNADCL DTTKDLNAEV MEQYSESNRL ILIRNTTDIL YYLVLDKNGK VLKQRSLNII NDGARDVDWK ERFRQVTKDR NEGYNEWDYS RTSNDLKEVY LNYALKEIAE AVIEYNAILI IEKMSNAFKD KYSFLDDVTF KGFETKLLAK LSDLHFRGIK DGEPCSFTNP LQLCQNDSNK ILQDGVIFMV PNSMTRSLDP DTGFIFAIND HNIRTKKAKL NFLSKFDQLK VSSEGCLIMK YSGDSLPTHN TDNRVWNCCC NHPITNYDRE TKKVEFIEEP VEELSRVLEE NGIETDTELN KLNERENVPG KVVDAIYSLV LNYLRGTVSG VAGQRAVYYS PVTGKKYDIS FIQAMNLNRK CDYYRIGSKE RGEWTDFVAQ LIN	
WP_081834226 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium MC2017].	MTMDYGNGQF ERRAPLTKTI TLRLKPIGET RETIREQKLL EQDAAFRKLV ETVTPIVDDC IRKIADNALC HFGTEYDFSC LGNAISKNDS KAIKKETEKV EKLLAKVLTE NLPDGLRKVN DINSAAFIQD TLTSFVQDDA DKRVLIQELK GKTVLMQRFL TTRITALTVW LPDRVFENFN IFIENAEKMR ILLDSPLNEK IMKFDPDAEQ YASLEFYGQC LSQKDIDSYN LIISGIYADD EVKNPGINEI VKEYNQQIRG DKDESPLPKL KKLHKQILMP VEKAFFVRVL SNDSDARSIL EKILKDTEML PSKIIEAMKE ADAGDIAVYG SRLHELSHVI YGDHGKLSQI IYDKESKRIS ELMETLSPKE RKESKKRLEG LEEHIRKSTY TFDELNRYAE KNVMAAYIAA VEESCAEIMR KEKDLRTLLS KEDVKIRGNR HNTLIVKNYF NAWTVFRNLI RILRRKSEAE IDSDFYDVLD DSVEVLSLTY KGENLCRSYI TKKIGSDLKP EIATYGSALR PNSRWWSPGE KFNVKFHTIV RRDGRLYYFI LPKGAKPVEL EDMDGDIECL QMRKIPNPTI FLPKLVFKDP EAFFRDNPEA DEFVFLSGMK APVTITRETY EAYRYKLYTV GKLRDGEVSE EEYKRALLQV LTAYKEFLEN RMIYADLNFG FKDLEEYKDS SEFIKQVETH NTFMCWAKVS SSQLDDLVKS GNGLLFEIWS ERLESYYKYG NEKVLRGYEG VLLSILKDEN LVSMRTLLNS RPMLVYRPKE SSKPMVVHRD GSRVVDRFDK DGKYIPPEVH DELYRFFNNL LIKEKLGEKA RKILDNKKVK VKVLESERVK WSKFYDEQFA VTFSVKKNAD CLDTTKDLNA EVMEQYSESN RLILIRNTTD ILYYLVLDKN GKVLKQRSLN IINDGARDVD WKERFRQVTK DRNEGYNEWD YSRTSNDLKE VYLNYALKEI AEAVIEYNAI LIIEKMSNAF KDKYSFLDDV TFKGFETKLL AKLSDLHFRG IKDGEPCSFT NPLQLCQNDS NKILQDGVIF MVPNSMTRSL DPDTGFIFAI NDHNIRTKKA KLNFLSKFDQ LKVSSEGCLI MKYSGDSLPT HNTDNRVWNC CCNHPITNYD RETKKVEFIE EPVEELSRVL EENGIETDTE LNKLNERENV PGKVVDAIYS LVLNYLRGTV SGVAGQRAVY YSPVTGKKYD ISFIQAMNLN RKCDYYRIGS KERGEWTDFV AQLIN	(SEQ ID NO: 156)
WP_027216152.1 type V CRISPR-associated protein Cpf1 [Butyrivibrio fibrisolvens]	MYYESLTKLY PIKKTIRNEL VPIGKTLENI KKNNILEADE DRKIAYIRVK AIMDDYHKRL INEALSGFAL IDLDKAANLY LSRSKSADDI ESFSRFQDKL RKAIAKRLRE HENFGKIGNK DIIPLLQKLS ENEDDYNALE SFKNFYTYFE SYNDVRLNLY SDKEKSSTVA YRLINENLPR FLDNIRAYDA VQKAGITSEE LSSEAQDGLF LVNTFNNVLI QDGINTYNED IGKLNVAINL YNQKNASVQG FRKVPKMKVL YKQILSDREE SFIDEFESDT ELLDSLESHY ANLAKYFGSN KVQLLFTALR ESKGVNVYVK NDIAKTSFSN VVFGSWSRID ELINGEYDDN NNRKKDEKYY DKRQKELKKN KSYTIEKIIT LSTEDVDVIG KYIEKLESDI DDIRFKGKNF YEAVLCGHDR SKKLSKNKGA VEAIKGYLDS VKDFERDLKL INGSGQELEK NLVVYGEQEA VLSELSGIDS LYNMTRNYLT KKPFSTEKIK LNFNKPTFLD GWDYGNEEAY LGFFMIKEGN YFLAVMDANW NKEFRNIPSV DKSDCYKKVI YKQISSPEKS IQNLMVIDGK TVKKNGRKEK EGIHSGENLI LEELKNTYLP KKINDIRKRR SYLNGDTFSK KDLTEFIGYY KQRVIEYYNG YSFYFKSDDD YASFKEFQED VGRQAYQISY VDVPVSFVDD LINSGKLYLF RVYNKDFSEY SKGRLNLHTL YFKMLFDERN LKNVVYKLNG QAEVFYRPSS IKKEELIVHR AGEEIKNKNP KRAAQKPTRR LDYDIVKDRR YSQDKFMLHT SIIMNFGAEE NVSFNDIVNG VLRNEDKVNV IGIDRGERNL LYVVVIDPEG KILEQRSLNC ITDSNLDIET DYHRLLDEKE SDRKIARRDW TTIENIKELK AGYLSQVVHI VAELVLKYNA IICLEDLNFG FKRGRQKVEK QVYQKFEKML	(SEQ ID NO: 157)

Table 5 - Cas12a orthologs		
	IDKLNYLVMD KSREQLSPEK ISGALNALQL TPDFKSFKVL GKQTGIIYYV PAYLTSKIDP MTGFANLFYV KYENVDKAKE FFSKFDSIKY NKDGKNWNTK GYFEFAFDYK KFTDRAYGRV SEWTVCTVGE RIIKFKNKEK NNSYDDKVID LTNSLKELFD SYKVTYESEV DLKDAILAID DPAFYRDLTR RLQQTLQMRN SSCDGSRDYI ISPVKNSKGE FFCSDNNDDT TPNDADANGA FNIARKGLWV LNEIRNSEEG SKINLAMSNA QWLEYAQDNT I	
WP_016301126.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium COE1]	MHENNGKIAD NFIGIYPVSK TLRFELKPVG KTQEYIEKHG ILDEDLKRAG DYKSVKKIID AYHKYFIDEA LNGIQLDGLK NYYELYEKKR DNNEEKEFQK IQMSLRKQIV KRFSEHPQYK YLFKKELIKN VLPEFTKDNA EEQTLVKSFQ EFTTYFEGFH QNRKNMYSDE EKSTAIAYRV VHQNLPKYID NMRIFSMILN TDIRSDLTEL FNNLKTKMDI TIVEEYFAID GFNKVVNQKG IDVYNTILGA FSTDDNTKIK GLNEYINLYN QKNKAKLPKL KPLFKQILSD RDKISFIPEQ FDSDTEVLEA VDMFYNRLLQ FVIENEGQIT ISKLLTNFSA YDLNKIYVKN DTTISAISND LFDDWSYISK AVRENYDSEN VDKNKRAAAY EEKKEKALSK IKMYSIEELN FFVKKYSCNE CHIEGYFERR ILEILDKMRY AYESCKILHD KGLINNISLC QDRQAISELK DFLDSIKEVQ WLLKPLMIGQ EQADKEEAFY TELLRIWEEL EPITLLYNKV RNYVTKKPYT LEKVKLNFYK STLLDGWDKN KEKDNLGIIL LKDGQYYLGI MNRRNNKIAD DAPLAKTDNV YRKMEYKLLT KVSANLPRIF LKDKYNPSEE MLEKYEKGTH LKGENFCIDD CRELIDFFKK GIKQYEDWGQ FDFKFSDTES YDDISAFYKE VEHQGYKITF RDIDETYIDS LVNEGKLYLF QIYNKDFSPY SKGTKNLHTL YWEMLFSQQN LQNIVYKLNG NAEIFYRKAS INQKDVVVHK ADLPIKNKDP QNSKKESMFD YDIIKDKRFT CDKYQFHVPI TMNFKALGEN HFNRKVNRLI HDAENMHIIG IDRGERNLIY LCMIDMKGNI VKQISLNEII SYDKNKLEHK RNYHQLLKTR EDENKSARQS WQTIHTIKEL KEGYLSQVIH VITDLMVEYN AIVVLEDLNF GFKQGRQKFE RQVYQKFEKM LIDKLNYLVD KSKGMDEDGG LLHAYQLTDE FKSFKQLGKQ SGFLYYIPAW NTSKLDPTTG FVNLFYTKYE SVEKSKEFIN NFTSILYNQE REYFEFLFDY SAFTSKAEGS RLKWTVCSKG ERVETYRNPK KNNEWDTQKI DLTFELKKLF NDYSISLLDG DLREQMGKID KADFYKKFMK LFALIVQMRN SDEREDKLIS PVLNKYGAFF ETGKNERMPL DADANGAYNI ARKGLWIIEK IKNTDVEQLD KVKLTISNKE WLQYAQEHIL	(SEQ ID NO: 158)
WP_035635841.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium ND2006]	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK IAISNKEWLE YAQTSVKH	(SEQ ID NO: 159)
WP_051666128.1 type V CRISPR-associated protein	MLKNVGIDRL DVEKGRKNMS KLEKFTNCYS LSKTLRFKAI PVGKTQENID NKRLLVEDEK RAEDYKGVKK LLDRYYLSFI NDVLHSIKLK NLNNYISLFR KKTRTEKENK ELENLEINLR KEIAKAFKGN EGYKSLFKKD IIETILPEFL DDKDEIALVN SFNGFTTAFT GFFDNRENMF SEEAKSTSIA FRCINENLTR	(SEQ ID NO: 160)

Table 5 - Cas12a orthologs		
Cpf1 [Lachnospiraceae bacterium ND2006]	YISNMDIFEK VDAIFDKHEV QEIKEKILNS DYDVEDFFEG EFFNFVLTQE GIDVYNAIIG GFVTESGEKI KGLNEYINLY NQKTKQKLPK FKPLYKQVLS DRESLSFYGE GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE KLFKNFDEYS SAGIFVKNGP AISTISKDIF GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE DDRRKSFKKI GSFSLEQLQE YADADLSVVE KLKEIIIQKV DEIYKVYGSS EKLFDADFVL EKSLKKNDAV VAIMKDLLDS VKSFENYIKA FFGEGKETNR DESFYGDFVL AYDILLKVDH IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG GWDKDKETDY RATILRYGSK YYLAIMDKKY AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA YDFNFSETEK YKDIAGFYRE VEEQGYKVSF ESASKKEVDK LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM YFKLLFDENN HGQIRLSGGA ELFMRRASLK KEELVVHPAN SPIANKNPDN PKKTTTLSYD VYKDKRFSED QYELHIPIAI NKCPKNIFKI NTEVRVLLKH DDNPYVIGID RGERNLLYIV VVDGKGNIVE QYSLNEIINN FNGIRIKTDY HSLLDKKEKE RFEARQNWTS IENIKELKAG YISQVVHKIC ELVEKYDAVI ALEDLNSGFK NSRVKVEKQV YQKFEKMLID KLNYMVDKKS NPCATGGALK GYQITNKFES FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA DSKKFISSFD RIMYVPEEDL FEFALDYKNF SRTDADYIKK WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT SAYKELFNKY GINYQQGDIR ALLCEQSDKA FYSSFMALMS LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY DSRNYEAQEN AILPKNADAN GAYNIARKVL WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA QTSVKH	
WP_015504779.1 type V CRISPR-associated protein Cpf1 [Candidatus Methanomethylophilus alvus]	MDAKEFTGQY PLSKTLRFEL RPIGRTWDNL EASGYLAEDR HRAECYPRAK ELLDDNHRAF LNRVLPQIDM DWHPIAEAFC KVHKNPGNKE LAQDYNLQLS KRRKEISAYL QDADGYKGLF AKPALDEAMK IAKENGNESD IEVLEAFNGF SVYFTGYHES RENIYSDEDM VSVAYRITED NFPRFVSNAL IFDKLNESHP DIISEVSGNL GVDDIGKYFD VSNYNNFLSQ AGIDDYNHII GGHTTEDGLI QAFNVVLNLR HQKDPGFEKI QFKQLYKQIL SVRTSKSYIP KQFDNSKEMV DCICDYVSKI EKSETVERAL KLVRNISSFD LRGIFVNKKN LRILSNKLIG DWDAIETALM HSSSSENDKK SVYDSAEAFT LDDIFSSVKK FSDASAEDIG NRAEDICRVI SETAPFINDL RAVDLDSLND DGYEAAVSKI RESLEPYMDL FHELEIFSVG DEFPKCAAFY SELEEVSEQL IEIIPLFNKA RSFCTRKRYS TDKIKVNLKF PTLADGWDLN KERDNKAAIL RKDGKYYLAI LDMKKDLSSI RTSDEDESSF EKMEYKLLPS PVKMLPKIFV KSKAAKEKYG LTDRMLECYD KGMHKSGSAF DLGFCHELID YYKRCIAEYP GWDVFDFKFR ETSDYGSMKE FNEDVAGAGY YMSLRKIPCS EVYRLLDEKS IYLFQIYNKD YSENAHGNKN MHTMYWEGLF SPQNLESPVF KLSGGAELFF RKSSIPNDAK TVHPKGSVLV PRNDVNGRRI PDSIYRELTR YFNRGDCRIS DEAKSYLDKV KTKKADHDIV KDRRFTVDKM MFHVPIAMNF KAISKPNLNK KVIDGIIDDQ DLKIIGIDRG ERNLIYVTMV DRKGNILYQD SLNILNGYDY RKALDVREYD NKEARRNWTK VEGIRKMKEG YLSLAVSKLA DMIIENNAII VMEDLNHGFK AGRSKIEKQV YQKFESMLIN KLGYMVLKDK SIDQSGGALH GYQLANHVTT LASVGKQCGV IFYIPAAFTS KIDPTTGFAD LFALSNVKNV ASMREFFSKM KSVIYDKAEG KFAFTFDYLD YNVKSECGRT LWTVYTVGER FTYSRVNREY VRKVPTDIIY DALQKAGISV EGDLRDRIAE SDGDTLKSIF YAFKYALDMR VENREEDYIQ SPVKNASGEF FCSKNAGKSL PQDSDANGAY NIALKGILQL RMLSEQYDPN AESIRLPLIT NKAWLTFMQS GMKTWKN	(SEQ ID NO: 161)
WP_044910713.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium MC2017]	MGLYDGFVNR YSVSKTLRFE LIPQGRTREY IETNGILSDD EERAKDYKTI KRLIDEYHKD YISRCLKNVN ISCLEEYYHL YNSSNRDKRH EELDALSDQM RGEIASFLTG NDEYKEQKSR DIIINERIIN FASTDEELAA VKRFRKFTSY FTGFFTNREN MYSAEKKSTA IAHRIIDVNL PKYVDNIKAF NTAIEAGVFD IAEFESNFKA ITDEHEVSDL LDITKYSRFI RNEDIIIYNT LLGGISMKDE KIQGLNELIN LHNQKHPGKK VPLLKVLYKQ ILGDSQTHSF VDDQFEDDQQ VINAVKAVTD TFSETLLGSL KIIINNIGHY DLDRIYIKAG QDITTLSKRA LNDWHIITEC LESEYDDKFP KNKKSDTYEE MRNRYVKSFK SFSIGRLNSL VTTYTEQACF LENYLGSFGG DTDKNCLTDF TNSLMEVEHL LNSEYPVTNR LITDYESVRI LKRLLDSEME VIHFLKPLLG NGNESDKDLV FYGEFEAEYE KLLPVIKVYN RVRNYLTRKP FSTEKIKLNF NSPTLLCGWS QSKEKEYMGV ILRKDGQYYL GIMTPSNKKI FSEAPKPDED CYEKMVLRYI PHPYQMLPKV FFSKSNIAFF NPSDEILRIK KQESFKKGKS FNRDDCHKFI DFYKDSINRH EEWRKFNFKF SDTDSYEDIS RFYKEVENQA FSMSFTKIPT VYIDSLVDEG	(SEQ ID NO: 162)

Table 5 - Cas12a orthologs		
	KLYLFKLHNK DFSEHSKGKP NLHTVYWNAL FSEYNLQNTV YQLNGSAEIF FRKASIPENE RVIHKKNVPI TRKVAELNGK KEVSVFPYDI IKNRRYTVDK FQFHVPLKMN FKADEKKRIN DDVIEAIRSN KGIHVIGIDR GERNLLYLSL INEEGRIIEQ RSLNIIDSGE GHTQNYRDLL DSREKDREKA RENWQEIQEI KDLKTGYLSQ AIHTITKWMK EYNAIIVLED LNDRFTNGRK KVEKQVYQKF EKMLIDKLNY YVDKDEEFDR MGGTHRALQL TEKFESFQKL GRQTGFIFYV PAWNTSKLDP TTGFVDLLYP 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_37_6]	MKNVFGGFTN LYSLTKTLRF ELKPTSKTQK LMKRNNVIQT DEEIDKLYHD EMKPILDEIH RRFINDALAQ KIFISASLDN FLKVVKNYKV ESAKKNIKQN QVKLLQKEIT IKTLGLRREV VSGFITVSKK WKDKYVGLGI KLKGDGYKVL TEQAVLDILK IEFPNKAKYI DKFRGFWTYF SGFNENRKNY YSEEDKATSI ANRIVNENLS RYIDNIIAFE EILQKIPNLK KFKQDLDITS YNYYLNQAGI DKYNKIIGGY IVDKDKKIQG INEKVNLYTQ QTKKKLPKLK FLFKQIGSER KGFGIFEIKE GKEWEQLGDL FKLQRTKINS NGREKGLFDS LRTMYREFFD EIKRDSNSQA RYSLDKIYFN KASVNTISNS WFTNWNKFAE LLNIKEDKKN GEKKIPEQIS IEDIKDSLSI IPKENLEELF KLTNREKHDR TRFFGSNAWV TFLNIWQNEI EESFNKLEEK EKDFKKNAAI KFQKNNLVQK NYIKEVCDRM LAIERMAKYH LPKDSNLSRE EDFYWIIDNL SEQREIYKYY NAFRNYISKK PYNKSKMKLN FENGNLLGGW SDGQERNKAG VILRNGNKYY LGVLINRGIF RTDKINNEIY RTGSSKWERL ILSNLKFQTL AGKGFLGKHG VSYGNMNPEK SVPSLQKFIR ENYLKKYPQL TEVSNTKFLS KKDFDAAIKE ALKECFTMNF INIAENKLLE AEDKGDLYLF EITNKDFSGK KSGKDNIHTI YWKYLFSESN CKSPIIGLNG GAEIFFREGQ KDKLHTKLDK KGKKVFDAKR YSEDKLFFHV SITINYGKPK NIKFRDIINQ LITSMNVNII GIDRGEKHLL YYSVIDSNGI ILKQGSLNKI RVGDKEVDFN KKLTERANEM KKARQSWEQI GNIKNFKEGY LSQAIHEIYQ LMIKYNAIIV LEDLNTEFKA KRLSKVEKSV YKKFELKLAR KLNHLILKDR NTNEIGGVLK AYQLTPTIGG GDVSKFEKAK QWGMMFYVRA NYTSTTDPVT GWRKHLYISN FSNNSVIKSF FDPTNRDTGI EIFYSGKYRS WGFRYVQKET GKKWELFATK ELERFKYNQT TKLCEKINLY DKFEELFKGI DKSADIYSQL CNVLDFRWKS LVYLWNLLNQ IRNVDKNAEG NKNDFIQSPV YPFFDSRKTD GKTEPINGDA NGALNIARKG LMLVERIKNN PEKYEQLIRD TEWDAWIQNF NKVN	(SEQ ID NO: 163)
WP_044919442.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium MA2020]	MYYESLTKQY PVSKTIRNEL IPIGKTLDNI RQNNILESDV KRKQNYEHVK GILDEYHKQL INEALDNCTL PSLKIAAEIY LKNQKEVSDR EDFNKTQDLL RKEVVEKLKA HENFTKIGKK DILDLLEKLP SISEDDYNAL ESFRNFYTYF TSYNKVRENL YSDKEKSSTV AYRLINENFP KFLDNVKSYR FVKTAGILAD GLGEEEQDSL FIVETFNKTL TQDGIDTYNS QVGKINSSIN LYNQKNQKAN GFRKIPKMKM LYKQILSDRE ESFIDEFQSD EVLIDNVESY GSVLIESLKS SKVSAFFDAL RESKGKNVYV KNDLAKTAMS NIVFENWRTF DDLLNQEYDL ANENKKKDDK YFEKRQKELK KNKSYSLEHL CNLSEDSCNL IENYIHQISD DIENIIINNE TFLRIVINEH DRSRKLAKNR KAVKAIKDFL DSIKVLEREL KLINSSGQEL EKDLIVYSAH EELLVELKQV DSLYNMTRNY LTKKPFSTEK VKLNFNRSTL LNGWDRNKET DNLGVLLLKD GKYYLGIMNT SANKAFVNPP VAKTEKVFKK VDYKLLPVPN QMLPKVFFAK SNIDFYNPSS EIYSNYKKGT HKKGNMFSLE DCHNLIDFFK ESISKHEDWS KFGFKFSDTA SYNDISEFYR EVEKQGYKLT YTDIDETYIN DLIERNELYL FQIYNKDFSM YSKGKLNLHT LYFMMLFDQR NIDDVVYKLN GEAEVFYRPA SISEDELIIH KAGEEIKNKN PNRARTKETS TFSYDIVKDK RYSKDKFTLH IPITMNFGVD EVKRFNDAVN SAIRIDENVN VIGIDRGERN LLYVVVIDSK GNILEQISLN SIINKEYDIE TDYHALLDER EGGRDKARKD WNTVENIRDL KAGYLSQVVN VVAKLVLKYN AIICLEDLNF GFKRGRQKVE KQVYQKFEKM LIDKLNYLVI DKSREQTSPK ELGGALNALQ LTSKFKSFKE LGKQSGVIYY VPAYLTSKID PTTGFANLFY MKCENVEKSK RFFDGFDFIR FNALENVFEF GFDYRSFTQR ACGINSKWTV CTNGERIIKY RNPDKNNMFD EKVVVVTDEM KNLFEQYKIP YEDGRNVKDM IISNEEAEFY RRLYRLLQQT LQMRNSTSDG TRDYIISPVK NKREAYFNSE LSDGSVPKDA DANGAYNIAR KGLWVLEQIR QKSEGEKINL AMTNAEWLEY	(SEQ ID NO: 164)

Table 5 - Cas12a orthologs

	AQTHLL	
WP_035798880.1 type V CRISPR-associated protein Cpf1 [Butyrivibrio sp. NC3005]	MYYQNLTKKY PVSKTIRNEL IPIGKTLENI RKNNILESDV KRKQDYEHVK GIMDEYHKQL INEALDNYML PSLNQAAEIY LKKHVDVEDR EEFKKTQDLL RREVTGRLKE HENYTKIGKK DILDLLEKLP SISEEDYNAL ESFRNFYTYF TSYNKVRENL YSDEEKSSTV AYRLINENLP KFLDNIKSYA FVKAAGVLAD CIEEEEQDAL FMVETFNMTL TQEGIDMYNY QIGKVNSAIN LYNQKNHKVE EFKKIPKMKV LYKQILSDRE EVFIGEFKDD ETLLSSIGAY GNVLMTYLKS EKINIFFDAL RESEGKNVYV KNDLSKTTMS NIVFGSWSAF DELLNQEYDL ANENKKKDDK YFEKRQKELK KNKSYTLEQM SNLSKEDISP IENYIERISE DIEKICIYNG EFEKIVVNEH DSSRKLSKNI KAVKVIKDYL DSIKELEHDI KLINGSGQEL EKNLVVYVGQ EEALEQLRPV DSLYNLTRNY LTKKPFSTEK VKLNFNKSTL LNGWDKNKET DNLGILFFKD GKYYLGIMNT TANKAFVNPP AAKTENVFKK VDYKLLPGSN KMLPKVFFAK SNIGYYNPST ELYSNYKKGT HKKGPSFSID DCHNLIDFFK ESIKKHEDWS KFGFEFSDTA DYRDISEFYR EVEKQGYKLT FTDIDESYIN DLIEKNELYL FQIYNKDFSE YSKGKLNLHT LYFMMLFDQR NLDNVVYKLN GEAEVFYRPA SIAENELVIH KAGEGIKNKN PNRAKVKETS TFSYDIVKDK RYSKYKFTLH IPITMNFGVD EVRRFNDVIN NALRTDDNVN VIGIDRGERN LLYVVVINSE GKILEQISLN SIINKEYDIE TNYHALLDER EDDRNKARKD WNTIENIKEL KTGYLSQVVN VVAKLVLKYN AIICLEDLNF GFKRGRQKVE KQVYQKFEKM LIEKLNYLVI DKSREQVSPE KMGGALNALQ LTSKFKSFAE LGKQSGIIYY VPAYLTSKID PTTGFVNLFY IKYENIEKAK QFFDGFDFIR FNKKDDMFEF SFDYKSFTQK ACGIRSKWIV YTNGERIIKY PNPEKNNLFD EKVINVTDEI KGLFKQYRIP YENGEDIKEI IISKAEADFY KRLFRLLHQT LQMRNSTSDG TRDYIISPVK NDRGEFFCSE FSEGTMPKDA DANGAYNIAR KGLWVLEQIR QKDEGEKVNL SMTNAEWLKY AQLHLL	(SEQ ID NO: 165)
WP_027109509.1 type V CRISPR-associated protein Cpf1 [Lachnospiraceae bacterium NC2008]	MENYYDSLTR QYPVTKTIRQ ELKPVGKTLE NIKNAEIIEA DKQKKEAYVK VKELMDEFHK SIIEKSLVGI KLDGLSEFEK LYKIKTKTDE DKNRISELFY YMRKQIADAL KNSRDYGYVD NKDLIEKILP ERVKDENSLN ALSCFKGFTT YFTDYYKNRK NIYSDEEKHS TVGYRCINEN LLIFMSNIEV YQIYKKANIK NDNYDEETLD KTFMIESFNE CLTQSGVEAY NSVVASIKTA TNLYIQKNNK EENFVRVPKM KVLFKQILSD RTSLFDGLII ESDDELLDKL CSFSAEVDKF LPINIDRYIK TLMDSNNGTG IYVKNDSSLT TLSNYLTDSW SSIRNAFNEN YDAKYTGKVN DKYEEKREKA YKSNDSFELN YIQNLLGINV IDKYIERINF DIKEICEAYK EMTKNCFEDH DKTKKLQKNI KAVASIKSYL DSLKNIERDI KLLNGTGLES RNEFFYGEQS TVLEEITKVD ELYNITRNYL TKKPFSTEKM KLNFNNPQLL GGWDVNKERD CYGVILIKDN NYYLGIMDKS ANKSFLNIKE SKNENAYKKV NCKLLPGPNK MFPKVFFAKS NIDYYDPTHE IKKLYDKGTF KKGNSFNLED CHKLIDFYKE SIKKNDDWKN FNFNFSDTKD YEDISGFFRE VEAQNYKITY TNVSCDFIES LVDEGKLYLF QIYNKDFSEY ATGNLNLHTL YLKMLFDERN LKDLCIKMNG EAEVFYRPAS ILDEDKVVHK ANQKITNKNT NSKKKESIFS YDIVKDKRYT VDKFFIHLPI TLNYKEQNVS RFNDYIREIL KKSKNIRVIG IDRGERNLLY VVVCDSDGSI LYQRSINEIV SGSHKTDYHK LLDNKEKERL SSRRDWKTIE NIKDLKAGYM SQVVNEIYNL ILKYNAIVVL EDLNIGFKNG RKKVEKQVYQ NFEKALIDKL NYLCIDKTRE QLSPSSPGGV LNAYQLTAKF ESFEKIGKQT GCIFYVPAYL TSQIDPTTGF VNLFYQKDTS KQGLQLFFRK FKKINFDKVA SNFEFVFDYN DFTNKAEGTK TNWTISTQGT RIAKYRSDDA NGKWISRTVH PTDIIKEALN REKINYNDGH DLIDEIVSIE KSAVLKEIYY GFKLTLQLRN STLANEEEQE DYIISPVKNS SGNYFDSRIT SKELPCDADA NGAYNIARKG LWALEQIRNS ENVSKVKLAI SNKEWFEYTQ NNIPSL	(SEQ ID NO: 166)
WP_049895985.1 type V CRISPR-associated protein Cpf1 [Oribacterium sp. NK2B42] WP_029202018	METEILKYDF FEREGKYMYY DGLTKQYALS KTIRNELVPI GKTLDNIKKN RILEADIKRK SDYEHVKKLM DMYHKKIINE ALDNFKLSVL EDAADIYFNK QNDERDIDAF LKIQDKLRKE IVEQLKGHTD YSKVGNKDFL GLLKAASTEE DRILIESFDN FYTYFTSYNK VRSNLYSAED KSSTVAYRLI NENLPKFFDN IKAYRTVRNA GVISGDMSIV EQDELFEVDT FNHTLTQYGI DTYNHMIGQL NSAINLYNQK MHGAGSFKKL PKMKELYKQL LTEREEEFIE EYTDDEVLIT SVHNYVSYLI DYLNSDKVES FFDTLRKSDG KEVFIKNDVS KTTMSNILFD NWSTIDDLIN HEYDSAPENV KKTKDDKYFE KRQKDLKKNK SYSLSKIAAL CRDTTILEKY IRRLVDDIEK IYTSNNVFSD IVLSKHDRSK KLSKNTNAVQ	(SEQ ID NO: 167)

Table 5 - Cas12a orthologs		
	AIKNMLDSIK DFEHDVMLIN GSGQEIKKNL NVYSEQEALA GILRQVDHIY NLTRNYLTKK PFSTEKIKLN FNRPTFLDGW DKNKEEANLG ILLIKDNRYY LGIMNTSSNK AFVNPPKAIS NDIYKKVDYK LLPGPNKMLP KVFFATKNIA YYAPSEELLS KYRKGTHKKG DSFSIDDCRN LIDFFKSSIN KNTDWSTFGF NFSDTNSYND ISDFYREVEK QGYKLSFTDI DACYIKDLVD NNELYLFQIY NKDFSPYSKG KLNLHTLYFK MLFDQRNLDN VVYKLNGEAE VFYRPASIES DEQIIHKSGQ NIKNKNQKRS NCKKTSTFDY DIVKDRRYCK DKFMLHLPIT VNFGTNESGK FNELVNNAIR ADKDVNVIGI DRGERNLLYV VVVDPCGKII EQISLNTIVD KEYDIETDYH QLLDEKEGSR DKARKDWNTI ENIKELKEGY LSQVVNIIAK LVLKYDAIIC LEDLNFGFKR GRQKVEKQVY QKFEKMLIDK MNYLVLDKSR KQESPQKPGG ALNALQLTSA FKSFKELGKQ TGIIYYVPAY LTSKIDPTTG FANLFYIKYE SVDKARDFFS KFDFIRYNQM DNYFEFGFDY KSFTERASGC KSKWIACTNG ERIVKYRNSD KNNSFDDKTV ILTDEYRSLF DKYLQNYIDE DDLKDQILQI DSADFYKNLI KLFQLTLQMR NSSSDGKRDY IISPVKNYRE EFFCSEFSDD TFPRDADANG AYNIARKGLW VIKQIRETKS GTKINLAMSN SEWLEYAQCN LL	
WP_028248456.1 type V CRISPR-associated protein Cpf1 [Pseudobutyrivibrio ruminis]	MYYQNLTKMY PISKTLRNEL IPVGKTLENI RKNGILEADI QRKADYEHVK KLMDNYHKQL INEALQGVHL SDLSDAYDLY FNLSKEKNSV DAFSKCQDKL RKEIVSLLKN HENFPKIGNK EIIKLLQSLY DNDTDYKALD SFSNFYTYFS SYNEVRKNLY SDEEKSSTVA YRLINENLPK FLDNIKAYAI AKKAGVRAEG LSEEDQDCLF IIETFERTLT QDGIDNYNAA IGKLNTAINL FNQQNKKQEG FRKVPQMKCL YKQILSDREE AFIDEFSDDE DLITNIESFA ENMNVFLNSE IITDFKIALV ESDGSLVYIK NDVSKTSFSN IVFGSWNAID EKLSDEYDLA NSKKKKDEKY YEKRQKELKK NKSYDLETII GLFDDNSDVI GKYIEKLESD ITAIAEAKND FDEIVLRKHD KNKSLRKNTN AVEAIKSYLD TVKDFERDIK LINGSGQEVE KNLVVYAEQE NILAEIKNVD SLYNMSRNYL TQKPFSTEKF KLNFNRATLL NGWDKNKETD NLGILFEKDG MYYLGIMNTK ANKIFVNIPK ATSNDVYHKV NYKLLPGPNK MLPKVFFAQS NLDYYKPSEE LLAKYKAGTH KKGDNFSLED CHALIDFFKA SIEKHPDWSS FGFEFSETCT YEDLSGFYRE VEKQGYKITY TDVDADYITS LVERDELYLF QIYNKDFSPY SKGNLNLHTI YLQMLFDQRN LNNVVYKLNG EAEVFYRPAS INDEEVIIHK AGEEIKNKNS KRAVDKPTSK FGYDIIKDRR YSKDKFMLHI PVTMNFGVDE TRRFNDVVND ALRNDEKVRV IGIDRGERNL LYVVVVDTDG TILEQISLNS IINNEYSIET DYHKLLDEKE GDRDRARKNW TTIENIKELK EGYLSQVVNV IAKLVLKYNA IICLEDLNFG FKRGRQKVEK QVYQKFEKML IDKLNYLVID KSRKQDKPEE FGGALNALQL TSKFTSFKDM GKQTGIIYYV PAYLTSKIDP TTGFANLFYV KYENVEKAKE FFSRFDSISY NNESGYFEFA FDYKKFTDRA CGARSQWTVC TYGERIIKFR NTEKNNSFDD KTIVLSEEFK ELFSIYGISY EDGAELKNKI MSVDEADFFR SLTRLFQQTM QMRNSSNDVT RDYIISPIMN DRGEFFNSEA CDASKPKDAD ANGAFNIARK GLWVLEQIRN TPSGDKLNLA MSNAEWLEYA QRNQI	(SEQ ID NO: 168)
WP_028830240 type V CRISPR-associated protein Cpf1 [Proteocatella sphenisci]	MENFKNLYPI NKTLRFELRP YGKTLENFKK SGLLEKDAFK ANSRRSMQAI IDEKFKETIE ERLKYTEFSE CDLGNMTSKD KKITDKAATN LKKQVILSFD DEIFNNYLKP DKNIDALFKN DPSNPVISTF KGFTTYFVNF FEIRKHIFKG ESSGSMAYRI IDENLTTYLN NIEKIKKLPE ELKSQLEGID QIDKLNNYNE FITQSGITHY NEIIGGISKS ENVKIQGINE GINLYCQKNK VKLPRLTPLY KMILSDRVSN SFVLDTIEND TELIEMISDL INKTEISQDV IMSDIQNIFI KYKQLGNLPG ISYSSIVNAI CSDYDNNFGD GKRKKSYEND RKKHLETNVY SINYISELLT DTDVSSNIKM RYKELEQNYQ VCKENFNATN WMNIKNIKQS EKTNLIKDLL DILKSIQRFY DLFDIVDEDK NPSAEFYTWL SKNAEKLDFE FNSVYNKSRN YLTRKQYSDK KIKLNFDSPT LAKGWDANKE IDNSTIIMRK FNNDRGDYDY FLGIWNKSTP ANEKIIPLED NGLFEKMQYK LYPDPSKMLP KQFLSKIWKA KHPTTPEFDK KYKEGRHKKG PDFEKEFLHE LIDCFKHGLV NHDEKYQDVF GFNLRNTEDY NSYTEFLEDV ERCNYNLSFN KIADTSNLIN DGKLYVFQIW SKDFSIDSKG TKNLNTIYFE SLFSEENMIE KMFKLSGEAE IFYRPASLNY CEDIIKKGHH HAELKDKFDY PIIKDKRYSQ DKFFFHVPMV INYKSEKLNS KSLNNRTNEN LGQFTHIIGI DRGERHLIYL TVVDVSTGEI VEQKHLDEII NTDTKGVEHK THYLNKLEEK SKTRDNERKS WEAIETIKEL KEGYISHVIN EIQKLQEKYN ALIVMENLNY GFKNSRIKVE KQVYQKFETA LIKKFNYIID KKDPETYIHG YQLTNPITTL DKIGNQSGIV LYIPAWNTSK	(SEQ ID NO: 169)

Table 5 - Cas12a orthologs		
	IDPVTGFVNL LYADDLKYKN QEQAKSFIQK IDNIYFENGE FKFDIDFSKW NNRYSISKTK WTLTSYGTRI QTFRNPQKNN KWDSAEYDLT EEFKLILNID GTLKSQDVET YKKFMSLFKL MLQLRNSVTG TDIDYMISPV TDKTGTHFDS RENIKNLPAD ADANGAYNIA RKGIMAIENI MNGISDPLKI SNEDYLKYIQ NQQE	
WP_084502895.1 type V CRISPR-associated protein Cpf1 [Proteocatella sphenisci]	MIILYISTSN MNMEGVFMEN FKNLYPINKT LRFELRPYGK TLENFKKSGL LEKDAFKANS RRSMQAIIDE KFKETIEERL KYTEFSECDL GNMTSKDKKI TDKAATNLKK QVILSFDDEI FNNYLKPDKN IDALFKNDPS NPVISTFKGF TTYFVNFFEI RKHIFKGESS GSMAYRIIDE NLTTYLNNIE KIKKLPEELK SQLEGIDQID KLNNYNEFIT QSGITHYNEI IGGISKSENV KIQGINEGIN LYCQKNKVKL PRLTPLYKMI LSDRVSNSFV LDTIENDTEL IEMISDLINK TEISQDVIMS DIQNIFIKYK QLGNLPGISY SSIVNAICSD YDNNFGDGKR KKSYENDRKK HLETNVYSIN YISELLTDTD VSSNIKMRYK ELEQNYQVCK ENFNATNWMN IKNIKQSEKT NLIKDLLDIL KSIQRFYDLF DIVDEDKNPS AEFYTWLSKN AEKLDFEFNS VYNKSRNYLT RKQYSDKKIK LNFDSPTLAK GWDANKEIDN STIIMRKFNN DRGDYDYFLG IWNKSTPANE KIIPLEDNGL FEKMQYKLYP DPSKMLPKQF LSKIWKAKHP TTPEFDKKYK EGRHKKGPDF EKEFLHELID CFKHGLVNHD EKYQDVFGFN LRNTEDYNSY TEFLEDVERC NYNLSFNKIA DTSNLINDGK LYVFQIWSKD FSIDSKGTKN LNTIYFESLF SEENMIEKMF KLSGEAEIFY RPASLNYCED IIKKGHHHAE LKDKFDYPII KDKRYSQDKF FFHVPMVINY KSEKLNSKSL NNRTNENLGQ FTHIIGIDRG ERHLIYLTVV DVSTGEIVEQ KHLDEIINTD TKGVEHKTHY LNKLEEKSKT RDNERKSWEA IETIKELKEG YISHVINEIQ KLQEKYNALI VMENLNYGFK NSRIKVEKQV YQKFETALIK KFNYIIDKKD PETYIHGYQL TNPITTLDKI GNQSGIVLYI PAWNTSKIDP VTGFVNLLYA DDLKYKNQEQ AKSFIQKIDN IYFENGEFKF DIDFSKWNNR YSISKTKWTL TSYGTRIQTF RNPQKNNKWD SAEYDLTEEF KLILNIDGTL KSQDVETYKK FMSLFKLMLQ LRNSVTGTDI DYMISPVTDK TGTHFDSREN IKNLPADADA NGAYNIARKG IMAIENIMNG ISDPLKISNE DYLKYIQNQQ E	(SEQ ID NO: 170)
WP_055225123.1 Eubacterium rectale	MNNGTNNFQN FIGISSLQKT LRNALIPTET TQQFIVKNGI IKEDELRGEN RQILKDIMDD YYRGFISETL SSIDDIDWTS LFEKMEIQLK NGDNKDTLIK EQTEYRKAIH KKFANDDRFK NMFSAKLISD ILPEFVIHNN NYSASEKEEK TQVIKLFSRF ATSFKDYFKN RANCFSADDI SSSSCHRIVN DNAEIFFSNA LVYRRIVKSL SNDDINKISG DMKDSLKEMS LEEIYSYEKY GEFITQEGIS FYNDICGKVN SFMNLYCQKN KENKNLYKLQ KLHKQILCIA DTSYEVPYKF ESDEEVYQSV NGFLDNISSK HIVERLRKIG DNYNGYNLDK IYIVSKFYES VSQKTYRDWE TINTALEIHY NNILPGNGKS KADKVKKAVK NDLQKSITEI NELVSNYKLC SDDNIKAETY IHEISHILNN FEAQELKYNP EIHLVESELK ASELKNVLDV IMNAFHWCSV FMTEELVDKD NNFYAELEEI YDEIYPVISL YNLVRNYVTQ KPYSTKKIKL NFGIPTLADG WSKSKEYSNN AIILMRDNLY YLGIFNAKNK PDKKIIEGNT SENKGDYKKM IYNLLPGPNK MIPKVFLSSK TGVETYKPSA YILEGYKQNK HIKSSKDFDI TFCHDLIDYF KNCIAIHPEW KNFGFDFSDT STYEDISGFY REVELQGYKI DWTYISEKDI DLLQEKGQLY LFQIYNKDFS KKSTGNDNLH TMYLKNLFSE ENLKDIVLKL NGEAEIFFRK SSIKNPIIHK KGSILVNRTY EAEEKDQFGN IQIVRKNIPE NIYQELYKYF NDKSDKELSD EAAKLKNVVG HHEAATNIVK DYRYTYDKYF LHMPITINFK ANKTGFINDR ILQYIAKEKD LHVIGIDRGE RNLIYVSVID TCGNIVEQKS FNIVNGYDYQ IKLKQQEGAR QIARKEWKEI GKIKEIKEGY LSLVIHEISK MVIKYNAIIA MEDLSYGFKK GRFKVERQVY QKFETMLINK LNYLVFKDIS ITENGGLLKG YQLTYIPDKL KNVGHQCGCI FYVPAAYTSK IDPTTGFVNI FKFKDLTVDA KREFIKKFDS IRYDSEKNLF CFTFDYNNFI TQNTVMSKSS WSVYTYGVRI KRRFVNGRFS NESDTIDITK DMEKTLEMTD INWRDGHDLR QDIIDYEIVQ HIFEIFRLTV QMRNSLSELE DRDYDRLISP VLNENNIFYD SAKAGDALPK DADANGAYCI ALKGLYEIKQ ITENWKEDGK FSRDKLKISN KDWFDFIQNK RYL	(SEQ ID NO: 171)
WP_055237260.1 Eubacterium rectale	MNNGTNNFQN FIGISSLQKT LRNALIPTET TQQFIVKNGI IKEDELRGEN RQILKDIMDD YYRGFISETL SSIDDIDWTS LFEKMEIQLK NGDNKDTLIK EQAEKRKAIY KKFADDDRFK NMFSAKLISD ILPEFVIHNN NYSASEKEEK TQVIKLFSRF ATSFKDYFKN RANCFSADDI SSSSCHRIVN DNAEIFFSNA LVYRRIVKNL SNDDINKISG DMKDSLKEMS LDEIYSYEKY GEFITQEGIS	(SEQ ID NO: 172)

