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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 September 2018 (20.09.2018)

W!P O | PCT

(10) International Publication Number
WO 2018/170015 Al

(51) International Patent Classification:
A61K 38/00 (2006.01) *C07K 16/28* (2006.01)
A61K 39/395 (2006.01) *C12N 5/10* (2006.01)
A61P 35/00 (2006.01) *C12N 15/09* (2006.01)

(21) International Application Number:
PCT/US2018/022258

(22) International Filing Date:
13 March 2018 (13.03.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/471,267 14 March 2017 (14.03.2017) US
62/614,875 08 January 2018 (08.01.2018) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, California 94607-5200 (US).

(72) Inventors: **MALI, Prashant;** c/o UC San Diego, 9500 Oilman Drive, Mail Code: 0910, La Jolla, California 92093 (US). **COLLADO, Ana Moreno;** c/o UC San Diego, 9500 Oilman Drive, Mail Code: 0910, La Jolla, California 92093 (US). **PALMER, Nathan;** c/o UC San Diego, 9500 Oilman Drive, Mail Code: 0910, La Jolla, California 92093 (US).

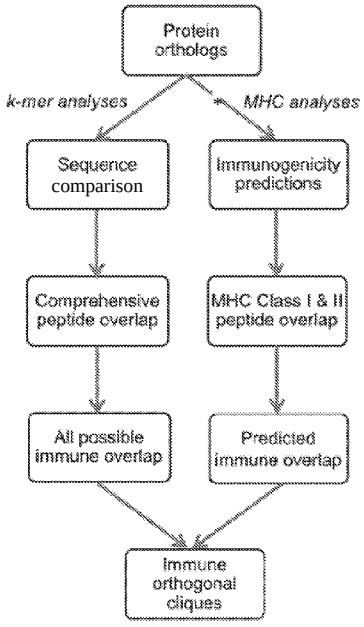
(74) Agent: **KONSKI, Antoinette F.** et al; FOLEY & LARDNER LLP, 3000 K Street, N.W., Suite 600, Washington, District of Columbia 20007-5 109 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BO, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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).

(54) Title: ENGINEERING CRISPR CAS9 IMMUNE STEALTH

FIGURE 5F



(57) Abstract: Described herein are methods of avoiding an immune response in a subject being administered a regimen requiring Cas9 in order to optimize and broaden the application of CRIPSR based therapeutics comprising administering immune orthogonal Cas9. Also described herein are methods to modify a Cas9 protein by swapping highly immunogenic peptides or amino acids with less immunogenic counterparts. These methods are particularly useful to enable the application of Cas9 arsenal for repeat treatments. Further provided are Cas9 proteins modified to reduce immunogenicity.

WO 2018/170015 A1

[Continued on next page]

Published:
— *with international search report (Art. 21(3))*

ENGINEERING CRISPR CAS9 IMMUNE STEALTH

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. 119(e) to U.S. Serial No. 62/471,267, filed March 14, 2017, and U.S. Serial No. 62/614,875, filed January 8, 2018, the entirety of which are incorporated by reference herein.

BACKGROUND

[0002] Immune responses against in vivo CRISPR/Cas9 for genome engineering purposes remain poorly characterized. Cas9 is a foreign protein, with prokaryotic origins, and could potentially elicit a strong immune response, which could ultimately result in the elimination of gene-edited cells or of the Cas9 protein by cytotoxic T cell mediated immune responses.

[0003] Cas9 specific cytotoxic cellular responses may be elicited due to the need of recurrent treatments for two reasons: 1) the current overall efficacy of in vivo CRISPRCas9 mediated genome editing is low which can require repetitive treatments, and 2) if genome regulation by dCas9 is a referred gene therapy method, repeat treatments will be necessary for continued repression/activation. Additionally, under certain delivery systems, such as AAV mediated delivery, Cas9 may have long term expression, further increasing the potential of Cas9 specific cytotoxic cellular responses, hampering long-term therapeutic efficacy. New methods of administering Cas9 that reduce immunogenicity to evade immune detection are needed. This disclosure addresses this need and provides related advantages as well.

SUMMARY

[0004] Novel methods to circumvent the problem of immune response to Cas9 include utilizing orthologous Cas9 proteins for each treatment and/or engineering a Cas9 that does not elicit an immune response. Thus, provided herein are methods of avoiding an immune response in a subject being administered a regimen requiring Cas9 in order to optimize and broaden the application of CRIPSR based therapeutics comprising administering immune orthogonal Cas9. Also provided herein are methods to modify a Cas9 protein by swapping highly immunogenic peptides or amino acids with less immunogenic counterparts. These

methods are particularly useful to enable the application of Cas9 arsenal for repeat treatments. Further provided are Cas9 proteins modified to reduce immunogenicity.

[0005] Aspects of the disclosure relate to a method of generating a protein comprising: identifying one or more regions of a protein with affinity for a major histocompatibility complex (MHC), and modifying the one or more regions of the protien with affinity for the MHC through one or more amino acid substitutions, such that the modified region has no affinity for the MHC, wherein the resulting modified protein is immunosilent upon administration of the modified protein or a polynucleotide encoding the modified protein to a subject. In some embodiments, the affinity for the MHC is high affinity. In some embodiments, at least one substituted amino acid is an amino acid which does not serve as an MHC protein core residue. In some embodiments, the protein is selected from the group of a cytidine deaminase, an adenosine deaminase, a zinc finger nuclease, a transcriptional activator-like effector nuclease, a Cas9, or an AAV capsid protein. In some embodiments, the protein is Cas9, optionally SpCas9.

[0006] Further aspects relate to a modified Cas9 protein produced according to the method disclosed above. Still further aspects relate to a modified Cas9 protein comprising one or more, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, fifteen or more, or twenty or more of the amino acid modifications provided in **Table 1.** Some embodiments relate to an isolated polynucleotide encoding the modified Cas9. Further embodiments, relate to a vector comprising the isolated polynucleotide, optionally an AAV vector, and still further optionally an AAV5 vector. Additional embodiments relate to an AAV capsid comprising the vector. In some embodiments, one or more of the AAV capsid proteins has been modified to be immunosilent.

[0007] Aspects of the disclosure relate to a method of identifying immune orthogonal orthologs comprising: determining a set of affinities of a protein or regions thereof to a plurality of major histocompatibility complexes (MHCs), comparing the set of affinities of the protein or regions thereof to sets of affinities of orthologs of the protein to the plurality of MHCs, and determining a set of immune orthogonal orthologs based on non-overlapping sets of affinites. In some embodiments, the affinity for the MHC is high affinity. In some

embodiments, the protein is selected from the group of a cytidine deaminase, an adenosine deaminase, a zinc finger nuclease, a transcriptional activator-like effector nuclease, a **Cas9,** or an AAV capsid protein. In some embodiments, the protein is **Cas9,** optionally **SpCas9** or **SaCas9.** In some embodiments, the **Cas9** proteins the orthologs are selected from *S. pyogenes* **Cas9 (spCas9),** *S. aureus* **Cas9 (saCas9),** *B. longum* **Cas9,** *A. muciniphilia* **Cas9,** or *O. laneus* **Cas9.**

[0008] Some aspects relate to a method of avoiding immune response in a subject being administered a regimen requiring a protein, the method comprising: administering to the subject, in sequence, two or more proteins that are immune orthogonal. In some embodiments, the proteins that are immune orthogonal do not share an amino acid sequence of greater than 5 consecutive amino acids. In some embodiments, the proteins that are immune orthogonal do not share affinity for a major histocompatibility complex (MHC). In some embodiments, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more proteins that are immune orthogonal are administered in sequence.

[0009] Non-limiting exemplary aspects relate to a method of avoiding immune response in a subject being administered a regimen requiring **Cas9** and/or gene editing or gene regulation in a subject and/or treating a subject in need of gene editing or gene regulation, the method comprising: administering to the subject, in sequence, two or more **Cas9** proteins that are immune orthogonal. In some embodiments, the **Cas9** proteins that are immune orthogonal do not share an amino acid sequence of greater than 5 consecutive amino acids. In some embodiments, the **Cas9** proteins that are immune orthogonal do not share affinity for a major histocompatibility complex (MHC). In some embodiments, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more **Cas9** proteins that are immune orthogonal are administered in sequence. In some embodiments, each **Cas9** protein that is immune orthogonal is a **Cas9** derived from a distinct species of bacteria. In some embodiments, the **Cas9** proteins that are immune orthogonal are selected from *S. pyogenes* **Cas9 (spCas9),** *S. aureus* **Cas9 (saCas9),** *B. longum* **Cas9,** *A. muciniphilia* **Cas9,** or *O. laneus* **Cas9.** In some embodiments, the **Cas9** proteins that are immune orthogonal comprise **spCas9** and **saCas9.** In some embodiments, at least one of the two or

more Cas9 proteins is modified to reduce immunogenicity upon administration to the subject. In some embodiments, at least one of the two or more Cas9 proteins is modified according the method disclosed above. In some embodiments, at least one of the two or more Cas9 proteins or polynucleotides encoding said Cas9 proteins is comprised in an AAV vector. In some embodiments, the AAV vector is an AAV5 vector. In some embodiments, the AAV vector is comprised in an AAV capsid. In some embodiments, two or more Cas9 proteins or polynucleotides encoding said Cas9 proteins are comprised in AAV vectors. In some embodiments, each AAV vector is comprised in an AAV capsid, optionally wherein the AAV capsids are immune orthogonal to one another. In some embodiments, the method further comprises administering one or more guide RNAs to the subject. In some embodiments, the guide RNA is selected to treat a disease, disorder, or condition selected from the group of achromatopsia, adenosine deaminase (ADA) deficiency, alpha-1-antitrypsin deficiency, Alzheimer's disease, amyotrophic lateral sclerosis, aromatic amino acid decarboxylase deficiency, Batten disease, choroideremia, Crigler Najjar syndrome, cystic fibrosis, fragile X syndrome, hemophilia, hepatitis B, hepatitis C,homozygous familial hypercholesteremia, Huntington's Disease, Leber congenital amaurosis, macular degeneration, maple syrup urine disease (MSUD), mucopolysarccharidosis (I-IX), multiple sclerosis, muscular dystrophy, myotonic dystrophy, neurofibramotosis type 1, ornithine transcarbamylase deficiency, pachyonychia congenita, Parkinson's disease, phenylketonuria, polycystic kidney disease, Pompe disease, retinal degeneration, Rett's syndrome, rickets, spinal muscular atrophy, severe combined immunodeficiency, sickle cell disease, Smith-Lemli-Opitz syndrome, Y-linked nonobstructive spermatogenic failure, thalassemia, Tay-Sachs disease, Wilson's disease, cardiovascular disease, metabolic syndrome, pain management, and X-linked retinoschisis.

BRIEF DESCRIPTION OF DRAWINGS

[0010] **FIG. 1:** is a flow diagram depicting the process described in Example 1.

[0011] **FIG. 2:** shows **(A)** sets of immune-orthogonal proteins, located with a recursive clique-finding algorithm (Bold outlines indicate top 4 sets of orthogonal proteins. Color indicates number of 5-mer overlaps between protein pairs. This method is guaranteed to find all maximal sets of orthogonal proteins. *Streptococcus pyogenes* belongs to a set of 5

mutually orthogonal proteins.) **(B)** the number of maximal cliques containing each protein, broken down by size (Cliques of size 4 are the most frequent.).

[0012] FIG. 3: shows **(A)** change in affinity resulting from swaps in each peptide position (Data are shown averaged over 98 high-affinity peptides found in Streptococcus pyogenes.) **(B)** after swapping, distribution of peptides in each affinity category, by swap position (Swapping out amino acids at the beginning of the high affinity peptide have the biggest effect.) **(C)** cumulative sum showing number of peptides with at least one no-affinity swap option (blue), or at least one no-affinity or low-affinity swap option (green) (There are 98 high affinity peptides in this protein (black dotted line).).

[0013] FIG. 4: shows a clique consists of strains of Cas9 with no high affinity peptides overlapping, accordingly providing five sets of five Cas9 proteins with no high affinity peptides overlapping.

[0014] FIGS. 5A - 5H: shows that protein Protein based therapeutics elicit an adaptive immune response: experimental and *in silico* analyses: **(FIG. 5A)** Proteins have substantial therapeutic potential, but a major drawback is the immune response to both the therapeutic protein and its delivery vehicle. **(FIG. 5B)** As a case study, we explored the CRISPR-Cas9 systems and corresponding delivery vehicles based on AAVs. **(FIG. 5C)** Mice were injected retro-orbital ly with 10^{12} vg/mouse of AAV8-SaCas9 targeting the PCSK9 gene or a non-targeting control (empty vector). A decrease in PCSK9 serum levels, due to successful gene targeting, can be seen in mice receiving AAV-SaCas9-PCSK9 virus (n=6 mice for each group). **(FIG. 5D)** Immune response to the payload was detected in ELISAs for the SaCas9 protein. (n=12) **(FIG. 5E)** Immune response to the delivery vehicle was detected in ELISAs for the AAV8 virus capsid (n=12 mice). **(FIG. 5F)** *In silico* workflow used to find immune orthogonal protein homolog cliques. **(FIG. 5G)** Immunologically uninformed sequence comparison was carried out by checking all &-mers in a protein for their presence in another protein sequence with either zero or one mismatch. The x-axis corresponds to *k*, while MHC I and MHC II show overlap only of peptides predicted to bind to MHC class I and class 11 molecules. 48% of Cas9 pairs show no 6-mer overlap, and 83% of pairs show no overlapping MHC-binding peptides. **(FIG. 5H)** Same as (g) but for AAV VPl capsid proteins. All AAV pairs contain overlapping MHC-binding peptides.

[0015] **FIGS. 6A - 6E:** shows experimental validation of Cas9 and AAV immunogenicity predictions. **(FIG. 6A)** Mice were exposed to antigens via retro-orbital injections at 10^{12} vg/mouse. Serum was harvested prior to injection on day 0, and at multiple points over the course of 4-6 weeks. **(FIG. 6B)** anti-SpCas9 antibodies generated in mice injected with SpCas9 (n=6) and SaCas9 (n=12), and anti-SaCas9 antibodies generated in mice injected with SpCas9 (n=6) and SaCas9 (n=12). **(FIG. 6C)** anti-SpCas9 and anti-SaCas9 antibodies generated by mice injected with AAV8 SpCas9 (n=12; left panel), or AAVDJ SpCas9 (n=12; right panel). **(FIG. 6D)** anti-AAV8/DJ/2/5 antibodies generated against mice injected with AAV8 or AAVDJ (n=4 for all panels). **(FIG. 6E)** anti-AAV8/DJ/2/5 antibodies generated against mice injected with AAV2 or AAV5 (n=5 for all panels).

[0016] **FIG. 7:** depicts Cas9 immune orthogonal cliques. Cliques corresponding to 6-mer overlaps are depicted. An example of an orthogonal clique is highlighted, which includes Cas9s from: *S. pyogenes, S. aureus, B. longum, A. muciniphila,* and *O. laneus.*

[0017] **FIGS. 8A - 8D:** show the results of in silico analyses and comparisons of immunogenicity of Cas9 and AAV orthologs. Linear regressions exclude pairs with no overlap. **(FIG. 8A)** Cas9 MHC class I peptide overlap vs. phylogenetic distance. **(FIG. 8B)** AAV MHC class I peptide overlap vs. phylogenetic distance. **(FIG. 8C)** Cas9 MHC class II peptide overlap vs. phylogenetic distance. **(FIG. 8D)** AAV MHC class II peptide overlap vs. phylogenetic distance.

[0018] **FIGS. 9A - 9B:** shows the major AAV serotype groups. **(FIG. 9A)** AAV immune orthogonal cliques over 81 HLA alleles. AAV5 is the most immune-divergent in comparison to the other serotypes. No orthogonal cliques exist. **(FIG. 9B)** AAV phylogeny showing major serotype groupings as well as the position of the reconstructed sequence Anc80L65.

[0019] **FIG. 10:** shows experimental validation of a MHCII peptide predictions via the ELISPOT assay.

[0020] **FIG. 11:** shows immune orthogonal cliques of extremophile Cas9s and peptide overlap with pools of Cas9s from commensal, pathogenic, and environmental species.

DETAILED DESCRIPTION

[0021] Embodiments according to the present disclosure will be described more fully hereinafter. Aspects of the disclosure may, however, be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art. The terminology used in the description herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

[0022] Unless otherwise defined, all terms (including technical and scientific terms) used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. It will be further understood that terms, such as those defined in commonly used dictionaries, should be interpreted as having a meaning that is consistent with their meaning in the context of the present application and relevant art and should not be interpreted in an idealized or overly formal sense unless expressly so defined herein. While not explicitly defined below, such terms should be interpreted according to their common meaning.

[0023] The terminology used in the description herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety.

[0024] The practice of the present technology will employ, unless otherwise indicated, conventional techniques of tissue culture, immunology, molecular biology, microbiology, cell biology, and recombinant DNA, which are within the skill of the art.

[0025] Unless the context indicates otherwise, it is specifically intended that the various features of the invention described herein can be used in any combination. Moreover, the disclosure also contemplates that in some embodiments, any feature or combination of features set forth herein can be excluded or omitted. To illustrate, if the specification states that a complex comprises components A, B and C, it is specifically intended that any of A, B or C, or a combination thereof, can be omitted and disclaimed singularly or in any combination.

[0026] Unless explicitly indicated otherwise, all specified embodiments, features, and terms intend to include both the recited embodiment, feature, or term and biological equivalents thereof.

[0027] All numerical designations, e.g., pH, temperature, time, concentration, and molecular weight, including ranges, are approximations which are varied (+) or (-) by increments of 1.0 or 0.1, as appropriate, or alternatively by a variation of +/- 15 %, or alternatively 10%, or alternatively 5%, or alternatively 2%. It is to be understood, although not always explicitly stated, that all numerical designations are preceded by the term "about". It also is to be understood, although not always explicitly stated, that the reagents described herein are merely exemplary and that equivalents of such are known in the art.

[0028] ***Definitions***

[0029] As used in the description of the invention and the appended claims, the singular forms "a," "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise.

[0030] The term "about," as used herein when referring to a measurable value such as an amount or concentration and the like, is meant to encompass variations of 20%, 10%, 5%, 1 %, 0.5%, or even 0.1 % of the specified amount.

[0031] The terms or "acceptable," "effective," or "sufficient" when used to describe the selection of any components, ranges, dose forms, etc. disclosed herein intend that said component, range, dose form, etc. is suitable for the disclosed purpose.

[0032] The term "adeno-associated virus" or "AAV" as used herein refers to a member of the class of viruses associated with this name and belonging to the genus dependoparvovirus, family Parvoviridae. Multiple serotypes of this virus are known to be suitable for gene delivery; all known serotypes can infect cells from various tissue types. At least 11 or 12, sequentially numbered, are disclosed in the prior art. Non-limiting exemplary serotypes useful in the methods disclosed herein include any of the 11 or 12 serotypes, e.g., AAV2, AAV5, and AAV8, or variant serotypes, e.g. AAV-DJ. The AAV structural particle is composed of 60 protein molecules made up of VP1, VP2 and VP3. Each particle contains approximately 5 VP1 proteins, 5 VP2 proteins and 50 VP3 proteins ordered into an

icosahedral structure. Non-limiting exemplary VP! sequences useful in the methods disclosed herein are provided below.

[0033] AAT46339.1 AAV-11

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTS
FGGNLGRAVFQAKKRVLEPLGLVEEGAKTAPGKKRPLESPQEPDSSSGIGKKGKQPA
RKRLNFEEDTGAGDGPPEGSDTSAMSSDIEMRAAPGGNAVDAGQGSDGVGNASGD
WHCDSTWSEGKVTTTSTRTWVLPTYNNHLYLRLGTTSSSNTYNGFSTPWGYFDFNR
FHCHFSPRDWQRLINNNWGLRPKAlviRVKIFNIQVKEVTTSNGETTVANNLTSTVQIF
ADSSYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGIVTGENQNQTDRNAFYCLEY
FPSQMLRTGNNFEMAYNFEKVPFHSMYAHSQSLDPXIVnv[PLLDQYLWHLQSTTSGET
LNQGNAATTFGKIRSGDFAFYRKNWLPGPCVKQQRFSKTASQNYKIPASGGNALLK
YDTHYTLNNRWSNIAPGPPMATAGPSDGDFSNAQLIFPGPSVTGNTTTSANNLLFTSE
EEIAATNPRDTDMFGQIADNNQNATTAPITGNVTAMGVLPGMVWQNRDIYYQGPIW
AKIPHADGHFHPSPLIGGFGLKHPPPQIFIKNTPVPANPATTFTAARVDSFITQYSTGQ
VAVQIEWEIEKERSKRWNPEVQFTSNYGNQSSMLWAPDTTGKYTEPRVIGSRYLTN
HL

[0034] pdb|4IOV|A AAV-rh32

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTS
FGGNLGRAVFQAKKRVLEPLGLVEEGAKTAPGKKRPLESPQEPDSSSGIGKKGKQPA
KKRLNFEEDTGAGDGPPEGSDTSAMSSDIEMRAAPGGNAVDAGQGSDGVGNASGD
WHCDSTWSEGKVTTTSTRTWVLPTYNNHLYLRLGTTSNSNTYNGFSTPWGYFDFNR
FHCHFSPPvDWQPvLINNNWGLRPKAIVniVKIFMQVKEVTTSNGETTVANNLTSTVQIF
ADSSYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGIVTGENQNQTDRNAFYCLEY
FPSQMLRTGNNFEMAYNFEKVPFHSMYAHSQSLDPXlviNPLLDQYLWHLQSTTSGET
LNQGNAATTFGKIRSGDFAFYRKNWLPGPCVKQQRFSKTASQNYKIPASGGNALLK
YDTHYTLNNRWSNIAPGPPMATAGPSDGDFSNAQLIFPGPSVTGNTTTSANNLLFTSE
EEIAATWRDTDMFGQIAD>mQNATTAPITGNVTAMGVLPGMVWQNRDIYYQGPIW

AKIPHADGHFHPSPLIGGFGLKHPPPQIFIKNTPVPANPATTFTAARVDSFITQYSTGQ
VAVQIEWEIEKERSKR\\WEVQFTSNYGNQSSMLWAPDTTGKYTEPRVIGSRYLTN
HL

[0035] ABI16639.1 AAV-12

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNGRGLVLPGYKYLG
PFNGLDKGEPVNEADAAALEHDKAYDKQLEQGDNPYLKYNHADAEFQQRLATDTS
FGGNLGRAVFQAKKRILEPLGLVEEGVKTAPGKKRPLEKTPNRPTNPDSGKAPAKKK
QKDGEPADSARRTLDFEDSGAGDGPPEGSSSGEMSHDAEMRAAPGGNAVEAGQGA
DGVGNASGDWHCDSTWSEGRVTTTSTRTWVLPTYNNHLYLRIGTTANSNTYNGFST
PWGYFDFNP^HCFIFSPRDWQRLINNNWGLRPKSMRVKIFNIQVKEVTTSNGETTVA
NNLTSTVQIFADSTYELPYVMDAGQEGSFPPFPNDVFMVPQYGYCGVVTGKNQNQT
DRNAFYCLEYFPSQMLRTG^FEVSYQFEKVPFHSMYAHSQSLDRMMNPLLDQYL
WHLQSTTTGNSLNQGTATTTYGKITTGDFAYYRKNWLPGACIKQQKFSKNANQNY
KIPASGGDALLKYDTHTTLNGRWSNMAPGPPMATAGAGDSDFSNSQLIFAGPNPSG
NTTTSSNT^LFTSEEEIATTTSTPPvDTDMFGQIADNNQNATTAPfflANLDAMGIVPGMV
WQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFGLKHPPPQIFIKNTPVPANPNTTFSA
ARINSFLTQYSTGQVAVQIDWEIQKEHSKJIWNPEVQFTSNYGTQNSMLWAPDNAGN
YHELRAIGSRFLTHHL

[0036] NP_044927.1 AAV-4

MTDGYLPDWLEDNLSEGVREWWALQPGAPKPKANQQHQDNARGLVLPGYKYLGP
GNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQQRLQGDTS
FGGNLGRAVFQAKKRVLEPLGLVEQAGETAPGKKRPLIESPQQPDSSTGIGKKGKQP
AKKKLVFEDETGAGDGPPEGSTSGAMSDDSEMRAAAGGAAVEGGQGADGVGNAS
GDWHCDSTWSEGHVTTTSTRTWVLPTYNNHLYKRLGESLQSNTYNGFSTPWGYFD
FNRFHCHFSPRDWQRLINNNWGMRPKAMRVKIFNIQVKEVTTSNGETTVANNLTST
VQTFADSSYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGLVTGNTSQQQTDRNAF
YCLEYFPSQMLRTGNNFEITYSFEKVPFHSMYAHSQSLDRLMNPLIDQYLWGLQSTT
TGTTLNAGTATTNFTKLRPTNFSNFKKNWLPGPSIKQQGFSKTANQNYKIPATGSDSL
IKYETHSTLDGRWSALTPGPPMATAGPADSKFSNSQLIFAGPKQNGNTATVPGTLIFT

SEEELAATNATDTDMWGNLPGGDQSNSNLPTVDRLTALGAVPGMVWQNPxDIYYQG
PIWAKIPHTDGHFHPSPLIGGFGLKHPPPQIFIKNTPVPANPATTFSSTPVNSFITQYSTG
QVSVQIDWEIQKERSKRWNPEVQFTSNYGQQNSLLWAPDAAGKYTEPRAIGTRYLT
HHL

[0037] YP_077178.1 AAV-7

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDNGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTS
FGGNLGRAVFQAKKRVLEPLGLVEEGAKTAPAKKRPVEPSPQRSPDSSTGIGKKGQQ
PARKRLNFGQTGDSESVPDPQPLGEPPAAPSSVGSGTVAAGGGAPMADNNEGADGV
GNASGNWHCDSTWLGDRVITTSTRTWALPTYNNHLYKQISSETAGSTNDNTYFGYS
TPWGYFDFNRFHCHPSPRDWQRLINNNWGFRPKKLRFKLFNIQVKEVTTNDGVTTIA
NNLTSTIQVFSDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQSVGRSS
FYCLEYFPSQMLRTGNNFEFSYSFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLART
QSNPGGTAGNRELQFYQGGPSTMAEQAKNWLPGPCFRQQRVSKTLDQNNTNSNFAW
TGATKYHLNGRNSLWPGVAMATHKI)DEDRFFPSSGVLIFGKTGATNKTTLENVLM
T^EEIRPTNPVATEEYGIVSS^QAANTAAQTQVVNNQGALPGMVWQNRDVYLQ
GPIWAKIPHTDGNFHPSPLMGGFGLKHPPPQILIKNTPVPANPPEVFTPAKFASFITQYS
TGQVSVEIEWELQKENSKRWNPEIQYTSNFEKQTGVDFAVDSQGVYSEPRPIGTRYL
TRNL

[0038] YP_077180.1 AAV-8

MAADGYLPDWLEDNLSEGIREWWALKPGAPKPKANQQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLQAGDNPYLRYNHADAEFQERLQEDTS
FGGNLGRAVFQAKKRVLEPLGLVEEGAKTAPGKKRPVEPSPQRSPDSSTGIGKKGQQ
PARKRLNFGQTGDSESVPDPQPLGEPPAAPSGVGPNTMAAGGGAPMADNNEGADG
VGSSSGNWHCDSTWLGDRVITTSTRTWALPTYNNHLYKQISNGTSGGATNDNTYFG
YSTPWGYFDFNRFHCHFSPRDWQRLJNNNWGFRPKRLSFKLFNIQVKEVTQNEGTKT
IANNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGR
SSFYCLEYFPSQMLRTGNNFQFTYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLSR
TQTTGGTANTQTLGFSQGGPNTMANQAKNWLPGPCYRQQRVSTTTGQNNNSNFAW

TAGTKYHLNGRNSLANPGIAMATHKDDEERFFPSNGILIFGKQNAARDNADYSDVM
LTSEEEIKTTNPVATEEYGIVAD^QQQNTAPQIGTWSQGALPGlVrVWQNRDVYLQ
GPIWAKIPHTDGNFHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFNQSKLNSFITQY
STGQVSVEIEWELQKENSKRWNPE1QYTSNYYKSTSVDFAVNTEGVYSEPRPIGTRYL
TRNL

[0039] AAT46337.1 AAV-10

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTS
FGGNLGRAVFQAKKRVLEPLGLVEEAAKTAPGKKRPVEPSPQRSPDSSTGIGKKGQQ
PAKKRLNFGQTGESESVPDPQPIGEPPAGPSGLGSGTMAAGGGAPMADNNEGADGV
GSSSGNWHCDSTWLGDRVITTSTRTWALPTYNNHLYKQISNGTSGGSTNDNTYFGY
STPWGYFDFNPJ:HCFiFSPRDWQRLiNNNWGFRPKRLSFKLFNIQVKEVTQNEGTKTI
ANNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGR
SSFYCLEYFPSQMLRTGNNFEFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLSR
TQSTGGTQGTQQLLFSQAGPANMSAQAKNWLPGPCYRQQRVSTTLSQNNNSNFAW
TGATKYHLNGRDSLVWGVAMATHKDDEERFFPSSGVLMFGKQGAGRDNVDYSSV
MLTSEEEIKTTNPVATEQYGVVAD^QQANTGPIVGNWSQGALPGMYWQNRDVY
LQGPIWAKIPHTDGNFHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFSQAKLASFIT
QYSTGQVSVEIEWELQKENSKRWWEIQYTSNYYKSTNVDFAVNTEGTYSEPRPIGT
RYLTRNL

[0040] AAS99264.1 AAV-9

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLG
PGNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQERLKEDTS
FGGNLGRAVFQAKKRLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQP
AKKRLNFGQTGDTESVPDPQPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVG
SSSGNWHCDSQWLGDRVITTSTRTWALPTYNNHLYKQISNSTSGGSSNDNAYFGYS
TPWGYFDFNRFHCHFSPPJ)WQRLIM^WGFRPKRLNFKLFNIQVKEVTDNNGVKTI
ANNLTSTVQVFTDSDYQLPYVLGSAHEGCLPPFPADVFMIPQYGYLTLNDGSQAVG
RSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLDRLMNPLIDQYLYYLS

KTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQNNNSEFAWP
GASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADKVMI
TNEEEIKTTNPVATESYGQVATNHQSAQAQAQTGWVQNQGILPGMVWQDRDVYLQ
GPIWAKIPHTDGOTHPSPLMGGFGMKHPPPQILIKNTPVPADPPTAFNKDKLNSFITQ
YSTGQVSVEIEWELQKENSKRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRP1GTR
YLTRNL

[0041] NP_049542.1 AAV-1

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTS
FGGNLGRAVFQAKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQQP
AKKRLNFGQTGDSESVPDPQPLGEPPATPAAVGPTTMASGGGAPMADNNEGADGV
GNASG>TWHCDSTWLGDRVITTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYS
TPWGYFDFNRFHCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVKEVTTNDGVTTIA
NNLTSTVQVFSDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRS
SFYCLEYFPSQMLRTGNNFTFSYTFEEVPFHSSYAHSQSLDRLMNPLIDQYLYYLNRT
QNQSGSAQNKDLLFSRGSPAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWT
GASKYNLNGRESIINPGTAMASHKDDEDKFFPMSGVMIFGKESAGASNTALDNYMIT
DEEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMVWQDRDVYLQG
PIWAKIPHTDGHFHPSPLMGGFGLKNPPPQILIKNTPVPANPPAEFSATKFASFITQYST
GQVSVEIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYL
TRPL

[0042] AAB95450.1 AAV-6

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTS
FGGNLGRAVFQAKKRVLEPFGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQQP
AKKRLNFGQTGDSESVPDPQPLGEPPATPAAVGPTTMASGGGAPMADNNEGADGV
GNASGNWHCDSTWLGDRVITTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYS
TPWGYFDFNRFHCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVKEVTTNDGVTTIA
NNLTSTVQVFSDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRS

SFYCLEWPSQMLRTGNKFTFSYTFEDVPFHSSYAHSQSLDPvLMNPLIDQYLYYLNRT
QNQSGSAQNKDLLFSRGSPAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWT
GASKYNLNGRESIINPGTAMASHKDDKDKFFPMSGVMIFGKESAGASNTALDNVMI
TDEEEIKATNPVATERFGTVAVNLQSSSTDPATGDVHVMGALPGMVWQDRDVYLQ
GPIWAKIPHTDGHFHPSPLMGGFGLKHPPPQILIKNTPVPANPPAEFSATKFASFITQYS
TGQVSVEIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRY
LTRPL

[0043] NP_043941.1 AAV-3

MAADGYLPDWLEDNLSEGIREWWALKPGVPQPKANQQHQDNRRGLVLPGYKYLG
PGNGLDKGEPVNEADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQERLQEDTS
FGGNLGRAVFQAKKRILEPLGLVEEAAKTAPGKKGAVDQSPQEPDSSSGVGKSGKQ
PARKRLNFGQTGDSESVPDPQPLGEPPAAPTSLGSNTMASGGGAPMADNNEGADGV
GNSSGNWHCDSQWLGDRVITTSTRTWALPTYNNHLYKQISSQSGASNDNHYFGYST
PWGWDFNRFHCHFSPRDWQRLINWJWGFRPKKLSFKLFMQVRGVTQNDGTTTIAN
NLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPQYGYLTLNNGSQAVGRS
SFYCLEWPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLNR
TQGTTSGTTOQSRLLFSQAGPQSMSLQARNWLPGPCYRQQRLSKTA>TONNNSNFPW
TAASKYHLNGRDSLVNPGPAMASHKDDEEKFFPMHGNLIFGKEGTTASNAELDNV
MITDEEEIRTTOPVATEQYGTVANNLQSSNTAPTTGTVNHQGALPGMVWQDRDVYL
QGPIWAKIPHTDGHFHPSPLMGGFGLKHPPPQIMIKNTPVPANPPTTFSPAKFASFITQ
YSTGQVSVEIEWELQKENSKRWNPEIQYTSNYNKSVNVDFTVDTNGVYSEPRPIGTR
YLTRNL

[0044] ABZ10812.1 AAV-13

MTDGYLPDWLEDNLSEGVREWWALQPGAPKPKANQQHQDNARGLVLPGYKYLGP
GNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQERLQEDTSF
GGNLGRAVFQAKKRILEPLGLVEEAAKTAPGKKRPVEQSPAEPDSSSGIGKSGQQPA
RKRLNFGQTGDTESVPDPQPLGQPPAAPSGVGSTTMASGGGAPMADNNEGADGVG
NSSGNWHCDSQWLGDRVITTSTRTWALPTYNNHLYKQISSQSGATNDNHYFGYSTP
WGYFDFNRFHCHFSPRDWQRLINNNWGFRPKRLNFKLFMQVKEVTQNDGTTTIAN

NLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPQYGYLTLNNGSQAVGRS
SFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLNR
TQTASGTQQSRLLFSQAGPTSMSLQAKNWLPGPCYRQQRLSKQANDNNNSNFPWTG
ATKYHLNGRDSLVNPGPAMASUKDDKEKFFPMHGTLIFGKEGTNANNADLENVMIT
DEEEIRTTWVATEQYGTVSNNLQNSNAGPTTGTVNHQGALPGMVWQDRDVYLQG
PIWAKJPHTDGFIFHPSPLMGGFGLKHPPPQIMIKNTPVPANPPTNFSAAKFASFITQYS
TGQVSVEIEWELQKENSKRWNPEIQYTSNYNKSVNVDFTVDTNGVYSEPRPIGTRYL
TRNL

[0045] YP_680426.1 AAV-2

MAADGYLPDWLEDTLSEGIRQWWKLKPGPPPPKPAERHKDDSRGLVLPGYKYLGPF
NGLDKGEPVNEADAAALEHDKAYDRQLDSGDNPYLKYNHADAEFQERLKEDTSFG
GNLGRAVFQAKKRVLEPLGL VEEPVKTAPGKKRPVEHSPVEPD SSSGTGKAGQQPA
RKRLNFGQTGDADSVPDPQPLGQPPAAPSGLGTNTMATGSGAPMADNNEGADGVG
NSSGNWHCDSTWMGDRVITTSTRTWALPTYNNHLYKQISSQSGASNDNHYFGYSTP
WGYFDFNRFHCHFSPRDWQRLINNNWGFRPKRLWKLFMQVKEVTQNDGTTTI^ N
NLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPQYGYLTLNNGSQAVGRS
SFYCLEYFPSQMLRTGNNFTFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLSRT
NTPSGTTTQSRLQFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNNNSEYSWTG
ATKYHLNGRDSLVNPGPAMASHKDDEEKFFPQSGVLIFGKQGSEKTNVDIEKVMITD
EEEIRTTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMVWQDRDVYLQGPI
WAKJPHTDGHFHPSPLMGGFGLKHPPPQILIKNTPVPANPSTTFSAAKFASFITQYSTG
QVSVEIEWELQKENSKRWNPEIQYTSNYNKSVNVDFTVDTNGVYSEPRPIGTRYLTR
NL

[0046] YP_068409.1 AAV-5

MSFVDUPPDWLEEVGEGLREFLGLEAGPPKPKPNQQHQDQARGLVLPGYNYLGPGN
GLDRGEPVNRADEVAREHDISYNEQLEAGDNPYLKYNHADAEFQEKLADDTSFGGN
LGKAVFQAKKRVLEPFGLVEEGAKTAPTGKRIDDHFPKRKKARTEEDSKPSTSSDAE
AGPSGSQQLQIPAQPASSLGADTMSAGGGGPLGDNNQGADGVGNASGDWHCDSTW
MGDRVVTKSTRTWVLPSYNNHQYREIKSGSVDGSNANAYFGYSTPWGYFDFNRFH

SHWSPPJ)WQPvLINNYWGFPvPRSLRVKIFMQVKEVTVQDSTTTIANlSLTSTVQVFTD
DDYQLPYVVGNGTEGCLPAFPPQVFTLPQYGYATLNRDNTENPTERSSFFCLEYFPS
KMLRTGN?^EFTYNFEEVPFHSSFAPSQNLFKLAW LVDQYLYRFVSTNNTGGVQFN
K^AGRYANTYKNWFPGPMGRTQGWNLGSGVNRASVSAFATTNRMELEGASYQV
PPQPNGMTW^QGSNTYALENTMIFNSQPANPGTTATYLEGNMLITSESETQPVNRV
AYNVGGQMATNNQSSTTAPATGTYNLQEIVPGSVWMERDVYLQGPIWAKIPETGAH
FHPSPAMGGFGLKHPPPMMLIKNTPVPGNITSFSDVPVSSFITQYSTGQVTVEMEWEL
KKENSKRWNPEIQYTONYNDPQFVDFAPDSTGEYRTTRPIGTRYLTRPL

[0047] 3J1Q_A AAV-DJ

MAADGYLPDWLEDTLSEGIRQWWKLKPGPPPPKPAERHKDDSRGLVLPGYKYLGPF
NGLDKGEPVNEADAAALEHDKAYDRQLDSGDNPYLKYNHADAEFQERLKEDTSFG
GNLGRAVFQAKXRLLEPLGLVEEAAKTAPGKKRPVEHSPVEPDSSSGTGKAGQQPA
RKRLWGQTGDADSVPDPQPIGEPPAAPSGVGSLTMAAGGGAPMADNNEGADGVG
NSSGNWHCDSTWMGDRVITTSTRTWALPTYTsnSTHLYKQISNSTSGGSSNDNAYFGYS
TPWGYFDFNP^HCHFSPRDWQRLINNNWGFRPKRLSFKLFNIQVKEVTQNEGTKTIA
NNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRS
SFYCLEYFPSQMLRTGNNFQFTYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLSRT
QTTGGTTNTQTLGFSQGGPNTMANQAKNWLPGPCYRQQRVSKTSADNNNSEYSWT
GATKYHLNGRDSLVNPGPAMASHKDDEEKFFPQSGVLIFGKQGSEKTNVDIEKVMIT
DEEEIRTTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMVWQDRDVYLQG
PIWAKIPHTDGHFHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFNQSKLNSFITQYST
GQVSVEIEWELQKENSKRWNPEIQYTSNYYKSTSVDFAVNTEGVYSEPRPIGTRYLT
RNL

[0048] AKU89595.1 Anc80

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTS
FGGNLGRAVFQAKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGKKGQQP
ARKRLNFGQTGDSESVPDPQPLGEPPAAPSGVGSNTMAAGGGAPMADNNEGADGV
GNASGNWHCDSTWLGDRVITTSTRTWALPTYNNHLYKQISSQSGGSTNDNTYFGYS

TPWGYFDFNRFHCHFSPPJ)WQRLINlWWGFPvPKKLNFKLFMQVKEVTTMDGTTTIA
NNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRS
SFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLSRT
QTTSGTAGNRTLQFSQAGPSSMANQAKNWLPGPCYRQQRVSKTTNQNNNSNFAWT
GATKYHLNGRDSLVNPGPAMATHKDDEDKFFPMSGVLIFGKQGAGNSNVDLDNVM
ITOEEEIKTmPVATEEYGTVATNLQSANTAPATGTVNSQGALPGMVWQDRDVYLQ
GPIWAKIPHTDGHFHPSPLMGGFGLKHPPPQILIKNTPVPANPPTTFSPAKFASFITQYS
TGQVSVEIEWELQKENSKRWNPEIQYTSNYNKSTNVDFAVDTNGVYSEPRPIGTRYL
TRNL

[0049] Also as used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative ("or").

[0050] The term "aptamer" as used herein refers to single stranded DNA or RNA molecules that can bind to one or more selected targets with high affinity and specificity. Non-limiting exemplary targets include by are not limited to proteins or peptides.

[0051] The term "Cas9" refers to a CRISPR-associated, RNA-guided endonuclease such as streptococcus pyogenes Cas9 (spCas9) and orthologs and biological equivalents thereof. Biological equivalents of Cas9 include but are not limited to C2cl from ***Alicyclobacillus acideterrestris*** and Cpfl (which performs cutting functions analogous to Cas9) from various bacterial species including ***Acidaminococcus spp.*** and ***Francisella novicida*** Ul 12. Cas9 may refer to an endonuclease that causes double stranded breaks in DNA, a nickase variant such as a RuvC or HNH mutant that causes a single stranded break in DNA, as well as other variations such as deadCas-9 or dCas9, which lack endonuclease activity. Cas9 may also refer to "split-Cas9" in which CAs9 is split into two halves - C-Cas9 and N-Cas9 - and fused with a two intein moieties. ***See,*** e.g., U.S. Pat. No. 9,074,199 Bl; Zetsche et al. (2015) Nat Biotechnol. 33(2): 139-42; Wright et al. (2015) PNAS 112(10) 2984-89. Non-limiting examples of commercially available sources of SpCas9 comprising plasmids can be found under the following AddGene reference numbers:

42230: PX330; SpCas9 and single guide RNA

48138: PX458; SpCas9-2A-EGFP and single guide RNA

62988: PX459; SpCas9-2A-Puro and single guide RNA

48873: PX460; SpCas9n (DIOA nickase) and single guide RNA

48140: PX461; SpCas9n-2A-EGFP (DIOA nickase) and single guide RNA

62987: PX462; SpCas9n-2A-Puro (DIOA nickase) and single guide RNA

48137: PX165; SpCas9

[0052] Further examples of Cas9 are provided in the table below:

Name	Protein Sequence
S. pyogenes Cas9	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS ITGLYETRIDLSQLGGD*

Staphylococcus aureus Cas9	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGAR RLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAA LLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKD GEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGP GEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI TRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPE FTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINE MQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN NPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYET FKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIAN ADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIK DFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL KKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGV YKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKING ELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYST DILGNLYEVKSKKHPQIIKKG*
S. thermophilus CRISPR 1 Cas9	MSDLVLGLDIGIGSVGVGILNKVTGE1IHKNSRIFPAAQAENNLVRRTNRQGRRL ARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKN MVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQ LRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILT GKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYT AQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLS CDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLN TEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELI PELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKI VNAA1KEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLK AANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINTMSN QFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFREL KAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQE HFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNL WKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSI LFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAF MKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINDKGKEVPCNPFLKYKE EHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADV YFNKTTGKYEILGLKYADLQFDKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTL YKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVL GNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF*

N . meningitidis **Cas9**	MAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRVFERAEVPKTG DSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPN TPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKG VADNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILL FEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAA KNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARK LLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLS PELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIV PLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPP1PADEIRNPVVLRALSQARK VINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREY FPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFSRT WDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRS KKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASN GQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMN AFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEK LRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHiVIETVKSAKRLDEGVS VLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKY DKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYY LVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKA RMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEI RPCRLKKRPPVR*
Parvibaculum lavamentivorans **Cas9**	MERIFGFDIGTTSIGFSVIDYSSTQSAGNIQRLGVRIFPEARDPDGTPLNQQRRQK RMMRRQLRRRRIRRKALNETLHEAGFLPAYGSADWPVVMADEPYELRRRGLE EGLSAYEFGRAIYHLAQHRHFKGRELEESDTPDPDVDDEKEAANERAATLKAL KNEQTTLGAWLARRPPSDRKRGIHAHRNVVAEEFERLWEVQSKFHPALKSEEM RARISDTIFAQRPVFWRKNTLGECRFMPGEPLCPKGSWLSQQRRMLEKLNNLAI AGGNARPLDAEERDAILSKLQQQASMSWPGVRSALKALYKQRGEPGAEKSLK FNLELGGESKLLGNALEAKLADMFGPDWPAHPRKQEIRHAVHERLWAADYGE TPDKKRVIILSEKDRKAHREAAANSFVADFGITGEQAAQLQALKLPTGWEPYSI PALNLFLAELEKGERFGALVNGPDWEGWRRTNFPHRNQPTGEILDKLPSPASKE ERERISQLRNPTVVRTQNELRKVVNNLIGLYGKPDRIRIEVGRDVGKSKREREEI QSGIRRNEKQRKKATEDLIKNGIANPSRDDVEKWILWKEGQERCPYTGDQIGFN ALFREGRYEVEHIWPRSRSFDNSPRNKTLCRKDVNIEKGNRMPFEAFGHDEDR WSAIQIRLQGMVSAKGGTGMSPGKVKRFLAKTMPEDFAARQLNDTRYAAKQI LAQLKRLWPDMGPEAPVKVEAVTGQVTAQLRKLWTLNNILADDGEKTRADH RHHAIDALTVACTHPGMTNKLSRYWQLRDDPRAEKPALTPPWDTIRADAEKA VSEIVVSHRVRKKVSGPLHKETTYGDTGTDIKTKSGTYRQFVTRKKIESLSKGEL DEIRDPRIKEIVAAHVAGRGGDPKKAFPPYPCVSPGGPEIRKVRLTSKQQLNLM AQTGNGYADLGSNHHIAIYRLPDGKADFEIVSLFDASRRLAQRNPIVQRTRADG ASFVMSLAAGEAIMIPEGSKKGIWIVQGVWASGQVVLERDTDADHSTTTRPMP NPILKDDAKKVSIDPIGRVRPSND*

Corynebacter diphtheria Cas9	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVTRL ASSGIARRTRRLYRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVRAELA ASYIADEKERGEKLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFKAIREEI KRASGQPVPETATVGQMVTLCELGTLKLRGEGGVLSARLQQSDYAREIQEICR MQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKASDAFQRYR1 AALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIAEILGIDRGQL IGTATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKAL SNAEVDDFDSPEGAKVQAFFADLDDDVHAKLDSLHLPVGRAAYSEDTLVRLTR RMLSDGVDLYTARLQEFGIEPSWTPPTPRIGEPVGNPAVDRVLKTVSRWLESAT KTWGAPERVIIEHVREGFVTEKRAREMDGDMRRRAARNAKLFQEMQEKLNVQ GKPSRADLWRYQSVQRQNCQCAYCGSPITFSNSEMDHIVPRAGQGSTNTRENL VAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAVERTRHWVTDTGMRSTDFK KFTKAVVERPQRATMDEEIDARSMESVAWMANELRSRVAQHFASHGTTVRVY RGSLTAEARRASG1SGKLKFFDGVGKSRLDRRHHAIDAAVIAFTSDYVAETLAV RSNLKQSQAHRQEAPQWREFTGKDAEHRAAWRVWCQKMEKLSALLTEDLRD DRVVVMSNVRLRLGNGSAHKETIGKLSKVKLSSQLSVSDIDKASSEALWCALT REPGFDPKEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAELGSSF HHARVYKITSGKKPAFAMLRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKL RDALATGNAEYLGWLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVDGFF SPSKLRLRP1.QMSKEGIKKESAPELSKIIDRPGWLPAVNKLFSDGNVTVVRRDSL GRVRLESTAHLPVTWKVQ*
Streptococcus pasteurianus Cas9	MTNGKILGLDIGIASVGVGIIEAKTGKVVHANSRLFSAANAENNAERRGFRGSR RLNRRKKHRVKRVRDLFEKYGIVTDFRNLNLNPYELRVKGLTEQLKNEELFAA LRTISKRRGISYLDDAEDDSTGSTDYAKSIDENRRLLKNKTPGQIQLERLEKYGQ LRGNFTVYDENGEAHRLINVFSTSDYEKEARKILETQADYNKKITAEFIDDYVEI LTQKRKYYHGPGNEKSRTDYGRFRTDGTTLENIFGILIGKCNFYPDEYRASKAS YTAQEYNFLNDLNNLKVSTETGKLSTEQKESLVEFAKNTATLGPAKLLKEIAKI LDCKVDEIKGYREDDKGKPDLHTFEPYRKLKFNLESINIDDLSREVIDKLADILT LNTEREGIEDAIKRNLPNQFTEEQISEIIKVRKSQSTAFNKGWHSFSAKLMNELIP ELYATSDEQMTILTRLEKFKVNKKSSKNTKTIDEKEVTDEIYNPVVAKSVRQTIK IINAAVKKYGDFDKIVIEMPRDKNADDEKKFIDKRNKENKKEKDDALKRAAYL YNSSDKLPDEVFHGNKQLETKIRLWYQQGERCLYSGKPISIQELVHNSNNFEID HILPLSLSFDDSLANKVLVYAWTNQEKGQKTPYQVIDSMDAAWSFREMKDYV LKQKGLGKKKRDYLLTTENIDKIEVKKKF1ERNLVDTRYASRVVLNSLQSALRE LGKDTKVSVVRGQFTSQLRRKWKIDKSRETYHHHAVDALIIAASSQLKLWEKQ DNPMFVDYGKNQVVDKQTGEILSVSDDEYKELVFQPPYQGFVNTISSKGFEDEI LFSYQVDSKYNRKVSDATIYSTRKAKIGKDKKEETYVLGKIKDIYSQNGFDTFIK KYNKDKTQFLMYQKDSLTWENVIEVILRDYPTTKKSEDGKNDVKCNPFEEYRR ENGLICKYSKKGKGTPIKSLKYYDKKLGNCIDITPEESRNKVILQSINPWRADVY FNPETLKYELMGLKYSDLSFEKGTGNYH1SQEKYDAIKEKEGIGKKSEFKFTLY RNDLILIKDIASGEQEIYRFLSRTMPNVNHYVELKPYDKEKFDNVQELVEALGE ADKVGRCIKGLNKPNISIYKVRTDVLGNKYFVKKKGDKPKLDFKNNKK*

Neisseria cinerea Cas9	MAAFKPNPlvrNYILGLDIGIASVGWAlVEIDEEENPIRLIDLGVRVFERAEVPKTG DSLAAARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPN TPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKG VADNTHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFNRKDLQAELNL LFEKQKEFGNPHVSDGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPTEPKA AKNTYTAERFVWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQA RKLLDLDDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPL NLSPELQDEIGTAFSLFKTDEDITGRLKDRVQPE1LEALLKHISFDKFVQISLKAL RRJVPLMEQGNRYDEACTEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQ ARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKSAAKF REYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALP FSRTWDDSFNNKVLALGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSR FPRSKKQRILLQKFDEDGFKERNLNDTRYINRFLCQFVADHMLLTGKGKRRVF ASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTIAMQQKITRFVRYKE MNAFDGKTIDKETGEVLHQKAHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADT PEKLRTLLAEKLSSRPEAVHKYVTPLFISRAPNRKMSGQGHMETVKSAKRLDE GISVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFY KYDKAGNRTQQVKAVRVEQVQKTGVWVHNHNGIADNATIVRVDVFEKGGKY YLVPIYSWQVAKGILPDRAVVQGKDEEDWTVMDDSFEFKFVLYANDLIKLTAK KNEFLGYFVSLNRATGAIDIRTHDTDSTKGKNGIFQSVGVKTALSFQKYQIDEL GKEIRPCRLKKRPPVR*
Campylobacter lari Cas9	MRJLGFDIGINSIGWAFVENDELKDCGVRIFTKAENPKNKESLALPRRNARSSRR RLKRRKARLIAIKRILAKELKLNYKDYVAADGELPKAYEGSLASVYELRYKALT QNLETKDLARVILHIAKHRGYMNKNEKKSNDAKKGKILSALKNNALKLENYQS VGEYFYKEFFQKYKKNTKNFIKIRNTKDNYNNCVLSSDLEKELKLILEKQKEFG YNYSEDFINEILKVAFFQRPLKDFSHLVGACTFFEEEKRACKNSYSAWEFVALT KIINEIKSLEKISGEIVPTQTINEVLNLILDKGSITYKKFRSCINLHESISFKSLKYDK ENAENAKLIDFRKLVEFKKALGVHSLSRQELDQISTHITLIKDNVKLKTVLEKYN LSNEQINNLLEIEFNDYINLSFKALGMILPLMREGKRYDEACEIANLKPKTVDEK KDFLPAFCDSIFAHELSNPVVNRAISEYRKVLNALLKKYGKVHKIHLELARDVG LSKKAREKIEKEQKENQAVNAWALKECENIGLKASAKN1LKLKLWKEQKEICIY SGNKISIEHLKDEKALEVDHIYPYSRSFDDSFINKVLVFTKENQEKLNKTPFEAF GKNIEKWSKIQTLAQNLPYKKKNKILDENFKDKQQEDFISRNLNDTRYIATLIAK YTKEYLNFLLLSENENANLKSGEKGSKIHVQTISGMLTSVLRHTWGFDKKDRN NHLHHALDAIIVAYSTNSIIKAFSDFRKNQELLKARFYAKELTSDNYKHQVKFFE PFKSFREKILSKIDEIFVSKPPRKRARRALHKDTFHSENKIIDKCSYNSKEGLQIAL SCGRVRKIGTKYVENDTIVRVDIFKKQNKFYAIPIYAMDFALGILPNKIVITGKD KNNNPKQWQTIDESYEFCFSLYKNDLILLQKKNMQEPEFAYYNDFSISTSSICVE KHDNKFENLTSNQKLLFSNAKEGSVKVESLGIQNLKVFEKYIITPLGDKIKADFQ PREN1SLK·1SKKYGLR*
T. **denticola Cas9**	MKKEIKDYFLGLDVGTGSVGWAVTDTDYKLLKANRKDLWGMRCFETAETAE VRRLHRGARRRIERRKKRJKLLQELFSQEIAKTDEGFFQRMKESPFYAEDKTILQ ENTLFNDKDFADKTYHKAYPTINHLIKAWIENKVKPDPRLLYLACHN1IKKRGH FLFEGDFDSENQFDTSIQALFEYLREDMEVDIDADSQKVKEILKDSSLKNSEKQS RLNKILGLKPSDKQKKAITNLISGNKINFADLYDNPDLKDAEKNSISFSKDDFDA LSDDLASILGDSFELLLKAKAVYNCSVLSKVIGDEQYLSFAKVKIYEKHKTDLT KLKNVIKKHFPKDYKKVFGYNKNEKNNNNYSGYVGVCKTKSKKLIINNSVNQ EDFYKFLKTILSAKSEIKEVNDILTEIETGTFLPKOISKSNAEIPYQLRKMELEKIL SNAEKHFSFLKQKDEKGLSHSEKIIMLLTFKIPYY1GPINDNHKKFFPDRCWVVK KEKSPSGKTTPWNFFDHIDKEKTAEAFITSRTNFCTYLVGESVLPKSSLLYSEYT VLNEINNLQIIIDGKNICDIKLKQKIYEDLFKKYKKITQKQISTFIKHEGICNK·1DE VIILGIDKECTSSLKSYIELKNIFGKQVDEISTKNMLEEIIRWATIYDEGEGKTILK TKIKAEYGKYCSDEQIKKILNLKFSGWGRLSRKFLETVTSEMPGFSEPVNIITAM RETQNNLMELLSSEFTFTENIKKINSGFEDAEKQFSYDGLVKPLFLSPSVKKML