Table 5 - Cas12a orthologs		
	FYNDICGKVN SFMNLYCQKN KENKNLYKLR KLHKQILCIA DTSYEVPYKF ESDEEVYQSV NGFLDNISSK HIVERLRKIG DNYNGYNLDK IYIVSRFYES VSQKTYRDWE TINTALEIHY NNILPGNGKS KADKVKKAVK NDLQKSITEI NELVSNYKLC PDDNIKAETY IHEISHILNN FEAQELKYNP EIHLVESELK ASELKNVLDV IMNAFHWCSV FMTEELVDKD NNFYAELEEI YDEIYPVISL YNLVRNYVTQ KPYSTKKIKL NFGIPTLADG WSKSKEYSNN AIILMRDNLY YLGIFNAKNK PDKKIIEGNT SENKGDYKKM IYNLLPGPNK MIPKVFLSSK TGVETYKPSA YILEGYKQNK HLKSSKDFDI TFCRDLIDYF KNCIAIHPEW KNFGFDFSDT STYEDISGFY REVELQGYKI DWTYISEKDI DLLQEKGQLY LFQIYNKDFS KKSTGNDNLH TMYLKNLFSE ENLKDIVLKL NGEAEIFFRK SSIKNPIIHK KGSILVNRTY EAEEKDQFGN IQIVRKTIPE NIYQELYKYF NDKSDKELSD EAAKLKNVVG HHEAATNIVK DYRYTYDKYF LHMPITINFK ANKTSFINDR ILQYIAKEND LHVIGIDRGE RNLIYVSVID TCGNIVEQKS FNIVNGYDYQ IKLKQQEGAR QIARKEWKEI GKIKEIKEGY LSLVIHEISK MVIKYNAIIA MEDLSYGFKK GRFKVERQVY QKFETMLINK LNYLVFKDIS ITENGGLLKG YQLTYIPEKL KNVGHQCGCI FYVPAAYTSK IDPTTGFANI FKFKDLTVDA KREFIKKFDS IRYDSEKNLF CFTFDYNNFI TQNTVMSKSS WSVYTYGVRI KRRFVNGRFS NESDTIDITK DMEKTLEMTD INWRDGHDLR QDIIDYEIVQ HIFEIFKLTV QMRNSLSELE DRDYDRLISP VLNENNIFYD SAKAGDALPK DADANGAYCI ALKGLYEIKQ ITENWKEDGK FSRDKLKISN KDWFDFIQNK RYL	
WP_055272206.1 Eubacterium rectale	MNNGTNNFQN FIGISSLQKT LRNALTPTET TQQFIVKNGI IKEDELRGEN RQILKDIMDD YYRGFISETL SSIDDIDWTS LFEKMEIQLK NGDNKDTLIK EQAEKRKAIY KKFADDDRFK NMFSAKLISD ILPEFVIHNN NYSASEKEEK TQVIKLFSRF ATSFKDYFKN RANCFSADDI SSSSCHRIVN DNAEIFFSNA LVYRRIVKNL SNDDINKISG DMKDSLKKMS LEKIYSYEKY GEFITQEGIS FYNDICGKVN SFMNLYCQKN KENKNLYKLR KLHKQILCIA DTSYEVPYKF ESDEEVYQSV NGFLDNISSK HIVERLRKIG DNYNGYNLDK IYIVSKFYES VSQKTYRDWE TINTALEIHY NNILPGNGKS KADKVKKAVK NDLQKSITEI NELVSNYKLC PDDNIKAETY IHEISHILNN FEAQELKYNP EIHLVESELK ASELKNVLDV IMNAFHWCSV FMTEELVDKD NNFYAELEEI YDEIYPVISL YNLVRNYVTQ KPYSTKKIKL NFGIPTLADG WSKSKEYSNN AIILMRDNLY YLGIFNAKNK PEKKIIEGNT SENKGDYKKM IYNLLPGPNK MIPKVFLSSK TGVETYKPSA YILEGYKQNK HLKSSKDFDI TFCRDLIDYF KNCIAIHPEW KNFGFDFSDT STYEDISGFY REVELQGYKI DWTYISEKDI DLLQEKGQLY LFQIYNKDFS KKSTGNDNLH TMYLKNLFSE ENLKDVVLKL NGEAEIFFRK SSIKNPIIHK KGSILVNRTY EAEEKDQFGN IQIVRKTIPE NIYQELYKYF NDKSDKELSD EAAKLKNAVG HHEAATNIVK DYRYTYDKYF LHMPITINFK ANKTSFINDR ILQYIAKEKD LHVIGIDRGE RNLIYVSVID TCGNIVEQKS FNIVNGYDYQ IKLKQQEGAR QIARKEWKEI GKIKEIKEGY LSLVIHEISK MVIKYNAIIA MEDLSYGFKK GRFKVERQVY QKFETMLINK LNYLVFKDIS ITENGGLLKG YQLTYIPEKL KNVGHQCGCI FYVPAAYTSK IDPTTGFVNI FKFKDLTVDA KREFIKKFDS IRYDSDKNLF CFTFDYNNFI TQNTVMSKSS WSVYTYGVRI KRRFVNGRFS NESDTIDITK DMEKTLEMTD INWRDGHDLR QDIIDYEIVQ HIFEIFKLTV QMRNSLSELE DRNYDRLISP VLNENNIFYD SAKAGDALPK DADANGAYCI ALKGLYEIKQ ITENWKEDGK FSRDKLKISN KDWFDFIQNK RYL	(SEQ ID NO: 173)
OLA16049.1 Eubacterium sp. 41_20	MNNGTNNFQN FIGISSLQKT LRNALIPTET TQQFIVKNGI IKEDELRGKN RQILKDIMDD YYRGFISETL SSIDDIDWTS LFEKMEIQLK NGDNKDTLIK EQAEKRKAIY KKFADDDRFK NMFSAKLISD ILPEFVIHNN NYSASEKKEK TQVIKLFSRF ATSFKDYFKN RANCFSADDI SSSSCHRIVN DNAEIFFSNA LVYRRIVKNL SNDDINKISG DMKDSLKEMS LEEIYSYEKY GEFITQEGIS FYNDICGKVN SFMNLYCQKN KENKNLYKLR KLHKQILCIA DTSYEVPYKF ESDEEVYQSV NGFLDNISSK HIVERLRKIG DNYNDYNLDK IYIVSKFYES VSQKTYRDWE TINTALEIHY NNILPGNGKS KADKVKKAVK NDLQKSITEI NELVSNYKLC SDDNIKAETY IHEISHILNN FEAHELKYNP EIHLVESELK ASELKNVLDI IMNAFHWCSV FMTEELVDKD NNFYAELEEI YDEIYPVISL YNLVRNYVTQ KPYSTKKIKL NFGIPTLADG WSKSKEYSNN AIILMRDNLY YLGIFNAKNK PDKKIIEGNT SENKGDYKKM IYNLLPGPNK MIPKVFLSSK TGVETYKPSA YILEGYKQNK HLKSSKDFDI TFCHDLIDYF KNCIAIHPEW	(SEQ ID NO: 174)

Table 5 - Cas12a orthologs		
	KNFGFDFSDT SAYEDISGFY REVELQGYKI DWTYISEKDI DLLQEKGQLY LFQIYNKDFS KKSTGNDNLH TMYLKNLFSE ENLKDIVLKL NGEAEIFFRK SSIKNPIIHK KGSILVNRTY EAEEKDQFGN IQIVRKTIPE NIYQELYKYF NDKSDKELSD EAAKLKNVVG HHEAATNIVK DYRYTYDKYF LHMPITINFK ANKTSFINDR ILQYIAKEKD LHVIGIDRGE RNLIYVSVID TCGNIVEQKS FNIVNGYDYQ IKLKQQEGAR QIARKEWKEI GKIKEIKEGY LSLVIHEISK MVIKYNAIIA MEDLSYGFKK GRFKVERQVY QKFETMLINK LNYLVFKDIS ITENGGLLKG YQLTYIPDKL KNVGHQCGCI FYVPAAYTSK IDPTTGFVNI FKFKDLTVDA KREFIKKFDS IRYDSEKNLF CFTFDYNNFI TQNTVMSKSS WSVYTYGVRI KRRFVNGRFS NESDTIDITK DMEKTLEMTD INWRDGHDLR QDIIDYEIVQ HIFEIFKLTV QMRNSLSELE DRDYDRLISP VLNENNIFYD SAKAGYALPK DADANGAYCI ALKGLYEIKQ ITENWKEDGK FSRDKLKISN KDWFDFIQNK RYL	

Table 6 - Cas12b (C2c1) orthologs		
Alicyclobacillus macrosporangiidus strain DSM 17980 WP_074948407.1	MVAVKSIKVK LMLGHLPEIR EGLWHLHEAV NLGVRYYTEW LALLRQGNLY RRGKDGAQEC YMTAEQCRQE LLVRLRDRQK RNGHTGDPGT DEELLGVARR LYELLVPQSV GKKGQAQMLA SGFLSPLADP KSEGGKGTSK SGRKPAWMGM KEAGDSRWVE AKARYEANKA KDPTKQVIAS LEMYGLRPLF DVFTETYKTI RWMPLGKHQG VRAWDRDMFQ QSLERLMSWE SWNERVGAEF ARLVDRRDRF REKHFTGQEH LVALAQRLEQ EMKEASPGFE SKSSQAHRIT KRALRGADGI IDDWLKLSEG EPVDRFDEIL RKRQAQNPRR FGSHDLFLKL AEPVFQPLWR EDPSFLSRWA SYNEVLNKLE DAKQFATFTL PSPCSNPVWA RFENAEGTNI FKYDFLFDHF GKGRHGVRFQ RMIVMRDGVP TEVEGIVVPI APSRQLDALA PNDAASPIDV FVGDPAAPGA FRGQFGGAKI QYRRSALVRK GRREEKAYLC GFRLPSQRRT GTPADDAGEV FLNLSLRVES QSEQAGRRNP PYAAVFHISD QTRRVIVRYG EIERYLAEHP DTGIPGSRGL TSGLRVMSVD LGLRTSAAIS VFRVAHRDEL TPDAHGRQPF FFPIHGMDHL VALHERSHLI RLPGETESKK VRSIREQRLD RLNRLRSQMA SLRLLVRTGV LDEQKRDRNW ERLQSSMERG GERMPSDWWD LFQAQVRYLA QHRDASGEAW GRMVQAAVRT LWRQLAKQVR DWRKEVRRNA DKVKIRGIAR DVPGGHSLAQ LDYLERQYRF LRSWSAFSVQ AGQVVRAERD SRFAVALREH IDNGKKDRLK KLADRILMEA LGYVYVTDGR RAGQWQAVYP PCQLVLLEEL SEYRFSNDRP PSENSQLMVW SHRGVLEELI HQAQVHDVLV GTIPAAFSSR FDARTGAPGI RCRRVPSIPL KDAPSIPIWL SHYLKQTERD AAALRPGELI PTGDGEFLVT PAGRGASGVR VVHADINAAH NLQRRLWENF DLSDIRVRCD RREGKDGTVV LIPRLTNQRV KERYSGVIFT SEDGVSFTVG DAKTRRRSSA SQGEGDDLSD EEQELLAEAD DARERSVVLF RDPSGFVNGG RWTAQRAFWG MVHNRIETLL AERFSVSGAA EKVRG	(SEQ ID NO: 175)
Bacillus hisashii strain C4 WP_095142515.1	MATRSFILKI EPNEEVKKGL WKTHEVLNHG IAYYMNILKL IRQEAIYEHH EQDPKNPKKV SKAEIQAELW DFVLKMQKCN SFTHEVDKDE VFNILRELYE ELVPSSVEKK GEANQLSNKF LYPLVDPNSQ SGKGTASSGR KPRWYNLKIA GDPSWEEEKK KWEEDKKKDP LAKILGKLAE YGLIPLFIPY TDSNEPIVKE IKWMEKSRNQ SVRRLDKDMF IQALERFLSW ESWNLKVKEE YEKVEKEYKT LEERIKEDIQ ALKALEQYEK ERQEQLLRDT LNTNEYRLSK RGLRGWREII QKWLKMDENE PSEKYLEVFK DYQRKHPREA GDYSVYEFLS KKENHFIWRN HPEYPYLYAT FCEIDKKKKD AKQQATFTLA DPINHPLWVR FEERSGSNLN KYRILTEQLH TEKLKKKLTV QLDRLIYPTE SGGWEEKGKV DIVLLPSRQF YNQIFLDIEE KGKHAFTYKD ESIKFPLKGT LGGARVQFDR DHLRRYPHKV ESGNVGRIYF NMTVNIEPTE SPVSKSLKIH RDDFPKVVNF KPKELTEWIK DSKGKKLKSG IESLEIGLRV MSIDLGQRQA AAASIFEVVD QKPDIEGKLF FPIKGTELYA VHRASFNIKL PGETLVKSRE VLRKAREDNL KLMNQKLNFL RNVLHFQQFE DITEREKRVT KWISRQENSD VPLVYQDELI QIRELMYKPY KDWVAFLKQL HKRLEVEIGK EVKHWRKSLS DGRKGLYGIS LKNIDEIDRT RKFLLRWSLR PTEPGEVRRL EPGQRFAIDQ LNHLNALKED RLKKMANTII MHALGYCYDV RKKKWQAKNP ACQIILFEDL SNYNPYEERS RFENSKLMKW SRREIPRQVA LQGEIYGLQV GEVGAQFSSR FHAKTGSPGI RCSVVTKEKL QDNRFFKNLQ REGRLTLDKI AVLKEGDLYP DKGGEKFISL SKDRKCVTTH	(SEQ ID NO: 176)

Table 6 - Cas12b (C2c1) orthologs		
	ADINAAQNLQ KRFWTRTHGF YKVYCKAYQV DGQTVYIPES KDQKQKIIEE FGEGYFILKD GVYEWVNAGK LKIKKGSSKQ SSSELVDSDI LKDSFDLASE LKGEKLMLYR DPSGNVFPSD KWMAAGVFFG KLERILISKL TNQYSISTIE DDSSKQSM	
Candidatus Lindowbacteria bacterium RIFCSPLOWO2 OGH55994.1	MPRDDLDLLT NLNSTAKGIR ERGKTKEGTD KKKSGRKSSW PMDKAAWETA KTSDSSAHFL EKLKQHPDLK DAFGNLSSGG SKKLEYYKKL AGSAPWKESQ SVILEKAARW KEAKQEREEK EQDSSEHGSK AAYRRLFDAG CLPMPEFAKY IDENQIEFGD LKLSDCGAEW KRGMWNQAGQ RVRSHMGWQR RREKENAVYS LRKELFEKGG AIRRKKSEEL TPEDILPGKA APDQNDWQER PAYGNQMWFI GLRSYEENEM AKYAEEAGMG SRSAPRIRRG TIKGWSKLRE RWLQILKRNP QATRDDLIGE LNALRSQDPR AYGDARLFDW LSKTDQRFLW DGFDADGKIL CGRDDRDCVS AFVAYNEEFA DEPSSITLTE TDERLHPVWP FFGESSAVPY EIEYDLETAC PTAIRLPLLV GKENGGYAER QGTRLPLAEY ADLASSFQLP TPVRLDVLVE IREVTRAGRK VTCPFSYFKQ NGVWYVREGE IPSGESIQIK QTDRKIENGK IFISSKLRMA YRDDLMVSPA TGDFGSIKIL WERIELASHV DQKKLPETAP ARSRVFVSFS CNVVERAPRK QLTRKPDAVV VTIPSGVDQG LVVVSTDVRT GKSKSSSAPP LPPGSRLWPA DAVHGDPPLR ILSVDLGHRH SAYAVWELGL QQKSWRAGVL KGSTQTPVYA DCTGTGLLCL PGDGEDTPAE EESLRLRSRQ IRRRLNLQNS ILRVSRLLSL DKFEKTIFEQ SDVRDRPNKK GLRIRRRCRT EKTPLSEAEV RKNCDKAAEI LIRWADTDAM AKSLAATGNA DISFWKYMAV KNPPLSAVVD VAPSTIVPDD GPDRETLKKK RQEEEEKFAS SIYENRVKLA GALCSGYDAD HRRPATGGLW HDLDRTLIRE ISYGDRGQKG NPRKLNNEGI LRLLRRPPRA RPDWREFHRT LNDANRIPKG RTLRGGLSMG RLNFLKEVGD FVKKWSCRPR WPGDRRHIPP GQLFDRQDAE HLEHLRDDRI KRLAHLIVAQ ALGFEPDIRR GLWKYVDGST GEILWQHPET RRFFAEGAAG ELREVSRPAE IDDDAAARPH TVSAPAHIVV FENLIRYRFQ SDRPKTENAG LMQWAHRQIV HFTKQVASLY GLKVAMVYAA FSSKFCSRCG SPGARVSRFD PAWRNQEWFK RRTSNPRSKV DHSLKRASED PTADETRPWV LIEGGKEFVC ANAKCSAHDE PLNADENAAA NIGLRFLRGV EDFRTKVNPA GALKGKLRFE TGIHSFRPPV SGSPFWSPMA EPAQKKKIGA AAPGADVDEA GDADESGVVV LFRDPSGAFR NKQYWYEGKI FWSNVMMAVE AKIAGASVGA KPVAASWGQA QPQSGPGLAK PGGD	(SEQ ID NO: 177)
Elusimicrobia bacterium RIFOXYA12 OGS02326.1	MNRIYQGRVT KVEVPDGKDE KGNIKWKKLE NWSDILWQHH MLFQDAVNYY TLALAAISGS AVGSDEKSII LREWAVQVQN IWEKAKKKAT VFEGPQKRLT SILGLEQNAS FDIAAKHILR TSEAKPEQRA SALIRLLEEI DKKNHNVVCG ERLPFFCPRN IQSKRSPTSK AVSSVQEQKR QEEVRRFHNM QPEEVVKNAV TLDISLFKSS PKIVFLEDPK KARAELLKQF DNACKKHKEL VGIKKAFTES IDKHGSSLKV PAPGSKPSGL YPSAIVFKYF PVDITKTVFL KATEKLAMGK DREVTNDPIA DARVNDKPHF DYFTNIALIR EKEKNRAAWF EFDLAAFIEA IMSPHRFYQD TQKRKEAARK LEEKIKAIEG KGGQFKESDS EDDDVDSLPG FEGDTRIDLL RKLVTDTLGW LGESETPDNN EGKKTEYSIS ERTLRIFPDI QKQWSELAEK GETTEGKLLE VLKHEQTEHQ SDFGSATLYQ HLAKPEFHPI WLKSGTEEWH AENPLKAWLN YKELQYELTD KKRPIHFTPA HPVYSPRYFD FPKKSETEEK EVSKNTHSLT TSLASEHIKN SLQFTAGLIR KTNVGKKAIK ARFSYSAPRL RRDCLRSENN ENLYKAPWLQ PMMRALGIDE EKADRQNFAN TRITLMAKGL DDIQLGFPVE ANSQELQKEV SNGISWKGQF NWGGIASLSA LRWPHEKKPK NPPEQPWWGI DSFSCLAVDL GQRYAGAFAR LDVSTIEKKG KSRFIGEACD KKWYAKVSRM GLLRLPGEDV KVWRDASKID KENGFAFRKE LFGEKGRSAT PLEAEETAEL IKLFGANEKD VMPDNWSKEL SFPEQNDKLL IVARRAQAAV SRLHRWAWFF DEAKRSDDAI REILESDDTD LKQKVNKNEI EKVKETIISL LKVKQELLPT LLTRLANRVL PLRGRSWEWK KHHQKNDGFI LDQTGKAMPN VLIRGQRGLS MDRIEQITEL RKRFQALNQS LRRQIGKKAP AKRDDSIPDC CPDLLEKLDH MKEQRVNQTA HMILAEALGL KLAEPPKDKK ELNETCDMHG AYAKVDNPVS FIVIEDLSRY RSSQGRSPRE NSRLMKWCHR AVRDKLKEMC EVFFPLCERR KAGSAWVSLP PLLETPAAYS SRFCSRSGVA GFRAVEVIPG FELKYPWSWL KDKKDKAGNL AKEALNIRTV SEQLKAFNQD KPEKPRTLLV PIAGGPIFVP ISEVGLSSFG LKPQVVQADI NAAINLGLRA ISDPRIWEIH PRLRTEKRDG RLFAREKRKY GEEKVEVQPS KNEKAKKVKD DRKPNYFADF SGKVDWGFGN IKNESGLTLV SGKALWWTIN QLQWERCFDI NKRHIEDWSN KQKQ	(SEQ ID NO: 178)

Table 6 - Cas12b (C2c1) orthologs

Omnitrophica WOR_2 bacterium RIFCSPHIGHO2 OGX36711.1	MNRIYQGRVT KVEKLKNGKS PDDREELKDW QTALWRHHEL FQDAVSYYTL ALAAMAEGLP DKHPINVLRK RMEEAWEEFP RKTVTPAKNL RDSVRPWLGL SESASFGDAL KKILPPAPEN KEVRALAVAL LAEKARTLKP QKTSASYWGR FCDDLKKKPN WDYSEEELAR KTGSGDWVAG LWSEDALNKI DELAKSLKLS SLVKCVPDGQ INPEGARNLV KEALDHLEGV SNGTKKEKND PGPAKKTNNW LRQHASDVRN FIHKNKNQFS SLPNGRLITE RARGGGININ KTYAGVLFKA FPCPFTFDYV RAAVPEPKVK KVDQEKKSEQ SATWTELEKR ILRIGDDPIE LARKNNKPIF KAFTALEKWS DQNSKSCWSD FDKCAFEEAL KTLNQFNQKT EEREKRRSEA EAELKYMMDE NPEWKPKKET EGDDVREVPI LKGDPRYEKL VKLFGDLDEE GSEHATGKIY GPSRASLRGF GKLRNEWVDL FTKANDNPRE QDLQKAVTGF QREHKLDMGY TAFFLKLCER DYWDIWRDDT EVEVKKIREK RWVKSVVYAA ADTRELAEEL ERLQEPVRYT PAEPQFSRRL FMFSDIKGKQ GAKHIREGLV EVSLAVKDQS GKYGTCRVRL HYSAPRLIRD HLSDGSSSMW LQPMMAALGL SSDARGCFTR DSKGNVKEPA VALMSDFVGR KRELRMLLNF PVDLDISKLE ENIGKKARWE KQMNTAYEKN KLKQRFHLIW PGMELKETQE PGQFWWDNPT IQKEGMYCLA IDLSQRRAAD YALLHAGVNR DSKTFVELGQ AGGQSWFTKL CAAGSLRLPG EDTEVIREGK RQIELSGKKG RNATQSEYDQ AIALAKQLLH NENSAELESA ARDWLGDNAK RFSFPEQNDK LIDLYYGALS RYKTWLRWSW RLTEQHKELW DKTLDEIRKV PYFASWGELA GNGTNEATVQ QLQKLIADAA VDLRNFLEKA LLHIAYRALP LRENTWRWIE NGKDGKGKPL HLLVSDGQSP AEIPWLRGQR GLSIARIEQL ENFRRAVLSL NRLLRHEIGT KPEFGSSTCG ESLPDPCPDL TDKIVRLKEE RVNQTAHLII AQSLGVRLKG HSLFTEEREK ADMHGEHEVI PGRSPVDFVV LEDLSRYTTD KSRSRSENSR LMKWCHRKIN EKVKLLAEPF GIPVIEVFAS YSSKFDARTG APGFRAVEVT SEDRPFWRKT IEKQSVAREV FDCLDNLVGK GLNGIHLVLP QNGGPLFIAA VKEDQPLPAI RQADINAAVN IGLRAIAGPS CYHAHPKVRL IKGESGTDKG KWLPRKGKEA NKRENAQFGN VDLDLEVKFN RLDIDSDVLK GDNTNLFHDP LNIACYGFAT IQNLQHPFLA HASAVFSRQK GAVARLQWEV CRAINSRRLE AWQKKAEKAA VKR	(SEQ ID NO: 179)
Phycisphaerae bacterium ST-NAGAB-D1 (transposase) AQT69685.1	MATKSYRARI LTDSRLAAAL DRTHVVFVES LKQMINTYLR MQNGKFGPDH KKLAQIMLSR SNTFAHGVMD QITRDQPTST LDEEWTDLAR RIHKTTGPLF LQAERFATVK NRAIHTKSRG KVIPSPETLA VPAKFWHQVC DSASAYIRSN RELMQQWRKD RAAWLKDKNE WQQKHPEFMQ FYNGPYQNFL KLCDDDRITS QLAAEQQPTA SKNNRPRKTG KRFARWHLWY KWLSENPEII EWRNKASASD FKTVTDDVRK QIITKYPQQN KYITRLLDWL EDNNPELKTL ENLRRTYVKK FDSFKRPPTL TLPSPYRHPY WFTMELDQFY KKADFENGTI QLLLIDEDDD GNWFFNWMPA SLKPDPRLVP SWRAETFETE GRFPPYLGGK IGKKLSRPAP TDAERKAGIA GAKLMIKNNR SELLFTVFEQ DCPPRVKWAK TKNRKCPADN AFSSDGKTRK PLRILSIDLG IRHIGAFALT QGTRNDSAWQ TESLKKGIIN SPSIPPLRQV RRHDYDLKRK RRRHGKPVKG QRSNANLQAH RTNMAQDRFK KGASAIVSLA REHSADLILF ENLHSLKFSA FDERWMNRQL RDMNRRHIVE LVSEQAPEFG ITVKDDINPW MTSRICSNCN LPGFRFSMKK KNPYREKLPR EKCTDFGYPV WEPGGHLFRC PHCDHRVNAD INAAANLANK FFGLGYWNNG LKYDAETKTF TVHTDKKTPP LIFKPRPQFD LWADSVKTRK QLGPDPF	(SEQ ID NO: 180)
Planctomycetes bacterium RBG_13_46_10 OHB62175.1	MSVRSFQARV ECDKQTMEHL WRTHKVFNER LPEIIKILFK MKRGECGQND KQKSLYKSIS QSILEANAQN ADYLLNSVSI KGWKPGTAKK YRNASFTWAD DAAKLSSQGI HVYDKKQVLG DLPGMMSQMV CRQSVEAISG HIELTKKWEK EHNEWLKEKE KWESEDEHKK YLDLREKFEQ FEQSIGGKIT KRRGRWHLYL KWLSDNPDFA AWRGNKAVIN PLSEKAQIRI NKAKPNKKNS VERDEFFKAN PEMKALDNLH GYYERNFVRR RKTKKNPDGF DHKPTFTLPH PTIHPRWFVF NKPKTNPEGY RKLILPKKAG DLGSLEMRLL TGEKNKGNYP DDWISVKFKA DPRLSLIRPV KGRRVVRKGK EQGQTKETDS YEFFDKHLKK WRPAKLSGVK LIFPDKTPKA AYLYFTCDIP DEPLTETAKK IQWLETGDVT KKGKKRKKKV LPHGLVSCAV DLSMRRGTTG FATLCRYENG KIHILRSRNL WVGYKEGKGC HPYRWTEGPD LGHIAKHKRE IRILRSKRGK PVKGEESHID LQKHIDYMGE DRFKKAARTI VNFALNTENA ASKNGFYPRA DVLLLENLEG LIPDAEKERG INRALAGWNR RHLVERVIEM AKDAGFKRRV FEIPPYGTSQ VCSKCGALGR RYSIIRENNR REIRFGYVEK LFACPNCGYC ANADHNASVN LNRRFLIEDS FKSYYDWKRL SEKKQKEEIE TIESKLMDKL CAMHKISRGS ISK	(SEQ ID NO: 181)

Table 6 - Cas12b (C2c1) orthologs

Spirochaetes bacterium GWB1_27_13 OHD16008.1	MSFTISYPFK LIIKNKDEAK ALLDTHQYMN EGVKYYLEKL LMFRQEKIFI GEDETGKRIY IEETEYKKQI EEFYLIKKTE LGRNLTLTLD EFKTLMRELY ICLVSSSMEN KKGFPNAQQA SLNIFSPLFD AESKGYILKE ENNNISLIHK DYGKILLKRL RDNNLIPIFT KFTDIKKITA KLSPTALDRM IFAQAIEKLL SYESWCKLMI KERFDKEVKI KELENKCENK QERDKIFEIL EKYEEERQKT FEQDSGFAKK GKFYITGRML KGFDEIKEKW LKEKDRSEQN LINILNKYQT DNSKLVGDRN LFEFIIKLEN QCLWNGDIDY LKIKRDINKN QIWLDRPEMP RFTMPDFKKH PLWYRYEDPS NSNFRNYKIE VVKDENYITI PLITERNNEY FEENYTFNLA KLKKLSENIT FIPKSKNKEF EFIDSNDEEE DKKDQKKSKQ YIKYCDTAKN TSYGKSGGIR LYFNRNELEN YKDGKKMDSY TVFTLSIRDY KSLFAKEKLQ PQIFNTVDNK ITSLKIQKKF GNEEQTNFLS YFTQNQITKK DWMDEKTFQN VKELNEGIRV LSVDLGQRFF AAVSCFEIMS EIDNNKLFFN LNDQNHKIIR INDKNYYAKH IYSKTIKLSG EDDDLYKERK INKNYKLSYQ ERKNKIGIFT RQINKLNQLL KIIRNDEIDK EKFKELIETT KRYVKNTYND GIIDWNNVDN KILSYENKED VINLHKELDK KLEIDFKEFI RECRKPIFRS GGLSMQRIDF LEKLNKLKRK WVARTQKSAE SIVLTPKFGY KLKEHINELK DNRVKQGVNY ILMTALGYIK DNEIKNDSKK KQKEDWVKKN RACQIILMEK LTEYTFAEDR PREENSKLRM WSHRQIFNFL QQKASLWGIL VGDVFAPYTS KCLSDNNAPG IRCHQVTKKD LIDNSWFLKI VVKDDAFCDL IEINKENVKN KSIKINDILP LRGGELFASI KDGKLHIVQA DINASRNIAK RFLSQINPFR VVLKKDKDET FHLKNEPNYL KNYYSILNFV PTNEELTFFK VEENKDIKPT KRIKMDKHEK ESTDEGDDYS KNQIALFRDD SGIFFDKSLW VDGKIFWSVV KNKMTKLLRE RNNKKNGSK	(SEQ ID NO: 182)
Verrucomicrobiaceae bacterium UBA2429 GCA_002343505.1	MPLSRIYQGR TNSLIILTPT PQEPWDHKAL ARFDSPLWRH HALFQDAVNY YQLCLVALAS SDGTRPLSKL HEQMKASWDE AKTDTEDSWR VRLARRLGIP AASLFEAALA KVLEGNEAPE RARELAGELL LDKIEGDIQQ AGRGYWPRFC DPKANPTYDY SATARASASG LTKLAAVIHA ENVTEEALKQ VAAEMDLSWT VKLQPDKNFV GAEARARLLE AAHHFIKVAE SPPTKLAEVL ARFPDGLALW QALPEKIAAL PEETQVPRNR KASPDLTFAT LLFQHFPSLF TAAVLGLSVG KPKSVKAPKV VEKVSARRKA NAVTQAVVIE EPEIDFAELG DDPIKLARGE RGFVFPAFTS LSFWAVPGPH VPVWKEFDIA AFKEALKTVN QFKLKTSERN ALLAEAQRRL DYMDEKTHDW KTGDSDEPGH IPPRLKSDPN FTLIQALTQD EGVSNKATGD QHIPKGVYTG GLRGFYAIKK DWCELWERKA DKSQGTPTEE ELISIVTDYQ RDHVYDVGDV GLFRALCEPR FWPLWQPLTD EQEAERIKAG RAKDMISAYR VWLELQEDVV RLAQPIRFTP AHAENSRRLF MFSDISGSHG AEFGSDGKSL EVSIAYDVDG KLQPVRAKLE FSAPRAARDE LEGLSGGSES MRWFQPMMKA LDCPEVEMPA LEKCAVSLMP DVVKKGGGKW VRLLLNFPAT LEPEGLIRHI GKQAMWYKQF NGTYKPRTQQ LDTGLHLYWP GLEKAPEAED AAAWWNREEI RAKGFSVLSV DLGQRDAGAW ALLESRSDKA FSRNRQPFIE LGEAGGKLWS TALLGLGMLR LPGEDARTGA LDDQGKRAVE FHGKAGRNAL EAEWQEAREM ALLFGGEEAK SRLGPGFDHL SHSKQNEELL RILSRAQSRL ARFHRWSCRI HEKPEATGDD VIDYGQVDEL LTKTAEAMLE NLKALYTNAG GILDSKSKQP LTLVGLRKKL EAQKVEPEKI AAVLKPHAEI IFQRLGTLIP ELKQHLRVSL ERLANRELPL RHREWVWNEA FEKLEQGNFK KEENPKWIRG QRGLSMARIE QIENLRKRFM SLRRQMSLIP GEQVKQGVED KGQRQPEPCE DILNKLDRMK QQRVNQTAHL ILAQALGLRL RPHLANDAER EEKDIHGEYE LIPGRKPVDF IVMEDLSRYL SSQGRAPSEN GRLMKWCHRA VLAKLKQMCE PFGIPVLEVP AAYSSRFCAL TGVPGFRAVE VHDGNAEDFR WKRLIKKAEK DKSSKDAEAA AMLFDQLHDL NIEAREARKQ DKKLPLRTLF APVAGGPLFI PMVGGGPRQA DMNAAINLGL RAIASPTCLR ARPKIRAELK DGKHQAMLGN KLEKAAALTL EPPKEPTKEL AAQKRTNFFL DEKFVGKFDT AHVTTSGKKL RLSGGMSLWK AIKDGAWQRV KKINDARIAK WKNNPPPEPD PDDEIQF	(SEQ ID NO: 183)
Alicyclobacillus kakegawensis WP_067936067.1	MAVKSIKVKL RLSECPDILA GMWQLHRATN AGVRYYTEWV SLMRQEILYS RGPDGGQQCY MTAEDCQREL LRRLRNRQLH NGRQDQPGTD ADLLAISRRL YEILVLQSIG KRGDAQQIAS SFLSPLVDPN SKGGRGEAKS GRKPAWQKMR DQGDPRWVAA REKYEQRKAV DPSKEILNSL DALGLRPLFA VFTETYRSGV DWKPLGKSQG VRTWDRDMFQ QALERLMSWE SWNRRVGEEY ARLFQQKMKF EQEHFAEQSH LVKLARALEA DMRAASQGFE AKRGTAHQIT RRALRGADRV FEIWKSIPEE ALFSQYDEVI RQVQAEKRRD FGSHDLFAKL AEPKYQPLWR ADETFLTRYA LYNGVLRDLE KARQFATFTL PDACVNPIWT RFESSQGSNL	(SEQ ID NO: 184)

Table 6 - Cas12b (C2c1) orthologs		
	HKYEFLFDHL GPGRHAVRFQ RLLVVESEGA KERDSVVVPV APSGQLDKLV LREEEKSSVA LHLHDTARPD GFMAEWAGAK LQYERSTLAR KARRDKQGMR SWRRQPSMLM SAAQMLEDAK QAGDVYLNIS VRVKSPSEVR GQRRPPYAAL FRIDDKQRRV TVNYNKLSAY LEEHPDKQIP GAPGLLSGLR VMSVDLGLRT SASISVFRVA KKEEVEALGD GRPPHYYPIH GTDDLVAVHE RSHLIQMPGE TETKQLRKLR EERQAVLRPL FAQLALLRLL VRCGAADERI RTRSWQRLTK QGREFTKRLT PSWREALELE LTRLEAYCGR VPDDEWSRIV DRTVIALWRR MGKQVRDWRK QVKSGAKVKV KGYQLDVVGG NSLAQIDYLE QQYKFLRRWS FFARASGLVV RADRESHFAV ALRQHIENAK RDRLKKLADR ILMEALGYVY EASGPREGQW TAQHPPCQLI ILEELSAYRF SDDRPPSENS KLMAWGHRGI LEELVNQAQV HDVLVGTVYA AFSSRFDART GAPGVRCRRV PARFVGATVD DSLPLWLTEF LDKHRLDKNL LRPDDVIPTG EGEFLVSPCG EEAARVRQVH ADINAAQNLQ RRLWQNFDIT ELRLRCDVKM GGEGTVLVPR VNNARAKQLF GKKVLVSQDG VTFFERSQTG GKPHSEKQTD LTDKELELIA EADEARAKSV VLFRDPSGHI GKGHWIRQRE FWSLVKQRIE SHTAERIRVR GVGSSLD	
Bacillus sp._V3-13 WP_101661451.1	MAIRSIKLKM KTNSGTDSIY LRKALWRTHQ LINEGIAYYM NLLTLYRQEA IGDKTKEAYQ AELINIIRNQ QRNNGSSEEH GSDQEILALL RQLYELIIPS SIGESGDANQ LGNKFLYPLV DPNSQSGKGT SNAGRKPRWK RLKEEGNPDW ELEKKKDEER KAKDPTVKIF DNLNKYGLLP LFPLFTNIQK DIEWLPLGKR QSVRKWDKDM FIQAIERLLS WESWNRRVAD EYKQLKEKTE SYYKEHLTGG EEWIEKIRKF EKERNMELEK NAFAPNDGYF ITSRQIRGWD RVYEKWSKLP ESASPEELWK VVAEQQNKMS EGFGDPKVFS FLANRENRDI WRGHSERIYH IAAYNGLQKK LSRTKEQATF TLPDAIEHPL WIRYESPGGT NLNLFKLEEK QKKNYYVTLS KIIWPSEEKW IEKENIEIPL APSIQFNRQI KLKQHVKGKQ EISFSDYSSR ISLDGVLGGS RIQFNRKYIK NHKELLGEGD IGPVFFNLVV DVAPLQETRN GRLQSPIGKA LKVISSDFSK VIDYKPKELM DWMNTGSASN SFGVASLLEG MRVMSIDMGQ RTSASVSIFE VVKELPKDQE QKLFYSINDT ELFAIHKRSF LLNLPGEVVT KNNKQQRQER RKKRQFVRSQ IRMLANVLRL ETKKTPDERK KAIHKLMEIV QSYDSWTASQ KEVWEKELNL LTNMAAFNDE IWKESLVELH HRIEPYVGQI VSKWRKGLSE GRKNLAGISM WNIDELEDTR RLLISWSKRS RTPGEANRIE TDEPFGSSLL QHIQNVKDDR LKQMANLIIM TALGFKYDKE EKDRYKRWKE TYPACQIILF ENLNRYLFNL DRSRRENSRL MKWAHRSIPR TVSMQGEMFG LQVGDVRSEY SSRFHAKTGA PGIRCHALTE EDLKAGSNTL KRLIEDGFIN ESELAYLKKG DIIPSQGGEL FVTLSKRYKK DSDNNELTVI HADINAAQNL QKRFWQQNSE VYRVPCQLAR MGEDKLYIPK SQTETIKKYF GKGSFVKNNT EQEVYKWEKS EKMKIKTDTT FDLQDLDGFE DISKTIELAQ EQQKKYLTMF RDPSGYFFNN ETWRPQKEYW SIVNNIIKSC LKKKILSNKV EL	(SEQ ID NO: 185)
Desulfatirhabdium butyrativorans WP_028326052.1	MPLSNNPPVT QRAYTLRLRG ADPSDLSWRE ALWHTHEAVN KGAKVFGDWL LTLRGGLDHT LADTKVKGGK GKPDRDPTPE ERKARRILLA LSWLSVESKL GAPSSYIVAS GDEPAKDRND NVVSALEEIL QSRKVAKSEI DDWKRDCSAS LSAAIRDDAV WVNRSKVFDE AVKSVGSSLT REEAWDMLER FFGSRDAYLT PMKDPEDKSS ETEQEDKAKD LVQKAGQWLS SRYGTSEGAD FCRMSDIYGK IAAWADNASQ GGSSTVDDLV SELRQHFDTK ESKATNGLDW IIGLSSYTGH TPNPVHELLR QNTSLNKSHL DDLKKKANTR AESCKSKIGS KGQRPYSDAI LNDVESVCGF TYRVDKDGQP VSVADYSKYD VDYKWGTARH YIFAVMLDHA ARRISLAHKW IKRAEAERHK FEEDAKRIAN VPARAREWLD SFCKERSVTS GAVEPYRIRR RAVDGWKEVV AAWSKSDCKS TEDRIAAARA LQDDSEIDKF GDIQLFEALA EDDALCVWHK DGEATNEPDF QPLIDYSLAI EAEFKKRQFK VPAYRHPDEL LHPVFCDFGK SRWKINYDVH KNVQAPFYRG LCLTLWTGSE IKPVPLCWQS KRLTRDLALG NNHRNDAASA VTRADRLGRA ASNVTKSDMV NITGLFEQAD WNGRLQAPRQ QLEAIAVVRD NPRLSEQERN LRMCGMIEHI RWLVTFSVKL QPQGPWCAYA EQHGLNTNPQ YWPHADTNRD RKVHARLILP RLPGLRVLSV DLGHRYAAAC AVWEAVNTET VKEACQNVGR DMPKEHDLYL HIKVKKQGIG KQTEVDKTTI YRRIGADTLP DGRPHPAPWA RLDRQFLIKL QGEEKDAREA SNEEIWALHQ MECKLDRTKP LIDRLIASGW GLLKRQMARL DALKELGWIP APDSSENLSR EDGEAKDYRE SLAVDDLMFS AVRTLRLALQ RHGNRARIAY YLISEVKIRP GGIQEKLDEN GRIDLLQDAL ALWHELFSSP GWRDEAAKQL WDSRIATLAG YKAPEENGDN VSDVAYRKKQ QVYREQLRNV AKTLSGDVIT CKELSDAWKE RWEDEDQRWK KLLRWFKDWV LPSGTQANNA	(SEQ ID NO: 186)

Table 6 - Cas12b (C2c1) orthologs		
	TIRNVGGLSL SRLATITEFR RKVQVGFFTR LRPDGTRHEI GEQFGQKTLD ALELLREQRV KQLASRIAEA ALGIGSEGGK GWDGGKRPRQ RINDSRFAPC HAVVIENLAN YRPDETRTRL ENRRLMTWSA SKVHKYLSEA CQLNGLYLCT VSAWYTSRQD SRTGAPGIRC QDVSVREFMQ SPFWRKQVKQ AEAKHDENKG DARERFLCEL NKTWKAKTPA EWKKAGFVRI PLRGGEIFVS ADSKSPSAKG IHADLNAAAN IGLRALTDPD WPGKWWYVPC DPVSFESKMD YVKGCAAVKV GQPLRQPAQT NADGAASKIR KGKKNRTAGT SKEKVYLWRD ISAFPLESNE IGEWKETSAY QNDVQYRVIR MLKEHIKSLD NRTGDNVEG	
Desulfonatronum thiodismutans WP_031386437.1	MVLGRKDDTA ELRRALWTTH EHVNLAVAEV ERVLLRCRGR SYWTLDRRGD PVHVPESQVA EDALAMAREA QRRNGWPVVG EDEEILLALR YLYEQIVPSC LLDDLGKPLK GDAQKIGTNY AGPLFDSDTC RRDEGKDVAC CGPFHEVAGK YLGALPEWAT PISKQEFDGK DASHLRFKAT GGDDAFFRVS IEKANAWYED PANQDALKNK AYNKDDWKKE KDKGISSWAV KYIQKQLQLG QDPRTEVRRK LWLELGLLPL FIPVFDKTMV GNLWNRLAVR LALAHLLSWE SWNHRAVQDQ ALARAKRDEL AALFLGMEDG FAGLREYELR RNESIKQHAF EPVDRPYVVS GRALRSWTRV REEWLRHGDT QESRKNICNR LQDRLRGKFG DPDVFHWLAE DGQEALWKER DCVTSFSLLN DADGLLEKRK GYALMTFADA RLHPRWAMYE APGGSNLRTY QIRKTENGLW ADVVLLSPRN ESAAVEEKTF NVRLAPSGQL SNVSFDQIQK GSKMVGRCRY QSANQQFEGL LGGAEILFDR KRIANEQHGA TDLASKPGHV WFKLTLDVRP QAPQGWLDGK GRPALPPEAK HFKTALSNKS KFADQVRPGL RVLSVDLGVR SFAACSVFEL VRGGPDQGTY FPAADGRTVD DPEKLWAKHE RSFKITLPGE NPSRKEEIAR RAAMEELRSL NGDIRRLKAI LRLSVLQEDD PRTEHLRLFM EAIVDDPAKS ALNAELFKGF GDDRFRSTPD LWKQHCHFFH DKAEKVVAER FSRWRTETRP KSSSWQDWRE RRGYAGGKSY WAVTYLEAVR GLILRWNMRG RTYGEVNRQD KKQFGTVASA LLHHINQLKE DRIKTGADMI IQAARGFVPR KNGAGWVQVH EPCRLILFED LARYRFRTDR SRRENSRLMR WSHREIVNEV GMQGELYGLH VDTTEAGFSS RYLASSGAPG VRCRHLVEED FHDGLPGMHL VGELDWLLPK DKDRTANEAR RLLGGMVRPG MLVPWDGGEL FATLNAASQL HVIHADINAA QNLQRRFWGR CGEAIRIVCN QLSVDGSTRY EMAKAPKARL LGALQQLKNG DAPFHLTSIP NSQKPENSYV MTPTNAGKKY RAGPGEKSSG EEDELALDIV EQAEELAQGR KTFFRDPSGV FFAPDRWLPS EIYWSRIRRR IWQVTLERNS SGRQERAEMD EMPY	(SEQ ID NO: 187)
Lentisphaeria bacterium DCFZ01000012.1	MAVELNRIYQ GRVNHVYIFD ENQNQVSVDN GDDLLFVHHE LYQDAINYYL VALAAMALDS KDSLFGKFKM QIRAVWNDFY RNGQLRPGLK HSLIRSLGHA AELNTSNGAD IAMNLILEDG GIPSEILNAA LEHLAEKCTG DVSQLGKTFF PRFCDTAYHG NWDVDAKSFS EKKGRQRLVD ALYSLHPVQA VQELAPEIEI GWGGVKTQTG KFFTGDEAKA SLKKAISYFL QDTGKNSPEL QEYFSVAGKQ PLEQYLGKID TFPEISFGRI SSHQNINISN AMWILKFFPD QYSVDLIKNL IPNKKYEIGI APQWGDDPVK LSRGKRGYTF RAFTDLAMWE KNWKVFDRAA FSDALKTINQ FRNKTQERND QLKRYCAALN WMDGESSDKK PPVEPADADA VDEAATSVLP ILAGDKRWNA LLQLQKELGI CNDFTENELM DYGLSLRTIR GYQKLRSMML EKEEKMRAKT ADDEEISQAL QEIIIKFQSS HRDTIGSVSL FLKLAEPKYF CVWHDADKNQ NFASVDMVAD AVRYYSYQEE KARLEEPIQI TPADARYSRR VSDLYALVYK NAKECKTGYG LRPDGNFVFE IAQKNAKGYA PAKVVLAFSA PRLKRDGLID KEFSAYYPPV LQAFLREEEA PKQSFKTTAV ILMPDWDKNG KRRILLNFPI KLDVSAIHQK TDHRFENQFY FANNTNTCLL WPSYQYKKPV TWYQGKKPFD VVAVDLGQRS AGAVSRITVS TEKREHSVAI GEAGGTQWYA YRKFSGLLRL PGEDATVIRD GQRTEELSGN AGRLSTEEET VQACVLCKML IGDATLLGGS DEKTIRSFPK QNDKLLIAFR RATGRMKQLQ RWLWMLNENG LCDKAKTEIS NSDWLVNKNI DNVLKEEKQH REMLPAILLQ IADRVLPLRG RKWDWVLNPQ SNSFVLQQTA HGSGDPHKKI CGQRGLSFAR IEQLESLRMR CQALNRILMR KTGEKPATLA EMRNNPIPDC CPDILMRLDA MKEQRINQTA NLILAQALGL RHCLHSESAT KRKENGMHGE YEKIPGVEPA AFVVLEDLSR YRFSQDRSSY ENSRLMKWSH RKILEKLALL CEVFNVPILQ VGAAYSSKFS ANAIPGFRAE ECSIDQLSFY PWRELKDSRE KALVEQIRKI GHRLLTFDAK ATIIMPRNGG PVFIPFVPSD SKDTLIQADI NASFNIGLRG VADATNLLCN NRVSCDRKKD CWQVKRSSNF SKMVYPEKLS LSFDPIKKQE GAGGNFFVLG CSERILTGTS EKSPVFTSSE MAKKYPNLMF GSALWRNEIL KLERCCKINQ SRLDKFIAKK EVQNEL	(SEQ ID NO: 188)