	WQTLKLVKEISHITQAPPKKIFIEMAKGAELEPARTKTRLKILQDLYNNCKNDA DAFSSEIKDLSGKIENEDNLRLRSDKLYLYYTQLGKCMYCGKPIEIGHVFDTSNY DIDHIYPQSKIKDDSISNRVLVCSSCNKNKEDKYPLKSEIQSKQRGFWNFLQRMN F1SLEKLNRLTRATPISDDETAKFIARQLVETRQATKVAAKVLEKMFPETKIVYS KAETVSMFRNKFDIVKCREINDFHHAHDAYLNIVVGNVYNTKFTNNPWNFIKE KRDNPKIADTYNYYKVFDYDVKRNNITAWEKGKTIITVKDMLKRNTPIYTRQA ACKKGELFNQTIMKKGLGQHPLKKEGPFSNISKYGGYNKVSAAYYTLIEYEEK GNKIRSLETIPLYLVKDIQKDQDVLKSYLTDLLGKKEFKILVPKIKINSLLKINGF PCHITGKTNDSFLLRPAVQFCCSNNEVLYFKKIIRFSEIRSQREKIGKTISPYEDLS FRSYIKENLWKKTKNDEIGEKEFYDLLQKKNLEIYDMLLTKHKDTIYKKRPNSA TIDILVKGKEKFKSLIIENQFEVILEILKLFSATRNVSDLQHIGGSKYSGVAKIGNK ISSLDNCILIYQSITGIFEKRIDLLKV*
S. mutans Cas9	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALL FDSGNTAEDRRLKRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFL VTEDKRGERHP1FGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAH IIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKI SKSAKKDRVLKLFPNEKSNGRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDT YEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVGTKAPLSASMIQRY NEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKDGYAGYIDGKTNQEAFYKYLK GLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFL ADNQDRIEKLLTFRIPYYVGPLARGKSDFAWLSRKSADKITPWNFDEIVDKESS AEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFD ANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASY GTYHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQ VKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDA LSFKEEIAKAQVIGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPE NIVVEN4ARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRL FLYYLQNGRDMYTGEELDIDYLSQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGK SDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKR QLVETRQITKHVARILDERFNTETDENNKKIRQVKIVTLKSNLVSNFRKEFELYK VREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATA KKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQVNIVKKVEEQ TGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKS KKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEF KELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAP ATFKFFDKNIDRKRYTSTTEILNATLIHQSITGLYETRIDLNKLGGD
S. thermophilus **CRISPR 3 Cas9**	MTKPYSIGLDIGTNSVGWAVTTDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLF DSGITAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVP DDKRDSKYPIFGNLVEEKAYHDEFPTIYHLRKYLADSTKKADLRLVYLALAHM IKYRGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKIS KLEKKDRILKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYD EDLETLLGYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYN EHKEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKKL LAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFYPFLA KNKERIEKILTFRJPYYVGPLARGNSDFAWSIRKRNEKITPWNFEDVIDKESSAE AFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSK QKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSLSTYHDLLN IINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLSRRHYT GWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQ IIGDEDKGNIKEVVKSLPGSPAIKKGILQSIKIVDELVKVMGGRKPESIVVEN4ARE NQYTNQGKSNSQQRLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLY YLQNGKDMYTGDDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSD DVPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLSPEDKAGFIQRQLV

	ETRQITKHVARLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVR EINDFHHAHDAYLNAVVASALLKKYPKLEPEFVYGDYPKYNSFRERKSATEKV YFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLATVRRVLSYPQ VNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNENLVGAKEYLDPKKYG GYAGISNSFTVLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLLEKGY KDIELIIELPKYSLFELSDGSRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYH AKRISNTINENHRKYVENHKKEFEELFYYILEFNENYVGAKKNGKLLNSAFQSW QNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKD ATLIHQSVTGLYETRIDLAKLGEG
C. jejuni **Cas9**	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSAR KRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRAL NELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVG EYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSF SKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIIN LLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKG TYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQ IDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFL PAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINIELAREVGKNH SQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYSGE KIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGN DSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVL NYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKD RNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRK FFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVL KALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNK AVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSST VSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVT KAEFRQREDFKK
P. multocida **Cas9**	MQTTNLSYILGLDLGIASVGWAVVEINENEDPIGLIDVGVRIFERAEVPKTGESL ALSRRLARSTRRLIRRRAHRLLLAKRFLKREGILSTIDLEKGLPNQAWELRVAGL ERRLSAIEWGAVLLHLIKHRGYLSKRKNESQTNNKELGALLSGVAQNHQLLQS DDYRTPAELALKKFAKEEGHIRNQRGAYTHTFNRLDLLAELNLLFAQQHQFGN PHCKEHIQQYMTELLMWQKPALSGEAILKMLGKCTHEKNEFKAAKHTYSAER FVWLTKLNNLRILEDGAERALNEEERQLLINHPYEKSKLTYAQVRKLLGLSEQA IFKHLRYSKENAESATFMELKAWHAIRKALENQGLKDTWQDLAKKPDLLDEIG TAFSLYKTDEDIQQYLTNKVPNSVINALLVSLNFDKFIELSLKSLRKILPLMEQG KRYDQACREIYGHHYGEANQKTSQLLPAIPAQEIRNPVVLRTLSQARKVINAIIR QYGSPARVHIETGRELGKSFKERREIQKQQEDNRTKRESAVQKFKELFSDFSSEP KSKDILKFRLYEQQHGKCLYSGKEINIHRLNEKGYVEIDHALPFSRTWDDSFNN KVLVLASENQNKGNQTPYEWLQGKINSERWKNFVALVLGSQCSAAKKQRLLT QVIDDNKFIDRNLNDTRYIARFLSNYIQENLLLVGKNKKNVFTPNGQITALLRSR WGLIKARENNMIHHALDAIVVACATPSMQQKITRFIRFKEVHPYKIENRYEMV DQESGEIISPHFPEPWAYFRQEVNIRVFDNHPDTVLKEMLPDRPQANHQFVQPL FVSRAPTRKMSGQGHMETIKSAKRLAEGISVLRIPLTQLKPNLLENMVNKEREP ALYAGLKARLAEFNQDPAKAFATPFYKQGGQQVKAIRVEQVQKSGVLVRENN GVADNASIVRTDVFIKNNKFFLVPIYTWQVAKGILPNKAIVAHKNEDEWEEMD EGAKFKFSLFPNDLVELKTKKEYFFGYYIGLDRATGNISLKEHDGEISKGKDGV YRVGVKLALSFEKYQVDELGKNRQICRPQQRQPVR
F. novicida **Cas9**	MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNGKVYELSKDSYTLLMNN RTARRHQRRGIDRKQLVKRLFKLIWTEQLNLEWDKDTQQAISFLFNRRGFSFIT DGYSPEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKISEIYNKLM QKILEFKLN4KLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYT DKQGNLKELSYYHHDKYNIQEFLKRHATINDRILDTLLTDDLDIWMFNFEKFDF

	DKNEEKLQNQEDKDHIQAHLHHFVFAVNKIKSEMASGGRHRSQYFQEITNVLD ENNHQEGYLKNFCENLHNKKYSNLSVKNLVNLIGNLSNLELKPLRKYFNDKIH AKADHWDEQKFTETYCHWILGEWRVGVKDQDKKDGAKYSYKDLCNELKQK VTKAGLVDFLLELDPCRTIPPYLDNNNRKPPKCQSLILNPKFLDNQYPNWQQYL QELKKLQSIQNYLDSFETDLKVLKSSKDQPYFVEYKSSNQQIASGQRDYKDLDA RILQFIFDRVKASDELLLNEIYFQAKKLKQKASSELEKLESSKKLDEVIANSQLSQ ILKSQHTNGIFEQGTFLHLVCKYYKQRQRARDSRLYIMPEYRYDKKLHKYNNT GRFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDLFISK WLVEHIRGFKKACEDSLKIQKDNRGLLNHKIN1ARNTKGKCEKEIFNLICKIEGS EDKKGNYKHGLAYELGVLLFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAER KGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKAQRLPAIPTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQLHIP11TESNAFEFEPALADVKGKS LKDRRKKALERISPENIFKDKNNRJKEFAKGISAYSGANLTDGDFDGAKEELDHI IPRSHKKYGTLNDEANLICVTRGDNKNKGNRIFCLRDLADNYKLKQFETTDDLE IEKKIADTIWDANKKDFKFGNYRSFINLTPQEQKAFRHALFLADENPIKQAVIRA INNRNRTFVNGTQRYFAEVLAMNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIA EIRQLYEKVDSDIQAYAKGDKPQASYSHLIDAMLAFCIAADEHRNDGSIGLEID KNYSLYPLDKNTGEVFTKDIFSQIKITDNEFSDKKLVRKKAIEGFNTHRQMTRD GIYAENYLPILIHKELNEVRKGYTWKNSEEIKIFKGKKYDIQQLNNLVYCLKFV DKPISIDIQISTLEELRNILTTNN1AATAEYYYINLKTQKLHEYYIENYNTALGYK KYSKEMEFLRSLAYRSERVKIKSIDDVKQVLDKDSNFIIGKITLPFKKEWQRLYR EWQNTTIKDDYEFLKSFFNVKSITKLHKKVRKDFSLPISTNEGKFLVKRKTWDN NFIYQILNDSDSRADGTKPFIPAFDISKNEIVEAIIDSFTSKNIFWLPKNIELQKVD NKNIFAIDTSKWFEVETPSDLRDIGIATIQYKIDNNSRPKVRVKLDYVIDDDSKIN YFMNHSLLKSRYPDKVLEILKQSTIIEFESSGFNKTIKEMLGMKLAGIYNETSNN
Lactobacillus buchneri Cas9	MKVNNYHIGLDIGTSSIGWVAIGKDGKPLRVKGKTAIGARLFQEGNPAADRRM FRTTRRRLSRRKWRLKLLEEIFDPYITPVDSTFFARLKQSNLSPKDSRKEFKGSM LFPDLTDMQYHKNYPTIYHLRHALMTQDKKFDIRMVYLAIHHIVKYRGNFLNS TPVDSFKASKVDFVDQFKKLNELYAAINPEESFKINLANSEDIGHQFLDPSIRKF DKKKQIPKIVPVMMNDKVTDRLNGKIASEIIHAILGYKAKLDVVLQCTPVDSKP WALKFDDEDIDAKLEKILPEMDENQQSIVAILQNLYSQVTLNQIVPNGMSLSES MIEKYNDHHDHLKLYKKLIDQLADPKKKAVLKKAYSQYVGDDGKVIEQAEFW SSVKKNLDDSELSKQIMDLIDAEKFMPKQRTSQNGVIPHQLHQRELDEIIEHQSK YYPWLVEINPNKHDLHLAKYKIEQLVAFRVPYYVGPMITPKDQAESAETVFSW MERKGTETGQITPWNFDEKVDRKASANRFIKRMTTKDTYLIGEDVLPDESLLYE KFKVLNELNMVRVNGKLLKVADKQAIFQDLFENYKHVSVKKLQNYIKAKTGL PSDPEISGLSDPEHFNNSLGTYNDFKKLFGSKVDEPDLQDDFEKIVEWSTVFEDK KILREKLNEITWLSDQQKDVLESSRYQGWGRLSKKLLTGIVNDQGERIIDKLWN TNKNFMQIQSDDDFAKR1HEANADQMQAVDVEDVLADAYTSPQNKKAIRQVV KVVDDIQKAMGGVAPKYISIEFTRSEDRNPRRTISRQRQLENTLKDTAKSLAKSI NPELLSELDNAAKSKKGLTDRLYLYFTQLGKDIYTGEPINIDELNKYDIDHILPQ AFIKDNSLDNRVLVLTAVNNGKSDNVPLRMFGAKMGHFWKQLAEAGLISKRK LKNLQTDPDTISKYAMHGFIRRQLVETSQVIKLVANILGDKYRNDDTKIIEITAR MNHQMRDEFGFIKNREINDYHHAFDAYLTAFLGRYLYHRYIKLRPYFVYGDFK KFREDKVTMRNFNFLHDLTDDTQEKIADAETGEVIWDRENSIQQLKDVYHYKF MLISHEVYTLRGAMFNQTVYPASDAGKRKLIPVKADRPVNVYGGYSGSADAY MAIVRIHNKKGDKYRVVGVPMRALDRLDAAKNVSDADFDRALKDVLAPQLT KTKKSRKTGEITQVIEDFEIVLGKVMYRQLMIDGDKKFMLGSSTYQYNAKQLV LSDQSVKTLASKGRLDPLQESNOYNNVYTEILDKVNQYFSLYDMNKFRHKLN LGFSKFISFPNHNVLDGNTKVSSGKREILQEILNGLHANPTFGNLKDVGITTPFG QLQQPNGILLSDETKIRYQSPTGLFERTVSLKDL
Listeria innocua	MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDSEKKQIKKNFWGVRL FDEGQTAADRRMARTARRRIERRRNRISYLQGIFAEEMSKTDANFFCRLSDSFY VDNEKRNSRHPFFATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHI

Cas9	IKYRGNFLIEGALDTQNTSVDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVA KILVEKVTRKEKLERILKLYPGEKSAGMFAQFISLIVGSKGNFQKPFDLIEKSDIE CAKDSYEEDLESLLALIGDEYAELFVAAKNAYSAVVLSSIITVAETETNAKLSAS MIERFDTHEEDLGELKAFIKLHLPKHYEEIFSNTEKHGYAGYIDGKTKQADFYK YMKMTLENIEGADYFIAKIEKENFLRKQRTFDNGAIPHQLHLEELEAILHQQAK YYPFLKENYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKADGEIRPWTMIEEKV DFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYLVYNELTKVRYINDQGK TSYFSGQEKEQIFNDLFKQKRKVKKKDLELFLRNMSHVESPTIEGLEDSFNSSYS TYHDLLKVGIKQEILDNPVNTEMLEN1VKILTVFEDKRMIKEQLQQFSDVLDGV VLKKLERRHYTGWGRLSAKLLMGIRDKQSHLTILDYLMNDDGLNRNLMQLIN DSNLSFKSIIEKEQVTTADKDIQSIVADLAGSPAIKKG1LQSLK1VDELVSVMGYP PQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQILKEHPTDNQELRNN RLYLYYLQNGKDMYTGQDLDIHNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGN REKGDDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKAR FIHRQLVETRQITKNVANILHQRFNYEKDDHGNTMKQVRIVTLKSALVSQFRKQ FQLYKVRDVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGDYHQFDWFK ANKATAKKQFYTO1MLFFAQKDRIIDENGEILWDKKYLDTVKXVMSYRQMNIV KKTEIQKGEFSKATIKPKGNSSKLIPRKTNWDPMKYGGLDSPNMAYAVVIEYA KGKNKLVFEKKIIRVT1MERKAFEKDEKAFLEEQGYRQPKVLAKLPKYTLYECE EGRRRMLASANEAQKGNQQVLPNHLVTLLHHAANCEVSDGKSLDYIESNREM FAELLAHVSEFAKRYTLAEANLNKINQLFEQNKEGDIKAIAQSFVDLMAFNAM GAPASFKFFETTIERKRYNNLKELLNSTIIYQSITGLYESRKRLDD
L. pneumophilia Cas9	MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQA QRRATRHRVRNKKRNQFVKRVALQLFQHILSRDLNAKEETALCHYLNNRGYT YVDTDLDEYIKDETTINLLKELLPSESEHNFIDWFLQKMQSSEFRKILVSKVEEK KDDKELKNAVKNIKNFITGFEKNSVEGHRHRKVYFENIKSDITKDNQLDSIKKKI PSVCLSNLLGHLSNLQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGS QESLAVRNLIQQLEQSQDYISILEKTPPEITIPPYEARTNTGMEKDQSLLLNPEKL NNLYPNWRNLIPGIIDAHPFLEKDLEHTKLRDRKRIISPSKQDEKRDSYILQRYLD LNKKIDKFKIKKQLSFLGQGKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNK EREDAAQGIWFDNAFSLCELSNINPPRKQKILPLLVGAILSEDFINNKDKWAKFK IFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNNKALIKIIQT IPDIIQAIQSHLGHNDSQALIYHNPFSLSQLYTILETKRDGFHKNCVAVTCENYW RSQKTEIDPEISYASRLPADSVRPFDGVLARMMQRLAYEIAMAKWEQIKHIPDN SSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAIEKQNIQWEEKFQR1INASMNI CPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYLL EHLSPLYLKHQFGTDNVSDIKNFISQNVANIKKYISFFiLLTPEQQKAARHALFLD YDDEAFKTITKFLMSQQKARVNGTQKFLGKQIMEFLSTLADSKQLQLEFSIKQIT AEEVHDHRELLSKQEPKLVKSRQQSFPSHAIDATLTMSIGLKEFPQFSQELDNS WFINHLMPDEVHLNPVRSKEKYNKPNISSTPLFKDSLYAERFIPVWVKGETFAIG FSEKDLFEIKPSNKEKLFTLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLAL EFLHHYFHKEIVTPDDTTVCHFINSLRYYTKKESITVKILKEPMPVLSVKFESSKK NVLGSFKHTIALPATKDWERLFNHPNFLALKANPAPNPKEFNEFIRKYFLSDNN PNSDIPNNGHNIKPQKHKAVRKVFSLPVIPGNAGTMMRIRRKDNKGQPLYQLQ TIDDTPSMGIQINEDRLVKQEVLMDAYKTRNLSTIDGINNSEGQAYATFDNWLT LPVSTFKPEIIKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDSIDDPLNMP NEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQ P

N. lactamica Cas9	MAAFKPNPMNYILGLDIGIASVGWAMVEVDEEENPIRLIDLGVRVFERAEVPKT GDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQDADFDENGLVKSL PNTPWQLRAAALDRKLTCLEWSAVLLHLVKHRGYLSQRKNEGETADKELGAL LKGVADNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAE LNLLFEKQKEFGNPHVSDGLKEDIETLLMAQRPALSGDAVQKMLGHCTFEPAE PKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYA QARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKS PLNLSTELQDEIGTAFSLFKTDKDITGRLKDRVQPEILEALLKHISFDKFVQISLK ALRRIVPLMEQGKRYDEACAEIYGDHYCKKNAEEKIYLPPIPADEIRNPVVLRA LSQARKVINCVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAA AKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDH ALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVE TSRFPRSKKQRILLQKFDEEGFKERNLNDTRYVNRFLCQFVADHILLTGKGKRR VFASNGQITNLLRGFWGLRKVRTENDRHHALDAVVVACSTVAMQQKITRFVR YKEMNAFDGKTIDKETGEVLHQKAHFPQPWEFFAQEVMIRVFGKPDGKPEFEE ADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKR LDEGISVLRVPLTQLKLKGLEKMVNREREPKLYDALKAQLETHKDDPAKAFAE PFYKYDKAGSRTQQVKAVRJEQVQKTGVWVRNHNGIADNATMVRVDVFEKG GKYYLVPIYSWQVAKGILPDRAVVAFKDEEDWTVMDDSFEFRFVLYANDLIKL TAKKNEFLGYFVSLNRATGAIDIRTHDTDSTKGKNGIFQSVGVKTALSFQKNQI DELGKEIRPCRLKKRPPVR
N. meningitides **Cas9**	1VLAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRVFERAEVPKTG DSLAN4ARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPN TPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKG VADNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILL FEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAA KNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLIVIDEPYRKSKLTYAQARK LLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLS PELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIV PLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARK VINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREY FPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFSRT WDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRS KKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASN GQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMN AFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEK LRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDEGVS VLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKY DKAGNRTQQVKAVRVEQVOKTGVWVRNHNGIADNATMVRVDVFEKGDKYY LVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKA RMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEI RPCRLKKRPPVR
B. longum Cas9	MLSRQLLGASHLARPVSYSYNVQDNDVHCSYGERCFMRGKRYRIG1DVGLNSV GLAAVEVSDENSPVRLLNAQSVIHDGGVDPQKNKEAITRKNMSGVARRTRRM RRRKRERLHKLDMLLGKFGYPVIEPESLDKPFEEWHVRAELATRYIEDDELRRE SISIALRHN4ARHRGWRNPYRQVDSLISDNPYSKQYGELKEKAKAYNDDATAAE EESTPAQLVVAMLDAGYAEAPRLRWRTGSKKPDAEGYLPVRLMQEDNANELK QIFRVQRVPADEWKPLFRSVFYAVSPKGSAEQRVGQDPLAPEQARALKASLAF QEYRIANVITNLRIKDASAELRKLTVDEKQSIYDQLVSPSSEDITWSDLCDFLGF KRSQLKGVGSLTEDGEERISSRPPRLTSVQRIYESDNKIRKPLVAWWKSASDNE HEAMIRLLSNTVDIDKVREDVAYASAIEFIDGLDDDAL TKLUSVDLPSGRAAYS VETLQKLTRQMLTTDDDLHEARKTLFNVTDSWRPPADPIGEPLGNPSVDRVLK NVNRYLMNCQQRWGNPVSVNIEHVRSSFSSVAFARKDKREYEKNNEKRSIFRS SLSEQLRADEQMEKVRESDLRRLEAIQRQNGQCLYCGRTITFRTCEMDHIVPRK

	GVGSTNTRTNFAAVCAECNRMKSNTPFAIWARSEDAQTRGVSLAEAKKRVTM FTFNPKSYAPREVKAFKQAVIARLQQTEDDAAIDNRSIESVAWMADELHRRID WYFNAKQYVNSASIDDAEAETMKTTVSVFQGRVTASARRAAGIEGKIHFIGQQ SKTRLDRRHHAVDASVIAMMNTAAAQTLMERESLRESQRLIGLMPGERSWKE YPYEGTSRYESFHLWLDNMDVLLELLNDALDNDR1AVMQSQRYVLGNSIAHD ATIHPLEKVPLGSAMSADLIRRASTPALWCALTRLPDYDEKEGLPEDSHREIRV HDTRYSADDEMGFFASQAAQIAVQEGSADIGSAIHHARVYRCWKTNAKGVRK YFYGMIRVFQTDLLRACHDDLFTVPLPPQSISMRYGEPRWQALQSGNAQYLG SLVVGDEIEMDFSSLDVDGQIGEYLQFFSQFSGGNLAWKHWVVDGFFNQTQLR IRPRYLAAEGLAKAFSDDVVPDGVQKIVTKQGWLPPVNTASKTAVRIVRRNAF GEPRLSSAHHMPCSWQWRHE
A . **muciniphila Cas9**	MSRSLTFSFDIGYASIGWAVIASASHDDADPSVCGCGTVLFPKDDCQAFKRREY RRLRRNIRSRRVRIERIGRLLVQAQIITPEMKETSGHPAPFYLASEALKGHRTLAP IELWHVLRWYAHNRGYDNNASWSNSLSEDGGNGEDTERVKHAQDLMDKHGT ATMAETICRELKLEEGKADAPMEVSTPAYKNLNTAFPRLIVEKEVRRILELSAPL IPGLTAEIIELIAQHHPLTTEQRGVLLQHGIKLARRYRGSLLFGQLIPRFDNRJISR CPVTWAQVYEAELKKGNSEQSARERAEKLSKVPTANCPEFYEYRMARILCNIR ADGEPLSAEIRRELMNQARQEGKLTKASLEKAISSRLGKETETNVSNYFTLHPD SEEALYLNPAVEVLQRSGIGQILSPSVYRIAANRLRRGKSVTPNYLLNLLKSRGE SGEALEKKIEKESKKKEADYADTPLKPKYATGRAPYARTVLKKVVEEILDGEDP TRPARGEAHPDGELKAHDGCLYCLLDTDSSVNQHQKERRLDTMTNNHLVRHR MLILDRLLKDLIQDFADGQKDRISRVCVEVGKELTTFSAMDSKKIQRELTLRQK SHTDAVNRLKRKLPGKALSANLIRKCRIAMDMNWTCPFTGATYGDHELENLEL EHIVPHSFRQSNALSSLVLTWPGVNRMKGQRTGYDFVEQEQENPVPDKPNLHI CSLNNYRELVEKLDDKKGHEDDRRRKKKRKALLMVRGLSHKHQSQNHEAMK EIGMTEGMMTQSSHLMKLACKSIKTSLPDAHIDMIPGAVTAEVRKAWDVFGVF KELCPEAADPDSGKILKENLRSLTHLHHALDACVLGLIPYIIPAHHNGLLRRVLA MRRIPEKLIPQVRPVANQRHYVLNDDGRMMLRDLSASLKENIREQLMEQRVIQ HVPADMGGALLKETMQRVLSVDGSGEDAMVSLSKKKDGKKEKNQVKASKLV GVFPEGPSKLKALKAAIEIDGNYGVALDPKPVVIRHIKVFKRIMALKEQNGGKP VRILKKGMLIHLTSSKDPKHAGVWRIESIQDSKGGVKLDLQRAHCAVPKNKTH ECNWREVDLISLLKKYQMKRYPTSYTGTPR
0 . laneus Cas9	METTLGIDLGTNSIGLALVDQEEHQILYSGVRIFPEGINKDTIGLGEKEESRNATR RAKRQMRRQYFRKKLRKAKLLELLIAYDMCPLKPEDVRRWKNWDKQQKSTV RQFPDTPAFREWLKQNPYELRKQAVTEDVTRPELGRILYQMIQRRGFLSSRKGK EEGKIFTGKDRMVGIDETRKNLQKQTLGAYLYDIAPKNGEKYRFRTERVRARY TLRDMYIREFEIIWQRQAGHLGLAHEQATRKKNIFLEGSATNVRNSKLITHLQA KYGRGHVLIEDTRITVTFQLPLKEVLGGKIEIEEEQLKFKSNESVLFWQRPLRSQ KSLLSKCVFEGRNFYDPVHQKWIIAGPTPAPLSHPEFEEFRAYQFINNIIYGKNEH LTAIQREAVFELMCTESKDFNFEKIPKHLKLFEKFNFDDTTKVPACTTISQLRKL FPHPVWEEKREEIWHCFYFYDDNTLLFEKLQKDYALQTNDLEKIKKIRLSESYG NVSLKAIRRINPYLKKGYAYSTAVLLGGIRNSFGKRFEYFKEYEPEIEKAVCRIL KEKNAEGEVIRKIKDYLVHNRFGFAKNDRAFQKLYHHSQAITTQAQKERLPET GNLRNPIVQQGLNELRRTVNKLLATCREKYGPSFKFDHIHVEMGRELRSSKTER EKQSRQIRENEKKNEAAKVKLAEYGLKAYRDNIQKYLLYKEIEEKGGTVCCPY TGKTLNISHTLGSDNSVQIEHIIPYSISLDDSLANKTLCDATFNREKGELTPYDFY QKDPSPEKWGASSWEEIEDRAFRLLPYAKAQRFIRRKPQESNEFISRQLNDTRYI SKKAVEYLSAICSDVKAFPGQLTAELRHLWGLNNILQSAPDITFPLPVSATENHR EYYVITNEQNEVIRLFPKQGETPRTEKGELLLTGEVERKVFRCKGMQEFQTDVS DGKYWRRIKLSSSVTWSPLFAPKPISADGQIVLKGRIEKGVFVCNQLKQKLKTG LPDGSYWISLPVISQTFKEGESVNNSKLTSQQVQLFGRVREGIFRCHNYQCPASG ADGNFWCTLDTDTAQPAFTPIKNAPPGVGGGQIILTGDVDDKGIFHADDDLHYE LPASLPKGKYYGIFTVESCDPTLIPIELSAPKTSKGENLIEGNIWVDEHTGEVRFD PKKNREDQRHHAIDAIVIALSSQSLFQRLSTYNARRENKKRGLDSTEHFPSPWP

	GFAQDVRQSVVPLLVSYKQNPKTLCKISKTLYKDGKKIHSCGNAVRGQLHKET VYGQRTAPGATEKSYHIRXDIPvELKTSKHIGKWDITIRQMLLKHLQENYHIDIT QEFNIPSNAFFKEGVYRJFLPNKHGEPVPIKKIRMKEELGNAERLKDNINQYVNP RKNHHVMIYQDADGNLKEEIVSFWSVIERQNQGQPIYQLPREGRNIVSILQINDT FLIGLKEEEPEVYRNDLSTLSKHLYRVQKLSGMYYTFRHHLASTLNNEREEFRI QSLEAWKRANPVKVQIDEIGRITFLNGPLC

[0053] Those Cas9 sequences used in the examples disclosed herein are provided below.

[0054] YP_898402.1 membrane protein [Francisella tularensis subsp. novicida U112]

MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNGKVYELSKDSYTLLMNNRTA
RRHQRRGIDRKQLVKRLFKXIWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSPEY
LMVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKJSEIYNKLMQKILEFKLMKL
CTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYTDKQGNLKELSYYH
HDKYMQEFLKJlHATrNDWLDTLLTDDLDIWNFNFEKFDFDKNEEKLQNQEDKDHI
QAHLHHFVFAVNKIKSEMASGGRHRSQWQEITWLDENNHQEGYLKNFCENLHMC
KYSNLSVKNLVNLIGNLSNLELKPLRKYFNDKIHAKADHWDEQKFTETYCHWILGE
WRVGVKDQDKKI)GAKYSYKDLCNELKQKVTKAGLVDFLLELDPCRTIPPYLDNNN
RKPPKCQSLILNPKFLDNQYPNWQQYLQELKKLQSIQNYLDSFETDLKVLKSSKDQP
YFVEYKSSNQQIASGQRDYKDLDARILQFIFDRVKASDELLLNEIYFQAKKLKQKASS
ELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYKQRQRARDSRLY
IMPEYRYDKKLHKY>WTGRFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNF
LKDKIGSDDDLFISKWLVEfflRGFKKACEDSLKIQKDNRGLLNHKINIARNTKGKCEK
EIFNLICKIEGSEDKKGNYKHGLAYELGVLLFGEPNEASKPEFDRKIKKFNSIYSFAQI
QQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKAQRLPAIPTRI
VDGAVKKMATILAKNIVDDNWQNIKQVLSAKHQLfflPIITESNAFEFEPALADVKGK
SLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPR
SHKKYGTLNDEANLICVTRGDNKNKGNRIFCLRDLADNYKLKQFETTDDLEIEKKIA
DTIWDANKKDFKFGNYRSFINLTPQEQKAFRHALFLADENPIKQAVIRAINNRNRTFV
NGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIAEIRQLYEKVDSDI

QAYAKGDKPQASYSHLIDAMLAFCIAADEHRNDGSIGLEIDKNYSLYPLDKNTGEVF
TKDffSQIKITDNEFSDKKLVRKKAIEGFNTHRQMTRDGIYAEWLPILIHKELNEVPvK
GYTWK^SEEIKIFKGKKYDIQQLNNLVYCLKFVDKPISIDIQISTLEELRMLTTNNIAA
TAEYYYINLKTQKLHEYYIENYNTALGYKKYSKEMEFLRSLAYRSERVKIKSIDDVK
QVLDKDSNFIIGKJTLPFKXEWQRLYREWQNTTIKDDYEFLKSFFNVKSITKLHKKVR
KDFSLPISTNEGKFLVKRKTWDNNFIYQILNDSDSRADGTKPFIPAFDISKNEIVEAIID
SFTSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEVETPSDLRDIGIATIQYKIDNNSRPK
VRVKLDYVIDDDSKINYFMNHSLLKSRYPDKVLEILKQSTIIEFESSGFNKTIKEMLG
MKLAGIYNETSNN

[0055] ZP_05061364.1 CRISPR-associated large protein (provisional), putative [gamma proteobacterium HTCC5015J

MTKNYISPIAIDLGAKFTGVALYQYLEGADCTQEVAKGLLVDDRGNVTWSQEGRRG
KRHQVRGYKJIRKMAKRLLWLILDSEYGIKREEVTEPLLKFINGLLNRRGYTYISEEV
DEESMm^SPLPFSEMMPDYFNSSAPLLEQLAKLLSDKNKLVRFRAEGKIPSNKNEFK
KLLDTALDGKYKDEKKELSEAWGMLIASENVLKSTVDGHKSRSEYLANIKEDIKSN
EELEKQISSKEIDGFYNLVGHLSNFQLRLLRKYFNDPNMSGVSYWDEKRLEKYFYQ
WVQGWHTKGGTDEAEKKMILKTKGAPLLKTLKSLSADLTIPPYEDQNNRRPPKCQS
VLLSDEKLTMHYPKWKEWVGQLVKQNDNAYLNENVTLANALHRIVERSRSIDPYQ
LRLLISITDAEKRNDLAGYKRLKLSLGSEVDEFLLLVKNIVDETKEAREGLWFETENK
LFFKCGKTPPRKEKLKSTLLSAVLGKNL SDDEQSSFIEEFWKSGTPKIERRNVRGWCR
LASQVQKTYGVYLKEYGLQQLHKLEAGKKLDDKPLALLYKNSGLIASKIGEALNIEP
DEVSPJ^ASPHSLAQIFNIIEGDVAGFNKTCRACTYENIWRMQEEKVESLLTNQLLSEIH
GERKVPLKSAMCTRLSADSTRPFDGQMASIIEfflAPvKIAQHKIAQINDVPKEFSIDIPIII
ESNQFSFTAELEEIKRGRGSAKAKKAKELGEKSKAGWVSKTERIKTSSEGICPYTGAP
LGGSGEIDHIIPRSLTGRTKKTVFNSEANLIYCSSKGNHDKGNRVYVIEQLNDKYLKK
QFSTSDVNLIKKKIKTTIQRFTEGGEKLRSFSELSREDQKAFRHALFVPELKSEVTSLL
AVKNITRVNGTQAWLAKKIASLLAEHLDKQGRDYTLSAHQIDPWSVSKQRKMLASA
EPIWAKKDPQPAASHVVDAVCTFLEALEQPHTASRLKTISSTSFEKTGWRSALIPDLIK
VDALDRRPKYRRYNIGSTSLFKDGIYAERFLPILIDENGLMAGYDIDNSLKAKGADV
VFESLSPFLLFKGEEVGAQSLSDWQERIDGRYLYMSIDKVKAFDYLQEKVGEKDIAA
ELLNSIHFTQRXTELRAKFSDDSGKKMKTLDAIRKSLKLTVTVNEIGKRKEKCGFSGT
IGIPAKSAWENLLDEPLLETYWGTKMPPQEIWEKVYRKFffPRMPNQAHRKVRKDFS

LPVVDSVSGGFRVKPvKTPNGYNYQLLAIDGYSAVGFKKEGDNA/DFKSPALVPQIAES
KSVTPISSELVHLDKNEIVYFDEWRKIDISDSDLKQFVSSLELAPGSQNRFYIRFTVDE
DQFERHFKSALRVNGIQDLDTWKTFDWNREIPSLLIPPRSNLFLLETGQKITFEYIAN
GANAEVKKAYSLRRA

[0056] ZP_08324662.1 CRISPR-associated protein, Csxl2 family [Parasutterella excrementihominis YIT 11859]

MGKTFniGVGLDLGGTYTGTFITSHPSDEAEHRDHSSAFTVVNSEKLSFSSKSRTAVR
HRVRSYKGFDLRRRLLLLVAEYQLLQKKQTLAPEERENLRIALSGYLKRRGYARTEA
ETDTSVLESLDPSVFSSAPSFTNFFNDSEPLNIQWEAIANSPETTKALNKELSGQKEAD
FKKYIKTSFPEYSAKEILANYVEGRRAILDASKYIANLQSLGHKHRSKYLSDILQDMK
RDSRITRLSEAFGSTDNLWRIIGNISNLQERAVRWYFNDAKFEQGQEQLDAVKLKNV
LVRALKYLRSDDKEWSASQKQIIQSLEQSGDVLDVLAGLDPDRTIPPYEDQNNRRPP
EDQTLYLNPKALSSEYGEKWKSWANKFAGAYPLLTEDLTEILKNTDRKSRIKIRSDV
LPDSDYRLAYILQRAFDRSIALDECSIRRTAEDFENGVV1KNEKLEDVLSGHQLEEFLE
FANRYYQETAKAKNGLWFPENALLERADLI1PPMKNKILNVIVGQALGVSPAEGTDFI
EEIWNSKVKGRSTVRSICNAIENERKTYGPYFSEDYKFVKTALKEGKTEKELSKKFA
AVIKVLKMVSEVVPFIGKELRLSDEAQSKFDNLYSLAQLYNLIETERNGFSKVSLAAH
LENAWRMTMTDGSAQCCRLPADCVRPFDGFIRKAIDRNSWEVAKRIAEEVKKSVDF
TNGTVKJPVAIEANSFNFTASLTDLKYIQLKEQKLKKKLEDIQRNEENQEKRWLSKEE
RIRADSHGICAYTGRPLDDVGEIDHIIPRSLTLKKSESIYNSEVNLIFVSAQGNQEKKN
NIYLLSNLAKNYLAAVFGTSDLSQITNEIESTVLQLKAAGRLGYFDLLSEKERACARH
ALFLNSDSEARRAVIDVLGSRRKASVNGTQAWFVRSIFSKVRQALAAWTQETGNELI
FDAISVPAADSSEMRXRFAEYRPEFRKPKVQPVASHSIDAMCIYLAACSDPFKTKRM
GSQLAIYEPINFDNLFTGSCQVIQNTPRNFSDKTNIANSPIFKETIYAERFLDIIVSRGEIF
IGYPSNMPFEEKPNRISIGGKDPFSILSVLGAYLDKAPSSEKEKLTIYRVVKNKAFELFS
KVAGSKFTAEEDKAAKILEALHFVTVKQDVAATVSDLIKSKKELSKDSIENLAKQKG
CLKKVEYSSKEFKFKGSLIIPAAVEWGKVLWNVFKENTAEELKDENALRKALEAAW
PSSFGTRNLHSKAKRVFSLPVVATQSGAVRIRRKTAFGDFVYQSQDTNNLYSSFPVK
NGKLDWSSPnHPALQ>m>n.TAYGYTRFVDHDRSISMSEFREVYNKDDLMRIELAQGT
SSRRYLRVEMPGEKFLAWFGENSISLGSSFKFSVSEVFDNKIYTENAEFTKFLPKPRED
NKHNGTIFFELVGPRVIFNYIVGGAASSLKEIFSEAGKERS

[0057] YP_122507.1 hypothetical protein lpp0160 [Legionella pneumophila str. Paris]

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRA
TRFmVRNKKRNQFVKJlVALQLFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDE
YIKOETTINLLKELLPSESEHM^IDWFLQKMQSSEFRXILVSKVEEKKDDKELKNAVK
MKNFITGFEKNSVEGHRHRKVYFEMKSDITKDNQLDSIKKKIPSVCLSNLLGHLSNL
QWKNLHRYLAKNPKQFDEQTFGNEFLRMLKM^RHLKGSQESLAVRNLIQQLEQSQD
YISILEKTPPEITIPPYEARTOTGMEKDQSLLLNPEKLNNLYPNWRNLIPGIIDAHPFLE
KDLEHTKLPJ)RKRIISPSKQDEKRDSYILQRYLDLNKKIDKFKIKKQLSFLGQGKQLP
AM.IETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSN1NPPRK
QKILPLLVGAILSEDFDslNKDKWAKFKIFWNTHKIGRTSLKSKCKEIEEARKNSGNAF
KIDYEEALNHPEHSNNKALIKIIQTIPDIIQAIQSHLGHNDSQALIYHNPFSLSQLYTILE
TKJRDGFHKNCVAVTCENYWRSQKTEIDPEISYASRLPADSVRPFDGVLARMMQRLA
YEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAIEKQNIQW
EEKJQWmASMMCPYKGASIGGQGEIDfflYPRSLSKKHFGVIFNSEVNLIYCSSQGNR
EKJ<£EHYLLEHLSPLYLKHQFGTDNVSDIKNFISQNVANIKKYISFHLLTPEQQKAAR
HALFLDYDDEAFKTITKFLMSQQKARVNGTQKFLGKQIMEFLSTLADSKQLQLEFSI
KQITAEEVHDHRELLSKQEPKLVKSRQQSFPSHAIDATLTMSIGLKEFPQFSQELDNS
WFnNiHLMPDEVHLNPVRSKEKYNKPNISSTPLFKDSLYAERFIPVWVKGETFAIGFSE
KJDLFEIKPSNKEKLFTLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHY
FHKEIVTPDDTTVCHFrNSLRYYTKKESITVKILKEPMPVLSVKFESSKKNVLGSFKHT
IALPATKDWERLF>^NFLALKANPAPNPKEFNEFIRKYFLSDNNPNSDIPNNGHN1K
PQKHKAVRKVFSLPVIPGNAGTMMRIRRKDNKGQPLYQLQTIDDTPSMGIQINEDRL
VKQEVLMDAYKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEIIKLEMKPHSK
TRRYIRITQSLADFIKTIDEALMIKPSDSlDDPLNMPNEIVCKNKL.FGNELKPRDGKMK
IVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

[0058] NP_907747.1 hypothetical protein WS1613 [Wolinella succinogenes DSM 1740]

MLVSPISVDLGGKNTGFFSFTDSLDNSQSGTVIYDESFVLSQVGRRSKRHSKRNNLRN
KLVKRLFLLlLQEiiHGLSIDVLPDEIRGLFNKRGYTYAGFELDEKKKDALESDTLKEF
LSEKLQSIDRDSDVEDFLNQIASNAESFKDYKKGFEAVFASATHSPNKKLELKDELKS
EYGENAKELLAGLRVTKEILDEFDKQENQGNLPRAKYFEELGEYIATNEKVKSFFDS
NSLKLTDMTKiIGNISNYQLKELRRYFM)KEMEKGDIWIPNKLHKITERFVRSWHPK

hTOADRQRRAELMKDLKSKEIMELLTTTEPVMTIPPYDDMNNRGAVKCQTLRLNEEY
LDKHLPNWRDIAKRLNHGKFNDDLADSTVKGYSEDSTLLHRLLDTSKEIDIYELRGK
KPNELLVKTLGQSDANRLYGFAQNYYELIRQKVRAGIWVPVKNKDDSLNLEDNSN
MLKRCNHNPPHKKNQIHNLVAGILGVKLDEAKFAEFEKELWSAKVGNKKLSAYCK
N1EELRKTHGNTFKIDIEELRKKDPAELSKEEKAKLRLTDDVILNEWSQKIANFFDIDD
KHRQRFNNLFSMAQLHTVIDTPRSGFSSTCKRCTAENRFRSETAFYNDETGEFHKKA
TATCQRLPADTQRPFSGKIERYIDKLGYELAKIKAKELEGMEAKEIKVPIILEQNAFEY
EESLRKSKTGSM)RVINSKKDRDGKKLAKAKENAEDRLKDKDKRIKAFSSGICPYCG
DTIGDDGEIDHILPRSHTLKIYGTVFNPEGNLIYVHQKCNQAKADSIYKLSDIKAGVSA
QWIEEQVAN1KGYKTFSVLSAEQQKAFRYALFLQNDNEAYKKVVDWLRTDQSARV
NGTQKYLAKKIQEKLTKMLPNKHLSFEFILADATEVSELRRQYARQNPLLAKAEKQA
PSSHAIDAVMAFVARYQKVFKDGTPPNADEVAKLAMLDSWNPASNEPLTKGLSTNQ
KIEKMIKSGDYGQKNMREVFGKSIFGENAIGERYKPIVVQEGGYYIGYPATVKKGYE
LKNCKVVTSK>TOIAKLEKJIKNQDLISLKENQYIKIFSINKQTISELSNRWNMNYKNL
VERDKEIVGLLEFIVENCRYYTKKVDVKFAPKYIHETKYPFYDDWRRFDEAWRYLQ
ENQNKTSSKDRFVIDKS SLNEYYQPDKNEYKLDVDTQPIWDDFCRWYFLDRYKTAN
DKKSIRIKARKTFSLLAESGVQGKVFRAKRKIPTGYAYQALPMDNNVIAGDYANILL
EANSKTLSLVPKSGISIEKQLDKKLDVIKKTDVRGLA1DNNSFFNADFDTHGIRLIVEN
TSVKVGNFPISAIDKSAKRMIFRALFEKEKGKiOCKKTTISFKESGPVQDYLKVFLKKI
VKIQLRTDGSISNIVVRKNAADFTLSFRSEHIQKLLK

[0059] ADX75954.1 CRISPR-associated protein, Csnl family [Staphylococcus pseudintermedius ED99]

MGRKPYILSLDIGTGSVGYACMDKGFNVLKYHDKDALGVYLFDGALTAQERRQFRT
SRRRKNRRIKJRLGLLQELLAPLVQNPNFYQFQRQFAWKNDNMDFKNKSLSEVLSFL
GYESKKYPTIYHLQEALLLKDEKFDPELIYMALYHLVKYRGHFLFDHLKIENLTOND
NMHDFVELIETYENLNMKLNLDYEKTKVIYEILKDNEMTKNDRAKRVKNMEKKLE
QFSIMLLGLKFNEGKLFNHADNAEELKGANQSHTFADNYEETMLTPFLTVEQSEFIERA
NKIYLSLTLQDILKGKKSMAMSKVAAYDKFRNELKQVKDIVYKADSTRTQFKKIFVS
SKKSLKQYDATPNDQTFSSLCLFDQYLIRPKKQYSLLIKELKKIIPQDSELYFEAENDT
LLKVLNTTDNASIPMQINLYEAETILRNQQKYHAEITDEMIEKVLSLIQFRIPYYVGPL
VM)HTASKFGWMEPvKSNESIKPWNFDEVVDRSKSATQFIRRMTOKCSYLr^ DVLP
KNSLLYQEMEVLNELNATQIRLQTDPKNRKYRMMPQIKLFAVEfflFKKYKTVSHSKF

LEIMLNSNHRENFMNHGEKL SIFGTQDDKKFASKL SSYQDMTKIFGDIEGKRAQIEEII
QWITIFEDKKILVQKLKECYPELTSKQINQLKKLNYSGWGRLSEKLLTHAYQGHSIIE
LLRHSDENFMEILTNDVYGFQNFIKEENQVQSNKIQHQDIANLTTSPALKKGIWSTIK
LVRELTSIFGEPEKIIMEFATEDQQKGKKQKSRKQLWDDN1KKNKLKSVDEYKYIIDV
ANKLNNEQLQQEKLWLYLSQNGKCMYSGQSIDLDALLSPNATKHYEVDfflFPRSFIK
DDSIDNKVLVIKKMNQTKGDQVPLQFIQQPYERIAYWKSLNKAGLISDSKLFIKLMKP
EFTAMDKEGFIQRQL**VETRQI**SVHYRDFLKEEYPNTKVIPMKAKMVSEFRKKFDIPKI
RQMNDAHHAIDAYLNGVVYHGAQLAYPNVDLFDFNFKWEKVREKWKALGEFNTK
QKSRELFFFKKLEKMEVSQGERLISKIKLDMh^KINYSRKLANTPQQFYNQTAVSPK
TAELKYESNKSNEVVYKGLTPYQTWVAIKSVNKKGKEKMEYQMIDHYVFDFYKF
QNGNEKELALYLAQRENKDEVLDAQIVYSLNKGDLLYFNNHPCYFVSRKEVINAKQ
FELTVEQQLSLYNVMNNKETNVEKLLIEYDFIAEKVINEYHHYLNSKLKEKRVRTFFS
ESNQTHEDFIKALDELFKVVTASATRSDKIGSRKNSMTHRAFLGKGKDVKIAYTSISG
LKTTKPKSLFKLAESRNEL

[0060] ZP_10206685.1 CRISPR-associated protein, Csnl family [Planococcus antarcticus DSM 14505]

MKNYTIGLDIGVASVGWVCIDENYKILNYNTSRHAFGVHEFESAESAAGRRLKRGMR
RRYNRRKKRLQLLQSLFDSYITDSGFFSKTDSQHFWKNNNEFENRSLTEVLSSLRISS
RKYPTIYHLRSDLIESNKKMDLRLVYLALHNLVKYRGHFLQEGNWSEAASAEGMDD
QLLELVTRYAELENLSPLDLSESQWKAAETLLLNRNLTKTDQSKELTAMFGKEYEPF
CKLVAGLGVSLHQLFPSSEQALAYKETKTKVQLSNENVEEVMELLLEEESALLEAVQ
PFYQQVVLYELLKGETYVAKAKVSAFKQYQKDMASLKNLLDKTFGEKVYRSYFISD
KNSQREYQKSHKVEVLCK1DQFNKEAKFAETFYKDLKKLLEDKSKTSIGTTEKDEM
LRIIKAIDSNQFLQKQKGIQNAAIPHQNSLYEAEKILRNQQAHYPFITTEWIEKVKQIL
AFRIPYYIGPLVKDTTQSPFSWVERKGDAPITPWNFDEQIDKAASAEAFISRMRKTCT
YLKGQEVLPKSSLTYERFEVLNELNGIQLRTTGAESDFRFERLSYEMKCWIIDNVFKQ
YKTVSTKRLLQELKKSPYADELYDEHTGEIKEVFGTQKENAFATSLSGYISMKSILGA
VVDDNPAMTEELIYWIAVFEDREILHLKIQEKYPSITDVQRQKLALVKLPGWGRFSRL
LIDGLPLDEQGQSVLDHMEQYSSVFMEVLKNKGFGLEKKIQKMNQHQVDGTKKIRY
EDIEELAGSPALKRGIWRSVKIVEELVSIFGEPANIVLEVAREDGEKKRTKSRKDQWE
ELTKTTLKNDPDLKSFIGEIKSQGDQRFNEQRFWLYVTQQGKCLYTGKALDIQNLSM
YEVDmLPQNFVKDDSLDNLALVMPEANQRKNQVGQNKMPLEIIEANQQYAMRTL

WERLHELKLISSGKLGPvLKXPSFDEVDKDKFIARQLVETRQIIKHVRDLLDERPSKSDI HLVKAGIVSKFPvRFSEIPKIRDYNNKHHAMDALFAAALIQSILGKYGKNFLAFDLSKK DRQKQWRSVKGSNKEFFLFKNFGNLRLQSPVTGEEVSGVEYMKHVYFELPWQTTK MTQTGDGMFYKESIFSPKVKQAKYVSPKTEKFVHDEVKNHSICLVEFTFMKKEKEV QETKFIDLKVIEHHQFLKEPESQLAKFLAEKETNSPIIHARIIRTIPKYQKIWIEHFPYYFI STRELHNARQFEISYELMEKVKQLSERSSVEELKIVFGLLIDQMNDNYPIYTKSSIQD RVQKFVDTQLYDFKSFEIGFEELKKAVAANAQRSDTFGSRISKKPKPEEVAIGYESIT GLKYRKPRSVVGTKR

[0061] ZP_16930555.1 csnl family CRISPR-associated protein [Streptococcus sanguinis SK49]

MTKFNKNYSIGLDIGVSSVGYAVVTEDYRVPAFKFKVLGNTEKEKJKKNLIGSTTFVS AQPAKGTRVFRVNRJIRIDRRNHRITYLRDIFQKEIEKVDKNFYRRLDESFRVLGDKSE DLQIKQPFFGDK£LETAYHKKYPT1YHLRKHLADADKNSPVAD1REVYMAISHILKY RGHFLTLDKINPNMNMQNSWIDFIESCQEVFDLEISDESKMADIFKSSEmQEKVKKI LPYFQQELLKKDKSIFKQLLQLLFGLKTKFKDCFELEEEPDLNFSKENYDENLENFLG SLEEDFSDVFAKLKVLRDTILLSGJVLTYTGATHARPSATMVERYEEHRKDLQRFKFF IKQNLSEQDYLDIFGRKTQNGFDVDKETKGYVGYITNKMVLTNPQKQKTIQQNFYD YISGKITGIEGAEYFLNKISDGTFLRKLRTSDNGAIPNQIHAYELEKIIERQGKDYPFLL ENKDKLLSILTFKIPYYVGPLAKGSNSPJ^AWIKRATSSDILDDNDEDTRNGKJRPWNY QKLINMDETRDAFITNLIG>TOIILLNEKVLPKRSLIYEEVMLQNELTRVKYKDKYGKA HFFDSELRQMINGLFKNNSKRVNAKSLIKYLSDNHKDLNAIEIVSGVEKGKSFNSTLK TYNDLKTIFSEELLDSEIYQKELEEnKVITVFDDKKSIKNYLTKFFGHLEILDEEKINQL SKLRYSGWGRYSAKLLLDIRDEDTGFNLLQFLRNDEENRNLTKLISDNTLSFEPKIKDI QSKSTIEDDIFDEIKKLAGSPAIKRGILNSIKIVDELVQIIGYPPHNIVIEMARENMTTEE GQKKAKTRKTKLESALKMENSLLENGKVPHSDEQLQSEKLYLYYLQNGKDMYTLD KTGSPAPLYLDQLDQYEVDHIIPYSFLPIDSIDNKVLTHRENNQQKLNNIPDKETVAN MKPFWEKLYNAKLISQTKYQRLTTSERTPDGVLTESMKAGFIERQLVETRQIIKHVA RILDNRFSDTKIITLKSQLITNFRNTFffIAKIRELNDYHHAFIDAYLAVVVGQTLLKVYP KLAPELIYGHHAOTNRmENKATLRKFILYSMMRFFh^DSKVSKDIWDCNRDLPIK DVIYNSQINFVKRTMIKKGAFYNQNPVGKFNKQLAANNRYPLKTKALCLDTSIYGG YGPMNSALSHIIAERFNEKXGKIETVKEFHDIFnDYEKFNNNPFQFLNDTSENGFLKK NNIi¾VLGFYWPKYSLMQKIDGTRMLFESKSNLHKATQFKLTKTQNELFFHMKRLL