Table 6 - Cas12b (C2c1) orthologs		
Laceyella sediminis WP_106341859.1	MSIRSFKLKI KTKSGVNAEE LRRGLWRTHQ LINDGIAYYM NWLVLLRQED LFIRNEETNE IEKRSKEEIQ GELLERVHKQ QQRNQWSGEV DDQTLLQTLR HLYEEIVPSV IGKSGNASLK ARFFLGPLVD PNNKTTKDVS KSGPTPKWKK MKDAGDPNWV QEYEKYMAER QTLVRLEEMG LIPLFPMYTD EVGDIHWLPQ ASGYTRTWDR DMFQQAIERL LSWESWNRRV RERRAQFEKK THDFASRFSE SDVQWMNKLR EYEAQQEKSL EENAFAPNEP YALTKKALRG WERVYHSWMR LDSAASEEAY WQEVATCQTA MRGEFGDPAI YQFLAQKENH DIWRGYPERV IDFAELNHLQ RELRRAKEDA TFTLPDSVDH PLWVRYEAPG GTNIHGYDLV QDTKRNLTLI LDKFILPDEN GSWHEVKKVP FSLAKSKQFH RQVWLQEEQK QKKREVVFYD YSTNLPHLGT LAGAKLQWDR NFLNKRTQQQ IEETGEIGKV FFNISVDVRP AVEVKNGRLQ NGLGKALTVL THPDGTKIVT GWKAEQLEKW VGESGRVSSL GLDSLSEGLR VMSIDLGQRT SATVSVFEIT KEAPDNPYKF FYQLEGTELF AVHQRSFLLA LPGENPPQKI KQMREIRWKE RNRIKQQVDQ LSAILRLHKK VNEDERIQAI DKLLQKVASW QLNEEIATAW NQALSQLYSK AKENDLQWNQ AIKNAHHQLE PVVGKQISLW RKDLSTGRQG IAGLSLWSIE ELEATKKLLT RWSKRSREPG VVKRIERFET FAKQIQHHIN QVKENRLKQL ANLIVMTALG YKYDQEQKKW IEVYPACQVV LFENLRSYRF SYERSRRENK KLMEWSHRSI PKLVQMQGEL FGLQVADVYA AYSSRYHGRT GAPGIRCHAL TEADLRNETN IIHELIEAGF IKEEHRPYLQ QGDLVPWSGG ELFATLQKPY DNPRILTLHA DINAAQNIQK RFWHPSMWFR VNCESVMEGE IVTYVPKNKT VHKKQGKTFR FVKVEGSDVY EWAKWSKNRN KNTFSSITER KPPSSMILFR DPSGTFFKEQ EWVEQKTFWG KVQSMIQAYM KKTIVQRMEE	(SEQ ID NO: 189)
Methylobacterium nodulans (long form)	MYEAIVLADD ANAQLANAFL GPLTDPNSAG FLEAFNKVDR PAPSWLDQVP ASDPIDPAVL AEANAWLDTD AGRAWLVDTG APPRWRSLAA KQDPIWPREF ARKLGELRKE AASGTSAIIK ALKRDFGVLP LFQPSLAPRI LGSRSSLTPW DRLAFRLAVG HLLSWESWCT RARDEHTARV QRLEQFSSAH LKGDLATKVS TLREYERARK EQIAQLGLPM GERDFLITVR MTRGWDDLRE KWRRSGDKGQ EALHAIIATE QTRKRGRFGD PDLFRWLARP ENHHVWADGH ADAVGVLARV NAMERLVERS RDTALMTLPD PVAHPRSAQW EAEGGSNLRN YQLEAVGGEL QITLPLLKAA DDGRCIDTPL SFSLAPSDQL QGVVLTKQDK QQKITYCTNM NEVFEAKLGS ADLLLNWDHL RGRIRDRVDA GDIGSAFLKL ALDVAHVLPD GVDDQLARAA FHFQSAKGAK SKHADSVQAG LRVLSIDLGV RSFATCSVFE LKDTAPTTGV AFPLAEFRLW AVHERSFTLE LPGENVGAAG QQWRAQADAE LRQLRGGLNR HRQLLRAATV QKGERDAYLT DLREAWSAKE LWPFEASLLS ELERCSTVAD PLWQDTCKRA ARLYRTEFGA VVSEWRSRTR SREDRKYAGK SMWSVQHLTD VRRFLQSWSL AGRASGDIRR LDRERGGVFA KDLLDHIDAL KDDRLKTGAD LIVQAARGFQ RNEFGYWVQK HAPCHVILFE DLSRYRMRTD RPRRENSQLM QWAHRGVPDM VGMQGEIYGI QDRRDPDSAR KHARQPLAAF CLDTPAAFSS RYHASTMTPG IRCHPLRKRE FEDQGFLELL KRENEGLDLN GYKPGDLVPL PGGEVFVCLN ANGLSRIHAD INAAQNLQRR FWTQHGDAFR LPCGKSAVQG QIRWAPLSMG KRQAGALGGF GYLEPTGHDS GSCQWRKTTE AEWRRLSGAQ KDRDEAAAAE DEELQGLEEE LLERSGERVV FFRDPSGVVL PTDLWFPSAA FWSIVRAKTV GRLRSHLDAQ AEASYAVAAG L	(SEQ ID NO: 190)
Opitutaceae bacterium WP_009513281.1	MSLNRIYQGR VAAVETGTAL AKGNVEWMPA AGGDEVLWQH HELFQAAINY YLVALLALAD KNNPVLGPLI SQMDNPQSPY HVWGSFRRQG RQRTGLSQAV APYITPGNNA PTLDEVFRSI LAGNPTDRAT LDAALMQLLK ACDGAGAIQQ EGRSYWPKFC DPDSTANFAG DPAMLRREQH RLLLPQVLHD PAITHDSPAL GSFDTYSIAT PDTRTPQLTG PKARARLEQA ITLWRVRLPE SAADFDRLAS SLKKIPDDDS RLNLQGYVGS SAKGEVQARL FALLLFRHLE RSSFTLGLLR SATPPPKNAE TPPPAGVPLP AASAADPVRI ARGKRSFVFR AFTSLPCWHG GDNIHPTWKS FDIAAFKYAL TVINQIEEKT KERQKECAEL ETDFDYMHGR LAKIPVKYTT GEAEPPPILA NDLRIPLLRE LLQNIKVDTA LTDGEAVSYG LQRRTIRGFR ELRRIWRGHA PAGTVFSSEL KEKLAGELRQ FQTDNSTTIG SVQLFNELIQ NPKYWPIWQA PDVETARQWA DAGFADDPLA ALVQEAELQE DIDALKAPVK LTPADPEYSR RQYDFNAVSK FGAGSRSANR HEPGQTERGH NTFTTEIAAR NAADGNRWRA THVRIHYSAP RLLRDGLRRP DTDGNEALEA VPWLQPMMEA LAPLPTLPQD LTGMPVFLMP DVTLSGERRI LLNLPVTLEP AALVEQLGNA GRWQNQFFGS REDPFALRWP ADGAVKTAKG KTHIPWHQDR DHFTVLGVDL GTRDAGALAL LNVTAQKPAK PVHRIIGEAD GRTWYASLAD ARMIRLPGED ARLFVRGKLV QEPYGERGRN ASLLEWEDAR NIILRLGQNP	(SEQ ID NO: 191)

Table 6 - Cas12b (C2c1) orthologs		
	DELLGADPRR HSYPEINDKL LVALRRAQAR LARLQNRSWR LRDLAESDKA LDEIHAERAG EKPSPLPPLA RDDAIKSTDE ALLSQRDIIR RSFVQIANLI LPLRGRRWEW RPHVEVPDCH ILAQSDPGTD DTKRLVAGQR GISHERIEQI EELRRRCQSL NRALRHKPGE RPVLGRPAKG EEIADPCPAL LEKINRLRDQ RVDQTAHAIL AAALGVRLRA PSKDRAERRH RDIHGEYERF RAPADFVVIE NLSRYLSSQD RARSENTRLM QWCHRQIVQK LRQLCETYGI PVLAVPAAYS SRFSSRDGSA GFRAVHLTPD HRHRMPWSRI LARLKAHEED GKRLEKTVLD EARAVRGLFD RLDRFNAGHV PGKPWRTLLA PLPGGPVFVP LGDATPMQAD LNAAINIALR GIAAPDRHDI HHRLRAENKK RILSLRLGTQ REKARWPGGA PAVTLSTPNN GASPEDSDAL PERVSNLFVD IAGVANFERV TIEGVSQKFA TGRGLWASVK QRAWNRVARL NETVTDNNRN EEEDDIPM	
Thermomonas hydrothermalis WP_072754838.1	MSEKTTQRAY TLRLNRASGE CAVCQNNSCD CWHDALWATH KAVNRGAKAF GDWLLTLRGG LCHTLVEMEV PAKGNNPPQR PTDQERRDRR VLLALSWLSV EDEHGAPKEF IVATGRDSAD DRAKKVEEKL REILEKRDFQ EHEIDAWLQD CGPSLKAHIR EDAVWVNRRA LFDAAVERIK TLTWEEAWDF LEPFFGTQYF AGIGDGKDKD DAEGPARQGE KAKDLVQKAG QWLSARFGIG TGADFMSMAE AYEKIAKWAS QAQNGDNGKA TIEKLACALR PSEPPTLDTV LKCISGPGHK SATREYLKTL DKKSTVTQED LNQLRKLADE DARNCRKKVG KKGKKPWADE VLKDVENSCE LTYLQDNSPA RHREFSVMLD HAARRVSMAH SWIKKAEQRR RQFESDAQKL KNLQERAPSA VEWLDRFCES RSMTTGANTG SGYRIRKRAI EGWSYVVQAW AEASCDTEDK RIAAARKVQA DPEIEKFGDI QLFEALAADE AICVWRDQEG TQNPSILIDY VTGKTAEHNQ KRFKVPAYRH PDELRHPVFC DFGNSRWSIQ FAIHKEIRDR DKGAKQDTRQ LQNRHGLKMR LWNGRSMTDV NLHWSSKRLT ADLALDQNPN PNPTEVTRAD RLGRAASSAF DHVKIKNVFN EKEWNGRLQA PRAELDRIAK LEEQGKTEQA EKLRKRLRWY VSFSPCLSPS GPFIVYAGQH NIQPKRSGQY APHAQANKGR ARLAQLILSR LPDLRILSVD LGHRFAAACA VWETLSSDAF RREIQGLNVL AGGSGEGDLF LHVEMTGDDG KRRTVVYRRI GPDQLLDNTP HPAPWARLDR QFLIKLQGED EGVREASNEE LWTVHKLEVE VGRTVPLIDR MVRSGFGKTE KQKERLKKLR ELGWISAMPN EPSAETDEKE GEIRSISRSV DELMSSALGT LRLALKRHGN RARIAFAMTA DYKPMPGGQK YYFHEAKEAS KNDDETKRRD NQIEFLQDAL SLWHDLFSSP DWEDNEAKKL WQNHIATLPN YQTPEEISAE LKRVERNKKR KENRDKLRTA AKALAENDQL RQHLHDTWKE RWESDDQQWK ERLRSLKDWI FPRGKAEDNP SIRHVGGLSI TRINTISGLY QILKAFKMRP EPDDLRKNIP QKGDDELENF NRRLLEARDR LREQRVKQLA SRIIEAALGV GRIKIPKNGK LPKRPRTTVD TPCHAVVIES LKTYRPDDLR TRRENRQLMQ WSSAKVRKYL KEGCELYGLH FLEVPANYTS RQCSRTGLPG IRCDDVPTGD FLKAPWWRRA INTAREKNGG DAKDRFLVDL YDHLNNLQSK GEALPATVRV PRQGGNLFIA GAQLDDTNKE RRAIQADLNA AANIGLRALL DPDWRGRWWY VPCKDGTSEP ALDRIEGSTA FNDVRSLPTG DNSSRRAPRE IENLWRDPSG DSLESGTWSP TRAYWDTVQS RVIELLRRHA GLPTS	(SEQ ID NO: 192)
Methylobacterium nodulans WP_043747912.1	MYEAIVLADD ANAQLANAFL GPLTDPNSAG FLEAFNKVDR PAPSWLDQVP ASDPIDPAVL AEANAWLDTD AGRAWLVDTG APPRWRSLAA KQDPIWPREF ARKLGELRKE AASGTSAIIK ALKRDFGVLP LFQPSLAPRI LGSRSSLTPW DRLAFRLAVG HLLSWESWCT RARDEHTARV QRLEQFSSAH LKGDLATKVS TLREYERARK EQIAQLGLPM GERDFLITVR MTRGWDDLRE KWRRSGDKGQ EALHAIIATE QTRKRGRFGD PDLFRWLARP ENHHVWADGH ADAVGVLARV NAMERLVERS RDTALMTLPD PVAHPRSAQW EAEGGSNLRN YQLEAVGGEL QITLPLLKAA DDGRCIDTPL	(SEQ ID NO: 193)
Chloracidobacterium thermophilum WP_058868187.1	MPQQAKPPVT QRAYTLRLRG ADSNDPSWRD ALWQTHEAVN RGAQAFGDWL LTLRGGLDHT LADTPVKGGK GKPDPDPTDE ERKARRILLA LSWLSVESKL GAPAGLIIAF GTEAAEERNR KVVAALEEIL KSRGVDQNEI NAWKKDCSAS LSAAIRDDAV WVNRSKAFDE AVESIGSSGS SGSSLTREEP WDMLERFFGS RDAYLAPAKG SEDESSEAKQ EDQAKDLVQK AGQWLSSRFG TGKGADFRRM ATVYEAIAKW DGKASLEMAG DKAIADLATA LSEFNPASND LQGVLGLISG PGYKSATRNF LNQLAAQTTV TQQDFVSLKD KANNDAQECK QNTGSKGQRP YSNSILEKVE SVCGFTYLQD GGPARHSEFA VILDHAARRV SLAHTWIKLA EAERRKFEED AKKIDQVPEA AKDWLDRFCL ERSGVSGALE PYRIRRRAVD GWKEVVAEWS KSDCKTVEDR IAAARALQDD PEIDKFGDIQ LFEALAEDDA VCVWHKDGDA AKAPDPQPLI DYALAAEAEF KKRHFKVPAY RHPDALLHPI	(SEQ ID NO: 194)

Table 6 - Cas12b (C2c1) orthologs

	FCDFGKSRWD ICFDVHKNMQ TPFPRALCLT LWTGSEMKRI PLCWQSKRLA RDLALGNNTG DAGASEVTRA DRLGRAASRA ASNVTKSDVV NIAGLFEQAD WNGRLQAPRQ QLEAIARYVE KHDWDQKAEK MRNAIQWLVT FSARLQPQGP WCAYAKIHGL KEDPQYWPHA DTNKNRKGHA RLILSRLPGL RVLAVDLGHR YAAACAVWEA LSTEAFQREI KGRTILRGRT DGNALYCHTR HKANGKERVT IYRRIGADTL PDGKPHPAPW ARLDRQFLIK LQGEEEGVRE ASNEEIWAVH QLEAALGRPV SLIDRLVASG WGGSDKQKAR LEGLKQLGWD PADKPSLSVD ELMSSAVRTM RLALKRHGDR ARIAHYLITD EKTTPGGIKE TLDEKGRIDL LQDALVLWHD LFSSRGWRDD TAKQLWNAHV AKLHGYKAPE EPGEDSSGAE RKKKQRENRE KLYDVAKALA QDVTLREALH DAWKKRWEND DERWKKQLRW FKDWVFPRGN HASDPTIRKR QLINPSGGNG RRGNHASDPT IRKRQLINPS GGNGRRGNHA SDPTIRKVGG LSLPRLATLT EFRRKVQVGF FTRLKPDGTR AETKEQFGQS ALDALEHLRE QRVKQLASRI AEAALGVGRV RRPVEGKDPK RPDVRVDEPC HAIVIEDLTH YRPEETRTRR ENRQLMTWSS SKVKKYLAEA CQLHGLHLRE VSASYTSRQD SRTGAPGVRC QDVPVKEFMR SPFWRKQVKQ AEAKQAANKG DARERLLCDL NARWKDRTAA DWEKAGAVRI PLQGGEIFVS ADANSPAAKG IQADLNAAAN IGLRALTDPD WAGKWWYVPC DPASFRPVRD KVDGSAVVNP DQPLRQSAQA QSGDAAKDKN GNKGAGKSKE VVNLWRDISS SPLECIEFGE WKEYAAYQNE VQCRVIRILK EQIKGRDKQP HEGSKEDDIP L	
Desulfovibrio inopinatus WP_027186183.1	MPTRTINLKL VLGKNPENAT LRRALFSTHR LVNQATKRIE EFLLLCRGEA YRTVDNEGKE AEIPRHAVQE EALAFAKAAQ RHNGCISTYE DQEILDVLRQ LYERLVPSVN ENNEAGDAQA ANAWVSPLMS AESEGGLSVY DKVLDPPPVW MKLKEEKAPG WEAASQIWIQ SDEGQSLLNK PGSPPRWIRK LRSGQPWQDD FVSDQKKKQD ELTKGNAPLI KQLKEMGLLP LVNPFFRHLL DPEGKGVSPW DRLAVRAAVA HFISWESWNH RTRAEYNSLK LRRDEFEAAS DEFKDDFTLL RQYEAKRHST LKSIALADDS NPYRIGVRSL RAWNRVREEW IDKGATEEQR VTILSKLQTQ LRGKFGDPDL FNWLAQDRHV HLWSPRDSVT PLVRINAVDK VLRRRKPYAL MTFAHPRFHP RWILYEAPGG SNLRQYALDC TENALHITLP LLVDDAHGTW IEKKIRVPLA PSGQIQDLTL EKLEKKKNRL YYRSGFQQFA GLAGGAEVLF HRPYMEHDER SEESLLERPG AVWFKLTLDV ATQAPPNWLD GKGRVRTPPE VHHFKTALSN KSKHTRTLQP GLRVLSVDLG MRTFASCSVF ELIEGKPETG RAFPVADERS MDSPNKLWAK HERSFKLTLP GETPSRKEEE ERSIARAEIY ALKRDIQRLK SLLRLGEEDN DNRRDALLEQ FFKGWGEEDV VPGQAFPRSL FQGLGAAPFR STPELWRQHC QTYYDKAEAC LAKHISDWRK RTRPRPTSRE MWYKTRSYHG GKSIWMLEYL DAVRKLLLSW SLRGRTYGAI NRQDTARFGS LASRLLHHIN SLKEDRIKTG ADSIVQAARG YIPLPHGKGW EQRYEPCQLI LFEDLARYRF RVDRPRRENS QLMQWNHRAI VAETTMQAEL YGQIVENTAA GFSSRFHAAT GAPGVRCRFL LERDFDNDLP KPYLLRELSW MLGNTKVESE EEKLRLLSEK IRPGSLVPWD GGEQFATLHP KRQTLCVIHA DMNAAQNLQR RFFGRCGEAF RLVCQPHGDD VLRLASTPGA RLLGALQQLE NGQGAFELVR DMGSTSQMNR FVMKSLGKKK IKPLQDNNGD DELEDVLSVL PEEDDTGRIT VFRDSSGIFF PCNVWIPAKQ FWPAVRAMIW KVMASHSLG	(SEQ ID NO: 195)
Desulfonatronum thiodismutans WP_031386437.1	MVLGRKDDTA ELRRALWTTH EHVNLAVAEV ERVLLRCRGR SYWTLDRRGD PVHVPESQVA EDALAMAREA QRRNGWPVVG EDEEILLALR YLYEQIVPSC LLDDLGKPLK GDAQKIGTNY AGPLFDSDTC RRDEGKDVAC CGPFHEVAGK YLGALPEWAT PISKQEFDGK DASHLRFKAT GGDDAFFRVS IEKANAWYED PANQDALKNK AYNKDDWKKE KDKGISSWAV KYIQKQLQLG QDPRTEVRRK LWLELGLLPL FIPVFDKTMV GNLWNRLAVR LALAHLLSWE SWNHRAVQDQ ALARAKRDEL AALFLGMEDG FAGLREYELR RNESIKQHAF EPVDRPYVVS GRALRSWTRV REEWLRHGDT QESRKNICNR LQDRLRGKFG DPDVFHWLAE DGQEALWKER DCVTSFSLLN DADGLLEKRK GYALMTFADA RLHPRWAMYE APGGSNLRTY QIRKTENGLW ADVVLLSPRN ESAAVEEKTF NVRLAPSGQL SNVSFDQIQK GSKMVGRCRY QSANQQFEGL LGGAEILFDR KRIANEQHGA TDLASKPGHV WFKLTLDVRP QAPQGWLDGK GRPALPPEAK HFKTALSNKS KFADQVRPGL RVLSVDLGVR SFAACSVFEL VRGGPDQGTY FPAADGRTVD DPEKLWAKHE RSFKITLPGE NPSRKEEIAR RAAMEELRSL NGDIRRLKAI LRLSVLQEDD PRTEHLRLFM EAIVDDPAKS ALNAELFKGF GDDRFRSTPD LWKQHCHFFH DKAEKVVAER FSRWRTETRP KSSSWQDWRE RRGYAGGKSY WAVTYLEAVR GLILRWNMRG RTYGEVNRQD KKQFGTVASA LLHHINQLKE	(SEQ ID NO: 187)

Table 6 - Cas12b (C2c1) orthologs		
	DRIKTGADMI IQAARGFVPR KNGAGWVQVH EPCRLILFED LARYRFRTDR SRRENSRLMR WSHREIVNEV GMQGELYGLH VDTTEAGFSS RYLASSGAPG VRCRHLVEED FHDGLPGMHL VGELDWLLPK DKDRTANEAR RLLGGMVRPG MLVPWDGGEL FATLNAASQL HVIHADINAA QNLQRRFWGR CGEAIRIVCN QLSVDGSTRY EMAKAPKARL LGALQQLKNG DAPFHLTSIP NSQKPENSYV MTPTNAGKKY RAGPGEKSSG EEDELALDIV EQAEELAQGR KTFFRDPSGV FFAPDRWLPS EIYWSRIRRR IWQVTLERNS SGRQERAEMD EMPY	
Tuberibacillus calidus WP_027726362.1	MATKSFILKM KTKNNPQLRL SLWKTHELFN FGVAYYMDLL SLFRQKDLYM HNDEDPDHPV VLKKEEIQER LWMKVRETQQ KNGFHGEVSK DEVLETLRAL YEELVPSAVG KSGEANQISN KYLYPLTDPA SQSGKGTANS GRKPRWKKLK EAGDPSWKDA YEKWEKERQE DPKLKILAAL QSFGLIPLFR PFTENDHKAV ISVKWMPKSK NQSVRKFDKD MFNQAIERFL SWESWNEKVA EDYEKTVSIY ESLQKELKGI STKAFEIMER VEKAYEAHLR EITFSNSTYR IGNRAIRGWT EIVKKWMKLD PSAPQGNYLD VVKDYQRRHP RESGDFKLFE LLSRPENQAA WREYPEFLPL YVKYRHAEQR MKTAKKQATF TLCDPIRHPL WVRYEERSGT NLNKYRLIMN EKEKVVQFDR LICLNADGHY EEQEDVTVPL APSQQFDDQI KFSSEDTGKG KHNFSYYHKG INYELKGTLG GARIQFDREH LLRRQGVKAG NVGRIFLNVT LNIEPMQPFS RSGNLQTSVG KALKVYVDGY PKVVNFKPKE LTEHIKESEK NTLTLGVESL PTGLRVMSVD LGQRQAAAIS IFEVVSEKPD DNKLFYPVKD TDLFAVHRTS FNIKLPGEKR TERRMLEQQK RDQAIRDLSR KLKFLKNVLN MQKLEKTDER EKRVNRWIKD REREEENPVY VQEFEMISKV LYSPHSVWVD QLKSIHRKLE EQLGKEISKW RQSISQGRQG VYGISLKNIE DIEKTRRLLF RWSMRPENPG EVKQLQPGER FAIDQQNHLN HLKDDRIKKL ANQIVMTALG YRYDGKRKKW IAKHPACQLV LFEDLSRYAF YDERSRLENR NLMRWSRREI PKQVAQIGGL YGLLVGEVGA QYSSRFHAKS GAPGIRCRVV KEHELYITEG GQKVRNQKFL DSLVENNIIE PDDARRLEPG DLIRDQGGDK FATLDERGEL VITHADINAA QNLQKRFWTR THGLYRIRCE SREIKDAVVL VPSDKDQKEK MENLFGIGYL QPFKQENDVY KWVKGEKIKG KKTSSQSDDK ELVSEILQEA SVMADELKGN RKTLFRDPSG YVFPKDRWYT GGRYFGTLEH LLKRKLAERR LFDGGSSRRG LFNGTDSNTN VE	(SEQ ID NO: 196)
Bacillus thermoamylovorans WP_041902512.1	MATRSFILKI EPNEEVKKGL WKTHEVLNHG IAYYMNILKL IRQEAIYEHH EQDPKNPKKV SKAEIQAELW DFVLKMQKCN SFTHEVDKDV VFNILRELYE ELVPSSVEKK GEANQLSNKF LYPLVDPNSQ SGKGTASSGR KPRWYNLKIA GDPSWEEEKK KWEEDKKKDP LAKILGKLAE YGLIPLFIPF TDSNEPIVKE IKWMEKSRNQ SVRRLDKDMF IQALERFLSW ESWNLKVKEE YEKVEKEHKT LEERIKEDIQ AFKSLEQYEK ERQEQLLRDT LNTNEYRLSK RGLRGWREII QKWLKMDENE PSEKYLEVFK DYQRKHPREA GDYSVYEFLS KKENHFIWRN HPEYPYLYAT FCEIDKKKKD AKQQATFTLA DPINHPLWVR FEERSGSNLN KYRILTEQLH TEKLKKKLTV QLDRLIYPTE SGGWEEKGKV DIVLLPSRQF YNQIFLDIEE KGKHAFTYKD ESIKFPLKGT LGGARVQFDR DHLRRYPHKV ESGNVGRIYF NMTVNIEPTE SPVSKSLKIH RDDFPKFVNF KPKELTEWIK DSKGKKLKSG IESLEIGLRV MSIDLGQRQA AAASIFEVVD QKPDIEGKLF FPIKGTELYA VHRASFNIKL PGETLVKSRE VLRKAREDNL KLMNQKLNFL RNVLHFQQFE DITEREKRVT KWISRQENSD VPLVYQDELI QIRELMYKPY KDWVAFLKQL HKRLEVEIGK EVKHWRKSLS DGRKGLYGIS LKNIDEIDRT RKFLLRWSLR PTEPGEVRRL EPGQRFAIDQ LNHLNALKED RLKKMANTII MHALGYCYDV RKKKWQAKNP ACQIILFEDL SNYNPYEERS RFENSKLMKW SRREIPRQVA LQGEIYGLQV GEVGAQFSSR FHAKTGSPGI RCSVVTKEKL QDNRFFKNLQ REGRLTLDKI AVLKEGDLYP DKGGEKFISL SKDRKLVTTH ADINAAQNLQ KRFWTRTHGF YKVYCKAYQV DGQTVYIPES KDQKQKIIEE FGEGYFILKD GVYEWGNAGK LKIKKGSSKQ SSSELVDSDI LKDSFDLASE LKGEKLMLYR DPSGNVFPSD KWMAAGVFFG KLERILISKL TNQYSISTIE DDSSKQSM	(SEQ ID NO: 197)
Bacillus sp. NSP2.1 WP_026557978.1	MAIRSIKLKL KTHTGPEAQN LRKGIWRTHR LLNEGVAYYM KMLLLFRQES TGERPKEELQ EELICHIREQ QQRNQADKNT QALPLDKALE ALRQLYELLV PSSVGQSGDA QIISRKFLSP LVDPNSEGGK GTSKAGAKPT WQKKKEANDP TWEQDYEKWK KRREEDPTAS VITTLEEYGI RPIFPLYTNT VTDIAWLPLQ SNQFVRTWDR DMLQQAIERL LSWESWNKRV QEEYAKLKEK MAQLNEQLEG GQEWISLLEQ YEENRERELR ENMTAANDKY RITKRQMKGW NELYELWSTF PASASHEQYK EALKRVQQRL RGRFGDAHFF QYLMEEKNRL IWKGNPQRIH	(SEQ ID NO: 198)

Table 6 - Cas12b (C2c1) orthologs		
	YFVARNELTK RLEEAKQSAT MTLPNARKHP LWVRFDARGG NLQDYYLTAE ADKPRSRRFV TFSQLIWPSE SGWMEKKDVE VELALSRQFY QQVKLLKNDK GKQKIEFKDK GSGSTFNGHL GGAKLQLERG DLEKEEKNFE DGEIGSVYLN VVIDFEPLQE VKNGRVQAPY GQVLQLIRRP NEFPKVTTYK SEQLVEWIKA SPQHSAGVES LASGFRVMSI DLGLRAAAAT SIFSVEESSD KNAADFSYWI EGTPLVAVHQ RSYMLRLPGE QVEKQVMEKR DERFQLHQRV KFQIRVLAQI MRMANKQYGD RWDELDSLKQ AVEQKKSPLD QTDRTFWEGI VCDLTKVLPR NEADWEQAVV QIHRKAEEYV GKAVQAWRKR FAADERKGIA GLSMWNIEEL EGLRKLLISW SRRTRNPQEV NRFERGHTSH QRLLTHIQNV KEDRLKQLSH AIVMTALGYV YDERKQEWCA EYPACQVILF ENLSQYRSNL DRSTKENSTL MKWAHRSIPK YVHMQAEPYG IQIGDVRAEY SSRFYAKTGT PGIRCKKVRG QDLQGRRFEN LQKRLVNEQF LTEEQVKQLR PGDIVPDDSG ELFMTLTDGS GSKEVVFLQA DINAAHNLQK RFWQRYNELF KVSCRVIVRD EEEYLVPKTK SVQAKLGKGL FVKKSDTAWK DVYVWDSQAK LKGKTTFTEE SESPEQLEDF QEIIEEAEEA KGTYRTLFRD PSGVFFPESV WYPQKDFWGE VKRKLYGKLR ERFLTKAR	
Alicyclobacillus acidoterrestris WP_021296342.1	MAVKSIKVKL RLDDMPEIRA GLWKLHKEVN AGVRYYTEWL SLLRQENLYR RSPNGDGEQE CDKTAEECKA ELLERLRARQ VENGHRGPAG SDDELLQLAR QLYELLVPQA IGAKGDAQQI ARKFLSPLAD KDAVGGLGIA KAGNKPRWVR MREAGEPGWE EEKEKAETRK SADRTADVLR ALADFGLKPL MRVYTDSEMS SVEWKPLRKG QAVRTWDRDM FQQAIERMMS WESWNQRVGQ EYAKLVEQKN RFEQKNFVGQ EHLVHLVNQL QQDMKEASPG LESKEQTAHY VTGRALRGSD KVFEKWGKLA PDAPFDLYDA EIKNVQRRNT RRFGSHDLFA KLAEPEYQAL WREDASFLTR YAVYNSILRK LNHAKMFATF TLPDATAHPI WTRFDKLGGN LHQYTFLFNE FGERRHAIRF HKLLKVENGV AREVDDVTVP ISMSEQLDNL LPRDPNEPIA LYFRDYGAEQ HFTGEFGGAK IQCRRDQLAH MHRRRGARDV YLNVSVRVQS QSEARGERRP PYAAVFRLVG DNHRAFVHFD KLSDYLAEHP DDGKLGSEGL LSGLRVMSVD LGLRTSASIS VFRVARKDEL KPNSKGRVPF FFPIKGNDNL VAVHERSQLL KLPGETESKD LRAIREERQR TLRQLRTQLA YLRLLVRCGS EDVGRRERSW AKLIEQPVDA ANHMTPDWRE AFENELQKLK SLHGICSDKE WMDAVYESVR RVWRHMGKQV RDWRKDVRSG ERPKIRGYAK DVVGGNSIEQ IEYLERQYKF LKSWSFFGKV SGQVIRAEKG SRFAITLREH IDHAKEDRLK KLADRIIMEA LGYVYALDER GKGKWVAKYP PCQLILLEEL SEYQFNNDRP PSENNQLMQW SHRGVFQELI NQAQVHDLLV GTMYAAFSSR FDARTGAPGI RCRRVPARCT QEHNPEPFPW WLNKFVVEHT LDACPLRADD LIPTGEGEIF VSPFSAEEGD FHQIHADLNA AQNLQQRLWS DFDISQIRLR CDWGEVDGEL VLIPRLTGKR TADSYSNKVF YTNTGVTYYE RERGKKRRKV FAQEKLSEEE AELLVEADEA REKSVVLMRD PSGIINRGNW TRQKEFWSMV NQRIEGYLVK QIRSRVPLQD SACENTGDI	(SEQ ID NO: 199)
Alicyclobacillus hesperidum WP_074693942.1	MTVRSIRVKL AVGSPQYRDV RRGLWKTHEI MNQGVRYYCE WLVLMRQEPI YDEDEHGLTV VQRTREDIQA ELLSRLRTLQ SAHQHSGDMG TDEELLSLMR QLYEQLVPSS VDKNKSGDAR MIARNFFNPL TNPNSQGGLG ISNAGRKPKW LLKKLSGDPT WEEDYKKAME QKQESSVSFL LLELRRFGLH PIFLPYTDTV LEVSWAPKKA RQWVRKWDYD LFQQSIERML SWESWTRRVK ERFEKLVESE KKFYDENFAT DPEFIKLAET LEGELQASSQ GFVAVDEHAF QIRPRSMRGF DRVADEWCKL ADDAPIEEYE AAIKRVQARL GRNFGSYVLF AHLAKPEYWS LWRSDPTKIL RFARLRALQR AVARAKRHAR LTLPDAIHHP IWIRYDAKGK NIYSYRLLIP EKRSKRYYVE FSSLIMPDGE NRWAEHRNIR VPLAFSRQWE RLHFSIMEDG SLCVQYRDPG VDEPLRAELG GAKIQFDRRY LIRRSSTLSA GECGPVYLNV SVDVNPAHRP DVQVLQSAKL VSVSRDTNRI YLRPENLSAY WKSQGDGTLP LRVMSVDLGV RSSAAVVICR LEHRDSVVSS GRRTATIYRI AGTDEFVAVQ ERAFLLRLPG EGKGTNEDAP LRDVYAQLGT IRQGIQILRS LLRLCDTKTP DERQEALHGL AQSLEPSGAW KDELHPHLVM LQGVVHDSVD NWKQKVISVH RQMERILGHA VREWKVARKN AGKPPIRRGA GGLSLRRIRQ LEQERRTLVA WSNHAREPGQ VVRIKRGTQV AQWLVERVNH LKEDRLKKLA DLLIMTALGY VYDETKPSGH KWDKRYPPCQ IILMEDLSRY RFQSDRPPSE NSQLMAWSHR RLLEILKLQA DLHKLIVGTV FPAFSSRFDA QSGAPGVRCR SVKKQDIENA AQGKGWLARE LQRLNWTLEW LQPNDLIPTG DGELFVTPAC CDRQKGIKIV HADLNAAQNL QRRFWGGHAE SLCRVTCDVV ERDGRRYAVP RISNAFADSF YKVFGQGVFV STDEEDVYRW MVGEKISSRG RSRGRTSDEE	(SEQ ID NO: 200)

Table 6 - Cas12b (C2c1) orthologs		
	AEAETWIDEA REQQGKVIAL FRDASGQIHG GDWLVAKVFW GWVERLVTAR LLSRMSEREA AAHKE	
Alicyclobacillus acidiphilus WP_067623834.1	MAVKSMKVKL RLDNMPEIRA GLWKLHTEVN AGVRYYTEWL SLLRQENLYR RSPNGDGEQE CYKTAEECKA ELLERLRARQ VENGHCGPAG SDDELLQLAR QLYELLVPQA IGAKGDAQQI ARKFLSPLAD KDAVGGLGIA KAGNKPRWVR MREAGEPGWE EEKAKAEARK STDRTADVLR ALADFGLKPL MRVYTDSDMS SVQWKPLRKG QAVRTWDRDM FQQAIERMMS WESWNQRVGE AYAKLVEQKS RFEQKNFVGQ EHLVQLVNQL QQDMKEASHG LESKEQTAHY LTGRALRGSD KVFEKWEKLD PDAPFDLYDT EIKNVQRRNT RRFGSHDLFA KLAEPKYQAL WREDASFLTR YAVYNSIVRK LNHAKMFATF TLPDATAHPI WTRFDKLGGN LHQYTFLFNE FGEGRHAIRF QKLLTVEDGV AKEVDDVTVP ISMSAQLDDL LPRDPHELVA LYFQDYGAEQ HLAGEFGGAK IQYRRDQLNH LHARRGARDV YLNLSVRVQS QSEARGERRP PYAAVFRLVG DNHRAFVHFD KLSDYLAEHP DDGKLGSEGL LSGLRVMSVD LGLRTSASIS VFRVARKDEL KPNSEGRVPF CFPIEGNENL VAVHERSQLL KLPGETESKD LRAIREERQR TLRQLRTQLA YLRLLVRCGS EDVGRRERSW AKLIEQPMDA NQMTPDWREA FEDELQKLKS LYGICGDREW TEAVYESVRR VWRHMGKQVR DWRKDVRSGE RPKIRGYQKD VVGGNSIEQI EYLERQYKFL KSWSFFGKVS GQVIRAEKGS RFAITLREHI DHAKEDRLKK LADRIIMEAL GYVYALDDER GKGKWVAKYP PCQLILLEEL SEYQFNNDRP PSENNQLMQW SHRGVFQELL NQAQVHDLLV GTMYAAFSSR FDARTGAPGI RCRRVPARCA REQNPEPFPW WLNKFVAEHK LDGCPLRADD LIPTGEGEFF VSPFSAEEGD FHQIHADLNA AQNLQRRLWS DFDISQIRLR CDWGEVDGEP VLIPRTTGKR TADSYGNKVF YTKTGVTYYE RERGKKRRKV FAQEELSEEE AELLVEADEA REKSVVLMRD PSGIINRGDW TRQKEFWSMV NQRIEGYLVK QIRSRVRLQE SACENTGDI	(SEQ ID NO: 201)
Alicyclobacillus macrosporangiidus SFU30094.1	MAVKSIKVKL MLGHLPEIRE GLWHLHEAVN LGVRYYTEWL ALLRQGNLYR RGKDGAQECY MTAEQCRQEL LVRLRDRQKR NGHTGDPGTD EELLGVARRL YELLVPQSVG KKGQAQMLAS GFLSPLADPK SEGGKGTSKS GRKPAWMGMK EAGDSRWVEA KARYEANKAK DPTKQVIASL EMYGLRPLFD VFTETYKTIR WMPLGKHQGV RAWDRDMFQQ SLERLMSWES WNERVGAEFA RLVDRRDRFR EKHFTGQEHL VALAQRLEQE MKEASPGFES KSSQAHRITK RALRGADGII DDWLKLSEGE PVDRFDEILR KRQAQNPRRF GSHDLFLKLA EPVFQPLWRE 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 FRVAHRDELT PDAHGRQPFF FPIHGMDHLV ALHERSHLIR LPGETESKKV RSIREQRLDR LNRLRSQMAS LRLLVRTGVL DEQKRDRNWE RLQSSMERGG ERMPSDWWDL FQAQVRYLAQ HRDASGEAWG RMVQAAVRTL WRQLAKQVRD WRKEVRRNAD KVKIRGIARD VPGGHSLAQL DYLERQYRFL RSWSAFSVQA GQVVRAERDS RFAVALREHI DNGKKDRLKK LADRILMEAL GYVYVTDGRR AGQWQAVYPP CQLVLLEELS EYRFSNDRPP SENSQLMVWS HRGVLEELIH QAQVHDVLVG TIPAAFSSRF DARTGAPGIR CRRVPSIPLK DAPSIPIWLS HYLKQTERDA AALRPGELIP TGDGEFLVTP AGRGASGVRV VHADINAAHN LQRRLWENFD LSDIRVRCDR REGKDGTVVL IPRLTNQRVK ERYSGVIFTS EDGVSFTVGD AKTRRRSSAS QGEGDDLSDE EQELLAEADD ARERSVVLFR DPSGFVNGGR WTAQRAFWGM VHNRIETLLA ERFSVSGAAE KVRG	(SEQ ID NO: 202)
Sulfobacillus thermosulfidooxidans PSR34340.1	RQSREDASPQ IIISASDLKA DLLYHARQQQ KEHVPRITGS DAEVLGALRQ VYELIVPSSV GKSGDSKTIA RKFLSPLTDP DSAGGRDQSA SGRKPTWTKM KAEGNPLWEE KFRQWKDRKD NDPTPFVLNQ LADYGLLPLI RLFTDVGENI FDPKKPGQFV RPWDRSMFQQ AIERLMSWES WNQRVRQEWE ALTQKHSAFY REQFTAEPDA ALYRVAQSLE EEMRKEHQGF ATDAPEAFRI RRVALKGFDR LLERWQKTLG KNGQSATLLD DIRRVQSDLG DKFGSAPLYQ KLVDERWQRL WTVDPTFLQR YAAFNDLTQR LQRAKRVANL TLPDAVAHPI WSRYEGPNAS SGNRYHIHLP TTGQPSSVTF DRILWPDGDG GWYERKRVTV FLRPSHQVDR IREAPTDSVV DNFPLVVEDQ SARTILRASW GGAKLEYDRN RLPRQLKKGV PDSIYLSLTL NLDTTKPSGL FHMQQNGRVW IRKDVVMQYY NEIPGDNVQF KPLYVMSVDL GIRSAAAVSI FSVQLKTGIE EHRLTYPVAD CPGLVAVHER SVLLTMPGER REQRDRRYEQ QRQGLRELRT DMRGMNDLLR GAYVDGDRRE	(SEQ ID NO: 203)