TKSW·MDLKSKSAIKESQOTILKHKEEFDMSNQLSAFSQKMLGNTTSLKNLIKGYNE
RKIK£IDIRDETIKYFYDNFIKMFSFVKSGAPKI)INDFFDNKCTVARMRPKPDKKLLN
ATLIHQSITGLYETRIDLSKLGED

[0062] AAK33936.1 conserved hypothetical protein [Streptococcus pyogenes M1 GAS]

MDKKYSIGLDIGTOSVGWAV1TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATPvLKJITARRRYTRRKNPJCYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGMVDEVAYFiEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWWEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRXPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAmFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPEMVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDfflVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDE>n3KLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYFfflAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSMMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVMVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKJT^LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEMIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0063] YP_820832.1 CRISPR-system-like protein [Streptococcus thermophilus LMD-9]

MTKPYSIGLDIGTOSVGWAVTTDWKVPSKKMKVLGNTSKXYIKKNLLGVLLFDSGI
TAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKPvDS
KYPIFG^VEEKAYHDEFPTIYHLPvKYLADSTKXADLRLVYLALAHMIKYRGHFLIE
GEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLFP
GEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGYIGDDYSD
VFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRMSLKT
YNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKKLLAEFEGADYFLEKIDREDFLRKQ
RTFDNGSIPYQIHLQEMRAILDKQAKFYPFLAKNKERIEKILTFRIPYYVGPLARGNSD
FAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFN
VYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDG
IELKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKF
EMFDKSVLKKLSRRHYTGWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLI
HDDALSFKKKIQKAQIIGDEDKGN1KEVVKSLPGSPAIKKGILQSIKIVDELVKVMGG
RKPESrVVEMARENQYTNQGKSNSQQRLKRLEKSLKELGSKILKENIPAKLSKIDNNA
LQNDRLYLYYLQNGKDMYTGDDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSAS
NRGKSDDVPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLSPEDKAGFIQR
QLVETRQITKHVARLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVR
EINDFHHAHDAYLNAVVASALLKKYPKLEPEFVYGDYPKYNSFRERKSATEKVYFY
SMJVINIFKKSISLADGRVIERPLIEX^ETGESVW^ KESDLATVRRVLSYPQVNVVKK
VEEQNHGLDRGKPKGLFNA^SSKPKPNSNENLVGAKEYLDPKKYGGYAGISNSFT
VLVKGTIEKGAKKKITWLEFQGISILDRTNYRKDKLNFLLEKGYKDIELIIELPKYSLF
ELSDGSPvRMLASILSTONKRGEIHKGNQIFLSQKFVK^LYHAKRISNTINENHRKYVE
NHKKEFEELFYYILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSERK
GLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGEG

[0064] NP_721764.1 hypothetical protein SMU_1405c [Streptococcus mutans UA159]

MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSfflEKNLLGALLFDSG
NTAEDRRLKRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRG
ERHPIFGNLEEEVKYHENFPT1YHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEG
KFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLF

PNEKSNGPvFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNY
AELFLSAKKLYDSILLSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLS
DKYNEVFSDVSKDGYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRK
QRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYYGPLARGKS
DFAWLSRKSADKITPWNFDEIVDKESSAEAFINRMTWDLYLPNQKVLPKJISLLYEK
FTVYNELTKVKYKTEQGKTAFFDA>MKQEIFDGVFKVYRKVTKDKLMDFLEKEFDE
FRIVDLTGLDKENKVFNASYGTYHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDRE
MIRKRLEWSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDG
NSNRNFMQLINDDALSFKEEIAKAQVIGETDNLNQVVSDIAGSPAIKKGILQSLKIVDE
LVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENS
QLQNDRLFLYYLQNGRDMYTGEELDIDYLSQYDIDHIIPQAFIKDNSIDNRVLTSSKE
NRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIK
RQLVETRQITKHVARILDERFNTETDENNKKIRQVKIVTLKSNLVSNFRKEFELYKVR
E1NDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYS
MMNFFKKDDVRTDKNGEnWKKDEfflSNIKKVLSYPQVMVKKVEEQTGGFSKESIL
PKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVG
VTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLENGRKRLLASARELQK
GNEIVLPNHLGTLLYHAKMHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLA
EG^EKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNIDRKRYTSTTEI
LNATLIHQSITGLYETRIDLNKLGGD

[0065] YP_004373648.1 CRISPR-associated protein, Csnl family [Coriobacterium glomerans PW2]

MKLRGIEDDYSIGLDMGTSSVGWAVTDERGTLAHFKRKPTWGSRLFREAQTAAVAR
MPRGQRRRWRRRWRLDLLQKLFEQQMEQADPDFFIRLRQSRLLRDDRAEEHADY
RWPLFNDCKFTERDYYQRFPTIYHVRSWLMETDEQADIRLIYLALHN1VKHRGNFLR
EGQSLSAKSARPDEALNHLRETLRVWSSERGFECSIADNGSILAMLTHPDLSPSDRRK
KIAPLFDVKSDDAAADKKLGIALAGAVIGLKTEFKNTFGDFPCEDSSIYLSNDEAVDA
VRSACPDDCAELFDRLCEVYSAYVLQGLLSYAPGQTISANMVEKYRRYGEDLALLK
KLVKIYAPDQYRMFFSGATYPGTGIYDAAQARGYTKYNLGPKKSEYKPSESMQYDD
FRKAVEKLFAKTDARADERYRMMMDRFDKQQFLRRLKTSDNGSIYHQLHLEELKAI
VENQGRFYPFLKRDADKLVSLVSFRIPYYVGPLSTRNARTDQHGENRFAWSERKPG
MQDEPIFPWNWESIIDRSKSAEKFILRMTGMCTYLQQEPVLPKSSLLYEEFCVLNELN

GAHWSIDGDDEHRFDAADREGIIEELFRRKRTVSYGDVAGWMEPvERNQIGAHVCGG
QGEKGFESKLGSYIFFCKDVFKVEPvLEQSDYPMIEPJILWNTLFEDPvKILSQPvLKEEYG
SPvLSAEQIKTICKXRFTGWGRLSEKFLTGITVQVDEDSVSIIVnDVLREGCPVSGKRGRA
MVMMEILRDEELGFQKKVDDFNRAFFAENAQALGVNELPGSPAVRRSLNQSIRIVDE
IASIAGKAPAMFIEVTRDEDPKKKGRJRTKRRYNDLKDALEAFKKEDPELWRELCETA
PNDMDEPvLSLYFMQRGKCLYSGPvAIDIHQLSNAGIYEVDHIIPRTYVKDDSLENKAL
VYREENQPJCTDMLLIDPEIRRRMSGYWRMLHEAKLIGDKKFRNLLRSRIDDKALKG
FIARQLVETGQMVKLVRSLLEARYPETNIISVKASISHDLRTAAELVKCREANDFHHA
HDAFLACRVGLFIQKRHPCVYENPIGLSQVVRNYVRQQADIFKRCRTIPGSSGFIVNS
FMTSGFDKETGEIFKDDWDAEAEVEGIRRSLNFRQCFISRMPFEDHGVFWDATIYSPR
AKKTAALPLKQGLNPSRYGSFSREQFAYFFIYKARNPRKEQTLFEFAQVPVRLSAQIR
QDENALERYARELAKDQGLEFIRIERSKILKNQLIEIDGDRLCITGKEEVRNACELAFA
QDEMRVIRMLVSEKPVSRECVISLFNRILLHGDQASRRLSKQLKLALLSEAFSEASDN
VQRNVVLGLIAIFNGSTOMV^SDIGGSKFAGNVRIKYKKELASPKVNYHLIDQSVT
GMFERRTKIGL

[0066] ZP_08576281.1 possible CRISPR associated protein [Lactobacillus farciminis KCTC 3681]

MTKKEQPYMGLDIGTSSVGWAVTODI^TOLLMK^ NLWGVRLFEEAQTAKETRLN
RSTRRRYRRRKNPJNWLNEIFSEELAKTDPSFLIRLQNSWVSKKDPDRKRDKYNLFID
GPYTDKEYYREFPTIFHLRKELILNKDKADIRLIYLALFIN1LKYRGNFTYEHQKFNISN
LNNNLSKELIELNQQLIKYDISFPDDCDWNHISDILIGRGNATQKSSNILKDFTLDKET
KKLLKEVINLILGNVAHLNTIFKTSLTKDEEKLNFSGKDIESKLDDLDSILDDDQFTVL
DAAMUYSTITLNEILNGESYFSMAKVNQYENHAIDLCKLRDMWHTTKNEEAVEQSR
QAYDDYFlv^KYGTK£LYTSLKXFLKVALPTNLAKEAEEKISKGTYLVKPRNSENGV
VPYQLNKJEMEKIIDNQSQYYPFLKENKEKLLSILSFRIPYYVGPLQSAEKNPFAWME
RKSNGHARPWNFDEIVDREKSSNKFIRRMTVTDSYLVGEPVLPKNSLIYQRYEVLNE
LNMRITENLKTNPIGSPXTVETKQMYNELFKKYKKVTVKKLTKWLIAQGYYKNPILI
GLSQKDEFNSTLTTYLDMKKIFGSSFMEDNKNYDQIEELIEWLTIFEDKQILNEKLHSS
KYSYTPDQIKKISNMRYKGWGRLSKKILMDITTETNTPQLLQLSNYSILDLMWATNN
NFISIMSNDKYDFKWIENHNLNKNEDQMSDLVNDIHVSPALKRGITQSIKIVQEIVK
FMGHAPKfflFIEVTPvETKKSEITTSREKmKPvLQSKlLNKANDFKPQLPvEYLVPNKKIQ
EELKKHK>TOLSSERIMLYFLQNGKSLYSEESLNINKLSDYQVDHILPRTYIPDDSLEN

KALVLAK^NQRKADDLLLNS^IDRNLERWTYIVO.NNNMIGLKKFKNLTRRVITDK
DKLGFIHRQLVQTSQMVKGVANILDNMYKNQGTTCIQARANLSTAFRKALSGQDDT
YHFKIiPELVK>mNV>roFHHAQDAYLASFLGTYRLRRFPTNEMLLMNGEYNKFYGQ
VKELYSKKKKLPDSRKNGFIISPLVNGTTQYDRNTGEIIWNVGFRDKILKIFNYHQCN
VTPJCTEIKTGQFYDQTIYSPKNPKYKKLIAQKKDMDPNIYGGFSGDNKSSITIVKIDN
NKJKPVAIPIPJLINDLKDKKTLQNWLEEm^KHKKSIQIIKNNVPIGQIIYSKKVGLLSLN
SDREVANRQQLILPPEHSALLRLLQIPDEDLDQILAFYDKNILVEILQELITKMKKFYPF
YKGEREFLIANIENFNQATTSEKVNSLEELITLLHANSTSAHLIFNNIEKKAFGRKTHG
LTLNNTDFIYQSVTGLYETRIHIE

[0067] ZP_03683851.1 hypothetical protein CATMIT_02512, partial [Catenibacterium mitsuokai DSM 15897]

IVDYCIGLDLGTGSVGWAVVDMNHRLMKRNGKHLWGSRLFSNAETAANRRASRSI
RRRYNKJIRERIRLLRAILQDMVLEKDPTFFIRLEHTSFLDEEDKAKYLGTDYKDNYN
LFIDEDFNDYTYYHKYPTIYHLRKALCESTEKADPRLIYLALHHIVKYRGNFLYEGQK
FNMDASMEDKLSDIFTQFTSFNMPYEDDEKKNLEILEILKKPLSKKAKVDEVMTLIA
PEKDYKSAFKELVTGIAGNKMNVTKMILCEPIKQGDSEIKLKFSDSNYDDQFSEVEK
DLGEYVEFVDALHNVYSWVELQTIMGATHTDNASISEAMVSRYNKHHDDLKLLKD
CIKNNVPNKYFDMFRNDSEKSKGYYNYINRPSKAPVDEFYKYYKKCIEKVDTPEAK
QILNDIELENFLLKQNSRTNGSVPYQMQLDEMIKIIDNQAEYYPILKEKREQLLSILTF
RIPYWGPLNETSEHAWIKRLEGKENQRILPWNYQDIVDVDATAEGFIKRMRSYCTY
FPDEEVLPKNSLIVSKYEVYNEL>¾IRVDDKLLEVDVKNDIYNELFMKNKTVTEKKL
KNWLVNNQCCSKDAEIKGFQKENQFSTSLTPWIDFTNIFGKIDQSNFDLIENIIYDLTV
FEDKKIMKRRLKKKYALPDDKVKQILKLKYKDWSRLSKKLLDGIVADNRFGSSVTV
LDVLEMSRLNLMEIINDKDLGYAQMIEEATSCPEDGKFTYEEVERLAGSPALKRGIW
QSLQIVEEITKVMKCRPKYIYIEFERSEEAKERTESKIKKLENYYKDLDEQTKKEYKS
VLEELKGFDNTKKISSDSLFLYFTQLGKCMYSGKKLDIDSLDKYQIDffIVPQSLVKDD
SFDNRVLVVPSENQRKLDDLVVPFDIRDKMYRFWKLLFDHELISPKKFYSLIKTEYTE
RDEERFINRQLVETRQITKNVTQnEDHYSTTKVAAIRANLSHEFRVKNHIYKNRDIND
YHHAHDAYIVALIGGFMRDRYPNMHDSKAVYSEYMKMFRKNK>TOQKRWKDGFV1
NSIVnWPYEVDGKLIWNPDLINElKXCFYYKDCYCTTKLDQKSGQLFNLTVLSNDAII
ADKGVTKAVVPVNKNRSDVHKYGGFSGLQYTIVAIEGQKKKGKKTELVKKISGVPL
iiLKAASINEKINYIEEKEGLSDVPJIKOMPVNQMIEMDGGEYLLTSPTEYVNARQLVL

NEKQCALIADIYNAIYKQDYDNLDD1LMIQLYIELTNKMKVLYPAYRGIAEKFESMN
ENYVVISKEEKAMIKQMLIVMHRGPQNGMVYDDFKISDWGRLKTKNHNLNMVFIS
QSPTGIYTKKYKL

[0068] YP_003171950.1 CRISPR-associated protein Csnl [Lactobacillus rhamnosus GG]

MTKLNQPYGIGLDIGSNSIGFAVVDANSHLLRLKGETAIGARLFREGQSAADRRGSRT
TRRRLSRTRWRLSFLRDFFAPHITKIDPDFFLRQKYSEISPKDKDRFKYEKRLFNDRTD
AEFYEDYPSMYHLRLHLMTHTHKADPREIFLAIHHILKSRGHFLTPGAAKDFNTDKV
DLEDIFPALTEAYAQVYPDLELTFDLAKADDFKAKLLDEQATPSDTQKALVNLLLSS
DGEKEIVKKRKQVLTEFAKAITGLKTKFNLALGTEVDEADASNWQFSMGQLDDKW
SN1ETSMTDQGTEIFEQIQELYRARLLNGIVPAGMSLSQAKVADYGQHKEDLELFKTY
LKKLNDHELAKTIRGLYDRYINGDDAKPFLREDFVKALTKEVTAHPNEVSEQLLNR
MGQANFMLKQRTKANGAIPIQLQQRELDQIIANQSKYYDWLAAPNPVEAHRWKMP
YQLDELLNFffIPYWGPLITPKQQAESGENVFAWMVRKDPSGNITPYNFDEKVDREA
SANTFIQRMKTTDTYLIGEDVLPKQSLLYQKYEVLNELNNVRINNECLGTDQKQRLI
REVFERHSSVTIKQVADNLVAHGDFARRPEIRGLADEKRFLSSLSTYHQLKEILHEAI
DDPTKLLDIEMITWSTVFEDHTIFETKLAEIEWLDPKKINELSGIRYRGWGQFSRKLL
DGLKLGNGHTVIQELMLSNHNLMQILADETLKETMTELNQDKLKTDDIEDVINDAY
TSPSNKKALRQVLRVVEDIKHAANGQDPSWLFIETADGTGTAGKRTQSRQKQIQTVY
ANAAQELIDSAVRGELEDKIADKASFTDRLVLYFMQGGRDIYTGAPLNIDQLSHYDI
DffILPQSLIKDDSLDNRVLVNATINREKNNVFASTLFAGKMKATWRKWHEAGLISGR
KLRNLMLRPDEIDKFAKGFVARQLVETRQIIKLTEQIAAAQYPNTKJIAVKAGLSHQL
REELDFPKNPxDVNHYHHAFDAFLAARIGTYLLKRYPKLAPFFTYGEFAKVDVKKFR
EFOTIGALTHAKKMIAKDTGEIVWDKERDIRELDRIYNFKRMLITHEVYFETADLFK
QTIYAAKDSKERGGSKQLIPKKQGYPTQVYGGYTQESGSYNALVRVAEADTTAYQV
IKISAQNASKIASANLKSREKGKQLLNEIVVKQLAKRRKNWKPSANSFKIVIPRFGMG
TLFQNAKYGLFMVNSDTYYRNYQELWLSRENQKLLKKLFSIKYEKTQMNHDALQV
YKAnDQVEKPFKLYDINQFRAKLSDAIERFEKLPINTDGNKIGKTETLRQILIGLQANG
TRSNVKNLGIKTDLGLLQVGSGIKLDKDTQIVYQSPSGLFKRRIPLADL

[0069] YP_003937986.1 CRISPR associated protein [Bifidobacterium bifidum S17]

MSRKNYVDDYAISLDIGNASVGWSAFTPNYRLVRAKGHELIGVRLFDPADTAESRR
MARTTRRRYSRRRWRLRLLDALFDQALSEIDPSFLARRKYSWVHPDDENNADCWY

GSVLFDSNEQDKRFYEKYPTIYHLRKALMEDDSQHDIREIYLAIHHMVKYRGNFLVE
GTLESSNAFKEDELLBCLLGRITRYEMSEGEQNSDIEQDDENKLVAPANGQLADALCA
TRGSRSMRVDNALEALSAVNDLSREQRAIVKAIFAGLEGNXLDLAKIFVSKEFSSEN
KKILGIYFNKSDYEEKCVQIVDSGLLDDEEREFLDRMQGQYNAIALKQLLGRSTSVS
DSKCASYDAHRANWNLIKLQLRTKENEKDINENYGILVGWKIDSGQRKSVRGESAY
ENMRKKANVFFKKMIETSDLSETDKNRLIHDIEEDKLFPIQRDSDNGVIPHQLHQNEL
KQIIKKQGKYYPFLLDAFEKDGKQINKIEGLLTFRVPYFVGPLVVPEDLQKSDNSENH
WMVRXKKGEITPWNFEEMVDKGASGRKFIERLVGTDSYLLGEPTLPKNSLLYQEYE
VLNELNNVRLSVRTGNHWNDKRRMRLGREEKTLLCQRLFMKGQTVTKRTAENLLR
KEYGRTYELSGLSDESKFTSSLSTYGKMCRIFGEKYVNEHRDLMEKIVELQTVFEDK
ETLLHQLRQLEGISEADCALLVNTHYTGWGRLSRKLLTTKAGECKISDDFAPRKHSII
EIMRAEDRNLMEIITDKQLGFSDWIEQENLGAENGSSLMEVVDDLRVSPKVKRGIIQS
IRLIDDISKAVGKRPSRIFLELADDIQPSGRTISRKSRLQDLYRNANLGKEFKGIADELN
ACSDKDLQDDRLFLYYTQLGKDMYTGEELDLDRLSSAYDIDHIIPQAVTQNDSIDNR
VLVARAENARKTDSFTYMPQIADRMRNFWQILLDNGLISRVKFERLTRQNEFSEREK
ERFVQRSLVETRQIMKNVATLMRQRYGNSAAVIGLNAELTKEMHRYLGFSHKNRDI
NDYHHAQDALCVGIAGQFAANRGFFADGEVSDGAQNSYNQYLRDYLRGYREKLSA
EDRKQGRAFGFIVGSMRSQDEQKRVNPRTGEVVWSEEDKDYLRKVMNYRKMLVT
QKVGDDFGALYDETRYAATDPKGIKGIPFDGAKQDTSLYGGFSSAKPAYAVLIESKG
KTRLVNVTMQEYSLLGDRPSDDELRXVLAKKKSEYAKAMLLRJTVPKMQLIRYGGG
LMVIKSAGELNNAQQLWLPYEEYCYFDDLSQGKGSLEKDDLKKLLDSILGSVQCLY
PWHRFTEEELADLFTVAFDKLPEDEKKNVITGIVSALHADAKTANLSIVGMTGSWRR
MNNKSGYTFSDEDEFIFQSPSGLFEKRVTVGELKJRKAKKEWSKYRTOTKRLPTLSG
ASQP

[0070] EHN59352.1 CRISPR-associated protein [Oenococcus kitaharae DSM 17330]

MARDYSVGLDIGTSSVGWAAIDNXYHLIRAKSKNLIGVRLFDSAVTAEKRRGYRTTR
RRLSRRHWRLRLLNDIFAGPLTDFGDENFLARLKYSWVHPQDQSNQAHFAAGLLFD
SK£QDKDFYRKYPTIYHLRLALMhTODQKHDLREVYLAIHHLVKYRGHFLIEGDVKA
DSAFDVHTFADAIQRYAES>WSDE>n.LGKIDEKKLSAALTDKHGSKSQRAETAETAF
DILDLQSKKQIQAILKSVVGNQANLMAIFGLDSSAISKDEQKNYKFSFDDADIDEKIA
DSEALLSDTEFEFLCDLKAAFDGLTLKMLLGDDKTVSAAMVRRFNEHQKDWEYIKS
fflRNAKNAGNGLYEKSKKFDGINAAYLALQSDNEDDRKKAKKIFQDEISSADIPDDV

KADFLKKIDDDQFLPIQRTKNNGTIPHQLHRNELEQIIEKQGIYYPFLKDTYQENSHEL
>mTALINFRVPYYVGPLVEEEQKIADDGKMPDPT ˙NHW]VrVRKSNDTITPWNLSQVV
DLDKSGRRFIERLTGTDTYLIGEPTLPKNSLLYQKFDVLQELNNIRVSGRRLDIRAKQ
DAFEHLFKVQKTVSATNLKDFLVQAGYISEDTQIEGLADVNGKNFNNALTTYNYLV
SVLGREFVENPSNEELLEEITELQTVFEDKKVLRRQLDQLDGLSDHNREKLSRKHYT
GWGRISKKLLTTKIVQNADKIDNQTFDVPRMNQSIIDTLYNTKMNLMEIINNAEDDF
GVRAWIDKQNTTDGDEQDVYSLIDELAGPKEIKRGIVQSFRILDDITKAVGYAPKRV
YLEFARKTQESHLTNSRKNQLSTLLKNAGLSELVTQVSQYDAAALQNDRLYLYFLQ
QGKDMY SGEKENLDNLSNYDIDKIPQ **AYTKDNSLDNRVL V** SNITNRRKSDSSNYLP
ALIDKMRPFWSVLSKQGLLSKHKFANLTRTRDFDDMEKERFIARSLVETRQIIKNVAS
LIDSHFGGETKAVAIRSSLTADMRRYVDIPKNRDINDYHHAFDALLFSTVGQYTENS
GLMKKGQLSDSAGNQYNRYIKEWIHAARLNAQSQRVNPFGFVVGSMRNAAPGKLN
PETGEITPEENADWSIADLDYLHKVMNFRKITVTRRLKDQKGQLYDESRYPSVLHDA
KSKASIWDKHKPVDLYGGFSSAKPAYAALIKFKNKFRLVNVLRQWTYSDKNSEDYI
LEQIRGKYPKAEMVLSH1PYGQLVKKDGALVTISSATELHNFEQLWLPLADYKLINTL
LKTKEDNLVDILHNRLDLPEMTIESAFYKAFDSILSFAFNRYALHQNALVKLQAHRD
DFNALNYEDKQQTLERJLDALHA SPASSDLKKINLS SGFGRLFSP SUFTL**ADTDEFIFQ**
SVTGLFSTQKTVAQLYQETK

[0071] ZP_08660870.1 possible CRISPR associated protein [Fructobacillus fructosus KCTC 3544]

MVYDVGLDIGTGSVGWVALDENGKLARAKGKNLVGVRLFDTAQTAADRRGFRTT
RRRLSRRKWPvLRLLDELFSAEINEIDSSFFQRLKYSYVHPKDEENKAHYYGGYLFPTE
EETKKFIIRSYPTIYHLRQELMAQP>nCRfDlREIYLAIFiiiLVKYRGHFLSSQEKITIGST
YNPEDLANAIEVYADEKGLSWELNNPEQLTEIISGEAGYGLNKSMKADEALKLFEFD
NNQDKVAIKTLLAGLTGNQIDFAKLFGKDISDKDEAKLWKLKLDDEALEEKSQTILS
QLTDEEIELFHAVVQAYDGFVLIGLLNGADSVSAAMVQLYDQHREDRKLLKSLAQK
AGLKHKRFSEIYEQLALATDEATIKNGISTARELVEESNLSKEVKEDTLRRLDENEFLP
KQRTKANSVIPHQLHLAELQKILQNQGQYYPFLLDTFEKEDGQDNKIEELLRFRIPYY
VGPLVTKKDVEHAGGDADNHWVERNEGFEKSRVTPWNFDKVFNRDKAARDFIERL
TGlvroTYLIGEKTLPQNSLRYQLFTVLNELNNVRVNGKKFDSKI ˈKADLINDLFKARKT
VSLSALKDYLKAQGKGDVTITGLADESKFNSSLSSYNDLKKTFDAEYLENEDNQETL
EKIIEIQTVFEDSKJASRELSKLPLDDDQVKKLSQTHYTGWGRLSEKLLDSKIIDERGQ

KVSILDKLKSTSQNFMSIINNDKYGVQAWITEQNTGSSKLTFDEKVNELTTSPANKRG
IKQSFAVLNDIKKAMKEEPRRVYLEFAREDQTSVRSVPRYNQLKEKYQSKSLSEEAK
VLKXTLDGNKNKMSDDRYFLYFQQQGKDMYTGRPINFERLSQDYDIDH1IPQAFTKD
DSLDNRVLVSRPENARKSDSFAYTDEVQKQDGSLWTSLLKSGFINRKKYERLTKAG
KYLDGQKTGFIARQLVETRQIIKNVASLIEGEYENSKAVAIRSEITADMRLLVGIKKH
mNSFHHAFDALLITAAGQYMQNRYPDRDSTNVYNEFDRYTNDYLKNLRQLSSRD
EVRRLKSFGFVVGTMRKGNEDWSEENTSYLRKVMMFKNILTTKKTEKDRGPLNKET
IFSPKSGKKLIPLNSKRSDTALYGGYSNVYSAYMTLVRANGKNLLIKIPISIANQIEVG
NLKIhTOYIVNWAIKKFEKILISKLPLGQLX^DG^IYLASNEYPvHNAKQLW LSTTD
ADKIASISENSSDEELLEAYDILTSENVKNRFPFFBCKDIDKLSQVRDEFLDSDKRIAVIQ
TILRGLQIDAAYQAPVKJISKKVSDWHKLQQSGGIKLSDNSEMIYQSATGIFETRVKJS
DLL

[0072] YP_001691366.1 hypothetical protein FMG_0058 [Finegoldia magna ATCC 29328]

MKSEKKYYIGLDVGTNSVGWAVTDEFYMLRAKGKDLWGVRLFEKADTAANTRIFR
SGRRRNDRKGMRLQILREIFEDEIKKVDKDFYDRLDESKFWAEDKKVSGKYSLFND
KNFSDKQYFEKFPTIFHLPJ<:YLMEEHGKVDIRYYFLAmQMMKRRGHFLIDGQISFIV
TDDKJLKEQLILLINDLLKIELEEELMDSIFEILADVNEKJITDKXNNLKELIKGQDFNK
QEGNILNSIFESIVTGKAKIKNIISDEDILEKIKEDNKEDFVLTGDSYEENLQYFEEVLQ
ENITLFNTLKSTYDFLILQSILKGKSTLSDAQVERYDEHKKDLEILKKVIKKYDEDGKL
FKQVFKEDNGNGYVSYIGYYLNKNKKITAKKKISNIEFTKYVKGILEKQCDCEDEDV
KYLLGKJEQENFLLKQISSrNSVIPHQIHLFELDKILENLAKNYPSFNNKKEEFTKIEKIR
KTFTFWPYYVGPL>TOYHKNNGGNAWIFR>¾GEKIRPWNFEKIVDLHKSEEEFIKJIM
LNQCTYLPEEWLPKSSILYSEYMVLNELNNLRINGKPLDTDVKLKLffiELFKKKTKV
TLKSIRDYMYR>^ADKEDFDNSEK>^EIASNMKSYIDFNMLEDKJ ?DVEMVEDLIE
KITIHTGNKKLLKKYIEETYPDLSSSQIQKIINLKYKDWGRLSRKLLDGIKGTKKETEK
TDTVINFLRNSSDM.MQIIGSQNYSFNEYIDKLRKKYIPQEISYEVVENLYVSPSVKKM
1WQVIRVTEEITKVMGYDPDKIFIEMAKSEEEKKTTISRKNKLLDLYKAIKKDERDSQ
YEKLLTGLNKLDDSDLRSRKLYLYYTQMGRDMYTGEKIDLDKLFDSTHYDKDHIIP
QSMKKDDSIINNLVLVNKNANQTTKGMYPVPSSIPvNNPKIYNYWKYLMEKEFISKE
KYNRLIRNTPLTNEELGGFINRQLVETRQSTKAIKELFEKFYQKSKIIPVKASLASDLR
KDlVmTLKSREVNDLHHAHDAFLNIVAGDVWNREFTSNPINYVKENREGDKVKYSLS

KDFTPJ'PJCSKGKVIWTPEKGRKLIVDTLNKPSVLISNESHVKKGELFNATIAGKKDY
KJCGKIYLPLKKDDRLQDVSKYGGYKAmGAFFFLVEHTKSKKRIRSIELFPLHLLSKF
YEDKNTVLDYAINVLQLQDPKIIIDKINYRTEIIIDNFSYLISTKSNDGSITVKPNEQMY
WRVDEISNLKJaENKYKKDAILTEEDRKIMESYIDKIYQQFKAGKYKNRRTTDTIIEK
YEIIDLDTLDNKQLYQLLVAFISLSYKTSNNAVDFTVIGLGTECGKPRITNLPDNTYLV
YKSITGIYEKRIRIK

[0073] ZP_07316256.1 CRISPR-associated protein, Csnl family [Veilionella atypica ACS-134-V-Coi7a]

METQTSNQLITSI1LKDYPKQDWVGLDIGTNSVGWAVTOTSYELLKFHSHKMWGSR
LFEEGESAVTRRGFRSMRRRLERRKLRLKLLEELFADAMAQVDSTFFIRLHESKYiiY
EDKTTGHSSKillLFIDEDYTDQDYFTEYPTIYHLRKDLMENGTDDIRKLFLAVHHILK
YRGNFLYEGATFNSNAFTFEDVLKQALVNITFNCFDTNSAISSISNILMESGKTKSDK
AKAIERLVDTYTVFDEVNTPDKPQKEQVKEDKKTLKAFANLVLGLSANLIDLFGSVE
DIDDDLKKLQIVGDTYDEKRDELAKVWGDEIFfflDDCKSVYDAIILMSIKEPGLTISQS
KVKAFDKFIK£DLVILKSLLKLDRNVYNEMFKSDKKGLFiNYVFIYIKQGRTEETSCSR
EDFYKYTKKIVEGLADSKDKEYILNEIELQTLLPLQRIKDNGVIPYQLHLEELKVILDK
CGPKFPFLHTVSDGFSVTEKLIKMLEFRIPYYVGPLNTHHNIDNGGFSWAVRKQAGR
VTPWWEEKJDREKSAAAFIKNLT>¾CTYLFGEDVLPKSSLLYSEFMLLNELNNYRID
GKALAQGVKQHLIDSIFKQDHKXMTKNRlELFLKDNWITKKHKPEITGLDGEIKND
LTSYRDMVRJLGNWDVSMAEDIITDITIFGESKKMLRQTLRNKFGSQLNDETIKKLS
KLRYRDWGRLSKJO.LKGIDGCDKAGNGAPKTIIELMRNDSYNLMEILGDKFSFMECI
EEENAK1AQGQVV1SIPFIDIIDELALSPAVKRAVWQALR1VDEVAHIKKALPSRIFVEV
ARTNKSEKICICKDSRQKRLSDLYSAIKKDDVLQSGI.QDKEFGALKSGLANYDDAALR
SKXLYLYYTQMGRCAYTGNIIDLNQLNTDNYDIDfflYPRSLTKDDSFDNLVLCERTA
NAKKSDIYPIDNRIQTKQKPFWAFLKHQGLISERKYERLTRIAPLTADDLSGFIARQLV
ETOQSVKATTTLLRRLYPDIDVVFVKAENYSDFRHNNWIKVRSLNHHHHAKDAYL
MVVG>TVYHEKFTRNFRLFFKKNGANRTYNLAKMFNYDVICTNAQDGKAWDVKTS
MNTVKi<MMASNDVRVTRRLLEQSGALADATIYKASVAAKAKDGAYIGMKTKYSV
FADVTKYGGMTKIKNAYSIIVQYTGKKGEEIKEIVPLPIYLINRNATDIELIDYVKSVIP
KAKDISIKYRKLCINQLVKWGFYYYLGGKTODKIYIDNAIELVVPHDIATYlKLLDK
YDLLRKENKTLKASSITTSIYNINTSTVVSLSNKVGIDVFDYFMSKLRTPLYMKMKGN

KVDELSSTGRSKFIKMTLEEQSIYLLEVLNLLTNSKTTFDVKPLGITGSRSTIGVKIHNL
DEFKIINESITGLYSNEVTIV

[0074] ZP_08029929.1 CRISPR-associated protein, Csnl family [Solobacterium moorei F0204]

MEGQMKNNGNNLQQGNYYLGLDVGTSSVGWAVTDTDYNVLKFRGKSMWGARLF
DEASTAEERRTHRGNRRRLARRKYRLLLLEQLFEKEIRKIDDNFFVRLHESNLWADD
KSKPSKPLLFM)TNFTDKDYLKKYPTIYHLRSDLIHNSTEHDIRLVFLALHHLIKYRG
fffIYDNSANGDVKTLDEAVSDFEEYLNENDIEFMENKKEFrNTVLSDKFILTKKEKKIS
LKKLYGDITDSENINISVLIEMLSGSSISLSNLFKDIEFDGKQNLSLDSDIEETLNDVVDI
LGDhnDLLIHAK£VYDIAVLTSSLGKFIKYLCDAKVELFEKNKKOLMILKXYIKKNHP
EDYKKIFSSPTEKKNYAAYSQTNSKNVCSQEEFCLFIKPYIRDMVKSENEDEVRIAKE
VEDKSFLTKLKGT^SVVPYQIHERELNQILKMVAYLPFMNDEQEDISVVDKIKLIFK
FKIPYYVGPLNTKSTRSWVYRSDEKIYP\\WS>rVlDLDKTAHEFMNRLIGRCTYTTW
PVLPMDSLLYSKYNVLNE1MPIKVNGKAIPVEVKQAIYTDLFENSKKKVTRKSIYIYLL
KNGYIEKEDIVSGIDIEIKSKLKSHHDFTQIVQENKCTPEEIERIIKGILVYSDDKSMLRR
WLKNMKGLSENDVKYLAKL>T*'KEWGRLSKTLLTDIYTINPEDGEACSILDIMWNTN
ATLMEILSNEKYQFKQhnENYKAENYDEKQNLHEELDDMYISPAARRSIWQALRJVD
EIVDIKKSAPKKJFIEMAREKXSAMKKKRTESRKDTLLELYKSCKSQADGFYDEELFE
KLSNESNSRLRRDQLYLYYTQMGRSMYTGKRIDFDKLINDKNTYDIDfflYPRSKIKD
DSITMlVLVEKDmGEKTDIYPISEDIRQKMQPFWKILKEKGLINEEKYKPvLTRNYELT
DEELSSFVARQLVETQQSTKALATLLKXEYPSAKJVYSKAGNVSEFRNRKDKELPKF
REINDLHHAKDAYLNIVVGNVYDTKFTEKFFNNIRNENYSLKRVFDFSVPGAWDAK
GSTFNTIKXYMAKNNPIIAFAPYEVKGELFDQQIVPKGKGQFPIKQGKDIEKYGGYNK
LSSAFLFAVEYKGKKARERSLETVYIKDVELYLQDPIKYCESVLGLKEPQIIKPKILMG
SLFSINMCKLVVTGRSGKQYVCHfflYQLSINDEDSQYLKNIAKYLQEEPDGNlERQNl
LMTSVNMKLFDVLCTKFNSNTYEIILNSLKNDVNEGREKFSELDILEQCNILLQLLKA
FKCNRESSNLEKLNNKKQAGVIVIPFILFTKCSVFKVIHQSITGLFEKEMDLLK

[0075] ZP_03989815.1 crispr-associated protein [Acidaminococcus sp. D21]

MGKMYYLGLDIGTNSVGYAVTDPSYHLLKFKGEPMWGAHVFAAGNQSAERRSFRT
SRRRLDRRQQRVKLVQEIFAPVISPIDPRFFIRLHESALWRDDVAE'I'**DKFflFFNDPTYT**
DKEYYSDYPTIHHLIVDLMESSEKFIDPRLVYLAVAWLVAHRGFIFLNEVDKDNIGDV
LSFDAFYPEFLAFLSDNGVSPWVCESKALQATLLSRNSVNDKYKALKSLIFGSQKPE

DNFDAMSEDGLIQLLAGKKVKVNKLFPQESNDASFTLNDKEDAIEEILGTLTPDECE
WIAfflPvPXFDWAIMKHALKDGRTlSESKVKLYEQHHHDLTQLKYFVKTYLAKEYDD
IFRNVDSETTKNYVAYSYIWKEVKGTLPKNKATQEEFCKYVLGKVKN1ECSEADKV
DFDEMIQRLTDNSFMPKQVSGENRVIPYQLYYYELKTILNKAASYLPFLTQCGKDAIS
NQDKLLSIMTFRIPYFVGPLRKDNSEHAWLERKAGKIYPWNFNDKVDLDKSEEAFIR
RMTNTCTYYPGEDVLPLDSLIYEKFMILNEINNIRIDGYPISVDVKQQVFGLFEKKRR
VTVKDIQNLLLSLGALDKHGKLTGIDTTIHSNYNTYHHFKSLMERGVLTRDDVERIV
ERMTYSDDTKRVRLWLNNNYGTLTADDVKHISRLRKHDFGRLSKMFLTGLKGVHK
ETGERASILDFMWNTODNLMQLLSECYTFSDEITKLQEAYYAKAQLSLNDFLDSMYI
SNAVKRPIYRTLAVVNDIRKACGTAPKRIFIEMARDGESKKKRSVTRREQIKNLYRSI
RKDFQQEVDFLEKILENKSDGQLQSDALYLYFAQLGRDMYTGDPIKLEHIKDQSFYN
IDfflYPQSMVKODSLDNKVLVQSEINGEKSSRYPLDAAIRNKMKPLWDAYYNHGLI
SLKKYQRLTRSTPFTDDEKWDFINRQLVETRQSTKALAILLKRKFPDTEIVYSKAGLS
SDFRHEFGLVKSRNTNDLHHAKDAFLAIVTGNVYFIERFNRRWFMVNQPYSVKTKTL
FTHSIKNGOTVAWNGEEDLGRIVKMLKQNKNTIHFTRFSFDRKEGLFDIQPLKASTGL
VPRKAGLDVVKYGGYDKSTAAYYLLVRFTLEDKKTQHKLMMIPVEGLYKARIDHD
KEFLTDYAQTTISEILQKDKQKVmiMFPMGTRHIKLNSMISIDGFYLSIGGKSSKGKS
VLCHAMVPLIVPHKIECYIKAMESFARKFKENNKLRIVEKFDKITVEDNLNLYELFLQ
KLQHNPYNKFFSTQFDVLTNGRSTFTKLSPEEQVQTLLNILSIFKTCRSSGCDLKSING
SAQAARIMISADLTGLSKKYSDIRLVEQSASGLFVSKSQNLLEYL

[0076] ZP_07455288.1 csnl family CRISPR-associated protein [Eubacterium yurii subsp. margaretiae ATCC 43715]

MENKQYYIGLDVGTNSVGWAVTDTSYNLLRAKGKDMWGARLFEKANTAAERRTK
RTSRRRSEREKARKAMLKELFADEINRVDPSFFIRLEESKFFLDDRSENNRQRYTLFN
DATFTDKDYYEKYKTIFHLRSALINSDEKFDVRLVFLAILNLFSHRGHFLNASLKGDG
DIQGMDVFYNDLVESCEYFEIELPRITNIDNFEKILSQKGKSRTKILEELSEELSISKKD
KSKYNLIKLISGLEASVVELYMEDIQDENKKIKJGFRESDYEESSLKVKEIIGDEYFDL
VERAKSVHDMGLLSMIGNSKYLCEARVEAYEhmHKDLLKIKELLKKYDKKAYNDM
FRKMTDKNYSAYVGSVNSNIAKERRSVDKRKIEDLYKYIEDTALKNIPDDNKDKIEIL
EKIKLGEFLKKQLTASNGVIPNQLQSRELRAILKKAENYLPFLKEKGEKNLTVSEMIIQ
LFEFQIPYYVGPLDKNPKKDNKANSWAKIKQGGRILPWNFEDKVDVKGSRKEFIEK
IVrVRKCTYISDEHTLPKQSLLYEKFMVLNEINNTKIDGEKISVEAKQKIYNDLFVKGKK

VSQKDIKKELISLNIMDKDSVLSGTDTVCNAYLSSIGKFTGVFKEEINKQSIVDMIEDII
FLKTVYGDEKRFVKEEIVEKYGDEIDKDKIKRILGFKFSNWGNLSKSFLELEGADVGT
GEVRSHQSLWETNFNLMELLSSRFTYMDELEKRVKiCLEKPLSEWTIEDLDDMYLSSP
VKRMIWQSMKIVDEIQTVIGYAPKRIFVEMTRSEGEKVRTKSRKDRLKELYNGIKED
SKQWVKELDSKDESYFRSKKMYLYYLQKGRClVrYSGEVIELDKLMDDNLYDIDfflYP
RSFVKDDSLDNLVLVKKErNNRKQNDPITPQIQASCQGFWKILHDQGFMSNEKYSRL
TRKTQEFSDEEKLSFINRQIVETGQATKCMAQILQKSMGEDVDVVFSKARLVSEFRH
KJELFKSRLINDFFfflANDAYLNlVVGNSYFVKFTRNPANFIKDARKNPDNPVYKYH
MDRFFERDVKSKSEVAWIGQSEGNSGT1VIVKKTMAKNSPLITKKVEEGHGSITKETI
VGVKEIKFGRNKVEKADKTPKKPNLQAYRPIKTSDERLCNILRYGGRTSISISGYCLV
EYVKKRKTIRSLEAIPVYLGRKDSLSEEKLLNYFRYNLNDGGKDSVSDIRLCLPFISTN
SLVKIDGYLYYLGGKNDDPJQLYTslAYQLKMKKEEVEYIRKIEKAVSMSKFDEIDREK
NPVLTEEKMELYNK1QDKFENTVFSKRMSLVKYNKKDLSFGDFLKNKKSKFEEIDLE
KQCKVLYMIFNLSNLKEVDLSDIGGSKSTGKCRCKKNITNYKEFKLIQQSITGLYSCE
KDLMTI

[0077] CBK78998.1 CRISPR-associated endonuclease, Csnl family [Coprococcus catus GD/7]

MKQEYFLGLDMGTGSLGWAVTDSTYQVMRKHGKALWGTRLFESASTAEERRMFR
TARRRLDRRNWRIQVLQEIFSEEISKVDPGFFLRMKESKYYPEDKRDAEGNCPELPY
ALFVDDWTDKWHKOYPTIYHLRKMLMETTEIPDIRLVYLVLHHMMKFIRGHFLLS
GDISQIKEFKSTFEQLIQNIQDEELEWfflSLDDAAIQFVEHVLKDRNLTRSTKKSRLIK
QLNAKSACEKAILNLLSGGTVKLSDIFNNKELDESERPKVSFADSGYDDYIGIVEAEL
AEQYYIIASAKAVYDWSVLVEILGNSVSISEAKIKVYQKHQADLKTLKKIVRQYMTK
EDYKRVFVDTEEKLNNYSAYIGMTKKNGKKVDLKSKQCTQADFYDFLKKNVIKVID
HKEITQEIESEIEKENFLPKQVTKDNGVIPYQVHDYELKXILDNLGTRMPFIKENAEKI
QQLFEFPJPYYVGPLNRVDDGKDGKFTWSVRXSDARJYPWNFTEVIDVEASAEKFIR
RMTNKCTYLVGEDVLPKDSLVYSKFMVLNELNI^RLNGEKISVELKQRIYEELF CKY
RKVTRKKLERYLVIEGIAKKGVEITGIDGDFKASLTAYHDFKERLTDVQLSQRAKEAI
VLNVVLFGDDKKLLKQRLSKMYPNLTTGQLKGICSLSYQGWGRLSKTFLEEITVPAP
GTGEVWMMTALWQTTvTDNLMQLLSRNYGFTNEVEEFNTLKKETDLSYKTVDELYV
SPAVKRQIWQTLKVVKEIQKVMGNAPKRVFVEMARKQEGKRSDSRKKQLVELYR
ACKNEERDWITELNAQSDQQLRSDKLFLYYIQKGRCMYSGETIQLDELWDNTKYDI

DHIYPQSKTMDDSLNTNRVLVKX>rWAIKSDTYPLSLDIQKKMMSFWKMLQQQGFIT
KEKYVRLVRSDELSADELAGFIERQIVETRQSTKAVATILKEALPDTEIVYVKAGNVS
WRQTYELLKVREMNDLHHAKDAYLNIVVGNAYFVKFTKNAAWFIRNNPGRSYNL
KEMFEFDIERSGEIAWKAGNKGSIVTVKKVMQKNMLVTRXAYEVKGGLFDQQIMK
KGKGQVPIKGMDEPvLADIEKYGGYNKAAGTYFMLVKSLDKKGKEIRTIEFVPLYLKN
QIEINHESAIQYLAQERGLNSPEILLSKIKIDTLFKVDGFKMWLSGRTGNQLIFKGANQ
LILSHQEAAILKGVVKWNRKNENKDAKLSERDGMTEEKLLQLYDTFLDKLSNTVY
SIRLSAQIKTLTEKRAKFIGLSNEDQCIVLNEILHMFQCQSGSANLKLIGGPGSAGILV
MNNNITACKQISVINQSPTGIYEKEIDLIKL

[0078] ZP_00143587.1 hypothetical protein [Fusobacterium nucleatum subsp. vincentii ATCC 49256]

MKKQKFSDYYLGFDIGTNSVGWCVTDLDYNVLRFNKKDMWGSRLFDEAKTAAER
RVQRNSRRRLKJIRKWRLNLLEEIFSDEIMKIDSNFFRRLKESSLWLEDKNSKEKFTLF
NDDNYKDYDFYKQYPTIFHLRDELIKNPEKKDIRLIYLALHSIFKSRGHFLFEGQNLK
EIKNFETLYNNLISFLEDNGINKSIDKDMEKLEKIICDSGKGLKDKEKEFKGIFNSDKQ
LVAIFKLSVGSSVSLNDLFDTDEYKKEEVEKEKISFREQIYEDDKPrYYSILGEKIELLD
IAKSFYDFMVLNNILSDSNYISEAKVKLYEEHKKDLKNLKYIIRKYNKENYDKLFKD
KNENWPAYIGLNKEKDKKEVVEKSRLKJDDLIKVIKGYLPKPERIEEKDKTIFNEILN
KIELKTILPKQRISDNGTLPYQIHEVELEKILENQSKYYDFLNYEENGVSTKDKLLKTF
KFPJPYYVGPLNSYHKDKGGNSWIVPvKEEGKILPWNFEQKVDIEKSAEEFIKRMTNK
CTYLNGEDVIPKDSFLYSEYIILNELNKVQVNDEFL>ffiENKPvKIIDELFKEM<XVSEKK
FKEYLLVNQIANRTVELKGIKDSFNSNYVSYIKFKDIFGEKLNLDIYKEISEKSILWKC
LYGDDKKIFEKKIKNEYGDILNKDEIKKINSFKFNTWGRLSEKLLTGIEFINLETGECY
SSVMEALRRTNYNLMELLSSKFTLQESIDNENKEMNEVSYRDLIEESYVSPSLKRAIL
QTLKJYEEIKKITGRVPKXVFIEMARGGDESMKNKKIPARQEQLKKLYDSCGNDIANF
SIDIKEMKNSLSSYDNNSLRQKXLYLYYLQFGKCMYTGREIDLDRLLQNNDTYDIDH
IYPRSKVIKDDSFDNLVLVLKNENAEKSNEYPVKKEIQEKMKSFWRFLKEKNFISDEK
YKRLTGKDDFELRGFMARQLVNVRQTTKEVGKILQQIEPEIKIVYSKAEIASSFREMF
DFIKVRELNDTHHAKDAYLNTVAG>WYNTKFTEKPYRYLQEIKENYDVKKIYNYDIK
NAWDKENSLEIVKKNMEKNT\O^TRFIKEEKGELFNLNPIKKGETSNEIISIKPKLYDG
KDNKLNEKYGYYTSLKAAYFIYVEHEKKNKKVKTFERITRIDSTLIKNEKNLIKYLVS
QKXLLNPKIIKKIYKEQTLIIDSYPYTFTGVDSNKKVELKNKKQLYLEKKYEQILKNA

LKFVEDNQGETEEWKFIYLKKRhWNEKNETIDAVKERYMEFNElVryTiKFLEKLSSK
DYKNYINNKLYTNFLNSKEKFKKLKLWEKSLILREFLKIFNKNTYGKYEIKDSQTKE
KLFSFPEDTGRIRLGQSSLGNNKELLEESVTGLFVKKIKL

[0079] YP_005054169.1 CRISPR-associated protein, Csnl family [Filifactor alocis ATCC 35896]

MTKEYYLGLDVGTNSVGWAVTDSQYNLCKFKKKDMWGIRLFESANTAKDRRLQR
GNRRRLERKKQRIDLLQEIFSPEICKIDPTFFIRLNESRLHLEDKSNDFKYPLFIEKDYS
DIEYYKEFPTIFHLRKHLIESEEKQDIRLIYLALHNIIKTRGHFLIDGDLQSAKQLRPILD
TFLLSLQEEQNLSVSLSENQKDEYEEILKNRSIAKSEKVKKLKNLFEISDELEKEEKKA
QSAVIENFCKFIVGNKGDVCKFLRVSKEELEIDSFSFSEGKYEDDIVKNLEEKVPEKV
YLFEQMKAMYDWNILVDILETEEYISFAKVKQYEKHKTNLRLLRDIILKYCTKDEYN
RMF>TOEKEAGSYTAYYGKLKKNNKKYWIEKKRNPEEFYKSLGKLLDKIEPLKEDLE
VLTMMIEECK>¾TLLPIQKNKDNGVIPHQVHEVELKKILENAKKYYSFLTETDKDGY
SWQKIESIFRFRIPYYVGPLSTRHQEKGSNVWMVRKPGREDRIYPWNMEEIIDFEKS
NENFITRMTmCTYLIGEDVLPKHSLLYSKYMVLNELNNVKVRGKKLPTSLKQKVFE
DLFENKSKVTGKNLLEYLQIQDKDIQIDDLSGFDKDFKTSLKSYLDFKKQIFGEEIEKE
SIQNMIEDHKWITIYGNDKEMLKRVIRANYSNQLTEEQMKKITGFQYSGWGNFSKMF
LKGISGSDVSTGETFDIITAMWETDNNLMQILSKKFTFMDNVEDFNSGKVGKIDKITY
DSTVKEMFLSPE>¾RAVWQTIQVAEEIKJCVMGCEPKKIFIEMARGGEKVKKRTKSR
KAQLLELYAACEEDCRELIKEIEDRDERDFNSMKLFLYYTQFGKCMYSGDDIDINELI
RGNSKWDRDffIYPQSKJKDDSIDNLVLVNKTYNAKKSl^LLSEDIQKKMHSFWLSLL
NKKLITKSKYDRLTRKGDFTDEELSGFIARQLVETRQSTKAIADIFKQIYS.SEVVYVKS
SLVSDFRKKPLNYLKSRRVNDYHHAKDAYLMVVG>^YNKKFTSNPIQWMKKNRD
TWSLNKVFEHDVVINGEVIWEKCTYHEDTNTYDGGTLDRIRKIVERDNILYTEYAY
CEKGELFNATIQNKNGNSTVSLKKGLDVKKYGGYFSANTSYFSLIEFEDKKGDRARH
IIGVPIYIANMLEHSPSAFLEYCEQKGYQNVRILVEKIKKNSLLIINGYPLRIRGENEVD
TSFKRAIQLKLDQKNYELVRMEKFLEKYVEKKGNYPIDENRDHITHEKMNQLYEVL
LSKMKKFNKKGMADPSDRIEKSKPKFIKLEDLIDK^^VINKMLNLLRCDNDTKADLS
LIELPKNAGSFVVKKNTIGKSKIILVNQSVTGLYENRREL