Table 6 - Cas12b (C2c1) orthologs		
	EFLARLSKLE ETSPELWEPV YRSLNDSKMA PAAEWERLVV YCHRQVEQSL SSRIQNLRSG RSAYRMSGGL SLDHVQDLER IRGIIASWTN HPRIPGSVVR WQQGRSHTVA LGRHILELKR DRVKKVANYL IMTALGYAYD SKRARGEKWV RRYPSCHLMV FEDLTRYRFR TDRPRSENRQ LMRWTHQELI AVTGIQAEPH GILVGTMYAG FSSRFDAVTK APGVRGATVR QILRTRGMVR LKEIAADVGV DINTLRPHDV LPTGDGEYLL SVVRHRDSYR LKQVHADINA AHNLQRRLWT QDEVFRVSCR LALNSERVVA TPPPSYNKRY GKGFFEKGDN GVYIWKTGGK IKISDMLEED MDIPEDTAEL LRGNSVTLFR DPSGTIAGGN WLEAKEFWGR VNSLVNKGVR DKILGGIPVD NSSAHAE	
Spirochaeta sp. LUC14_002_19_P3 OQX29950.1	MGLLLPSLSR TVNVTIHLIL HPRKKGSRHR EYAVMLDHAV RKIFLAHNWI KRAEAERQKF EADLYKIDRV PQEARDWLDE FCRERTESTG SIDGYHIRRK AVLGWEALVE AWDQKDCLSV EDRIAAARDL QDNPGMDKFG DIWLYEALAS APCVWQKDGE PNAQILLDYV DAGEAEYKRS HYKVPAYRHP DPLLHPIFCD FGQSRWSISF DIHEFKKNGE KNPVNIHALT MGLVSKKRIV KTELKWSSKR LNSNLALSLE SPEDAIEVSR ATRLGRAAVG ASQDRAVNIA GLFESAGWNG RLQAPRKQLE ALAKLEEDKS AEALAKALRN RIKWFITFSP KLQPHGPWME YAERFSGEAP SRAAVIKGKY TVIHQDKTRR RPLAKLHLCR MPGLRVLSVD LGHRHAAACA VWETLSSESM EKKCREAGCL PPAPEDLYLH LKKKNKTAVY RRIGGNFLPD GNEHPAPWAK LDRQFIIDLQ GEEGCTRMAL AGEIWQVHCM EKVFGRSIPL VDRLVRAGWG EKNKQPEILQ ELKQKGWVPL EVSKTNTGYH YSLCVDSLMT LAVNTVRFAL RRHACRARIA YYMEGGAIPE GGLPENSGNK DFIVEALMLW YELATDSRWN GSWEANFWDE NFDKKLAEIQ DAVNEREGDK AKIIKQKERK ELLKKEFIPL AEGLLENSRR ISIASQWRMV WNEEDAIWQS ELRSLRDWIL PKGTRGKKRT IRHVGGLSLS RLAVIKSLYR VQKSFYTRMK PEGEPMDGTM AVGEGFGQKI LDDLETMKEQ RVKQLASRVV EAALGTGRIK KPENNKTPKR PFTAVDEPCH AVVIENLTHY RPENKRTRRE NRQLMTWSSS KVKKYLFESC QLHGLYLFEV QASYTSRQDS RTGAPGVRCS ELSVKKFLES PFRQREIAHA EENMAQENPC NRYLIALHNK WKNREYDKTA PPLRIPHWGG EIFVSALTGN TLQADLNAAA NIGLQALLDP DWPGRWWYVP AVKGCDGRRI PHSKCSGAAC LDNWRVGLKN NLYTGVRTPL PGKNKGSTSG EDVHKSNAVE KSTINLWRDI SVLPLTEGQW	(SEQ ID NO: 204)
Bacillus hisashii strain C4 v4 mutant of WP_095142515.1 K846R S893R E837G	MATRSFILKI EPNEEVKKGL WKTHEVLNHG IAYYMNILKL IRQEAIYEHH EQDPKNPKKV SKAEIQAELW DFVLKMQKCN SFTHEVDKDE VFNILRELYE ELVPSSVEKK GEANQLSNKF LYPLVDPNSQ SGKGTASSGR KPRWYNLKIA GDPSWEEEKK KWEEDKKKDP LAKILGKLAE YGLIPLFIPY TDSNEPIVKE IKWMEKSRNQ SVRRLDKDMF IQALERFLSW ESWNLKVKEE YEKVEKEYKT LEERIKEDIQ ALKALEQYEK ERQEQLLRDT LNTNEYRLSK RGLRGWREII QKWLKMDENE PSEKYLEVFK DYQRKHPREA GDYSVYEFLS KKENHFIWRN HPEYPYLYAT FCEIDKKKKD AKQQATFTLA DPINHPLWVR FEERSGSNLN KYRILTEQLH TEKLKKKLTV QLDRLIYPTE SGGWEEKGKV DIVLLPSRQF YNQIFLDIEE KGKHAFTYKD ESIKFPLKGT LGGARVQFDR DHLRRYPHKV ESGNVGRIYF NMTVNIEPTE SPVSKSLKIH RDDFPKVVNF KPKELTEWIK DSKGKKLKSG IESLEIGLRV MSIDLGQRQA AAASIFEVVD QKPDIEGKLF FPIKGTELYA VHRASFNIKL PGETLVKSRE VLRKAREDNL KLMNQKLNFL RNVLHFQQFE DITEREKRVT KWISRQENSD VPLVYQDELI QIRELMYKPY KDWVAFLKQL HKRLEVEIGK EVKHWRKSLS DGRKGLYGIS LKNIDEIDRT RKFLLRWSLR PTEPGEVRRL EPGQRFAIDQ LNHLNALKED RLKKMANTII MHALGYCYDV RKKKWQAKNP ACQIILFEDL SNYNPYGERS RFENSRLMKW SRREIPRQVA LQGEIYGLQV GEVGAQFSSR FHAKTGSPGI RCRVVTKEKL QDNRFFKNLQ REGRLTLDKI AVLKEGDLYP DKGGEKFISL SKDRKCVTTH ADINAAQNLQ KRFWTRTHGF YKVYCKAYQV DGQTVYIPES KDQKQKIIEE FGEGYFILKD GVYEWVNAGK LKIKKGSSKQ SSSELVDSDI LKDSFDLASE LKGEKLMLYR DPSGNVFPSD KWMAAGVFFG KLERILISKL TNQYSISTIE DDSSKQSM	(SEQ ID NO: 205)

Table 7- Cas12c (C2c3) orthologs		
OspCas12c AWU30132.1 KZX85786.1	MTKLRHRQKK LTHDWAGSKK REVLGSNGKL QNPLLMPVKK GQVTEFRKAF SAYARATKGE MTDGRKNMFT HSFEPFKTKP SLHQCELADK AYQSLHSYLP GSLAHFLLSA HALGFRIFSK SGEATAFQAS SKIEAYESKL ASELACVDLS IQNLTISTLF NALTTSVRGK GEETSADPLI ARFYTLLTGK PLSRDTQGPE RDLAEVISRK IASSFGTWKE MTANPLQSLQ FFEEELHALD ANVSLSPAFD VLIKMNDLQG DLKNRTIVFD PDAPVFEYNA EDPADIIIKL TARYAKEAVI KNQNVGNYVK NAITTTNANG LGWLLNKGLS LLPVSTDDEL LEFIGVERSH PSCHALIELI AQLEAPELFE KNVFSDTRSE VQGMIDSAVS NHIARLSSSR NSLSMDSEEL ERLIKSFQIH TPHCSLFIGA QSLSQQLESL PEALQSGVNS ADILLGSTQY MLTNSLVEES IATYQRTLNR INYLSGVAGQ INGAIKRKAI DGEKIHLPAA WSELISLPFI GQPVIDVESD LAHLKNQYQT LSNEFDTLIS ALQKNFDLNF NKALLNRTQH FEAMCRSTKK NALSKPEIVS YRDLLARLTS CLYRGSLVLR RAGIEVLKKH KIFESNSELR EHVHERKHFV FVSPLDRKAK KLLRLTDSRP DLLHVIDEIL QHDNLENKDR ESLWLVRSGY LLAGLPDQLS SSFINLPIIT QKGDRRLIDL IQYDQINRDA FVMLVTSAFK SNLSGLQYRA NKQSFVVTRT LSPYLGSKLV YVPKDKDWLV PSQMFEGRFA DILQSDYMVW KDAGRLCVID TAKHLSNIKK SVFSSEEVLA FLRELPHRTF IQTEVRGLGV NVDGIAFNNG DIPSLKTFSN CVQVKVSRTN TSLVQTLNRW FEGGKVSPPS IQFERAYYKK DDQIHEDAAK RKIRFQMPAT ELVHASDDAG WTPSYLLGID PGEYGMGLSL VSINNGEVLD SGFIHINSLI NFASKKSNHQ TKVVPRQQYK SPYANYLEQS KDSAAGDIAH ILDRLIYKLN ALPVFEALSG NSQSAADQVW TKVLSFYTWG DNDAQNSIRK QHWFGASHWD IKGMLRQPPT EKKPKPYIAF PGSQVSSYGN SQRCSCCGRN PIEQLREMAK DTSIKELKIR NSEIQLFDGT IKLFNPDPST VIERRRHNLG PSRIPVADRT FKNISPSSLE FKELITIVSR SIRHSPEFIA KKRGIGSEYF CAYSDCNSSL NSEANAAANV AQKFQKQLFF EL	(SEQ ID NO: 206)
QFN42172.1	MRSNYHGGRN ARQWRKQISG LARRTKETVF TYKFPLETDA AEIDFDKAVQ TYGIAEGVGH GSLIGLVCAF HLSGFRLFSK AGEAMAFRNR SRYPTDAFAE KLSAIMGIQL PTLSPEGLDL IFQSPPRSRD GIAPVWSENE VRNRLYTNWT GRGPANKPDE HLLEIAGEIA KQVFPKFGGW DDLASDPDKA LAAADKYFQS QGDFPSIASL PAAIMLSPAN STVDFEGDYI AIDPAAETLL HQAVSRCAAR LGRERPDLDQ NKGPFVSSLQ DALVSSQNNG LSWLFGVGFQ HWKEKSPKEL IDEYKVPADQ HGAVTQVKSF VDAIPLNPLF DTTHYGEFRA SVAGKVRSWV ANYWKRLLDL KSLLATTEFT LPESISDPKA VSLFSGLLVD PQGLKKVADS LPARLVSAEE AIDRLMGVGI PTAADIAQVE RVADEIGAFI GQVQQFNNQV KQKLENLQDA DDEEFLKGLK IELPSGDKEP PAINRISGGA PDAAAEISEL EEKLQRLLDA RSEHFQTISE WAEENAVTLD PIAAMVELER LRLAERGATG DPEEYALRLL LQRIGRLANR VSPVSAGSIR ELLKPVFMEE REFNLFFHNR LGSLYRSPYS TSRHQPFSID VGKAKAIDWI AGLDQISSDI EKALSGAGEA LGDQLRDWIN LAGFAISQRL RGLPDTVPNA LAQVRCPDDV RIPPLLAMLL EEDDIARDVC LKAFNLYVSA INGCLFGALR EGFIVRTRFQ RIGTDQIHYV PKDKAWEYPD RLNTAKGPIN AAVSSDWIEK DGAVIKPVET VRNLSSTGFA GAGVSEYLVQ APHDWYTPLD LRDVAHLVTG LPVEKNITKL KRLTNRTAFR MVGASSFKTH LDSVLLSDKI KLGDFTIIID QHYRQSVTYG GKVKISYEPE RLQVEAAVPV VDTRDRTVPE PDTLFDHIVA IDLGERSVGF AVFDIKSCLR TGEVKPIHDN NGNPVVGTVA VPSIRRLMKA VRSHRRRRQP NQKVNQTYST ALQNYRENVI GDVCNRIDTL MERYNAFPVL EFQIKNFQAG AKQLEIVYGS	(SEQ ID NO: 207)
QFN42158.1	MKKFELKQNF RNNYSGKTLR NFRQTLAQIA NKKSSDSILT IKFKLDCSKT GKLPKYENLI SLYDTIEDIK KGTLSYYLFT LIVSGFKFFG SASQAKAFST KDIFKDNDFY NQFKIQSHLD LPDFVPSKIY QRLKKNVRST NGKDNAFKAS VIVAEYRKEI GKLKNKDESS EHQCEELFKK IGTALETRFS SWQDLINNCS TGCEIIDEIL NDSFGTLPSI KKMVLASTTQ SSDGEQDGIA IAYDPDSTFI KSDELLNPYF AVATILKSMP PEIQQDKKSA YVKANLTTPT HNALSWIFGK GLTLFQTEST EKLCAMFNVS DKRVIEQVQD AAKAVKLPAE LDLNHCTLKF QDFRSSLGGH LDSWTTNYLK RLDELNDLLL NLPKNLSLPD IFMIDGKDFI EYSGCNRDEI QQMIDFVVNE QNRIKLQESL NALLGKGNNQ ICSDDISTVK DFSEIVNSLH SFVQQIDNSL EQSSNEANSI FSELKKKIEK NEKWDIWKNN LKKIPKLNKL SGGVPDAWKE IREIEQKFHE ISENQKKHFT EVMEWIDAGN GTIDIFESRF KYDELLKKSK KNNLQSADEL AFRSVLNKLG RFARQGNDLV CEKIKNWFKE QNIFDSSKDF NRYFINQKGF IFKHPSSKKD NSPYNLSANL LEKRYEVTNT VGALLEQCES DPAIVNDPFS MRSLVEFRAL WFSINISGIS	(SEQ ID NO: 208)

Table 7- Cas12c (C2c3) orthologs		
	KEQHIPTKIA QPKLDDSTYQ ESVSPTLKYR LEKEQITSSE LNSIFTVYKS LLSGLSIRLS RNSFYLRTKF SWIGNNSLIY CPKETTWKIP AAYFKSDLWN EYKDKQILIV NEEYDVDVVK TFESVYKIVK SKDNNEKNRI LPLLKQLPHD WMFKLPFGAS NAEKCKVLKL EKNNKKFKPL SVSKDSLARL SGPSTYFNQI DEIMMNDESE LSEMTLLADE PVRQQMSNGK IEIIPDDYVM SLAIPITRSL KKGNTESFPF KNIVSIDQGE AGFAYAVFKL SDCGNERAEP IATGLIPIPS IRRLIHSVKK YRGKKQRIQN FNQKFDSTMF TLRENVTGDI CGLIVALMKK YNAFPILEKQ VGNLESGSKQ LMLVYKAVNS KFLAAKVDMQ NDQRRSWWYQ GNSWNTPILR ISNPNQSNNK NIVKNINGKK YEELKIYPGY SVSAYMTSCI CHVCGRNALE LLKNDDSTGK VKKYQINQDG EVTIGGEVIK LYRKPDRLTP VKNLAKKGNR ERTYASINER APMSKDTTQS RYFCVFKNCP CHNKEQHADV NAAINIGRRF LKDCILDDNK EKD	
QFN42173.1	MNARDWRKHV GVLAQQHKET TRTYTFPLDT TGSAIDFDAA LQAYNAVEGV GYGSLLGLAC AVHLSGFRLF STGKEAATFR NRARYPNAAF QAALRKELGT TITTLTPETL DRLFSSRPKR RNGVPLPWNQ DSIRDRLYTN WVKPRPGDTP DAVLFQIATG IAQEITEDVS SWTDLAKNSD RGLKAAHRYF ARVGGFPAFD NLTPPATVQP TDTTIDYDPN APFHLVSHAD QTLIHQSISL CAHRIRQEDP ALDPNKSGFI KQLQNNFLSQ TFYGLSWLFG AGYVHFRECT ANDLAIQYGI PNNCRDGIHQ IKSFADAILP NTFFEKKHYR KDSRSVGKKA KSWISNYWQR LLQLQTWVDD HTWVTLPQEL TEAQFKPLFR GLLVDAVELM AIAERLPQRL ADCRDSLDCL MGKGPQAATK NDVEIVEKVR EEIESFVGQI EQLGNQLRHQ LENENNDQVH RDNLHQLKNR LPLDLRRPQA LNKISGGVPD VAKSIRGLET QLDQVLKERR SHFGRLTKWA KECGITLDPL QPLIESEKQR VAERGSAHDA KELAIRLLLQ RIGRLGHRLS PTNATAIQEL LRPVFAVKRE FNLFFHNHMG ALYRSPYSTS RHQPFQINVD VAHGTDWIGT IETLIQNLFT QIQDDALLRD LVQLEGFVFS HKLRALPGVI PSELARPNNL QQMGLPALLL VLLQADQVHR ETVLRVFNLY GSAINGYLFQ ALRPGFIVRA GFQRLETKKL RYVPKAQSWQ YPDRLHHAKS AIKNSLSAGW IKKNHQGAIL PQKTLTALVK QKSLKDTGVP EYLVQAPHDW YVPIDLRGPA IPIEGLTVGT EGPELTQLGP MKDDCAFRAI GPSSFKSKID AGLLPQDVKY GDMTLIFDQH YQQSISFANG TFSIQYQPTS LQVKAAIPVV DKRPRDTRNN SHLYDRIVAI DLGERKIGYA IFDLKQVLKS EQLEPMREDG KPLIGSISIR SIRGLMKAVQ THRNRRQPNY RIDQTYSKAL MHYRESVIGD VCNAIDTLCA RYGGFPVLES SVRNFEVGSA QLKTVYGSVS RRYTWSAVDA HKNQRQQYWL GGTKDKIPIW THPYLMTREW DEKNSKWSNR SKPLKMHPGV EVHPAGTSQI CHQCKRNPIG ALWNVADTVV LDDQGQLDLD DGTIRLNSGY IDTTEIKRAR RKKIRLPENK PLTGSHKTSH VRAVARRNLR QPPKSTRAKD TTQSRYTCLY VDCGHECHAD ENAAINIGRK YLQERIHIEA SRQALSTR	(SEQ ID NO: 209)
QFN42174.1	MVAGLKKIKR DGVTMKSNYH GGVKARAWRK RIGGLARRQK ETVFTYKFPL ETEEAGIDFD KAVQTYGIAE GISQGSLIGL VCAFHLSGFR LFSKADETKA FCNQGRYPNQ AFAEKLRNEL SVTLPKLSPQ SLDVLFQSSP KSKNGVAPEW SKNAIRNRLY TNWTGKGAGT NPDEHLLEIA EDIAAEIDSD LDGWKDLEEH PEKGLSAADR YFQAQGDFPS LTGLPPSVPL TPQNSTVAFE GDPVCLNPSD NTLLHQAVAR CAGRILQEQP NLSPDKNRFI NQLQDELVSS QNNGLSWLFG VGFKYWKEMS VDQLADDYKV KSTDLDALKQ VKSFIDAIPL NPLFDTPHYG EFRASVAGKM RSWVKNYWKR LLDLKSQLGT ANINLPEGLD EQRAENLFSG LLIDSKGLRQ VTDKLPSRLK KAEDTIDRLM GDGNPTSDDI EQVETVAAEI SAFIGQVEQF NNQLEQRLEN PLEGDDETFL KQLKIDLPAE FKKPPAINRI SGGSPDPTAE IAELEEKLDR LMSARKEHYE TIAEWASANK VTLDPMEAMT TLEAQRLTER GAEGDQEEFA LRLLLQRIGR LANRLSPQGA TAIRDLLRPV FTEKREFNLF FHNRMGSLYR SPYSTSRHQP FTIDVAVAKN TDWMDALDGI AETIMKGLSQ AGDELSLRQL EEDEVSREVC LKAFNLYVSA INGCLFRALR EGFIVRTKFQ RLERDVLSYV PKTKLWNYPQ RLDTARGPIH SALAAAWINK EGSVIDPVET VTALSDTGFS DDGIPEYLVQ APHDWYLRDW INISGFSLSQ RLRGLPDTVP GELALVRSAD DVRIPPMLAL TPIDLRDISK PVSGLPVKKN ITGLKRQKKQ TAFRMVGPSS FKSHLDSTLL SEEVKLGDFT LIFDQYYKQR VSYNGRVKIT FEPDRLHVEA AVPVIDKRVR PSTEEDALFD HLLAIDLGEK RVGYAVYDIK ACLRTGDIKP LEDGDGKPIV GSVAVPSIRR LMKAVRSHRQ QRQPNQKVNQ TYSTALMNYR ENVIGDVCNR IDTLMEKYNA FPVLESSVMN FEAGSRQLEM VYGSVLHRYT YSKIDAHTAK RKEYWYTGEY WDHPYLMAHK WNERTRSYSG SLSALTLYPG VMVHPAGTSQ RCHQCKRNPM VEIKQLTGQV	(SEQ ID NO: 210)

Table 7- Cas12c (C2c3) orthologs		
	EINADGSLEL DDGTICLYEG YDYSPEEYKK AKREKRRLDP NVPLSGRHQA KHVSAVAKRN LRRPTVSMMS GDTTQARYVC LYTDCDFTGH ADENAAINIG WKYLTERIAL SESKDKAGV	

Table 8 - Cas12e (CasY) orthologs		
APG80656.1 GI: 1110962136 QFN42175.1	MSKRHPRISG VKGYRLHAQR LEYTGKSGAM RTIKYPLYSS PSGGRTVPRE IVSAINDDYV GLYGLSNFDD LYNAEKRNEE KVYSVLDFWY DCVQYGAVFS YTAPGLLKNV AEVRGGSYEL TKTLKGSHLY DELQIDKVIK FLNKKEISRA NGSLDKLKKD IIDCFKAEYR ERHKDQCNKL ADDIKNAKKD AGASLGERQK KLFRDFFGIS EQSENDKPSF TNPLNLTCCL LPFDTVNNNR NRGEVLFNKL KEYAQKLDKN EGSLEMWEYI GIGNSGTAFS NFLGEGFLGR LRENKITELK KAMMDITDAW RGQEQEEELE KRLRILAALT IKLREPKFDN HWGGYRSDIN GKLSSWLQNY INQTVKIKED LKGHKKDLKK AKEMINRFGE SDTKEEAVVS SLLESIEKIV PDDSADDEKP DIPAIAIYRR FLSDGRLTLN RFVQREDVQE ALIKERLEAE KKKKPKKRKK KSDAEDEKET IDFKELFPHL AKPLKLVPNF YGDSKRELYK KYKNAAIYTD ALWKAVEKIY KSAFSSSLKN SFFDTDFDKD FFIKRLQKIF SVYRRFNTDK WKPIVKNSFA PYCDIVSLAE NEVLYKPKQS RSRKSAAIDK NRVRLPSTEN IAKAGIALAR ELSVAGFDWK DLLKKEEHEE YIDLIELHKT ALALLLAVTE TQLDISALDF VENGTVKDFM KTRDGNLVLE GRFLEMFSQS IVFSELRGLA GLMSRKEFIT RSAIQTMNGK QAELLYIPHE FQSAKITTPK EMSRAFLDLA PAEFATSLEP ESLSEKSLLK LKQMRYYPHY FGYELTRTGQ GIDGGVAENA LRLEKSPVKK REIKCKQYKT LGRGQNKIVL YVRSSYYQTQ FLEWFLHRPK NVQTDVAVSG SFLIDEKKVK TRWNYDALTV ALEPVSGSER VFVSQPFTIF PEKSAEEEGQ RYLGIDIGEY GIAYTALEIT GDSAKILDQN FISDPQLKTL REEVKGLKLD QRRGTFAMPS TKIARIRESL VHSLRNRIHH LALKHKAKIV YELEVSRFEE GKQKIKKVYA TLKKADVYSE IDADKNLQTT VWGKLAVASE ISASYTSQFC GACKKLWRAE MQVDETITTQ ELIGTVRVIK GGTLIDAIKD FMRPPIFDEN DTPFPKYRDF CDKHHISKKM RGNSCLFICP FCRANADADI QASQTIALLR YVKEEKKVED YFERFRKLKN IKVLGQMKKI	(SEQ ID NO: 211)

6.4. Protospacer Adjacent Motif

[0086] 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.

[0087] 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.

[0088] 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 (NmCas) recognizes NNNNGATT. In another example, Cas9 from Streptococcus thermophilis (StCas9) recognizes NNAGAAW. In still another example, Cas9 from Treponema denticola (TdCas) 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 disclosure. 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.

[0089] 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

[0090] "**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 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. 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).

[0091] 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.

[0092] As used herein, "PE2" refers to a PE complex comprising a fusion protein comprising Cas9(H840A) and a variant MMLV RT having the following N-terminus to C-terminus structure: [NLS]-[Cas9(H840A)]- [linker]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)] + a desired atgRNA (PEgRNA). In various embodiments, the prime editors disclosed herein are comprised of PE2.

[0093] In various embodiments, the prime editors disclosed herein are comprised of PE2 and co-expression of MMR protein MLH1dn, that is PE4.

[0094] As used herein, "PE3" refers to PE2 plus a second-strand nicking guide RNA that complexes with the PE2 and introduces a nick in the non-edited DNA strand. The induction of the second nick increases the chances of the unedited strand, rather than the edited strand, to be repaired. In various embodiments, the prime editors disclosed herein are comprised of PE3.

[0095] In various embodiments, the prime editors disclosed herein is comprised of PE3 and co-expression of MMR protein MLH1dn, that is PE5.

[0096] 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

[0097] 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.

[0098] 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).

6.7. Attachment Site-Containing Guide RNA (atgRNA)

[0099] 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).

[0100] As used herein, the term "cognate integration recognition site" or "integration cognate" or "cognate pair" refers to a first integration recognition site (e.g., any of the integration recognition sites described herein) and a second integration recognition site (e.g., any of the integration recognition sites described herein) that can be recombined. Recombination between a first integration recognition site (e.g., any of the integration recognition sites described herein) and a second recognition site (e.g., any of the integration recognition sites described herein) is mediated by functional symmetry between the two integration recognition sites and the central dinucleotide of each of the two integration recognition sites. In some cases, a first integration recognition site (e.g., any of the integration recognition sites described herein) that can be recombined with a second integration recognition site (e.g., any of the integration recognition sites described herein) are referred to as a "cognate pair." A non-limiting example of a cognate pair include an attB site and an attP site, whereby a serine integrase mediates recombination between the attB site and the attP site. FIGs. 1A-1E show optimization of the integration recognition site.

[0101] In typical embodiments, an atgRNA comprises a reverse transcriptase template that encodes, partially or in its entirety, an integration recognition site (also referred to as an integration target recognition site) or a recombinase recognition site (also referred to as a recombinase target recognition site). The integration target recognition site, which is to be placed at a desired location in the genome or intracellular nucleic acid, is referred to as a "beacon," a "beacon" site or an "attachment site" or a "landing pad" or "landing site." An integration target recognition site or recombinase target recognition site incorporated into the pegRNA is referred to as an attachment site containing guide RNA (atgRNA).

[0102] 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.

[0103] In some embodiments, where the system contains a first atgRNA and a second atgRNA (see FIG. 2), 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.

[0104] 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 complementarity to each other and comprised of consecutive bases of an integrase target recognition site are referred to as dual, paired, annealing, complementary, or twin attachment site-containing guide RNAs (atgRNAs). In certain embodiments, use of two guide RNAs that are (or encode DNA that is) full complementarity to each other and comprised of consecutive bases of an integrase target recognition site are referred to as dual, paired, annealing, complementary, or twin attachment site-containing guide RNAs (atgRNAs).

[0105] In some embodiments, upon introducing the nucleic acid construct into a cell, the first atgRNA incorporates the first integration recognition site into the cell's genome at the target sequence.

6.7.1. AtgRNAs Comprising a Clamp Segment

[0106] This disclosure features an attachment site containing guide RNA (atgRNA) comprising a clamp segment (see, e.g., FIGs. 3A-3C and FIGs. 4A-4C). The clamp segment is designed, at least in part, to enhance integration efficiency of the integration recognition site at the double-stranded target DNA sequence (e.g., as compared to integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment).

[0107] In one embodiment, this disclosure features an attachment site containing guide RNA (atgRNA) comprising from 5' to 3': a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase binding domain to the target DNA sequence; a reverse transcriptase template comprising: an integration recognition site; and a clamp segment comprising a sequence where upon being reverse transcribed the sequence is at least partially complementary (e.g., the reverse transcribed clamp segment comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity) to a second genomic site of the first strand of the double-stranded target DNA sequence; and a primer binding site complementary to a third genomic site of the double-stranded target DNA sequence, wherein the third genomic site is on a second strand of the target DNA sequence. In some embodiments, the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site. In some embodiments, the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site and the clamp segment. In some embodiments, upon introducing into a cell with a gene editor polypeptide, all or a portion of the single-stranded DNA sequence comprising the integration recognition site is site-specifically integrated into the double-stranded target DNA sequence. In some embodiments, upon introducing the atgRNA into a cell with a gene editor polypeptide, the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment.

[0108] In one embodiment, this disclosure features an attachment site containing guide RNA (atgRNA) comprising from 5' to 3': a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase binding domain to the target DNA sequence; a reverse transcriptase template comprising: an integration recognition site; and a clamp segment comprising a sequence where upon being reverse transcribed the sequence is at least partially complementary (e.g., the reverse transcribed clamp segment comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity) to a second genomic site of the first strand of the double-stranded target DNA sequence; and a primer binding site comprising: a sequence complementary to a third genomic site the double-stranded DNA target sequence, wherein the third genomic site is on a second strand of the target DNA sequence; and an anti-reverse transcriptase template (anti-RTT) sequence comprising a sequence at least partially complementary (e.g., the anti-RTT sequence comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity) to all or a portion of the RT template. In some embodiments, the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site. In some embodiments, the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site and the clamp segment. In some embodiments, upon introducing into a cell with a gene editor polypeptide, all or a portion of the single-stranded DNA sequence comprising the integration recognition site is site-specifically integrated into the double-stranded target DNA sequence. In some embodiments, upon introducing the atgRNA into a cell with a gene editor polypeptide, the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment.

[0109] In one embodiment, this disclosure features an attachment site containing guide RNA (atgRNA), comprising from 5' to 3': a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase binding domain to the target DNA sequence; a clamp segment at least partially complementary (e.g., the clamp segment comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity) to a second genomic site of the first strand of the double-stranded target DNA sequence; a reverse transcriptase (RT) template comprising an integration

recognition site; and a primer binding site complementary to a third genomic site of the double-stranded target DNA sequence, wherein the third genomic site is on a second strand of the target DNA sequence; wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site. In some embodiments, upon introducing into a cell with a gene editor polypeptide, all or a portion of the single-stranded DNA sequence comprising the integration recognition site is site-specifically integrated into the double-stranded target DNA sequence. In some embodiments, upon introducing the atgRNA into a cell with a gene editor polypeptide, the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment. In some embodiments, the atgRNA includes a transcription termination signal that prevents reverse transcription of the clamp segment. In one embodiment, the internucleoside linkage connecting the RT template to the clamp signal comprises a moiety that prevents reverse transcription of the clamp segment.

[0110] In one embodiment, this disclosure features an attachment site containing guide RNA (atgRNA) comprising from 5' to 3': a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence; a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase binding domain to the target DNA sequence; a clamp segment at least partially complementary (e.g., the clamp segment comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity) to a second genomic site of the first strand of the double-stranded target DNA sequence; a reverse transcriptase (RT) template comprising an integration recognition site; and a primer binding site comprising: a sequence complementary to a third genomic site of the double-stranded DNA target sequence, wherein the third genomic site is on a second strand of the target DNA sequence; and an anti-reverse transcriptase template (anti-RTT) sequence comprising a sequence at least partially complementary (e.g., the anti-RTT sequence comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity) to all or a portion of the RT template, wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site. In some embodiments, upon introducing into a cell with a gene editor polypeptide, all or a portion of the single-stranded DNA sequence comprising the integration recognition site is site-specifically integrated into the double-stranded target DNA sequence. In some embodiments, upon introducing the atgRNA into a cell with a gene editor polypeptide, the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site

at the double-stranded target DNA sequence as compared to integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment. In some embodiments, the atgRNA includes a transcription termination signal that prevents reverse transcription of the clamp segment. In one embodiment, the internucleoside linkage connecting the RT template to the clamp signal comprises a moiety that prevents reverse transcription of the clamp segment.

[0111] In some embodiments, the atgRNA has a length of 150 to 210 nucleotides (e.g., a length of 150 to 200 nucleotides, a length of 150 to 190 nucleotides, a length of 150 to 180 nucleotides, a length of 150 to 170 nucleotides, a length of 150 to 160 nucleotides, a length of 160 to 210 nucleotides, a length of 160 to 200 nucleotides, a length of 160 to 190 nucleotides, a length of 160 to 180 nucleotides, a length of 160 to 170 nucleotides, a length of 170 to 210 nucleotides, a length of 170 to 200 nucleotides, a length of 170 to 190 nucleotides, a length of 170 to 180 nucleotides, a length of 180 to 210 nucleotides, a length of 180 to 200 nucleotides, a length of 180 to 190 nucleotides, a length of 190 to 210 nucleotides, a length of 190 to 200 nucleotides, or a length of 200 to 210 nucleotides.

[0112] In some embodiments, the RT template has a length of 10 nucleotides to 200 nucleotides (e.g., a length 10 to 180 nucleotides, a length of 10 to 160 nucleotides, a length of 10 to 140 nucleotides, a length of 10 to 120 nucleotides, a length of 10 to 100 nucleotides, a length of 10 to 80 nucleotides, a length of 10 to 60 nucleotides, a length of 10 to 40 nucleotides, a length of 10 to 20 nucleotides, a length of 20 to 200 nucleotides, a length of 20 to 180 nucleotides, a length of 20 to 160 nucleotides, a length of 20 to 140 nucleotides, a length of 20 to 120 nucleotides, a length of 20 to 100 nucleotides, a length of 20 to 80 nucleotides, a length of 20 to 60 nucleotides, a length of 20 to 40 nucleotides, a length of 40 to 200 nucleotides, a length of 40 to 180 nucleotides, a length of 40 to 160 nucleotides, a length of 40 to 140 nucleotides, a length of 40 to 120 nucleotides, a length of 40 to 100 nucleotides, a length of 40 to 80 nucleotides, a length of 40 to 60 nucleotides, a length of 60 to 200 nucleotides, a length of 60 to 180 nucleotides, a length of 60 to 160 nucleotides, a length of 60 to 140 nucleotides, a length of 60 to 120 nucleotides, a length of 60 to 100 nucleotides, a length of 60 to 80 nucleotides, a length of 80 to 200 nucleotides, a length of 80 to 180 nucleotides, a length of 80 to 160 nucleotides, a length of 80 to 140 nucleotides, a length of 80 to 120 nucleotides, a length of 80 to 100 nucleotides, a length of 100 to 200 nucleotides, a length of 100 to 180 nucleotides, a length of 100 to 160 nucleotides, a length of 100 to 140 nucleotides, a length of 100 to 120 nucleotides, a length of 120 to 200 nucleotides, a length of 120 to 180 nucleotides, a length of 120 to 160 nucleotides, a length of 120 to 140 nucleotides, a length

of 140 to 200 nucleotides, a length of 140 to 180 nucleotides, a length of 140 to 160 nucleotides, a length of 160 to 200 nucleotides, a length of 160 to 180 nucleotides, or a length of 180 to 200). In some embodiments, the RT template has a length of 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 or more nucleotides.

[0113] In typical embodiments, the atgRNA binds at least a first target genomic site (via spacer sequence), a second genomic site (via clamp sequence), and a third genomic site (via primer binding site). In some embodiments, the atgRNA is designed to optimize the distance between the second genomic site and the third genomic site. Optimization of the distance between the second genomic site and the third genomic site can be based on the length of the RT template to be integrated into the target sequence, the sequence of the target DNA sequence, among other factors. In some cases, the distance between the second genomic site and the third genomic site can be referred to as the "separation" between the second genomic site and the third genomic site.

[0114] In some embodiments, the atgRNA is designed such that there is partial genomic replacement of the target DNA sequence upon incorporation of the integration recognition sequence into the target DNA sequence. In such cases, the atgRNA is designed so that the second genomic site and the third genomic site are separated by one or more genomic bases (e.g., one or more base pairs) in the double-stranded target DNA sequence, whereby the one or more genomic bases are deleted upon incorporation of the integration recognition site.

[0115] In some embodiments, the atgRNA is designed such that partial genomic replacement comprises deletion of 1 to 10 genomic bases (e.g., base pairs), 11 to 20 genomic bases, 21 to 30 genomic bases, 31 to 40 genomic bases 41 to 50 genomic bases, 51 to 60 genomic bases, 61 to 70 genomic bases, 71 to 80 genomic bases, 81 to 90 genomic bases, or 91-100 genomic bases. In some embodiments, the atgRNA is designed such that partial genomic replacement comprises deletion of 100 to 1000 genomic bases (e.g., base pairs). In some embodiments, the atgRNA is designed such that partial genomic replacement comprises deletion of 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, or 45 genomic bases (e.g., base pairs).

[0116] In some embodiments, the second genomic site and the third genomic site are separated by one or more genomic bases (e.g., one or more base pairs) in the double-stranded target DNA sequence. In some embodiments, by relaxing the requirement that the clamp binding site (i.e., second genomic site) be immediately adjacent to the primer binding site (i.e., third genomic site), this disclosure provides for guide RNA compositions which are capable of being reduced in total

sequence length as compared to standard prime editing guide RNAs or gene editing RNAs capable of integrase target site recognition site insertion (i.e., standard atgRNAs). The clamp segment binding site (i.e., second genomic site) can be selected from within a large genomic window, wherein for example the window frame can be shifted to screen for variable sequence distance from the third genomic site and/or constrained to select for efficient hybridization to a defined length clamp segment. In some embodiments, the atgRNA clamp segment is less than 50 nucleotides in length. In some embodiments, the atgRNA clamp segment is less than 25 nucleotides in length. In some embodiments, the atgRNA clamp segment is less than 15 nucleotides in length. In some embodiments, the atgRNA clamp segment is less than 10 nucleotides in length. In some embodiments, by relaxing the requirement that the clamp binding site (i.e., second genomic site) be immediately adjacent to the primer binding site (i.e., third genomic site), this disclosure provides for atgRNA compositions which are capable of increasing efficiency of the integration recognition site genomic incorporation (i.e., beacon placement). In some embodiments, increased efficiency is due to favored repair by endogenous cellular machinery. In some embodiments, increased efficiency is due to reduced 3' flap thermodynamic strain during genomic binding (i.e., during "flap equilibration" processes see FIGs. 4A-4C)).

[0117] In some embodiments, the second genomic site and the third genomic site are separated by 1 to 10 genomic bases, 11 to 20 genomic bases, 21 to 30 genomic bases, 31 to 40 genomic bases, or 41 to 45 genomic bases. In some embodiments, the second genomic site and the third genomic site are separated by 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, or 45 genomic bases (e.g., base pairs). In some embodiments, the second genomic site and the third genomic site are separated by at least 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 (1 kilobase distance) genomic bases. In some embodiments, the second genomic site and the third genomic site are separated by more than 1 kilobase. In some embodiments, second genomic site and the third genomic site are located on different chromosomes.

[0118] In some embodiments, the second genomic site and the third genomic site are adjacent in the double-stranded target DNA sequence. In some embodiments, the genomic bases (e.g., base pairs) that are between the second genomic site and the third genomic site are not incorporated into the target DNA sequence, thereby removing (e.g., deleting) these genomic bases from the target DNA sequence.

[0119] In certain embodiments, the guide RNA is circularized. In some embodiments, guide RNA circularization is mediated by 5' and 3' terminus dimerization motifs. In some embodiments, the

dimerization motifs are kissing loops. In some embodiments, guide RNA circularization is mediated by one or more of a ribozyme or a ligase. In some embodiments, guide RNA circularization is mediated by covalently or non-covalently linking the 5' and 3' termini.

[0120] In typical embodiments, one or more guide RNA is comprised in one or more RNA polynucleotides or DNA polynucleotides.

6.7.2. Clamp Segment

[0121] This disclosures features an attachment site containing guide RNA (atgRNA) comprising a clamp segment that is at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence. In some instances, the clamp is designed to enhance efficiency of integrating the integration recognition site into the target DNA Sequence. For example, upon introducing the atgRNA comprising the clamp into a cell with a gene editor polypeptide, the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to the integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment. In some embodiments, inclusion of a clamp segment in the atgRNA allows for a reduction in the number of nucleotides in the primer binding site (PBS) need to bind the third genomic site and facilitate reverse transcription. This in turn enables design of atgRNA that are shorter than guide RNAs used for other gene editing methods, for example, prime editing.

[0122] In some embodiments, the clamp segment has a length of 1 to 10 nucleotides, 11 to 20 nucleotides, 21 to 30 nucleotides, or 31 to 40 nucleotides. In some embodiments, the clamp segment has a length of 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, or more nucleotides. In some embodiments, the clamp segment has a length of 10 nucleotides.

[0123] In some embodiments, the clamp segment comprises a sequence that is partially complementary to the second genomic site of the first strand of the double-stranded target DNA sequence. In some embodiments, a clamp segment that is partially complementary o the second genomic site of the first strand of the double-stranded target DNA sequence includes a clamp segment having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity to the second genomic site. In some embodiments, the clamp segment comprises a sequence that is complementary to the second genomic site of the first strand of the double-stranded target DNA sequence.

[0124] In some embodiments, the clamp segment comprises a sequence that is capable of binding to the second genomic site of the first strand of the double-stranded target DNA sequence with at least a first nucleotide mismatch. In some embodiments, the clamp segment comprises a sequence that is capable of binding to the second genomic site of the first strand of the double-stranded target DNA sequence with a first nucleotide mismatch and a second nucleotide mismatch. In some embodiments, the first nucleotide mismatch and the second nucleotide mismatch are consecutive. In some embodiments, the first nucleotide mismatch and the second nucleotide mismatch are non-consecutive. In some embodiments, the clamp segment comprises a sequence that is capable of binding to the second genomic site of the first strand of the double-stranded target DNA sequence with the first nucleotide mismatch, the second nucleotide mismatch, and a third nucleotide mismatch. In some embodiments, the clamp segment comprises a sequence that is capable of binding to the second genomic site of the first strand of the double-stranded target DNA sequence with the first nucleotide mismatch, the second nucleotide mismatch, the third nucleotide mismatch, and a fourth nucleotide mismatch.

[0125] In one embodiment, the reverse transcriptase template comprises an integration recognition site and a clamp segment. In such embodiments, the clamp segment includes a sequence where upon being reverse transcribed comprises a sequence at least partially complementary (e.g., the reverse transcribed clamp segment comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity) to a second genomic site of the first strand of the double-stranded target DNA sequence. In some embodiments, the clamp segment comprises a sequence (e.g., a reverse complement of the clamp "rcClamp") that upon being reverse transcribed is complementary to the second genomic site of the first strand of the double-stranded target DNA sequence.

6.7.3. Anti-Reverse Transcriptase Template (anti-RTT) Sequence

[0126] In some embodiments, an atgRNA comprises an anti-reverse transcriptase template (anti-RTT) sequence (see, e.g., FIGs. 6A-6B). In some embodiments, the anti-RTT sequence is capable of binding to all or a portion of the RT template, thereby blocking the 3' end of the atgRNA from degradation (e.g., degradation by exonucleases). In such embodiments, upon introducing the atgRNA into a cell along with a gene editor polypeptide comprising a DNA binding nickase domain linked to a reverse transcriptase domain, and after binding of the PBS to the third genomic site and initiation of a reverse transcription reaction, the reverse transcriptase can displace the anti-RTT sequence, thereby making the RT template sequence accessible for reverse transcription.

[0127] In some embodiments, upon being introduced into a cell, the binding of the anti-RTT sequence to all or a portion of the RT template increases stability and half-life of the atgRNA by protecting the atgRNA from exonuclease activity. In some embodiments, upon introducing the atgRNA into a cell, the anti-RTT sequence increases stability and half-life of the atgRNA as compared to an atgRNA that does not comprise the anti-RTT sequence.

[0128] In some embodiments, the anti-RTT sequence comprises a sequence that is at least partially complementary to the RT template. For example, the anti-RTT sequence comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity to the RT template. In some embodiments, the anti-RTT sequence comprises a sequence that is complementary to all or a portion of the RT template.

[0129] In some embodiments, the anti-RTT sequence has a length of 1 to 20 nucleotides, 5 to 15 nucleotides, or 10 to 20 nucleotides. In some embodiments, the anti-RTT sequence has a length of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more nucleotides. In one embodiment, the anti-RTT sequence has a length of 10 nucleotides.