[0080] ZP_07398877.1 csnl family CRISPR-associated protein [Peptoniphilus duerdenii ATCC BAA-1640]

MKNLKEYYIGLDIGTASVGWAVTDESYMPKFNGKICMWGVPvLFDDAKTAEERRTQ
RGSPJIRLNRPvKERINLLQDLFATEISKVDPNFFLRLDNSDLYREDKDEKLKSKYTLFN
DKDFKI)RDYHKKYPTIIiHLIMDLIEDEGKKDIRLLYLACHYLLKNRGHFIFEGQKFD
TKNSFDKSI>TOLKJFILRDEYMDLEFWreDLIEIITDTTLNKTNKK^ ELKNIVGDTKFL
KAISAIMIGSSQKLVDLFEDGEFEETTVKSVDFSTTAFDDKYSEYEEALGDTISLLNIL
KSIYDSSILENLLKI)ADKSKDGNKYISKAFVKKFNKHGKDLKTLKRIIKKYLPSEYAN
IFRNKSrhTONYVAYTKSMTSNKRTKASKFTKQEDFYKFIKKFILDTIKETKLNSSENED
LKLIDEMLTDIEFKTFIPKLKSSDNGV1PYQLKLMELKKILDNQSKYYDFLNESDEYGT
VKOKVESIMEFRIPYYVGPLNPDSKYAWIKRENTKITPWWKDIVDLDSSREEFIDRLI
GRCTYLKEEKVLPKASLIYNEFMVLNELNNLKLNEFLITEEMKKAIFEELFKTKKKVT
LKAVSNLLKKEFNLTGDILLSGTDGDFKQGLNSYIDFKNIIGDKVDRDDYRIKIEEIIK
LIVLYEDDKTYLKXKJKSAYK>TOFTDDEIKXIAALNYKDWGRLSKRFLTGIEGVDKT
TGEKGSIIYFMREYNLNLMELMSGHYTFTEEVEKLNPVENRELCYEMVDELYLSPSV
KRMLWQSLRVVDEIKRIIGKDPKKIFIEMARAKEAKNSRXESRKNKLLEFYKFGKKA
FINEIGEERYNYLLNEINSEEESKFRWDNLYLYYTQLGRCMYSLEPIDLADLKSNMY
DQDfflYPKSKIYDDSLENRVLVKKNLNHEKGNQYPIPEKVLNKNAYGFWKILFDKGL
IGQKKYTRLTRRTPFEERELAEFIERQIVETRQATKETANLLKNICQDSEIVYSKAENA
SRFRQEFDIIKCRTVNDLHHMHDAYLMVVGNVYNTFCFTKNPLNFIKDKDNVRSYNL
ENMFKYDVVRGSYTAWIADDSEGNVKAATIKKVKRELEGKNYRFTRMSYIGTGGL
YDQNLMRKGKGQIPQKENTNKSNIEKYGGYNKASSAYFALIESDGKAGRERTLETIPI
MVYNQEKYGNTEAVDKYLKDNLELQDPKILKDKIKINSLIKLDGFLYNIKGKTGDSL
SIAGSVQLIVNKEEQKLIKKMDKFLVKXKDNKDIKVTSFDMKEEELIKLYKTLSDKL
NNGIYSNKRNNQAKMSEALDKFKEISIEEKJDVLNQHLLFQSYNNGCNLKSIGLSAKT
GVVFIPKKLNYKECKLINQSITGLFENEVDLLNL

[0081] NP_970941.1 CRISPR-associated Cas5e [Treponema denticola ATCC 35405]

MKKEIKDYFLGLDVGTGSVGWAVTDTDYKLLKANPvKDLWGMRCFETAETAEVRR
LHRGARRRIERRKKRIKLLQELFSQEIAKTDEGFFQRMKESPFYAEDKTILQENTLFN
DKDFADKTYHKAYPTmHLIKAWIENKVKPDPPJ ^LYLACHNlIKKRGFIFLFEGDFDSE
NQFDTSIQALFEYLREDMEVDIDADSQKVKEILKDSSLKNSEKQSRLNKILGLKPSDK

QKKAITNLISGNKINFADLYDNPDLKDAEKNSISFSKDDFDALSDDLASILGDSFELLL
KAKAVYNCSVLSKVIGDEQYLSFAKVKIYEKHKTDLTKLKNVIKKHFPKDYKKVFG
YMCNEKNMWYSGYVGVCKTKSKKLIINNSVNQEDFYKFLKTILSAKSEIKEVNDILT
EIETGTFLPKQISKSNAEIPYQLRKMELEKILSNAEKHFSFLKQKDEKGLSHSEKIIMLL
TFKIPYYIGPINDNHKKFFPDRCWVVKKEKSPSGKTTPWNFFDfflDKEKTAEAFITSR
TNFCTYLVGESVLPKSSLLYSEYTVLNErNNLQIIIDGKNICDIKLKQKIYEDLFKKYK
KITQKQISTFIKHEGICNKTDEVnLGIDKECTSSLKSYIELKNIFGKQVDEISTKNMLEEI
IRWATIYDEGEGKTILKTKIKAEYGKYCSDEQIKKILNLKFSGWGRLSRKFLETVTSE
MPGFSEPVMITAMRETQNNLMELLSSEFTFTENIKKINSGFEDAEKQFSYDGLVKPLF
LSPSVKKMLWQTLKLVKEISHITQAPPKKIFIEMAKGAELEPARTKTRLKILQDLYNN
CKNDADAFSSEIKDLSGKJE>¾DNLRLRSDKLYLYYTQLGKCMYCGKPIEIGHVFDTS
NYDIDHIYPQSKIKDDSISNRVLVCSSCNKNKEDKYPLKSEIQSKQRGFWNFLQRNNF
ISLEKLNRLTRATPISDDETAKFIARQLVETRQATKVAAKVLEKMFPETKIVYSKAET
VSMFRNKFDIVKCREINDFHHAHDAYLNIVVGNVYNTKFTNNPWNFIKEKRDNPKIA
DTYNYYKVFDYDVKRNMTAWEKGKTIITVKDMLKRNTPIYTRQAACKKGELFNQT
IMKKGLGQHPLKXEGPFSMSKYGGYNKVSAAYYTLIEYEEKGNKIRSLETIPLYLVK
DIQKDQDVLKSYLTDLLGKKEFKILVPKIKINSLLKINGFPCHITGKTNDSFLLRPAVQ
FCCSNNEVLYFKKILRFSEIRSQREKJGKTISPYEDLSFRSYIKENLWKKTKNDEIGEKE
FYDLLQKKNLEIYDMLLTKFnCDTIYKKRPNSATIDILVKGKEKFKSLIIENQFEVILEIL
KLFSATRNVSDLQHIGGSKYSGVAKIGNKISSLDNCILIYQSITGIFEKRIDLLKV

[0082] ZP_07912707.1 conserved hypothetical protein [Staphylococcus lugdunensis M23590]

MNQKEILGLDIGITSVGYGLIDYETKMIDAGVRLFPEANVENNEGRRSKRGSRRLKR
RRIHRLERVKKLLEDYNLLDQSQIPQSTNPYAIRVKGLSEALSKDELVIALLfflAKRRG
IHKJDVIDSNDDVGNELSTKEQLNKNSKLLKDKFVCQIQLERMNEGQVRGEKNRFKT
ADIIKEIIQLLNVQKNFHQLDENFINKYIELVEMRREYFEGPGKGSPYGWEGDPKAW
YETLMGHCTYFPDELRSVKYAYSADLFNALNDLNNLVIQRDGLSKLEYHEKYFniEN
VFKQKKKPTLKQIANEINVNPEDIKGYRITKSGKPQFTEFKLYHDLKSVLFDQSILENE
DVLDQIAEILTIYQDKDSIKSKLTELDILLNEEDKENIAQLTGYTGTHRLSLKCIRLVLE
EQWYSSRNQMEIFTHLNlKJKKlNXTAANKIPKAMIDEFILSPVVKRTFGQAINLINiai
EKYGVPEDIIIELARENNSKDKQKFINEMQKKNENTRKRINEIIGKYGNQNAKRLVEK
IRLFlDEQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSK

KSNLTPYQYFNSGKSKLSYNQFKQfflLNLSKSQDmSKKKKEYLLEERDINKFEVQKE
FINW^VDTRYATRELTNYLKAWSANNIVnWKVKTINGSFTDYLRKVWK^ KKERNH
GYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIESKQLDIQVDSEDNYSEMFIIP
KQVQDIKDFRWKYSHRVDKXPNRQLnNTOTLYSTRXKDNSTYIVQTIKDIYAKDNTT
LKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYANEKNPLAKYHEETGEYLTKYSKK
^GPIVKSLKYIGNKLGSHLDVTHQFKSSTKKLVKLSIKPYRFDVYLTDKGYKFITIS
YLDVLKXDNYYYIPEQKYDKLKLGKAIDKNAKFIASFYKNDLIKLDGEIYKIIGVNSD
TRNMIELDLPDIRYKEYCELNNIKGEPRIKKTIGKKVNSIEKLTTDVLGNVFTNTQYT
KPQLLFKRGN

[0083] ZP_02077990.1 hypothetical protein EUBDOL_01797 [Eubacterium dolichum DSM 3991]

MMEVFMGRLVLGLDIGITSVGFGIIDLDESEIVDYGVRLFKEGTAAENETRRTKRGGR
RLKRRRVTRREDMLHLLKQAGIISTSFHPLNNPYDVRVKGLNERLNGEELATALLHL
CKHRGSSVETIEDDEAKAKEAGETKKVLSMNDQLLKSGKYVCEIQKERLRTNGfflRG
HENNFKTPvAYVDEAFQILSHQDLSNELKSAHTnSRKRMYYDGPGGPLSPTPYGRYTY
FGQKEPIDLIEKMRGKCSLFPNEPRAPKLAYSAELF>n ,LNDLNNLSIEGEKLTSEQKA
MILKIVHEKGKITPKQLAKEVGVSLEQIRGFRIDTKGSPLLSELTGYKMIREVLEKSND
EHLEDHVFYDEIAEILTKTKDIEGRKKQISELSSDLNEESVHQLAGLTKFTAYHSLSFK
ALRLINEEMLKTELNQMQSITLFGLKQWffiLSVKGMKMQADDTAILSPVAKRAQRE
TFKVVNRLREIYGEFDSIVVEMAREKNSEEQRKAIRERQKFFEMRNKQVADIIGDDR
KINAKLREK1VLYQEQDGKTAYSLEPIDLKLLIDDPNAYEVDHIIPISISLDDSITNKVL
VTHPvENQEKGNLTPISAFVKGRFTKGSLAQYKAYCLKLKEKNIKTNKGYRKKVEQY
LL>ffiNDIYKYDIQKEFINRNLVDTSYASRVVL,NTLTTYFKQNEIPTKVFTVKGSLTNA
FRRKINLKKDRDEDYGHHAIDALIIASMPKMRLLSTIFSRYKIEDIYDESTGEVFSSGD
DSMYYDDRYFAFIASLKAIKVRKFSHKIDTKPNRSVADETIYSTRVIDGKEKVVKKY
KDIYDPKFTALAEDILNNAYQEKYLMALHDPQTFDQIVKVVNYYFEEMSKSEKYFT
KBKKGRIKISGMNPLSLYPvDEHGMLKKYSKKGDGPAITQMKYFDGVLGNHIDISAH
YQVRDKKVVLQQISPYRTDFYYSKENGYKFVTIRYKDVRWSEKKKKYVIDQQDYA
MKKAEKKIDDTYEFQFSMHRDELIGITKAEGEALIYPDETWFINFNFFFHAGETPEILK
FTATONDKSNKIEVKPIHCYCKMPvLMPTISKKIVPJDKYATDVVGNLYKVK^ TLKF
EFD

[0084] YP_820161.1 CRISPR-system-like protein [Streptococcus thermophilus LMD-9]

MSDLVLGLDIGIGSVGVGIL^VTGEHHKNSRIFPAAQAENNLVRRTORQGRPxLARR
KXHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHR
GISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEK
DGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNE
KSRTDYGRYRTSGETLDMFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVP
TETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEI
HTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQK
QVDELVQFRKANSSIFGKGWHWSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSS
NKTKYIDEKLLTEEIYNPVVAKSVRQAIKI\nSTAAIK£YGDFDNIVIEMARETNEDDEK
KAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGER
CLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQA
LDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYA
SRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAA
SSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFE
DSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFM
KJYKKDKSKFLMYRHDPQTFEK'VIEPILENYPNKQINEKGKEVPCWFLKYKEEHGYI
RKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTG
KYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKD
TETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGL
GKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF

[0085] EFT93846.1 CRISPR-associated protein, Csnl family [Enterococcus faecalis TX0012]

MYSIGLDLGISSVGWSVIDERTGNVIDLGVRLFSAKNSEKNLERRTNRGGRRLIRRKT
NRLKDAKKILAAVGFYEDKSLKNSCPYQLRVKGLTEPLSRGEIYKVTLHILKKRGISY
LDEVDTEAAKESQDYKEQVRKNAQLLTKYTPGQIQLQRLKENNRVKTGINAQGNYQ
LNVFKVSAYANELATILKTQQAFYPNELTDDWIALFVQPGIAEEAGLIYRKRPYYHG
PGNEANNSPYGRWSDFQKTGEPATOIFDKLIGKDFQGELRASGLSLSAQQYNLLNDL
TNLKIDGEVPLSSEQKEYILTELMTKEFTRFGVNDVVKLLGVKKERLSGWRLDKKGK
PEIHTLKGYRNWRKIFAEAGIDLATLPTETIDCLAKVLTLNTEREGIENTLAFELPELSE
SVKLLVLDRYKELSQSISTQSWHRFSLKTLHLLIPELMNATSEQNTLLEQFQLKSDVR

KilYSEYKKLPTKDVLAEIYNPTVNKTVSQAFKVIDALLVKYGKEQIRYITIEMPRDDN
EEDEKKRIKELHAKNSQRXhTOSQSYFMQKSGWSQEKFQTTIQKNRilFLAKLLYYYE
QDGICAYTGLPISPELLVSDSTEIDHIIPISISLDDSINNKVLVLSKANQVKGQQTPYDA
W1VTOGSFKXTNGKFSNWDDYQKWVESRHFSHKKENM.LETRNIFDSEQVEKFLARNL
M3TRYASRLVLNTLQSFFTNQETKVRVVNGSFTHTLRKKWGADLDKTRETFIHHHA
VDATLCAVTSFVKVSRYHYAVKEETGEKVMREIDFETGEIVNEMSYWEFKKSKKYE
RKTYQVKWPNFREQLKPX^FIPRIKFSHQVDRKANRKLSDATIYSVREKTEVKTLKS
GKQKITTDEYTIGKIKDIYTLDGWEAFKKKQDKLLMKDLDEKTYERLLSIAETTPDFQ
EVEEKNGKVKRVKRSPFAVYCEENDIPAIQKYAKKNNGPLIRSLKYYDGKLNKHINI
TKDSQGRPVEKTKNGRKVTLQSLKPYRYDIYQDLETKAYYTVQLYYSDLRFVEGKY
GITEKEYMKKVAEQTKGQVVRFCFSLQKNDGLEIEWKDSQRYDVRFYNFQSANSIN
FKGLEQEMMPAENQFKQKPYNNGAINLN1AKYGKEGKKLPJCFNTDILGKKHYLFYE
KEPKNIIK

[0086] YP_002937591.1 CRISPR-system related protein [Eubacterium rectale ATCC 33656]

MWTEKEKLFMKYILALDIGIASVGWAILDKESETVIEAGSNIFPEASAADNQLRRDM
RGAKR^NRRLKTRINDFIKLWENNNLSIPQFKSTEIVGLKVRAITEEITLDELYLILYSY
LKHRGISYLEDALDDTVSGSSAYANGLKLNAKELETHYPCEIQQERLNTIGKYRGQS
QIINENGEVLDLSNVFTIGAYRXEIQRVFEIQKXYHPELTDEFCDGYMLIFNRKRKYY
EGPGNEKSRTDYGRFTTKLDANGNYITEDNIFEKLIGKCSVYPDELRAAAASYTAQE
YTWLNDLNmTINGRKLEENEKHEIVERIKSSNTI>nVlRKIISDCMGEMDDFAGARIDK
SGK^IFHKFEVYNKMRKALLEIGIDISNYSREELDEIGYIMTINTDKEAMMEAFQKSW
IDLSDDVKQCLFNnviRKTOGALF^WQSFSLKIMNELIPEMYAQPK^ QMTLLTEMGV
TKGTQEEFAGLKYIPVDVVSEDIFNPVVRRSVRISFKILNAVLKKYKALDTIVIEMPRD
RNSEEQKKRINDSQKLNEKEMEYIEKXLAVTYGIKLSPSDFSSQKQLSLKLKLWNEQ
DGICLYSGKTIDPNDIINiSrPQLFEIDHIIPRSISFDDARSNKVLVYRSENQKKGNQTPYY
YLTHSHSEWSFEQYKATVMNLSKXKEYAISRKKIQNLLYSEDITKMDVLKGFINRNI
>TOTSYASRLVLNTIQNFFMANEADTKVKVIKGSYTHQMRCNLKLDKNRDESYSHFIA
VDAMLIGYSELGYEAYHKLQGEFIDFETGEILRKDMWDENMSDEVYADYLYGKKW
AMRNEVVKAEK>TVKYWHYVMRKS>¾GLCNQTIRGTREYDGKQYKINKLDIRTKE
GIKVFAKLAFSKKDSDRERLLVYLNDRRTFDDLCKIYEDYSDAANPFVQYEKETGDII
RKYSKKHNGPPJDKLKYKDGEVGACIDISFQCYGFEKGSKKVILESLVPYPxMDVYYKE

ENHSYYLVGVKQSDIKPEKGPJWIDEEAYAPJLVNEKMIQPGQSRADLENLGFKFKL
SFYKNDnEYEKDGKIYTERLVSRTMPKQRNYIETKPIDKAKFEKQNLVGLGKTKFIK
KYRYDILGNKYSCSEEKFTSFC

[0087] YP_015730.1 hypothetical protein MMOB0330 [Mycoplasma mobile 163K]

MYFYKNKENKLNKKVVLGLDLGIASVGWCLTDISQKEDNKFPIILHGVRLFETVDDS
DDKLLNETRRKKRGQRRRNRRLFTRKRDFIKYUDNNIIELEFDKNPKILVRNFIEKYI
NPFSKNLELKYKSVTNLPIGFHNLRKAAINEKYKLDKSELIVLLYFYLSLRGAFFDNP
EDTKSKEMNKNEIEIFDKNESIKNAEFPIDKIIEFYKISGKIRSTINLKFGHQDYLKEIKQ
VFEKQMDFMNYEKFAMEEKSFFSRIRNYSEGPGNEKSFSKYGLYANENGNPELIINE
KGQKJYTKIFKTLWESKIGKCSYDKKLYRAPKNSFSAKVFDITOKLTDWKHKNEYIS
ERLKRKILLSRFLNKDSKSAVEKILKEEWKFE^SEIAYNKDDhraM.PnNAYHSLTT
IFKKHLINFENYLISNENDLSKLMSFYKQQSEKLFVPNEKGSYEINQNNNVLHIFDAIS
MLNKFSTIQDRIRILEGYFEFSNLKKDVKSSEIYSEIAKLREFSGTSSLSFGAYYKFIPN
LISEGSKNYSTISYEEKALQNQKNWSHSNLFEKTWVEDLIASPTVKRSLRQTMNLLK
EIFKYSEK>WLEIEKIVVEVTRSSNNKFffiRKKIEGINKYRKEKYEELKKVYDLPNENT
TLLKKLWLLRQQQGYDAYSLRKIEA>TOVINKPWNYDIDHIVPRSISFDDSFSNLVIVN
KLDNAKXSNDLSAKQFIEKIYGIEKLKEAKENWGNWYLRNANGKAFNDKGKFIKLY
TIDNLDEFDNSDFINRNLSDTSYITNALVNHLTFSNSKYKYSVVSVNGKQTSNLRNQI
AFVGIKNNKETEREWKRPEGFKSINSNDFLIREEGKNDVKDDVLIKDRSFNGHHAED
AYFITIISQYFRSFKRIERLNVNYRKETRELDDLEKNNIKFKEKASFDNFLLINALDELN
EKLNQMRFSRMVITKKNTQLFNETLYSGKYDKGKNTIKKVEKLNLLDNRTDKIKKIE
EFFDEDKLKENELTKLHIFNHDKNLYETLKIIWNEVKIEIKNKNLNEKNYFKYFVNKK
LQEGKISFNEWVPILDNDFKIIRKIRYIKFSSEEKETDEIIFSQSNFLKIDQRQNFSFHNT
LYWVQIWVYKNQKDQYCFISIDARNSKFEKDEIKINYEKLKTQKEKLQIINEEPILKIN
KGDLFENEEKELFYIVGRDEKPQKLEIKYILGKKIKDQKQIQKPVKKYFPNWKKVNL
TYMGEIFKK

[0088] ZP_09312133.1 hypothetical protein MoviS_00710 [Mycoplasma ovipneumoniae SCOl]

MFINKXMTIGFDLGIASIGWAIIDSTTSKJLDWGTRTFEERXTANERRAFRSTRRMRR
KAYRNQRFINLILKYKDLFELKNTSDIQRANKKDTEN YEKIISFFTEIYKKCAAKHSNIL
EVKVKALDSKIEKLDLIWILHDYLENRGFFYDLEEENVADKYEG1EHPSILLYDFFKK
NGFFKSNSSIPKDLGGYSFSNLQWVNEIKJCLFEVQEINPEFSEKFLNLFTSVRDYAKGP

GSEHSASEYGIFQKDEKGKVFKKYDNIWDKTIGKCSFFVEENRSPVNYPSYEIFNLLN
QLINLSTDLKTTOKKIWQLSSNDP^ LLDELLKVKEKAKIISISLKKNEIKKIILKDFGF
EKSDIDDQDTIEGRKIIK.EEPTTKLEVTKHLLATIYSHSSDSNWINTNMLEFLPYLDAIC
IILDREKSRGQDEVLKKLTEKNIFEVLKIDREKQLDFVKSIFSNTKFNFKKIGNFSLKAI
REFLPKMFEQNKNSEYLKWKDEEIPvRKWEEQKSKLGKTDKKTKYLNPRIFQDEIISP
GTKNTFEQAVLVLNQIIKKYSKEMIDAIHESPPvEKNDKKTIEEIKKRNKKGKGKTLEK
LFQILNLENKGYKLSDLETKPAKLLDRLPvFYHQQDGIDLYTLDKIMDQLINGSQKYEI
EffllPYSMSYDNSQANKILTEKAENLKKGKLIASEYIKRNGDEFYNKYYEKAKELFIN
KYKKNKKLDSYVDLDEDSAKNRFRFLTLQDYDEFQVEFLARNLNDTRYSTKLFYHA
LVEHFE>^FFTYIDENSSKHKVKISTIKGHVTKYFRAKPVQKNNGPNENLNNNKPE
KlEKNRENNEHHAVDAAIVAnGNKNPQIANLLTLADNKTDKKFLLHDENYKENlETG
ELVKIPKFEVDKLAKVEDLKKIIQEKYEEAKKHTAIKFSRKTRTILNGGLSDETLYGF
KYDEKEDKYFKIIKKJO.VTSKNEELKKYFENPFGKKADGKSEYTVLMAQSHLSEFNK
LKEIFEKYNGFSNKTGNAFVEYMNDLALKEPTLKAEIESAKSVEKLLYYNFKPSDQF
TYHDNTNNKSFKPJ^YKMRIIEYKSIPIKFKILSKHDGGKSFKDTLFSLYSLVY^ VYEN
GKESYKSIPVTSQMRNFGIDEFDFLDENLYNKEKLDIYKSDFAKPIPVNCKPVFVLKK
GSILKKKSLDIDDFKETKETEEGhTA'FISTISKRFNRDTAYGLKPLKLSVVKPVAEPST
NPIFKEYIPIHLDELGNEYPVKIKEHTDDEKLMCTIK

[0089] ADC31648.1 Csnl family CRISPR-associated protein [Mycoplasma gallisepticum str. F]

MNNS1KSKPEVTIGLDLGVGSVGWAIVDNETMIHHLGSRLFSQAKTAEDRRSFRGVR
mRRRKYKLKRFVNLIWKYNSYFGFKNKEDILNNYQEQQKLHNTVLNLKSEALNA
KIDPKALSWILHDYLKNRGHFYEDNRDFNVYPTKELAKYFDKYGYYKGIIDSKEDN
DNKLEEELTKYXFSNKHWLEEVKXVLSNQTGLPEKFKEEYESLFSYVRNYSEGPGSI
NSVSPYGIYHLDEKEGKVVQKYNNIWDKTIGKCNIFPDEYRAPKNSPIAMIFNEINELS
TIRSYSIYLTGWFINQEFKKAYLNKLLDLLIKTNGEKPIDARQFKKLREETIAESIGKET
LKDVENEEKLEiCEDHKWKLKGLKLNTNGKIQYNDLSSLAKFVHKLKQHLKLDFLLE
DQYATLDKINFLQSLFVYLGKHLRYSNRVDSANLKEFSDSNKLFER1LQKQKDGLFK
LFEQTDKDDEKILAQTHSLSTKAMLLAITPvMTNLDNDEDNQKNNDKG\VNFEAIKNF
DQKFIDITKKNWn_SLKQNKRYLDDRFIND A1LSPG VKRILREATKVFNAILKQFSEEY
DVTKVVIELARELSEEKELENTKNYKKLIKKNGDKISEGLKALGISEDEIKDILKSPTK
SYKFLLWLQQDfflDPYSLKEIAFDDIFTKTEKFEIDHIIPYSISFDDSSSNKLLVLAESNQ

AKSNQTPYEFISSGNAGIKWEDYEAYCRKFKDGDSSLLDSTQRSKKFAKMMKTDTS
SKYDIGFLARNLNDTRYATIVFRDALEDYANNHLVEDKPMFKVVCINGSVTSFLRKN
FDDSSYAKKDRDKNIHHAVDASIISIFSNETKTLFNQLTQFADYKLFKNTDGSWKKID
PKTGVVTEVTDENWKQIRVRNQVSEIAKVIEKYIQDSNIERKARYSRKIENKTMSLFN
DTVYSAKKVGYEDQIKJRKNLKTLDIFIESAKENKNSKVKRQFVYRKLVNVSLLNNDK
LADLFAEKEDILMYRANPWVI^AEQIFNEYTENKKIKSQNVFEKYMLDLTKEFPEK
FSEFLVKSJV^RNKTAnYDDKKNIVHRIKRLKMLSSELKENKLSNVIIRSKNQSGTKLS
YQDTINSLALJVDMRSIDPTAKKQYIRVPLNTLNLHLGDHDFDLHNMDAYLKKPKFVK
YLKA>ffiIGDEYXPWRVLTSGTLLIFIKKDKKLMYISSFQNLNDVIEIKNLIETE YKEND
DSDSKXKKKANRf LMTLSTILNDYILLDAKDNFDILGLSKNRIDEILNSKLGLDKJV K

[0090] YP_278700 .1 hypothetical protein MS53_0582 [Mycoplasma synoviae 53]

MLRLYCANNLVL>mVQNLWKYLLLLIFDKKnFLFKIKVILIRRYMENWJKEKIVIGF
DLGVASVGWSIVNAETKEVIDLGVRLFSEPEKADYRRAKRTTRRLLRRKKFKREKFH
KLILKNAEIFGLQSRNEILNVYKDQSSKYRN1LKLKINALKEEIKPSELVWILRDYLQN
RGYFYKNEKLTDEFVSNSFPSKKiHEHYEKYGFFRGSVKLDNKLDNKKDKAKEKDE
EEESDAKKESEELIFSNKQWINEIVKVFENQSYLTESFKEEYLKLFNYVRPFNKGPGS
KNSRTAYGVFSTDIDPETOKFKDYSN1WDKTIGKCSLFEEEIRAPKNLPSALIFNLQNEI
CTIKNEFTEFKNWWLNAEQKSEILKFVFTELFNWKDKKYSDKXFNKNLQDKIKKYL
LNFALENFNLNEEILKNRDLENDTVLGLKGVKYYEKSNATADAALEFSSLKPLYVFI
KF·LKEKXLDL>TYLLGLENTEILYFLDSIYLAISYSSDLKERNEWFKKLLKELYPKIKN
NNLEIIENVEDIFEITDQEKFESFSKTHSLSREAFNHIIPLLLSNNEGKNYESLKHSNEEL
KKRTEKAELKAQQNQKYLKDNFLKEALVPLSVKTSVLQAIKIFNQIIKNFGKKYEISQ
VVIEMARELTKPM.EKLLNNATNSMKTTXEKLDQTEKFDDFTKKKFIDKIENSVVFR
NKLFLWFEQDPJCDPYTQLDIK1NEIEDETEIDHVIPYSKSADDSWFNKLLVKKSTNQL
KKNKTVWEYYQNESDPEAKWNKFVAWAKRIYLVQKSDKESKDNSEKNSIFKNKKP
NLKFKNITKKLFDPYKI)LGFLAR>n.NDTRYATKVFRDQLNNYSKHHSKDDENKLFK
VVCMNGSITSFLRKSMWRKNEEQVYRFNFWKKDRDQFFHHAVDASIIAIFSLLTKTL
YNKLRVYESYDVQRREDGVYLINKETGEVKKADKDYWKDQHNFLKIRENAIEIKNY
LNNVDFQNQVRYSRKANTKLNTQLFNETLYGVKEFENNFYKLEKVNLFSRKDLRKF
ILEDLNEESEKNKKNENGSRKRILTEKYIVDEILQILENEEFKDSKSDINALNKYMDSL
PSKFSEFFSQDFI>¾CKKENSLILTFDAIKIINroPKKVIKIKNLKFFREDATLKNKQAVH
KDSKNQIKSFYESYKCVGFIWLKNKhTOLEESIFVPINSRVIHFGDKDKDIFDFDSYNKE

KLLNEINLKRPENKKFNSINEIEFVKFVKPGALLLNFENQQIYYISTLESSSLRAKIKLL
NKMDKGKAVSMKKITNPDEYKIIEHVNPLGINLNWTKKLENNN

[0091] EIE39736.1 Csnl family CRISPR-associated protein [Mycoplasma canis PG 14]

MEKKPJCVTLGFDLGIASVGWAIVDSETNQVYKLGSRLFDAPDTOLERRTQRGTRRL
LRPvRKYRNQKPYNLVKRTEVFGLSSREAIENRFRELSIKYPNIIELKTKALSQEVCPDE
IAWILHDYLKNRGYFYDEKETKEDFDQQTVESMPSYKLNEFYKKYGYFKGALSQPT
ESENlKDNKDLKEAFFFDFSNKEWLKEINYFFm^QKMLSETFIEEFKJaFSFTRDISKG
PGSDNMPSPYGIFGEFGDNGQGGRYEHIWDKNIGKCSIFTNEQRAPKYLPSAEIFNFL
NELAMPvLYSTDKKMQPLWKLSSVDKLMLLNLFM ,PISEKKKKLTSTNTNDIVKKESI
KSIMISVEDIDNDKDEWAGKEPNVYGVGLSGLNIEESAKENKFKFQDLKILhTVLINLL
D>WGIKFEFKDR>ronKNLELLD NI.YLFLIYQKESNNKJDSSIDLFIAKNESLMENLKLK
LKEFLLGAGNEFENHNSKTHSLSKKAIDEILPKLLD>^GWNLEAIKNYDEEIKSQIE
DNSSLMAKQDKKYLNDNFLKDAILPPNVKVTFQQAILIFNKIIQKFSKDFEIDKVVIEL
AREMTQDQENDALKGIAKAQKSKXSLVEERLEANMDKSVFNDKYEKLIYKIFLWIS
QDFKDPYTGAQISVNEIVNNKVEIDHIIPYSLCFDDSSANXVLVHKQSNQEKSNSLPY
EYIKQGHSGWNWDEFTKYVKRVFVM^DSILSKKERLKKSENLLTASYDGYDKLGF
LARNLNDTRYATILFRDQLNWAEHHLIDNKKMFKVIAMNGAVTSFIRKNMSYDNK
LRLKDRSDFSHHAYDAAIIALFSNKTKTLYNLIDPSLNGIISKRSEGYWVIEDRYTGEI
KELKKEDWTSIKNNVQARKIAKEIEEYLIDLDDEVFFSRKTKRKTNRQLYNETIYGIA
TKTDEDGITNYYKKEKFSILDDKDIYLRLLREREKFVINQSNPEVIDQIIEIIESYGKEN
hnPSRDEAIMKYTKNlQNYNLYLKQYMRSLTKSLDQFSEEFINQMIANKTFVLYNPT
KNTTRKIKFLRLVNDViCINDIRKNQVTNKFNGKNNEPKAFYENINSLGAIVFKNSANN
FKTLSINTQIAIFGDKNWDIEDFKTYNMEKIEKYKEIYGIDKTYNFHSFIFPGTILLDKQ
N¾EFYYISSIQTVRDn EIKFLKKJEFKI)ENKNQDTSKTPKRLMFGIKSIMN^ YEQVDIS
PFGINKKIFE

[0092] NP_907605.1 hypothetical protein WS1445 [Wolinella succinogenes DSM 1740]

MIERILGVDLGISSLGWAIVEYDKDDEAANPJIDCGVRLFTAAETPKKKESPNKARRE
ARGIRRVLNRRRVRMNMIKXLFLRAGLIQDVDLDGEGGMFYSKANRADVWELRHD
GLYRLLKGDELARVLIffl AKHRGYKFIGDDEADEESGKVKKAGVVLRQNFEAAGCR
TVGEWLWRERGANGKKRNKHGDYEISIHRDLLVEEVEAIFVAQQEMRSTIATDALK
AAYREIAFFVRPMQRIEKMVGHCTYFPEERRAPKSAPTAEKFIAISKFFSTVIIDNEGW

EQKIIERKTLEELLDFAVSREKVEFPvHLRKFLDLSDNEIFKGLHYKGKPKTAKKREAT
LFDPNEPTELEFDKVEAEKKAWISLRGAAKLPvEALGNEFYGRFVALGKHADEATKIL
TYYKDEGQKJIRELTKLPLEAEMVERLVKIGFSDFLKLSLKAIRDILPAMESGARYDE
AVLMLGVPHKEKSAILPPLNKTDIDILNPTVIRAFAQFRKVANALVRKYGAFDRVHF
ELARErNTKGEIEDIKESQRKNEKERKEAADWIAETSFQVPLTRKNlLKKRLYIQQDG
RCAYTGDVIELERLFDEGYCEIDHILPRSRSADDSFANKVLCLARANQQKTDRTPYE
WFGHDAARWNAFETRTSAPSNRVRTGKGKIDRLLKKNFDENSEMAFKDRNLNDTR
YMARAIKTYCEQYWVFKNSHTKAPVQVRSGKLTSVLRYQWGLESKDRESHTHHAV
DAHIAFSTQGIVrVQKLSEYYRFKETHREKERPKLAVPLANFRDAVEEATRIENTETVK
EGVEVKRLLISRPPRARVTGQAHEQTAKPYPRIKQVKNKKKWRLAPIDEEKFESFKA
DRVASANQKOTYETSTIPRVDVYHKKGKFHLVPIYLHEMVLNELPNLSLGTNPEAM
DENFFKFSIFKDDLISIQTQGTPKKPAKIIMGYFKNMHGANMVLSSINNSPCEGFTCTP
VSMDKKHKDKCKLCPEENRIAGRCLQGFLDYWSQEGLRPPRKEFECDQGVKFALDV
KKYQIDPLGYYYEVKQEKRLGTIPQMRSAKKLVKK

[0093] YP_002344900.1 CRISPR-associated protein [Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819]

MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRL
ARRJCARLhfflLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRALNELLSK
QDFARVILfflAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQ
KFKENSKEFTNVRNKKESYERCIAQSFLKOELmFKKQREFGFSFSKKFEEEVLSVAF
YKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKD
DLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGEHN
LSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDIILNISFKALKLVT
PLMLEGKKYDEACNEL^KVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRK
VLNALLKKYGKVFnaMELAREVGKNHSQRAKJEKEQNENYKAKKDAELECEKLGL
KWSKMLKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDfflYPYSRSFDDSYMNKVL
VFTKQNQEKLNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQK
NFKDR^>TOTRYIARLVLNYTKI)YLDFLPLSDDENTKX>roTQKGSKVFrVEAKSGML
TSALP^TWGFSAKDRNNHLHHAIDAVnAYANNSIVKAFSDFKKEQESNSAELYAKK
ISELDYKNKRKJFEPFSGFRQKVLDKIDEIFVSKPEPJCKPSGALHEEIFRKEEEFYQSY
GGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKFIKKTNKFYAVPIYTMDFALKV
LPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSST

VSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAE
FRQREDFKK

[0094] YP_003516037.1 CRISPR associated protein [Helicobacter mustelae 12198]

MRTLGIDIGIASIGWAVIEGEYTDKGLENKEIVASGVRVFTKAENPKNKESLALPRTL
ARSARRRNARKKGRIQQVKHYLSKALGLDLECFVQGEKLATLFQTSKDFLSPWELR
ERALYRVLDKEELARVILHIAKRRGYDDITYGVEDNDSGKIKKAIAENSKRIKEEQCK
TIGEMMYKLYFQKSLNVRNKKESYNRCVGRSELREELKTIFQIQQELKSPWVNEELI
YKLLGNPDAQSKQEREGLIFYQRPLKGFGDKIGKCSHIKKGENSPYRACKHAPSAEE
FVALTKSINFLKNLTNRHGLCFSQEDMCVYLGKILQEAQKNEKGLTYSKLKLLLDLP
SDFEFLGLDYSGKNPEKAVFLSLPSTFKLNKITQDRKTQDKIANILGANKDWEAILKE
LESLQLSKEQIQTIKDAKLNFSKHINLSLEALYHLLPLMREGKRYDEGVEILQERGIFS
KPQPKNRQLLPPLSELAKEESYFDIPNPVLRRALSEFRKVVNALLEKYGGFHYFHIEL
TRDVCKAKSARMQLEKINKKNKSENDAASQLLEVLGLPNTYNNRLKCKLWKQQEE
YCLYSGEKITIDHLKDQRALQIDHAFPLSRSLDDSQSNKVLCLTSSNQEKSNKTPYEW
LGSDEKXWDIVTYVGRVYSSNFSPSKKRKLTQKWK^RNEEDFLARNLVDTGYIGRVT
KEYIKHSLSFLPLPDGKXEfflRIISGSMTSTMRSFWGVQEKNRDHHLHHAQDAIIIACI
EPSMIQKYTTYLKI)KETHRLKSHQKAQILREGDFIKLSLRWPMSNFKDKIQESIQMIP
SHHVSHKVTGELHQETVRTKEFYYQAFGGEEGVKKALKFGKIREINQGIVDNGAMV
RVDIFKSKDKGKFYAVPIYTYDFAIGKiPNKAIVQGKKNGIIKDWLENlDENYEFCFSL
FKNDCIKIQTKEMQEAVLAIYKSTNSAKATIELEHLSKYALKNEDEEKMFTDTDKEK
NKTMTRESCGIQGLKVFQKVKLSVLGEVLEHKPRNRQNIALKTTPKHV

[0095] ZP_06887976.1 CRISPR-associated protein, Csnl family [Methylosinus trichosporium OB3b]

MRVLGLDAGIASLGWALIEIEESNRGELSQGTIIGAGTWMFDAPEEKTQAGAKLKSE
QRRTFRGQRRVVRRRRQRMNEVRRILHSHGLLPSSDRDALKQPGLDPWRIRAEALD
RLLGPVELAVALGHIARHRGFKSNSKGAKT>roPADDTSKMKRAVNETREKLARFGS
AAKMLVEDESFVLRQTPTKNGASEIVRRFRNREGDYSRSLLRDDLAAEMRALFTAQ
ARFQSAIATADLQTAFTKAAFFQRPLQDSEKLVGPCPFEVDEKRAPKRGYSFELFRFL
SRLNHVTLRDGKQERTLTRDELALAAADFGAAAKVSFTALRKKLKLPETTVFVGVK
ADEESKLDVVARSGKAAEGTARLRSVIVDALGELAWGALLCSPEKLDKIAEVISFRS
DIGRISEGLAQAGCNAPLVDALTAAASDGRFDPFTGAGHISSKAARNILSGLRQGMT
YDKACCAADYDHTASRERGAFDVGGHGREALKRILQEERISRELVGSPTARKALIESI

KQVKAIVERYGVPDRIHVELARDVGKSIEEREEITRGIEKRNRQKDKLRGLFEKEVGR
PPQDGARGKEELLRFELWSEQMGRCLYTDDY1SPSQLVATDDAVQVDHILPWSRFA
DDSYANKTLCMAKANQDKKGRTPYEWFKAEKTDTEWDAFIVRVEALADMKGFKK
R>rm.RNAEEAAAKFRNRM.NDTRWACRLLAEALKQLYPKGEKDKDGKERRRVFS
RPGALTDRLRRAWGLQWMKKSTKGDRIPDDRHHALDAIVIAATTESLLQRATREVQ
EIEDKGLHYDLVKNVTPPWPGFREQAVEAVEKVFVARAERRRARGKAHDATIRHIA
VREGEQRVYERRKVAELKLADLDRVKDAERNARLIEKLRNWIEAGSPKDDPPLSPK
GDPIFKVRLVTKSKVMALDTGNPKRPGTVDRGEMARVDVFRKASKKGKYEYYLVP
IYPHDIATMKTPPIRAVQAYKPEDEWPEMDSSYEFCWSLVPMTYLQVISSKGEIFEGY
YRGMNRSVGAIQLSAHSNSSDVVQGIGARTLTEFKKFNVDRFGRXHEVERELRTWR
GETWRGKAYI

[0096] YP_003968716.1 CRISPR-associated protein, Csnl family (plasmid) [Ilyobacter polytropus DSM 2926]

IVKYSIGLDIGIASVGWSVINKDKERIEDMGVWFQKAENPKDGSSLASSPJIEKRGSRR
RMIRKKHRLDRIKMLCESGLVKKNEIEKIYKNAYLKSPWELRAKSLEAKISNKEIAQI
LLfflAKRRGFKSFRKTDRNADDTGKLLSGIQENKKIMEEKGYLTIGDMVAKDPKFNT
HVRNKAGSYLFSFSRKLLEDEVRKIQAKQKELGNTHFTDDVLEKYIEVFNSQRNFDE
GPSKPSPYYSEIGQIAKMIGNCTFESSEKRTAKNTWSGERFVFLQKLNNFRIVGLSGK
RPLTEEEPJDIVEKEVYLKKEVRYEKLRKILYLKEEERFGDLNYSKDEKQDKKTEKTK
FISLIGWTIKKLNLSEKLKSEIEEDKSKLDKIIEILTFNKSDKTIESNLKKLELSREDIEIL
LSEEFSGTLNLSLKAIKKILPYLEKGLSYNEACEKADYDYKNNGIKFKRGELLPVVDK
DLIANPVVLRAISQTRKVVNAIIRKYGTPHTIHVEVARDLAKSYDDRQTIIKENKKRE
LENEKTKKFISEEFGIKNVKGKLLLKYRLYQEQEGRCAYSRKELSLSEVILDESMTDI
DfflPYSRSMDDSYSNKVLVLSGENRKKSNLLPKEYFDRQGRDWDTFVLNVKAMKI
HPRKKSNLLKEKFTREDNKDWKSRALNDTRY1SRFVANYLENALEYRDDSPKKRVF
MIPGQLTAQLRARWRLNKVRENGDLHHALDAAVVAVTDQKAINMSNISRYKELKN
CKDVIPSIEYHADEETGEVYFEEVKDTRFPMPWSGFDLELQKRLESENPREEFYNLLS
DKRYLGWFWEEGFIEKLRPVFVSRMPNRGVKGQAHQETIRSSKKISNQIAVSKKPL
NSIKLKDLEKMQGRDTDRKLYEALKNRLEEYDDKPEKAFAEPFYKPTNSGKRGPLV
RGIKVEEKQNVGVYVNGGQASNGSMVRIDVFRKNGKFYTVPIYVHQTLLKELPNRA
INGKPYKDWDLIDGSFEFLYSFYPNDLIEIEFGKSKSIKNDNKLTKTEIPEVNLSEVLG
YYRGMDTSTGAATIDTQDGKIQMRIGIKTVKNIKKYQVDVLGNVYKVKREKRQTF

[0097] ZP_09352959.1 CRISPR-associated protein cas9/csnl, subtype II/nmemi [Bacillus smithii 7_3_47FAA]

MNYKMGLDIGIASVGWAVINLDLKPvIEDLGVRIFDKAEHPQNGESLALPRRIARSAR
RRLRRRKHRLERIRRLLVSENVLTKEEMNLLFKQKKQIDVWQLRVDALERKLNNDE
LARVLLHLAKRRGFKSNRKSERNSKESSEFLKN1EENQSILAQYRSVGEMIVKDSKFA
YHKRh^DSYSNMIARDDLEREIKLIFEKQREFNNPVCTERLEEKYLNIWSSQRPFAS
KEDIEKKVGFCTFEPKEKRAPKATYTFQ**SFIVWEHΓNKLRLV**SPDETRALTEIERNLL**Y**
KQAFSKNKMTYYDIRKLLNLSDDIHFKGLLYDPKSSLKQIENIRFLELDSYHKIRKCIE
NVYGKDGIRMFNETDIDTFGYALTIFKDDEDIVAYLQNEYITKNGKRVSNLANKVYD
KSLIDELLNLSFSKFAHLSMKAIRNILPYMEQGEIYSKACELAGYNFTGPKKKEKALL
LPVIPMANPVVMRALTQSRXVVNAnKKYGSPVSIfflELARDLSHSFDERKKIQKDQT
ENRKKNETAIKQLIEYELTKNPTGLDIVKFKLWSEQQGRCMYSLKPIELERLLEPGYV
EVDHILPYSRSLDDSYANKVLVLTKENREKGNHTPVEYLGLGSERWKKFEKFVLAN
KQFSKKKKQNLLRLRYEETEEKEFKERNLNDTRYTSKFFANFIKEHLKFADGDGGQK
VYTINGKITAHLRSRWDFNKNREESDLHHAVDAVIVACATQGMIKKITEFYKAREQN
KESAKKKEPIFPQPWPHFADELKARLSKFPQESIEAFALGNYDRKKLESLRPVFVSRM
PKRSVTGAAHQETLRRCVGIDEQSGKIQTAVKTKLSDIKLDKDGFIFPMYQKESDPRT
YEAIRQRLLEHN^JDPKKAFQEPLYKPKKNGEPGPVIRTVKIIDTKNKVVHLDGSKTV
AYNSNWRTDVFEKDGKYYCVPVYTMDIMKGTLPNKAIEANKPYSEWKEMTEEYTF
QFSLFPNDLVRIVLPREKTIKTSTNEEIIIKDIFAYYKTIDSATGGLELISHDRNFSLRGV
GSKTLKRFEKYQVDVLGNIUKVKGEKRVGLAAPTNQKKGKTVDSLQSVSD

[0098] YP_002507391.1 CRISPR-associated protein, Csnl family [Clostridium cellulolyticum H10]

MKYTLGLDVGIASVGWAVIDKDNNKIIDLGVRCFDKAEESKTGESLATARRIARGM
RRRISRRSQRLRLVKKLFVQYEIIKDSSEFNRIFDTSRDGWKDPWELRYNALSRILKPY
ELVQVLTHITKRRGFKSNRKEDLSTTKEGVVITSIKNNSEMLRTKNYRTIGEMIFMET
PENSNKRNKVDEYIHTIAREDLLNEIKYIFSIQRKLGSPFVTEKLEHDFLNIWEFQRPFA
SGDSILSKVGKCTLLKEELRAPTSCYTSEYFGLLQSINNLVLVEDNNTLTLNNDQRAK
nEYAHFKNEIKYSEIRKLLDIEPEILFKAFINLTHKNPSGNNESKKFYEMKSYHKLKST
LPiDiWGKLHS>^SLDNLFYCLTVYKNDNEIKDYLQA>WLDYLIEYIAKLPTFNKF
KHLSLVAMKJUIPFMEKGYKYSDACNMAELDFTGSSKLEKCh^TVEPnE>TVTNPV
VIRALTQARKVINAIIQKYGLPYIVm^ELAREAGMTRQDRDNLKKEHENNRKAREKI

SDLIRQNGRVASGLDILKWRLWEDQGGRCAYSGKPIPVCDLLNDSLTQIDHIYPYSRS
MDDSYMNKVLVLTDENQNKRSYTPYEVWGSTEKWEDFEARIYSMHLPQSKEKRLL
NRNFITKDLDSFISRNLNDTRYISRFLKNYIESYLQFSNDSPKSCWCVNGQCTAQLRS
RWGLNKNREESDLHHALDAAVIACADRKIIKEITNYYNERENHNYKVKYPLPWHSF
RQDLMETLAGVFISRAPRRKITGPAFIDETIRSPKHFNKGLTSVKIPLTTVTLEKLETMV
KNTKGGISDKAVYlWLKNRilEHNNKPLKAFAEKIYKPLKNGTNGAIIRSIRVETPSY
TGVFRNEGKGISDNSLMVRVDVFKKKDKYYLVPIYVAHMIKKELPSKAIVPLKPESQ
WELIDSTHEFLFSLYQNDYLVIKTKKGITEGYYRSCmGTGSLSLMPHFANNKNVKID
IGVRTAISIEKYNVDILGNKSIVKGEPRRGMEKYNSFKSN

[0099] YP_00255 1549.1 crispr-associated protein, csnl family [Acidovorax ebreus TPSY]

MAQHVFGLDIGIASVGWAILGEQRIIDLGVRCFDKAETAKEGDPLNLTRRQARLLRR
RLYRRAWRLTQLRLLKRKGLIADAKLFAKAPSYGDSAWELRRQGLDRLLTPLEWAR
VIYHQCKHRGFHWTSKAEEAKADSDAEGGRVKQGLAHTKALMQAKNYRSAAEMV
LAEFPDAQRNKRGQYDKALSRVLLGEELALLFATQRRLGNPHASDFFEKLILGDGDR
KSGLFWQQKPALSGADLLKMLGKCTFEKGEYRAPKASFSVERHVWLTRLNNLRIVV
DGRSRPLNEAERQAALLLPYQTETSKYKTLKNAFIKAGLWGDGVRFGGLAYPSQAQI
DAEKTKDPEDQFLVKLPAWHELRKAFKAAGHEALWQQISTPALDGDPTLLDQIATV
LSVYKDGAEVVQQLRQLALPEPAASIAVLEKISFDKFSSLSLKALRRIVPLMQSGLRY
DEAVAQIPEYGHHSQRIEPGAAKHLYLPPFYEAQRKYAGKGDfflGSMQFRDDADIPR
NPWLRALNQARKVVNALIREYGSPIAVMEMARDLSRPLDERNKVKRAQEEFRDRN
DRARSEFERDFGYKPKAAAFEKWMLYREQLGQCAYSQQPLDIQRVLDDHNYAQVD
HALPYSRSYDDSKNNKVLVLTHENQ>fKGNRTAFEYLTSFPDGEDGERWRTFVAWV
QGNKAYRMAKRNRLLRKNYGVDESKGFIDRNL>TOTRYICKFFKNWEEFD**ŁQLAAR**
ADGDTARRCVVVNGQLTAFLRARWGLTKVRGDSDRHHALDAAVVAACTHGMVK
ALADYSRRKEISFLQEGFPDPETGEILNPAAFDRARQHFPEPWTHFAHELKARLFTDD
LAALREDMQRLGSYTTEDLGRLRTLFVSRAPQRRSGGAVFIKETIYAQPESLKQQGG
VIEKJLLTSLKLQDFDKLLNPESNDHFVEPHRNERLYAAIRQRLEQFGGRADKAFGPD
NLFHKPDKNNQPTGPVVRSIKLVRGKQTGIPIRGGLAKNDSMLRVDIFTKAGKFHLV
PVYVRHR VlGLPNRAIVAFKDEDEWTLIDESFAFLFSVYPNDYVKVTLKKEQQSGYY
SGADRSTGAMNLWAHDRAASVGKDGLIRGIGVKTALSVEKFNVDVLGRIYLAPPET

RSGLA[0100] YP_002342100.1 hypothetical protein NMA0631 [Neisseria meningitidis Z2491]

MAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRVFERAEVPKTGDSL
AMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLR
AAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNAHAL
QTGDFRTPAELALNKFEKESGffIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHV
SGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLN
NLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDN
AEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRL
KDRIQPEILEALLKffISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKN
TEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDR
KEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEI
NLGRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNS
REWQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRM
RLTGKGKKRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQ
KITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPE
FEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRL
DEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYK
YDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYYLV
PIYSWQVAKGILPDPvAVVQGKDEEDWQLIDDSFNFKPSLHPNDLVEVITKXAPvMFGY
FASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKR
PPVR

[0101] NP_246064.1 hypothetical protein PM1127 [Pasteurella multocida subsp. multocida str. Pm70]

MQTTNLSYILGLDLGIASVGWAVVEINENEDPIGLIDVGVRIFERAEVPKTGESLALSR
RLARSTRRLIRRRAHRLLLAKRFLKREGILSTIDLEKGLPNQAWELRVAGLERRLSAIE
WGAVLLHLIKmGYLSKRKNESQTNNKELGALLSGVAQNHQLLQSDDYRTPAELAL
KKFAKEEGffIRNQRGAYTHTFNRLDLLAELNLLFAQQHQFGNPHCKEHIQQYMTEL
LMWQKPALSGEAILKMLGKCTHEKNEFKAAKHTYSAERFVWLTKLNNLRILEDGAE
RALNEEERQLLINHPYEKSKLTYAQVRKLLGLSEQAIFKHLRYSKENAESATFMELK
AWHAIRKALENQGLKDTWQDLAKKPDLLDEIGTAFSLYKTDEDIQQYLTNKVPNSVI
NALLVSLNFDKFIELSLKSLRKILPLMEQGKRYDQACREIYGHHYGEANQKTSQLLP