[0130] In some embodiments, the atgRNA further comprises a 3' extension (also referred to as anti-RTT sequence) capable of binding one or more consecutive nucleotides of the reverse transcriptase template first segment. In some embodiments, the atgRNA further comprises a 3' extension capable of binding one or more consecutive nucleotides of the reverse transcriptase template second segment. In some embodiments, the atgRNA further comprises a 3' extension capable of binding one or more consecutive nucleotides of the reverse transcriptase template first and second segment. In some embodiments, the atgRNA further comprises a 5' extension capable of binding one or more consecutive nucleotides of the reverse transcriptase template first segment. In some embodiments, the atgRNA further comprises a 5' extension capable of binding one or more consecutive nucleotides of the reverse transcriptase template second segment. In some embodiments, the atgRNA further comprises a 5' extension capable of binding one or more consecutive nucleotides of the reverse transcriptase template first and second segment. Such 5' and/or 3' extensions may increase the stability of the guide RNA and/or prevent spurious guide RNA genomic binding. In some embodiments, the stability is increased from cellular exonucleases, endonucleases, or hydrolysis.

6.7.4. Additional Embodiments for atgRNA

[0131] The present disclosure contemplates a split atgRNA comprised of two or more polynucleotides that are capable of forming a atgRNA complex (see Liu et al; doi:

10.1038/s41587-022-01255-9). In some embodiments, the atgRNA may be a atgRNA complex, wherein the atgRNA complex is comprised of at least two polynucleotide components. In certain embodiments, the atgRNA complex is comprised of a first polynucleotide component and a second polynucleotide component. In certain embodiments, the first polynucleotide component comprises a spacer sequence complementary to a target first genomic and a scaffold capable of binding to a DNA binding nickase, and wherein the second polynucleotide component comprises a reverse transcriptase template comprised of at least a first segment and a second segment, wherein said first segment encodes a clamp segment complementary to a second genomic site and said second segment encodes an integrase target recognition site, a primer binding site complementary to a third genomic site, and a RNA-protein recruitment domain, and wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence. In some embodiments, the RNA-protein recruitment domain is a MS2 hairpin.

[0132] In some embodiments, at least one of the first polynucleotide component and the second polynucleotide component are circularized. In some embodiments, circularization is mediated by a ribozyme or ligase. In some embodiments, circularization is mediated by covalently or non-covalently linking the 5' and 3' termini.

[0133] In certain embodiments, the atgRNA or the atgRNA complex further comprises one or more of an RNA-protein recruitment domain, RNA-RNA recruitment domain, a transcriptional termination signal, a reverse transcription termination signal, an RNA ribozyme, or a chemical linker.

[0134] In typical embodiments, one or more atgRNA complex is comprised in one or more RNA polynucleotides or DNA polynucleotides.

[0135] In some embodiments, the atgRNA or the atgRNA complex 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.

[0136] In some embodiments, the integrase target recognition site is an Bxb1 attB or attP sequence. In some embodiments, the att-site central dinucleotide is not GT or CA.

[0137] In typical embodiments, one or more atgRNA or the atgRNA complex is comprised in one or more RNA polynucleotides or DNA polynucleotides.

[0138] In some embodiments, the atgRNA or the atgRNA complex 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.

[0139] In some embodiments, the integrase target recognition site is an Bxb1 attB or attP sequence. In some embodiments, the att-site central dinucleotide is not GT or CA.

[0140] In typical embodiments, one or more atgRNA or the atgRNA complex is comprised in one or more RNA polynucleotides or DNA polynucleotides.

6.7.5. Non-limiting Example AtgRNAs

[0141] 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	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGAGCTGCGAG AA*	212
ACTB N-term PBS 13 RT 29 AttB 46 atgRNA with v2 scaffold	GCTATTCTCGCAGCTCACCAgtttgagagctatgctggaaacagcatagcaagttcaaataag gctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATC ATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGAG CTGCGA GAA*	213
ACTB N-term PBS_13_RT_29_with TP901-1 minimal AttB f atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tc**aacttgaaaaagtggcaccgagtcggtgcGAGTCGGTGCGACGAGCGCGGCGA TATCATCATCCATGGcacaattaacatctcaatcaaggtaaa***TGCTTGAGCTGCGA GAA*	214
ACTB N-term PBS_13_RT_29_with TP901-1 minimal AttB rc atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tc**aacttgaaaaagtggcaccgagtcggtgcGAGTCGGTGCGACGAGCGCGGCGA TATCATCATCCATGGagcatttaccttgattgagatgttaattgtg***TGAGCTGCGAGA A*	215
ACTB N-term PBS_13_RT_29_with PhiBT1 minimal AttB f atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tc**aacttgaaaaagtggcaccgagtcggtgcGAGTCGGTGCGACGAGCGCGGCGA TATCATCATCCATGGcaggtttttgacgaaagtgatccagatgatccag***TGAGCTGCG AGAA*	216
ACTB N-term PBS_13_RT_29_with PhiBT1 minimal AttB rc atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tc**aacttgaaaaagtggcaccgagtcggtgcGAGTCGGTGCGACGAGCGCGGCGA TATCATCATCCATGGctggatcatctggatcactttcgtcaaaaacctg***TGAGCTGCG AGAA*	217
ACTB N-term Nicking	GAAGCCGGCCTTGCACATGCgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc	218

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
guide 1 +48 guide		
ACTB N-term PBS_18_RT_16_with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**ATATCATCATCCATGGtaccgttcgtatagcatacattatacgaagttat***TGAGCTGCGAGAATAGCC*	219
ACTB N-term PBS_13_RT_29_with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCCATGGtaccgttcgtatagcatacattatacgaagttat***TGAGCTGCGAGAA*	220
ACTB N-term PBS 13 RT 34 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**TCGACGACGAGCGCGGCGATATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGAGCTGCGAGAA*	221
ACTB N-term PBS 13 RT 26 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GAGCGCGGCGATATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGAGCTGCGAGAA*	222
ACTB N-term PBS 13 RT 23 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**CGCGGCGATATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGAGCTGCGAGAA*	223
ACTB N-term PBS 13 RT 20 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GGCGATATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGAGCTGCGAGAA*	224
ACTB N-term PBS 13 RT 16 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**ATATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGAGCTGCGAGAA*	225
ACTB N-term PBS 18 RT 34 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**TCGACGACGAGCGCGGCGATATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGAGCTGCGAGAATAGCC*	226
ACTB N-term PBS 18 RT 29 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCCATGGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGAGCTGCGAGAATAGCC*	227

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 18 RT 16 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**ATATCATCATCCATGGccggatgatcctgacg acggagaccgccgtcgtcgacaagccggcc***TGAGCTGCGAGAATAGCC*	228
LMNB1 N-term PBS 13 RT 39 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**CTGCCCATCCGCGGCGGCACGGGGG TCGCAGTCGCCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc** *CGGGCGGCGGAGA*	229
LMNB1 N-term PBS 13 RT 34 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**CATCCGCGGCGGCACGGGGGTCGCA GTCGCCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***CGGGC GGCGGAGA*	230
LMNB1 N-term PBS 13 RT 29 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GCGGCGGCACGGGGGTCGCAGTCGC CATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***CGGGCGGCGGA GA*	231
LMNB1 N-term PBS 13 RT 24 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GGCACGGGGGTCGCAGTCGCCATGcc ggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***CGGGCGGCGGAGA*	232
LMNB1 N-term PBS 13 RT 19 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GGGGGTCGCAGTCGCCATGccggatgatc ctgacgacggagaccgccgtcgtcgacaagccggcc***CGGGCGGCGGAGA*	233
LMNB1 N-term PBS 18 RT 39 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**CTGCCCATCCGCGGCGGCACGGGGG TCGCAGTCGCCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc** *CGGGCGGCGGAGACAGCG*	234
LMNB1 N-term PBS 18 RT 34 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**CATCCGCGGCGGCACGGGGGTCGCA GTCGCCATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***CGGGC GGCGGAGACAGCG*	235
LMNB1 N-term PBS 18 RT 29 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GCGGCGGCACGGGGGTCGCAGTCGC CATGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***CGGGCGGCGGA GACAGCG*	236
LMNB1 N-term PBS 18 RT 24 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GGCACGGGGGTCGCAGTCGCCATGcc ggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***CGGGCGGCGGAGACAG CG*	237

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
LMNB1 N-term PBS 18 RT 19 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GGGGGTCGCAGTCGCCATGccggatgatc ctgacgacggagaccgccgtcgtcgacaagccggcc***CGGGCGGCGGAGACAGC*G	238
LMNB1 N-term Nicking guide 1 +46	GCGTGGTGGGGCCGCCAGCGgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc	239
ACTB N-term PBS 13 RT 29 AttB 42 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgatgatcctgacgacggagaccgccgtcgtcgacaagccgg***TGAGCTGCGAGAA*	240
ACTB N-term PBS 13 RT 29 AttB 40 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgatgatcctgacgacggagaccgccgtcgtcgacaagccg***TGAGCTGCGAGAA*	241
ACTB N-term PBS 13 RT 29 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGatgatcctgacgacggagaccgccgtcgtcgacaagcc***TGAGCTGCGAGAA*	242
ACTB N-term PBS 13 RT 29 AttB 36 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtgatcctgacgacggagaccgccgtcgtcgacaagc***TGAGCTGCGAGAA*	243
LMNB1 N-term PBS 13 RT 29 AttB 44 atgRNA v2	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GCGGCGGCACGGGGGTCGCAGTCGC CATGcggatgatcctgacgacggagaccgccgtcgtcgacaagccggc***CGGGCGGCGGAG A*	244
LMNB1 N-term PBS 13 RT 29 AttB 42 atgRNA v2	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GCGGCGGCACGGGGGTCGCAGTCGC CATGgatgatcctgacgacggagaccgccgtcgtcgacaagccgg***CGGGCGGCGGAGA*	245
LMNB1 N-term PBS 13 RT 29 AttB 40 atgRNA v2	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GCGGCGGCACGGGGGTCGCAGTCGC CATGgatgatcctgacgacggagaccgccgtcgtcgacaagccg***CGGGCGGCGGAGA*	246

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
LMNB1 N-term PBS 13 RT 29 AttB 38 atgRNA v2	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GCGGCGGCACGGGGGTCGCAGTCGC CATGatgatcctgacgacggagaccgccgtcgtcgacaagcc***CGGGCGGCGGAGA*	247
NOLC1 N-term PBS 18 RT 29 AttB 46 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**GAACCACGCGGCGAATGCCGGCGT CCGCCccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TCCTCCAGGC AATACGCG*	248
NOLC1 N-term PBS 13 RT 29 AttB 46 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**GAACCACGCGGCGAATGCCGGCGT CCGCCccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TCCTCCAGGCA AT*	249
NOLC1 N-term PBS 13 RT 29 AttB 44 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**GAACCACGCGGCGAATGCCGGCGT CCGCCcggatgatcctgacgacggagaccgccgtcgtcgacaagccggc***TCCTCCAGGCAA T*	250
NOLC1 N-term PBS 13 RT 29 AttB 42 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**GAACCACGCGGCGAATGCCGGCGT CCGCCggatgatcctgacgacggagaccgccgtcgtcgacaagccgg***TCCTCCAGGCAAT*	251
NOLC1 N-term PBS 13 RT 29 AttB 40 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**GAACCACGCGGCGAATGCCGGCGT CCGCCgatgatcctgacgacggagaccgccgtcgtcgacaagccg***TCCTCCAGGCAAT*	252
NOLC1 N-term PBS 13 RT 29 AttB 38 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**GAACCACGCGGCGAATGCCGGCGT CCGCCatgatcctgacgacggagaccgccgtcgtcgacaagcc***TCCTCCAGGCAAT*	253
NOLC1 nicking guide -43	GAGCCGAGCACGAGGGGATACgttttagagctagaaatagcaagttaaaataaggctagtcc gttatcaacttgaaaaagtggcaccgagtcggtgc	254
ACTB N-term PBS 13 RT 20 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GGCGATATCATCATCCATGGatgatcctg acgacggagaccgccgtcgtcgacaagcc***TGAGCTGCGAGAA*	255
ACTB N-term PBS 13 RT 15 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**TATCATCATCCATGGatgatcctgacgacgga gaccgccgtcgtcgacaagcc***TGAGCTGCGAGAA*	256

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 10 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**TCATCCATGGatgatcctgacgacggagaccgcc gtcgtcgacaagcc***TGAGCTGCGAGAA*	257
ACTB N-term PBS 9 RT 20 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GGCGATATCATCATCCATGGatgatcctg acgacggagaccgccgtcgtcgacaagcc***TGAGCTGCG*	258
ACTB N-term PBS 9 RT 15 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**TATCATCATCCATGGatgatcctgacgacgga gaccgccgtcgtcgacaagcc***TGAGCTGCG*	259
ACTB N-term PBS 9 RT 10 AttB 38 atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**TCATCCATGGatgatcctgacgacggagaccgcc gtcgtcgacaagcc***TGAGCTGCG*	260
LMNB1 N-term PBS 13 RT 20 AttB 38 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**CGGGGGTCGCAGTCGCCATGatgatcctg acgacggagaccgccgtcgtcgacaagcc***CGGGCGGCGGAGA*	261
LMNB1 N-term PBS 13 RT 15 AttB 38 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GTCGCAGTCGCCATGatgatcctgacgacgg agaccgccgtcgtcgacaagcc***CGGGCGGCGGAGA*	262
LMNB1 N-term PBS 13 RT 10 AttB 38 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**AGTCGCCATGatgatcctgacgacggagaccgcc gtcgtcgacaagcc***CGGGCGGCGGAGA*	263
LMNB1 N-term PBS 9 RT 20 AttB 38 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**CGGGGGTCGCAGTCGCCATGatgatcctg acgacggagaccgccgtcgtcgacaagcc***CGGGCGGCG*	264
LMNB1 N-term PBS 9 RT 15 AttB 38 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GTCGCAGTCGCCATGatgatcctgacgacgg agaccgccgtcgtcgacaagcc***CGGGCGGCG*	265
LMNB1 N-term PBS 9 RT 10 AttB 38 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**AGTCGCCATGatgatcctgacgacggagaccgcc gtcgtcgacaagcc***CGGGCGGCG*	266

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
SUPT16H N-term PBS 13 RT 24 Bxb1-GT_Initial length	GAGAAGCGGCGTCCGGGGCTAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**TCTTTGTCCAGAGTCACAGCCATAccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***CCCCGGACGCCGC*	267
SRRM2 N-term PBS 13 RT 24 Bxb1 Initial length	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GGCGTCGGCAGCCCGATCCCGTTGccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TACATGGCCCCGT*	268
DEPDC4 N-term PBS 18 RT 24 Bxb1 Initial length	GTGTCAGGTGGGGCGGGGCTAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GCTGGCTCCTCCCCTGGCACCATAccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***CCCCGCCCCACCTGACAC*	269
NES N-term PBS 13 RT 29 Bxb1 Initial length	GAGTGGGTCAGACGAGCAGGAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**CGACTCCTCCCCCATGCAGCCCTCCATCccggatgatcctgacgacggagaccgccgtcgtcgacaagccggcc***TGCTCGTCTGACC*	270
SUPT16H nicking guide - 53	GCAGCCACCCGCTCTCGGCCCgttttagagctagaaatagcaagttaaaataaggctagtccgtatcaacttgaaaaagtggcaccgagtcggtgc	271
SRRM2 N-term nicking guide 1 +87	GTGTAGTCAGGCCGCTCACCCgttttagagctagaaatagcaagttaaaataaggctagtccgtatcaacttgaaaaagtggcaccgagtcggtgc	272
DEPDC4 N-term Nicking guide 1 +59	GCTGACAAGTCTACGGAACCTgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	273
NES N-term Nicking guide 2 +79	GCTCCTCCAGCGCCTTGACCgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	274
HITI_ACTB_guide	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	275
HITI_SUPTH16_guide	AGAAGCGGCGTCCGGGGCTAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	276
HITI_SRRM2_guide	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	277
HITI_NOLC1_guide	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	278
HITI_DEPDC4_guide	TGTCAGGTGGGGCGGGGCTAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	279

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
HITI_NES_guide	AGTGGGTCAGACGAGCAGGAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	280
HITI_LMNB1_guide	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	281
HDR Cas9 ACTB guide	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	275
HDR Cas9 LMNB1 guide	GGGGTCGCAGTCGCCATGGCgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	282
ACTB N-term PBS 13 RT 29 AttB original length atgRNAs for dinucleotides	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGccggatgatcctgacgacggag*XX*cgccgtcgtcgacaagccggcc***TGAGCTGCGAG*AA* *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	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGccggatgatcctgacgacggagACcgccgtcgtcgacaagccggcc***TGAGCTGCGAGAA*	284
ACTB N-term PBS 13 RT 29 atgRNA with AttB 46 CT for multiplexing	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGccggatgatcctgacgacggagAGcgccgtcgtcgacaagccggcc***TGAGCTGCGAGAA*	285
NOLC1 N-term PBS 18 RT 29 atgRNA with AttB 46 GA for multiplexing	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GAACCACGCGGCGAATGCCGGCGT CCGCCccggatgatcctgacgacggagTCcgccgtcgtcgacaagccggcc***TCCTCCAGGCAATACGCG*	286
LMNB1 N-term PBS 18 RT 29 atgRNA with AttB 46	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GCGGCGGCACGGGGGTCGCAGTCGC CATGccggatgatcctgacgacggagCTcgccgtcgtcgacaagccggcc***CGGGCGGCGGAGACAGCG*	287

Table 9		
Description	**Sequence (5'-3')**	**SEQ ID NO:**
AG for multiplexing		
EMX1 Cas9 guide 1	GTCACCTCCAATGACTAGGGgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc	288
EMX1 Cas9 guide 2	GGGCAACCACAAACCCACGAgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc	289
ACTB N-term PBS 13 RT 29 AttB 56 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGctatgccggatgatcctgacgacggagtccgccgtcgtcgacaagccggccctagc***TGAGC TGCGAGAA*	290
ACTB N-term PBS 13 RT 29 AttB 51 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtgccggatgatcctgacgacggagtccgccgtcgtcgacaagccggcccta***TGAGCTGC GAGAA*	291
ACTB N-term PBS 13 RT 29 AttB 46 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGccggatgatcctgacgacggagtccgccgtcgtcgacaagccggcc***TGAGCTGCGAGA A*	292
ACTB N-term PBS 13 RT 29 AttB 41 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCAT GG**ggatgatcctgacgacggagtccgccgtcgtcgacaagccg***TGAGCTGCGAGAA*	293
ACTB N-term PBS 13 RT 29 AttB 36 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtgatcctgacgacggagtccgccgtcgtcgacaagc***TGAGCTGCGAGAA*	294
ACTB N-term PBS 13 RT 29 AttB 31 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGatcctgacgacggagtccgccgtcgtcgaca***TGAGCTGCGAGAA*	295
ACTB N-term PBS 13 RT 29 AttB 26 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGcctgacgacggagtccgccgtcgtcg***TGAGCTGCGAGAA*	296

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 21 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgcGACGAGCGCGGCGATATCATCATCCAT GG**tgacgacggagtccgccgtcg***TGAGCTGCGAGAA*	297
ACTB N-term PBS 13 RT 29 AttB 16 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGacgacggagtccgccg***TGAGCTGCGAGAA*	298
ACTB N-term PBS 13 RT 29 AttB 11 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgacggagtccg***TGAGCTGCGAGAA*	299
ACTB N-term PBS 13 RT 29 AttB 6 GA atgRNA	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGcggagt***TGAGCTGCGAGAA*	300
ACTB N-term PBS_18_RT_34_with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**TCGACGACGAGCGCGGCGATATCA TCATCCATGGtaccgttcgtatagcatacattatacgaagttat***TGAGCTGCGAGAATA GCC*	301
ACTB N-term PBS_18_RT_29_with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtaccgttcgtatagcatacattatacgaagttat***TGAGCTGCGAGAATAGCC*	302
ACTB N-term PBS_13_RT_34_with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**TCGACGACGAGCGCGGCGATATCA TCATCCATGGtaccgttcgtatagcatacattatacgaagttat***TGAGCTGCGAGAA*	303
ACTB N-term PBS_13_RT_16_with_Lo x71_Cre atgRNA	GAAGCCGGCCTTGCACATGCgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc**ATATCATCATCCATGGtaccgttcgtatagc atacattatacgaagttat***TGAGCTGCGAGAA*	304

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term Nicking guide 2 +93 guide	CCCCACGATGGAGGGGAAGAgttttagagctagaaatagcaagttaaaataaggctagtccgt tatcaacttgaaaaagtggcaccgagtcggtgc	305
LMNB1 N-term Nicking guide 2 +87 guide	CCTTCTCCTGGAGCCGCGACgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc	306
ACTB N-term PBS 13 RT 29 AttB 46 N191352_143_72 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGcattatatgttcttacagtatggcggcccggattgtaaaaacatataatg***TGAGCTGCGA GAA*	307
ACTB N-term PBS 13 RT 29 AttB 46 N684346_90_69 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGcgttatagggtattacagtatggcggtcggtactgcaataccctataacg***TGAGCTGCGA GAA*	308
ACTB N-term PBS 13 RT 29 AttB 46 N675015_95_5 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtgtatcattttcatatagttagcacctgcacactatatgaaaatgataca***TGAGCTGCGAG AA*	309
ACTB N-term PBS 13 RT 29 AttB 46 N189929_49_54 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtgtctactatctgtatatgcgacacatgtggcataaagacatagtagaca***TGAGCTGCGA GAA*	310
ACTB N-term PBS 13 RT 29 AttB 46 N203911_45186_6 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGcatcgaccctgacgcatgcggaggcggcgctccatgcgtctgacctcatt***TGAGCTGCGA GAA*	311

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
ACTB N-term PBS 13 RT 29 AttB 46 N687663_53_29 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgttagtacccaaatgacaaaaggtcatccttttatcatttgggtactaac***TGAGCTGCGA GAA*	312
ACTB N-term PBS 13 RT 29 AttB 46 N687611_90_68 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGcttattaaaacccgttccgcttctgtcaaagcggcatcggttttataaac***TGAGCTGCGAG AA*	313
ACTB N-term PBS 13 RT 29 AttB 46 N190156_234_12 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGggcgtgatggtcgtgaacctcaacatgacgacgaacacgacctcgcggcc***TGAGCTGCG AGAA*	314
ACTB N-term PBS 13 RT 29 AttB 46 N191533_224_76 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtctacatcttgaatatatcaagttataactttgaattatatcagtttata***TGAGCTGCGAGA A*	315
ACTB N-term PBS 13 RT 29 AttB 46 N208621_9_15 integrase	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGaattatatctaaaagcactaagctccgccatactgcttttagatataata***TGAGCTGCGA GAA*	316
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH187_38 bp_Att	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgatatggggaagtgaatcagtacaaccgccacagtacc***TGAGCTGCGAGAA*	317
ACTB N-term PBS 13 RT 29 AttB 46	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgtactgtggcggttgtactgattcacttccccatatc***TGAGCTGCGAGAA*	318

Table 9		
Description	**Sequence (5'-3')**	**SEQ ID NO:**
Bacillus_cereus_AH187_38 bp_Att_rc		
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugdunensis_N920143_38bp_Att	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtgggtggtacaggtgccacattagttgtaccatttatg***TGAGCTGCGAGAA*	319
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugdunensis_N920143_38bp_Att_rc	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGcataaatggtacaactaatgtggcacctgtaccaccca***TGAGCTGCGAGAA*	320
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_38bp_Att	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgttgtttttccagatccagttggtcctgtaaatataag***TGAGCTGCGAGAA*	321
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_38bp_Att_rc	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGcttatatttacaggaccaactggatctggaaaaacaac***TGAGCTGCGAGAA*	322
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH187_Att 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgtactgtggcggttgtactgattcacttccccatat***TGAGCTGCGAGAA*	323

Table 9		
Description	**Sequence (5'-3')**	**SEQ ID NO:**
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH187_Att 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCCATGGtactgtggcggttgtactgattcacttccccata***TGAGCTGCGAGAA*	324
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH187_Att 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCCATGGactgtggcggttgtactgattcacttccccat***TGAGCTGCGAGAA*	325
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH187_Att _rc 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCCATGGatatggggaagtgaatcagtacaaccgccacagtac***TGAGCTGCGAGAA*	326
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH187_Att _rc 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCCATGGtatggggaagtgaatcagtacaaccgccacagta***TGAGCTGCGAGAA*	327
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cereus_AH187_Att _rc 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCCATGGatggggaagtgaatcagtacaaccgccacagt***TGAGCTGCGAGAA*	328
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugdunensis	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCCATGGataaatggtacaactaatgtggcacctgtaccaccc***TGAGCTGCGAGAA*	329

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
_N920143_ Att 36bp		
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugdunensis _N920143_ Att 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtaaatggtacaactaatgtggcacctgtaccacc***TGAGCTGCGAGAA*	330
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugdunensis _N920143_ Att 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGaaatggtacaactaatgtggcacctgtaccac***TGAGCTGCGAGAA*	331
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugdunensis _N920143_ Att_rc 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgggtggtacaggtgccacattagttgtaccatttat***TGAGCTGCGAGAA*	332
ACTB N-term PBS 13 RT 29 AttB 46 Staphylococcus_lugdunensis _N920143_ Att_rc 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGggtggtacaggtgccacattagttgtaccattta***TGAGCTGCGAGAA*	333
ACTB N-term PBS 13 RT 29 AttB 46	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgtta tcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGgtggtacaggtgccacattagttgtaccattt***TGAGCTGCGAGAA*	334

Table 9		
Description	**Sequence (5'-3')**	**SEQ ID NO:**
Staphylococcus_lugdunensis_N920143_Att_rc 32bp		
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_Att 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGttatatttacaggaccaactggatctggaaaaacaa***TGAGCTGCGAGAA*	335
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_Att 34bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtatatttacaggaccaactggatctggaaaaaca***TGAGCTGCGAGAA*	336
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_Att 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGatatttacaggaccaactggatctggaaaaac***TGAGCTGCGAGAA*	337
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_Att_rc 36bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGttgtttttccagatccagttggtcctgtaaatataa***TGAGCTGCGAGAA*	338
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCC ATGGtgtttttccagatccagttggtcctgtaaatata***TGAGCTGCGAGAA*	339

Table 9		
Description	**Sequence (5'-3')**	**SEQ ID NO:**
98_Att_rc 34bp		
ACTB N-term PBS 13 RT 29 AttB 46 Bacillus_cytotoxicus_NVH_391-98_Att_rc 32bp	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GACGAGCGCGGCGATATCATCATCCATGGgtttttccagatccagttggtcctgtaaatat***TGAGCTGCGAGAA*	340
Bacillus_cereus_AH187_Att_rc_36 LMNB1 PBS 9 RT 10 AttB 36 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**AGTCGCCATGatatggggaagtgaatcagtacaaccgccacagtac***CGGGCGGCG*	341
Bacillus_cereus_AH187_Att_rc_36 NOLC1 PBS 18 RT 29 AttB 36 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GAACCACGCGGCGAATGCCGGCGTCCGCCatatggggaagtgaatcagtacaaccgccacagtac***TCCTCCAGGCAATACGCG*	342
Bacillus_cereus_AH187_Att_rc_36 SUPT16H PBS 13 RT 24 AttB 36 atgRNA	GAGAAGCGGCGTCCGGGGCTAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**TCTTTGTCCAGAGTCACAGCCATAatatggggaagtgaatcagtacaaccgccacagtac***CCCCGGACGCCGC*	343
Bacillus_cereus_AH187_Att_rc_36 SRRM2 PBS 13 RT 24 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GGCGTCGGCAGCCCGATCCCGTTGatatggggaagtgaatcagtacaaccgccacagtac***TACATGGCCCCGT*	344

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
Bacillus_cereus_AH187_Att_rc_36 DEPDC4 PBS 18 RT 24 AttB 36 atgRNA	GTGTCAGGTGGGGCGGGGCTAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GCTGGCTCCTCCCCTGGCACCATAatatggggaagtgaatcagtacaaccgccacagtac***CCCCGCCCCACCTGACAC*	345
Bacillus_cereus_AH187_Att_rc_36 NES PBS 13 RT 28 AttB 36 atgRNA	GAGTGGGTCAGACGAGCAGGAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**CGACTCCTCCCCCATGCAGCCCTCCATCatatggggaagtgaatcagtacaaccgccacagtac***TGCTCGTCTGACC*	346
B. cereus LMNB1_PBS 9 RT 20 AttB 36 atgRNA	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatca acttgaaaaagtggcaccgagtcggtgc**CGGGGGTCGCAGTCGCCATGatatggggaagtgaatcagtacaaccgccacagtac***CGGGCGGCG*	347
B. cereus LMNB1_PBS 13 RT 20 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcCGGGGGTCGCAGTCGCCATG**atatggggaagtgaatcagtacaaccgccacagtac***CGGGCGGCGGAGA*	348
B. cereus LMNB1_PBS 13 RT 29 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GCGGCGGCACGGGGGTCGCAGTCGCCATGatatggggaagtgaatcagtacaaccgccacagtac***CGGGCGGCGGAGA*	349
B. cereus NOLC1_PBS 13 RT 29 AttB 36 atgRNA	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcGAACCACGCGGCGAATGCCGGCGTCCGCC**atatggggaagtgaatcagtacaaccgccacagtac***TCCTCCAGGCAAT*	350
B. cereus NOLC1_PBS 13 RT 20 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GGCGAATGCCGGCGTCCGCCatatgggaagtgaatcagtacaaccgccacagtac***TCCTCCAGGCAAT*	351
B. cereus NOLC1_PBS 18 RT 20 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GGCGAATGCCGGCGTCCGCCatatgggaagtgaatcagtacaaccgccacagtac***TCCTCCAGGCAATACGCG*	352

Table 9

Description	Sequence (5'-3')	SEQ ID NO:
B. cereus SRRM2_PBS 9 RT 24 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GGCGTCGGCAGCCCGATCCCGTTGatatggggaagtgaatcagtacaaccgccacagtac***TACATGGCC*	353
B. cereus SRRM2_PBS 9 RT 10 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GATCCCGTTGatatggggaagtgaatcagtacaaccgccacagtac***TACATGGCC*	354
B. cereus SRRM2_PBS 13 RT 10 AttB 36 atgRNA	GGGCACGGGGCCATGTACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc**GATCCCGTTGatatggggaagtgaatcagtacaaccgccacagtac***TACATGGCCCCCGT*	355
Screen validation guides ACTB_1_11_24_38	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcgcgcggcgatatcatcatccatggatgatcctgacgacggagaccgccgtcgtcgacaagcctgagctgcgag	356
Screen validation guides ACTB_1_16_18_43	GCTATTCTCGCAGCTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgccgatatcatcatccatggcggatgatcctgacgacggagaccgccgtcgtcgacaagccggctgagctgcgagaatag	357
Screen validation guides LMNB1_1_8_26_38	GCTGTCTCCGCCGCCCGCCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcgcggcacgggggtcgcagtcgccatgatgatcctgacgacggagaccgccgtcgtcgacaagcccgggcggc	358
Screen validation guides NOLC1_1_15_16_43	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcaatgccggcgtccgcccggatgatcctgacgacggagaccgccgtcgtcgacaagccggctcctccaggcaatac	359
Screen validation guides NOLC1_1_14_10_38	GCGTATTGCCTGGAGGATGGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcggcgtccgccatgatcctgacgacggagaccgccgtcgtcgacaagcctcctccaggcaata	360

Table 9		
Description	**Sequence (5'-3')**	**SEQ ID NO:**
Screen validation guides SERPIN_13_32_38	GGGAAATGCATCTTGCACAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcagcccctccatgctctctagctgttgccattgggcttgtcgacgacggcggtctccgtcgtcaggatcattgcaagatgcatt	361
Screen validation guides DEPDC4_8_10_38	GTGTCAGGTGGGGCGGGGCTAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctggcaccataatgatcctgacgacggagaccgccgtcgtcgacaagcccccccgccc	362
SERPIN Nicking guide -107 guide	GTGGGGACAGCCCCGTCTCTgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	363
SERPIN Nicking guide -91 guide	GCTCTTGGGAAAAAAACCCTAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	364
SERPIN Nicking guide -90 guide	GTCTTGGGAAAAAAACCCTAAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	365
SERPIN Nicking guide -84 guide	GAAAAAAACCCTAAGGGCTGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	366
SERPIN Nicking guide -67 guide	GCTGAGGATCCTTGTGAGTGTgttttagagctagaaatagcaagttaaaataaggctagtccgtatcaacttgaaaaagtggcaccgagtcggtgc	367
SERPIN Nicking guide -66 guide	GTGAGGATCCTTGTGAGTGTTgttttagagctagaaatagcaagttaaaataaggctagtccgtatcaacttgaaaaagtggcaccgagtcggtgc	368
SERPIN Nicking guide -63 guide	GGATCCTTGTGAGTGTTGGGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	369
SERPIN Nicking guide -62 guide	GATCCTTGTGAGTGTTGGGTgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgc	370

Table 9		
Description	**Sequence (5'-3')**	**SEQ ID NO:**
SERPIN Nicking guide -49 guide	GTTGGGTGGGAACAGCTCCCgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc	371
SERPIN Nicking guide -46 guide	GGGTGGGAACAGCTCCCAGGgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc	372
SERPIN Nicking guide +34 guide	GCTTCTGTGCAGCAGTTTCCCgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc	373
SERPIN Nicking guide +48 guide	GTTTCCCTGGCCACTAAATAGgttttagagctagaaatagcaagttaaaataaggctagtccgt tatcaacttgaaaaagtggcaccgagtcggtgc	374
SERPIN Nicking guide +49 guide	GTTCCCTGGCCACTAAATAGTgttttagagctagaaatagcaagttaaaataaggctagtccgt tatcaacttgaaaaagtggcaccgagtcggtgc	375
SERPIN Nicking guide +71 guide	GATTAGATAGAAGCCCTCCAgttttagagctagaaatagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgc	376
SERPIN Nicking guide +72 guide	GATTAGATAGAAGCCCTCCAAgttttagagctagaaatagcaagttaaaataaggctagtccg ttatcaacttgaaaaagtggcaccgagtcggtgc	377

6.8. Systems and Methods Comprising an atgRNA having a Clamp

[0142] This disclosure also features systems for site-specifically incorporating an integration recognition site into a mammalian cell genome at a desired target site (e.g., a double stranded target DNA sequence), comprising: any of the attachment site containing guide RNAs (atgRNAs) described in Section 6.7; any of the gene editor polypeptides described herein, for example, gene editor polypeptides comprising a DNA binding nickase domain linked to a reverse transcriptase domain that is capable of incorporating the integration recognition site into the target DNA sequence, wherein, upon introducing into the cell, the system integrates the integration recognition site into the target DNA sequence. In some embodiments, upon introducing the system into the

cell, the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to the integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment.

[0143] This disclosure also features methods for site-specifically incorporating an integration recognition site into a mammalian cell genome at a desired target site (e.g., a double stranded target DNA sequence), comprising: incorporating an integration target recognition site into the genome by delivering into the cell any of the attachment site containing guide RNAs (atgRNAs) described in Section 6.7; and any of the gene editor polypeptides described herein or any of the polynucleotides encoding a gene editor polypeptide, for example, a gene editor polypeptide comprising a DNA binding nickase domain linked to a reverse transcriptase domain, wherein the integration recognition site is incorporated into the target DNA sequence by the gene editor polypeptide; and optionally, a nicking gRNA. In some embodiments, the method using the atgRNA comprising the clamp segment enhances integration efficiency of the integration recognition site at the target DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment.

[0144] This disclosure also features a system for site-specifically integrating a donor polynucleotide template into a mammalian cell genome at a desired target site (e.g., a double stranded target DNA sequence), comprising: any of the attachment site containing guide RNAs (atgRNAs) described in Section 6.7; any of the gene editor polypeptides described herein, for example, a gene editor polypeptide comprising a DNA binding nickase domain linked to a reverse transcriptase domain that is capable of incorporating the integration recognition site into the target DNA sequence; an integration enzyme; and an donor polynucleotide template linked to a sequence that is an integration cognate of the integration recognition site present in the atgRNA, whereby the integration enzyme integrates the donor polynucleotide template into the target DNA sequence. In some embodiments of the system, upon introducing the system into the cell, the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to the integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment.

[0145] This disclosure also features a method for site-specifically integrating a donor polynucleotide template into a genome of a cell, comprising: incorporating an integration target

recognition site into the genome by delivering into the cell: any of the attachment site containing guide RNAs (atgRNAs) described in Section 6.7; and any of the gene editor polypeptides described herein or any of the polynucleotides encoding a gene editor polypeptide, for example, a gene editor polypeptide comprising a DNA binding nickase domain linked to a reverse transcriptase domain that is capable of incorporating the integration recognition site into the target DNA sequence; and optionally, a nicking gRNA; and integrating the donor polynucleotide template into the genome by delivering into the cell: an integration enzyme; and a donor polynucleotide template, wherein the donor polynucleotide template is linked to a sequence that is an integration cognate of the integration recognition site present in the atgRNA, wherein the donor polynucleotide is integrated into the cellular genome at the incorporated genomic integration recognition site by the integration enzyme. In some embodiments, the method using the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment. In some embodiments, the method using the atgRNA comprising the clamp enhances integration efficiency of the donor polynucleotide template site at the target DNA sequence as compared to the integration efficiency of the donor polynucleotide template at the target DNA sequence when using an atgRNA that does not comprise the clamp segment. In some embodiments, the atgRNA, the gene editor polypeptide (or the polynucleotide encoding the gene editor polypeptide), the integration enzyme, and the donor polynucleotide template are introduced into the cell concurrently.

6.9. Integrases/Recombinases and Integration/Recombination Sites

[0146] In typical embodiments, a gene editor polypeptide described herein contains an integrase or recombinase. In some embodiments, the integrase is delivered as a protein or the integrase is encoded in a delivered polynucleotide. In some embodiments, the integration enzyme is selected from the group consisting of Dre, Vika, Bxb1, φC31, RDF, φBTl, R1, R2, R3, R4, R5, TP901-1, A118, φFCl, φC1, MR11, TG1, φ370.l, Wβ, 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, φRV, 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, φC31, φBT1, Bxb1, SPBc, TP901-1 and Wβ 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.

[0147] 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 disclosure 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 disclosure.

[0148] In typical embodiments, the integrase used herein is selected from below (Table 10).