AIPAQEIRNPVVLRTLSQARKVINAIIRQYGSPARVHIETGPvELGKSFKERREIQKQQE
DNRTKRESAVQKFKELFSDFSSEPKSKDILKFRLYEQQHGKCLYSGKEINIHRLNEKG
WEIDHALPFSRTWDDSFNNKVLVLASENQNKGNQTPYEWLQGKINSERWKNFVAL
VLGSQCSAAKKQRLLTQVIDDNKFIDRNLNDTRYIARFLSNYIQENLLLVGKNKKNV
FTPNGQITALLRSRWGLIKARENNNRHHALDAIVVACATPSMQQKITRFIRFKEVHPY
KIE>mYEMVDQESGEnSPHFPEPWAYFRQEWIRVFDNHPDTVLKEMLPDRPQANH
QFVQPLFVSRAPTRKMSGQGHMETIKSAKRLAEGISVLRIPLTQLKP^LENMVNKE
REPALYAGLKARLAEFNQDPAKAFATPFYKQGGQQVKAIRVEQVQKSGVLVRENN
GVADNASIVRTDVFIKNNKFFLVPIYTWQVAKGILPNKAIVAHKNEDEWEEMDEGA
KFKFSLFPNDLVELKTKKEYFFGYYIGLDRATGNISLKEHDGEISKGKDGVYRVGVK
LALSFEKYQVDELGKNRQICRPQQRQPVR

[0102] ZP_07738815.1 CRISPR-associated protein, Csnl family [Aminomonas paucivorans DSM 12260]

MIGEHVRGGCLFDDHWTPNWGAFRLPNTVRTFTKAENPKDGSSLAEPRRQARGLRR
RLRRKTQRLEDLRRLLAKEGVLSLSDLETLFRETPAKDPYQLRAEGLDRPLSFPEWV
RVLYfflTKHRGFQSNRRNPVEDGQERSRQEEEGKLLSGVGENERLLREGGYRTAGE
MLARDPKFQDHRRNRAGDYSHTLSRSLLLEEARRLFQSQRTLGNPHASSKLEEAFLH
LVAFQNPFASGEDIRNKAGHCSLEPDQIRAPRRSASAETFMLLQKTGNLRLIHRRTGE
ERPLTDKEREQIHLLAWKQEKVTHKTLRRHLEIPEEWLFTGLPYFIRSGDKAEEKLFV
HLAGIHEIRKALDKGPDPAVWDTLRSRRDLLDSIADTLTFYKNEDEILPRLESLGLSPE
NARALAPLSFSGTAHLSLSALGKLLPHLEEGKSYTQARADAGYAAPPPDRHPKLPPL
EEADWRNPVVFRALTQTRKVVNALVRRYGPPWCIHLETARELSQPAKVRRRIETEQ
QANEKKKQQAEREFLDIVGTAPGPGDLLKMRLWREQGGFCPYCEEYLNPTRLAEPG
YAEMDHILPYSRSLDNGWHNRVLVHGKDNRDKGNRTPFEAFGGDTARWDRLVAW
VQASFILSAPKKRNLLREDFGEEAERELKDRNLTDTRFITKTAATLLRDRLTFHPEAPK
DPVMTLNGRLTAFLRKQWGLHKNRKNGDLFfflALDAAVLAVASRSFVYRLSSHNAA
WGELPRGREAENGFSLPYPAFRSEVLARLCPTREEILLRLDQGGVGYDEAFRNGLRP
VFVSRAPSRRLRGKAHMETLRSPKWKDHPEGPRTASRIPLKDLNLEKLERMVGKDR
DRKLYEALRERLAAFGGNGKKAFVAPFRKPCRSGEGPLVRSLRIFDSGYSGVELRDG
GEVYAVADHESMVRVDVYAKKNRFYLVPVYYVADVARGIVKNRArVAHKSEEEWD
LVDGSFDFRFSLFPGDLVEIEKKDGAYLGYYKSCHRGDGRLLLDRHDRMPRESDCG
TFYVSTRKDVLSMSKYQVDPLGEIRLVGSEKPPFVL

[0103] ZP_08574780.1 CRISPR-associated protein, Csnl family [Lactobacillus coryniformis subsp. torquens KCTC 3535]

MGYRIGLDVGITSTGYAVLKTDKNGLPYKILTLDSVIYPRAENPQTGASLAEPRRIKR
GLPvPxRTRRTKFPvKQRTQQLFIHSGLLSKPEIEQILATPQAKYSVYELRVAGLDRRLTN
SELFRVLYFFIGHRGFKSNRKAELNPENEADKKQMGQLLNSIEEIRKAIAEKGYRTVG
ELYLKDPKYTNJDHKRNKGYIDGYLSTPNRQMLVDEIKQILDKQRELGNEKLTDEFYA
TYLLGDENRAGIFQAQRDFDEGPGAGPYAGDQIKKMVGKDIFEPTEDRAAKATYTF
QYFNLLQKMTSLNYQNTTGDTWHTLNGLDRQAIIDAVFAKAEKPTKTYKPTDFGEL
PJO.LKLPDDARFNLVNYGSLQTQKEIETVEKKTRFVDFKAYHDLVKVLPEEMWQSR
QLLDfflGTALTLYSSDKJUlRRYFAEELNLPAELIEKLLPLWSKFGHLSIKSMQNIIPYL
EMGQVYSEATTmTGYDFPvKKQISKDTIREEITNPVVRRAVTKTIKIVEQIIRRYGKPDG
INIELARELGRNFKERGDIQKRQDKNRQTNDKIAAELTELGIPVNGQNIIRYKLHKEQ
NGVDPYTGDQIPFERAFSEGYEVDHIIPYSISWDDSYTNKVLTSAKCNREKGNRIPMV
YLANNEQRLNALTNIADMIRNSRKJIQKLLKQKLSDEELKDWKQRMNDTRFITRVL
YNYFRQAIEFNPELEKKQRVLPLNGEVTSKIRSRWGFLKVREDGDLHHAIDATVIAAI
TPKFIQQVTKYSQHQEVKNNQALWHDAEIKDAEYAAEAQRMDADLFNKIFNGFPLP
WPEFLDELLAPJSDNPVEMMKSRSWNTYTPIEIAKLKPVFVVRLANHKISGPAHLDTI
RSAKLFDEKGIVLSRVSITKLKINKKGQVATGDGIYDPENSNNGDKVVYSAIRQALEA
HNGSGELAFPDGYLEYVDHGTKKLVRKVRVAKKVSLPVRLKNKAAADNGSMVRID
VFNTGKKFVFVPIYIKDTVEQVLPNKAIARGKSLWYQITESDQFCFSLYPGDMVHIES
KTGIKPKYSNKENNTSVVPIKNFYGYFDGADIATASILVRAHDSSYTARSIGIAGLLKF
EKYQVDYFGRYHKVHEKKRQLFVKRDE

[0104] ZP_03755025.1 hypothetical protein ROSEINA2194_03455 [Roseburia inulinivorans DSM 16841]

MNAEHGKEGLLIMEENFQYRIGLDIGITSVGWAVLQNNSQDEPVRITDLGVRIFDVA
ENPKNGDALAAPPJIDARTTRRRLRRRRHRLERIKFLLQENGLIEMDSFMERYYKGN
LPDVYQLRYEGLDRKLKDEELAQVLIHIAKFIRGFRSTRKAETKEKEGGAVLKATTEN
QKIMQEKGYRTVGEMLYLDEAFHTECLWNEKGYVLTPRNRPDDYKHTILRSMLVEE
VHAIFAAQRAHGNQKATEGLEEAYVEIMTSQRSFDMGPGLQPDGKPSPYAMEGFGD
RVGKCTFEKDEYRAPKATYTAELFVALQKINHTKLIDEFGTGRFFSEEERK I'11GLLLS
SKELKYGTIRKKLNIDPSLKFNSLNYSAKKEGETEEERVLDTEKAKFASMFWTYEYS
KCLKDRTEEMPVGEKADLFDRIGEILTAYKNDDSRSSRLKELGLSGEEIDGLLDLSPA

KYQRVSLKAMRKMQPYLEDGLIYDKACEAAGYDFRALNDGNKKHLLKGEE1NAIV
NDIT^VVKRSVSQTIKVINAHQKYGSPQAVNIELAREMSKNFQDRTNLEKEMKKRQ
QE>ffiRAKQQIIELGKQNPTGQDILKYRJLWNDQGGYCLYSGKKIPLEELFDGGYDIDHI
LPYSITFDDSYRNKVLVTAQENRQKGNRTPYEYFGADEKRWEDYEASVRLLVRDYK
KQQKLLKKNFTEEERKEFKJERNLNDTKYTTRVVYNMIRQNLELEPFNHPEKKKQVW
AVNGAVTSYLRKRWGLMQKDRSTDRHHAMDAWIACCTDGMIHKISRYMQGREL
AYSRNFKFPDEETGEILNRDNFTREQWDEKFGVKVPLPWNSFRDELDIRLLNEDPKN
FLLTHADVQRELDYPGWMYGEEESPIEEGRYINYIRPLFVSRMPNHKVTGSAHDATI
RSARDYETRGVVITKVPLTDLKLNKDNEIEGYYDKDSDRLLYQALVRQLLLHGNDG
KKAFAEDFHKPKADGTEGPVVRKVKIEKKQTSGVMVRGGTGIAANGEMVRIDVFRE
NGKYYFVPVYTADVVRKVLPNRAATHTKPYSEWRVMDDANFVFSLYSRDLIHVKS
KKDIKTNLVNGGLLLQKEIFAYYTGADIATASIAGFANDSNFKFRGLGIQSLEIFEKCQ
VDILGNISVVRHENRQEFH

[0105] ZP_10953934.1 HNH endonuclease [Alicyclobacillus hesperidum URH17-3-68]

MAYRLGLDIGITSVGWAVVALEKDESGLKPVRIQDLGVRIFDKAEDSKTGASLALPR
REARSARRRTRRRRHRLWRVKRLLEQHGILSMEQIEALYAQRTSSPDVYALRVAGL
DRCLIAEEIARVLIfflAHRRGFQSNRKSEIKDSDAGKLLKAVQENENLMQSKGYRTV
AEMLVSEATKTDAEGKLVHGKKHGYVSNVRNKAGEYRHTVSRQAIVDEVRKIFAA
QRALGNDVMSEELEDSYLKILCSQRNFDDGPGGDSPYGHGSVSPDGVRQSIYERMV
GSCTFETGEKRAPRSSYSFERFQLLTKVVNLRIYRQQEDGGRYPCELTQTERARVIDC
AYEQTKITYGKLRKLLDMKDTESFAGLTYGLNRSRNKTEDTVFVEMKFYFDEVRKAL
QRAGVFIQDLSIETLDQIGWILSVWKSDDNRRKKLSTLGLSDNVIEELLPLNGSKFGH
LSLKAIRKJLPFLEDGYSYDVACELAGYQFQGKTEYVKQRLLPPLGEGEVTNPVVRR
ALSQAIKVVNAVIRKHG SPE SIFIIELAREL SKNLDERRKIEK AQKENQKNNEQIKDE1R
EILGSAHVTGRDIVKYKLFKQQQEFCMYSGEKLDVTRLFEPGYAEVDHIIPYGISFDD
SYDNKVLVKTEQNRQKGNRTPLEYLRDKPEQKAKFIALVESIPLSQKKKNHLLMDK
RAIDLEQEGFRERNLSDTRYITRALMNHIQAWLLFDETASTRSKRVVCVNGAVTAY
MRARWGLTKDRDAGDKHHAADAVVVACIGDSLIQRVTKYDKFKRNALADRNRYV
QQVSKSEGITQYVDKETGEVFTWESFDERKFLPNEPLEPWPFFRDELLARLSDDPSKN
IRAIGLLTYSETEQIDPIFVSRMPTRKVTGAAF£KETIRSPRIVKVDDNKG ' 1E1QWVSK
VALTELKLTKDGEIKDYFRPEDDPRLYNTLRERLVQFGGDAKAAFKEPVYKISKDGS
VRTPVRKVKIQEKLTLGVPVHGGRGIAENGGMVRIDVFAKGGKYYFVPIYVADVLK

PvELPNPJ_ATAHKPYSEWRVVDDSYQFKFSLYPNDAVMIKPSREVDITYICDPvKEPVG
CRIMYFVSANIASASISLRTHDNSGELEGLGIQGLEVFEKYVVGPLGDTUPVYKERRM
PFRVERKMN

[0106] ADI19058.1 uncharacterized protein conserved in bacteria [uncultured delta proteobacterium HF0070_07E19]

MSSKAIDSLEQLDLFKPQEYTLGLDLGIKSIGWAILSGERIANAGVYLFETAEELNSTG
NKLI SKA AERGRKRJRJRRMLDRKARRGRFIIR YLLEREGLPTDELEEV VVHQ SNRTL W
DVRAEAVERKLTKQELAAVLFHLVRmGYFPNTKXLPPDDESDSADEEQGKINTlATS
PvLREELKASDCKTIGQFLAQNRDRQRNREGDYSNLMARKLVFEEALQILAFQRKQG
HELSKDFEKTYLDVLMGQRSGRSPKLGNCSLIPSELRAPSSAPSTEWFKFLQNLGNLQ
ISNAYREEWSIDAPRRAQIIDACSQRSTSSYWQIRJIDFQIPDEYPvFNLVNYERRDPDV
DLQEYLQQQERXTLA>fFRNWKQLEKIIGTGHPIQTLDEAARLITLIKDDEKLSDQLAD
LLPEASDKAITQLCELDFTTAAKJSLEAMYRILPHMNQGMGFFDACQQESLPEIGVPP
AGDRVPPFDEMYNPVVNRVLSQSRKLINAVIDEYGMPAKIRVELARDLGKGRELRE
RIKLDQLDKSKQNDQRAEDFRAEFQQAPRGDQSLRYRLWKEQNCTCPYSGRMIPVN
SVLSEDTQIDHILPISQSFDNSLSNKVLCFTEENAQKSNRTPFEYLDAADFQRLEAISG
NWPEAKJlNKLLHKSFGKVAEEWKSRALhTOTRYLTSALADHLRJHHLPDSKIQTVNGR
ITGYLRKQWGLEKDRDKHTHHAVDAIVVACTTPAIVQQVTLYHQDIRRYKKLGEKR
PTPWPETFRQDVLDVEEEIFITRQPKKVSGGIQTKDTLRKHRSKPDRQRVALTKVKLA
DLERLVEKDASNRNLYEHLKQCLEESGDQPTKAFKAPFYMPSGPEAKQRPILSKVTL
LREKPEPPKQLTELSGGPJIYDSMAQGRLDIYRYKPGGKJRXDEYRWLQRMIDLMRG
EENVHVFQKGVPYDQGPEIEQNYTFLFSLYFDDLVEFQRSADSEVIRGYYRTFNIANG
QLKISTYLEGRQDFDFFGANRLAHFAKVQVNLLGKVIK

[0107] ZP_08157403.1 CRISPR-associated protein, Csnl family [Ruminococcus albus 8]

MGNYYLGLDVGIGSIGWAVINIEKKRIEDFNVRIFKSGEIQEKNRNSRASQQCRRSRG
LRRLYRRKSFIPvKLRLKNYLSnGLTTSEKIDYYYETADNNVIQLRNKGLSEKJLTPEEIA
ACLIfflCNNRGYKJ)FYEVNVEDIEDPDERNEYKEEHDSIVLISNLMNEGGYCTPAEMI
CNCREFDEPNSVYRKFHNSAASKNHYLITRHMLVKEVDLILENQSKYYGILDDKTIA
KJKDIIFAQPvDFEIGPGKNEPJ^RRFTGYLDSIGKCQFFKDQERGSRPTVIADIYAFVNV
LSQYTYTONRGESVFDTSFANDLINSALKNGSMDKRELKAIAKSYfflDISDKNSDTSL
TKCFKYIKVVKPLFEKYGYDWDKLIENYTDTDNNVLNPJGIVLSQAQTPKRRREKLK

ALNIGLDDGLINELTKLKLSGTANVSYKYMQGSlEAFCEGDLYGKYQAKFNKEIPDID
ENAKPQKLPPFKNEDDCEFFKNPVVFRSINETRKLINAIIDKYGYPAAVNIETADELNK
TFEDRAIDTKRN>JDNQKiNDRIVKEIIECIKCDEVHARHLIEKYKLWEAQEGKCLYSG
ETITKEDMLPJ)KDKl.FEVDfflVPYSLILDNTINMCALVYAEENQKKGQRTPLMYMNE
AQAADYRVRVNTMFKSKXCSKKKYQYLMLPDLNDQELLGGWRSRNLNDTRYICK
YLV?^LRKNLRPDRSYESSDEDDLKIRDFTYRVFPVKSRFTSMFRRWWLNEKTWGR
YDKAELKKLTYLDHAADAIIIANCRPEYVVLAGEKLKLNKMYHQAGKRITPEYEQS
KKACID>n.**YKLFPvMDRRTAEKLLSGHGRLTPnP** ^SEEVDKRLWDKNIYEQFWKDD
KDKKSCEELYRENVASLYKGDPKFASSLSMPVISLKPDHKYRGTITGEEAIRVKEIDG
KLIKLKRKSISEITAESINSIYTDDKILIDSLKTIFEQADYKDVGDYLKKTNQHFFTTSS
GKRVNKVTVIEKVPSRWLRKEIDDNNFSLLNDSSYYCIELYKDSKGDNNLQGIAMSD
IVFTORKTKJCLYLKiDFWPDDYYTHVMYIFPGDYLRIKSTSKKSGEQLKFEGYFISVK
NVNENSFmSDNKPCAKDKRVSITKKDIVIKLAVDLMGKVQGENNGKGISCGEPLSL
LKEKN

[0108] ZP_10010146.1 CRISPR-associated protein Cas9/Csnl, subtype II/NMEMI [Treponema sp. JC4]

MIMKLEKWRLGLDLGTNSIGWSVFSLDKDNSVQDLIDMGVRIFSDGRDPKTKEPLA
VARRTARSQRKLIYRRKLRRKQVFKFLQEQGLFPKTKEECMTLKSLNPYELRIKALD
EKLEPYELGRALFNLAVRRGFKS>niKDGSREEVSEKKSPDEIKTQADMQTHLEKAIK
ENGCRTITEFLYKNQGENGGIRFAPGRMTYYPTRKMYEEEFNLIRSKQEKYYPQVDW
DDIYKAIFYQRPLKPQQRGYCIYENDKERTFKAMPCSQKLRILQDIGNLAYYEGGSK
KRVELNDNQDKVLYELLNSKDKVTFDQMRKALCLADSNSFNLEENRDFLIGNPTAV
KMRSKNRFGKLWDEIPLEEQDLIIETIITADEDDAVYEVIKKYDLTQEQRDFIVKNTIL
QSGTSMLCKEVSEKLVKRLEEIADLKYHEAVESLGYKFADQTVEKYDLLPYYGKVL
PGSTMEIDLSAPETNPEKHYGKISNPTVHVALNQTRVVVNALIKEYGKPSQIA1ELSRD
LK>nWEKKAEIARKQNQRAKEMAlM)TISALYHTAFPGKSFYPMlNDRMKYRLWSE
LGLGNKCIYCGKGISGAELFTKEIEIEHILPFSRTLLDAESNLTVAHSSCNAFKAERSPF
EAFGTNPSGYSWQEIIQRANQLKNTSKKNKFSPNAMDSFEKDSSFIARQLSDNQYIAK
AALRYLKCLVENPSDVWTTNGSMTKLLRDKWEMDSILCRKFTEKEVALLGLKPEQI
GNYKKNRFDHRHHAIDAVVIGLTDRSMVQKLATKNSFDCGNRJEIPEFPILRSDLIEKV
KMVVSFKPDHGAEGKLSKETLLGKIKLHGKETFVCRENIVSLSEKNLDDIVDEKIKS
KVKDYVAKHKGQKIEAVLSDFSKENGIKKVRCVNRVQTPIEITSGKISRYLSPEDYFA

AVIWEIPGEKKTFKAQYIRRNEVEKNSKGLNVVKPAVLENGKPHPAAKQVCLLHKD
DYLEFSDKGKMYFCRIAGYAATNNKLDIRPVYAVSYCADWINSTNETMLTGYWKPT
PTQNWVSVNVLFDKQKARLVTVSPIGRVFPvK

[0109] ZP_11150502.1 CRISPR-associated protein, Csnl family [Alcanivorax pacificus W11-5]

NmYRVGLDLGTASVGAAVFSMDEQGNPMELIWHYERLFSEPLVPDMGQLKPKKAA
RRLARQQRRQIDRRASRLRRIAIVSRRLGIAPGRNDSGVHGNDVPTLRAMAVNERIEL
GQLRAVLLRMGKKRGYGGTFKAVRKVGEAGEVASGASRLEEEMVALASVQNKDS
VTVGEYLAARVEHGLPSKLKVAANNEYYAPEYALFRQYLGLPAIKGRPDCLPNMYA
LRHQIEHEFERIWATQSQFHDVMKDHGVKEEIRNAIFFQRPLKSPADKVGRCSLQTN
LPRAPRAQIAAQNFRIEKQMADLRWGMGRRAEMLNDHQKAVIRELLNQQKELSFRK
IYKELERAGCPGPEGKGLNMDRAALGGRDDLSGNTTLAAWRKLGLEDRWQELDEV
TQIQVINFLADLGSPEQLDTDDWSCRFMGKNGRPRNFSDEFVAFMNELRMTDGFDR
LSKMGFEGGRSSYSIKALKALTEWMIAPHWRETPETHRVDEEAAIRECYPESLATPA
QGGRQSKLEPPPLTGNEVVDVALRQVRHTINMMIDDLGSVPAQIWEMAREMKGGV
TRRNDIEKQNKRFASERKKAAQSIEENGKTPTPARILRYQLWIEQGHQCPYCESNISL
EQALSGAYTNFEHILPRTLTQIGRKRSELVLAHRECNDEKGNE.TPYQAFGHDDRRWR
IVEQRANALPKKSSRKTRLLLLKDFEGEALTDESIDEFADRQLHESSWLAKVTTQWL
SSLGSDVYVSRGSLTAELRRRWGLDTVIPQVRFESGMPVVDEEGAEITPEEFEKFRLQ
WEGHRVTREMRTDRRPDKRIDHRHHLVDAIVTALTSRSLYQQYAKAWKVADEKQR
HGRVDVKVELPMPILTIRDIALEAVRSVRISHKPDRYPDGRFFEATAYGIAQRLDERS
GEKVDWLVSRKSLTDLAPEKKSIDVDKVRANISRIVGEAIRLHISNIFEKRVSKGMTP
QQALREPIEFQGMLRKVRCFYSKADDCVRIEHSSRRGHHYKMLLNDGFAYMEVPC
KEGILYGVPNLVRPSEAVGIKRAPESGDFIRFYKGDTVKNIKTGRVYTIKQILGDGGG
KLILTPVTETKPADLLSAKWGRLKVGGRNIFILLRLCAE

[0110] ZP_18919511.1 hypothetical protein C882_0672 [Caenispirillum salinarum AK4]

MPVLSPLSPNAAQGRRRWSLALDIGEGSIGWAVAEVDAEGRVLQLTGTGVTLFPSA
WSNENGTYVAHGAADRAVRGQQQRHDSRRRRLAGLARLCAPVLERSPEDLKDLTR
TPPKADPRAIFFLRADAARRPLDGPELFRVLHHMAAHRGIRLAELQEVDPPPESDAD
DAAPAATEDEDGTPJRAAADERAFRRLMAEFIMHRHGTQPTCGEIMAGRLRETPAGA
QPVTRARDGLRVGGGVAVPTRALIEQEFDAIRAIQAPRHPDLPWDSLRRLVLDQAPI

AVPPATPCLFLEELRRRGETFQGRTITREAIDRGLTVDPLIQALRIRETVGNLRLHERIT
EPDGRQRYVPRAMPELGLSHGELTAPERDTLVRALMHDPDGLAAKDGRIPYTRLRK
LIGYDNSPVCFAQERDTSGGGITVNPTDPLMARWIDGWVDLPLKARSLYVRDVVAR
GADSAALARLLAEGAHGVPPVAAAAVPAATAAILESD1MQPGRYSVCPWAAEAILD
AWANAPTEGFYDVTRGLFGFAPGEIVLEDLRRARGALLAHLPRTMAAARTPNRAAQ
QRGPLPAYESVIPSQLITSLRRAHKGRAADWSAADPEERNPFLRTWTGNAATDHILN
QVRKTANEVITKYGNRRGWDPLPSRITVELAREAKHGVIRRNEIAKENRENEGRRKK
ESAALDTFCQDNTVSWQAGGLPKERAALRLRLAQRQEFFCPYCAERPKLRATDLFSP
AETEIDHVIERRMGGDGPDNLVLAHKDCNNAKGKKTPHEHAGDLLDSPALAALWQ
GWRKENADRLKGKGHKARTPPxEDKDFMDRVGWRFEEDARAKAEENQERRGRRML
HDTARATRLARLYLAAAVMPEDPAEIGAPPVETPPSPEDPTGYTAIYRTISRVQPVNG
SVTHMLRQRLLQRDKNRDYQTHHAEDACLLLLAGPAVVQAFNTEAAQHGADAPDD
RPVDLMPTSDAYHQQRRARALGRVPLATVDAALADIVMPESDRQDPETGRVHWRL
TRAGRGLKRRIDDLTRNCVILSRPRRPSETGTPGALHNATHYGRREITVDGRTDTVVT
QRMNARDLVALLDNAKIVPAARLDAAAPGDTILKEICTEIADRHDRVVDPEGTHARR
WISARLAALVPAHAEAVARDIAELADLDALADADRTPEQEARRSALRQSPYLGRAIS
AKXADGRARAREQEILTRALLDPHWGPRGLRFILIMREARAPSLVRIRANKTDAFGRP
VPDAAVWVKTDGNAVSQLWRLTSVVTDDGRRIPLPKPIEKRIEISNLEYARLNGLDE
GAGVTGNNAPPRPLRQDIDRLTPLWRDHGTAPGGYLGTAVGELEDKARSALRGKA
MRQTLTDAGITAEAGWRLDSEGAVCDLEVAKGDTVKKDGKTYKVGVITQGIFGMP
VDAAGSAPRTPEDCEKFEEQYGIKPWKAKGIPLA

[0111] YP_425545.1 CRISPR-associated endonuclease Csnl family protein [Rhodospirilhim rubrum ATCC 11170]

MRPIEPWILGLDIGTDSLGWAVFSCEEKGPPTAKELLGGGVRLFDSGRDAKDHTSRQ
AERGAFRRARRQTRTWPWRRDRLIALFQAAGLTPPAAETRQIALALRREAVSRPLAP
DALWAALLHLAHHRGFRSNRIDKRERAAAKALAKAKPAKATAKATAPAKEADDEA
GFWEGAEAALRQRMAASGAPTVGALLADDLDRGQPVRMRYNQSDRDGVVAPTRA
LIAEELAEIVARQSSAYPGLDWPAVTRLVLDQRPLRSKGAGPCAFLPGEDRALRALP
TVQDFIIRQTLANLRLPSTSADEPRPLTDEEHAKALALLSTARFVEWPALRRALGLKR
GVKFTAETERNGAKQAARGTAGNLTEAILAPLIPGWSGWDLDRKDRVFSDLWAAR
QDRSALLALIGDPRGPTRVTEDETAEAVADAIQIVLPTGRASLSAKAARAIAQAMAP
GIGYDEAVTLALGLHHSHRPRQERLARLPYYAAALPDVGLDGDPVGPPPAEDDGAA

AEAYYGWGMSVfflALNETRKIVNALLHRHGPILRLVMVETTRELKAGADEPvKRMIA
EQAERERENAEIDVELRKSDRWMANARERRQRVRLARRQNNLCPYTSTPIGHADLL
GDAYDIDHVIPLARGGRDSLDNMVLCQSDANKTKGDKTPWEAFHDKPGWIAQRDD
FLARLDPQTAKALAWRFADDAGERVARKSAEDEDQGFLPRQLTDTGYIARVALRYL
SLVTNEPNAVVATNGRLTGLLRLAWDITPGPAPRDLLPTPRDALRDDTAARRFLDGL
TPPPLAKAVEGAVQARLAALGRSRVADAGLADALGLTLASLGGGGKNRADHRHHFI
DAAMlAVTTRGLINQINQASGAGRILDLRKWPRTNFEPPYPTFRAEVMKQWDfflHPSI
RPAHRDGGSLHAATVFGVRNRPDARVLVQRKPVEKLFLDANAKPLPADKIAEIIDGF
ASPRMAKRFKALLARYQAAHPEVPPALAALAVARDPAFGPRGMTANTVIAGRSDG
DGEDAGLITPFRANPKAAVRTMGNAVYEVWEIQVKGRPRWTHRVLTRFDRTQPAPP
PPPENARLVMRLRRGDLVYWPLESGDRLFLVKKMAVDGRLALWPARLATGKATAL
YAQLSCPNINLNGDQGYCVQSAEGIRKEKIRTTSCTALGRLRLSKKAT

[0112] CCA84553.1 conserved hypothetical protein [Ralstonia syzygii R24]

MAEKQHRWGLDIGTNSIGWAVIALIEGRPAGLVATGSRIFSDGRNPKDGSSLAVERR
GPRQMRPxJUUDRYLRRRDRFMQALrNVGLMPGDAAARKALVTENPYVLRQRGLDQA
LTLPEFGRALFHLNQRRGFQSNRKTDRATAKESGKVKNAIAAFRAGMGNARTVGEA
LARRLEDGRPVRARMVGQGKDEHYELYIAREWIAQEFDALWASQQRFHAEVLADA
ARDRLRAILLFQRKLLPVPVGKCFLEPNQPRVAAALPSAQRFRLMQELNHLRVMTLA
DKRERPLSFQERNDLLAQLVARPKCGFDMLRKTVFGANKEAYRFTIESERRKELKGC
DTAAKLAKVNALGTRWQALSLDEQDRLVCLLLDGENDAVLADALREHYGLTDAQI
DTLLGLSFEDGHMRLGRSALLRVLDALESGRDEQGLPLSYDKAVVAAGYPAHTADL
ENGERDALPYYGELLWRYTQDAPTAKhTOAERKFGKIANPTVHIGLNQLRKLVNALI
QRYGKPAQIVVELAPvNLKAGLEEKERIKKQQTANLERNERIRQKLQDAGVPDNREN
RLRMRLFEELGQGNGLGTPCIYSGRQISLQRLFSNDVQVDHILPFSKTLDDSFANKVL
AQHDANRYKGNRGPFEAFGANRDGYAWDDIRARAAVLPPvNKRNRFAETAMQDWL
HNETDFLARQLTDTAYLSRVARQYLTAICSKDDVYVSPGRLTAMLRAKWGLNRVL
DGVMEEQGRPAVKNRDDHRHHAIDAVVIGATDRAMLQQVATLAARAREQDAERLI
GDMPTPWPNFLEDVRAAVARCVVSHKPDHGPEGGLHNDTAYGIVAGPFEDGRYRV
RHRVSLFDLKPGDLSNVRCDAPLQAELEPIFEQDDARAREVALTALAERYRQRKVW
LEELMSVLPIRPRGEDGKTLPDSAPYKAYKGDSNYCYELFINERGRWDGELISTFRAN
QAAYRRFRNDPARFRRYTAGGRPLLMRLCINDYIAVGTAAERTIFRVVKMSENKITL
AEHFEGGTLKQRDADKDDPFKYLTKSPGALRDLGARRIFVDLIGRVLDPGIKGD

[0113] ZP_10898214.1 CRISPR-associated protein, Csnl family [Rhodovulum sp. PH10]

MGIRFAFDLGTNSIGWAVWRTGPGVFGEDTAASLDGSGVLIFKDGRNPKDGQSLAT
MPJIVPRQSRKPJU^RFVLRRRDLLAALRKAGLFPVDVEEGRRLAATDPYHLRAKAL
DESLTPHEMGRVIFHLNQRRGFRSNRKADRQDREKGKIAEGSKRLAETLAATNCRTL
GEFLWSRHRGTPRTRSPTRIRMEGEGAKALYAFYPTREMVRAEFERLWTAQSRFAP
DLLTPERHEEIAGILFRQRDLAPPKIGCCTFEPSERRLPRALPSVEARGIYERLAHLRIT
TGPVSDRGLTRPERDVLASALLAGKSLTFKAVRKTLKILPHALVNFEEAGEKGLDGA
LTAKLLSKPDHYGAAWHGLSFAEKDTFVGKLLDEADEERLIRRLVTENRLSEDAAR
RCASIPLADGYGRLGRTANTEILAALVEETDETGTVVTYAEAVRRAGERTGRNWHH
SDERDGVILDRLPYYGEILQRHVVPGSGEPEEKNEAARWGRLANPTVfflGLNQLRKV
VNRLIAAHGRPDQIVVELARELKLNREQKERLDRENRKNREENERRTAILAEHGQRD
TAENKIRLRLFEEQARANAGIALCPYTGRAIGIAELFTSEVEIDHILPVSLTLDDSLANR
VLCRREANREKJRRQTPFQAFGATPAWNDIVARAAKLPPNKRWRFDPAALERFEREG
GFLGRQLNETKYLSRLAKIYLGKICDPDRVYVTPGTLTGLLRARWGLNSILSDSNFKN
RSDHRHHAVDAVVIGVLTRGMIQRIAHDAARAEDQDLDRVFRDVPVPFEDFRDHVR
ERVSTITVAVKPEHGKGGALHEDTSYGLVPDTDPNAALGNLVVRKPIRSLTAGEVDR
VRDRALRARLGALAAPFRDESGRVRDAKGLAQALEAFGAENGIRRVRILKPDASVV
TIADRRTGVPYRAVAPGENHHVDIVQMRDGSWRGFAASVFEVNRPGWRPEWEVKK
LGGKLVMRLHKGDMVELSDKDGQRRVKVVQQIEISANRVRLSPHNDGGKLQDRHA
DADDPFRWDLATIPLLKDRGCVAVRVDPIGVVTLRRSNV

[0114] YP_004386148.1 CRISPR-associated protein, Csnl family [Alicycliphilus denitrificans K601]

MRSLRYRLALDLGSTSLGWALFRLDACNRPTAVIKAGVRIFSDGRNPKDGSSLAVTR
RAARAMRRRRDRLLKRKTRMQAKLVEHGFFPADAGKRKALEQLNPYALRAKGLQE
ALLPGEFARALFFnNQRRGFKSNRKTDKKDNDSGVLKXAIGQLRQQMAEQGSRTVG
EYLWTRLQQGQGVRARYREKPYTTEEGKKRIDKSYDLY1DRAMIEQEFDALWAAQA
AFNPTLFHEAARADLKDTLLHQRPLRPVKPGRCTLLPEEERAPLALPSTQRFRIHQEV
>^RLLDENLREVALTLAQRDAVVTALETKAKLSFEQIRKLLKLSGSVQFNLEDAKR
TELKGNATSAALARKELFGAAWSGFDEALQDEIVWQLVTEEGEGAL1AWLQTHTGV
DEARAQAIVDVSLPEGYGNLSRKALARIVPALRAAVITYDKAVQAAGFDHHSQLGFE
YDASEVEDLVHPETGEIRSVFKQLPYYGKALQRHVAFGSGKPEDPDEKRYGKIANPT

VfflGLNQVRMVVNALIRRYGRPTEVVIELARDLKQSREQKVEAQRRQADNQRRNAR
IRRSIAEVLGIGEERVRGSDIQKWICWEELSFDAADRRCPYSGVQISAAMLLSDEVEV
EHILPFSKTLDDSLNNRTVAMRQANRIKRNRTPWDARAEFEAQGWSYEDILQRAER
MPLRKRYRFAPDGYERWLGDDKDFLARALNDTRYLSRVAAEYLRLVCPGTRVIPGQ
LTALLRGKFGLNDVLGLDGEKNPvNDHRHHAVDACVIGVTDQGLMQRFATASAQAR
GDGLTRLVDGMPMPWPTYPJDHVERAVRmWVSHRPDHGFEGAMMEETSYGIRKDG
SIKQRRKADGSAGREISNLIRIHEATQPLRHGVSADGQPLAYKGYVGGSNYCIEITVN
DKGKWEGEVISTFRAYGVVRAGGMGRLRNPHEGQNGRKLIMRLVIGDSVRLEVDG
AERTMRIVKISGSNGQIFMAPIHEANVDARNTDKQDAFTYTSKYAGSLQKAKTRRVT
ISPIGEVRDPGFKG

[0115] YP_003552871.1 CRISPR-associated protein, Csnl family [Candidatus Puniceispirillum marinum IMCC1322]

MRRLGLDLGTNSIGWCLLDLGDDGEPVSIFRTGARIFSDGRDPKSLGSLKATRREARL
TRRRRDRFIQRQKNLINALVKYGLMPADEIQRQALAYKDPYPIRKKALDEAIDPYEM
GRAIFfflNQRRGFKSNRKSADNEAGVVKQSIADLEMKLGEAGARTIGEFLADRQATN
DTVRARRLSGTNALYEFYPDRYMLEQEFDTLWAKQAAFNPSLYTEAARERLKEIVFF
QRKLKPQEVGRCIFLSDEDRISKALPSFQRFRIYQELSNLAWIDFIDGVAHRITASLALR
DHLFDELEHKKKLTFKAMRAILRKQGVVDYPVGFNLESDh^HLIGNLTSCIMRDA
KKMIGSAWDRLDEEEQDSFILMLQDDQKGDDEVRSILTQQYGLSDDVAEDCLDVRL
PDGHGSLSKKAIDRILPVLRDQGLIYYDAVKEAGLGEANLYDPYAALSDKLDYYGK
ALAGHVMGASGKFEDSDEKRYGTISNPTVHIALNQVRAVVNELIRLHGKPDEVVIEI
GPJ)LPMGADGKJlELERFQKEGRAKKERARDELKKLGfflDSRESRQKFQLWEQLAKE
PVDRCCPFTGKMMSISDLFSDKVEIEHLLPFSLTLDDSMANKTVCFRQANRDKGNRA
PFDAFGNSPAGYDWQEILGRSQNLPYAKRWRFLPDAMKRFEADGGFLERQLNDTRY
ISRYTTEYISTIIPKNKIWVVTGRLTSLLRGFWGLNSILRGHNTDDGTPAKKSRDDHRH
HAIDAIVVGMTSRGLLQKVSKAARRSEDLDLTRLFEGRIDPWDGFRDEVKKfflDAIIV
SHRPRKKSQGALHNDTAYGIVEHAENGASTVVHRVPITSLGKQSDIEKVRDPLIKSAL
LNETAGLSGKSFENAVQKWCADNSIKSLRIVETVSIIPITDKEGVAYKGYKGDGNAY
MDIYQDPTSSKWKGEIVSRFDANQKGFIPSWQSQFPTARLIMRLRINDLLKLQDGEIE
EIYRVQRLSGSKILMAPHTEANYDARDRDKNDTFKLTSKSPGKLQSASARKVfflSPT
GLIREG

[0116] YP_003448082.1 CRISPR-associated protein, Csnl family [Azospirillum sp. B510]

MARPAFRAPRJIEHVNGWTPDPHPJSKPFFILVSWHLLSRVVIDSSSGCFPGTSRDHTD
KFAEWECAVQPYRLSFDLGTNSIGWGLLNLDRQGKPREIRALGSRIFSDGRDPQDKA
SLAVARRLARQMRRRRDRYLTRRTRLMGALVRFGLMPADPAARKRLEVAVDPYLA
RERATRERLEPFEIGRALFHLNQRRGYKPVRTATKPDEEAGKVKEAVERLEAAIAAA
GAPTLGAWFAWPvKTRGETLRAPvLAGKGKEAAYPFYPARRMLEAEFDTLWAEQARH
HPDLLTAEAREILRHRIFHQRPLKPPPVGRCTLYPDDGRAPRALPSAQRLRLFQELAS
LRVIHLDLSERPLTPAERDRIVAFVQGRPPKAGRKPGKVQKSVPFEKLRGLLELPPGT
GFSLESDKRPELLGDETGARIAPAFGPGWTALPLEEQDALVELLLTEAEPERAIAALT
ARWALDEATAAKLAGATLPDFHGRYGRRAVAELLPVLERETRGDPDGRVRPIRLDE
AVKLLRGGKDHSDFSREGALLDALPYYGAVLERHVAFGTGNPADPEEKRVGRVAN
PTVfflALNQLRHLVNAILARHGRPEEIVIELARDLKRSAEDRRREDKRQADNQKRNE
EPJCRLILSLGERPTPRNLLKLRLWEEQGPVENPJICPYSGETISMRMLLSEQVDIDHILP
FSVSLDDSAAh¾VVCLREANPJKPvMlSPWEAFGHDSERWAGILARAEALPKNKRWR
FAPDALEKLEGEGGLRARHLNDTRHLSRLAVEYLRCVCPKVRVSPGRLTALLRRRW
GIDAILAEADGPPPEVPAETLDPSPAEKNRADHRHHALDAVVIGCIDRSMVQRVQLA
AASAEREAAAREDNIRRVLEGFKEEPWDGFRAELERRARTIVVSHPvPEHGIGGALHK
ETAYGPVDPPEEGFNLVVRJ<PIDGLSKDEmSVRDPRLPvRALIDRLAIRRRDANDPAT
ALAKAAEDLAAQPASRGIRRVRVLKKESNPIRVEHGGNPSGPRSGGPFHKLLLAGEV
HFIVDVALRADGRRWVGHWVTLFEAHGGRGADGAAAPPRLGDGERFLMRLHKGDC
LKLEFDCGRVRVMQVVKLEPSSNSVVVVEPHQVKTDRSKHVKISCDQLRARGARRV
TVDPLGRVRVHAPGARVGIGGDAGRTAMEPAEDIS

[0117] YP_57 1550.1 hypothetical protein Nham_4054 (plasmid) [Nitrobacter hamburgensis X14]

MHVEIDFPHFSRGDSHLAMMCNEILRGSSVLYRLGLDLGSNSLGWFVTHLEKRGDR
HEPVALGPGGVRIFPDGRDPQSGTSNAVDRRMARGARKRRDRFVERRKELIAALIKY
NLLPDDARERRALEVLDPYALRKTALTDTLPAHHVGRALFHLNQRRGFQSNRKTDS
KQSEDGAIKQAASRLATDKGNETLGVFFADMHLRKSYEDRQTAIRAELVRLGKDHL
TGNARKKIWAKVRKRLFGDEVLPRADAPHGVRARA'IlTGTKASYDYYPTRDMLRD
EFNAIWAGQSAHHATITDEARTEIEHIIFYQRPLKPAIVGKCTLDPATRPFKEDPEGYR
APWSHPLAQRFRJLSEARNLEIRDTGKGSRRLTKEQSDLVVAALLANREVKFDKLRT

LLKLPAEARFNLESDRRAALDGDQTAARLSDKKGFNKAWRGFPPERQIAIVARLEET
EDENELIAWLEKECALDGAAAARVANTTLPDGHCRLGLRAIKKIVPIMQDGLDEDG
VAGAGYfflAAKRAGYDHAKLPTGEQLGRLPYYGQWLQDAVVGSGDARDQKEKQY
GQFPNPTVHIGLGQLRRVVNDLIDKYGPPTEISIEFTRALKLSEQQKAERQREQRRNQ
DKNXARAEELAKJFGRPAWR^LKMRLWEELAHDPLDPvKCVYTGEQISIERLLSDEV
DIDFflLPVAMTLDDSPANKJICMRYANRHKRKQTPSEAFGSSPTLQGHRYNWDDIAA
RATGLPRNKRWRFDANAREEFDKRGGFLARQLNETGWLARLAKQYLGAVTDPNQI
WVVPGRLTSMLRGKWGLNGLLPSDNYAGVQDKAEEFLASTDDMEFSGVKNRADH
RHHAIDGLVTALTDRSLLWKMANAYDEEHEKFVIEPPWPTMRDDLKAALEKMVVS
HKPDHGIEGKLHEDSAYGFVKPLDATGLKEEEAGNLVYRKAIESLNENEVDRIRDIQ
LRTIVRDHVNVEKTKGVALADALRQLQAPSDDYPQFKHGLRHVRILKKEKGDYLVP
IANPvASGVAYKAYSAGEWCVEVFETAGGKWDGEAVRRFDANKKNAGPKIAHAPQ
WRDANEGAKLVMRIHKGDLIRLDHEGRARIMVVHRLDAAAGRFKLADHNETGNLD
KJIHATONDIDPFRWLMASYTSITLKXLAAVPVRVDELGRVWRVMPN

[0118] YP_001239928.1 hypothetical protein BBta_3952 [Bradyrhizobium sp. BTAil]

MKRTSLRAYRLGVDLGANSLGWFVVWLDDHGQPEGLGPGGVRIFPDGRNPQSKQS
NAAGRRLARSARRRRDRYLQRRGKLMGLLVKHGLMPADEPARKRLECLDPYGLRA
KALDEVLPLHHVGRALFHLNQRRGLFANRAIEQGDKDASAIKAAAGRLQTSMQACG
ARTLGEFLNRRHQLRATVRARSPVGGDVQARYEFYPTRAMVDAEFEAIWAAQAPH
HPTMTAEAHDTIREAIFSQRAMKRPSIGKCSLDPATSQDDVDGFRCAWSHPLAQRFRI
WQDVRNLAVVETGPTSSRLGKEDQDKVARALLQTDQLSFDEIRGLLGLPSDARFNLE
SDRRDHLKGDATGAILSARRHFGPAWHDRSLDRQIDIVALLESALDEAAIIASLGTTH
SLDEAAAQRALSALLPDGYCRLGLRAIKRVLPLMEAGRTYAEAASAAGYDHALLPG
GKLSPTGYLPYYGQWLQNDVVGSDDERDTNERRWGRLPNPTVfflGIGQLRRVVNEL
IRWHGPPAEITVELTPvDLKLSPRRLAELEREQAENQRKNDKRTSLLRKLGLPASTHNL
LKLRLWDEQGDVASECPYTGEAIGLERLVSDDVDIDHLIPFSISWDDSAANKVVCMR
YANREKGNRTPFEAFGHRQGRPYDWADIAERAARLPRGKRWRFGPGARAQFEELG
DFQARLLNETSWLARVAKQYLAAVTHPHRIHVLPGRLTALLRATWELNDLLPGSDD
RAAKSRKDHRHHAIDALVAALTDQALLRRMANAHDDTRRKIEVLLPWPTFRIDLET
RLKAMLVSHKPDHGLQARLHEDTAYGTVEHPETEDGANLVYRKTFVDISEKEIDRIR
DRRLRDLVRAFtVAGERQQGKTLKAAVLSFAQRRDIAGHPNGIRHVRLTKSIKPDYL
VPIRDKAGRIYKSYNAGENAFVDILQAESGRWIARATTVFQANQANESHDAPAAQPI

MRVFKGDMLPJDHAGAEKFVKJVRLSPSNNLLYLVEHHQAGVFQTRHDDPEDSFRW
LFASFDKiREWNAELVRlDTLGQPWPvRKRGLETGSEDATRIGWTRPKKWP

[0119] YP_001531750.1 CRISPR-associated protein [Dinoroseobacter shibae DFL 12 = DSM 16493]

MRLGLDIGTSSIGWWLYETDGAGSDARITGVVDGGVRIFSDGRDPKSGASLAVDRR
AARAMRRRRDRYLRRRATLMKVLAETGLMPADPAEAKALEALDPFALRAAGLDEP
LPLPHLGPvALFHLNQRRGFKSNRKTDRGDNESGKIKDATARLDMEMMANGARTYG
EFLHKRRQKATDPRHVPSVRTRLSIANRGGPDGKEEAGYDFYPDPJIHLEEEFHKLW
AAQGAHHPELTETLRDLLFEKIFFQRPLKEPEVGLCLFSGHHGVPPKDPRLPKAHPLT
QRRVLYETVNQLRVTADGREARPLTREERDQVIHALDNKKPTKSLSSMVLKLPALA
KVLKLRDGERFTLETGVRDAIACDPLRASPAHPDRFGPRWSILDADAQWEVISRIRR
VQSDAEHAALVDWLTEAHGLDRAHAEATAHAPLPDGYGRLGLTATTRILYQLTAD
VVTYADAVKACGWHHSDGRTGECFDRLPYYGEVLERHVIPGSYHPDDDDITRFGRI
TWTVfflGLNQLRPvLVNPJIETHGKPHQIWELARDLKKSEEQKRADIKRIRDTTEAA
KKRSEKLEELEIEDNGRNRMLLRLWEDLNPDDAMRRFCPYTGTRISAAMIFDGSCDV
DfflLPYSRTLDDSFPNRTLCLREANRQKRNQTPWQAWGDTPHWHAIAANLKNLPEN
KRWRFAPDAMTRFEGENGFLDRALKDTQYLARISRSYLDTLFTKGGHVWVVPGRFT
EMLRRHWGLNSLLSDAGRGAVKAKNRTDHRHHAIDAAVIAATDPGLLNRISRAAGQ
GEAAGQSAELIARDTPPPWEGFRDDLRVRLDRIIVSHRADHGRIDHAARKQGRDSTA
GQLHQETAYSIVDDIHVASRTDLLSLKPAQLLDEPGRSGQVRDPQLRKALRVATGGK
TGKDFENALRYFASKPGPYQAIRRVRIIKPLQAQARVPVPAQDPIKAYQGGSNHLFEI
WRLPDGEIEAQVITSFEAHTLEGEKRPHPAAKRLLRVHKGDMVALERDGRRVVGHV
QKMDIANGLFIVPHNEANADTRNNDKSDPFKWIQIGARPAIASGIRRVSVDEIGRLRD
GGTRPI

[0120] YP_001411379.1 CRISPR-associated endonuclease Csnl family protein [Parvibaculum lavamentivorans DS-1]

MEWFGFDIGTTSIGFSVIDYSSTQSAGNIQRLGVRIFPEARDPDGTPLNQQRRQKRMM
RRQLRRRRIRRKALNETLHEAGFLPAYGSADWPVVMADEPYELRRRGLEEGLSAYE
FGRAIYHLAQHRHFKGRELEESDTPDPDVDDEKEAANERAATLKALKNEQTTLGAW
LARRPPSDRKRGIHAHRNVVAEEFERLWEVQSKFHPALKSEEMRARISDTIFAQRPVF
WPvKNTLGECRFMPGEPLCPKGSWLSQQRRMLEKLNNLAIAGGNARPLDAEERDAIL
SKLQQQASMSWPGVRSALKALYKQRGEPGAEKSLKFNLELGGESKLLGNALEAKLA

DMFGPDWPAHPPvKQEIRHAVHEPvLWAADYGETPDKKRVIILSEKDRKAHREAAANS
FVADFGITGEQAAQLQALKLPTGWEPYSIPALNLFLAELEKGERFGALVNGPDWEG
WPJlTWPHRNQPTGEILDKLPSPASKEEPvEPJSQLPJ^TWRTQNELRKVVNNLIGLY
GKPDRIRIEVGRDVGKSKREREEIQSGIRRNEKQRKKATEDLIKNGIANPSRDDVEKW
ILWKEGQERCPYTGDQIGFNALFREGRYEVEfflWPRSRSFDNSPRNKTLCRKDVNIEK
GNRMPFEAFGHDEDRWSAIQIRLQGMVSAKGGTGMSPGKVKRFLAKTMPEDFAAR
QLNDTRYAAKQILAQLKRLWPDMGPEAPVKVEAVTGQVTAQLRKLWTLNNILADD
GEKTRADHRifflAIDALTVACTFIPGMTNKLSRYWQLRDDPRAEKPALTPPWDTIRAD
AEKAVSEIVVSHRVRKKVSGPLHKETTYGDTGTDIKTKSGTYRQFVTRKKIESLSKGE
LDEIRDPRIKEIVAAHVAGRGGDPKKAFPPYPCVSPGGPEIRKVRLTSKQQLNLMAQT
GNGYADLGSNHfflAIYRLPDGKADFEIVSLFDASRRLAQRNPIVQRTRADGASFVMS
LAAGEAIMIPEGSKKGIWIVQGVWASGQVVLERDTDADHSTTTRPMPNPILKDDAKK
VSIDPIGRVRPSND

[0121] ZP_17295095.1 CRISPR-associated protein cas9/csnl, subtype II/nmemi [Bergeyella zoohelcum ATCC 43767]

IVlKfflLGLDLGTNSIGWALIERNIEEKYGKIIGMGSRIVPMGAELSKFEQGQAQTKNAD
RRTNRGARRLNKRYKQRRNKLIYILQKLDMLPSQIKLKEDFSDPNKIDKITILPISKKQ
EQLTAFDLVSLRVKALTEKVGLEDLGKIIYKYNQLRGYAGGSLEPEKEDIFDEEQSKD
KKNKSFIAFSKIVFLGEPQEEIFKNKKLNRRAIIVETEEGNFEGSTFLENIKVGDSBELLI
MSASKSGDTITIKLPNKTNWRKXMENIENQLKEKSKEMGREFYISEFLLELLKENRW
AKJRNNTILRARYESEFEAIW>ffiQVKHYPFLENLDKKTLIEIVSFIFPGEKESQKKYRE
LGLEKGLKYIIKNQVVFYQRELKDQSHLISDCRYEPNEKAIAKSHPVFQEYKVWEQIN
KLIVNTKJEAGTNRKGEKXYKYIDRPIPTALKEWIFEELQNKKEITFSAIFKKLK^EFD
LREGIDFLNGMSPKDK1KGNETKLQLQKSLGELWDVLGLDSINRQIELWNILYNEKG
NEYDLTSDRTSKVLEFINKYGNNIVDDNAEETAIRISKIKFARAYSSLSLKAVERILPL
VRAGKYFNNDFSQQLQSKILKLLNENVEDPFAKAAQTYLDNNQSVLSEGGVGNSIAT
ILVYDKMTAKEYSHDELYKSYKEINLLKQGDLRNPLVEQIINEALVLIRDIWKNYGIK
PNEIRVELAPJDLKNSAKERATIHKRNKDNQTINNKIKETLVKNKKELSLANIEKVKL
WEAQRHLSPYTGQPIPLSDLFDKEKYDVDHIIPISRYFDDSFTNKVISEKSVNQEKANR
TAMEYFEVGSLKYSIFTKEQFIAIWNEYFSGVKRKNLLATSIPEDPVQRQIKDTQYIAI
RVKEELNKWGNENVKTTTGSITDYLRNHWGLTDKFKLLLKERYEALLESEKFLEAE
YDNYKiCDFDSRKKEYEEKEVLFEEQELTREEFIKEYKENYIRYKKNKLIIKGWSKRID