Table 10 - Integrases

Database	SRA accession	bioproject_acc	nucleotide accession	protein accession or ORF ID	internal protein ID	Proposed names	Alternative names	organism/ source	description	Sequence	SEQ ID NO:	Length	Group
ENA	SRS12052 98	PRJEB2 6277	NA	NA	N18992 9_49_5 4	SsuINT	NA	human gutmetag enome	stool sample from male in USA	MEKNRAVLYLRLSKEDVDK VNKGDDSSSIKSQRLLLTDF ALERGFKIVGVYSDDDESGL YDDRPDFERMMTDAKLDEF DIIIAKTQSRFSRNMEHIEKY LHHDLPNLGIRFIGAVDGVD TESDENKKSRQINGLVNEW YCEDLSKNIRSAFKAKMKD GQFLGSSCPYGYKKDPQNH NHLVVDDYAAKVVQKIFNL YLEGYGKAKIGSILSSEGILIP TLYKKDILKQNYHNSKALD TTQNWSYQTIHTILNNEVYL GHLIQNKVNTMSYKDKNKR ILPKEKWIIVRNTHEPIITEE MFQDVQKLQKNRTRSVENI EPNGLFSGLIFCADCKHAMS RKYARRGEKGFVGYVCKTY KTQGKNFCESHSIDYDELEE AVLFSIKNEARSILQQEEIDE LRKVQAYDETKSYYEMQLE NIKSRMEKIEKYKKKTYDN YMDDLISRDDYKKYVTEYD KEIGGLKQQQELINSKTDLE KEISTQYDEWVEAFINYVDI DKLTREIVIELIEKIEVNKDG SINIYYKFKNPYIS	378	527	INTc
ENA	ERS3964 61	PRJEB2 6280	NA	NA	N19015 6_234_ 12	SssINT	NA	human gut metageno me	stool sample from Spain	MNTVIYARYSAGPRQTDQSI DGQLRVCTEFCKQRGLTVV DTYCDRHISGRTDERPEFQR LIADAKAHKFEAVVVYKTD RFARNKYDSAIYKRELRRN GIQIFYAAEAIPEGPEGIILES LMEGLAEYYSAELAQKIKR GLNESALKCQSLGSGRPLGY TVDEQKHFQIDPESSQAVKT IFEMYIKGESNAAICDYLNA RGLRTSQGNLFNKNSINRIIK NRKYIGEYRYNDIVVEGGM PAIISKETFCMAQAEMERRR THRAPVSPKAEYLLAGKLFC GHCKGPMQGVSGTGKSGN KWYYYYCANTRGKERTCD KKQVSRDRLEKAVVDFTVR YILQENVLEELSKKVYAAQE RQNNTASEIAFYEKKLAENK KAIANILRAIESGAMTQALP ARLQELENEQTVIQGELSYL KGARLAFTEDQILFALLQHL DPRPGESERDYHRRIITDFVS EVYLYDDRMLIYFNISSADG KLKHADLSAIESGVFDAGLI SSSSRASSFSTRCALI	379	510	INTd

ENA	ERS1015837	PRJEB26832	NA	NA	N191352_143_72	SscINT	NA	human gutmetagenome	stool sample from China	MNEKNLEIGAAYIRVSTDDQTELSPDAQLRVILEAAKKDGIIIPQEFVFMEDRGRSGRRADNRPEFQRMISTARQNPSPFRYLYLWKFSRFARNQEESAFYKGILRKKCGVTIKSVSEPIMEGMFGRLVEMIIEWSDEFYSVNLSGEVLRGMTQKALEHGYQLTPCLGYDAVGHGRPYVINEEQYQIVEFIHRSFFDGKDMTWIAREANRRGYHTRRGNPFDTRAVRIILTNSFYVGLVKWNDVTFQGTHECRESVTSVFSANQERLNRIHRPRGRRQASSCKHWLSGLLKCSICGASLGYNQTKDLTKRGHAFQCWKYTKGIHPGSCSVSSLKAEAAVLESLQMILETGEVEYTYEQREKHLDDNKLTLIQKSLERLDTKELRIREAYESGIDTLDEFKTNKARLQRERDQLMEELEELHSQEEPEDVPGKEILIERIQNVYDLLQSPDVDNDDKGNAVRSIIKKIVYIKESKTFCFYYYV	380	482	INTd
ENA	ERS1289677	PRJEB26924	NA	NA	N191533_224_76	Ssc2INT	NA	human gutmetagenome	stool sample from China	MERTIKVIQPGTVKIPTKKRVAAYARVSSGKDAMLHSLSAQVSYYSNMIQQKNEWSYVGIYADEAITGTKDRRVEFNRLIQDCTDGKIDMIITKSISRFARNTLTMLEVVRKLKNINVDVYFEKENIHSISGDGELMLTILASFAQEESRSVSENCKWRIRKGFEQGELINLRFLYGYRINKGKIEIYEKEAEIVRMIFDDYLNGEGCTRIGNKLRKMKVNKLRGGMWNSERVVDIIKNEKYTGNALLQKKYVKDHLSKKLVRNKGILTQYYAEGTHPAIIDIKTFEIAQKIMEANRTKFQGKCGSNRYLFTSKIECGICGKNYRHKDREGKSTWVCANHLKYGNSRCIAKPLNEEKLKKLINEALELKYFDEEIFIRNIKRIKVTGNQTIEFILKDGKVIEEGMI	381	406	INTc
ENA	ERS2655827	PRJEB28245	NA	NA	N203911_45186_6	SsdINT	NA	human gut metagenome	stool sample from Denmark	MKKIKIDRAIQERPATRKQTRNEKIRQSLTEHVDVQVIPAITDREGYEKPKLRVCAYCRVSTDMDTQALSYELQVQNYTDYIRGNDEWRFAGIYADRGISGTSLKHRDEFNRMIEDCKAGKIDLIITKAVTRFARNVLDCISTIRMLKQLEHPVAVYFETERINTLDTTSETYLGLISLFAQGESESKSESLKWSYIRRWKRGTGIYPAWSLLGYEMGEDGKWQIVEAEAELVRIIYDMYLNGYSSPQIAEILTRSGVPTATNQTVWSSGGVLGILRNEKYCGNVLCQKTMTVDVFSHKAIKNTGQKTQYFIEGHHDPIILRSDWDRVQQMIDEKYYRKRRGRRTKPRIVLKGCLAGFTQIDLDWDEDDIARIFYSTTPAAEVATPAMADHIEIIKVKGEN	382	401	INTc

ENA	SRS 294942	PRJ EB30046	NA	NA	N208621_9_15	SmcINT	NA	human gut metagenome	sample from 72-year-old male from China	MKTAAAYIRVSTDDQVEYS PDSQIKLIRDYAKRNDYILP DEFIFRDDGISGKSAKHRPEF TKMIALAKSPEHPFDAILVW KFSRFARNQEESIVFKNILRK IGVEVRSVSEPISEDPFGSLV ERIIEWTDEYYIINLSGEVKR GMLEKISRGQPVVPPPVGYK MENGQYIPDENAHFIKEIFE AYAAGEGARHIAQRLAAQG CLTKRGNPIDNRFVDYVLH NPVYIGKLRWSVNSHAASS RHYDSADIIVFDGTHEPLISS ELWESVQKRLHEVKTLYPK YQRREQPVSFMLKGLVRCS SCGSTLCYCRTSEPSLQCHS YARGSCRQSHSINIATANEA VIKGLQLAVDKLDFAIAPAK PHYSADAPGTNKLLAAEYK KMERIKAAYANGTDTLEEY AANKKKISAEIARLEAELQQ ESNVKPINKKAFAKRVSEIIK YISDPHNSEAAKNQALRTVI SYIIFDRAATTFNIIFHF	383	476	INTd
MetaSUB	NA	NA	NA	NA	N675015_95_5	UhmINT	NA	urban human microbiome	NA	MKIAIYARKSKYSPTGESVE NQIQLCKEYLQAKYKSETLE IDEYKDEGYSGGNTNRPDF KKLIAQIEDYDMLICYRLDR ISRNVADFSSTLTLLQNNKC DFVSIKEQFDTTSPMGRAMI YISSVFAQLERETIAERIRDN MMELAKMGRWLGGTIPMG FDSEPITFIDENMKERSMTK LIPNVEELKVIELIYEKYLQL GSMGKVVTYLLQNNIKTKK GKDFTLGSIKVILTNPIYVKA NQEVVNHLKTQGITICGDV DGKKALLTYNKTTGISNDV GTKTIVKDKSEWIAAVANH KGIIPADKWLQAQNIKDKN KDSFPALGRSNTTIASRVLR CDKCESTMGVTHGHINPVT GKKHYYYNCTLKKRSKGV RCDNKPAKAAEVDEAILITL ENMFKAKSSIIDNLKAKNKA RRIEMISSNRVDVINKIIEDK TKQIDNLVNKLSLDDDLTDI LFKKIKGLKAEIKELEDELLT LTSDNIKLNEDEVVLDFTEK LLEKCSIIRTLDILEQQQIVD ALIPLVTWNGDTEVLNIYPL GSPELELKEAESKKK	384	550	INTd

Segata-Pasolli	NA	PRJNA422434	NA	NA	N684346_90_69	SacINT	NA	human gut metagenome	stool sample from adult in China	MKEKVSERKTGAIYIRVSTDKQEELSPDAQLRLLLDYAKKDSIDVPKEYIFQDNGISGRKANKRPAFQNMIALAKSKEHPIDTIIVWKFSRFARNQEESIVYKSLLKKNNVDVVSVSEPLIDGPFGSLIERIIEWMDEYYSIRLSGEVMRGMTQNAMRGHYQSDAPIGYTSPGDKKPPVINPDTVQIPLMIKDMFLSGSTQLQIARKLNDSGYRTKRGNLWDARGVRYVLENPFYIGKSRWNYTERGRRLKPADEVIYADGNWEALWDEDTFKEIQKRLALNMRKSKSRDISAAKHWLSGLLICSSCGGTLAFGGAHNMRGFQCWKYSKGFCSESHYISTGPIEKMVLEYLEAVMHSPALSYTVISSSSVDASSKLSDLERQLQKIDAKEKRIKAAYLNEIDTLEEYKANKTALEEERRTVEKEIEELTLSDVKYSKEDLDKKMKQNISDLLRVLRDESADYIQKGNMMRNVVDHIVFNRKNTSLDVFLKLVV	385	493	INTd
Segata-Pasoli	ERR1136864	PRJEB11532	NA	NA	N687611_90_68	RsaINT	NA	human gut metagenome	rectal swab from adult in Isreal	MKITKKQPLRPRGRSEDKRQSTKNVIRDAYINGPQKEVQIIPAKRDMEAETEKKKLRVCAYCRVSTDEDTQASSYELQVQNYTRMIRENPEWEFAGIFADEGISGTSVLHREHFLEMIEKCKAGEIDLIITKQVSRFARNVLDSLNYIFMLRKLDPPVGVYFETEKLNTLDKSSDMVITVLSLVAQSESEQKSNSLKWSFKRRRAQGLGIYPSWALLGYRLDDEKNWEIVEDEADIVRTIYSLYLDGYSSTQIAELLTKSGIPTVKGLSVWSSGSVLGILKNEKFCGDALCQKTVTIDFFTHKSVKNNGIEPQYFVEGHHIPIIEKNDWLLAQQIRKERYRKRRSTHRKPRIVVKGALSGFMIVDTSWDEEYVDSLLISATQKPEPAPVIAEEDENFIVIEKE	386	404	INTc
Segata-Pasolli	ERR1136737	PRJEB11532	NA	NA	N687663_53_29	Rsa2INT	NA	human gut metagenome	rectal swab from adult in Isreal	MADIQPVKNGALYIRVSTHLQEELSPDAQKRLLMEYAEAHNIIVLKEHIYIDSGISGRSARQRPQFNNMIAEAKSKEHPFDVILVWKYSRFARNQEESIVYKSMLKRENVDVISVSEPISDDPFGSLIERIIEWMDEYYSIRLSGEVSRGMAENAMRGNYQARPPLGYRIPGYRQTPVIVPEEAELIQLIFDLYTEKKMGIFEIVRYLNEHGYQTGHKKPFQRRSVTYILKNPTYIGKTIWNQHDQDHKLRDKSEWIIADGKHEPIISKEQFDKAQKRIESTYKPAYRKPTSVCHHWLSSLLKCSSCGRTLVVKRTASKKKDRMYVNFQCYGYQKGICNTNQSISAIKLEPVIMHALEDAMTSGKIHFDVLNPTTLDSSQKQQFLTRLNEIEKKEERIKRAYRDGIDTLEEYKENKSIIQTEKEMLLKKIEHIEEPALSPEEAKPIMMDRIKNVYEIITNPDIGMEEKNKAARSIIEKIVFDRATGSVNIFFYLAHCP	387	498	INTd

NCBI	NA	NA	NC_002656.1	NP_075302.1	NA	BxbINT	Bxb1 integrase	Mycobacterium phage Bxb1	NA	MRALVVIRLSRVTDATTSPERQLESCQQLCAQRGWDVVGVAEDLDVSGAVDPFDRKRRPNLARWLAFEEQPFDVIVAYRVDRLTRSIRHLQQLVHWAEDHKKLVVSATEAHFDTTTPFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRGSLPPWGYLPTRVDGEWRLVPDPVQRERILEVYHRVVDNHEPLHLVAHDLNRRGVLSPKDYFAQLQGREPQGREWSATALKRSMISEAMLGYATLNGKTVRDDDGAPLVRAEPILTREQLEALRAELVKTSRAKPAVSTPSLLLRVLFCAVCGEPAYKFAGGGRKHPRYRCRSMGFPKHCGNGTVAMAEWDAFCEEQVLDLLGDAERLEKVWVAGSDSAVELAEVNAELVDLTSLIGSPAYRAGSPQREALDARIAALAARQEELEGLEARPSGWEWRETGQRFGDWWREQDTAAKNTWLRSMNVRLTFDVRGGLTRTIDFGDLQEYEQHLRLGSVVERLHTGMS*	388	501	INTa
NCBI	NA	NA	NC_002747.1	NP_112664.1	NA	Tp9INT	TP901-1integrase	Lactococcus phage TP901-1	NA	MTKKVAIYTRVSTTNQAEEGFSIDEQIDRLTKYAEAMGWQVSDTYTDAGFSGAKLERPAMQRLINDIENKAFDTVLVYKLDRLSRSVRDTLYLVKDVFTKNKIDFISLNESIDTSSAMGSLFLTILSAINEFERENIKERMTMGKLGRAKSGKSMMWTKTAFGYYHNRKTGILEIVPLQATIVEQIFTDYLSGISLTKLRDKLNESGHIGKDIPWSYRTLRQTLDNPVYCGYIKFKDSLFEGMHKPIIPYETYLKVQKELEERQQQTYERNNNPRPFQAKYMLSGMARCGYCGAPLKIVLGHKRKDGSRTMKYHCANRFPRKTKGITVYNDNKKCDSGTYDLSNLENTVIDNLIGFQENNDSLLKIINGNNQPILDTSSFKKQISQIDKKIQKNSDLYLNDFITMDELKDRTDSLQAEKKLLKAKISENKFNDSTDVFELVKTQLGSIPINELSYDNKKKIVNNLVSKVDVTADNVDIIFKFQLA*	389	486	INTd

NCBI	NA	NA	NC_004664.2	NP_813744.2	NA	Bt1INT	PhiBTintegrase	Streptomyces virus phiBT1	NA	MSPFIAPDVPEHLLDTVRVF LYARQSKGRSDGSDVSTEA QLAAGRALVASRNAQGGA RWVVAGEFVDVGRSGWDP NVTRADFERMMGEVRAGE GDVVVVNELSRLTRKGAHD ALEIDNELKKHGVRFMSVL EPFLDTSTPIGVAIFALIAAL AKQDSDLKAERLKGAKDEI AALGGVHSSSAPFGMRAVR KKVDNLVISVLEPDEDNPDH VELVERMAKMSFEGVSDNA IATTFEKEKIPSPGMAERRAT EKRLASIKARRLNGAEKPIM WRAQTVRWILNHPAIGGFA FERVKHGKAHINVIRRDPGG KPLTPHTGILSGSKWLELQE KRSGKNLSDRKPGAEVEPTL LSGWRFLGCRICGGSMGQS QGGRKRNGDLAEGNYMCA NPKGHGGLSVKRSELDEFV ASKVWARLRTADMEDEHD QAWIAAAAERFALQHDLAG VADERREQQAHLDNVRRSI KDLQADRKAGLYVGREELE TWRSTVLQYRSYEAECTTR LAELDEKMNGSTRVPSEWF SGEDPTAEGGIWASWDVYE RREFLSFFLDSVMVDRGRHP ETKKYIPLKDRVTLKWAEL LKEEDEASEATERELAAL*	390	595	INTa
NCBI	NA	NA	NC_011658.1	WP_000286206.1	NA	BceINT	NA	Bacillus cereus AH187	NA	MYPYDVPDYAGSYRPESLD VCIYLRKSRKDVEEERRAIE EGSSYNALERHRKRLFAIAK AENHNIIDIFEEVASGESIQE RPQMQQLLRKLEGNEIDGV LVIDLDRLGRGDMLDAGMI DRAFRYSSTKIITPTDVYDPD DESWELVFGIKSLISRQELKS ITKRLQNGRIDSVKEGKHIG KKPPYGYLKDENLRLYPDP EKAWIVKKIFELMCDGKGR QMIAAELDRLGIDPPVTKRG AWDSSTITSIIKNEVYTGVIV WGKFKHKKRNGKYTRHKN PQEKWIMYENAHEPIISKEL FDAANEAHSSRHKPAVITSK KLTNPLAGILKCKLCG YTMLIQTRKDRPHNYLRCN NPACKGKQKQSVFNLVEEK LLYSLQQIVDEY QAQKVEEVEIDDSKLISFKE KAIISKE KELKELQAQKGNLHDLLEQ GIYTVE IFLERQKNLVERITSIENDIEV LQKEIE TEQIKEHNKTEFIPALKTVIE SYHKTT NIELKNQLLKTILSTVTYYR HPDWKT NEFEIQVYFKIS*	391	529	INTc

NCBI	NA	NA	NC_009674.1	WP_012095429.1	NA	BcyINT	NA	Bacillus cytotoxicus NVH391-98	NA	MYPYDVPDYAGSAVGIYIR VSTQEQASEGHSIESQKKKL ASYCEIQGWDDYRFYIEEGI SGKNTNRPKLKLLMEHIEK GKINILLVYRLDRLTRSVIDL HKLLNFLQEHGCAFKSATET YDTTTANGRMSMGIVSLLA QWETENMSERIKLNLEHKV LVEGERVGAIPYGFDLSDDE KLVKNEKSAILLDMVERVE NGWSVNRIVNYLNLTNNDR NWSPNGVLRLLRNPALYGA TRWNDKIAENTHEGIISKER FNRLQQILADRSIHHRRDVK GTYIFQGVLRCPVCDQTLSV NRFIKKRKDGTEYCGVLYR CQPCIKQNKYNLAIGEARFL KALNEYMSTVEFQTVEDEVI PKKSEREMLESQLQQIARKR EKYQKAWASDLMSDDEFE KLMVETRETYDECKQKLES CEDPIKIDETYLKEIVYMFH QTFNDLESEKQKEFISKFIRT IRYTVKEQQPIRPDKSKTGK GKQKVIITEVE FYQS*	392	487	INTd
NCBI	NA	NA	NC_017353.1	WP_014533238.1	NA	SluINT	NA	Staphylococcus lugdunensis N920143	NA	MYPYDVPDYAGSKVAIYTR VSSAEQANEGYSIHEQKKKL ISYCEIHDWNEYKVFTDAGI SGGSMKRPALQKLMKHLSS FDLVLVYKLDRLTRNVRDL LDMLEEFEQYNVSFKSATE VFDTTSAIGKLFITMVGAMA EWERETIRERSLFGSRAAVR EGNYIREAPFCYDNIEGKLH PNEYAKVIDLIVSMFKKGIS ANEIARRLNSSKVHVPNKKS WNRNSLIRLMRSPVLRGHT KYGDMLIENTHEPVLSEHD YNAINNAISSKTHKSKVKHH AIFRGALVCPQCNRRLHLYA GTVKDRKGYKYDVRRYKC ETCSKNKDVKNVSFNESEV ENKFVNLLKSYELNKFHIRK VEPVKKIEYDIDKINKQKIN YTRSWSLGYIEDDEYFELME EINATKKMIEEQTTENKQSV SKEQIQSINNFILKGWEELTI KDKEELILSTVDKIEFNFIPK DKKHK TNTLDINNIHFKFS*	393	473	INTd

[0149] Sequences of insertion sites (i.e., recognition target sites) suitable for use in embodiments of the disclosure are presented below (Table 11). FIGs. 1A-1E show analysis of effect of variant AttP sites on integration efficiency.

Table 11

Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bxb1_AttP_GT_original_site	GTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCA	394	TGGGTTTGTACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCAC	473
Bxb1_AttP_CG_site	GTGGTTTGTCTGGTCAACCACCGCGcgCTCAGTGGTGTACGGTACAAACCCA	395	TGGGTTTGTACCGTACACCACTGAGCGCGCGGTGGTTGACCAGACAAACCAC	474

Table 11

Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bxb1_AttP_GC_site	GTGGTTTGTCTGGTCAACCACCGCGgcCTCAGTGGTGTACGGTACAAACCCA	396	TGGGTTTGTACCGTACACCACTGAGGCCGCGGTGGTTGACCAGACAAACCAC	475
Bxb1_AttP_AT_site	GTGGTTTGTCTGGTCAACCACCGCGatCTCAGTGGTGTACGGTACAAACCCA	397	TGGGTTTGTACCGTACACCACTGAGATCGCGGTGGTTGACCAGACAAACCAC	476
Bxb1_AttP_TA_site	GTGGTTTGTCTGGTCAACCACCGCGtaCTCAGTGGTGTACGGTACAAACCCA	398	TGGGTTTGTACCGTACACCACTGAGTACGCGGTGGTTGACCAGACAAACCAC	477
Bxb1_AttP_GG_site	GTGGTTTGTCTGGTCAACCACCGCGggCTCAGTGGTGTACGGTACAAACCCA	399	TGGGTTTGTACCGTACACCACTGAGCCCGCGGTGGTTGACCAGACAAACCAC	478
Bxb1_AttP_TT_site	GTGGTTTGTCTGGTCAACCACCGCGttCTCAGTGGTGTACGGTACAAACCCA	400	TGGGTTTGTACCGTACACCACTGAGAACGCGGTGGTTGACCAGACAAACCAC	479
Bxb1_AttP_GA_site	GTGGTTTGTCTGGTCAACCACCGCGgaCTCAGTGGTGTACGGTACAAACCCA	401	TGGGTTTGTACCGTACACCACTGAGTCCGCGGTGGTTGACCAGACAAACCAC	480
Bxb1_AttP_AG_site	GTGGTTTGTCTGGTCAACCACCGCGagCTCAGTGGTGTACGGTACAAACCCA	402	TGGGTTTGTACCGTACACCACTGAGCTCGCGGTGGTTGACCAGACAAACCAC	481
Bxb1_AttP_CC_site	GTGGTTTGTCTGGTCAACCACCGCGccCTCAGTGGTGTACGGTACAAACCCA	403	TGGGTTTGTACCGTACACCACTGAGGGCGCGGTGGTTGACCAGACAAACCAC	482
Bxb1_AttP_TC_site	GTGGTTTGTCTGGTCAACCACCGCGtcCTCAGTGGTGTACGGTACAAACCCA	404	TGGGTTTGTACCGTACACCACTGAGGACGCGGTGGTTGACCAGACAAACCAC	483
Bxb1_AttP_CT_site	GTGGTTTGTCTGGTCAACCACCGCGctCTCAGTGGTGTACGGTACAAACCCA	405	TGGGTTTGTACCGTACACCACTGAGAGCGCGGTGGTTGACCAGACAAACCAC	484
Bxb1_AttP_AA_site	GTGGTTTGTCTGGTCAACCACCGCGaaCTCAGTGGTGTACGGTACAAACCCA	406	TGGGTTTGTACCGTACACCACTGAGTTCGCGGTGGTTGACCAGACAAACCAC	485
Bxb1_AttP_CA_site	GTGGTTTGTCTGGTCAACCACCGCGcaCTCAGTGGTGTACGGTACAAACCCA	407	TGGGTTTGTACCGTACACCACTGAGTGCGCGGTGGTTGACCAGACAAACCAC	486
Bxb1_AttP_AC_site	GTGGTTTGTCTGGTCAACCACCGCGacCTCAGTGGTGTACGGTACAAACCCA	408	TGGGTTTGTACCGTACACCACTGAGGTCGCGGTGGTTGACCAGACAAACCAC	487
Bxb1_AttP_TG_site	GTGGTTTGTCTGGTCAACCACCGCGtgCTCAGTGGTGTACGGTACAAACCCA	409	TGGGTTTGTACCGTACACCACTGAGCACGCGGTGGTTGACCAGACAAACCAC	488
Bxb1_AttB_46_GT_original_site	GGCCGGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGG	410	CCGGATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCC	489
Bxb1_AttB_46_AA_site	GGCCGGCTTGTCGACGACGGCGaaCTCCGTCGTCAGGATCATCCGG	411	CCGGATGATCCTGACGACGGAGTTCGCCGTCGTCGACAAGCCGGCC	490

Table 11

Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bxb1_AttB_46_GA_site	GGCCGGCTTGTCGACGACGGCGgaCTCCGTCGTCAGGATCATCCGG	412	CCGGATGATCCTGACGACGGAGTCCGCCGTCGTCGACAAGCCGGCC	491
Bxb1_AttB_46_CA_site	GGCCGGCTTGTCGACGACGGCGcaCTCCGTCGTCAGGATCATCCGG	413	CCGGATGATCCTGACGACGGAGTGCGCCGTCGTCGACAAGCCGGCC	492
Bxb1_AttB_46_TA_site	GGCCGGCTTGTCGACGACGGCGtaCTCCGTCGTCAGGATCATCCGG	414	CCGGATGATCCTGACGACGGAGTACGCCGTCGTCGACAAGCCGGCC	493
Bxb1_AttB_46_AG_site	GGCCGGCTTGTCGACGACGGCGagCTCCGTCGTCAGGATCATCCGG	415	CCGGATGATCCTGACGACGGAGCTCGCCGTCGTCGACAAGCCGGCC	494
Bxb1_AttB_46_GG_site	GGCCGGCTTGTCGACGACGGCGggCTCCGTCGTCAGGATCATCCGG	416	CCGGATGATCCTGACGACGGAGCCCGCCGTCGTCGACAAGCCGGCC	495
Bxb1_AttB_46_CG_site	GGCCGGCTTGTCGACGACGGCGcgCTCCGTCGTCAGGATCATCCGG	417	CCGGATGATCCTGACGACGGAGCGCGCCGTCGTCGACAAGCCGGCC	496
Bxb1_AttB_46_TG_site	GGCCGGCTTGTCGACGACGGCGtgCTCCGTCGTCAGGATCATCCGG	418	CCGGATGATCCTGACGACGGAGCACGCCGTCGTCGACAAGCCGGCC	497
Bxb1_AttB_46_AC_site	GGCCGGCTTGTCGACGACGGCGacCTCCGTCGTCAGGATCATCCGG	419	CCGGATGATCCTGACGACGGAGGTCGCCGTCGTCGACAAGCCGGCC	498
Bxb1_AttB_46_GC_site	GGCCGGCTTGTCGACGACGGCGgcCTCCGTCGTCAGGATCATCCGG	420	CCGGATGATCCTGACGACGGAGGCCGCCGTCGTCGACAAGCCGGCC	499
Bxb1_AttB_46_CC_site	GGCCGGCTTGTCGACGACGGCGccCTCCGTCGTCAGGATCATCCGG	421	CCGGATGATCCTGACGACGGAGGGCGCCGTCGTCGACAAGCCGGCC	500
Bxb1_AttB_46_TC_site	GGCCGGCTTGTCGACGACGGCGtcCTCCGTCGTCAGGATCATCCGG	422	CCGGATGATCCTGACGACGGAGGACGCCGTCGTCGACAAGCCGGCC	501
Bxb1_AttB_46_AT_site	GGCCGGCTTGTCGACGACGGCGatCTCCGTCGTCAGGATCATCCGG	423	CCGGATGATCCTGACGACGGAGATCGCCGTCGTCGACAAGCCGGCC	502
Bxb1_AttB_46_CT_site	GGCCGGCTTGTCGACGACGGCGctCTCCGTCGTCAGGATCATCCGG	424	CCGGATGATCCTGACGACGGAGAGCGCCGTCGTCGACAAGCCGGCC	503
Bxb1_AttB_46_TT_site	GGCCGGCTTGTCGACGACGGCGttCTCCGTCGTCAGGATCATCCGG	425	CCGGATGATCCTGACGACGGAGAACGCCGTCGTCGACAAGCCGGCC	504
Bxb1_AttB_38_GT_site	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCAT	426	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCC	505
Bxb1_AttB_38_AA_site	GGCTTGTCGACGACGGCGaaCTCCGTCGTCAGGATCAT	427	ATGATCCTGACGACGGAGTTCGCCGTCGTCGACAAGCC	506

Table 11

Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bxb1_AttB_38_GA_site	GGCTTGTCGACGACGGCGgaCTCCGTCGTCAGGATCAT	428	ATGATCCTGACGACGGAGTCCGCCGTCGTCGACAAGCC	507
Bxb1_AttB_38_CA_site	GGCTTGTCGACGACGGCGcaCTCCGTCGTCAGGATCAT	429	ATGATCCTGACGACGGAGTGCGCCGTCGTCGACAAGCC	508
Bxb1_AttB_38_TA_site	GGCTTGTCGACGACGGCGtaCTCCGTCGTCAGGATCAT	430	ATGATCCTGACGACGGAGTACGCCGTCGTCGACAAGCC	509
Bxb1_AttB_38_AG_site	GGCTTGTCGACGACGGCGagCTCCGTCGTCAGGATCAT	431	ATGATCCTGACGACGGAGCTCGCCGTCGTCGACAAGCC	510
Bxb1_AttB_38_GG_site	GGCTTGTCGACGACGGCGggCTCCGTCGTCAGGATCAT	432	ATGATCCTGACGACGGAGCCCGCCGTCGTCGACAAGCC	511
Bxb1_AttB_38_CG_site	GGCTTGTCGACGACGGCGcgCTCCGTCGTCAGGATCAT	433	ATGATCCTGACGACGGAGCGCGCCGTCGTCGACAAGCC	512
Bxb1_AttB_38_TG_site	GGCTTGTCGACGACGGCGtgCTCCGTCGTCAGGATCAT	434	ATGATCCTGACGACGGAGCACGCCGTCGTCGACAAGCC	513
Bxb1_AttB_38_AC_site	GGCTTGTCGACGACGGCGacCTCCGTCGTCAGGATCAT	435	ATGATCCTGACGACGGAGGTCGCCGTCGTCGACAAGCC	514
Bxb1_AttB_38_GC_site	GGCTTGTCGACGACGGCGgcCTCCGTCGTCAGGATCAT	436	ATGATCCTGACGACGGAGGCCGCCGTCGTCGACAAGCC	515
Bxb1_AttB_38_CC_site	GGCTTGTCGACGACGGCGccCTCCGTCGTCAGGATCAT	437	ATGATCCTGACGACGGAGGGCGCCGTCGTCGACAAGCC	516
Bxb1_AttB_38_TC_site	GGCTTGTCGACGACGGCGtcCTCCGTCGTCAGGATCAT	438	ATGATCCTGACGACGGAGGACGCCGTCGTCGACAAGCC	517
Bxb1_AttB_38_AT_site	GGCTTGTCGACGACGGCGatCTCCGTCGTCAGGATCAT	439	ATGATCCTGACGACGGAGATCGCCGTCGTCGACAAGCC	518
Bxb1_AttB_38_CT_site	GGCTTGTCGACGACGGCGctCTCCGTCGTCAGGATCAT	440	ATGATCCTGACGACGGAGAGCGCCGTCGTCGACAAGCC	519
Bxb1_AttB_38_TT_site	GGCTTGTCGACGACGGCGttCTCCGTCGTCAGGATCAT	441	ATGATCCTGACGACGGAGAACGCCGTCGTCGACAAGCC	520
Cre Lox 66 site	TACCGTTCGTATAATGTATGCTATACGAAGTTAT	442	ATAACTTCGTATAGCATACATTATACGAACGGTA	521
Cre Lox 71 site	ATAACTTCGTATAATGTATGCTATACGAACGGTA	443	TACCGTTCGTATAGCATACATTATACGAAGTTAT	522
TP901-1 minimal AttB site	TTTACCTTGATTGAGATGTTAATTGTG	444	CACAATTAACATCTCAATCAAGGTAAA	523

Table 11

Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
TP901-1 minimal AttP site	GCGAGTTTTTATTTCGTTTATTTCAATTAAGGTAACTAAAAAACTCCTTT	445	AAAGGAGTTTTTTAGTTACCTTAATTGAAATAAACGAAATAAAAACTCGC	524
PhiBT1 minimal AttB site	CTGGATCATCTGGATCACTTTCGTCAAAAACCTG	446	CAGGTTTTTGACGAAAGTGATCCAGATGATCCAG	525
PhiBT1 minimal AttP site	TTCGGGTGCTGGGTTGTTGTCTCTGGACAGTGATCCATGGGAAACTACTCAGCACCA	447	TGGTGCTGAGTAGTTTCCCATGGATCACTGTCCAGAGACAACAACCCAGCACCCGAA	526
Bacillus_cereus_AH187_Int30_38bp_Att	gatatggggaagtgaatcagtacaaccgccacagtacc	448	ggtactgtggcggttgtactgattcacttccccatatc	527
Staphylococcus_lugdunensis_N920143_Int12_38bp_Att	tgggtggtacaggtgccacattagttgtaccattatatg	449	cataaatggtacaactaatgtggcacctgtaccaccca	528
Bacillus_cytotoxicus_NVH_391-98_Int13_38bp_Att	gttgtttttccagatccagttggtcctgtaaatataag	450	cttatatttacaggaccaactggatctggaaaaacaac	529
Bacillus_cereus_AH187_Int30_Att_30	tggggaagtgaatcagtacaaccgccacag	451	ctgtggcggttgtactgattcacttcccca	454
Bacillus_cereus_AH187_Int30_Att_28	ggggaagtgaatcagtacaaccgccaca	452	tgtggcggttgtactgattcacttcccc	455
Bacillus_cereus_AH187_Int30_Att_26	gggaagtgaatcagtacaaccgccac	453	gtggcggttgtactgattcacttccc	456
Bacillus_cereus_AH187_Int30_Att_rc_30	ctgtggcggttgtactgattcacttcccca	454	tggggaagtgaatcagtacaaccgccacag	451

Table 11

Description	Forward Sequence (5'-3')	SEQ ID NO:	Reverse Sequence (5'-3')	SEQ ID NO:
Bacillus_cereus_AH187_Int30_Att_rc_28	tgtggcggttgtactgattcacttcccc	455	ggggaagtgaatcagtacaaccgccaca	452
Bacillus_cereus_AH187_Int30_Att_rc_26	gtggcggttgtactgattcacttccc	456	gggaagtgaatcagtacaaccgccac	453
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_30	tttttccagatccagttggtcctgtaaata	457	tatttacaggaccaactggatctggaaaaa	460
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_28	ttttccagatccagttggtcctgtaaat	458	atttacaggaccaactggatctggaaaa	461
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_26	tttccagatccagttggtcctgtaaa	459	tttacaggaccaactggatctggaaa	462
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_rc_30	tatttacaggaccaactggatctggaaaaa	460	tttttccagatccagttggtcctgtaaata	457
Bacillus_cytotoxicus_NVH_391-98_Int13_Att_rc_28	atttacaggaccaactggatctggaaaa	461	ttttccagatccagttggtcctgtaaat	458
Bacillus_cytotoxicus_NVH_391-98_Int13	tttacaggaccaactggatctggaaa	462	tttccagatccagttggtcctgtaaa	459

Table 11				
Description	**Forward Sequence (5'-3')**	**SEQ ID NO:**	**Reverse Sequence (5'-3')**	**SEQ ID NO:**
_Att_rc_26				
N680429_560_31_50bp	CATTATATGTTTTTACAATCCGGGCCGCCATACTGTAAGAACATATAATG	463	cattatatgttcttacagtatggcggcccggattgtaaaaacatataatg	530
N191607_8_101_50bp	CGTTATAGGGTATTGCAGTACCGACCGCCATACTGTAATACCCTATAACG	464	cgttatagggtattacagtatggcggtcggtactgcaataccctataacg	531
N674992_1_1308_50bp	TGTATCATTTTCATATAGTGTGCAGGTGCTAACTATATGAAAATGATACA	465	tgtatcattttcatatagttagcacctgcacactatatgaaaatgataca	532
N684613_54_96_50bp	TGTCTACTATGTCTTTATGCCACATGTGTCGCATATACAGATAGTAGACA	466	tgtctactatctgtatatgcgacacatgtggcataaagacatagtagaca	533
N252616_121_74_50bp	AATGAGGTCAGACGCATGGAGCGCCGCCTCCGCATGCGTCAGGGTCGATG	467	catcgaccctgacgcatgcggaggcggcgctccatgcgtctgacctcatt	534
N683040_222_19_50bp	GTTAGTACCCAAATGATAAAAGGATGACCTTTTGTCATTTGGGTACTAAC	468	gttagtacccaaatgacaaaaggtcatccttttatcatttgggtactaac	535
N687537_173_59_50bp	GTTTATAAAACCGATGCCGCTTTGACAGAAGCGGAACGGGTTTTAATAAG	469	cttattaaaacccgttccgcttctgtcaaagcggcatcggttttataaac	536
N183629_47_40_50b p	GGCCGCGAGGTCGTGTTCGTCGTCATGTTGAGGTTCACGACCATCACGCC	470	ggcgtgatggtcgtgaacctcaacatgacgacgaacacgacctcgcggcc	537
N191533_224_76_50bp	TATAAACTGATATAATTCAAAGTTATAACTTGATATATTCAAGATGTAGA	471	tctacatcttgaatatatcaagttataactttgaattatatcagttttata	538
N682356_188_20_50 bp	TATTATATCTAAAAGCAGTATGGCGGAGCTTAGTGCTTTTAGATATAATT	472	aattatatctaaaagcactaagctccgccatactgcttttagatataata	539

6.10. Gene Editor Polypeptide Compositions

[0150] This disclosure features a gene editor polypeptide useful for site-specifically integrating an integration recognition site into a target nucleic acid sequence (i.e., beacon placement), for example by using a polypeptide comprising a DNA binding nuclease linked to a reverse transcriptase, and integrating an exogenous nucleic acid sequence at the integrated integration recognition site (i.e., PASTE) by introducing into the cell the exogenous nucleic acid and an integration enzyme. The gene editor polypeptide includes a polypeptide comprised of a (i) DNA binding nickase (optionally with nickase activity), (ii) a reverse transcriptase, and (iii) an integration enzyme (e.g., an integrase or a recombinase), wherein at least any two elements of (i),

(ii), or (iii) are linked via at least a first C-terminal linker. In some embodiments, the gene editor polypeptide includes a second linker. For example, where a gene editor polypeptide includes a first and a second linker, the first linker links a first domain (e.g., a DNA binding) to a second domain (e.g., a RT) and the second linker links the second domain to a third domain (e.g., an integration enzyme). In some embodiments, the orientation of the DNA binding nuclease, the RT, and the integration enzyme are configured for example from N-terminus to C-terminus: a DNA binding nuclease domain, a RT, and an integration enzyme; a DNA binding nuclease, an integration enzyme, and a RT; an integration enzyme, a DNA binding nuclease, and a RT; an integration enzyme, a RT, and a DNA binding nuclease; a RT, a DNA binding nuclease, and an integration enzyme; and a RT, an integration enzyme, and a DNA binding nuclease, where each of the domains are linked via a linker. In one embodiment, the C-terminus of the DNA binding nuclease is linked to the reverse transcriptase, wherein the C-terminus of the reverse transcriptase is linked to the integration enzyme. In one embodiment, the C-terminus of the integration enzyme is linked to the DNA binding nuclease, wherein the C-terminus of the DNA binding nuclease is linked to a reverse transcriptase.

[0151] In some embodiments, the gene editor polypeptide is comprised of at least two of any one of (i) a DNA binding nickase, (ii) a reverse transcriptase, and (iii) an integration enzyme wherein at least any two elements of (i), (ii), or (iii) are linked via at least a first C-terminal linker.

[0152] In certain embodiments, the gene editor polypeptide can be comprised of (i) a DNA binding nickase, (ii) a reverse transcriptase, and (iii) an integration enzyme (e.g., an integrase or a recombinase), wherein at least any two elements of (i), (ii), or (iii) may be linked via two or more C-terminal linkers at the same linking position. For example, a gene editor polypeptide includes a DNA binding nickase linked via two linkers to (a reverse transcriptase, and the reverse transcriptase domain fused or linked to an integration enzyme.

[0153] In some embodiments, a linker (e.g., the first linker, the second linker, or both) is selected from a flexible linker, a semi-flexible linker, a rigid linker, a flexible-rigid linker.

[0154] In some embodiments, the gene editor polypeptide includes at least one flexible linker. Non-limiting examples of flexible linkers include L1 (SEQ ID NO: 15), L2 (SEQ ID NO: 551), L5 (SEQ ID NO: 19), L6 (SEQ ID NO: 18), and L13 (SEQ ID NO: 558). In some embodiments, the gene editor polypeptide includes at least one flexible linker selected from: L1, L2, L5, L6, and L13. In some embodiments, the gene editor polypeptide includes at least a L1 linker (SEQ ID NO: 15). In some embodiments, the gene editor polypeptide includes at least a L2 linker (SEQ ID NO:

551). In some embodiments, the gene editor polypeptide includes at least a L5 linker (SEQ ID NO: 19) linker. In some embodiments, the gene editor polypeptide includes at least a L6 linker (SEQ ID NO: 18). In some embodiments, the gene editor polypeptide includes at least a L13 linker (SEQ ID NO: 558).

[0155] In some embodiments, the gene editor polypeptide includes at least one semi-flexible linker. Non-limiting examples of semi-flexible linkers include L7 (SEQ ID NO: 552) and L8 (SEQ ID NO: 553). In some embodiments, the gene editor polypeptide includes at least a L7 linker (SEQ ID NO: 552). In some embodiments, the gene editor polypeptide includes at least a L7 linker (SEQ ID NO: 553).

[0156] In some embodiments, the gene editor polypeptide includes at least one rigid linker. Non-limiting examples of rigid linkers include L3 (SEQ ID NO: 16), L4 (SEQ ID NO: 17), and L9 (SEQ ID NO: 554). In some embodiments, the gene editor polypeptide includes at least a L3 linker (SEQ ID NO: 16). In some embodiments, the gene editor polypeptide includes at least a L4 linker (SEQ ID NO: 17). In some embodiments, the gene editor polypeptide includes at least a L9 linker (SEQ ID NO: 554).

[0157] In some embodiments, the gene editor polypeptide includes at least one flexible-rigid linker. Non-limiting examples of flexible-rigid linkers include L11 (SEQ ID NO: 557) and L12 (SEQ ID NO: 558). In some embodiments, the gene editor polypeptide includes at least a L11 linker (SEQ ID NO: 557). In some embodiments, the gene editor polypeptide includes at least a L12 linker (SEQ ID NO: 558).

[0158] In some embodiments, the gene editor polypeptide includes a linker (e.g., a first linker, a second linker, or both) where the length of the linker is selected from short, medium, and long. Non-limiting examples of short linkers include: L1 (SEQ ID NO: 15), L3 (SEQ ID NO: 16), L7 (SEQ ID NO: 552), and L8 (SEQ ID NO: 553). Non-limiting examples of a medium linker include: L2 (SEQ ID NO: 551), L4 (SEQ ID NO: 17), L5 (SEQ ID NO: 19), L6 (SEQ ID NO:18), L10 (SEQ ID NO: 555), L11 (SEQ ID NO: 556), L12 (SEQ ID NO: 557), and L13 (SEQ ID NO: 558). A non-limiting example of a long linker is L9 (SEQ ID NO: 554).

[0159] In some embodiments, the gene editor polypeptide includes a linker (e.g., the first linker, the second linker, or both) that is a non-cleavable linker. Non-limiting examples of non-cleavable linkers include: L1 (SEQ ID NO: 15), L2 (SEQ ID NO: 551), L3 (SEQ ID NO:16), L4 (SEQ ID NO: 17), L5 (SEQ ID NO: 19), L6 (SEQ ID NO: 18), L7 (SEQ ID NO: 552), L8 (SEQ ID NO:

553), L9 (SEQ ID NO: 554), L10 (SEQ ID NO: (555), L11 (SEQ ID NO: 556), L12 (SEQ ID NO: 557), and L13 (SEQ ID NO: 558).

[0160] In some embodiments, the gene editor polypeptide include a linker (e.g., the first linker, the second linker, or both) that is a cleavable linker. Cleavable linkers can facilitate the co-expression of multiple proteins within a single open reading frame, effectively separating them post-translationally. Non-limiting examples of self-cleaving peptides include 2A peptides (derived from the foot-and-mouth disease virus), such as F2A, T2A, E2A, and P2A; the "Inteins" that facilitate protein splicing; and the tobacco etch virus (TEV) protease recognition site. In some embodiments, the cleavable linker is 2A peptide. In some embodiments, the cleavable linker is a P2A peptide.

[0161] In some embodiments, the gene editor polypeptide includes at least a first linker and a second linker where: the at least first linker is a non-cleavable linker and the second linker is a cleavable linker; the at least first linker is a cleavable linker and the second linker is a non-cleavable linker; the at least first linker is a cleavable linker and the second linker is a cleavable linker; or the at least first linker is a non-cleavable linker and the second linker is a non-cleavable linker.

[0162] In some embodiments, the gene editor polypeptide includes the DNA binding nuclease, the RT, and the integration enzyme, where the C-terminus of the DNA binding nuclease is linked to the reverse transcriptase, wherein the C-terminus of the reverse transcriptase is linked to the integration enzyme, and where any one of L7 (SEQ ID NO: 552), L10 (SEQ ID NO: 555), or L12 (SEQ ID NO: 557) links the DNA binding nuclease to the reverse transcriptase.

[0163] In some embodiments, the gene editor polypeptide includes the DNA binding nuclease, the RT, and the integration enzyme, where the C-terminus of the DNA binding nuclease is linked to the reverse transcriptase, wherein the C-terminus of the reverse transcriptase is linked to the integration enzyme, wherein any one of L8 (SEQ ID NO: 553), L9 (SEQ ID NO: 554), L10 (SEQ ID NO: 555), L12 (SEQ ID NO: 557), or L13 (SEQ ID NO: 558) links the reverse transcriptase to the integration enzyme.

[0164] In some embodiments, the gene editor polypeptide includes the DNA binding nuclease, the RT, and the integration enzyme, where the C-terminus of the integration enzyme is linked to the DNA binding nuclease by L13 (SEQ NO: 558), and where the C-terminus of the DNA binding nuclease is linked to the reverse transcriptase by L13 (SEQ ID NO: 558).

[0165] In some embodiments, the gene editor polypeptide is not comprised of a N-most N-terminal reverse transcriptase.

[0166] In some embodiments, the gene editor polypeptide includes the DNA binding nuclease, the RT, and the integration enzyme, where the C-terminus of the DNA binding nuclease is linked to the reverse transcriptase by L13 (SEQ ID NO: 558), and where the C-terminus of the reverse transcriptase is linked to the integration enzyme by a cleavable linker (e.g., a P2A (SEQ ID NO: 13).

[0167] In some embodiments, the gene editor polypeptide includes a nuclear localization signal (NLS). In some embodiments, the gene editor polypeptide comprises a mitochondrial localization signal.

[0168] In one embodiment, the gene editor polypeptide comprises: from N-terminus to C-terminus: a DNA binding nickase domain; a P2A linker (SEQ ID NO: 13), a reverse transcriptase; a mXTEN linker, and an integration enzyme.

[0169] In one embodiment, the gene editor polypeptide comprises: from N-terminus to C-terminus: an integration enzyme, a mXTEN linker (SEQ ID NO: 558), a DNA binding nuclease domain; a mXTEN linker (SEQ ID NO: 558), and a reverse transcriptase.

[0170] This disclosure also features systems capable of site-specifically integrating at least a first integration recognition site into the genome of a cell where the system includes any of the gene editor polypeptides described herein.

[0171] This disclosure also features systems capable of site-specifically integrating an exogenous nucleic acid sequence into the genome of a cell where the system includes any of the gene editor polypeptides described herein.

6.11. Gene Editor Polynucleotide Compositions

[0172] In certain embodiments, a polynucleotide encodes the gene editor polypeptide. In some embodiments, the polynucleotide is an RNA or DNA. In some embodiments, the polynucleotide is a mRNA.

[0173] In certain embodiments, an expression vector comprises the polynucleotide encoding the gene editor polypeptide. In some embodiments, the expression vector further comprises any one or more of a one or more of a promoter sequence, atgRNA, nicking gRNA (ngRNA), donor

polynucleotide template, or donor polynucleotide template complement. In some embodiments, two or more expression vectors are delivered.

[0174] In typical embodiments, the expression vector is recombinant adenovirus, helper dependent adenovirus, AAV, lentivirus, HSV, annelovirus, retrovirus, Doggybone DNA, minicircle, plasmid, miniDNA, nanoplasmid, exosome, fusosome, or lipid nanoparticle.

6.12. Programmable Gene Insertion

[0175] In some embodiments, the polypeptides described herein are used for programmable polynucleotide (e.g., genes or gene fragments) insertion. Programmable polynucleotide/gene insertion includes site-specifically integrating an exogenous nucleic acid sequence (e.g., a donor polynucleotide template) into a mammalian cell genome at a desired target site (e.g., a target DNA sequence). In typical embodiments, the method of site-specifically integrating an exogenous nucleic acid sequence into a target nucleic acid comprises incorporating an integration recognition site into a cell by delivering: (i) a polynucleotide encoding a gene editor polypeptide (e.g., any of the gene editor polypeptides described herein), (ii) at least a first target specific atgRNAs comprising a first integration recognition site, for example, as described in Section 6.7, and (iii) optionally, a nicking guide RNA. In some embodiments, the first atgRNA comprises: (i) a domain that is capable of guiding the prime editor system to a target sequence; and (ii) a reverse transcriptase (RT) template that comprises at least a portion of an at least first integration recognition site, whereby the at least portion of the at least first integration recognition site is integrated into the genome of the cell at the target sequence.