HPJiHAIDALIVACTEPAfflKRLNDLNKVLQDWLVEHKSEFMPNFEGSNSELLEEILSL
PENERTEIFTQIEKFRAIEMPWKGFPEQVEQKLKEIIISHKPKDKLLLQYNKAGDRQIK
LRGQLHEGTLYGISQGKEAYRIPLTKFGGSKFATEKNTQKIVSPFLSGFIANHLKEYNN
KKEEAFSAEGIMDL>WKLAQYRNEKGELKPHTPISTVKIYYKDPSKNKKKKDEEDLS
LQKLDREKAFNEKLYVKTGDNYLFAVLEGEIKTKKTSQIKRLYDIISFFDATNFLKEE
FRNAPDKKTFDKDLLFRQYFEERNKAKLLFTLKQGDFVYLPNENEEVILDKESPLYN
QYWGDLKERGKNIYVVQKFSKKQIYFIKHTIADIIKKDVEFGSQNCYETVEGRSIKEN
CFKLEIDRLGNIVKVIKR

[0122] ZP_07217791.1 conserved hypothetical protein [Bacteroides sp. 20_3]

MKKIVGLDLGTNSlGWALİNAYINKEFiLYGIEACGSRIlPMDAAILGNFDKGNSISQTA
DRTSYRGIRRLRERHLLRRERLHRILDLLGFLPKHYSDSLNRYGKFLNDIECKLPWVK
DETGSYKFIFQESFKEMLANFTEifflPILIANNKKVPYDWTIYYLRKKALTQKISKEEL
AWILLNFNQKRGYYQLRGEEEETPNKLVEYYSLKVEKVEDSGERKGKDTWYNVHL
ENGMIYRRTSNIPLDWEGKTKEFIVTTDLEADGSPKKDKEGNIKRSFRAPKDDDWTLI
KKKTEADIDKIKMTVGAYIYDTLLQKPDQKIRGKLVRTIERKYYKNELYQILKTQSEF
HEELRDKQLYIACLNELYPNNEPRRNSISTRDFCHLFIEDIIFYQRPLKSKKSLIDNCPY
EENRYIDKESGEIKHASIKCIAKSHPLYQEFRLWQFIVNLRIYRKETDVDVTQELLPTE
ADYVTLFEWLNEKKEIDQKAFFKYPPFGFKKTTShTlfRW^YVEDKPYPCNETHAQIIA
RLGKAfflPKAFLSKEKEETLWFflLYSIEDKQEIEKALHSFANKNNLSEEFIEQFKNFPPF
KKEYGSYSAKAIKKLLPLMRMGKYWSIENIDNGTRIRINKIIDGEYDENIRERVRQKA
INLTDITHFRALPLWLACYLVYDRHSEVKDIVKWKTPKDIDLYLKSFKQHSLRNPIVE
QVITETLRTVRDIWQQVGfflDEIfflELGREMKNPADKRARMSQQMIKNENTNLRIKA
LLTEFLNPEFGIENYRPYSPSQQDLLRIYEEGVLNSILELPEDIGIILGKFNQTDTLKRPT
RSEILRYKLWLEQKYRSPYTGEMIPLSKLFTPAYEIEHIIPQSRYFDDSLSNKVICESEI
NKLKDRSLGYEFIKNF1HGEKVELAFDKPVEVLSVEAYEKLVHESYSHNRSKMKKLL
MEDIPDQFIERQLNDSRYISKVVKSLLSNIVREENEQEAISKNVIPCTGGITDRLKKDW
GINDVWNKIVLPRFIRLNELTESTRFTSINTONTMIPSMPLELQKGFNKXWDHRHHA
MDAHIACANRMVWLNIWSASKNTKITRRDL^ TLLCHKDKTDNNGNYKWVIDKP
WETFTQDTLTALQKITVSFKQNLRVINKTTNHYQHYENGKKIVSNQSKGDSWAIRKS
MHKETVHGEVNLRMIKTVSFNEALKKPQAIVEMDLKKKILAMLELGYDTKJRIKWF
EENKDTWQDINPSKIKVYYFTKETKDRYFAVRKPIDTSFDKKKIKESITDTGIQQIMLR
HLETKDhTOPTLAFSPDGIDEMNRNILILNKGKKHQPIYKVRVYEKAEKFTVGQKGNK

RTKFVEAAKGTNLFFAIYETEEIDKDTKKVIRKRSYSTIPLNVVIERQKQGLSSAPEDE
NG^PKYILSPNDLVYVPTQEEINKGEVVMPIDRDRIYKMVDSSGITANFIPASTANLI
FALPKATAEIYCNGENCIQNEYGIGSPQSKNQKAITGEMVKEICFPIKVDRLGMIQVG
SCILTN

[0123] YP_005848005.1 hypothetical protein IALB_3034 [Ignavibacterium album JCM 16511]

MEFKKVLGLDIGTNSIGCALLSLPKSIQDYGKGGRLEWLTSRVIPLDADYMKAFIDG
KNGLPQVITPAGKRRQKRGSRRLKIiRYKLRRSRLIRVFKTLNWLPEDFPLDNPKRIK
ETISTEGKPSFRISDYVPISDESYREFYREFGYPENEIEQVIEEINFRRKTKGKNKNPMI
KLLPEDWWYYLRKKALIKPTTKEELIRIIYLFNQRRGFKSSRKDLTETAILDYDEFAK
RLAEKEKYSAE>TYETKFVSITKVKEVVELKTDGRKGKKRFKVILEDSRIEPYEIERKE
KPDWEGKEYTFLVTQKLEKGKFKQNKPDLPKEEDWALCTTALDNRMGSKHPGEFFF
DELLKAFK^KRGYKIRQYPVmWRYKKELEFIWTKQCQLNPELNNLMNKEILP^ A
TVLYPSQSKFFGPKIKEFENSDVLFflISEDIIYYQRDLKSQKSLISECRYEKRKGIDGEIY
GLKCIPKSSPLYQEFRIWQDIHNIKVIRKESE\n^GKKKIMDETQLYINENlKEKLFELF
NSKDSLSEKDILELISLNIINSGIKJSKKEEETTHRINLFANRKELKGNETKSRYRKVFK
KLGFDGEYILNFIPSKLNRLWHSDYSNDYADKEKTEKSILSSLGWKNRNGKWEKSKN
YDVFNLPLEVAKAIANLPPLKKEYGSYSALAIRKMLVVMRDGKYWQHPDQIAKDQE
NTSLIVnLFDKNLIQLTONQRKVLNKYLLTLAEVQKRSTLIKQKLNEIEIIWYKlELVS
DQDLEKQVLKSFLEKKNESDYLKGLKTYQAGYLIYGKHSEKDVPIVNSPDELGEYIR
KKLPNNSLRNPIVEQVIRETIFIVRDVWKSFGIIDEIHIELGRELKNNSEERKKTSESQE
KOTQEKERARXLLKELLNSSNFEHYDENGNXIFSSFTVNPNPDSPLDIEKFRIWKNQS
GLTDEELNKXLKDEKIPTEIEVKKYILWLTQKCRSPYTGKIIPLSKLFDSNVYEIEffllP
RSKMKNDSTNmVICELGVNKAKGDRLAANFISESNGKCKFGEVEYTLLKYGDYLQ
YCKDTFKYQKAKYKNLLATEPPEDFIERQINDTRYIGRKLAELLTPVVKDSKNIIFTIG
SITSELKITWGLNGVWKDILRPRFKRLESIINKKLIFQDEDDPNKYHFDLSINPQLDKE
GLKRLDHRHHALDATnAATTREHVRYLNSLNAADNDEEKREYFLSLCNHKIRDFKL
PWENFTSEVKSICLLSCVVSYKESKPILSDPFNKYLKWEYKNGKWQKVFAIQIKNDR
WKAVRRSMFKEPIGTVWIKKIKEVSLKJEAIKIQAIWEEVKNDPVRKKKEKYIYDDYA
QKVIAKIVQELGLSSSMRKQDDEKLNKF1NEAKVSAGVNKNLNTTOKTIYNLEGRFY
EKIKVAEYVLYKAKilMPLNKKEYIEKLSLQKMFNDLPNFILEKSILDNYPEILKELES
DNKYHEPHKKNNPVNRLLLEffllEYHNNPKEAFSTEGLEKLMCKAINKIGKPIKYITR

LDGDINEEEIFRGAVFETDKGSNVYFVMYENNQTKDREFLKPNPSISVLKAIEHKNKI
DFFAPNRLGFSmiLSPGDLVYVPTNDQYVLIKDNSSNETIINWDDNEFISNRIYQVKK
FTGNSCYFLKNDIASLILSYSASNGVGEFGSQN1SEYSVDDPPIRJKDVCIKIRVDRLGN
VRPL

[0124] YP_213533.1 conserved hypothetical protein [Bacteroides fragilis NCTC 9343]

MKPJLGLDLGTNSIGWALWEAENKDERSSIVKLGVRVNPLTVDELTWEKGKSI^
NADRTLKRGMRRNLQRYKLRRETLTEVLKEHKLITEDTILSENGNRTTFETYRLRAK
AVTEEISLEEFARVLLMINKKRGYKSSRKAKGVEEGTLIDGMDIARELYNNNLTPGEL
CLQLLDAGKKFLPDFYRSDLQNELDRIWEKQKEYYPEILTDVLKEELRGKKRDAVW
AICAKYFVWKEhT/TEWNK^KGKTEQQEREHKLEGIYSKRKRDEAKRENLQWRVNG
LKEKLSLEQLVIVFQEMNTQINNSSGYLGAISDRSKELYFNKQTVGQYQMEMLDKNP
NASLRNMVFYRQDYLDEFNMLWEKQAVYHKELTEELKKEIRDI1IFYQRRLKSQKGL
IGFCEFESRQIEVDIDGKKKJKTVGNRVISRSSPLFQEFKIWQILhWIEVTVVGKKRKRR
KLKENYSALFEELNDAEQLELNGSRRLCQEEKELLAQELFIRDKMTKSEVLKLLFDN
PQELDLNFKTIDGNKTGYALFQAYSKMIEMSGHEPVDFKKPVEKVVEYIKAVFDLLN
WNTDILGFNSNEELDNQPYYKLWHLLYSFEGDNTPTGNGRLIQKMTELYGFEKEYA
Tn¸ANVSFQDDYGSLSAKAIHKILPHLKEGNRYDVACVYAGYRHSESSLTREEIANKV
LKDRLMLLPKNSLHM'VVEKILNQIVrvT^INVnDIYGKPDEIRVELARELKKNAKERE
ELTKSIAQTTKAHEEYKTLLQTEFGLTNVSRTDILRYKLYKELESCGYKTLYSNTYIS
REKLFSK£FDIEfflIPQAPXFDDSFSNKTLEARSVMEKGNKTAYDFVKEKFGESGADN
SLEHYLNMEDLFKSGKISKTKYNKLKMAEQDIPDGFIERDLRNTQYIAKKALSMLNE
ISHRVVATSGSVTDKLREDWQLIDVMKELNWEKYKALGLVEYFEDRDGRQIGRIKD
WTKJmDllRHHAMDALTVAFTKDVFIQYFNNKNASLDPNANEHAIKNKYFQNGRAI
APMPLPvEFRAEAKKHLENTLISIKAKNKVITGMNKTRKKGGVNKNMQQTPRGQLHL
ETIYGSGKQYLTKEEKVNASFDMRKIGTVSKSAYRDALLKRLYENDNDPKKAFAGK
NSLDKQPIWLDKEQMRKVPEKVKIVTLEAIYTIRKEISPDLKVDKVIDVGVRKILIDRL
NEYGNDAKKAFSNLDKNPIWLNKEKGISIKRVTISGISNAQSLFTVKKDKDGKPILDEN
GRN1PVDFVNTGNNHHVAVYYRPVIDKRGQLVVDEAGNPKYELEEVWSFFEAVTR
ANLGLPIIDKDYKTTEGWQFLFSMKQNEYFVFPNEKTGFNPKEIDLLDVENYGLISPN
LFRVQKFSLKNYVFRHHLETTIKDTSSILRGITWIDFRSSKGLDTIVKVRVNHIGQIVS
VGEY

[0125] ZP_10895610.1 CRISPR-associated protein Cas9/Csnl, subtype II/NMEMI [Porphyromonas sp. oral taxon 279 str. F0450]

MLMSKHVLGLDLGVGSIGWCLIALDAQGDPAEILGMGSRVVPLNNATKAIEAFNAG
AAFTASQERTARRTMRRGFARYQLRRYRLRRELEKVGMLPDAALIQLPLLELWELR
ERAATAGRRLTLPELGRVLCHINQKRGYRHVKSDAAAIVGDEGEKKKDSNSAYLAG
IRANDEKLQAEHKTVGQYFAEQLRQNQSESPTGGISYRIKDQIFSRQCYIDEYDQIMA
VQRVHYPDILTDEFIRMLRDEVIFMQRPLKSCKHLVSLCEFEKQERVMRVQQDDGK
GGWQLVERRVKFGPKVAPKSSPLFQLCCIYEAVNNIRLTRPNGSPCDITPEERAKIVA
HLQSSASLSFAALKKLLKEKALIADQLTSKSGLKGNSTRVALASALQPYPQYHHLLD
IVffiLETRMMTVQLTDEETGEVTEREVAVVTDSYVRKPLYRLWHILYSIEEREAMRRA
LITQLGMKEEDLDGGLLDQLYRLDFVKPGYGNKSAKFICKLLPQLQQGLGYSEACA
AVGYRHSNSPTSEEITERTLLEKIPLLQRNELRQPLVEKILNQMINLVNALKAEYGIDE
VRVELARELKMSREERERMARNNKDREERNKGVAAKIRECGLYPTKPRIQKYMLW
KEAGRQCLYCGRSIEEEQCLREGGMEVEHIIPKSVLYDDSYGNKTCACRRCNKEKGN
RTALEYIRAKGREAEYMKRINDLLKEKKISYSKHQRLRWLKEDIPSDFLERQLRLTQ
YISRQAMAILQQGIRRVSASEGGVTARLRSLWGYGKILHTLNLDRYDSMGETERVSR
EGEATEELmTNWSKRMDHRHHAIDALVVACTRQSYIQRLNRLSSEFGREDKKKEDQ
EAQEQQATETGRLSNLERWLTQRPHFSVRTVSDKVAEILISYRPGQRWTRGRNIYR
KKMADGREVSCVQRGVLVPRGELMEASFYGKILSQGRVRJVKRYPLHDLKGEVVDP
HLRELITTYNQELKSREKGAPIPPLCLDKDKKQEVRSVRCYAKTLSLDKAIPMCFDEK
GEPTAFVKSASNFIHLALYRTPKGKLVESIVTFWDAVDRARYGIPLVITHPREVMEQV
LQRGDIPEQVLSLLPPSDWVFVDSLQQDEMVVIGLSDEELQRALEAQNYRKISEHLY
RVQKMSSSYYVFRYHLETSVADDKNTSGRIPKFHRVQSLKAYEERMRKVRVDLLG
RISLL

[0126] ZP_11022414.1 CRISPR-associated protein cas9/csnl, subtype II/nmemi [Barnesiella intestinihominis YIT 11860]

MKN1LGLDLGLSSIGWSVIRENSEEQELVAMGSRVVSLTAAELSSFTQGNGVSINSQR
TQKRTQRKGYDRYQLRRTLLRNKLDTLGMLPDDSLSYLPKLQLWGLRAKAVTQRIE
LNELGRVLLHLNQKRGYKSIKSDFSGDKKITDYVKTVKTRYDELKEMRLTIGELFFR
RLTENAFFRCKEQVYPRQAYVEEFDCIMNCQRKFYPDILTDETIRCIRDEIIYYQRPLK
SCKYLVSRCEFEKRFYLNAAGKKTEAGPKVSPRTSPLFQVCRLWESINNIVVKDRRN
EIVFISAEQRAALFDFLNTHEKLKGSDLLKLLGLSKTYGYRLGEQFKTGIQGNKTRVE

IEPvALGNYPDKXRLLQFNLQEESSSMVNTETGEIIPMISLSFEQEPLYRLWHVLYSIDD
REQLQSVLRQKFGIDDDEVLERLSAIDLVKAGFGNKSSKAIRRILPFLQLGMNYAEAC
EAAGYNHSNNYTKAENEAPvALLDRLPAIKKNELRQPVVEKILNQMVNVVNALMEK
YGRFDEIRVELARELKQSKEERSNTYKSINKNQRENEQ1AKRIVEYGVPTRSRIQKYK
MWEESKHCCIYCGQPVDVGDFLRGFDVEVEHIIPKSLYFDDSFANKVCSCRSCNKEK
NNRTAYDYMKSKGEKALSDYVERWTMYTNNQISKTKWQNLLTPVDKISIDFIDRQ
LRESQYIARKAKEILTSICYNVTATSGSVTSFLRHVWGWDTVLHDLNFDRYKKVGLT
EVIEVNmGSVIRREQIKDWSKJlFDHRHHAIDALTIACTKQAYIQPvLNNLRAEEGPDF
NKMSLERYIQSQPHFSVAQVREAVDRILVSFRAGKRAVTPGKRYIRKNRKRISVQSV
LIPRGALSEESVYGVIHVWEKDEQGHVIQKQRAVMKYPITSINREMLDKEKVVDKRI
HRILSGRLAQYNDNPKEAFAKPVYIDKECRIPIRTVRCFAKPA1NTLVPLKKDDKGNP
VAWVNPGNNHHVAIYRDEDGKYKERTVTFWEAVDRCRVGIPAIVTQPDTIWDNILQ
RhnDISEhTVLESLPDVKWQFVLSLQQNEMFILGMNEEDYRYAMDQQDYALLNKYLY
RVQKLSKSDYSFRYHTETSVEDKYDGKPNLKiSMQMGKLKRVSIKSLLGLNPFIKVH
ISVLGEIKEIS

[0127] ZP_09642280.1 CRISPR-associated protein cas9/csnl, subtype II/nmemi [Odoribacter Ianeus YIT 12061]

METTLGIDLGTOSIGLALVDQEEHQILYSGVmFPEGrNfKDTIGLGEKEESRNATRRAK
RQMRRQYFRKKLRKAKLLELLIAYDMCPLKPEDVRRWKNWDKQQKSTVRQFPDTP
AFREWLKQNPYELRKQAVTEDVTRPELGRILYQMIQRRGFLSSRKGKEEGKIFTGKD
RMVGIDETRKNLQKQTLGAYLYDIAPKNGEKYRFRTERVRARYTLRDMYIREFEIIW
QRQAGHLGLAHEQATRKKNIFLEGSATNYRNSKLITHLQAKYGRGFTVLIEDTRITVT
FQLPLKEVLGGKIEIEEEQLKFKSNESVLFWQRPLRSQKSLLSKCVFEGRNFYDPVHQ
KWIIAGPTPAPLSHPEFEEFRAYQFINNIIYGKNEHLTAIQREAVFELMCTESKDFNFE
KIPKHLKLFEKFNFDDTTKVPACTTISQLRKLFPHPVWEEKREEIWHCFYFYDDNTLL
FEKLQKDYALQTNDLEKJKXIRLSESYGNVSLKAIRRIWYLKKGYAYSTAVLLGGIR
NSFGKRPEWKEYEPEIEKAVCPJLKEKNAEGEVIRKIKDYLVHNRFGFAKNDRAFQK
LYHHSQAITTQAQKERLPETGNLRNPIVQQGLNELRJITVNKLLATCREKYGPSFKFD
fflHVEMGRELRSSKTEREKQSRQIRENEKKNEAAKVKLAEYGLKAYRDNIQKYLLY
KEIEEKGGTVCCPYTGKTLNISHTLGSDNSVQIEHIIPYSISLDDSLANKTLCDATFNRE
KGELTPYDFYQKDPSPEKWGASSWEEIEDRAFRLLPYAKAQRFIRRKPQESNEFISRQ
LM)TRYISKKAVEYLSAICSDVKAFPGQLTAELRHLWGLNNILQSAPDITFPLPVSATE

NHPvEYYVITNEQNEVIRLFPKQGETPRTEKGELLLTGEVERKVFRCKGMQEFQTDVS
DGKYWRRIKLSSSVTWSPLFAPKPISADGQIVLKGRIEKGVFVCNQLKQKLKTGLPD
GSYWISLPVISQTFKEGESVNNSKLTSQQVQLFGRVREGIFRCHNYQCPASGADGNF
WCTLDTDTAQPAFTPIKNAPPGVGGGQIILTGDVDDKGIFHADDDLHYELPASLPKG
KYYGIFTVESCDPTLIPIELSAPKTSKGENLIEGNIWVDEHTGEVRFDPKKNREDQRHH
AIDAIVIALSSQSLFQRLSTYNARRENKKRGLDSTEHFPSPWPGFAQDVRQSVVPLLV
SYKQNPKTLCKISKTLYKDGKKIHSCGNAVRGQLHKETVYGQRTAPGATEKSYHIRK
DIRELKTSKfflGKVVDITIRQMLLKHLQENYmDITQEFMPSNAFFKEGVYRJFLPNKH
GEPVPIKKIRMKEELGNAERLKDMNQYVNPRNNH1WMIYQDADGNLKEEIVSFWSV
IERQNQGQPIYQLPREGRMVSILQINDTFLIGLKEEEPEVYRNDLSTLSKHLYRVQKLS
GMYYTFRHHLASTLNTvlEREEFRIQSLEAWKJRANPVKVQIDEIGRITFLNGPLC

[0128] YP_004843922.1 putative CRISPR-associated (Cas) protein [Flavobacterium branchiophilum FL-15]

MAKILGLDLGTNSIGWAVVEREN1DFSLIDKGVRIFSEGVKSEKGIESSRAAERTGYRS
ARKIKYRRXLRXYETLKVLSLNRMCPLSIEEVEEWKKSGFKDYPLNPEFLKWLSTDE
ESNVNPYFFRDRASKHKVSLFELGRAFYHIAQRRGFLSNRLDQSAEGILEEHCPKIEAI
VEDLISIDEISTNITDYFFETGILDSNEKNGYAKDLDEGDKKLVSLYKSLLAILKKNES
DFENCKSEIIERLNKKDVLGKVKGKIKDISQAMLDGNYKTLGQYFYSLYSKEKIRNQ
YTSREEFIYLSEFITICKVQGIDQINEEEKINEKKFDGLAKDLYKAIFFQRPLKSQKGLIG
KCSFEKSKSRCAISHPDFEEYPJvlWTYLNTIKIGTQSDKKLRFLTQDEKLKLVPKFYRK
NDFNFDVLAKELIEKGSSFGFYKSSKKNDFFYWFNYKPTDTVAACQVAASLKNAIGE
DWKTKSFKYQTrNSNKEQVSRTVDYKDLWHLLTVATSDVYLYEFAIDKLGLDEKNA
KAFSKTKLKJCDFASLSLSAINKILPYLKEGLLYSHAVFVANIENIVDENTWKDEKQRD
YIKTQISEIIENYTLEKSRFEIINGLLKEYKSENEDGKJIVYYSKEAEQSFENDLKKKLV
LFYKSNEIENKEQQETIFNELLPIFIQQLKDYEFIKIQRLDQKVLIFLKGKNETGQIFCTE
EKGTAEEKFKKTKTKNRLKKLYHPSDIEKFKKKIIKDEFGNEKI VLG SPLTP SIKNPMAMR
ALHQLRKVLNALILEGQIDEKTIIfflEMARELNDANKRKGIQDYQNDNKKFREDAIKE
IKKLYFEDCKKEVEPTEDDILRYQLWMEQNRSEIYEEGKNISICDIIGSNPAYDIEHTIP
RSRSQDNSQMNKTLCSQRFNPvEVKKQSMPIELNNHLEILPRIAHWKEEADNLTREIEII
SRSIKAAATKEIKDKXIRRRHYL 'rLKRDYLQGKYDRFIWEEPKVGFKNSQIPDTGIITK
YAQAYLKSYFKKVESVKGGIVrVAEFPvKIWGIQESFIDENGMKFIYKVKDRSKHTHHTI
DAITIACMTKEKYDVLAHAWTLEDQQNKKEARSIIEASKPWKTFKEDLLKIEEEILVS

HYTPDNVKKQAKKIVRVRGKKQFVAEVERDVNGKAVPKKAASGKTIYKLDGEGKK
LPRLQQGDTIRGSLHQDSIYGAIKNPLNTDEIKYVIRKDLESIKGSDVESIVDEVVKEKI
KEAIANKVLLLSSNAQQKNKLVGTVWMNEEKRIAINKVRIYANSVKNPLHIKEHSLL
SKSKHVHKQKVYGQ>TOENYAMAIYELDGKJIDFELIN1FNLAKLIKQGQGFYPLHKK
KEIKGKIVFVPIEKRNKRDVVLKRGQQVVFYDKEVENPKDISE1VDFKGRIYIIEGLSIQ
RIVRPSGKVDEYGVIMLRYFKEARKADDIKQDOTKPDGVFKLGENKPTPvKMNF^QF
TAFVEGIDFKVLPSGKFEKI

[0129] ZP_08837074.1 hypothetical protein HMPREF0666_03250 [Prevotella sp. C561]

MTQKVLGLDLGTNSIGSAVRNLDLSDDLQWQLEFFSSDIFRSSVNKESNGREYSLAA
QRSAHRRSRGLNEVRRRRLWATLNLLIKHGFCPMSSESLMRWCTYDKRKGLFREYP
IDDKDFNAWILLDFNGDGRPDYSSPYQLRRELVTRQFDFEQPIERYKLGRALYfflAQH
RGFKSSKGETLSQQETNSKPSSTDEIPDVAGAMKASEEKLSKGLSTYMKEHNLLTVG
AAFAQLEDEGVRVRNNNDYRAIRSQFQHEIETIFKFQQGLSVESELYERLISEKKNVG
TIFYKRPLRSQRGNYGKCTLERSKPRCAIGHPLFEKFRAWTLINMKVRMSVDTLDEQ
LPMKLRLDLYNECFLAFVRTEFKFEDIRKYLEKRLGIHFSYNDKTINYKDSTSVAGCP
ITARFRKMLGEEWESFRVEGQKERQAHSKNNISFHRVSYSIEDIWFIFCYDAEEPEAVL
AFAQETLRLERKKAEELVRIWSAMPQGYAMLSQKAIRNINKILMLGLKYSDAVILAK
VPELVDVSDEELLSIAKDYYLVEAQVWDKJRINSIVNGLIAKYKSVSEEYRFADHNY
EYLLDESDEKDIIRQIENSLGARRWSLMDANEQTDILQKVRDRYQDFFRSHERKFVES
PKLGESFEWLTKKFPIVrVEREQWKKLYHPSQITIYRPVSVGKDRSVLRLGNPDIGAIK
NPTVLRVLNTLRRRVNQLLDDGVISPDETRVVVETARELNDANRKWALDTYNRIRH
DEhffiKIKKILEEFYPKRDGISTDDIDKARYVIDQREVDYFTGSKTYNKDIKKYKFWLE
QGGQCMYTGRTI^SNLFDPNAFDIEHTIPESLSFDSSDMNLTLCDAim ↓RFIKKNHIP
TDMPNYDKAITIDGKEYPAITSQLQRWVERVERLNRNVEYWKGQARRAQNKDRKD
QCMREMHLWKMELEYWKKKLERFTVTEVTDGFKNSQLVDTRVITRHAVLYLKSIFP
HVDVQRGDVTAKFRKILGIQSVDEKKDRSLHSHHAIDATTLTIIPVSAKRDRMLELFA
KmEINKMLSFSGSEDRTGLlQELEGLKNKLQMEVKVCRIGHNVSEIGTFI NDNIIVNH
mKNQALTPVPxRRLRKKGYIVGGVDNPRWQTGDALRGEIHKASYYGAITQFAKDDE
GKVLMKEGRPQVNPTIKFVIRRELKYKKSAADSGFASWDDLGKAIVDKELFALMKG
QFPAETSFKDACEQGIYMIKKGKNGMPDIKLHHIRUVRCEAPQSGLKIKEQTYKSEKE
YKJIYFYAAVGDLYAMCCYTOGKIREFRIYSLYDVSCHRKSDIEDIPEFITDKKGNRL

MLDYKLRTGDMILLYKDNPAELYDLDNVNLSRRLYKINRFESQSNLVLMTHHLSTS
KERGRSLGKTVDYQNLPESIRSSVKSLNFLIMGENRDFVIKNGKIIFNHR

[0130] ZP_06288774.1 CRISPR-associated protein, Csnl family [Prevotella timonensis CRIS 5C-B1]

MNKRILGLDTGTNSLGWAVVDWDEHAQSYELIKYGDVIFQEGVKIEKGIESSKAAER
SGYKAIRKQYFRRRLRKIQVLKVLVKYHLCPYLSDDDLRQWHLQKQYPKSDELML
WQRTSDEEGKNPYYDRHRCLHEKLDLTVEADRYTLGRALYHLTQRRGFLSNRLDTS
ADNKEDGWKSGISQLSTEMEEAGCEYLGDYFYKLYDAQGNKVRIRQRYTDRMCH
YQHEFDAICEKQELSSELIEDLQRAIFFQLPLKSQRHGVGRCTFERGKPRCADSHPDY
EEFRMLCFVNMQVKGPHDLELRPLTYEEREKIEPLFFRKSKPNFDFEDIAKALAGKK
NYAWIHDKEERAYKFNYRMTQGVPGCPTIAQLKSIFGDDWKTGIAETYTLIQKKNGS
KSLQEMVDDVWAALYSFSSVEKI.KEFA^IHKLQLDEESAEKFAKIKLSHSFAALSLKA
IRKFLPFLRKGIVr^THASFFAMPTIVGKElWNKEQNP^YIMEWGELVFNYQPKHR
EVQGTIEMLIKDFLANNFELPAGATDKLYHPSMIETYPNAQRNEFGILQLGSPRTNAI
RNPMAMRSLmLRRVVNQLLKESIIDENTEVHVEYARELNDANKRRAIADRQKEQD
KQHKKYGDEIRKLYKEETGKDIEPTQTDVLKFQLWEEQNHHCLYTGEQIGITDFIGSN
PKFDIEHTIPQSVGGDSTQMNLTLCDNRFNREVKKAKLPTELANHEEILTRIEPWKNK
YEQLVKERDKQRTFAGMDKAVKDIRIQKRHKLQMEIDYWRGKYERFTMTEVPEGFS
RRQGTGIGLISRYAGLYLKSLFHQADSRNKSNVYVVKGVATAEFRKMWGLQSEYEK
KCRDNHSHHCMDAITIACIGKREYDLMAEYYRMEETFKQGRGSKPKFSKPWATFTE
DVLMYKNLLWHDTPNNMPKHTKKYVQTSIGKVLAQGDTARGSLHLDTYYGAIER
DGEIRYVVRRPLSSFTKPEELEMVDETVKRTIKEAIADKNFKQAIAEPIYMNEEKGILI
KKVRCFAKSVKQPIMRQHRDLSKKEYKQQYFTVMNENNYLLAIYEGLVKNKVVREF
EIVSYIEAAKYYKRSQDRNIFSSIVPTHSTKYGLPLKTKLLMGQLVLMFEENPDEIQV
DNTKDLVKRLYKVVGIEKDGRIKFKYHQEARKEGLPIFSTPYKNNDDYAPIFRQSINN
INILVDGIDFTIDILGKVTLKE

[0131] YP_001875142.1 CRISPR-associated endonuclease Csnl family protein [Elusimicrobium minutum Peil91]

MQKNINTKQNFnYIKQAQKIKEKLGDKPYRIGLDLGVGSIGFAIVSMEENDGNVLLPK
EIIMVGSWFKASAGAADRKLSRGQRNNHRHTRERMRYLWKVLAEQKLALPVPADL
DRKENSSEGETSAKRFLGDVLQKDIYELRVKSLDERLSLQELGYVLYfflAGHRGSSAI
RTFENDSEEAQKENTENKKIAGNIKRLMAKKNYRTYGEYLYKEFFENKEKHKREKIS

NAANNHKFSPTRDLVIKEAEAILKKQAGKDGFHKELTEEYIEKLTKAIGYESEKLIPES
GFCPYLKDEKRLPASHKLNEERRLWETLNNARYSDPIVDIVTGEITGYYEKQFTKEQ
KQKLFDYLLTGSELTPAQTKKLLGLKNTWEDnLQGRDKKAQKIKGYKLIKLESMPF
WARLSEAQQDSFLYDWNSCPDEKLLTEKLSNEYHLTEEEIDNAFNEIVLSSSYAPLGK
SAMLIILEKIKNDLSYTEAVEEALKEGKLTKEKQAIKDRLPYYGAVLQESTQKIIAKG
FSPQFKI)KGYKTPHT>¾YELEYGRIAWVVHQTLNELRKLVNEIIDILGKKPCEIGLET
ARELKKSAEDRSKLSPvEQNDNES NRNRIYEIYIRPQQQV IITRR£ NPRNYILKFELLEE
QKSQCPFCGGQISPNDIINNQADIEHLFPIAESEDNGRNNLVISHSACNADKAKRSPW
AAFASAAKDSKYDYNRILS>TVKEMPHKAWRFNQGAFEKFIENKPMAARFKTDNSYI
SKVAHKYLACLFEKPNIICVKGSLTAQLRMAWGLQGLMIPFAKQLITEKESESFNKD
VNSNKKIRLDhmHHALDAIVIAYASRGYGhnXNKMAGKDY KINYSERNWLSKILLPP
NNIVWENIDADLESFESSVKTALKNAFISVKHDHSDNGELVKGTMYKIFYSERGYTL
TT YKKLSALKLTDPQKKKTPKDFLETAL^ KFKGRESEMKNEKIKSAIENNKRLFDVIQ
DM.EKAKKLLEEE^KSKAEGK^KNINDASIYQKAISLSGDKYVQLSKKEPGKFFAI
SKPTPTTTGYGYDTGDSLCVDLYYDNKGKLCGEIIRKIDAQQKNPLKYKEQGFTLFE
RIYGGDILEVDFDIHSDKNSFRhWTGSAPEmVFIKVGTFTEITONMQIWFGNIIKSTG
GQDDSFTFNSMQQYNPRKLILSSCGFIKYRSPILKNKEG

[0132] YP_004248194.1 CRISPR-associated protein, Csnl family [Sphaerochaeta globosa str. Buddy]

MSKKVSRRYEEQAQEICQRLGSRPYSIGLDLGVGSIGVAVAAYDPIKKQPSDLVFVSS
RIFIPSTGAAERRQKRGQRNSLRHRANRLKFLWKLLAERNLMLSYSEQDVPDPARLR
FEDAVVRANPYELRLKGLNEQLTL SELGY**ALYiD** ANHRG SSS**VRTFLDEEKS SDDKK**
LEEQQAMTEQLAKEKGISTFIEVLTAFNTNGLIGYRNSESVKSKGVPVPTRDIISNEID
VLLQTQKQFYQEILSDEYCDRIVSAILFENEKIVPEAGCCPYFPDEKKLPRCHFLNEER
RLWEAINNARIKMPMQEGAAKRYQSASFSDEQRFflLFHlARSGTDITPKLVQKEFPAL
KTSnVLQGKEK^IQKJAGFRFRRLEEKSFWKRLSEEQKJDDFFSAWTNTPDDKRLSKY
LMKHLLLTENEWDALKTVSLIGDYGP1GKTATQLLMKHLEDGLTYTEALERGMET
GEFQELSVWEQQSLLPYYGQILTGSTQALMGKYWHSAFKEKRDSEGFFKPNTNSDE
EKYGWANPVVHQTLNELRKLMNELITILGAKPQEITVELARELKVGAEKREDIIKQQ
TKQEKEAVLAYSKYCEPNNLDKRY1ERFRLLEDQAFVCPYCLEHISVADIAAGRADV
DfflFPRDDTADNSYGNKVVAHRQCNDIKGKRTPYAAFSNTSAWGPIMHYLDETPGM
WRKRRKFETNEEEYAKYLQSKGFVSRFESDNSYIAKAAKEYLRCLFNPNNVTAVGS

LKGMETSILRKAWM.QGIDDLLGSRHWSKDADTSPTMRKNRDDNRHHGLDAIVAL
YCSRSLVQMINTMSEQGKRAVEIEAMIPIPGYASEPNLSFEAQPvELFPvKKILEFMDLH
AFVSMKTDND ANG ALLKDT VY SILG ADTQGEDL VFVVKKKIKDIG VKIGD YEE VAS
AIRGRITDKQPKWYPMEMKDKIEQLQSKNEAALQKYKESLVQAAAVLEESNRKLIES
GKXPIQLSEKTISKKALELVGGYYYLIS>WKRTKTFVVK£PSNEVKGFAFDTGSNLCL
DFYHDAQGKLCGEnRKIQAMNPSYKPAYMKQGYSLYVRLYQGDVCELRASDLTEA
ESNLAKTTHVRLPNAKPGRTFVIIITFTEMGSGYQIYFSNLAKSKKGQDTSFTLTTIKN
YDVRKVQLSSAGLVRYVSPLLVDKIEKDEVALCGE

[0133] YP_873709.1 HNH endonuclease [Acidothermus cellulolyticus 11B]

MGGSEVGTVPVTWRLGVDVGERSIGLAAVSYEEDKPKEILAAVSWIHDGGVGDERS
GA SPvL ALRGM ARRARRLRRFRRARLRDLDMLLSELG WTPLPDKNV SPVDA WL ARK
RLAEEYVVDETERRRLLGYAVSHMARHRGWRNPWTTIKDLKNLPQPSDSWERTRES
LEARYSVSLEPGTVGQWAGYLLQRAPGIRLNPTQQSAGRRAELSNATAFETRLRQED
VLWELRCIADVQGLPEDVVSNVIDAVFCQKRPSVPAERIGRDPLDPSQLRASRACLEF
QEYRIVAAVANLRIRDGSGSRPLSLEERNAVIEALLAQTERSLTWSDIALEILKLPNES
DLTSVPEEDGPSSLAYSQFAPFDETSARIAEFIAKNRRKIPTFAQWWQEQDRTSRSDL
VAALADNSIAGEEEQELLVHLPDAELEALEGLALPSGRVAYSRLTLSGLTRVMRDDG
VDVFiNARKTCFGVDDNWRPPLPALF£EATGFIPVVDRNLAILRKFLSSATMRWGPPQS
IVVELARGASESRERQAEEEAARRAHRKANDRIRAELRASGLSDPSPADLVRARLLE
LYDCHCMYCGAPISWENSELDfflVPRTDGGSNRHENLAITCGACNXEKGRRPFASW
AETSNRVQLRDVIDRVQKLKYSGNMYWTRDEFSRYKKSVVARLKRRTSDPEVIQSIE
STGYAAVALRDRLLSYGEKNGVAQVAVFRGGVTAEARRWLDISIERLFSRVAIFAQS
TSTKRLDRRHHAVDAVVLTTLTPGVAKTLADARSRRVSAEFWRRPSDVNRHSTEEP
QSPAYRQWKESCSGLGDLLISTAARDSIAVAAPLRLRPTGALHEETLRAFSEHTVGA
AWKGAELRRIVEPEVYAAFLALTDPGGRFLKVSPSEDVLPADENRfflVLSDRVLGPR
DRVKLFPDDRGSIRVRGGAAYIASFHHARVFRWGSSHSPSFALLRVSLADLAVAGLL
RDGVDVFTAELPPWTPAWRYASIALVKAVESGDAKQVGWLVPGDELDFGPEGVTT
AAGDLSMFLKYFPERHWVVTGFEDDKRINLKPAFLSAEQAEVLRTERSDRPDTLTEA
GEILAQFFPRCWRATVAKVLCHPGLTVIRRTALGQPRWRRGHLPYSWRPWSADPWS
GGTP

[0134] ZP_07880770.1 conserved hypothetical protein [Actinomyces sp. oral taxon 180 str. F0310]

MLHCIAVIRVPPSEEPGFFETHADSCALCHHGCMTYAANDKAIRYRVGIDVGLRSIGF
CAVEVDDEDHPIRILNSVVHVHDAGTGGPGETESLRKRSGVAARARRRGRAEKQRL
KKLDVLLEELGWGVSSNELLDSHAPWfflRKRLVSEYIEDETERRQCLSVAMAfflARH
RGWRNSFSKVDTLLLEQAPSDRMQGLKERVEDRTGLQFSEEVTQGELVATLLEHDG
DVTIRGFVRKGGKATKVHGVLEGKYMQSDLVAELRQICRTQRVSETTFEKLVLSIFH
SKEPAPSAARQRERVGLDELQLALDPAAKQPRAERAHPAFQKFKWATLANMRIRE
QSAGERSLTSEELNRVARYLLNHTESESPTWDDVARKLEVPRHRLRGSSRASLETGG
GLTYPPVDDTTVRVMSAEVDWLADWWDCANDESRGHMIDAISNGCGSEPDDVEDE
EVNELISSATAEDMLKLELLAKKLPSGRVAYSLKTLREVTAAILETGDDLSQAITRLY
GVDPGWVPTPAPIEAPVGNPSVDRVLKQVARWLKFASKRWGVPQTVN1EHTREGLK
SASLLEEERERWERFEARREIRQKEMYKRLGISGPFRRSDQVRYEILDLQDCACLYCG
NEINFQTFEVDHIIPRVDASSDSRRTNLAAVCHSCNSAKGGLAFGQWVKRGDCPSGV
SLENAIKRVRSWSKDRLGLTEKAMGKRKSEVISRLKTEMPYEEFDGRSMESVAWMA
IELKKRIEGYFNSDRPEGCAAVQVNAYSGRLTACARRAAHVDKRVRLIRLKGDDGH
HKNRFDRRNHAMDALVIALMTPAIARTIAVREDRREAQQLTRAFESWKNFLGSEER
MQDRWESWIGDVEYACDRLNELIDADKIPVTENLRLRNSGKLHADQPESLKKARRG
SKRPRPQRYVLGDALPADVINRVTDPGLWTALVRAPGFDSQLGLPADLNRGLKLRG
KRISADFPIDYFPTDSPALAVQGGYVGLEFHHARLYR1IGPKEKVKYALLRVCAIDLC
GIDCDDLFEVELKPSSISMRTADAKLKEAMGNGSAKQIGWLVLGDEIQIDPTKFPKQS
IGKFLKECGPVSSWRVSALDTPSKITLKPRLLSNEPLLKTSRVGGHESDLVVAECVEK
IMKKTGWVVEINALCQSGLIRVIRRNALGEVRTSPKSGLPISLNLR

[0135] ZP_03925169.1 conserved hypothetical protein [Actinomyces coleocanis DSM 15436]

MDNKNYRIGIDVGLNSIGFCAVEVDQHDTPLGFLNLSVYRHDAGIDPNGKKTNTTRL
AMSGVARRTRRLFRKPvKPvRLAALDRFIEAQGWTLPDHADYKDPYTPWLVRAELAQ
TPIRDE>iDLHEKLAIAVRfflARHRGWRSPWVPVRSLHVEQPPSDQYLALKERVEAKT
LLQMPEGATPAEMVVALDLSVDWLPvPKNREKTDTRPENKKPGFLGGKLMQSDNA
NELRKIAKIQGLDDALLRELIELVFAADSPKGASGELVGYDVLPGQHGKRRAEKAHP
AFQRYRIASIVSNLRIRHLGSGADERLDVETQKRVFEYLLNAKPTADITWSDVAEEIG
VERNLLMGTATQTADGERASAKPPVDV'1NVAFATCKIKPLKEWWLNADYEARCVM
VSALSHAEKLTEGTAAEVEVAEFLQNLSDEDNEKLDSFSLPIGRAAYSVDSLERLTKR
MIENGEDLFEARVNEFGVSEDWRPPAEPIGARVGNPAVDRVLKAVNRYLMAAEAE

WGAPLSVNIEHYREGFISKRQAVEIDPvENQKRYQRNQAVRSQIADfflNATSGVRGSD
VTRYLAIQRQNGECLYCGTAITFVNSElVroffIVPRAGLGSTOTRDNLVATCERCNKSK
SNKPFAVWAAECGIPGVSVAEALKRVDFWIADGFASSKEHRELQKGVKDRLKRKVS
DPEIDNRSMESVAWMARELAHRVQYYFDEKHTGTKVRVFRGSLTSAARKASGFESR
VNFIGGNGKTRLDRRHHAMDAATVAMLRNSVAKTLVLRGNIRASERAIGAAETWK
SFRGENVADRQIFESWSENMRVLVEKFNLALYNDEVSIFSSLRLQLGNGKAHDDTIT
KLQMHKVGDAWSLTEIDRASTPALWCALTRQPDFTWKDGLPANEDRTIIVNGTHYG
PLDKVGIFGKAAASLLVRGGSVDIGSAIHHARIYRIAGKKPTYGMVRVFAPDLLRYR
NEDLFNVELPPQSVSMRYAEPKVREAIREGKAEYLGWLVVGDELLLDLSSETSGQIA
ELQQDFPGTTHWTVAGFFSPSRLRLRPVYLAQEGL
GEDVSEGSKSIIAGQGWRPAVNKVFGSAMPEVIRRDGLGRKRRFSYSGLPVSWQG

[0136] YP_001955845.1 restriction endonuclease [Bifidobacterium longum DJO10A]
MLSRQLLGASHLARPVSYSYNVQDNDVHCSYGERCFMRGKRYRIGIDVGLNSVGLA
AVEVSDENSPVRLLNAQSVIHDGGVDPQKNKEAITRKNMSGVARRTRRMRRRKRER
LFIKiDMLLGKFGYPVIEPESLDKPFEEWHVRAELATRYIEDDELRRESISIALRHMAR
HRGWRNPYRQVDSLISDNPYSKQYGELKEKAKAYNDDATAAEEESTPAQLVVAML
DAGYAEAPRLRWRTGSKKPDAEGYLPVRLMQEDNANELKQIFRVQRVPADEWKPL
FRSVFYAVSPKGSAEQRVGQDPLAPEQARALKASLAFQEYRIANVITNLRIKDASAEL
RKLTVDEKQSIYDQLVSPSSEDITWSDLCDFLGFKRSQLKGVGSLTEDGEERISSRPPR
LTSVQRIYESDNKIRKPLVAWWKSASDNEHEAMIRLLSNTVDIDKVREDVAYASAIE
FIDGLDDDALTKLDSVDLPSGRAAYSVETLQKLTRQMLTTDDDLHEARKTLFNVTDS
WRPPADPIGEPLGNPSVDRVLKNVNRYLMNCQQRWGNPVSVMEHYRSSFSSVAFA
RKDKREYEKNNEKRSIFRSSLSEQLRADEQMEKVRESDLRRLEAIQRQNGQCLYCGR
TITFRTCEMDfflVPRKGVGSTNTRTNFAAVCAECNRMKSNTPFAIWARSEDAQTRGV
SLAEAKKRVTMFTFNPKSYAPREVKAFKQAVIARLQQTEDDAAIDNRSIESVAWMA
DELHRiaDWYFNAKQYVNSASIDDAEAETMKTTVSVFQGRVTASARRAAGIEGKIHF
IGQQSKTRLDREHHAVDASVIAMMNTAAAQTLMERESLRESQRLIGLMPGERSWKE
YPYEGTSRYESFHLWLDNMDVLLELLNDALDNDRIAVMQSQRYVLGNSIAHDATIH
PLEKVPLGSAMSADLIRRASTPALWCALTRLPDYDEKEGLPEDSHREIRVHDTRYSA
DDEMGFFASQAAQIAVQEGSADIGSAIHHARVYRCWKTNAKGVRKYFYGMIRVFQT
DLLRACHDDLFTVPLPPQSISMRYGEPRVVQALQSGNAQYLGSLWGDEIEMDFSSL

DVDGQIGEYLQFFSQFSGGNLAWKHWVVDGFFNQTQLRIRPRYLAAEGLAKAFSDD
VVPDGVQKIVTKQGWLPPVNTASKTAVRIVRRNAFGEPRLSSAHHMPCSWQWRHE

[0137] YP_001878601.1 hypothetical protein Amuc_2010 [Akkermansia muciniphila ATCC BAA-835]

MSRSLTFSFDIGYASIGWAVIASASHDDADPSVCGCGTVLFPKDDCQAFKRREYRRL
PvRMRSRRVRJERIGRLLVQAQnTPEMKETSGHPAPFYLASEALKGHRTLAPIELWFTV
LRWYAH>¾GYDNNASWSNSLSEDGGNGEDTERVKHAQDLMDKHGTATMAETICR
ELKLEEGKADAPMEVSTPAYKNLNTAFPRLIVEKEVRRILELSAPLIPGLTAEIIELIAQ
HHPLTTEQRGVLLQHGIKLARRYRGSLLFGQLIPRFDNRI1SRCPVTWAQVYEAELKK
GNSEQSARERAEKLSKVPTANCPEFYEYRMARILCNIRADGEPLSAEIRRELMNQAR
QEGKLTKASLEKAISSRLGKETETNVSNYFTLHPDSEEALYLNPAVEVLQRSGIGQILS
PSVYmAANRLRRGKSVTPNYLL^LKSRGESGEALEKKIEKESKKKEADYADTPLK
PKYATGRAPYARTVLKKVVEEILDGEDPTRPARGEAHPDGELKAHDGCLYCLLDTD
SSVNQHQKERRLDTMTNNHLVRHRMLILDRLLKDLIQDFADGQKDRI SRVCVEVGK
ELTTFSAIVTOSKKIQRELTLRQKSHTDAVNRLKRKLPGKALSANLIRKCRIAMDMNW
TCPFTGATYGDHELE^ELEFflVPHSFRQSNALSSLVLTWPGVNRMKGQRTGYDFVE
QEQENPVPDKJNLfflCSLNNYRELVEKLDDKKGHEDDRRRKKKRKALLMVRGLSH
KHQSQNIiEAMKEIGMTEGMMTQSSHLMKLACKSIKTSLPDAfflDMIPGAVTAEVRK
AWDVFGVFKELCPEAADPDSGKILKENLRSLTHLHHALDACVLGLIPYIIPAHHNGLL
RRVLAMRRIPEKLIPQVRPVANQRHYVLNDDGRMMLRDLSASLKENIREQLMEQRV
IQHVPADMGGALLKETMQRVLSVDGSGEDAMVSLSKKKDGKKEKNQVKASKLVG
VFPEGPSKLKALKAAIEIDGNYGVALDPKPVVIRHIKVFKRIMALKEQNGGKPVRILK
KGMLIHLTSSKDPKHAGVWRIESIQDSKGGVKLDLQRAHCAVPKNKTHECNWREVD
LISLLKKYQMKRYPTSYTGTPR

[0138] YP_004168469.1 CRISPR-associated protein, csnl family [Nitratifractor salsuginis DSM 16511]

MKKILGVDLGITSFGYAILQETGKDLYRCLDNSVVMRNNPYDEKSGESSQSIRSTQKS
MRRLIEKRKKRIRCVAQTMERYGILDYSETMKINDPKNNPIKNRWQLRAVDAWKRP
LSPQELFAIFAHMAKHRGYKSIATEDLIYELELELGLNDPEKESEKKADERRQVYNAL
RHLEELRKKYGGETIAQTIHRAVEAGDLRSYRNHDUYEKMIRREDIEEEIEKVLLRQA
ELGALGLPEEQVSELIDELKACITDQEMPTIDESLFGKCTFYKDELAAPAYSYLYDLY
RLYKKLADLNIDGYEVTQEDREKVIEWVEKKIAQGKNLKKITFDCDLRKILGLAPEQK

IFGVEDERIVKGKKEPRTFVPFFFLADIAKFKELFASIQKHPDALQIFRELAEILQRSKT
PQEALDRLRALMAGKGIDTDDRELLELFKNKRSGTRELSHRYILEALPLFLEGYDEKE
VQmLGFDDREDYSRYPKSLRHLHLREGNLFEKEENPrNNHAVKSLASWALGLIADLS
WRYGPFDEHLETTRDALPEKIPJCEIDKAMREPvEKALDKnGKYKKEFPSIDKRLARKI
QLWERQKGLDLYSGKVINLSQLLDGSADIEHIVPQSLGGLSTDYNTIVTLKSVNAAK
GNRLPGDWLAGNPDYREPJGMLSEKGLIDWKKRKNLLAQSLDEIYTENTHSKGIRAT
SYLEALVAQVLKRYYPFPDPELRKNGIGVRMIPGKVTSKTRSLLGIKSKSRETNFHHA
EDALILSTLTRGWQNRLHRMLRDNYGKSEAELKELWKKYMPHIEGLTLADYIDEAF
RRFMSKGEESLFYPvDMFDTIRSISYWVDKKPLSASSFiKETVYSSRHEVPTLRKNILEA
FDSLNVIKDRFIKLTTEEFMKRYDKEIRQKLWLIimGNTNDESYRAVEERATQIAQILT
RYQLMDAQNDKEIDEKJQQALKELITSPIEVTGKLLRKMRPVYDKLNAMQIDRGLV
ETDK>MLGIfflSKGPNEKLIFRRMDV>WAHELQKERSGILCYL **Nf MLFlFNKKGLIFTY**
GCLRSYLEKGQGSKYIALFNPRFPANPKAQPSKFTSDSKIKQVGIGSATGIIKAHLDLD
GHVRSYEVFGTLPEGSIEWFKEESGYGRVEDDPHH

[0139] ZP_08015909.1 hypothetical protein HMPREF9464_01128 [Sutterella wadsworthensis 3_1_45B]

MTQSERRFSCSIGIDMGAKYTGVFYALFDREELPTNLNSKAMTLVMPETGPRYVQA
QRTAVRHRLRGQKRYTLARKLAFLVVDDMIKKQEKRLTDEEWKRGREALSGLLKR
RGYSRPNADGEDLTPLENVRADVFAAHPAFSTYFSEVRSLAEQWEEFTAN1SNVEKF
LGDPN1PADKEFIEFAVAEGLIDKTEKKAYQSALSTLRANANVLTGLRQMGHKPRSE
WKAIEADLKKDSRLAKINEAFGGAERLARLLGNLSNLQLRAERWYFNAPDIMKDR
GWEPDRFKXTLVRAFKFFHPAKDQNKQHLELIKQIENSEDIIETLCTLDPNRTIPPYED
QNNRRPPLDQTLLLSPEKLTRQYGEIWKTWSARLTSAEPTLAPAAEILERSTDRKSRV
AVNGHEPLPTLAYQLSYALQRAFDRSKALDPYALRALAAGSKSNKLTSARTALENCI
GGQNVKTFLDCARRYYREADDAKVGLWFDNADGLLERSDLHPPMKKKILPLLVAN1
LQTDETTGQKFLDEIWRKQIKGRETVASRCARIETVRKSFGGGFNIAYNTAQYREVN
KLPRNAQDKELLTIRDRVAETADFIAANLGLSDEQKRKFANPFSLAQFYTLIETEVSG
FSATTLAVFILENAWRMTIKDAVINGETVRAAQCSRLPAETARPFDGLVRRLVDRQA
WEIAKRVSTDIQSKVDFSNGIVDVSIFVEENKFEFSASVADLKKNKRVKDKMLSEAE
KLETRWLIKNERIKKASRGTCPYTGDRLAEGGEIDHILPRSLIKDARGIVFNAEPNLIY
ASSRGNQLKKNQRYSLSDLKANYRNEIFKTSN1AAITAEIEDVVTKLQQTHRLKFFDL
LNEHEQDCVRHALFLDDGSEARDAVLELLATQRRTRVNGTQIWMIKNLANKIREEL

QNWCKTTNN^HFQAAATNYSDAKNLRLKLAQNQPDFEKPDIQPIASHSIDALCSFA
VGSADAERDQNGFDYLDGKTVLGLYPQSCEVIHLQAKPQEEKSHFDSVAIFKEGIYA
EQFLPIFTLNEKIWIGYETLNAKGERCGAIEVSGKQPKELLEMLAPFFNKPVGDLSAH
ATYRILKKPAYEFLAKAALQPLSAEEKRLAALLDALRYCTSRKSLMSLFMAANGKSL
KKREDVLKPKLFQLKVELKGEKSFKLNGSLTLPVKQDWLRICDSPELADAFGKPCSA
DELTSKIAPJWKRPVMRDLAHAPVRREFSLPAIDNPSGGFRIRRTNLFGNELYQVHAI
NAKKYRGFASAGSNVDWSKGILFNELQHENLTECGGRFITSADVTPMSEWRKVVAE
DNLSIWIAPGTEGRRYVRVETTFIQASHWFEQSVENWAITSPLSLPASFKVDKPAEFQ
KAVGTELSELLGQPRSEIFIENVGNAKfflRFWIVVSSNKJKMNESYN^ SKS

[0140] J7RUA5.1 CRISPR-associated endonuclease Cas9 [Staphylococcus aureus]

MKRNYILGLDIGITSVGYGUDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKR
RRRHWQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRR
GVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKT
SDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEW
YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIEN
VFKQKXKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTT^TLKVYPIDIKDITARKEIIENA
ELLDQIAKILTIYQSSEDIQEELimNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE
LWHTNDNQIAIFNPvLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK
KYGLPNDniELAREKNSKDAQmiNEMQKRhniQTOEPJEEnRTTGKENAKYLIEKIK
LHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKK
GNRTPFQYLSSSDSKJSYETFKKfflLNLAKGKGPJSKTKKEYLLEERDINRFSVQKDFI
NRNLVDTRYATRGLMNLLRSYFRVN^DVKVK^ INGGFTSFLRRKWKFKKERNKG
YKHHAEDÀLIIANADFIFKEWKKLDKAKXVMENQMFEEKQAESMPEIETEQEYKEIF
ITPHQIKFnKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDK
DhTOKiKJCLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTK
YSKKDNGPVIKXIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVY
KFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRV
IGVNNDLLNRIEVNMIDITYREYLENMNDKRPPPJIKTIASKTQSIKKY STDILGNLYE
VKSKKHPQIIKKG

[0141] AEX66236.1 CRISPR-associated endonuclease [Corynebacterium diphtheriae C7 (beta)]

MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSfflHDSGLDPDKIKSAVTRLASSG
IAPJITRRLYRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVRAELAASYIADE
KERGEKLSVALRHIARHRGWRNPYAKVSSLYLPDEPSDAFKAIREEIKRASGQPVPET
ATVGQMVTLCELGTLKLRGEGGVLSARLQQSDHAREIQEICRMQEIGQELYRKIIDV
VFAAESPKGSASSRVGKDPLQPGKNRALKASDAFQRYRIAALIGNLRVRVDGEKRIL
SVEEKNLVFDHLVNLAPKKEPEWVTIAEILGIDRGQLIGTATMTDDGERAGARPPTH
DTNRSIVNSRIAPLVDWWKTASALEQHAMVKALSNAEVDDFDSPEGAKVQAFFADL
DDDVHAKLDSLHLPVGRAAYSEDTLVRLTRRMLADGVDLYTARLQEFGIEPSWTPP
APRIGEPVGNPAVDRVLKTVSRWLESATKTWGAPERVIIEHVREGFVTEKRAREMDG
DMRRRAARNAKLFQEMQEKLNVQGKPSRADLWRYQSVQRQNCQCAYCGSPITFSN
SEMDfflVPRAGQGSTNTRENLVAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAVER
TRHWVTDTGMRSTDFKKFTKAVVERFQRATMDEEIDARSMESVAWMANELRSRVA
QHFASHGTTVRVYRGSLTAEARRASGISGKLEFLDGVGKSRLDRRHHAIDAAVIAFT
SDYVAETLAVRSNLKQSQAHRQEAPQWREFTGKDAEHRAAWRVWCQKMEKLSAL
LTEDLRDDRVVVMSNVRLRLGNGSAHEETIGKLSKVKLGSQLSVSDIDKASSEALWC
ALTREPDFDPKDGLPANPERfflRVNGTHVYAGDMGLFPVSAGSIALRGGYAELGSSF
HHARVYKITSGKXPAFAIVnRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKLRDA
LATGNAEYLGWLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVDGFFGDTRLRL
RPLQMSKEGIKXESAPELSKIIDRPGWLPAVNKLFSEGNVTVVRRDSLGRVRLESTAH
LPVTWKVQ

[0142] WP_013852048.1 type II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pasteurianus]

MTNGKILGLDIGIASVGVGIIEAKTGKVVHANSRLFSAANAENNAERRGFRGSRRLN
RRXKHRVKRVRDLFEKYGIVTDFRNLNLNPYELRVKGLTEQLKNEELFAALRTISKR
RGISYLDDAEDDSTGSTDYAKSIDENRRLLKNKTPGQIQLERLEKYGQLRGNFTVYD
ENGEAHRLINVFSTSDYEKEARKJLETQADYNKKITAEFIDDYVEILTQKRKYYHGPG
NEKSRTDYGRFRTDGTrLENIFGILIGKCNFYPDEYRASKASYTAQEYNFLNDLNM,K
VSTETGKLSTEQKESLVEFAKNTATLGPAKLLKEIAKILDCKVDEIKGYREDDKGKPD
LHTFEPYRKLKFNLESINIDDLSREVIDKLADILTLNTEREGIEDAIKRNLPNQFTEEQIS

EIIKVRKSQSTAFNKGWHSFSAKLMNELIPELYATSDEQMTILTRLEKFKVNKKSSKN
TKTIDEKEVTDEIYNPVVAKSVRQTIKJrNAAVKKYGDFDKIVIEMPRDKNADDEKKF
IDKRNl^NKKEKDDALKRAAYLYNSSDKLPDEVFHGNKQLETKIRLWYQQGERCLY
SGmSIQELVHNSNNFEIDHILPLSLSFDDSLANKVLVYAWTNQEKGQKTPYQVIDS
MDAAWSFREMKDYVLKQKGLGKKKRDYLLTTENIDKIEVKKKFIERNLVDTRYASR
VVLNSLQSALRELGKDTKVSVVRGQFTSQLRRKWKIDKSRETYHHHAVDALIIAASS
QLKLWEKQDNPMFVDYGKNQVVDKQTGEILSVSDDEYKELVFQPPYQGFVNTISSK
GFEDEILFSYQVDSKYNRKVSDATIYSTRKAK1GKDKKEETYVLGKIKD1YSQNGFDT
FIKKYNKDKTQFLMYQKDSLTWENVIEVILRDYPTTKKSEDGKNDVKCNPFEEYRRE
NGLICKYSKKGKGTPIKSLKYYDKKLGNCIDITPEESRNKVILQSINPWRADVYFNPE
TLKYELMGLKYSDLSFEKGTGNYHISQEKYDAIKEKEG1GKKSEFKFTLYRNDLILIK
DIASGEQEIYRFLSRTMPNVNHYVELKPYDKEKFDNVQELVEALGEADKVGRCIKGL
NKPN1SIYKVRTDVLGNKYFVKKKGDKPKLDFKNNKK

[0143] EEZ71796.1 CRISPR-associated protein, Csnl family [Neisseria cinerea ATCC 14685]

MAAFKPNPMNYILGLDIGIASVGWAIVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLA
AARRLARSVRilLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRA
AALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNTHALQT
GDFRTPAELALNKFEKESGfflP^QRGDYSHTFNRKDLQAELNLLFEKQKEFGNPHVS
DGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPTEPKAAKNTYTAERFVWLTKLN
NLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLDLDDTAFFKGLRYGKDN
AEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRL
KDRVQPEILEALLKfflSFDKFVQISLKALRRIVPLMEQGNRYDEACTEIYGDHYGKKN
TEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDR
KEIEKRQEENRKDREKSAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEIN
LGRLNEKGYVEIDHALPFSRTWDDSFNNKVLALGSENQNKGNQTPYEYFNGKDNSR
EWQEFKARVETSRFPRSKKQmLLQKFDEDGFKJERNLhTOTRYINRFLCQFVADHMLL
TGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTIAMQQKITR
FVRYKEMNAFDGKTIDKETGEVLHQKAHFPQPWEFFAQEVMIRVFGKPDGKPEFEE
ADTPEKLRTLLAEKLSSRPEAVHKYVTPLFISRAPNRKMSGQGFIMETVKSAKRLDEG
ISVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKYDK
AGNRTQQVKAVRVEQVQKTGVWVHNHNGIADNATIVRVDVFEKGGKYYLVPIYS

WQVAKGILPDRAVVQGKDEEDWTVMDDSFEFKFVLYANDLIKLTAKKNEFLGYFV
SLNRATGAIDIRTHDTDSTKGKNGIFQSVGVKTALSFQKYQIDELGKEIRPCRLKKRPP
VR

[0144] BAK69486.1 putative CRISPR associated protein [Campylobacter lari]

MRILGFDIGINSIGWAFVENDELKDCGVRIFTKAENPKNKESLALPRRNARSSRRRLK
RRKARLIAIKRILAKELKLNYKDYVAADGELPKAYEGSLASVYELRYKALTQNLETK
DLARVILHIAKFTRGYMNKNEKKSNDAKKGKILSALKNNALKLENYQSVGEYFYKEF
FQKYKKNTKNFIKIRNTKDNYNNCVLSSDLEKELKLILEKQKEFGYNYSEDFINEILK
VAFFQRPLKDFSF1LVGACTFFEEEKRACKNSYSAWEFVALTKIINEIKSLEKISGEIVP
TQTINEVLNLILDKGSITYKKFRSCINLHESISFKSLKYDKENAENAKLIDFRKLVEFK
KALGVHSLSRQELDQISTHITLIKDNVKLKTVLEKYNLSNEQINNLLEIEFNDYINLSF
KALGMILPLMREGKRYDEACEIANLKPKTVDEKKDFLPAFCDSIFAFEELSNPWNRAI
SEYRKVLNALLKKYGKVHKIHLELARDVGLSKKAREKIEKEQKENQAVNAWALKE
CEMGLKASAKMLKLKlWKEQKEICIYSGNKISIEHLKDEKALEVDFnYPYSRSFDDS
FINKVLVFTKENQEKL>¾TPFEAFGKMEKWSKIQTLAQNLPYKKKNKILDENFKDK
QQEDFISRNLNDTRYIATLIAKYTKEYLNFLLLSENENANLKSGEKGSKIHVQTISGM
LTSVLRHTWGFDKKDRNNHLHHALDAIIVAYSTNSIIKAFSDFRKNQELLKARFYAK
ELTSDNYKHQVKFFEPFKSFREKILSKIDEIFVSKPPRKRARRALHKDTFHSENXIIDK
CSYNSKEGLQIALSCGRVRKIGTKYVENDTIVRVDIFKKQNKFYAIPIYAMDFALGILP
NKIVITGKDKNNNPKQWQTI^ ESYEFCFSLYKNDLILLQKKNMQEPEFAYYNDFSIST
SSICVEKHDNKPENLTSNQKLLFSNAKEGSVKVESLGIQNLKVFEKYIITPLGDKIKAD
FQPRENISLKTSKKYGLR

[0145] OJI07263.1 hypothetical protein BK997_03320 [Candidatus Micrarchaeum acidiphiium ARMAN-1]

MRDSITAPRYSSALAARIKEFNSAFKLGIDLGTKTGGVALVKDNKVLLAKTFLDYHK
QTLEERRJHRRNRRSRLARRKRIARLRSWILRQKIYGKQLPDPYKIKKMQLPNGVRK
GENWIDLVVSGRDLSPEAFVRAITLIFQKRGQRYEEVAKEIEEMSYKEFSTfflKALTS
VTF.F.EFTALAAEIERRQDVVDTDKEAERYTQLSELLSKVSESKSESKDRAQRKEDLG
KVVNAFCSAHRIEDKDKWCKELMKLLDRPVRHARFLNKVLIRCNICDRATPKKSRP
DVRELLYFDTVRNFLKAGRVEQNPDVISYYKKIYMDAEVIRVKILNKEKLTDEDKKQ
KRKLASELNRYKNKEWTDAQKKMQEQLKTLLFMKLTGRSRYCMAHLKERAAGK
DVEEGLHGVVQKRHDRNIAQRNHDLRVINLIESLLFDQNKSLSDAIRKNGLMYVTIE

APEPKTKHAKKGAAVVRDPRKLKEKLFDDQNGVCIYTGLQLDKLEISKYEKDfflFPD
SPJ^GPSIRDNLVLTrKEINSDKGDRTPWEWMHDNPEKWKAFERRVAEFYKKGRINE
RKRELLLNKGTEYPGDNPTELARGGARVNNFITEFNDRLKTHGVQELQTIFERNKPIV
QVVRGEETQRLRRQWNALNQNFIPLKDRAMSFNHAEDAAIAASMPPKFWREQIYRT
AWHFGPSGNERPDFALAELAPQWNDFFMTKGGPIIAVLGKTKYSWKHSIIDDTIYKP
FSKSAYYVGIYKKPNAITSNAIKVLRPKLLNGEHTMSKNAKYYHQKIGNERFLMKSQ
KGGSnTVKPHDGPEKVLQISPTYECAVLTKHDGKIIVKFKPIKPLRDMYARGVIKAM
DKELETSLSSMSK^AKYKELHTHDnYLPATKKHVDGYFIITKLSAKHGIKALPESMV
KVKYTQIGSENNSEVKLTKPKPEITLDSEDITNIYNPTR

[0146] APG80630.1 CRISPR-associated endonuclease Cas9 [Candidatus Parvarchaeum acidiphilum ARMAN-4]

MLGSSRYLRY^TSFEGKEIPFLIMGYYKEYNKELSSKAQKEFNDQISEFNSYYKLGID
LGDKTGIAIVKGNKIILAKTLIDLHSQKLDKRJIEARJINRRTRLSRXKRLARLRSWVM
RQKVGNQRLPDPYKIMiiDNKYWSIYNKSNSANKKNWIDLLIHSNSLSADDFVRGLTI
IFRKRGYLAFKYLSRLSDKEFEKYIDNLKPPISKYEYDEDLEELSSRVENGEIEEKKFE
GLKNKLDKIDKESKDFQVKQREEVKKELEDLVDLFAKSVDNKIDKARWKRELNNLL
DKKVRKIRFDNRFILKCKIKGCNKNTPKXEKVPJ)FELKMVLNNARSDYQISDEDLNS
FRNEVINIFQKK£NLKKGELKGVTIEDLRKQLNKTFNKAKIKKGIREQIRSIVFEKISGR
SKfCKEHLKEFSEKPAPSDPvIWGVNSAREQHDFRVLWIDKKIFKDKLIDPSKLRYITI
ESPEPETEKLEKGQISEKSFETLKEKLAKETGGIDIYTGEKLKKDFEIEHIFPRARMGPS
IRENEVASNLETTSTKEKADRTPWEWFGQDEKRWSEFEKRVNSLYSKKKISERKREILL
NKSNEYPGLNPTELSRIPSTLSDFVESIRKMFVKYGYEEPQTLVQKGKPIIQVVRGRDT
QALRWRWHALDSNIIPEKDRKSSFNHAEDAVIAACMPPYYLRQKIFREEAKIKRKVS
MCEKEVTRPDIVffTKJKJAPNWSEFMKTRNEPVIEVIGKVKPSWKNSIMDQTFYKYLLK
PFKDNLIKIPNVKNTYKWIGVNGQTDSLSLPSKVLSISNKKVDSSTVLLVHDKKGGK
PxNWVPKSIGGLLVYITPKDGPKRIVQVKPATQGLLIYPxNEDGRVDAVREFINPVIEMY
NNGKLAFVEKENEEELLKYFNLLEKGQKFERIRRYDMITYNSKFYYVTKINKNHRVT
IQEESKIKAESDKVKSSSGKEYTRKETEELSLQKLAELISI

[0147] tr|I0AP30|I0AP30_IGNAJ CRISPR-associated endonuclease Cas9 OS=Ignavibacterium album (strain DSM 19864 / JCM 16511 / NBRC 101810 / Mat9-16) OX=945713 GN=cas9 PE=3 SV=1

MEFKKVLGLDIGTNSIGCALLSLPKSIQDYGKGGPvLEWLTSRVIPLDADYMKAFIDG
KNGLPQVITPAGKRRQKRGSRRLKHRYKLRRSRLIRVFKTLNWLPEDFPLDNPKPvIK
ETISTEGKFSFRISDYVPISDESYREFYREFGYPENEIEQVIEEINFRRKTKGKNKNPMI
KLLPEDWVVYYLRKKALIKPTTKEELIRIIYLFNQRRGFKSSRKDLTETAILDYDEFAK
RLAEKEKYSAENYETKFVSITKVKEVVELKTDGRKGKKRFKVILEDSRIEPYEIERKE
KPDWEGKEYTFLVTQKLEKGKFKQNKPDLPKEEDWALCTTALDNRMGSKHPGEFFF
DELLKAFKEKRGYKIRQYPX^WRYKKELEFIWTKQCQLNPELNNLNINKEILRKLA
TVLYPSQSKFFGPKIKEFENSDVLHIISEDIIYYQRDLKSQKSLISECRYEKRKGIDGEIY
GLKCIPKSSPLYQEFRIWQDIHMKVIRKESEWGKKKT^DETQLYINEMKEKLFELF
NSKDSLSEKDILELISLMINSGIKISKKEEETTHRINLFANRKELKGNETKSRYRKVFK
KLGFDGEYILNFIPSKLNRLWHSDYSNDYADKEKTEKSILSSLGWKNRNGKWEKSKN
YDVFM.PLEVAKAIA NLPPLKKEYGSYSALAIRXMLWMPJ)GKYWQHPDQIAKDQE
NTSLMLFDKNLIQLTNNQRKVLNKYLLTLAEVQKRSTLIKQKLNEIEHNPYKLELVS
DQDLEKQVLKSFLEKKNESDYLKGLKTYQAGYLIYGKHSEKDVPIVNSPDELGEYIR
KKLPhWSLRNPIVEQVIRETIFIVRDVWKSFGIIDEIHIELGRELKNNSEERKKTSESQE
KNFQEKERARKLLKELLNSSNFEHYDENGNKIFSSFTVNPNPDSPLDIEKFPJWKNQS
GLTDEELNKKLKDEKIPTEIEVKKYILWLTQKCRSPYTGKIIPLSKLFDSNVYEIEHIIP
RSKMKNDSTNI^VICELGVNKAKGDRLAANFISESNGKCKFGEVEYTLLKYGDYLQ
YCKDTFKYQKAKYKNLLATEPPEDFIERQINDTRYIGRKLAELLTPVVKDSKNIIFTIG
SITSELKJTWGLNGVWKDILRPRJKJRLESIINKKLIFQDEDDPNKYHFDLSrKPQLDK^
GLKRLDHRHHALDATIIAATTREHVRYLNSLNAADNDEEKREYFLSLCNHKIRDFKL
PWENFTSEVKSKLLSCVVSYKESKPILSDPFNKYLKWEYKNGKWQKVFAIQIKNDR
WKAVRRSMFKEPIGTVWIKKIKEVSLKEAIKIQAIWEEVKNDPVRXKKEKYIYDDYA
QKVIAKIVQELGLSSSMRKQDDEKLNKFINEAKVSAGVNKNLNTTNKTIYNLEGRFY
EKIKVAEYVLYKAKRMPLNKKEYIEKLSLQKMFNDLPNFILEKSILDNYPEILKELES
DNKYIIEPHKKNNPVT4RLLLEFnLEYHNNPKEAFSTEGLEKLN^ AINKIGKPIKYITR
LDGDINEEEIFRGAVFETDKGSNVYFVMYENNQTKDREFLKPNPSISVLKAIEFIKNKI
DFFAPNRLGFSRIILSPGDLVYVPTNDQYVLIKDNSSNETIINWDDNEFISNRIYQVKK
FTGNSCYFLKM)IASLILSYSASNGVGEFGSQNISEYSVDDPPIRIKDVCIKIRVDRLGN
VRPL

[0148] Ga0054994_10813 Geobacillus stearothermophilus Cas9

MRYKIGLDIGITSVGWAVMNLDIPPJEDLGVPJFDRAEWQTGESLALPRRLARSAPvR
RLRPvRXHRLERIRRLVIPJEGILTKEELDKLFEEKJffilDVWQLRVEALDRKLNNDELAR
VLLHLAKRRGFKSNPvKSERSNKENSTMLKfflEENRAILSSYRTVGEMIVKDPKFALH
KRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEEFENEYITIWASQRPVASKD
DIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHINKLRLISPSGARGLTDEERRLLYEQ
AFQKNKITYHDIRTLLHLPDDTYFKGIVYDRGESRKQNENIRFLELDAYHQIPvKAVDK
VYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLANKVYDN
ELIEELLNLSFTKFGiiLSLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLL
PNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIfflELARDLSQTFDERRKTKKEQDE
NRKKNETAIRQLMEYGLTLNPTGHDIVKEKLWSEQNGRCAYSLQPIEIERLLEPGYVE
VDHVIPYSRSLDDSYTNKVLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQ
FSKXKRDRLLRLFIYDENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQKVY
T\nsfGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFYQRREQNK
ELAKKTEPHFPQPWPiiFADELRARLSKHPKESIKALNLGNYDDQKLESLQPVFVSRM
PKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDASGHFPMYGKESDPRT
YEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPV1RTVKIIDTKNQVIPLNDGKTVA
YNSMVRVDVFEKDGKYYCVPVYTMD1MKGILPNKAIEPNKPYSEWKEMTEDYTFR
FSLYPNDLIRIELPREKTVKTAAGEEINVKDVFVYYKTIDSANGGLELISHDHRFSLRG
VGSRTLKRFEKYQVDVLGNIYKVRGEKRVGLASSAHSKPGKTIRPLQSTRD

[0149] WP_036475267.1 type II CRISPR RNA-guided endonuclease Cas9 [Neisseria lactamica]

MAAFKPNPMNYILGLDIGIASVGWAMVEVDEEENPIRLIDLGVRVFERAEVPKTGDS
LAMARPXARSVRRLTRJIRAHRLLRARRLLKREGVLQDADFDENGLVKSLPNTPWQ
LRAAALDRKLTCLEWSAVLLHLVKHRGYLSQRKNEGETADKELGALLKGVADNAH
ALQTGDFRTPAELALNKFEKESGfflRNQRGDYSHTFSRKDLQAELNLLFEKQKEFGN
PHVSDGLKEDIETLLMAQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWL
TKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRY
GKDNAEASTLMEMKAYHAJSRALEKEGLKDKKSPLNLSTELQDEIGTAFSLFKTDKD
ITGRLKDRVQPEILEALLKJilSFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDH
YCKKNAEEKIYLPPIPADEIRNPVVLRALSQARKVINCVVRRYGSPARIfflETAREVGK

SFKDRXEIEKRQEENPJO)PvEKAAAKFPvEYFPNFVGEPKSKDILKLRLYEQQHGKCLY
SGKEIHLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFN
GKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEEGFKERNL>JDTRYVNRFLCQFV
ADHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRTENDRHHALDAVVVACSTVA
MQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKAHFPQPWEFFAQEVMIRVFGKPD
GKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKS
AKJU.DEGISVLRVPLTQLKLKGLEKMVNREREPKLYDALKAQLETF1KDDPAKAFAE
PFYKYDKAGSRTQQVKAVRIEQVQKTGVWVRNFfNGIADNATMVRVDVFEKGGKY
YLVPIYSWQVAKGILPDRAVVAFKDEEDWTVMDDSFEFRFVLYANDLIKLTAKKNE
FLGYFVSLNRATGAIDIRTHDTDSTKGKNGIFQSVGVKTALSFQKNQIDELGKEIRPC
RLKKRPPVR

[0150] The term "cell" as used herein may refer to either a prokaryotic or eukaryotic cell, optionally obtained from a subject or a commercially available source.

[0151] As used herein, the term "CRISPR" refers to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR). CRISPR may also refer to a technique or system of sequence-specific genetic manipulation relying on the CRISPR pathway. A CRISPR recombinant expression system can be programmed to cleave a target polynucleotide using a CRISPR endonuclease and a guideRNA. A CRISPR system can be used to cause double stranded or single stranded breaks in a target polynucleotide. A CRISPR system can also be used to recruit proteins or label a target polynucleotide. In some aspects, CRISPR-mediated gene editing utilizes the pathways of nonhomologous end-joining (NHEJ) or homologous recombination to perform the edits. These applications of CRISPR technology are known and widely practiced in the art. *See,* e.g., U.S. Pat. No. 8,697,359 and Hsu et al. (2014) *Cell* 156(6): 1262-1278.

[0152] As used herein, the term "comprising" is intended to mean that the compositions and methods include the recited elements, but do not exclude others. As used herein, the transitional phrase "consisting essentially of (and grammatical variants) is to be interpreted as encompassing the recited materials or steps "and those that do not materially affect the basic and novel characteristic(s)" of the recited embodiment. *See, In re Herz,* 537 F.2d 549, 551-52, 190 U.S.P.Q. 461, 463 (CCPA 1976) (emphasis in the original); *see also* MPEP § 2111.03. Thus, the term "consisting essentially of as used herein should not be interpreted as

equivalent to "comprising." "Consisting of shall mean excluding more than trace elements of other ingredients and substantial method steps for administering the compositions disclosed herein. Aspects defined by each of these transition terms are within the scope of the present disclosure.

[0153] The term "encode" as it is applied to nucleic acid sequences refers to a polynucleotide which is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, can be transcribed and/or translated to produce the mRNA for the polypeptide and/or a fragment thereof. The antisense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

[0154] The terms "equivalent" or "biological equivalent" are used interchangeably when referring to a particular molecule, biological, or cellular material and intend those having minimal homology while still maintaining desired structure or functionality.

[0155] As used herein, the term "expression" refers to the process by which polynucleotides are transcribed into mRNA and/or the process by which the transcribed mRNA is subsequently being translated into peptides, polypeptides, or proteins. If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell. The expression level of a gene may be determined by measuring the amount of mRNA or protein in a cell or tissue sample; further, the expression level of multiple genes can be determined to establish an expression profile for a particular sample.

[0156] As used herein, the term "functional" may be used to modify any molecule, biological, or cellular material to intend that it accomplishes a particular, specified effect.

[0157] The term "gRNA" or "guide RNA" as used herein refers to the guide RNA sequences used to target specific genes for correction employing the CRISPR technique. Techniques of designing gRNAs and donor therapeutic polynucleotides for target specificity are well known in the art. For example, Doench, J., et al. Nature biotechnology 2014; 32(12): 1262-7, Mohr, S. et al. (2016) FEBS Journal 283: 3232-38, and Graham, D., et al. Genome Biol. 201 5; 16: 260. gRNA comprises or alternatively consists essentially of, or yet further consists of a fusion polynucleotide comprising CRISPR RNA (crRNA) and trans-activating CRIPSPR RNA (tracrRNA); or a polynucleotide comprising CRISPR RNA

(crRNA) and trans-activating CRIPSPR RNA (tracrRNA). In some aspects, a gRNA is synthetic (Kelley, M. et al. (2016) J of Biotechnology 233 (2016) 74-83).

[0158] As used herein, the term "immune orthogonal" refers to a lack of immune cross-reactivity between two or more antigens. In some embodiments, the antigens are proteins (e.g., Cas9). In some embodiments, the antigens are viruses (e.g., AAV). In some embodiments, antigens that are immune orthogonal do not share an amino acid sequence of greater than 5, greater than 6, greater than 7, greater than 8, greater than 9, greater than 10, greater than 11, greater than 12, greater than 13, greater than 14, greater than 15, or greater than 16 consecutive amino acids. In some embodiments, antigens that are immune orthogonal do not share any highly immunogenic peptides. In some embodiments, antigens that are immune orthogonal do not share affinity for a major histocompatibility complex (e.g., MHC class I or class 11). Antigens that are immune orthogonal are amenable for sequential dosing to evade a host immune system.

[0159] The term "immunosilent" refers to an antigen that does not elicit an immune response from a host upon administration. In some embodiments, the antigen does not elicit an adaptive immune response. In some embodiments, the antigen does not elicit an innate immune response. In some embodiments, the antigen does not elicit either an adaptive or an innate immune response. In some embodiments, an immunosilent antigen has reduced immunogenicity.

[0160] The term "intein" refers to a class of protein that is able to excise itself and join the remaining portion(s) of the protein via protein splicing. A "split intein" comes from two genes. A non-limiting example of a "split-intein" are the C-intein and N-intein sequences originally derived from *N. punctiforme.*

[0161] The term "isolated" as used herein refers to molecules or biologicals or cellular materials being substantially free from other materials.

[0162] As used herein, the terms "nucleic acid sequence" and "polynucleotide" are used interchangeably to refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA

hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, nnn-natural, or derivatized nucleotide bases.

[0163] The term "Major Histocompatibility Complex" (MHC) refers to a family of proteins responsible for the presentation of peptides, including self and non-self (antigenic) to T-cells. T-cells recognize antigenic peptides and trigger a cascade of events which leads to the destruction of pathogens and infected cells. The MHC family is divided into three subgroups: class I, class II, and class III. Class I MHC molecules have β2 subunits that are only recognized by CD8 co-receptors. Class II MHC molecules have βĩ and β2 subunits that are only recognized by CD4 co-receptors. In this way MHC molecules chaperone which type of lymphocytes may bind to the given antigen with high affinity, since different lymphocytes express different T-Cell Receptor (TCR) co-receptors. In general, MHC class I molecules bind short peptides, whose N- and C-terminal ends are anchored into pockets located at the ends of a peptide binding groove. While the majority of the peptides are nine amino acid residues in length, longer peptides can be accommodated by the bulging of their central portion, resulting in binding peptides of length 8 to 15. Peptides binding to class II proteins are not constrained in size and can vary from 11 to 30 amino acids long. The peptide binding groove in the MHC class II molecules is open at both ends, which enables binding of peptides with relatively longer length. The "core" refers to the amino acid residues that contribute the most to the recognition of the peptide. In some embodiments, the core is nine amino acids in length. In addition to the core, the flanking regions are also important for the specificity of the peptide to the MHC molecule.

[0164] As used herein, the term "organ" a structure which is a specific portion of an individual organism, where a certain function or functions of the individual organism is locally performed and which is morphologically separate. Non-limiting examples of organs include the skin, blood vessels, cornea, thymus, kidney, heart, liver, umbilical cord, intestine, nerve, lung, placenta, pancreas, thyroid and brain.

[0165] The term "ortholog" is used in reference of another gene or protein and intends a homolog of said gene or protein that evolved from the same ancestral source. Orthologs may or may not retain the same function as the gene or protein to which they are orthologous. Non-limiting examples of Cas9 orthologs include *S. aureus* Cas9 ("spCas9"), *S. thermophiles*

Cas9, ***L. pneumophilia*** Cas9, ***N. lactamica*** Cas9, ***N. meningitides*** Cas9, ***B. longum*** Cas9, ***A. muciniphilu*** Cas>9, and ***O. luneus*** Cas9.

[0166] The term "promoter" as used herein refers to any sequence that regulates the expression of a coding sequence, such as a gene. Promoters may be constitutive, inducible, repressible, or tissue-specific, for example. A "promoter" is a control sequence that is a region of a polynucleotide sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind such as RNA polymerase and other transcription factors. Non-limiting exemplary promoters include CMV promoter and U6 promoter.

[0167] The term "protein", "peptide" and "polypeptide" are used interchangeably and in their broadest sense to refer to a compound of two or more subunits of amino acids, amino acid analogs or peptidomimetics. The subunits may be linked by peptide bonds. In another aspect, the subunit may be linked by other bonds, e.g., ester, ether, etc. A protein or peptide must contain at least two amino acids and no limitation is placed on the maximum number of amino acids which may comprise a protein's or peptide's sequence. As used herein the term "amino acid" refers to either natural and/or unnatural or synthetic amino acids, including glycine and both the D and L optical isomers, amino acid analogs and peptidomimetics.

[0168] As used herein, the term "recombinant expression system" refers to a genetic construct for the expression of certain genetic material formed by recombination.

[0169] As used herein, the term "subject" is intended to mean any animal. In some embodiments, the subject may be a mammal; in further embodiments, the subject may be a bovine, equine, feline, murine, porcine, canine, human, or rat.

[0170] The term "tissue" is used herein to refer to tissue of a living or deceased organism or any tissue derived from or designed to mimic a living or deceased organism. The tissue may be healthy, diseased, and/or have genetic mutations. The biological tissue may include any single tissue (e.g., a collection of cells that may be interconnected) or a group of tissues making up an organ or part or region of the body of an organism. The tissue may comprise a homogeneous cellular material or it may be a composite structure such as that found in regions of the body including the thorax which for instance can include lung tissue, skeletal tissue, and/or muscle tissue. Exemplary tissues include, but are not limited to those derived

from liver, lung, thyroid, skin, pancreas, blood vessels, bladder, kidneys, brain, biliary tree, duodenum, abdominal aorta, iliac vein, heart and intestines, including any combination thereof.

[0171] As used herein, "treating" or "treatment" of a disease in a subject refers to (1) preventing the symptoms or disease from occurring in a subject that is predisposed or does not yet display symptoms of the disease; (2) inhibiting the disease or arresting its development; or (3) ameliorating or causing regression of the disease or the symptoms of the disease. As understood in the art, "treatment" is an approach for obtaining beneficial or desired results, including clinical results. For the purposes of the presenttechnology, beneficial or desired results can include one or more, but are not limited to, alleviation or amelioration of one or more symptoms, diminishment of extent of a condition (including a disease), stabilized *{i.e.,* not worsening) state of a condition (including disease), delay or slowing of condition (including disease), progression, amelioration or palliation of the condition (including disease), states and remission (whether partial or total), whether detectable or undetectable.

[0172] As used herein, the term "vector" intends a recombinant vector that retains the ability to infect and transduce non-dividing and/or slowly-dividing cells and integrate into the target cell's genome. The vector may be derived from or based on a wild-type virus. Aspects of this disclosure relate to an adeno-associated virus vector.

[0173] It is to be inferred without explicit recitation and unless otherwise intended, that when the present disclosure relates to a polypeptide, protein, polynucleotide or antibody, an equivalent or a biologically equivalent of such is intended within the scope of this disclosure. As used herein, the term "biological equivalent thereof is intended to be synonymous with "equivalent thereof when referring to a reference protein, antibody, polypeptide or nucleic acid, intends those having minimal homology while still maintaining desired structure or functionality. Unless specifically recited herein, it is contemplated that any polynucleotide, polypeptide or protein mentioned herein also includes equivalents thereof. For example, an equivalent intends at least about 70% homology or identity, or at least 80 % homology or identity and alternatively, or at least about 85 %, or alternatively at least about 90 %, or alternatively at least about 95 %, or alternatively 98 % percent homology or identity and exhibits substantially equivalent biological activity to the reference protein, polypeptide or

nucleic acid. Alternatively, when referring to polynucleotides, an equivalent thereof is a polynucleotide that hybridizes under stringent conditions to the reference polynucleotide or its complement.

[0174] Applicants have provided herein the polypeptide and/or polynucleotide sequences for use in gene and protein transfer and expression techniques described below. It should be understood, although not always explicitly stated that the sequences provided herein can be used to provide the expression product as well as substantially identical sequences that produce a protein that has the same biological properties. These "biologically equivalent" or "biologically active" polypeptides are encoded by equivalent polynucleotides as described herein. They may possess at least 60%, or alternatively, at least 65%, or alternatively, at least 70%, or alternatively, at least 75%, or alternatively, at least 80%, or alternatively at least 85%, or alternatively at least 90%, or alternatively at least 95% or alternatively at least 98%, identical primary amino acid sequence to the reference polypeptide when compared using sequence identity methods run under default conditions. Specific polypeptide sequences are provided as examples of particular embodiments. Modifications to the sequences to amino acids with alternate amino acids that have similar charge. Additionally, an equivalent polynucleotide is one that hybridizes under stringent conditions to the reference polynucleotide or its complement or in reference to a polypeptide, a polypeptide encoded by a polynucleotide that hybridizes to the reference encoding polynucleotide under stringent conditions or its complementary strand. Alternatively, an equivalent polypeptide or protein is one that is expressed from an equivalent polynucleotide.

"Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson-Crick base pairing, Hoogstein binding, or in any other sequence-specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PC reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme.

[0175] Examples of stringent hybridization conditions include: incubation temperatures of about 25°C to about 37°C; hybridization buffer concentrations of about 6x SSC to about lOx SSC; formamide concentrations of about 0% to about 25%; and wash solutions from about 4x SSC to about 8x SSC. Examples of moderate hybridization conditions include: incubation temperatures of about 40°C to about 50°C; buffer concentrations of about 9x SSC to about 2x SSC; formamide concentrations of about 30% to about 50%; and wash solutions of about 5x SSC to about 2x SSC. Examples of high stringency conditions include: incubation temperatures of about 55°C to about 68°C; buffer concentrations of about lx SSC to about O.lx SSC; formamide concentrations of about 55% to about 75%; and wash solutions of about lx SSC, Q1x SSC, or deionized water. In general, hybridization incubation times are from 5 minutes to 24 hours, with 1, 2, or more washing steps, and wash incubation times are about 1, 2, or 15 minutes. SSC is 0.1 5 MNaCl and 15 mM citrate buffer. It is understood that equivalents of SSC using other buffer systems can be employed.

[0176] "Homology" or "identity" or "similarity" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40% identity, or alternatively less than 25% identity, with one of the sequences of the present invention.

[0177] ***Modes of Carrying out the Disclosure***

[0178] **Methods of Generating Immunosilent Proteins and Identifying Immune Orthogonal Proteins**

[0179] Disclosed herein are methods of identifying or modifying a protein sequence to reduce immunogenicity, and optionally be immunosilent. In some aspects, the method comprises, consists of, or consists essentially of identifying affinity for a major histocompatibility complex (MHC) for one or more regions of a protein. Those protein regions which have no affinity to an MHC may be immunosilent without further modification. In contrast, those protein regions which have affinity, optionally high affinity,

to an MHC may be modified through one or more amino acid substitutions, such that the modified region has no affinity for the MHC. In some embodiments the MHC is MHC class I. In some embodiments, the MHC is MHC class II.

[0180] Simultaneously or sequentially, orthologs of the protein may be identified, optionally through alignment or alignment free methods *(**e.g.*** k-mer analysis. Regions of the orthologous may, thus, be targeted for similar modifications or may be considered immunosilent without further modification based on the results above. Alternatively, orthologs may be selected for sequential administration based on the fact that they are immune orthogonal, for example having affinity for different MHCs from those for the initially screened protein. Sequential administration of such immune orthogonal proteins an alternative **FIG. 5F** provides an exemplary schematic of the workflow to identify and/or modify these proteins.

[0181] Techniques to identify orthologous proteins are known in the art and include but are not limited to both traditional alignment based methods and alignment free methods. Further, databases of orthologous proteins are well known and include but are not limited to COGs, eggNOG, InParanoid, OrthoDB, Ortholuge, CDD, Ensmbl Compara, and KEGG. Thus, it is appreciated that one of ordinary skill may readily identify orthologs. For example, k-mer analysis is a computational method that identifies all possible substrings of a length k that are contained in a string, ***e.g.*** a sequence. The frequency of k-mers creates a "signature" of an underlying sequence, which in turn may be utilized as an alignment free means of comparing sequences and determining comprehensive peptide overlap. Other computations methods include those based on alignments, for example BLOSM (block substitution matrix) or PAM (point accepted mutation) matricies.

[0182] Methods of determining MHC affinity are likewise known in the art and may include computational methods available through software or publicly accessible databases or "wet lab" assays. Examples of computational methods of predicting MHC affinity include but are not limited to the MHC binding prediction model available through the IEDB Analysis Resource (http://tools.immuneepitope.org/mhci/ (MHC I) and http://tools.immuneepitope.org/mhcii/ (MHC II)) or NetMHC (http://www.cbs.dtu.dk/services/NetMHC/). Alternatively or in addition, MHC affinity can be determined or computational predictions thereof can be validated using assays, such as but

not limited to immunoassays, such as ELISA, microarray, tetramer assay, and peptide-induced MHC stabilization assay. Using such assays and computational methods can further be adapted to account for the MHC profile of a specific subject or patient being treated. Thus, modifications in the proteins can be optimized to be immunosilent in a particular subject or patient. Similarly the comparisons can be host-restricted, such that the protein is identified or modified to be specific to a particular host, ***e.g.,*** a mouse or a human.

[0183] Applicants contemplate use of this method for a variety of proteins that present a risk of eliciting an immune response. Non-limiting exemplary proteins of interest include cytidine deaminases, which can be used for gene editing via catalysis of DNA base change from C to T ***(e.g.*** APOBEC - Conserved across many species e.g. Rat APOBEC3, Rat APOBEC1, Resus Macaque APOBEC3G, human APOBEC1 (Al), AID, APOBEC2 (A2), APOBEC3A (A3A), APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3DE (A3DE), APOBEC3F (A3F), APOBEC3G (A3G), APOBEC3H (A3H) and APOBEC4 (A4)); adenosine deaminases, which can be used for gene editing via catalysis of DNA base change from A to G (e.g. ADA (DNA editor) - Widely conserved across virtually all species and ADAR (RNA editor) - Conserved across most metazoan species); Zing Finger nucleases (ZFNs), which can be used for genome engineering in a similar manner to CRISPR/Cas9 and are engineered site-specific nucleases consisting of: 3-6 repeated zinc finger domains, which is a widely conserved DNA-binding motif and a nuclease domain; transcriptional activator-like effector nucleases (TALENs), which be used for genome engineering in a similar manner to CRISPR/Cas9 and are similar to ZFNs in that they are engineered site-specific nucleases consisting of: a TAL effector DNA binding domain (generally derived from a species of ***Xanthomonas proteobacteria)*** and a nuclease domain. The domains of the site specific enzymes mentioned above (ZFNs and TALENs) are well characterized and subject of extensive engineering to generate the desired specificity. Thus, many variants exist of such proteins. Additional proteins for which MHC affinity analysis is relevant include Cas9 proteins and AAV capsids, both of which are used in CRISPR based gene editing.

[0184] Aspects of the disclosure relate to a method of generating a protein comprising: identifying one or more regions of a protein with affinity for a major histocompatibility complex (MHC), and modifying the one or more regions of the protien with affinity for the MHC through one or more amino acid substitutions, such that the modified region has no

affinity for the MHC, wherein the resulting modified protein is immunosilent upon administration of the modified protein or a polynucleotide encoding the modified protein to a subject. In some embodiments, the affinity for the MHC is high affinity. In some embodiments, at least one substituted amino acid is an amino acid which does not serve as an MHC protein core residue. In some embodiments, the protein is selected from the group of a cytidine deaminase, an adenosine deaminase, a zinc finger nuclease, a transcriptional activator-like effector nuclease, a Cas9, or an AAV capsid protein. In some embodiments, the protein is Cas9, optionally SpCas9.

[0185] For example, in order to optimize and broaden the application of CRIPSR based therapeutics the inventors correspondingly developed a couple of technologies: 1) "humanize" the Cas9 protein by swapping high immunogenic domains or peptides with less immunogenic counterparts. This is particularly useful to enable the application of Cas9 arsenal for repeat treatments. Upon mapping highly immunogenic peptides in SpCas9, Applicants computed single amino acid swaps at each position in these immunogenic peptides that are predicted to lower overall immunogenicity without potentially modifying the activity. The disclosure teaches which region to mutate and what to mutate to. In addition, applicants identified natural Cas9 ortholog proteins that are orthogonal in the immune space i.e. that do not share any highly immunogenic peptides, and are thus amenable for sequential dosing to evade host immune system and improve therapeutic regimen.

[0186] Thus, aspects of the disclosure relate to a modified Cas9 for immune stealth and use of a Cas9 ortholog to enhance immune evasion. The modified Cas9 can replace the existing wildtype Cas9 for any application requiring in vivo delivery, which would potentially have no loss of efficacy after repetitive use. The Cas9 proteins that are orthologous in the immune space can also be utilized for in vivo applications, where Cas9 proteins that are orthologous in the immune space can be utilized sequentially, if repetitive treatments are required. Such non-limiting aspects relating to Cas9 are described herein below.

[0187] Some embodiments disclosed herein relate to a method of generating a modified Cas9 comprising: identifying one or more regions of a Cas9 with high affinity for a major histocompatibility complex (MHC), and modifying the one or more regions of the Cas9 with high affinity for the MHC through one or more amino acid substitutions, such that the

modified region has no affinity for the MHC, wherein the resulting modified Cas9 is immunosilent upon administration to a subject. In some embodiments, the Cas9 is SpCas9. Further embodiments relate to a modified Cas9 generated according to this method. Some embodiments disclosed herein relate to a modified SpCas9 comprising one or more of the amino acid modifications provided in **Table 1.** Some embodiments disclosed herein relate to a method of avoiding an immune response in a subject being administering a regimen requiring Cas9 comprising: administering, in sequence, each of a group of orthologous Cas9 proteins with no shared affinity for a major histocompatibility complex (MHC). In some embodiments, the group of Cas9 proteins is selected from the groups of Cas9 proteins provided in **Figure 4.**

[0188] In some aspects, provided herein are methods of generating a modified Cas9 comprising, consisting of, or consisting essentially of: identifying one or more regions of a Cas9 with affinity for a major histocompatibility complex (MHC), and modifying the one or more regions of the Cas9 with affinity for the MHC through one or more amino acid substitutions, such that the modified region has no affinity for the MHC, wherein the resulting modified Cas9 has reduced immunogenicity upon administration to a subject. In some embodiments, the affinity for an MHC is high affinity. In some embodiments, the Cas9 is SpCas9. In some embodiments, at least one substituted amino acid is an amino acid which does not serve as an MHC protein core residue. In some aspects, provided herein is a modified Cas9 generated by identifying one or more regions of a Cas9 with affinity for a major histocompatibility complex (MHC), and modifying the one or more regions of the Cas9 with affinity for the MHC through one or more amino acid substitutions, such that the modified region has no affinity for the MHC, wherein the resulting modified Cas9 has reduced immunogenicity upon administration to a subject.

[0189] In some aspects, provided herein is a modified Cas9 comprising, consisting of, or consisting essentially of one or more, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, fifteen or more, or twenty or more of the amino acid modifications provided in **Table 1.**

[0190] In some aspects, provided herein are isolated polynucleotides encoding a modified Cas9 protein, wherein the modified Cas9 is generated by identifying one or more regions of a

Cas9 with affinity for a major histocompatibility complex (MHC), and modifying the one or more regions of the Cas9 with affinity for the MHC through one or more amino acid substitutions, such that the modified region has no affinity for the MHC, wherein the resulting modified Cas9 has reduced immunogenicity upon administration to a subject. In some aspects, provided herein are isolated polynucleotides encoding a modified Cas9 protein, wherein the modified Cas9 comprises, consists of, or consists essentially of one or more, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, fifteen or more, or twenty or more of the amino acid modifications provided in Table 1. In some aspects, provided herein are vectors comprising the isolated polynucleotide. In some embodiments, the vector is an AAV vector, optionally wherein the AAV vector is AAV5.

[0191] It is further appreciated that the AAV capsid may be modified to be immunosilent according to the same method, *i.e.* identifying one or more regions of one or more AAV capsid proteins with affinity for a major histocompatibility complex (MHC), and modifying the one or more regions of the one or more AAV capsid proteins with affinity for the MHC through one or more amino acid substitutions, such that the modified region has no affinity for the MHC, wherein the resulting capsid comprising the one or more AAV capsid proteins has reduced immunogenicity upon administration to a subject. A modified AAV generated according to this method may be employed in any one or the embodiments disclosed herein to evade the immune system.

[0192] Further, immune orthogonal AAV may be identified according to the method disclosed herein. Thus, contemplated herein are embodiments in which the immune orthogonal Cas9 is comprised in an immune orthogonal AAV.

[0193] Additional aspects to a method of identifying immune orthogonal orthologs comprising: determining a set of affinities of a protein or regions thereof to a plurality of major histocompatibility complexes (MHCs), comparing the set of affinities of the protein or regions thereof to sets of affinities of orthologs of the protein to the plurality of MHCs, and determining a set of immune urthogonal orthologs based on non-overlapping sets of affinit.es. In some embodiments, the affinity for the MHC is high affinity. In some embodiments, the protein is selected from the group of a cytidine deaminase, an adenosine deaminase, a zinc

finger nuclease, a transcriptional activator-like effector nuclease, a Cas9, or an AAV capsid protein. In some embodiments, the protein is Cas9, optionally SpCas9 or SaCas9. In some embodiments, the Cas9 proteins the orthologs are selected from *S. pyogenes* Cas9 (spCas9), *S. aureus* Cas9 (saCas9), *B. longum* Cas9, *A. muiciniphilia* Cas9, or *O. laneus* Cas9.

[0194] Not to be bound by theory, Applicants contemplate that even after MHC screening, a subject may still have a repitoire of pre-existing immunity that could result in cross-reactivity against protiens or their orthologs. Thus, there exists some risk of confounding in sequential administration of proteins that are immune orthogonal. Non-limiting exemplary protiens which may present this concern are those derived from organisms that are pathogenic in a subject *(e.g. S. aureus* or *S. pyogenes* in humans). Accordingly, Applicants propose identifying immune orthogonal orthologs of such proteins that are extremophiles (and, thus, unlikely to come into contact with humans or other subjects under normal circumstances) and/or highly abundant commensal species for which the subject's immune system has developed tolerance. Species abundant in a normal microbiome or in the particular subject's microbiome can be determined based on the literature and/or based on sampling over a population of subjects or the particular subjects. In some embodiments, the commensal species is one present at early stages of development, when tolerance is established.

[0195] Proteins and Vectors

[0196] Further aspects relate to a modified Cas9 protein produced according to the method disclosed above. Still further aspects relate to a modified Cas9 protein comprising one or more, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, fifteen or more, or twenty or more of the amino acid modifications provided in **Table 1.** Some embodiments relate to an isolated polynucleotide encoding the modified Cas9. Further embodiments, relate to a vector comprising the isolated polynucleotide, optionally an AAV vector, and still further optionally an AAV5 vector. Additional embodiments relate to an AAV capsid comprising the vector. In some embodiments, one or more of the AAV capsid proteins has been modified to be immunosilent.

[0197] In general methods of packaging genetic material such as RNA into one or more vectors is well known in the art. For example, the genetic material may be packaged using a packaging vector and cell lines and introduced via traditional recombinant methods.

[0198] In some embodiments, the packaging vector may include, but is not limited to retroviral vector, lentiviral vector, adenoviral vector, and adeno-associated viral vector (optionally AAV8). The packaging vector contains elements and sequences that facilitate the delivery of genetic materials into cells. For example, the retroviral constructs are packaging plasmids comprising at least one retroviral helper DNA sequence derived from a replication-incompetent retroviral genome encoding in trans all virion proteins required to package a replication incompetent retroviral vector, and for producing virion proteins capable of packaging the replication-incompetent retroviral vector at high titer, without the production of replication-competent helper virus. The retroviral DNA sequence lacks the region encoding the native enhancer and/or promoter of the viral 5' LTR of the virus, and lacks both the psi function sequence responsible for packaging helper genome and the 3' LTR, but encodes a foreign polyadenylation site, for example the SV40 polyadenylation site, and a foreign enhancer and/or promoter which directs efficient transcription in a cell type where virus production is desired. The retrovirus is a leukemia virus such as a Moloney Murine Leukemia Virus (MMLV), the Human Immunodeficiency Virus (FHV), or the Gibbon Ape Leukemia virus (GALV). The foreign enhancer and promoter may be the human cytomegalovirus (HCMV) immediate early (IE) enhancer and promoter, the enhancer and promoter (U3 region) of the Moloney Murine Sarcoma Virus (MMSV), the U3 region of Rous Sarcoma Virus (RSV), the U3 region of Spleen Focus Forming Virus (SFFV), or the HCMV IE enhancer joined to the native Moloney Murine Leukemia Virus (MMLV) promoter.

[0199] The retroviral packaging plasmid may consist of two retroviral helper DNA sequences encoded by plasmid based expression vectors, for example where a first helper sequence contains a cDNA encoding the gag and pol proteins of ecotropic MMLV or GALV and a second helper sequence contains a cDNA encoding the env protein. The Env gene, which determines the host range, may be derived from the genes encoding xenotropic, amphotropic, ecotropic, polytropic (mink focus forming) or 10A1 murine leukemia virus env proteins, or the Gibbon Ape Leukemia Virus (GALV env protein, the Human

Immunodeficiency Virus env (gpl60) protein, the Vesicular Stomatitus Virus (VSV) G protein, the Human T cell leukemia (HTLV) type I and II env gene products, chimeric envelope gene derived from combinations of one or more of the aforementioned env genes or chimeric envelope genes encoding the cytoplasmic and transmembrane of the aforementioned env gene products and a monoclonal antibody directed against a specific surface molecule on a desired target cell. Similar vector based systems may employ other vectors such as sleeping beauty vectors or transposon elements.

[0200] The resulting packaged expression systems may then be introduced via an appropriate route of administration, discussed in detail with respect to the method aspects disclosed herein.