[0176] In some embodiments, the method of site-specifically integrating an exogenous nucleic acid sequence into a target nucleic acid, the method comprises incorporating an integration recognition site into a cell by delivering: (i) a polynucleotide encoding a gene editor polypeptide (e.g., any of the gene editor polypeptides described herein), (ii) a first target specific atgRNA and a second target specific atgRNA. In such embodiments, the first atgRNA and the second atgRNA are an at least first pair of atgRNAs, wherein the at least first pair of atgRNAs have domains that are capable of guiding the prime editor system to a target sequence; the first atgRNA further includes a first RT template that comprises at least a portion of an at least first integration recognition site; the second atgRNA further includes a second RT template that comprises at least a portion of the first integration recognition site, and the first atgRNA and the second atgRNAs collectively encode the entirety of the first integration recognition site, whereby the first integration recognition site is integrated into the genome of the cell at the target sequence.

[0177] In some embodiments, the method includes integrating the polynucleotide in the cellular genome by delivering a donor polynucleotide template, wherein the donor polynucleotide template is comprised of a second integration recognition site that is a cognate pair of the first integration recognition site, wherein the donor polynucleotide is integrated into the cellular genome at the first integration target recognition site by the gene editor polynucleotide..

[0178] In some embodiments, the method of site-specifically integrating an exogenous nucleic acid sequence into a mammalian cell genome at a desired target site (e.g., a target DNA sequence) comprises incorporating an integration target site into a cell by delivering: an expression vector of comprising a polynucleotide encoding a gene editor polypeptide (e.g., any of the gene editor polypeptides described herein), further comprising a promoter, one or more cell genome target specific atgRNA comprising an integration target recognition site, for example, as described in Section 6.7, a donor polynucleotide comprising an integration target site, and optionally, a ngRNA. In some embodiments, the method of polynucleotide integration into a cell genome comprises incorporating an integration target site into a cell by delivering: an expression vector of comprising a polynucleotide encoding a gene editor polypeptide (e.g., any of the gene editor polypeptides described herein), further comprising a promoter, two cell genome target specific atgRNA where the two atgRNA collectively encode the entirety of the an integration recognition site, and a donor polynucleotide comprising an integration target site.

[0179] In some embodiments, the method of site-specifically integrating an exogenous nucleic acid sequence into a genome of a cell includes incorporating an integration recognition site into a cell by delivering: any of the polynucleotides described herein; at least a first genome target specific atgRNA comprising a first integration recognition site, for example, as described in Section 6.7; optionally, a nicking gRNA; and integrating an exogenous nucleic acid sequence into the cellular genome by delivering into the cell: a donor polynucleotide template, wherein the donor polynucleotide template is comprised of the exogenous nucleic acid sequence positioned relative to a second integration recognition site that is a cognate pair of the first integration recognition site such that the exogenous nucleic acid sequence is integrated into the cellular genome at the first integration target recognition site when contacted with the integration enzyme and the first integration recognition site. In some embodiments, the polynucleotides and the at least first atgRNA are delivered concurrently with the donor polynucleotide template.

[0180] In some embodiments, the method of site-specifically integrating an exogenous nucleic acid sequence into a target nucleic acid includes: incorporating an integration recognition site into a cell by delivering: any of the expression vectors described herein; and integrating an exogenous

nucleic acid sequence into the cellular genome by integrating a donor polynucleotide template, wherein the donor polynucleotide template is comprised of the exogenous nucleic acid sequence positioned relative to a second integration recognition site that is a cognate pair of the first integration recognition site such that the exogenous nucleic acid sequence is integrated into the cellular genome at the first integration target recognition site when contacted with the integration enzyme and the first integration recognition site. In some embodiments, the expression vector and the donor polynucleotide template are delivered to the cell concurrently.

6.13. Genes and Targets

[0181] This disclosure provides compositions, systems 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 disclosure comprises recombining corrective gene fragments into a defective locus.

[0182] 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 disclosure are adapted to target organoids, including patient derived organoids.

[0183] In certain embodiments, methods and compositions of the disclosure 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 of generate a premature stop codon. An aspect of DMD that lends it to gene editing as a therapeutic approach is the modular structure of the dystrophin protein. Redundancy in the central rod domain permits the deletion of internal segments of the gene that may harbor loss-of-function mutations, thereby restoring the open reading frame (ORFs). In some embodiments, the methods and systems described herein are used to treat DMD by site-

specifically integrating in the genome a polynucleotide template that repairs or replaces all or a portion of the defective DMD gene.

[0184] 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)

Autoimmune diseases:

- IgA Nephropathy (Berger's disease)
- Anti-Neutrophil Cytoplasmic Antibody (ANCA) Vasculitis
- Systemic Lupus Erythematosus (SLE) / Lupus Nephritis (LN)
- Membranous Nephropathy (MN)
- C3 glomerulonephritis (C3GN)

Blood disorders:

- Sickle Cell
- Hemophilia

- Factor VIII or
- Factor IX
- Ornithine transcarbamylase deficiency (OTCD)
- Homocystinuria (HCU)
- Phenylketonuria (PKU)

Cancer

- Prostate cancer
- Renal cell cancer
- Thyroid cancer

[0185] CFTR (cystic fibrosis transmembrane conductance regulator). The most common cystic fibrosis (CF) mutation F508del removes a single amino acid. In some embodiments, recombining human CFTR into an SHS of a cell that expresses CFTR F508del is a corrective treatment path. In some embodiments, the methods and systems described herein are used to CF by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing CF. Proposed validation is detection of persistent CFTR mRNA and protein expression in transduced cells.

[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. In some embodiments, the methods and systems described herein are used to sickle cell disease by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing the disease. In some embodiments, validation is detection of persistent HBB mRNA and protein expression in transduced cells.

[0187] DMD — Duchenne Muscular Dystrophy. 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 of 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).

[0188] 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 e 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.

[0189] In some embodiments, correcting mutations in exon 44 (or 51) by recombining in a corrective coding sequence downstream of exon 43 (or 50), using the methods and systems described herein is a corrective option. Proposed validation is detection of persistent DMD mRNA and protein expression in transduced cells.

[0190] **F8 (Factor VIII).** 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.

[0191] In some embodiments, correcting factor VIII deficiency by recombining the FVIII gene into an SHS is a corrective path. In some embodiments, the methods and systems described herein are used to correct factor VIII deficiency by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing the FIX deficiency. Proposed validation is detection of persistent FVIII mRNA and protein expression in transduced cells.

[0192] **Factor 9 (Factor IX)** Hemophilia B, also called factor IX (FIX) deficiency is a genetic disorder caused by missing or defective factor IX, a clotting protein.

[0193] In some embodiments, the methods and systems described herein are used to correct factor IX deficiency by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing the FIX deficiency. Proposed validation is detection of persistent FiX mRNA and protein expression in transduced cells.

[0194] **Ornithine transcarbamylase deficiency (OTCD).** Ornithine transcarbamylase deficiency is a rare genetic condition that causes ammonia to build up in the blood. The condition

– more commonly called OTC deficiency – is more common in boys than girls and tends to be more severe when symptoms emerge shortly after birth.

[0195] In some embodiments, the methods and systems described herein are used to correct OTC deficiency by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing the OTC deficiency or integrates a polynucleotide encoding a functional ornithine transcarbamylase enzyme. Proposed validation is detection of persistent OTC mRNA and protein expression in transduced cells.

[0196] **Phenylketonuria**, also called PKU, is a rare inherited disorder that causes an amino acid called phenylalanine to build up in the body. PKU is caused by a change in the phenylalanine hydroxylase (PAH) gene. This gene helps create the enzyme needed to break down phenylalanine.

[0197] In some embodiments, the methods and systems described herein are used to correct PKU by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing the PKU deficiency or integrates a polynucleotide encoding a functional phenylalanine hydroxylase (PAH) gene. Proposed validation is detection of persistent PAH mRNA and protein expression in transduced cells.

[0198] **Homocystinuria (HCU).** Homocystinuria is elevation of the amino acid, homocysteine (protein building block coming from our diet) in the urine or blood. Common causes of HCU include: problems with the enzyme cystathionine beta synthase (CBS), which converts homocysteine to the amino acid cystathionine (which then becomes cysteine) and needs the vitamin B6 (pyridoxine); and roblems with converting homocysteine to the amino acid methionine.

[0199] In some embodiments, the methods and systems described herein are used to correct HCU by site-specifically integrating in the genome a polynucleotide template that corrects the mutation causing the HCU or integrates a polynucleotide encoding a functional copy of a gene (e.g., CBS) able to reduce or prevent buildup of homocysteine in the urine. Proposed validation is detection of persistent CBS mRNA and protein expression in transduced cells.

[0200] IgA Nephropathy (Berger's disease). IgA nephropathy, also known as Berger's disease, is a kidney/autoimmune disease that occurs when an antibody called immunoglobulin A (IgA) builds up in the kidneys.

[0201] In some embodiments, the methods and systems described herein are used to treat Berger's disease by administering to a patient an iPSC-derived Natural Killer cell that includes a polynucleotide site-specifically integrated in the genome of the cell using the methods described herein. Upon administering the iPSC-NK cell to the patient, the iPSC-NK cell is capable of removing native cells (e.g., B cells) that are responsible, at least in part, for the symptoms of Berger's disease.

[0202] **Anti-Neutrophil Cytoplasmic Antibody (ANCA) Vasculitis.** ANCA vasculitis is an autoimmune disease affecting small blood vessels in the body. It is caused by autoantibodies called ANCAs, or Anti-Neutrophilic Cytoplasmic Autoantibodies. ANCAs target and attack a certain kind of white blood cells called neutrophils.

[0203] In some embodiments, the methods and systems described herein are used to treat ANCA vasculitis by administering to a patient an iPSC-derived Natural Killer cell that includes a polynucleotide site-specifically integrated in the genome of the cell using the methods described herein. Upon administering the iPSC-NK cell to the patient, the iPSC-NK cell is capable of removing native cells (e.g., B cells) that are responsible, at least in part, for the symptoms of ANCA vasculitis.

[0204] **Systemic Lupus Erythematosus (SLE) / Lupus Nephritis (LN).** Lupus is an autoimmune—a disorder in which the body's immune system attacks the body's own cells and organs.

[0205] In some embodiments, the methods and systems described herein are used to treat SLE/LN by administering to a patient an iPSC-derived Natural Killer cell that includes a polynucleotide site-specifically integrated in the genome of the cell using the methods described herein. Upon administering the iPSC-NK cell to the patient, the iPSC-NK cell is capable of removing native cells (e.g., B cells) that are responsible, at least in part, for the symptoms of SLE/LN.

[0206] **Membranous Nephropathy (MN).** MN is a kidney disease that affects the filters (glomeruli) of the kidney and can cause protein in the urine, as well as decreased kidney function and swelling. It can sometimes be called membranous glomerulopathy as well (these terms can be used interchangeably and mean the same thing).

[0207] In some embodiments, the methods and systems described herein are used to treat MN by administering to a patient an iPSC-derived Natural Killer cell that includes a polynucleotide

site-specifically integrated in the genome of the cell using the methods described herein. Upon administering the iPSC-NK cell to the patient, the iPSC-NK cell is capable of removing native cells (e.g., B cells) that are responsible, at least in part, for the symptoms of MN.

[0208] **C3 glomerulonephritis (C3GN).** C3 glomerulopathy is a group of related conditions that cause the kidneys to malfunction. The major features of C3 glomerulopathy include high levels of protein in the urine (proteinuria), blood in the urine (hematuria), reduced amounts of urine, low levels of protein in the blood, and swelling in many areas of the body. Affected individuals may have particularly low levels of a protein called complement component 3 (or C3) in the blood.

[0209] In some embodiments, the methods and systems described herein are used to treat C3 glomerulopathy by administering to a patient an iPSC-derived Natural Killer cell that includes a polynucleotide site-specifically integrated in the genome of the cell using the methods described herein. Upon administering the iPSC-NK cell to the patient, the iPSC-NK cell is capable of removing native cells (e.g., B cells) that are responsible, at least in part, for the symptoms of C3 glomerulopathy.

6.14. Methods of treatment

[0210] In another aspect, methods of treatment are presented. The method comprises administering an effective amount of the pharmaceutical composition comprising the nucleic acid construct or vectorized nucleic acid construct described above to a patient in need thereof. In typical embodiments, the polynucleotide encoding a gene editor polypeptide are delivered by a vector, wherein the vector is selected from the group consisting of recombinant adenovirus, helper dependent adenovirus, AAV, lentivirus, HSV, annelovirus, retrovirus, Doggybone^{TM} DNA (dbDNA), minicircle, plasmid, miniDNA, exosome, fusosome, lipid nanoparticle, or nanoplasmid.

[0211] 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.

[0212] Methods of non-viral delivery of the polynucleotide encoding a gene editor polypeptide or the gene editor polypeptide 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).

6.14.1. Lipid Nanoparticle Delivery

[0213] In some embodiments, the gene editor polypeptide or polynucleotide encoding the gene editor polypeptide is packaged in a LNP and administered intravenously. In some embodiments, the gene editor polypeptide or polynucleotide encoding the gene editor polypeptide is packaged in a LNP and administered intrathecally. In some embodiments, the gene editor polypeptide or polynucleotide encoding the gene editor polypeptide is packaged in a LNP and administered by intracerebral ventricular injection. In some embodiments, the gene editor polypeptide or polynucleotide encoding the gene editor polypeptide is packaged in a LNP and administered by intracisternal magna administration. In some embodiments, the gene editor polypeptide or polynucleotide encoding the gene editor polypeptide is packaged in a LNP and administered by intravitreal injection.

[0214] 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).

[0215] 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, acetampinophen, 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.

[0216] 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-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoleyl-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.

[0217] 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-piperazinediethanami- ne (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), 1,2-dioleyloxy-N,N-dimethylaminopropane (DODMA), 2-({8-[(3)-cholest-5-en-3-yloxy]octyl}oxy)-N,N-dimethyl-3-[(9Z,12Z)-octad- eca-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-y loxy]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.

[0218] 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-piperazinediethanami- ne (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-dioleyl-N,N-dimethylammonium chloride ("DODAC"); N-(2,3-dioleyloxy)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-Dioleyloxy-3-trimethylaminopropane chloride salt ("DOTAP.Cl"); 3-.beta.-(N—(N',N'-dimethylaminoethane)-carbamoyl)cholesterol ("DC-Chol"), N-(1-(2,3-dioleyloxy)propyl)-N-2-(sperminecarboxamido)ethyl)-N,N-dimethyl- -ammonium trifluoracetate ("DOSPA"), dioctadecylamidoglycyl carboxyspermine ("DOGS"), 1,2-dioleoyl-3-dimethylammonium propane ("DODAP"), N,N-dimethyl-2,3-dioleyloxy)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.

[0219] 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).

[0220] Anionic lipids suitable for use in lipid nanoparticles include, but are not limited to, phosphatidylglycerol, cardiolipin, diacylphosphatidylserine, diacylphosphatidic acid, N-dodecanoyl phosphatidylethanoloamine, N-succinyl phosphatidylethanolamine, N-glutaryl phosphatidylethanolamine, lysylphosphatidylglycerol, and other anionic modifying groups joined to neutral lipids.

[0221] 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 cerebrosides. 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, dihydrosphingomyeline, or phospholipids with other head groups, such as serine and inositol.

[0222] 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.

[0223] 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.

[0224] 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.

[0225] 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.

[0226] 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.

[0227] 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).

[0228] Phospholipids include, but are not limited to, glycerophospholipids such as phosphatidylcholines, phosphatidylethanolamines, phosphatidylserines, phosphatidylinositols, phosphatidy glycerols, and phosphatidic acids. Phospholipids also include phosphosphingolipid, such as sphingomyelin.

[0229] In some embodiments, the LNP composition comprises one or more phospholipids. In some embodiments, the phospholipid is selected from the group consisting of 1,2-dilinoleoyl-sn-glycero-3-phosphocholine (DLPC), 1,2-dimyristoyl-sn-glycero-phosphocholine (DMPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1,2-dipalmitoyl-sn-glycero-3-phosphocholine

(DPPC), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), 1,2-diundecanoyl-sn-glycero-phosphocholine (DUPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1,2-di-O-octadecenyl-sn-glycero-3-phosphocholine (18:0 Diether PC), 1-oleoyl-2-cholesterylhemisuccinoyl-sn-glycero-3-phosphocholine (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-didocosahexaenoyl-sn-glycero-3-phosphocholine, 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (ME 16:0 PE), 1,2-distearoyl-sn-glycero-3-phosphoethanolamine, 1,2-dilinoleoyl-sn-glycero-3-phosphoethanolamine, 1,2-dilinolenoyl-sn-glycero-3-phosphoethanolamine, 1,2-diarachidonoyl-sn-glycero-3-phosphoethanolamine1,2-didocosahexaenoyl—sn-glycero-3-phosphoethanolamine, 1,2-dioleoyl-sn-glycero-3-phospho-rac-(1-glycerol) sodium salt (DOPG), sphingomyelin, and any mixtures thereof.

[0230] 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.

[0231] In some embodiments, the LNP composition comprises one or more helper lipids. The term "helper lipid" as used herein refers to lipids that enhance transfection (e.g., transfection of an LNP comprising an mRNA that encodes a site-directed endonuclease, such as a SpCas9 polypeptide). In principle, there are no specific limitations concerning the helper lipids of the LNP compositions disclosed herein. Without being bound to any particular theory, it is believed that the mechanism by which the helper lipid enhances transfection includes enhancing particle stability. In some embodiments, the helper lipid enhances membrane fusogenicity. Generally, the helper lipid of the LNP compositions disclosure herein can be any helper lipid known in the art. Non-limiting examples of helper lipids suitable for the compositions and methods include steroids, sterols, and alkyl resorcinols. Particularly helper lipids suitable for use in the present disclosure include, but are not limited to, saturated phosphatidylcholine (PC) such as distearoyl-PC (DSPC) and dipalymitoyl-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.

[0232] 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.

[0233] 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-dipalmetoleyl, PEG-dioleyl, PEG-distearyl, PEG-diacylglycamide (PEG-DAG), PEG-dipalmitoyl phosphatidylethanolamine (PEG-DPPE), or PEG-1,2-dimyristyloxlpropyl-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.

[0234] 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.

[0235] 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.

[0236] 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).

[0237] 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.

[0238] 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.

[0239] 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).

[0240] 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.

[0241] 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.

[0242] 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.

[0243] 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.

[0244] 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.

[0245] 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.

[0246] 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.

[0247] 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.

6.14.2. Viral Vector Delivery

[0248] In certain aspects this disclosure includes involves vectors, e.g. for delivering or introducing in a cell, but also for propagating these components (e.g. in prokaryotic cells). A used

herein, a "vector" is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Vectors for and that result in expression in a eukaryotic cell can be referred to herein as "eukaryotic expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0249] Recombinant expression vectors can comprise a nucleic acid of the disclosure 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.

[0250] 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.

[0251] 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.

[0252] In an embodiment herein the delivery is via an adenovirus, which may be at a single booster dose containing at least 1 x 10^5 particles (also referred to as particle units, pu) of adenoviral vector.

In some embodiments, the dose preferably is at least about 1 x 10^6 particles (for example, about 1 x 10^6-1 x 10^{11} particles), more preferably at least about 1 x 10^7 particles, more preferably at least about 1 x 10^8 particles (e.g., about 1 x 10^8-1 x 10^{11} particles or about 1 x 10^9-1 x 10^{12} particles), and most preferably at least about 1 x 10^{10} particles (e.g., about 1 x 10^9-1 x 10^{10} particles or about 1 x 10^9-1 x 10^{12} particles), or even at least about 1 x 10^{10} particles (e.g., about 1 x 10^{10}-1 x 10^{12} particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1 x 10^{14} particles, preferably no more than about 1 x 10^{13} particles, even more preferably no more than about 1 x 10^{12} particles, even more preferably no more than about 1 x 10^{11} particles, and most preferably no more than about 1 x 10^{10} particles (e.g., no more than about 1 x 10^9 particles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1 x 10^6 particle units (pu), about 2 x 10^6 pu, about 4 x 10^6 pu, about 1 x 10^7 pu, about 2 x 10^7 pu, about 4 x 10^7 pu, about 1 x 10^8 pu, about 2 x 10^8 pu, about 4 x 10^8 pu, about 1 x 10^9 pu, about 2 x 10^9 pu, about 4 x 10^9 pu, about 1 x 10^{10} pu, about 2 x 10^{10} pu, about 4 x 10^{10} pu, about 1 x 10^{11} pu, about 2 x 10^{11} pu, about 4 x 10^{11} pu, about 1 x 10^{12} pu, about 2 x 10^{12} pu, or about 4 x 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 some embodiments, the adenovirus is delivered via multiple doses.

[0253] In some embodiments, 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 x 10^{10} to about 1 x 10^{10} functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In some embodiments, the AAV dose is generally in the range of concentrations of from about 1 x 10^5 to 1 x 10^{50} genomes AAV, from about 1 x 10^8 to 1 x 10^{20} genomes AAV, from about 1 x 10^{10} to about 1 x 10^{16} genomes, or about 1 x 10^{11} to about 1 x 10^{16} genomes AAV. A human dosage may be about 1 x 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.

[0254] 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.

[0255] 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)

[0256] 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.

[0257] 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.

[0258] 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 disclosure include utilizing homologs of nucleic acid-targeting effector protein (such as a Type V protein such as C2c1 or C2c3) that are shorter.

[0259] 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.

[0260] 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.

[0261] 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 x 10^8 vp or 1.8 x 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 disclosure, contemplating a dose of about 2 x 10^{11} to about 6 x 10^{11} vp administered to a human.

[0262] 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.times.10.sup.12 vg/ml. The AAV vectors of Dalkara et al. may be applied to the nucleic acid-targeting system of the present disclosure, contemplating a dose of about 1 x 10^{15} to about 1 x 10^{16} vg/ml administered to a human.

[0263] 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 immune 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).

[0264] 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.

[0265] 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.

[0266] 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.

[0267] 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.

[0268] 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.

[0269] In one aspect, the disclosure 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).

[0270] In plants, pathogens are often host-specific. For example, *Fusariumn 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 disclosure, 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 disclosure to induce the rise of resistance genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

[0271] 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. ADDITIONAL EMBODIMENTS

[0272] Embodiment 1. A guide RNA, comprising from 5' to 3': a spacer complementary to a first target genomic site; a scaffold capable of binding to a DNA binding nickase; a reverse transcriptase template comprised of at least a first segment and a second segment, wherein the first segment encodes a clamp segment complementary to a second genomic site and the second segment encodes an integrase target recognition site; and a primer binding site complementary to a third genomic site; wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence.

[0273] Embodiment 2. The guide RNA of any one of Embodiment 1, further comprising a 3' extension capable of binding one or more consecutive nucleotides of the reverse transcriptase template first segment.

[0274] Embodiment 3. The guide RNA of any one of Embodiments 1-2, wherein the second genomic site and the third genomic site are separated by one or more genomic bases.

[0275] Embodiment 4. The guide RNA of Embodiment 3, wherein the second genomic site and the third genomic site are separated by 15-21, 33-39, or 35-41 genomic bases.

[0276] Embodiment 5. The guide RNA of any one of Embodiments 1-4, wherein the clamp segment binds to the second genomic site with one or more nucleotide mismatches.

[0277] Embodiment 6. The guide RNA of Embodiment 5, wherein the mismatches are consecutive or non-consecutive.

[0278] Embodiment 7. The guide RNA of any one of Embodiments 1-6, wherein the guide RNA is circularized.

[0279] Embodiment 8. A guide RNA complex, comprising: a spacer complementary to a first target genomic site; a scaffold capable of binding to a DNA binding nickase; a reverse transcriptase template comprised of at least a first segment and a second segment, wherein the first segment

encodes a clamp segment complementary to a second genomic site and the second segment encodes an integrase target recognition site; a primer binding site complementary to a third genomic site; wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence.

[0280] Embodiment 9. The guide RNA complex of Embodiment 8, wherein the guide RNA complex is comprised of a first polynucleotide component and a second polynucleotide component.

[0281] Embodiment 10. The guide RNA complex of Embodiment 9, wherein the first polynucleotide component comprises a spacer complementary to a first target genomic site and a scaffold capable of binding to a DNA binding nickase, and wherein the second polynucleotide component comprises a reverse transcriptase template comprised of at least a first segment and a second segment, wherein said first segment encodes a clamp segment complementary to a second genomic site and said second segment encodes an integrase target recognition site, a primer binding site complementary to a third genomic site, and a RNA-protein recruitment domain, and wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence.

[0282] Embodiment 11. The guide RNA complex of Embodiment 10, wherein one or more of the first polynucleotide component and the second polynucleotide component are circularized.

[0283] Embodiment 12. The guide RNA or the guide RNA complex of any one of Embodiments 1-11, further comprises one or more of an RNA-protein recruitment domain, a transcriptional termination signal, a reverse transcription termination signal, a nucleobase analog, an RNA ribozyme, or a chemical linker.

[0284] Embodiment 13. The guide RNA or the guide RNA complex of any one of Embodiments 1-12, wherein the guide RNA reverse transcriptase template encodes for a first single-stranded DNA sequence that contains a complementary region to a second single-stranded DNA sequence encoded by a second guide RNA comprised of a reverse transcriptase template.

[0285] Embodiment 14. The guide RNA or the guide RNA complex of any one of Embodiments 1-13, wherein an RNA polynucleotide comprises the guide RNA or a DNA polynucleotide encodes the guide RNA.

[0286] Embodiment 15. The guide RNA of Embodiment 1 or the guide RNA complex of Embodiment 2, wherein the guide RNA 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.

[0287] Embodiment 16. The guide RNA or the guide RNA complex of Embodiment 14, wherein the guide RNA or the guide RNA complex is introduced into a cell using a recombinant adenovirus, helper dependent adenovirus, AAV, lentivirus, HSV, annelovirus, retrovirus, Doggybone DNA, minicircle, plasmid, miniDNA, nanoplasmid, exosome, fusosome, mRNA, RNP, or lipid nanoparticle.

[0288] Embodiment 17. A method of polynucleotide integration into a cellular genome, the method comprising: incorporating an integration target recognition site into a cell genome by delivering: a gene editor protein or polynucleotide encoding a gene editor protein, wherein the gene editor protein is capable of polynucleotide genomic integration; one or more guide RNA or guide RNA complex selected from any one of Embodiments 1-16; optionally, a nicking gRNA; and integrating the polynucleotide into the cellular genome by delivering: a donor polynucleotide template, wherein the donor polynucleotide template is comprised of an integration target recognition site, wherein the donor polynucleotide is integrated into the cellular genome at the incorporated genomic integration target recognition site by the gene editor protein.

8. EXAMPLES

8.1. Example 1: Attachment Site-Containing guide RNA (atgRNAs) capable of partial genomic replacement demonstrate improved attachment site genomic placement

[0289] In this example, attachment site-containing guide RNAs (atgRNAs) that are capable of site-specifically incorporating an integrase recognition site into a genomic locus were assessed for percent beacon placement (i.e., percent integration efficiency of integration recognition site). In particular, an atgRNA targeting the *LMBN1* locus and capable of site-specifically incorporating an integrase recognition site into the *LMBN1* locus were synthesized to allow for a 0 nucleotide separation (see atgRNA (v1) in FIGs. 3A-3C), 18 nucleotide separation (see, e.g., atgRNA (v2) in FIGs. 4A-4C where the separation is 18 nucleotides), or 36 nucleotide separation (see, e.g., atgRNA (v2) in FIGs. 4A-4C where the separation is 36 nucleotides), respectively, between the third genomic site (i.e., PBS binding site) and the second genomic site (i.e., clamp binding site). The synthesized atgRNAs incorporated a clamp segment 10 nucleotides long and a 38 nucleotide integration recognition site (e.g., AttB or AttP).

[0290] AtgRNAs were electroporated along with mRNA encoding a gene editor polypeptide (i.e., a Cas9-nickase-reverse transcriptase fusion) and a nicking guide RNA (ngRNA) into HEK 293 cells. In particular, the Neon electroporation system was used. AtgRNA and nicking gRNAs were ordered from Integrated DNA Technologies as Alt-R gRNA. As shown in Table 11, 1 µg of mRNA encoding the gene editor polypeptide, 120 pmol of atgRNA, 40 pmol of nicking sgRNA and 50,000 HEK293T cells were mixed in Buffer R and electroporated using 10-µl Neon tips using the following electroporation parameters: 1,150 V, 20 ms, two pulses. After electroporation, cells were plated in prewarmed 48-well plates with DMEM containing 10% FBS and incubated for 72 h before analysis.

Table 11.

#	Buffer R µL	atgRNA (100 µM)	sgRNA µL	nicking (100 µM)	nicki ng µL	DNA donor	Buffer T µL	mRNA 1 (gene editor)	µL	mRNA 2	µL	Cells per µl	uL of Cells	Total µL
5	4.4	LMNB1 atgRNA	1.2	LMNB1	0.4	-	0	1000ng/µl	1	-	0	1X10^4	5	12
6	2.5	LMNB1 atgRNA	1.2	LMNB1	0.4	552ng/µl	0.9	1000ng/µl	1	Bxb1 1000ng/µl	1	1X10^4	5	12
7	4.4	LMNB1 dual atg	0.8 x2	-	0.8	-	0	1000ng/µl	1	-	0	1X10^4	5	12
8	2.5	LMNB1 dual atg	0.8 x 2	-	0.8	552ng/µl	0.9	1000ng/µl	1	Bxb1 1000ng/µl	1	1X10^4	5	12
9	4.4	LMNB1 atg-RTm1	1.2	LMNB1	0.4	-	0	1000ng/µl	1	-	0	1X10^4	5	12
10	4.4	LMNB1 atg-RTm2	1.2	LMNB1	0.4	-	0	1000ng/µl	1	-	0	1X10^4	5	12

*Cells are either HEK293 or Jurkat cells

[0291] As shown in FIG. 5, an atgRNA having a clamp segment at least partially complementary to a second genomic site and a primer binding site complementary to a third genomic site, where the second genomic site and the third genomic site are separated by 18 base pairs demonstrated an 82% beacon placement efficiency as measured by droplet digital PCR (ddPCR). In contrast, an atgRNA having a clamp segment at least partially complementary to a second genomic site and a primer binding site complementary to a third genomic site, where the second genomic site are adjacent (i.e., a 0-base separation) demonstrated a 74% beacon placement efficiency.

8.2. Example 2: Attachment Site-Containing guide RNA (atgRNAs) capable of partial genomic replacement demonstrate improved attachment site genomic placement

[0292] In this example, attachment site-containing guide RNAs (atgRNAs) that are capable of site-specifically incorporating an integrase recognition site into a genomic locus were assessed for percent beacon placement. The AtgRNA were designed to target the *LMNB1* locus, the *hF9* locus, and the *ACTB* locus and with each atgRNA capable of site-specifically incorporating an integrase recognition site into the *LMBN1* locus, the *hF9* locus, and the *ACTB* locus, respectively. The atgRNAs targeting the *LMNB1* locus were synthesized to allow for a 0-36 base pair separation between the second genomic site and the third genomic site. The atgRNAs targeting the *hF9* locus were synthesized to allow for a 0-42 base pair separation between the second genomic site and the third genomic site. The atgRNAs targeting the *ACTB* locus were synthesized to allow for a 0-33 base pair separation between the second genomic site and the third genomic site. The synthesized atgRNAs incorporate the use of a 10 nucleotide clamp segment and a 38 nucleotide integration recognition site (e.g., AttB or AttP). atgRNA and nicking gRNAs were ordered from Integrated DNA Technologies as Alt-R gRNA.

[0293] The atgRNAs were transfected along with mRNA encoding a gene editor polypeptide (i.e., a Cas9-nickase-reverse transcriptase fusion) and a nicking guide RNA (ngRNA) into HEK 293 cells using Lipofectamine MessengerMax (Thermo Fisher) according to the manufacturer's instructions. In particular, 1 μg of a gene editor polypeptide (Cas9-nickase linked to a reverse transcriptase) encoding mRNA, 120 pmol of atgRNA, 40 pmol of nicking sgRNA were transfected into 50,000 HEK293T cells. After transfection, cells were allowed to recover for 72 hours in DMEM containing 10% FBS before analysis.

[0294] FIG. 7A shows percent beacon placement (i.e., efficiency of integration of the integration recognition site) for atgRNAs targeting the *LMNB1* locus having a clamp segment at least partially complementary to a second genomic site and a primer binding site complementary to a third genomic site, where the second genomic site and the third genomic site are separated by 0 to 36 base pairs (see FIG. 7A "clamp+0" (i.e., 0 base pair separation) to "clamp+36" (i.e., 36 base pair separation). Percent beacon placement was highest for atgRNA having a 10 base pair separation between the second and third genomic sites (see "clamp+10" in FIG. 7A).

[0295] FIG. 7B shows percent beacon placement (i.e., efficiency of integration of the integration recognition site) for atgRNAs targeting the *hF9* locus having a clamp segment at least partially complementary to a second genomic site and a primer binding site complementary to a

third genomic site, where the second genomic site and the third genomic site are separated by 0 to 42 base pairs (see FIG. 7B "clamp+0" (i.e., 0 base pair separation) to "clamp+42" (i.e., 42 base pair separation). Percent beacon placement was highest for atgRNA having a 31 base pair separation between the second and third genomic sites (see "clamp+31" in FIG. 7B).

[0296] FIG. 7C shows percent beacon placement (i.e., efficiency of integration of the integration recognition site) for atgRNAs targeting the *ACTB* locus having a clamp segment at least partially complementary to a second genomic site and a primer binding site complementary to a third genomic site, where the second genomic site and the third genomic site are separated by 0 to 33 base pairs (see FIG. 7C "clamp+0" (i.e., 0 base pair separation) to "clamp+33" (i.e., 33 base pair separation). Percent beacon placement is highest for atgRNA having a 16 base pair separation between the second and third genomic sites (see "clamp+16" in FIG. 7C).

[0297] Overall, this data showed the distance between the second genomic site and the third genomic site can impact percent beacon placement.

8.3. Example 3: PASTE with atgRNAs capable of partial genomic replacement

[0298] A PASTE readout in HEK293 cells is used to identify functionally active atgRNAs capable of inserting an integration recognition site where insertion includes partial genomic replacement. Droplet digital PCR (ddPCR) in a 96-well plate format is applied to verify gene insertion into the ACTB locus. Using Lipofectamine 3000 transfection reagent, 100ng total DNA at a plasmid ratio of 1:1:4 of Cas9(nickase)-RT-BxB1 : nicking guide : atgRNA is delivered to 30,000 HEK293 cells. The percent integration across tested atgRNAs capable of partial genomic replacement is verified via ddPCR.

9. EQUIVALENTS AND INCORPORATION BY REFERENCE

[0299] 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.

[0300] While the disclosure 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 disclosure.

WHAT IS CLAIMED IS:

1. An attachment site containing guide RNA (atgRNA) comprising from 5' to 3':
    a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence;
    a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence;
    a reverse transcriptase (RT) template comprising:
        an integration recognition site; and
        a clamp segment comprising a sequence where upon being reverse transcribed the sequence is at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence; and
    a primer binding site complementary to a third genomic site of the double-stranded target DNA sequence, wherein the third genomic site is on a second strand of the target DNA sequence,
    wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site and the clamp segment.

2. An attachment site containing guide RNA (atgRNA) comprising from 5' to 3':
    a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence;
    a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence;
    a reverse transcriptase (RT) template comprising:
        an integration recognition site; and
        a clamp segment comprising a sequence where upon being reverse transcribed comprises a sequence at least partially complementary to a second genomic site of first strand of the double-stranded target DNA sequence; and
    a primer binding site comprising:
        a sequence complementary to a third genomic site of the double-stranded DNA target sequence, wherein the third genomic site is on a second strand of the target DNA sequence; and

an anti-reverse transcriptase template (anti-RTT) sequence comprising a sequence at least partially complementary to all or a portion of the RT template,

wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site and the clamp segment.

3. An attachment site containing guide RNA (atgRNA), comprising from 5' to 3':

a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence;

a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence;

a clamp segment at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence;

a reverse transcriptase (RT) template comprising an integration recognition site; and

a primer binding site complementary to a third genomic site of the double-stranded target DNA sequence, wherein the third genomic site is on a second strand of the target DNA sequence;

wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site.

4. An attachment site containing guide RNA (atgRNA) comprising from 5' to 3':

a spacer sequence complementary to a first genomic site of a first strand of a double-stranded target DNA sequence;

a scaffold capable of associating with a DNA binding nickase domain and guiding the DNA binding nickase domain to the target DNA sequence;

a clamp segment at least partially complementary to a second genomic site of the first strand of the double-stranded target DNA sequence;

a reverse transcriptase (RT) template comprising an integration recognition site; and

a primer binding site comprising:

a sequence complementary to a third genomic site of the double-stranded DNA target sequence, wherein the third genomic site is on a second strand of the target DNA sequence; and

an anti-reverse transcriptase template (anti-RTT) sequence comprising a sequence at least partially complementary to all or a portion of the RT template,

wherein the reverse transcriptase template encodes for a first single-stranded DNA sequence comprising the integration recognition site.

5. The atgRNA of any one of claims 1-4, wherein, upon introducing the atgRNA into a cell with a gene editor polypeptide, all or a portion of the single-stranded DNA sequence comprising the integration recognition site is site-specifically integrated into the double-stranded target DNA sequence.

6. The atgRNA of any one of claims 1-5, wherein, upon introducing the atgRNA into a cell with a gene editor polypeptide, the atgRNA comprising the clamp segment enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to the integration efficiency of the integration recognition site at the double-stranded target DNA sequence when using an atgRNA that does not comprise the clamp segment.

7. The atgRNA of any one of claims 1-6, wherein the atgRNA has a length of 150 to 190 nucleotides.

8. The atgRNA of claim 7, wherein the atgRNA has a length of 160 to 180 nucleotides.

9. The atgRNA of any one of claims 1-8, wherein the clamp segment has a length of 1 nucleotide to 40 nucleotides.

10. The atgRNA of claim 9, wherein the clamp segment has a length of 1 to 10 nucleotides, 11 to 20 nucleotides, 21 to 30 nucleotides, or 31 to 40 nucleotides.

11. The atgRNA of claim 10, wherein the clamp segment has a length of 10 nucleotides.

12. The atgRNA of any one of claims 1-11, wherein the RT template has a length of 10 nucleotides to 200 nucleotides.

13. The atgRNA of claim 12, wherein the RT template has a length of 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 or 50 nucleotides.

14. The atgRNA of any one of claims 1-13, wherein the second genomic site and the third genomic site are separated by one or more genomic bases in the double-stranded target DNA sequence.

15. The atgRNA of any one of claims 1-14, wherein the second genomic site and the third genomic site are separated by 1 to 10 genomic bases, 11 to 20 genomic bases, 21 to30 genomic bases, 31 to 40 genomic bases, or 41 to 45 genomic bases in the double-stranded target DNA sequence.

16. The atgRNA of any one of claims 1-14, wherein the second genomic site and the third genomic site are separated by 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, or 45 genomic bases in the double-stranded target DNA sequence.

17. The atgRNA of any one of claims 1-13, wherein the second genomic site and the third genomic site are adjacent in the double-stranded target DNA sequence.

18. The atgRNA of any one of claims 1-17, wherein the clamp segment comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity to the second genomic site.

19. The atgRNA of any one of claims 1-17, wherein the clamp segment comprises a sequence that is capable of binding to the second genomic site of the first strand of the double-stranded target DNA sequence with at least a first nucleotide mismatch.

20. The atgRNA of claim 19, wherein the clamp segment comprises a sequence that is capable of binding to the second genomic site of the first strand of the double-stranded target DNA sequence with a first nucleotide mismatch and a second nucleotide mismatch.

21. The atgRNA of claim 20, wherein the first nucleotide mismatch and the second nucleotide mismatch are consecutive.

22. The atgRNA of claim 20, wherein the first nucleotide mismatch and the second nucleotide mismatch are non-consecutive.

23. The atgRNA of any one of claims 2 or 4-22, wherein the anti-RTT sequence comprises a sequence having 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence complementarity to all or a portion the RT template.

24. The atgRNA of claim 23, wherein the anti-RTT sequence has a length between 1 nucleotide to 20 nucleotides, 5 nucleotides to 15 nucleotides, or 10 nucleotides to 20 nucleotides.

25. The atgRNA of claim 24, wherein the anti-RTT sequence has a length of 10 nucleotides.

26. The atgRNA of any one of claims 2 or 4-25, wherein, upon introducing the atgRNA into a cell, the anti-RTT sequence increases stability of the atgRNA as compared to an atgRNA that does not comprise the anti-RTT sequence.

27. A system for site-specifically incorporating an integration recognition site into a mammalian cell genome at a target DNA sequence, comprising:

the attachment site containing guide RNA (atgRNA) of any one of claims 1-26; and

a gene editor polypeptide comprising a DNA binding nickase domain linked to a reverse transcriptase domain capable of incorporating the integration recognition site into the target DNA sequence,

wherein, upon introducing into the cell, the system integrates the integration recognition site into the target DNA sequence.

28. The system of claim 27, wherein, upon introducing the system into the cell, the atgRNA comprising the clamp segment enhances integration efficiency of the integration recognition site at the target DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment.

29. A method for site-specifically incorporating an integration recognition site into a mammalian cell genome at a target DNA sequence, comprising:

incorporating an integration target recognition site into the genome by delivering into the cell:

the attachment site containing guide RNA (atgRNA) of any one of claims 1-26; and

a gene editor polypeptide or polynucleotide encoding a gene editor polypeptide, wherein the gene editor polypeptide comprises a DNA binding nickase domain linked to a reverse transcriptase domain, wherein the integration recognition site is incorporated into the target DNA sequence by the gene editor polypeptide; and

optionally, a nicking gRNA.

30. The method of claim 29, wherein the method using the atgRNA comprising the clamp segment enhances integration efficiency of the integration recognition site at the target DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment.

31. A system for site-specifically integrating a donor polynucleotide template into a mammalian cell genome at a target DNA sequence, comprising:

the attachment site containing gRNA (atgRNA) of any one of claims 1-26;

a gene editor polypeptide comprising a DNA binding nickase domain linked to a reverse transcriptase domain capable of incorporating the integration recognition site into the target DNA sequence,

an integration enzyme; and

a donor polynucleotide template linked to a sequence that is an integration cognate of the integration recognition site present in the atgRNA,

whereby the integration enzyme integrates the donor polynucleotide template into the target DNA sequence.

32. The system of claim 31, wherein, upon introducing the system into the cell, the atgRNA comprising the clamp enhances integration efficiency of the integration recognition site at the target DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment.

33. A method for site-specifically integrating a donor polynucleotide template into a mammalian cell genome at a target DNA sequence, comprising:

incorporating an integration recognition site into the genome by delivering into the cell:

the attachment site containing guide RNA (atgRNA) of any one of claims 1-26; and

a gene editor polypeptide or polynucleotide encoding a gene editor polypeptide, wherein the gene editor polypeptide comprises a DNA binding nickase domain linked to a reverse transcriptase domain and is capable of incorporating the integration recognition site into the target DNA sequence; and

optionally, a nicking gRNA; and

integrating the donor polynucleotide template into the genome by delivering into the cell:

an integration enzyme; and

a donor polynucleotide template, wherein the donor polynucleotide template is linked to a sequence that is an integration cognate of the integration recognition site present in the atgRNA, and wherein the donor polynucleotide template is integrated into the genome at the incorporated genomic integration recognition site by the integration enzyme.

34. The method of claim 33, wherein the method enhances integration efficiency of the integration recognition site at the double-stranded target DNA sequence as compared to the integration efficiency of the integration recognition site at the target DNA sequence when using an atgRNA that does not comprise the clamp segment.

AttP mutants for improved kinetics

WT AttP

GTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCA

AttP1

GTGGTTTGTTTCATGAACAACTGCAGTCTCAGTGGTGTACGGTACAAACCCA

AttP2

GTGGTTTGTTTCTTAACCACACAGCGGTCTCAGTGGTGTACGGTACAAACCCA

AttP3

GTGGTTTGTTTCATGAACCACTCAAGTCTCAGTGGTGTACGGTACAAACCCA

FIG. 1A

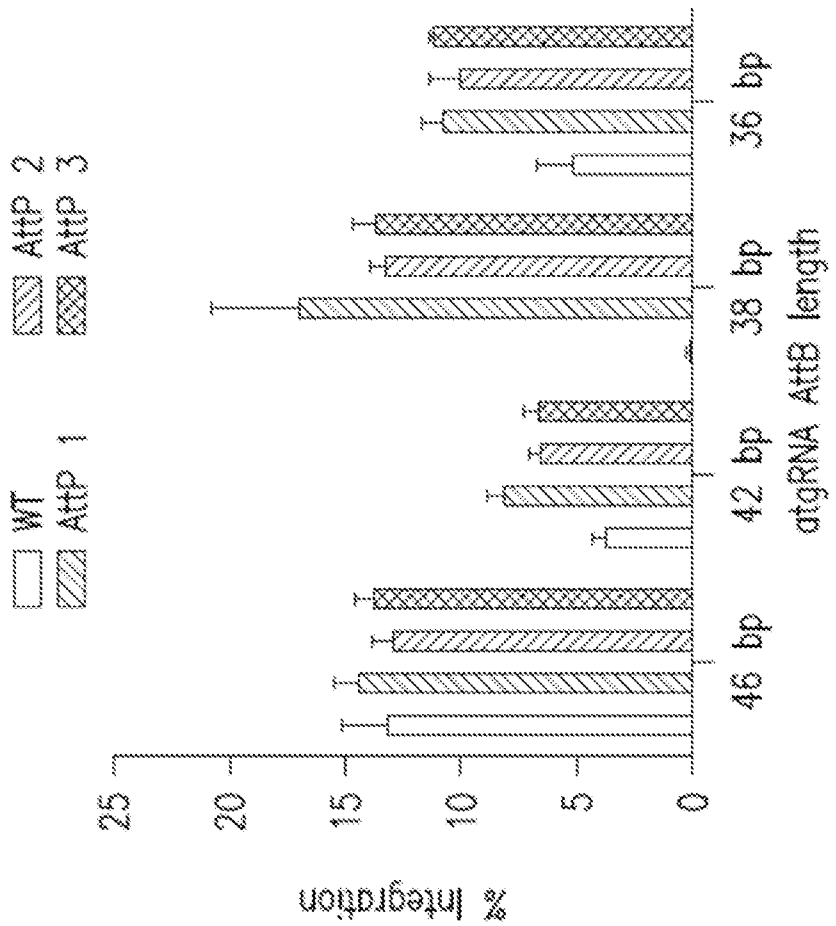
% Integration
25
20
15
10
5
0
WT
AttP 1
AttP 2
AttP 3
46 bp
42 bp
38 bp
36 bp
otgRNA AttB length

FIG. 1B

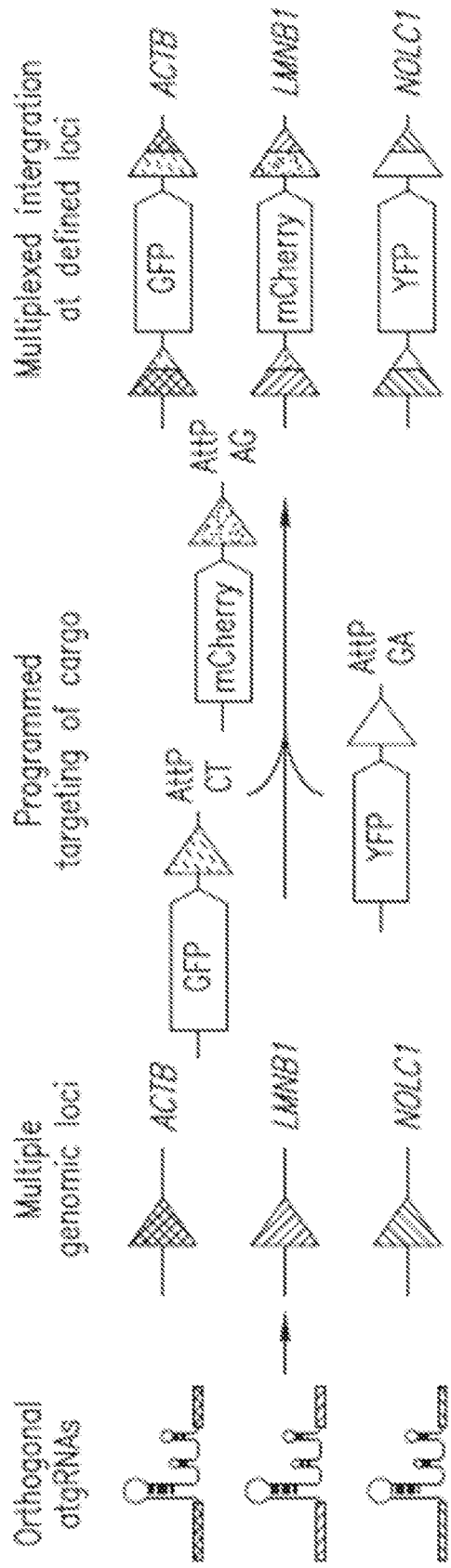
Orthogonal dtgRNAs
Multiple genomic loci
ACTB
LMNB1
NOLC1
Programmed targeting of cargo
GFP
AttP CT
mCherry
AttP AG
YFP
AttP GA
Multiplexed integration at defined loci
GFP
ACTB
mCherry
LMNB1
YFP
NOLC1

FIG. 1C

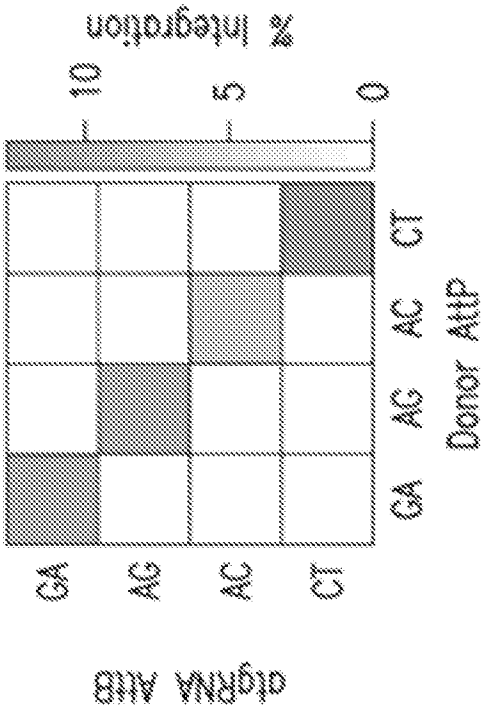
% integration
10
5
0
atgRNA AttB
GA
AG
AC
CT
GA
AG
AC
CT
Donor AttP

FIG. 1D

% integration
20
15
10
5
0
Locus
Cargo
Size (bp)
ACTB
GFP
779
LMNB1
mCherry
797
NOLC1
YFP
815
ACTB atgRNA
LMNB1 atgRNA
NOLC1 atgRNA
SpCas9–RT

FIG. 1E

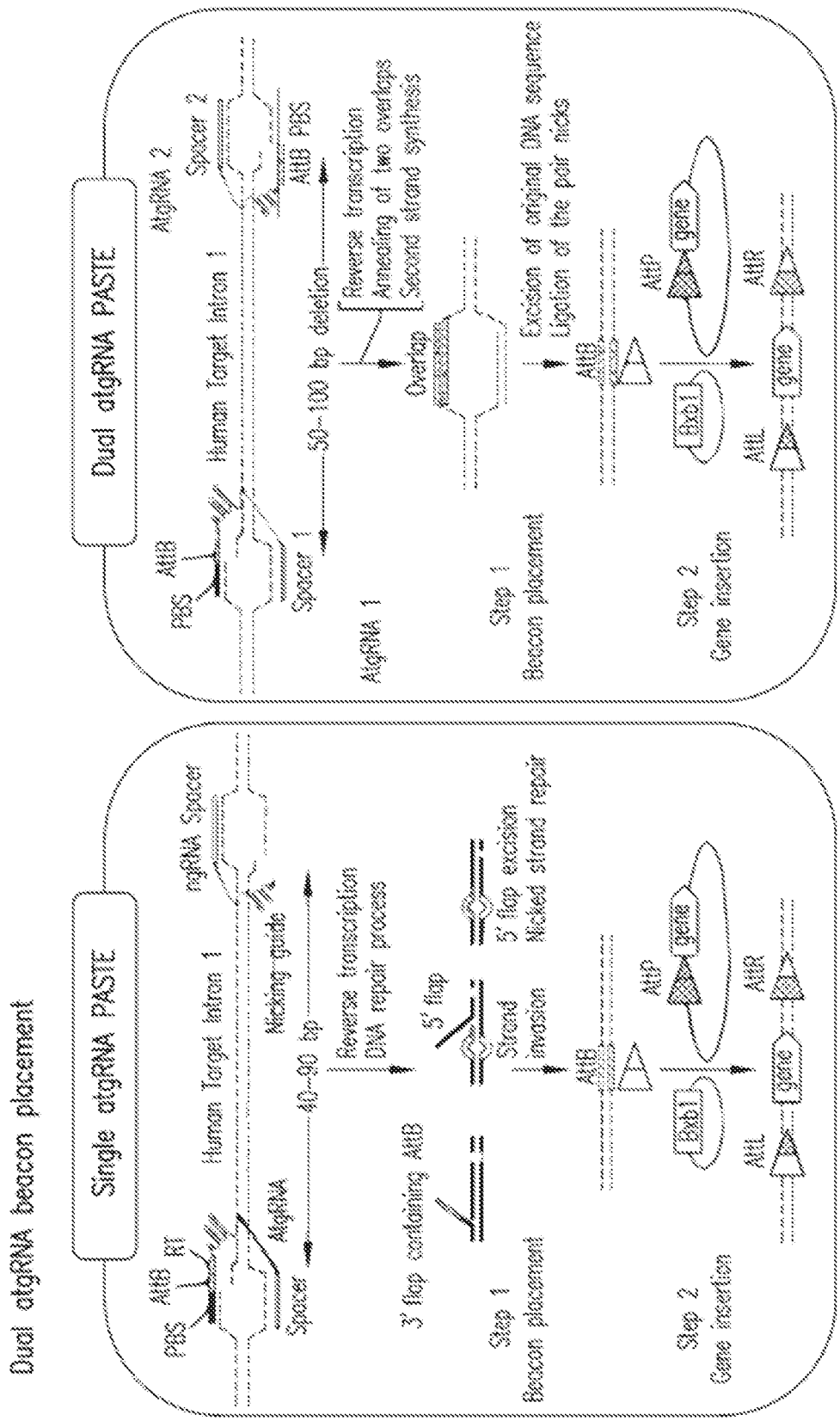
Dual atgRNA beacon placement
Single atgRNA PASTE
PBS
AttB
RT
Human Target Intron 1
ngRNA Spacer
AtgRNA
Spacer
Nicking guide
40–90 bp
Reverse transcription
DNA repair process
3' flap containing AttB
5' flap
Step 1
Beacon placement
Strand
invasion
5' flap excision
Nicked strand repair
AttB
AttP
gene
ExoI
Step 2
Gene insertion
AttL
AttR
gene
Dual atgRNA PASTE
AtgRNA 2
Spacer 2
Human Target Intron 1
AttB
PBS
Spacer 1
50–100 bp deletion
AttB PBS
AtgRNA 1
Reverse transcription
Annealing of two overlaps
Second strand synthesis
Overlap
Step 1
Beacon placement
Excision of original DNA sequence
Ligation of the pair nicks
AttB
AttP
gene
ExoI
Step 2
Gene insertion
AttL
AttR
gene

FIG. 2

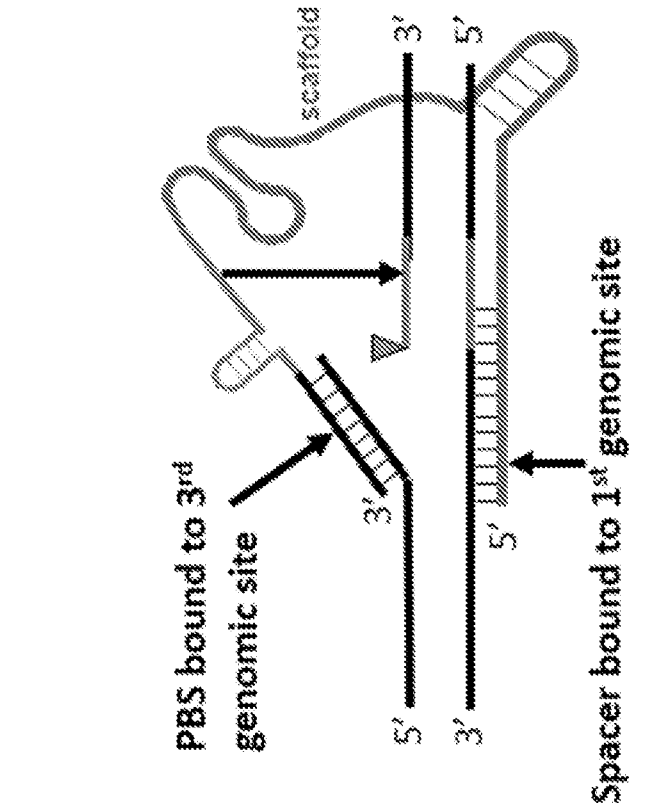
PBS bound to 3rd genomic site
scaffold
3'
5'
3'
5'
5'
3'
Spacer bound to 1st genomic site

FIG. 3B

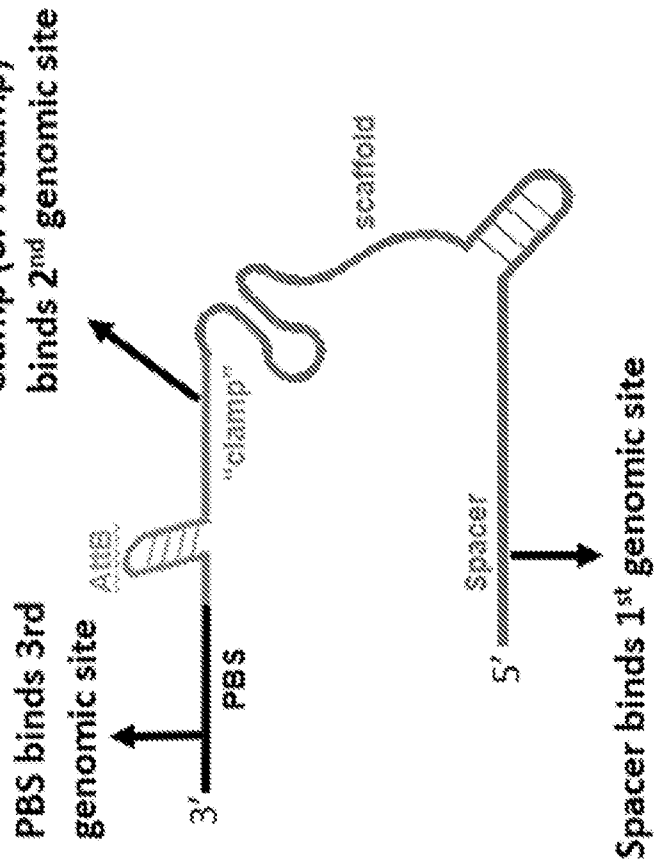
Clamp (or rcClamp) binds 2nd genomic site
PBS binds 3rd genomic site
Anti
"clamp"
scaffold
3'
PBS
Spacer
5'
Spacer binds 1st genomic site

FIG. 3A

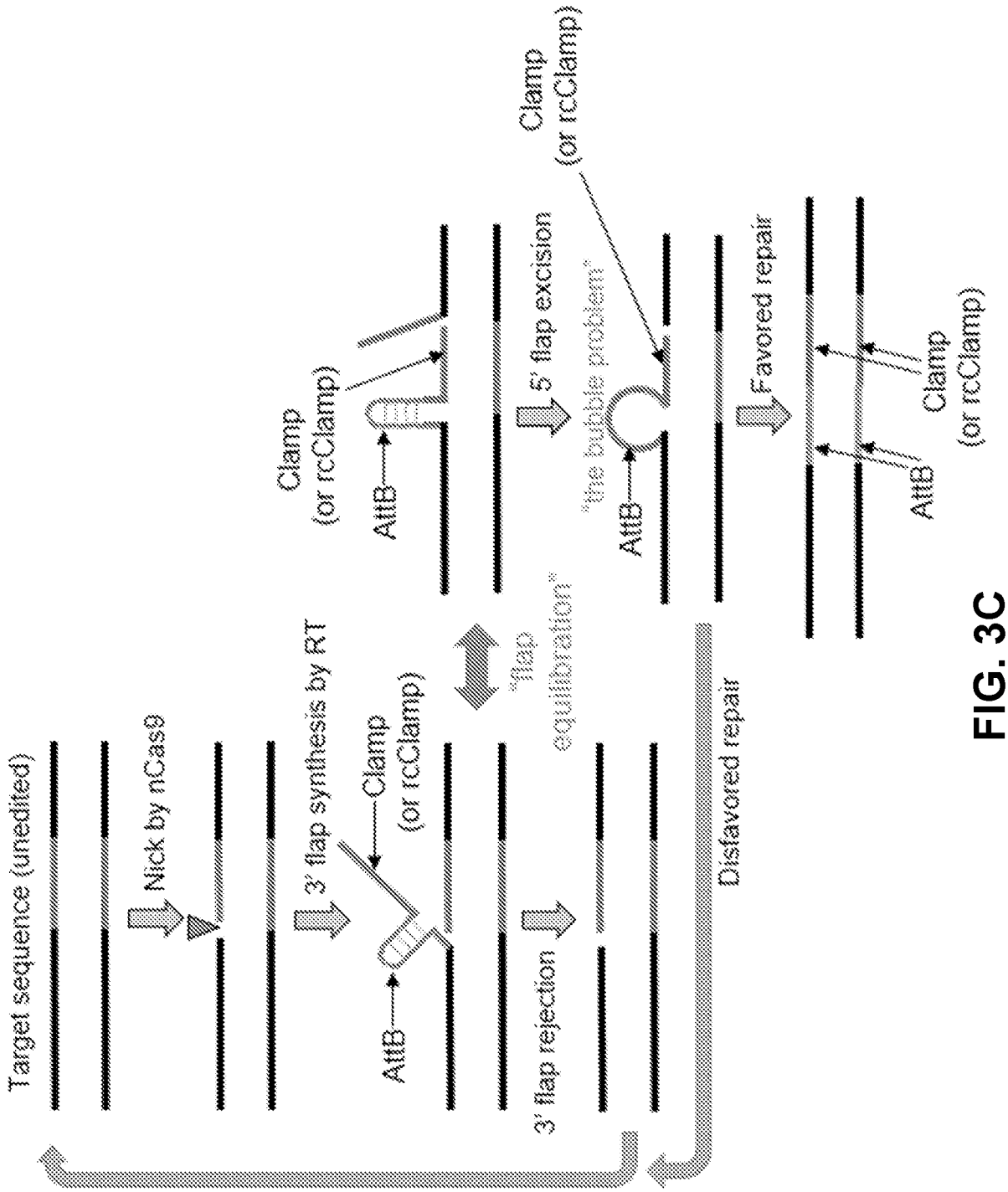
Target sequence (unedited)
Nick by nCas9
3' flap synthesis by RT
Clamp
(or rcClamp)
AttB
3' flap rejection
"flap equilibration"
Clamp
(or rcClamp)
AttB
5' flap excision
"the bubble problem"
Clamp
(or rcClamp)
AttB
Favored repair
Disfavored repair
AttB
Clamp
(or rcClamp)

FIG. 3C

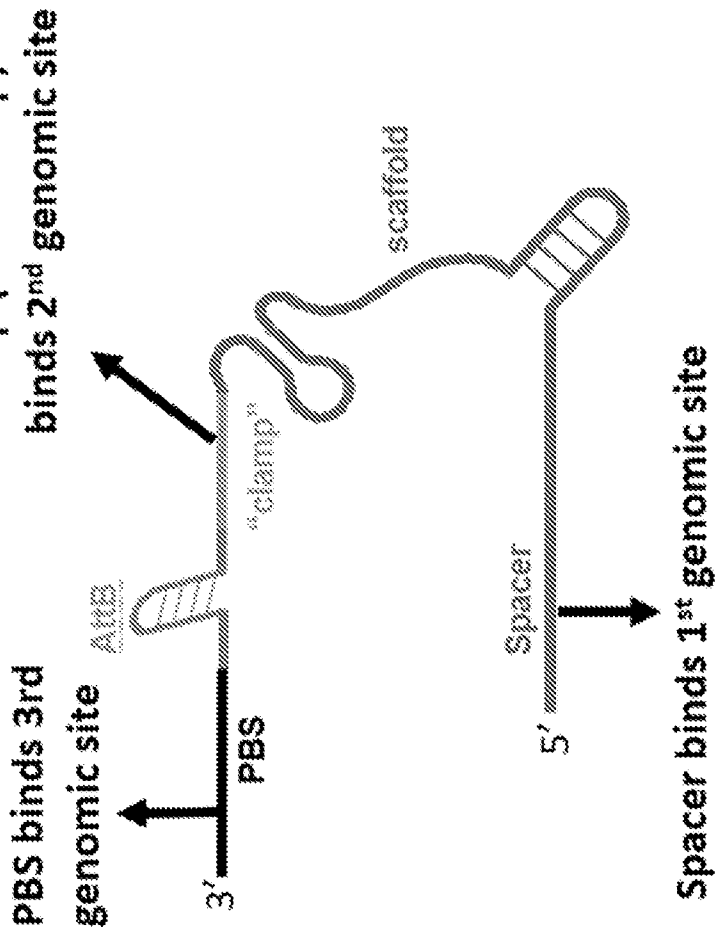
PBS binds 3rd genomic site
3'
PBS
Anti
"clamp"
Clamp (or rcClamp) binds 2nd genomic site
scaffold
Spacer
5'
Spacer binds 1st genomic site

FIG. 4A

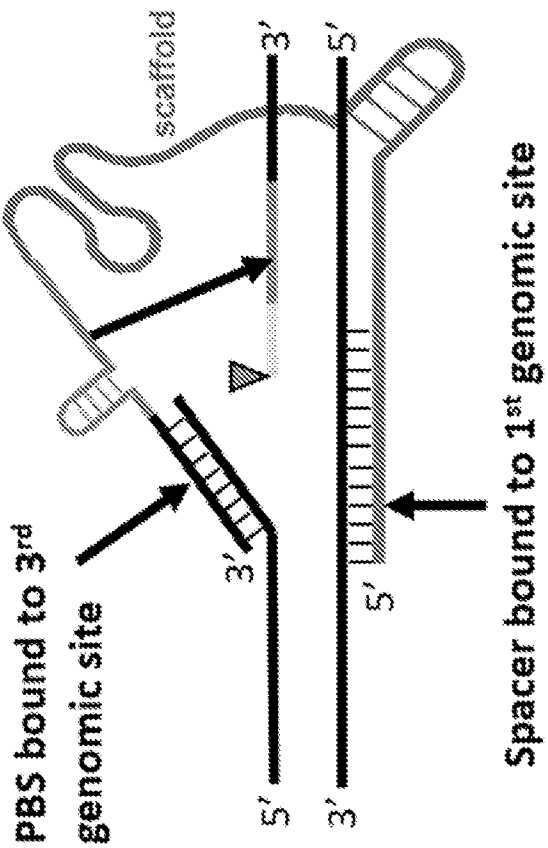
PBS bound to 3rd genomic site
3'
5'
scaffold
3'
5'
5'
3'
Spacer bound to 1st genomic site

FIG. 4B

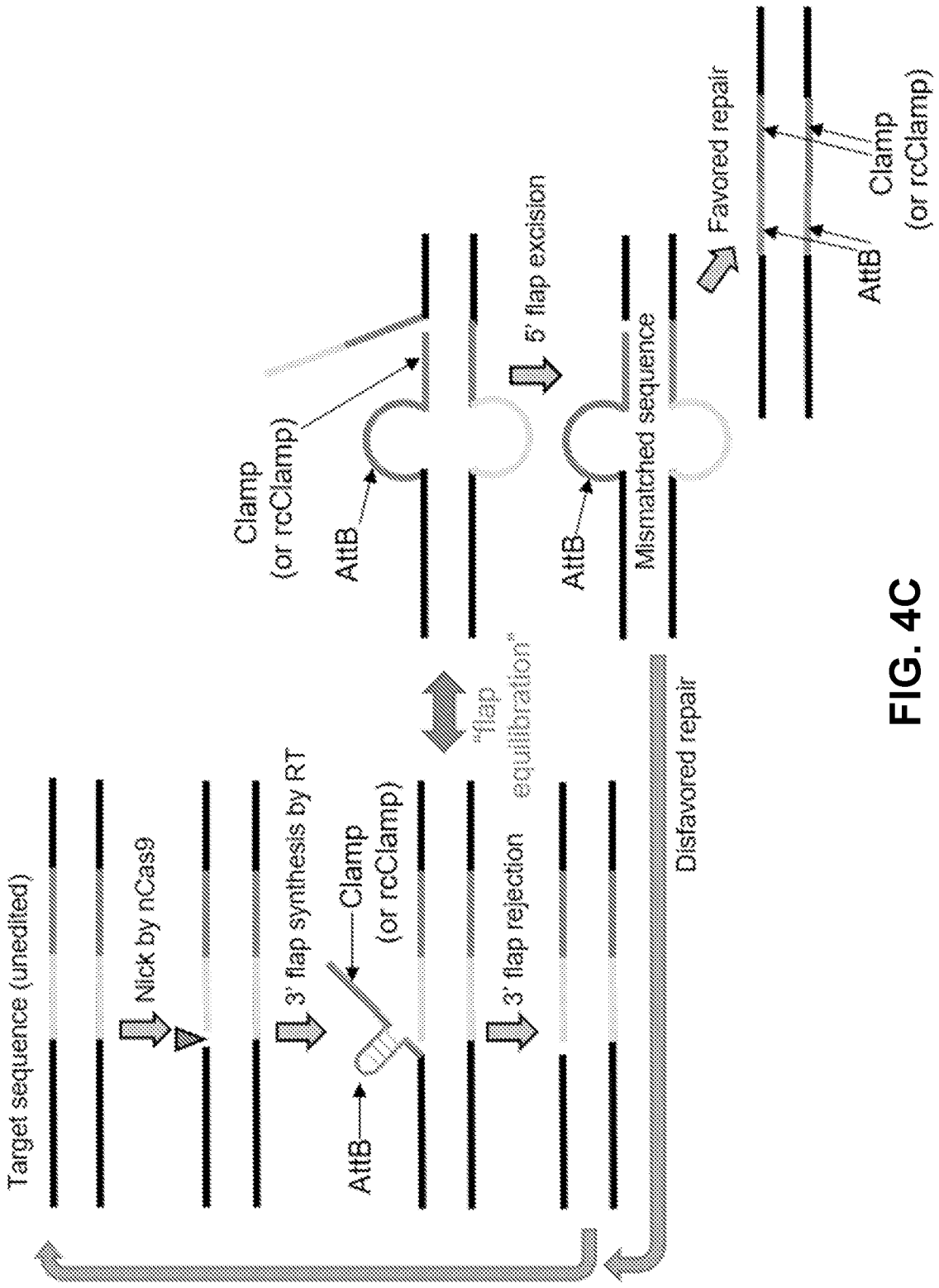
Target sequence (unedited)
Nick by nCas9
3' flap synthesis by RT
Clamp
(or rcClamp)
AttB
3' flap rejection
"flap equilibration"
Clamp
(or rcClamp)
AttB
5' flap excision
AttB
Mismatched sequence
Disfavored repair
Favored repair
AttB
Clamp
(or rcClamp)

FIG. 4C

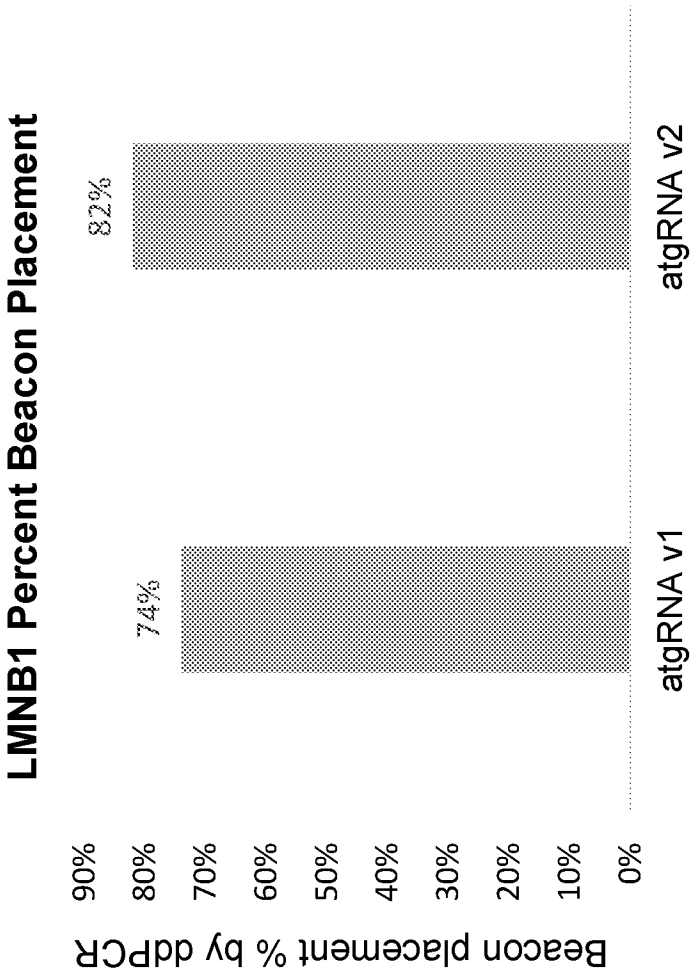
LMNB1 Percent Beacon Placement
Beacon placement % by ddPCR
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
74%
82%
atgRNA v1
atgRNA v2

FIG. 5

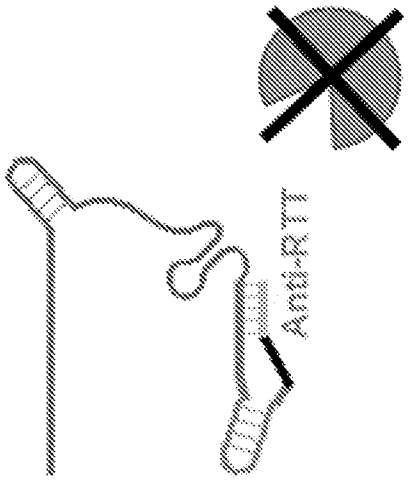

FIG. 6B

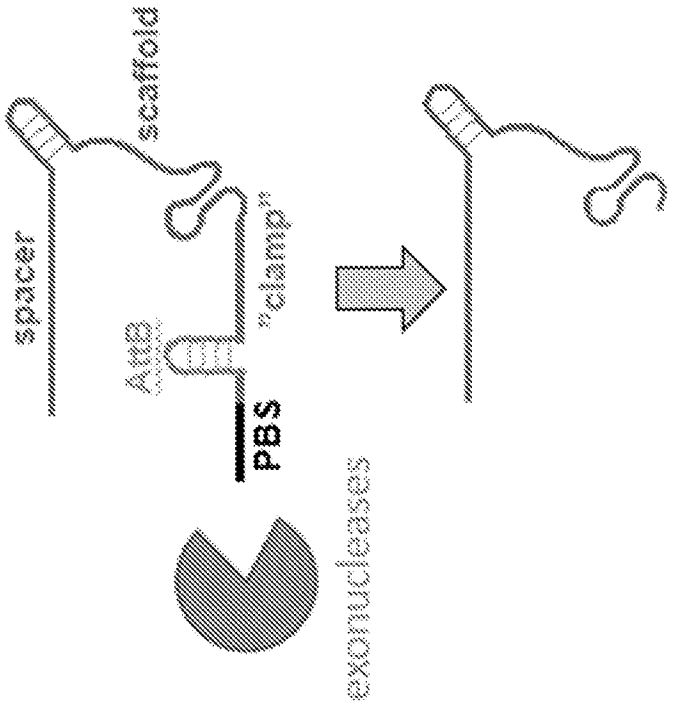

FIG. 6A

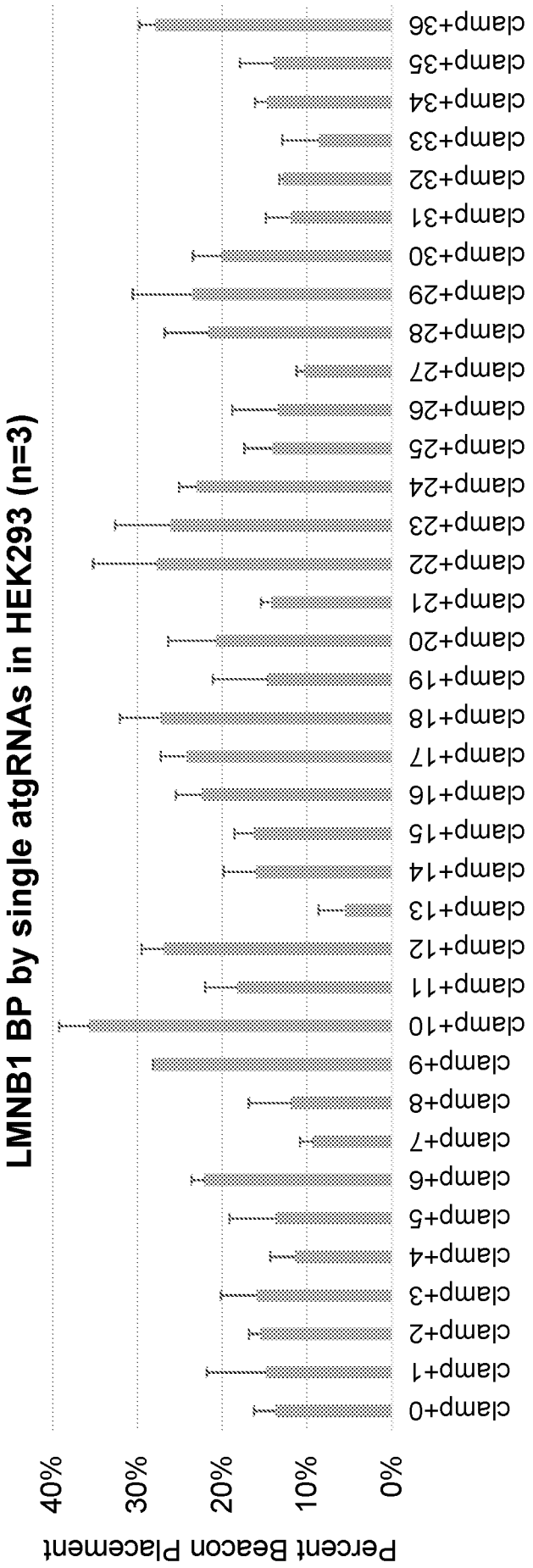
LMNB1 BP by single atgRNAs in HEK293 (n=3)
Percent Beacon Placement
40%
30%
20%
10%
0%
clamp+0
clamp+1
clamp+2
clamp+3
clamp+4
clamp+5
clamp+6
clamp+7
clamp+8
clamp+9
clamp+10
clamp+11
clamp+12
clamp+13
clamp+14
clamp+15
clamp+16
clamp+17
clamp+18
clamp+19
clamp+20
clamp+21
clamp+22
clamp+23
clamp+24
clamp+25
clamp+26
clamp+27
clamp+28
clamp+29
clamp+30
clamp+31
clamp+32
clamp+33
clamp+34
clamp+35
clamp+36

FIG. 7A

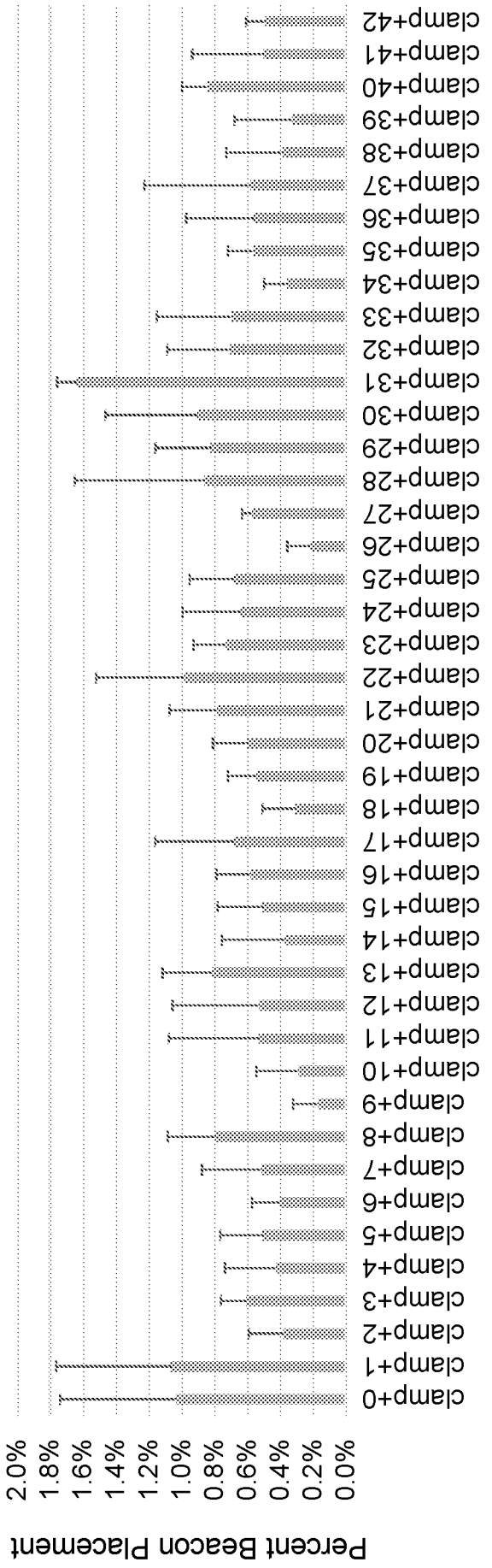
hF9 s17 bp by single atgRNAs in HEK293 (n=3)
Percent Beacon Placement
2.0%
1.8%
1.6%
1.4%
1.2%
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
clamp+0
clamp+1
clamp+2
clamp+3
clamp+4
clamp+5
clamp+6
clamp+7
clamp+8
clamp+9
clamp+10
clamp+11
clamp+12
clamp+13
clamp+14
clamp+15
clamp+16
clamp+17
clamp+18
clamp+19
clamp+20
clamp+21
clamp+22
clamp+23
clamp+24
clamp+25
clamp+26
clamp+27
clamp+28
clamp+29
clamp+30
clamp+31
clamp+32
clamp+33
clamp+34
clamp+35
clamp+36
clamp+37
clamp+38
clamp+39
clamp+40
clamp+41
clamp+42

FIG. 7B

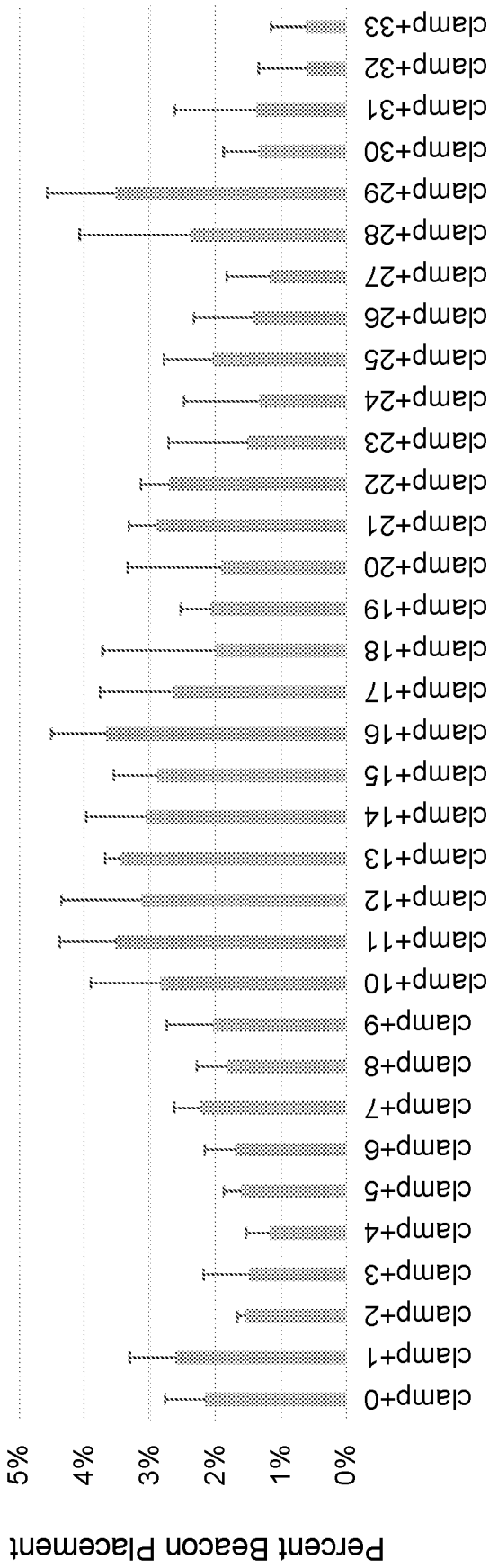
ACTB bp by single atgRNAs in HEK293 (n=3)
Percent Beacon Placement
0%
1%
2%
3%
4%
5%
clamp+0
clamp+1
clamp+2
clamp+3
clamp+4
clamp+5
clamp+6
clamp+7
clamp+8
clamp+9
clamp+10
clamp+11
clamp+12
clamp+13
clamp+14
clamp+15
clamp+16
clamp+17
clamp+18
clamp+19
clamp+20
clamp+21
clamp+22
clamp+23
clamp+24
clamp+25
clamp+26
clamp+27
clamp+28
clamp+29
clamp+30
clamp+31
clamp+32
clamp+33

FIG. 7C

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2023/066609

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/113 C12N15/90
ADD. C12N9/12

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Ioannidi Eleonora I. ET AL: "Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed integrases", bioRxiv, 1 November 2021 (2021-11-01), pages 1-61, XP093015571, DOI: 10.1101/2021.11.01.466786 Retrieved from the Internet: URL:https://www.biorxiv.org/content/10.1101/2021.11.01.466786v1.full.pdf [retrieved on 2023-01-19] figure 1 the whole document -/--	1-34

[x] Further documents are listed in the continuation of Box C.

[x] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 September 2023	22/09/2023

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Spindler, Mark-Peter

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	-& Ioannidi Eleonora I ET AL: "Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed integrases - Supplementary Information", bioRXiV, 1 November 2021 (2021-11-01), pages 1-37, XP093079476, Retrieved from the Internet: URL:https://www.biorxiv.org/content/biorxiv/early/2021/11/01/2021.11.01.466786/DC1/embed/media-1.pdf?download=true [retrieved on 2023-09-06] table S3 the whole document -----	
X	WO 2022/087235 A1 (MASSACHUSETTS INST TECHNOLOGY [US]) 28 April 2022 (2022-04-28) figures 1, 29; examples 2, 3, 13, 14, 18; table 4; sequences 84-141 the whole document -----	1-34
X	ANZALONE ANDREW V. ET AL: "Search-and-replace genome editing without double-strand breaks or donor DNA", NATURE, vol. 576, no. 7785, 21 October 2019 (2019-10-21), pages 149-157, XP055980447, London ISSN: 0028-0836, DOI: 10.1038/s41586-019-1711-4 figures 5, S3 the whole document -& ANZALONE ANDREW V. ET AL: "Search-and-replace genome editing without double-strand breaks or donor DNA - SUPPLEMENTARY INFORMATION", NATURE, vol. 576, no. 7785, 21 October 2019 (2019-10-21), pages 149-157, XP055899878, London ISSN: 0028-0836, DOI: 10.1038/s41586-019-1711-4 page 12; table S3 -----	1-34
X	WO 2022/032085 A1 (JACKSON LAB [US]) 10 February 2022 (2022-02-10) figures 1-3; examples 1-3; sequences 3, 13 the whole document -----	1-34

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2023/066609

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NELSON JAMES W ET AL: "Engineered pegRNAs improve prime editing efficiency", NATURE BIOTECHNOLOGY, NATURE PUBLISHING GROUP US, NEW YORK, vol. 40, no. 3, 4 October 2021 (2021-10-04), pages 402-410, XP037720612, ISSN: 1087-0156, DOI: 10.1038/S41587-021-01039-7 [retrieved on 2021-10-04] figure 1 the whole document -----	1-34
A	WO 2022/067130 A2 (BROAD INST INC [US]; HARVARD COLLEGE [US]) 31 March 2022 (2022-03-31) figures 3, 71, 92, 108; example 3 the whole document -----	1-34

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2023/066609

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed.

   b. [ ] furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

      [ ] accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. [ ] With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2023/066609

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2022087235 A1	28-04-2022	AU 2021364781 A1	01-06-2023
		CA 3196116 A1	28-04-2022
		CN 116419975 A	11-07-2023
		EP 4232583 A1	30-08-2023
		IL 301368 A	01-05-2023
		KR 20230091894 A	23-06-2023
		US 2022145293 A1	12-05-2022
		US 2022154224 A1	19-05-2022
		US 2023135673 A1	04-05-2023
		US 2023279391 A1	07-09-2023
		WO 2022087235 A1	28-04-2022
WO 2022032085 A1	10-02-2022	EP 4176058 A1	10-05-2023
		WO 2022032085 A1	10-02-2022
WO 2022067130 A2	31-03-2022	AU 2021350835 A1	27-04-2023
		CA 3193099 A1	31-03-2022
		EP 4217490 A2	02-08-2023
		WO 2022067130 A2	31-03-2022