[0201] Methods of Treatment

[0202] Some aspects relate to a method of avoiding immune response in a subject being administered a regimen requiring a protein, the method comprising: administering to the subject, in sequence, two or more proteins that are immune orthogonal. In some embodiments, the proteins that are immune orthogonal do not share an amino acid sequence of greater than 5 consecutive amino acids. In some embodiments, the proteins that are immune orthogonal do not share affinity for a major histocompatibility complex (MHC). In some embodiments, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more proteins that are immune orthogonal are administered in sequence.

[0203] Non-limiting exemplary aspects relate to Cas9. In some embodiments, the Cas9 proteins that are immune orthogonal do not share an amino acid sequence of greater than 5 consecutive amino acids. In some embodiments, the Cas9 proteins that are immune orthogonal do not share affinity for a major histocompatibility complex (MHC). In some embodiments, at least one of the two or more Cas9 proteins is modified according the method disclosed above. In some embodiments, at least one of the two or more Cas9 proteins or polynucleotides encoding said Cas9 proteins is comprised in an AAV vector. In some embodiments, the AAV vector is an AAV5 vector. In some embodiments, the AAV vector is comprised in an AAV capsid. In some embodiments, two or more Cas9 proteins or polynucleotides encoding said Cas9 proteins are comprised in AAV vectors. In some

embodiments, each AAV vector is comprised in an AAV capsid, optionally wherein the AAV capsids are immune orthogonal to one another.

[0204] Disclosed herein is a method of gene editing comprising contacting a cell sequentially with two or more immune orthogonal Cas9s or polynucleotides encoding said Cas9s, optionally comprised in an AAV capsid. In some embodiments, the AAV capsids comprising each of the Cas9 or the polynucleotides encoding them may be immune orthogonal. In some aspects, the contact is in vitro. In other aspects, the contact is in vivo. In some aspects, the contact is in vivo or in vitro. In some aspects, at least one of the polynucleotides comprises or consists essentially of, or yet further consists of a polynucleotide encoding a guide RNA (gRNA). In some aspects, at least one of the polynucleotides comprises or alternatively consists essentially of, or yet further consists of a therapeutic polypeptide.

[0205] Further disclosed herein is a method of gene editing in a subject in need thereof, comprising administering sequentially to the subject an effective amount of two or more immune orthogonal Cas9 or polynucleotides encoding said Cas9s, optionally comprised in an AAV. In some embodiments, the AAV capsids comprising each of the Cas9 or the polynucleotides encoding them may be immune orthogonal. In some aspects, at least one of the polynucleotides comprises or consists essentially of, or yet further consists of a polynucleotide encoding a guide RNA (gRNA). In some aspects, at least one of the polynucleotides comprises or alternatively consists essentially of, or yet further consists of a therapeutic polypeptide.

[0206] In some aspects, the polynucleotide encoding the gRNA comprises or alternatively consists essentially of, or yet further consists of a fusion polypeptide comprising CRISPR RNA (crRNA) and trans-activating CRIPSPR RNA (tracrRNA); or a polypeptide comprising CRISPR RNA (crRNA) and trans-activating CRIPSPR RNA (tracrRNA). In one aspect, the polynucleotide encoding the gRNA comprises or consists of one or more sequence from **Table 2** or **Table 3** or an equivalent each thereof. In some aspects, the gRNA is specific for a region of DNA that is in need of gene editing in the subject or cell in need thereof.

[0207] In some aspects, provided herein are methods of treating a subject in need of gene editing or gene regulation, the method comprising: administering to the subject, in sequence, two or more Cas9 proteins that are immune orthogonal. In some embodiments, the Cas9 proteins that are immune orthogonal do not share an amino acid sequence of greater than 5 consecutive amino acids. In some embodiments, the Cas9 proteins that are immune orthogonal do not share affinity for a major histocompatibility complex (MHC). In some embodiments, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more Cas9 proteins that are immune orthogonal are administered in sequence. In some embodiments, each Cas9 protein that is immune orthogonal is a Cas9 derived from a distinct species of bacteria. In some embodiments, the Cas9 proteins that are immune orthogonal are selected from ***S. pyogenes*** Cas9 (spCas9), ***S. aureus*** Cas9 (saCas9), ***B. longum*** Cas9, ***A. muiciniphilia*** Cas9, or ***O. laneus*** Cas9. In particular embodiments, the Cas9 proteins that are immune orthogonal comprise spCas9 and saCas9. In some embodiments, at least one Cas9 is modified to reduce immunogenicity upon administration to the subject. In some embodiments, the methods further comprise administering at least one of the two or more Cas9 proteins in an AAV5 vector. In some embodiments, the methods further comprise administering one or more guide RNAs to the subject.

[0208] In some embodiments, the guide RNA is selected to treat a disease, disorder, or condition selected from the group of achromatopsia, adenosine deaminase (ADA) deficiency, alpha-1-antitrypsin deficiency, Alzheimer's disease, amyotrophic lateral sclerosis, aromatic amino acid decarboxylase deficiency, Batten disease, choroideremia, Crigler Najjar syndrome, cystic fibrosis, fragile X syndrome, hemophilia, hepatitis B, hepatitis C,homozygous familial hypercholesteremia, Huntington's Disease, Leber congenital amaurosis, macular degeneration, maple syrup urine disease (MSUD), mucopolysarccharidosis (I-IX), multiple sclerosis, muscular dystrophy, myotonic dystrophy, neurofibramotosis type 1, ornithine transcarbamylase deficiency, pachyonychia congenita, Parkinson's disease, phenylketonuria, polycystic kidney disease, Pompe disease, retinal degeneration, Rett's syndrome, rickets, spinal muscular atrophy, severe combined immunodeficiency, sickle cell disease, Smith-Lemli-Opitz syndrome, Y-linked nonobstructive spermatogenic failure, thalassemia, and X-linked retinoschisis.

[0209] In some aspects, the guide RNA is designed and/or selected to target or repair a gene selected from the group of: Nav 1.7 (SCN9A), Nav **1.8** (SCNIOA gene), 1.9 (SCNl 1A gene) and 1.3 (SCN3A gene); transient receptor potential cation channel subfamily V member 1 (TrpVl), also known as the capsaicin receptor and the vanilloid receptor 1; PRDM12; or HCN2.

[0210] It is appreciated by those skilled in the art that gRNAs can be generated for target specificity to target a specific gene, optionally a gene associated with a disease, disorder, or condition. Thus, in combination with Cas9, the guide RNAs facilitate the target specificity of the CRISPR/Cas9 system. Further aspects such as promoter choice, as discussed above, may provide additional mechanisms of achieving target specificity - e.g., selecting a promoter for the guide RNA encoding polynucleotide that facilitates expression in a particular organ or tissue. Accordingly, the selection of suitable gRNAs for the particular disease, disorder, or condition is contemplated herein. Non-limiting examples of suitable gRNA for genes in humans are provided in **Table 2** and in mice in **Table 3.**

[0211] Administration of the modified AAV or compositions can be effected in one dose, continuously or intermittently throughout the course of treatment. Administration may be through any suitable mode of administration, including but not limited to: intravenous, intra¬ arterial, intramuscular, intracardiac, intrathecal, subventricular, epidural, intracerebral, intracerebroventricular, sub-retinal, intravitreal, intraarticular, intraocular, intraperitoneal, intrauterine, intradermal, subcutaneous, transdermal, transmuccosal, and inhalation.

[0212] Methods of determining the most effective means and dosage of administration are known to thqse of skill in the art and will vary with the composition used for therapy, the purpose of the therapy and the subject being treated. Single or multiple administrations can be carried out with the dose level and pattern being selected by the treating physician. It is noted that dosage may be impacted by the route of administration. Suitable dosage formulations and methods of administering the agents are known in the art. Non-limiting examples of such suitable dosages may be as low as 1E+9 vector genomes to as much as 1E+17 vector genomes per administration.

[0213] In a further aspect, the modified viral particle and compositions of the invention can be administered in combination with other treatments, ***e.g.*** those approved treatments suitable for the particular disease, disorder, or condition. A non-limiting example includes the treatment of muscular dystrophy with a combination of the modified viral particle and one or more steroids.

[0214] This administration of the modified viral particle or compositions of the invention can be done to generate an animal model of the desired disease, disorder, or condition for experimental and screening assays.

[0215] Doses suitable for uses herein may be delivered via any suitable route, ***e.g.*** intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods, and/or via single or multiple doses. It is appreciated that actual dosage can vary depending on the recombinant expression system used ***(e.g.*** AAV or lentivirus), the target cell, organ, or tissue, the subject, as well as the degree of effect sought. Size and weight of the tissue, organ, and/or patient can also affect dosing. Doses may further include additional agents, including but not limited to a carrier. Non-limiting examples of suitable carriers are known in the art: for example, water, saline, ethanol, glycerol, lactose, sucrose, dextran, agar, pectin, plant-derived oils, phosphate-buffered saline, and/or diluents. Additional materials, for instance those disclosed in paragraph [00533] of WO 2017/070605 may be appropriate for use with the compositions disclosed herein. Paragraphs [00534] through [00537] of WO 2017/070605 also provide non-limiting examples of dosing conventions for CRJSPR-Cas systems which can be used herein. In general, dosing considerations are well understood by those in the art.

[0216] Compositions and Kits

[0217] Also provided by this invention is a composition or kit comprising any one or more of the immunosilent and/or immune orthogonal proteins. In one aspect, the carrier is a pharmaceutically acceptable carrier. These compositions can be used therapeutically as described herein and can be used in combination with other known therapies and/or according to the method aspects described herein.

[0218] Briefly, pharmaceutical compositions of the present invention may comprise an immunosilent and/or immune orthogonal Cas9 or a polynucleotide encoding said Cas9,

optionally comprised in an AAV, which is optionally also immune orthogonal, in combination with one or more pharmaceutically or physiologically acceptable carriers, diluents or excipients. Such compositions may comprise buffers such as neutral buffered saline, phosphate buffered saline and the like; carbohydrates such as glucose, mannose, sucrose or dextrans, mannitol; proteins; polypeptides or amino acids such as glycine; antioxidants; chelating agents such as EDTA or glutathione; adjuvants (e.g., aluminum hydroxide); and preservatives. Compositions of the present disclosure may be formulated for oral, intravenous, topical, enteral, and/or parenteral administration. In certain embodiments, the compositions of the present disclosure are formulated for intravenous administration.

[0219] ***Examples***

[0220] The following examples are non-limiting and illustrative of procedures which can be used in various instances in carrying the disclosure into effect. Additionally, all reference disclosed herein are incorporated by reference in their entirety.

[0221] **Example 1 - Immunogenicity of Cas9 proteins**

[0222] Several in silico epitope binding prediction methods have been developed that employ machine learning methods to predict peptide-MHC class I binding affinity. Applicants have utilized the NetMHC 4.0 Server 4, a neural network and weight matrix based predictive algorithm, to determine the immunogenic level of peptides in previously identified Cas9 protein sequences from 88 strains 6, over all HLA allele supertypes.

[0223] NetMHC was run with default parameters, predicting immunogenic scores for each allele over peptide sequences of 8 to 11 amino acids. Highly immunogenic peptides were defined as having an affinity score < 50nM and intermediate as 50nM500nM.

[0224] After identifying the most immunogenic peptides, Applicants utilized two in silico methods to determine which modifications were necessary to reduce SpCas9 immunogenicity

[0225] 1) determined the effect that single amino acid swaps in each highly immunogenic peptide would have on reducing immunogenicity

[0226] 2) found which Cas9 orthologs are the closest in their 'immunogenic space' to determine which Cas9 proteins could be utilized sequentially for repetitive treatments.

[0227] An overall workflow is described in **Figure 1.**

[0228] Example 2 - Effect of single amino acid swaps in immunogenic peptides in SpCas9

[0229] After mapping the highly immunogenic peptides in SpCas9, Applicants did single amino acid swaps at each position in these immunogenic peptides to determine whether these swaps would lower the peptides' overall immunogenicity. This new list of peptides was first submitted to the NetMHC server to predict their immunogenicity scores. The goal was to find if changing the single AA in such peptides would significantly modify the affinity.

[0230] Affinity scores were calculated for every single amino acid swap in an immunogenic peptide. For example, the peptide 'HHQDLTLL', located at amino acid position 327-334 in the original protein, has 32 no-affinity scoring peptides with a single amino acid swap (e.g. 'HHQDLTLK','HHQDLTLN','HHQDLTLD'). Top scoring peptides were defined as those that displayed the lowest affinity value out of all possible peptide swaps. Subsequently, the 'no' affinity peptides were submitted to the PROVEAN Server, which predicts the effect that single amino acid changes at certain positions can have on a protein's functionality. 7 The single amino acid swaps leading to 'no' or 'low' immunogenicity and that are non-deleterious will subsequently be utilized for experimental mutagenesis of SpCas9. These mutations are listed in **Table 1,** with the matching colors corresponding to peptides whose immunogenicity can change with the same AA swap.

[0231] One can then use this mutated SpCas9 sequentially for in vivo genome therapy. Not to bound by theory it is believed this may be accomplished without lowering its efficacy after repetitive treatments without eliciting an immunogenic response.

[0232] Example 3 - Orthogonality of Cas9 proteins for sequential dosing to evade host immune system

[0233] The goal was to determine Cas9 orthologs that are orthogonal in the 'immunogenicity space'. This will allow Applicants to prescribe a sequential regimen of Cas9s for therapeutic interventions. The analysis reveals that for the most conservative data, there are always at the very least groups of 35 proteins that are mutually orthogonal and that include SpCas9. The methodology implemented goes as follows: high affinity peptides from one protein were selected and the number of times those exact peptide sequences occurred in the entire other sequence was determined. If no peptides were found, the proteins are

determined to be orthogonal. The peptides selected, usually composed of 8 to 11 amino acids, were further split up into subpeptides of lengths 5 to 11. This allowed for the identification of more subtle similarities between protein sequences. This analysis was carried over every possible protein pair. The groups of mutually orthogonal proteins here presented had no matches of even length 5. The algorithm used to determined mutual orthogonality, `find_cliques`, is provided in the Python package Networkx.

[0234] Applicants created a network where two proteins (nodes) were connected by an edge if they were orthogonal. Applicants then applied the clique-finding algorithm to locate all maximal cliques in the graph, where a maximal clique is a complete subgraph such that no other node may be added while maintaining completeness. ***See, e.g.*** **Figure 4.**

[0235] Example 4 - Mouse experiments

[0236] Two month old mice are injected with AAV virus at 6E+1 lGC/mouse. Applicants will be testing two different AAV capsids, AAV8 and AAVDJ, as well as two orthogonal Cas9 proteins, SpCas9 and SaCas9, to test whether sequential rounds of AAV virus injections with differing capsid or differing SpCas9 proteins has any effect on reducing efficacy of genome editing, due to an immunogenic response.

Week 0	Week 3	Week 6
A1	B2	Assay (baseline and role of AAVs)
A2	A1	Assay (baseline and role of AAVs)
B1	B2	Assay (baseline and role of AAVs)
B2	B1	Assay (baseline and role of AAVs)
A1	B2	Assay (Cas9 orthogonality)
B2	A1	Assay (Cas9 orthogonality)
A2	B1	Assay (Cas9 orthogonality)
B1	A2	Assay (Cas9 orthogonality)

[0237] Legend:

Al: AAV8 SpCas9 CD81; A2: AAVDJ SpCas9 Scarbl; Bl: AAV8 SaCas9 CD81; B2: AAVDJ SaCas9 Scarbl

[0238] Example 5 - Determining presence of memory T-celi populations to predicted peptides

[0239] Memory T-cell populations present in the human populations are assessed for the presence of T-cells directed to any of the predicted Cas9 orthologs. In particular, *S. aureus* peptides are studied, as approximately 30% of the human population is colonized with this pathogen.

[0240] Example 6 - Screening for "Immune Orthogonal" Orthologs

[0241] A major hurdle in protein-based therapeutics is the interaction with the adaptive immune system, which can lead to neutralization by circulating antibodies and clearance of treated cells by cytotoxic T-lymphocytes. One method of circumventing these issues is to use human or humanized proteins which avoid the immune response by self-recognition. However, this approach limits potential protein therapeutics to those of human origin, excluding many exciting effectors and delivery vehicles such as CRISPR-Cas9 and adeno-associated viruses (AAVs). To address this issue, Applicants propose here the sequential use of orthologous proteins whose function is constrained by natural selection, but whose structure is subject to diversification by genetic drift. This would, in principle, allow for repeated treatments by 'immune orthogonal' orthologs without reduced efficacy due to lack of immune cross-reactivity among the proteins. To explore and validate this concept, Applicants chose 9 1 Type II CRJSPR-Cas9 orthologs and 167 AAV capsid protein orthologs, and developed a pipeline to compare total sequence similarity as well as predicted binding to class I and class II Major Histocompatibility Complex (MHC) proteins. Interestingly, MHC binding predictions revealed wide diversity among the set of Cas9 orthologs, with 83% of pairs predicted to have non cross-reacting immune responses, while no global immune orthogonality among AAV serotypes was observed. To confirm these findings Applicants selected two Cas9 orthologs, from *S. pyogenes* and *S. aureus*, predicted to be orthogonal in immune space, and delivered them into mice via multiple AAV serotypes. Applicants observed cross-reacting antibodies against AAV but not Cas9 orthologs in sera from immunized mice, validating the computationally predicted immune orthogonality among these proteins. Moving forward, Applicants anticipate this framework can be applied to rationally engineer immune orthogonality among protein orthologs.

[0242] Protein therapeutics, including protein-based gene therapy, have several advantages over small-molecule drugs. They generally serve complex, specific functions, and have minimal off-target interference with normal biological processes. However, one of the

fundamental challenges to any protein-based therapeutic is the interaction with the adaptive immune system. Neutralization by circulating antibodies through B-cell activation and clearance of treated cells by CD8+ cytotoxic T-lymphocytes (CTLs) create a substantial barrier to effective protein therapies[10]. Although the delay in the adaptive immune response to novel proteins may allow sufficient time for the initial dose to work, subsequent doses face faster and stronger secondary immune responses due to the presence of memory T- and B-cells. In addition, gene transfer studies have shown that host immune responses against the delivery vector and/or therapeutic transgene can eliminate treated cells, thus limiting the efficacy of the treatment[11-16].

[0243] A common approach to circumventing these issues has been to utilize human proteins, or to humanize proteins by substitution of non-human components[17,18]. However, this approach is limited to a small set of therapeutic proteins naturally occurring in humans or closely related species. In addition, although the humanization of proteins can result in a significantly less immunogenic product, they still carry immunological risk[18]. Another way to circumvent an immune response to protein therapeutics is the removal of immunogenic T cell epitopes.[19,20] Once immunogenic T cell epitopes are identified, substitution of key amino acids may reduce the protein's immunogenicity since modification of amino acids at critical anchor residues can abrogate binding to MHC molecules and prevent antigen presentation. However, this can prove difficult due to the massive diversity at HLA loci. As epitope engineering must account for the substrate specificity of each different HLA allele, therapeutics would likely have to be uniquely modified for each patient. All the same, epitope deletion has been successfully applied to several proteins,[21] but can only preserve protein function when limited to small numbers of HLA alleles unrepresentative of the full diversity. Structural modifications such as PEGylation have also been known to reduce immunogenicity by interfering with antigen-processing mechanisms. However, there is evidence that PEG-specific antibodies are elicited in patients treated with PEGylated therapeutic enzymes[22-25].

[0244] Furthermore, protein therapies have required repeated treatments due to degradation of the protein or turnover of treated cells, or, in the case of gene therapy, reduced expression of the transgene[26,27]. This provides an even greater challenge as repeated exposure to the same antigen can elicit a more robust secondary immune response[28], which may completely

inhibit subsequent dosage or even sensitize the immune system to antigens remaining from the initial exposure. In order to facilitate efficacious repeat protein therapies, Applicants propose the use of orthologous proteins whose function is constrained by natural selection, but whose structure is subject to diversification by genetic drift. An ortholog, given sufficient sequence divergence, will not cross-react with the immune response generated by exposure to the others, allowing repeat doses to avoid neutralization by existing antibodies and treated cells to avoid clearance by activated CTLs.

[0245] As a case study for exploring this approach, Applicants focused on the CRISPR-Cas9 system, perhaps the most anticipated therapeutic for gene editing[29-36]. Comparative genomics has demonstrated that Cas9 proteins are widely distributed across bacterial species and have diversified over an extensive evolutionary history[37-39]. Applicants hypothesized this diversity could provide a mechanism to circumvent inducing immunological memory by utilizing orthologous Cas9 proteins for each treatment. Additionally, the immunogenicity due to the delivery vehicle or administration route for the Cas9 and the associated guide RNA (gRNA) must also be considered. In this regard, adeno-associated viruses (AAVs) have emerged as a highly preferred vehicle for gene delivery, as these are associated with low immunogenicity and toxicity[14,15], which promotes long-term transgene expression[40,41] and treatment efficacy. Despite the relatively low immunogenicity of AAV vectors, antibodies against both the capsid and transgene may still be elicited[42-46]. Additionally, the prevalence of neutralizing antibodies (NAB) against AAVs in the human population[47] and cross-reactivity between serotypes[48] remains a hurdle for efficacious AAV therapy. Although AAVs were initially considered non-immunogenic due to their poor transduction of antigen-presenting cells (APCs)[49], it is now known that they can transduce dendritic cells (DCs)[50] and trigger innate immune responses through Toll-like receptor (TLR) signaling pathways[51]. The ability to transduce DCs is dependent on AAV serotype and genome, and may be predictive of overall immunogenicity[52].

[0246] To evaluate the immune orthogonality of AAV-delivered CRISPR-Cas systems, Applicants analyzed 91 Cas9 orthologs, and 167 AAV VP1 orthologs. By comparing total sequence similarity as well as predicted binding strengths to class I and class II MHC molecules, Applicants constructed graphs of immune cross-reactivity and computed cliques of proteins that are orthogonal in immunogenicity profiles. Although MHC epitopes do not

predict antibody epitopes, the induction of the more powerful memory response is primarily dependent on reactivation of memory B-cells with help from memory T-cells through the presentation of antigens on class II MHC molecules.[53'54] Finally, Applicants experimentally confirmed these immunological predictions by assaying treated mice for induction of protein-targeting antibodies.

[0247] Humoral immune response to AAV and Cas9

[0248] One of the major obstacles for sequential gene therapy treatments is the presence of neutralizing antibodies against the delivery vehicle and transgene cargo induced by the first administration of the therapy. To determine the humoral immune response kinetics to the AAV-8 capsid and the Cas9 transgene, Applicants first injected C57BL/6J mice retro-orbitally with 10^{12} vg of AAV-8-SaCas9 targeting proprotein convertase subtilisin/kexin type 9 (PCSK9), a promising gene target that when disrupted can reduce Low Density Lipoprotein (LDL) levels and protect against cardiovascular disease. Consistent with a previous study[55], mice had reduced PCSK9 serum levels as early as one week post-injection due to successful SaCas9 mediated gene-editing, which was sustained for the entire duration of the experiment (4 weeks) **(FIG. 5C).** Notably, mice developed humoral immunity to the AAV8 capsid within one week post-injection **(FIG. 5D).** Additionally, Applicants noted that a subset of the mice developed IgGl antibodies against the SaCas9 protein **(FIG. 5E).** To evaluate the feasibility of multiple dosing with AAV-Cas9, Applicants next investigated whether immune orthogonal sets of AAV and Cas9 orthologs exist.

[0249] Identifying immune-orthogonal proteins

[0250] Natural selection produces diverse structural variants with conserved function in the form of orthologous genes. Applicants assayed the relevance of this diversity for immunological cross-reactivity of 9 1 Type II Cas9 orthologs and 167 AAV orthologs by first comparing their overall amino acid sequence similarities, and second, using a more specific constraint of how their respective amino acid sequences are predicted to bind MHC Type I and II molecules **(FIG. 5F).** From these analyses Applicants obtained first an estimate of the comprehensive immune overlap among Cas9 and AAV orthologs based purely at the sequence level, and second a more stringent estimate of predicted immune overlap based on predicted MHC binding. By sequence-level clustering and clique finding methods, Applicants defined many sets of Cas9 orthologs containing up to 9 members with no 6-mer overlap

(FIG. 7). Notably, based on MHC-binding predictions, Applicants find among the set of Cas9 orthologs that 83% of pairs are predicted to have non cross-reacting immune responses, i.e. they are predicted to be orthogonal in immune space **(FIG. 5G).** On the contrary, among AAV capsid (VPl protein) orthologs, Applicants did not find full orthogonality up to the 16-mer level, even when restricting predictions with MHC-binding strengths **(FIG. 5H),** likely reflecting the strong sequence conservation and shorter evolutionary history of AAVs[56]. This analysis suggests, consistent with previous observations[57,58], that exposure to one AAV serotype can induce broad immunity to all AAVs, which presents a significant challenge to AAV delivery platforms, as some serotypes are prevalent in human populations. Despite the most divergent AAV serotype (AAV-5) showing the fewest shared immunogenic peptides, there remain tracts of sequences fully conserved within the VPl orthologs. As expected, predicted immune cross-reaction negatively correlates with phylogenetic distance **(FIG. 8),** though there is significant variation not captured by that regression, suggesting that MHC-binding predictions can refine the choice of sequential orthologs beyond phylogenetic distance alone.

[0251] Confirming humoral immune-orthogonality among Cas9 proteins

[0252] To test these immunological predictions and to establish the utility of this approach, Applicants narrowed in on a 5-member clique containing the ubiquitously used ***S. pyogenes*** Cas9 in addition to the well-characterized ***S. aureus*** Cas9 **(FIG. 7).** To determine whether either of these proteins have cross-reacting antibody responses, Applicants injected mice with 10^{12} vg of either AAV8-SaCas9 or AAV8-SpCas9 via retro-orbital injections and harvested serum at days 0 (pre-injection), and periodically over 4-6 weeks **(FIG. 6A).** SpCas9-specific antibodies were detected in the plasma of all mice injected with SpCas9 (n=6), and notably none of the mice injected with SaCas9 (n=12) **(FIG. 6B).** Although SaCas9 appeared to induce a weaker response, as only half of the mice injected with SaCas9 AAVs (n=12) developed detectable antibodies against SaCas9, none of the mice injected with SpCas9 AAVs (n=6) developed an antibody response against SaCas9. These results were confirmed in an independent study in which SpCas9-specific antibodies, but not SaCas9-specific antibodies, were detected in the plasma of mice injected with AAV-SpCas9 (n=12). These mice were injected retro-orbitally with 10^{12} vg of AAV8-SpCas9 or AAVDJ-SpCas9, and

also received an additional intramuscular injection with 10^{11} vg at week 4. **(FIG. 6C).** Taken together, this data confirms that SpCas9 and SaCas9 have humoral immune-orthogonality.

[0253] Broad cross-reactivity among AAV serotypes

[0254] AAVs are becoming a preferred delivery vehicle due to their ability to avoid induction of a strong CD8+ T-cell response, however, the presence of neutralizing antibodies remains a significant barrier to successful application of AAV therapies. Consistent with previous results,[57] Applicants found shared immunogenic peptides among all the various human AAV serotypes, **(FIG. 9).** Applicants confirmed the lack of orthogonality for two serotypes, AAV8 and AAVDJ, in which Applicants found that antibodies produced in mice injected with AAV8 and AAVDJ react to both AAV8 and AAVDJ antigens **(Figure 6D).** This analysis suggests that there are no two known AAVs for which exposure to one would guarantee immune naiveté to another across all HLA genotypes. However, immune cross-reaction could be minimized through the use of AAV5[58,59], the most phylogenetically divergent serotype. These predictions identify only a single shared highly immunogenic peptide between AAV5 and the commonly used AAV2 and AAV8 in the mouse model (though several other shared peptides of mild MHC affinity exist). Applicants confirmed this via ELISAs, where mice injected with AAV2 did not elicit antibodies against AAV5 and AAV8, and mice injected with AAV5 did not elicit antibodies against AAVDJ and AAV8 **(Figure 6E).**

[0255] The use of protein therapeutics requires ways to evade the host's immune response. Cas9, as an example, has prokaryotic origins and can evoke a T-cell response, which may lead to clearance of transduced cells. In addition, circulating antibodies can neutralize the AAV vector and prevent efficient transduction upon repeated doses. Immunosuppressive drugs could mitigate some of these aspects, but not without significant side-effects, as well as not being applicable to patients in poor health[60-63]. Similar to what has been done in cancer antibody therapeutics[64], the SpCas9 protein could also be de-immunized by swapping high-immunogenicity domains. This is a promising approach, however, it will be complex and laborious as Applicants anticipate tens of mutations to achieve stealth, and could result in a reduction in activity and an overall less effective therapy.

[0256] To circumvent this issue, Applicants developed here a framework to compare protein orthologs and their predicted binding to MHC I and MHC II by checking a sliding window of all k-mers in a protein for their presence in another, focusing on peptides predicted to bind to at least one MHC allele. Through this analysis, Applicants identified cliques of Cas9 proteins that are immune orthogonal. Based on these predictions, specific T-cell responses from one ortholog would not cross-react with another ortholog of the same clique, preventing the re-activation of CD8+ cytotoxic T-cells, as well as the CD4+ T-cell help necessary to re-activate memory B-cells. Applicants confirmed these results through ELISAs, and verified two well-characterized Cas9 proteins to be immune orthogonal, SpCas9 and SaCas9. Therefore, Applicants expect that proteins belonging to the same clique can be used sequentially without eliciting memory T- and B- cell responses.

[0257] Due to the importance of AAVs as a delivery agent in gene therapy, Applicants also analyzed AAV serotypes through this MHC I and II comparison framework, and have demonstrated that no two AAVs are mutually immune orthogonal. However, with a known HLA genotype, it may be possible to define a personalized regimen of immune orthogonal AAVs using currently defined serotypes. For instance, use of AAV5 minimizes immune cross-reactivity in mice and primates, as demonstrated by a recent study in which chimeric-AAV5 immunized mice and primates successfully received a second dose of treatment with AAV1[59]. However, in the human setting Applicants predict that there will be substantially more immune overlap between AAV5 and other AAVs. This analysis suggests that creating a pair of globally orthogonal AAV capsids for human application would require [0053] 10 mutations in one of the two proteins. This hypothetical orthogonal AAV capsid presents a substantial engineering challenge, as it requires mutating many of the most conserved regions to achieve immune orthogonality.

[0258] Previous work has identified that MHC affinity is highly dependent on anchor residues at either end of the binding pocket[56]. Residue diversity is more tolerated in the center of the binding pocket, though it may be these residues that most impact antigen specificity, as it is thought that they are central to interaction with the T-cell receptor (TCR). Comparing the number of orthologous pairs in 9-mer space with the number of predicted orthologous pairs based on class II binding predictions suggests that only approximately 65% of 9-mer peptides serve as appropriate MHC class II binding cores, even across the thousands of HLA-2

combinations Applicants explore here. This under-sampling of peptide space by MHC molecules likely reflects the requirement for hydrophobic anchor residues and leaves some space for protein de-immunization by mutation of immunogenic peptides to ones which never serve as MHC binding cores. Achieving this while preserving protein function however, has proven difficult even for few HLA alleles, and remains a significant protein engineering challenge.

[0259] Applicant also notes some limitations to this work. Mainly, Applicants have used inbred C57BL/6J as the mice model, which have very limited MHC diversity,[66] and might not recapitulate other human immunological features, such as differences in antigen processing and presentation. In this regard, Applicants attempted to measure the T-cell response with the ELISPOT assay for a subset of predicted MHC II peptides and indeed confirmed immunogenicity against some, although Applicants also noted the C57BL/6J mice did not show robust responses in general to the AAV-CRJSPRs (FIG. 10). Moving forward, this work can be potentially repeated using other mouse models, such as mice expressing human HLA allotypes, however, these models come with their own technical challenges, such as restricted HLA alleles (representing only main MHC II subgroups) as well as a restricted TCR repertoire[66]. In addition, B-cell epitopes can also be predicted and incorporated into immune orthogonality analysis. However, since B-cell epitopes may be both linear and conformational, these are more difficult to predict. Advances and further validation of these ***in silico*** models will allow for better predictions in the future[67-71]. Finally, recent work has indicated that MHC class I peptides may have significant contribution from spliced host and pathogen-derived peptides created by proteasomal processing[72]. It is unclear how this may affect cross-recognition of proteins Applicants predict to be immune orthogonal. On the one hand, it provides a mechanism whereby very short antigenic sequences spliced to the same host protein may result in cross-recognition of substantially different foreign antigens, however, Applicants expect this to be unlikely due to the massive number of possible spliced peptides between the antigen and entire host proteome.

[0260] Overall, Applicants believe this framework provides a potential solution for efficacious gene therapy, not solely for Cas9-mediated genome engineering, but also for other protein therapeutics that might necessitate repetitive treatments. Although using this approach still requires mitigating the primary immune response, particularly CTL clearance, Applicants

expect that epitope deletion and low-immunogenicity delivery vectors such as AAVs will mitigate this problem, and the potential for repeated dosage will reduce the need for very high first-dose efficiency.

[0261] Computational Methods

[0262] For Cas9, Applicants chose 9 1 orthologs cited in exploratory studies cataloguing the diversity of the Cas9 protein,[73] including several that are experimentally well-characterized. For AAVs, Applicants analyzed 167 sequences, focusing in on all 13 characterized human serotypes, as well as one isolate from rhesus macaque (rh32), one engineered variant (DJ), and one reconstructed ancestral protein (Anc80L65). Applicants then compared total sequence similarity (immunologically uninformed) as well as predicted binding to class I and class II MHC molecules (immunologically informed) between these proteins. Immunologically uninformed sequence comparison was carried out by checking a sliding window of all contiguous k-mers in a protein for their presence in another protein sequence with either zero or one mismatch. Immunologically informed comparison was done in a similar fashion, but using only those k-mers predicted to bind to at least one of 8 1 HLA-1 alleles using netMHC 4.0[74] for class I (alleles can be found at http://www.cbs.dtu.dk/services/NetMHC/MHC_allele_names.txt), and at least one of 5,620 possible MHC II molecules based on 936 HLA-2 alleles using netMHCIIpan 3.1[75] for class II (alleles can be found at http://www.cbs.dtu.dk/services/NetMHCIIpan-3. l/alleles_name.Iist). Applicants compared the use of netMHC to alternative immune epitope prediction platforms such as the Immune Epitope Database (iedb.org)[76] and found very strong agreement across software. Ultimately, Applicants chose netMHC because of the larger number of HLA alleles it supports. Sequences were defined as binding if the predicted affinity ranked in the top 2% of a test library of 400,000 random peptides as suggested in the software guidelines. Generation of immune orthogonal cliques was carried out using the Bron-Kerbosch algorithm. Briefly, a graph was constructed with each ortholog as a vertex, where the edges are defined by the number of shared immunogenic peptides between the connecting vertices. Sets of proteins for which every pair in the set is immune orthogonal constitutes a clique. Phylogenetic distance between protein sequences was measured using the BLOSUM 62 matrix excluding indels. All software, input and output files are available at GitHub.

[0263] Experimental Methods

[0264] AAV Production

[0265] AAV2/8, AAV2/2, AAV2/DJ virus particles were produced using HEK293T cells via the triple transfection method and purified via an iodixanol gradient (Grieger et al., 2006). Confluency at transfection was between 80% and 90%. Media was replaced with pre-warmed media 2 hours before transfection. Each virus was produced in 5 x 15 cm plates, where each plate was transfected with 7.5 μg of pXR-capsid (pXR-8, pXR-2, pXR-DJ), 7.5 of μg recombinant transfer vector, and 22.5 μg of pAd5 helper vector using PEI (lug/uL linear PEI in lx DPBS pH 4.5, using HC1) at a PEI:DNA mass ratio of 4:1. The mixture was incubated for 10 minutes at RT and then applied dropwise onto the media. The virus was harvested after 72 hours and purified using an iodixanol density gradient ultracentrifugation method. The virus was then dialyzed with lx PBS (pH 7.2) supplemented with 50 mM NaCl and 0.0001% of Pluronic F68 (Thermo Fisher) using lOOkDA filters (Millipore), to a final volume of ~ 1 mL and quantified by qPCR using primers specific to the ITR region, against a standard (ATCC VR-1616).

AAV-ITR-F: 5'-CGGCCTCAGTGAGCGA-3' and

AAV-ITR-R: 5'-GGAACCCCTAGTGA TGGAGTT-3'.

[0266] Animal studies

[0267] All animal procedures were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) of the University of California, San Diego. All mice were acquired from Jackson labs. AAV injections were done in adult C57BL/6J mice (10 weeks) through retro-orbital injections using $lxlO^{12}$ vg/mouse.

[0268] ELISA

[0269] *PCSK9:* Levels of serum PCSK9 were measured using the Mouse Proprotein Convertase 9/PCSK9 Quantikine ELISA kit (R&D Systems) according to manufacturer's guidelines. Briefly, serum samples were diluted 1:200 in Calibrator diluent and allowed to bind for 2 h onto microplate wells that were precoated with the capture antibody. Samples were then sequentially incubated with PCSK9 conjugate followed by the PCSK9 substrate solution with extensive intermittent washes between each step. The amount of PCSK9 in serum was estimated colorimetrically using a standard microplate reader (BioRad iMark).

[0270] *Cas9 and AAV:* Recombinant SpCas9 protein (PNA Bio, cat. no. CPOl), or SaCas9 protein (ABM good, cat no. K144), was diluted in lx coating buffer (Bethyl), and 0.5 μg was used to coat each well of 96-well Nunc MaxiSorp Plates (ab210903) overnight at 4 °C. For AAV experiments, 10^9 vg of AAV-2, -5, -8 or -DJ in lx coating buffer was used to coat each well of 96-well Nuc MaxiSorp Plates. Plates were washed three times for 5 min with 350 μï of lx Wash Buffer (Bethyl) and blocked with 300 μï of lx BSA Blocking Solution (Bethyl) for 2 h at RT. The wash procedure was repeated. Serum samples were added at 1:40 dilution, and plates were incubated for 5 h at 4 °C with shaking. Wells were washed three times for 5 min, and 100 μï of HRP-labeled goat anti-mouse IgGl (Bethyl; diluted 1:100,000 in 1% BSA Blocking Solution) was added to each well. After incubating for lhr at RT, wells were washed four times for 5 min, and 100 μï of TMB Substrate (Behtyl) was added to each well. Optical density (OD) at 450 nm was measured using a plate reader (BioRad iMark).

[0271] EXAMPLE 7 - Extremophile Cas9

[0272] Applicants explored the strategy of selecting additional orthologs from extremophile species which would not be expected to come into contact with humans under normal circumstances and/or orthologs from commensal species which are highly abundant in the normal microbiome, perhaps especially at early stages of development, to which the immune system has developed tolerance.

[0273] Applicants mined Cas9 sequences from species fitting into these categories of extremophiles, commensals, pathogens, and non-extreme environmental species. Using these sequences, Applicants explored the orthogonality of Cas9s across these categories to identify orthologs which are good candidates to not cross-react with pre-existing immunity **(FIG. 11).** Although there is broad orthogonality among the extremophile Cas9s, some overlapping peptides are observed when comparing to the larger groups of commensals, pathogens, and environmental species. A few Cas9 orthologs do not show substantial overlap, and these may be useful candidates for characterization, testing, and future use. Furthermore, exploring the diversity of Cas9 orthologs in extreme environments may well provide additional promising targets for immune orthogonality.

Equivalents

[0274] Unless otherwise defined, all technical and scientific terms used herein have the

same meaning as commonly understood by one of ordinary skill in the art to which this technology belongs.

[0275] The present technology illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising," "including," "containing," *etc.* shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the present technology claimed.

[0276] Thus, it should be understood that the materials, methods, and examples provided here are representative of preferred aspects, are exemplary, and are not intended as limitations on the scope of the present technology.

[0277] The present technology has been described broadly and generically herein. Each of the narrower species and sub-generic groupings falling within the generic disclosure also form part of the present technology. This includes the generic description of the present technology with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[0278] In addition, where features or aspects of the present technology are described in terms of Markush groups, those skilled in the art will recognize that the present technology is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[0279] All publications, patent applications, patents, and other references mentioned herein are expressly incorporated by reference in their entirety, to the same extent as if each were incorporated by reference individually. In case of conflict, the present specification, including definitions, will control.

[0280] Other aspects are set forth within the following claims.

References

1. Chew W, et al. (2016) A multifunctional AAV-CRISPR-Cas9 and its host response. Nature Methods, 13(10):868-874.

2. Wang D, Mou H, Li S, Li Y, Hough S, Tran K, et al. AdenovirusMediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. Hum Gene Ther. 2015;26

3. Riechmann L, et al. (1988) Reshaping human antibodies for therapy. Nature 332:323-327.

4. Lundegaard C, et al. (201 0) "Major Histocompatibility Complex Class I Binding Predictions as a Tool in Epitope Discovery." Immunology 130.3 (2010): 309-318. PMC. Web. 7 Nov. 2016.

5. Massimo A, et al. (201 6) Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics, 32(4):51 17.

6. Fonfara I, et al. (2014) Phylogeny of Cas9 Determines Functional Exchangeability of Dual-RNA and Cas9 among Orthologous Type II CRISPR-Cas Systems. Nucleic Acids Research 42.4: 2577-2590.

7. Choi Y and Chan AP (201 5) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31(16): 27452747.

8. Massimo Andreatta and Morten Nielsen. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics, Feb 15;32(4):51 l7 2016.

9. Tong, SYC et al. (201 5) Staphylococcus aureus Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. Clinical Microbiology Reviews. 28: 603661 .

10. Mingozzi, F. & High, K. A. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. ***Blood*** **122,** 23-36 (2013).

11. Mays, L. E. & Wilson, J. M. The Complex and Evolving Story of T cell Activation to AAV Vector-encoded Transgene Products. ***Mol. Ther.*** **19,** 16-27 (201 1).

12. Basner-Tschakarjan, E., Bijjiga, E. & Martino, A. T. Pre-clinical assessment of immune responses to adeno-associated virus (AAV) vectors. ***Front. Immunol.*** **5,** (2014).

13. Ertl, H. C. J. & High, K. A. Impact of AAV Capsid-Specific T-Cell Responses on Design and Outcome of Clinical Gene Transfer Trials with Recombinant Adeno-Associated Viral Vectors: An Evolving Controversy. ***Hum. Gene Ther.*** **28,** 328-337 (2017).

14. Kotterman, M. A., Chalberg, T. W. & Schaffer, D. V. Viral Vectors for Gene Therapy: Translational and Clinical Outlook. ***Annu. Rev. Biomed. Eng.*** **17,** 63-89 (2015).

15. Mingozzi, F. & High, K. A. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. ***Nat. Rev. Genet.*** **12,** 341-355 (201 1).

16. Manno, C. S. ***et al.*** Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. ***Nat. Med.*** **12,** 342-347 (2006).

17. Sathish, J. G. ***et al.*** Challenges and approaches for the development of safer immunomodulatory biologies. ***Nat Rev Drug Discov*** **12,** 306-324 (2013).

18. Harding, F. A., Stickler, M. M., Razo, J. & DuBridge, R. B. The immunogenicity of humanized and fully human antibodies: Residual immunogenicity resides in the CDR regions. ***MAbs*** **2,** 256-265 (2010).

19. De Groot, a S., Knopp, P. M. & Martin, W. De-immunization of therapeutic proteins by T-cell epitope modification. ***Dev. Biol. (Basel).*** **122,** 171-194 (2005).

20. Tangri, S. *et al.* Rationally Engineered Therapeutic Proteins with Reduced Immunogenicity. *J. Immunol.* **174,** 3187 IJP-3196 (2005).

21. Salvat, R. S., Choi, Y., Bishop, A., Bailey-Kellogg, C. & Griswold, K. E. Protein deimmunization via structure-based design enables efficient epitope deletion at high mutational loads. ***Biotechnol. Bioeng. Ill,*** 1306-1318 (2015).

22. Armstrong, J. K. *et al.* Antibody against poly(ethylene glycol) adversely affects PEG-asparaginase therapy in acute lymphoblastic leukemia patients. ***Cancer*** **110,** 103-1 11 (2007).

23. Ganson, N. J., Kelly, S. J., Scarlett, E., Sundy, J. S. & Hershfield, M. S. Control of hyperuricemia in subjects with refractory gout, and induction of antibody against poly(ethylene glycol) (PEG), in a phase I trial of subcutaneous PEGylated urate oxidase. ***Arthritis Res. Ther.*** 8, R12-R12 (2006).

24. Veronese, F. M. & Mero, A. The impact of PEGylation on biological therapies. ***BioDrugs*** **22,** 315-329 (2008).

25. Jevsevar, S., Kunstelj, M. & Porekar, V. G. PEGylation of therapeutic proteins. ***Biotechnol. J.*** **5,** 113-128 (2010).

26. Jacobs, F., Gordts, S. C , Muthuramu, I. & De Geest, B. The liver as a target organ for gene therapy: state of the art, challenges, and future perspectives. ***Pharmaceuticals (Basel).*** **5,** 1372-92 (2012).

27. Kok, C. Y. *et al.* Adeno-associated Virus-mediated Rescue of Neonatal Lethality in Argininosuccinate Synthetase-deficient Mice. ***Mol. Ther.*** **21,** 1823-1831 (2013).

28. Courtenay-Luck, N. S., Epenetos, A. A. & Moore, R. Development of primary and secondary immune responses to mouse monoclonal antibodies used in the diagnosis and therapy of malignant neoplasms. ***Cancer Res.*** **46,** 6489-6493 (1986).

29. Jinek, M. *et al.* A Programmable Dual-RNA - Guided DNA Endonuclease in Adaptice Bacterial Immunity. ***Science*** **337,** 816-822 (2012).

30. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. ***Science*** **339,** 823-6 (2013).

31. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. ***Proc. Natl. Acad. Sci.*** **109,** E2579-E2586 (2012).

32. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. ***Science*** **339**, 819-23 (2013).

33. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. ***Nature*** **520,** 186-190 (2015).

34. Jinek, M. *et al.* RNA-programmed genome editing in human cells. ***Elife*** **2013,** (2013).

35. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. ***Nat. Methods*** **10,** 957-963 (2013).

36. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRJSPR-Cas9 for genome engineering. ***Cell*** **157,** 1262-1278 (2014).

37. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. ***Nat. Rev. Microbiol.*** **13,** 722-736 (2015).

38. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. ***Nucleic Acids Research*** **42,** 6091-6105 (2014).

39. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. ***Nat. Rev. Microbiol.*** **15,** 169-182 (2017).

40. Wagner, J. a *et al.* Safety and biological efficacy of an adeno-associated virus vector-cystic fibrosis transmembrane regulator (AAV-CFTR) in the cystic fibrosis maxillary sinus. ***Laryngoscope*** **109,** 266-74 (1999).

41. Song, S. *et al.* Sustained secretion of human alpha-1-antitrypsin from murine muscle transduced with adeno-associated virus vectors. ***Proc. Natl. Acad. Sci. U. S. A.* 95,** 14384-8 (1998).

42. Chirmule, N. *et al.* Humoral Immunity to Adeno-Associated Virus Type 2 Vectors following Administration to Murine and Nonhuman Primate Muscle. ***J. Virol.* 74,** 2420-2425 (2000).

43. Fields, P. a *et al.* Risk and prevention of anti-factor IX formation in AAV-mediated gene transfer in the context of a large deletion of F9. ***Mol. Ther.* 4,** 201-210 (2001).

44. Herzog, R. W. *et al.* Influence of vector dose on factor IX-specific T and B cell responses in muscle-directed gene therapy. ***Hum. Gene Ther.* 13,** 1281-91 (2002).

45. Lozier, J. N., Tayebi, N. & Zhang, P. Mapping of genes that control the antibody response to human factor IX in mice. ***Blood* 105,** 1029-1035 (2005).

46. Zhang, H. G. *et al.* Genetic analysis of the antibody response to AAV2 and factor IX. ***Mol.Ther.* 11,** 866-874 (2005).

47. Benveniste, O. *et al.* Prevalence of Serum IgG and Neutralizing Factors Against Adeno-Associated Virus (AAV) Types 1,2,5,6,8, and 9 in the Healthy Population: Implications for Gene Therapy Using AAV Vectors. ***Hum. Gene Ther.* 21,** 704-712 (2010).

48. Gao, G.-P. *et al.* Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. ***Proc. Natl. Acad. Sci.* 99,** 1 1854-1 1859 (2002).

49. Jooss, K., Yang, Y., Fisher, K. J. & Wilson, J. M. Transduction of Dendritic Cells by DNA Viral Vectors Directs the Immune Response to Transgene Products in Muscle Fibers. ***J. Virol.* 72,** 4212^1223 (1998).

50. Gernoux, G. *et al.* Early Interaction of Adeno-Associated Virus Serotype 8 Vector with the Host Immune System Following Intramuscular Delivery Results in Weak but

Detectable Lymphocyte and Dendritic Cell Transduction. ***Hum. Gene Ther.*** **26,** 1-13 (2015).

51. Zhu, J., Huang, X. & Yang, Y. The TLR9-MyD88 pathway is critical for adaptive immune responses to adeno-associated virus gene therapy vectors in mice. ***J. Clin. Invest.*** **119,** 2388-2398 (2009).

52. Gernoux, G., Wilson, J. M. & Mueller, C. Regulatory and Exhausted T Cell Responses to AAV Capsid. ***Hum. Gene Ther.*** **28,** 338-349 (2017).

53. Kurosaki, T., Kometani, K. & Ise, W. Memory B cells. ***Nat. Rev. Immunol.*** **15,** 149—159 (2015).

54. Zabel, F. ***et al.*** Distinct T helper cell dependence of memory B-cell proliferation versus plasma cell differentiation. ***Immunology*** **150,** 329-342 (2017).

55. Ding, Q. ***et al.*** Permanent Alteration of PCSK9 With In Vivo CRISPR-Cas9 Genome Editing. ***Circ. Res.*** **115,** 488-492 (2014).

56. Zinn, E. ***et al.*** In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. ***Cell Rep.*** **12,** 1056-1068 (2017).

57. Calcedo, R. & Wilson, J. M. AAV Natural Infection Induces Broad Cross-Neutralizing Antibody Responses to Multiple AAV Serotypes in Chimpanzees. ***Hum. Gene Ther. Clin. Dev.*** **27,** 79-82 (2016).

58. Harbison, C. E. ***et al.*** Examining the cross-reactivity and neutralization mechanisms of a panel of mabs against adeno-associated virus serotypes 1 and 5. ***J. Gen. Virol.*** **93,** (2012).

59. Majowicz, A. ***et al.*** Successful Repeated Hepatic Gene Delivery in Mice and Non-human Primates Achieved by Sequential Administration of AAV5^{ch} and AAV1. ***Mol. Ther.*** **25,** 1831-1842 (2017).

60. Mcintosh, J. H. *et al.* Successful attenuation of humoral immunity to viral capsid and transgenic protein following AAV-mediated gene transfer with a non-depleting CD4 antibody and cyclosporine. *Gene Ther* **19,** 78-85 (2012).

61. Mingozzi, F. *et al.* Prevalence and pharmacological modulation of humoral immunity to AAV vectors in gene transfer to synovial tissue. *Gene Ther* **20,** 417-424 (2013).

62. Mingozzi, F. *et al.* Pharmacological Modulation of Humoral Immunity in a Nonhuman Primate Model of AAV Gene Transfer for Hemophilia B. *Mol. Ther.* **20,** 1410-1416 (2017).

63. Unzu, C. *et al.* Transient and intensive pharmacological immunosuppression fails to improve AAV-based liver gene transfer in non-human primates. *J. Transl. Med.* **10,** 122 (2012).

64. Riechmann, L., Clark, M., Waldmann, H. & Winter, G. Reshaping human antibodies for therapy. *Nature* **332,** 323-7 (1988).

65. Ruppert, J. *et al.* Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. **Ce// 74,** 929-937 (2017).

66. Baker, M. P., Reynolds, H. M., Lumicisi, B. & Bryson, C. J. Immunogenicity of protein therapeutics: The key causes, consequences and challenges. *Self Nonself* **1**, 314-322 (2010).

67. EL-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.* **21,** 243-255 (2008).

68. Larsen, J. E. P., Lund, O. & Nielsen, M. Improved method for predicting linear B-cell epitopes. *Immunome Res. 2, 2* (2006).

69. Sollner, J. *et al.* Analysis and prediction of protective continuous B-cell epitopes on pathogen proteins. *Immunome Res. 4,* 1 (2008).

70. Dalkas, G. A. & Rooman, M. SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence. *BMC Bioinformatics* **18,** 95 (2017).

71. Sun, P. *et al.* Bioinformatics resources and tools for conformational B-cell epitope prediction. *Computational and Mathematical Methods in Medicine* **2013,** (2013).

72. Liepe, J. *et al.* A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science (80-.)*. **354,** (2016).

73. Fonfara, I. *et al.* Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **42,** 2577-2590 (2014).

74. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32,** 511-517 (2015).

75. Andreatta, M. *et al.* Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* 67, 641-650 (2015).

76. Vita, R. *et al.* The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **43,** D405-12 (2015).

77. Giiell, M., Yang, L. & Church, G. M. Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics* **30,** 2968-2970 (2014).

Table 1

Peptide	Pos	Peptide	ID	Allele	Affinity L	n-mer	Score	Actual po	Surface	Mutation		Pos	Peptide	nM	Rank	ID	Allele	Affinity Level	n-mer
1	197	IVDEVA-Y	Streptococcus_P	HLA-A01:01	High	8	0	121-128	Yes	Y128N	142	0	IVDEVA-N	38015	80	Streptococcus_I	HLA-A0101	No	E
2	1196	LFGNIJAL	Streptococcus_P	HLA-A02:01	High	9	0	236-244	No	L236H	28	0	-HFGNIJAL	6395.2	11	Streptococcus_-	HLA-A0201	No	S
3	2523	LEDIVLTL	Streptococcus_P	HLA-A02:01	High	9	0	614-623	No	L615D and E616V	23	0	IDEDIVLTL	14725.8	21	Streptococcus_I	HLA-A0201	No	S
4	2454	GTYHDL-L-R	Streptococcus_P	HLA-A03:01	High	10	0	591-599	No	K599D	183	0	GTYHDL-L-D	12613.3	11	Streptococcus_G	HLA-A0301	No	LD
5	2189	ETITPWNF	Streptococcus_P	HLA-A26:01	High	8	0	470-478	Yes	T471C	24	0	ECITPWNF	34785.2	60	Streptococcus_E	HLA-A2601	No	E
6	196	NIVDEVA-Y	Streptococcus_P	HLA-A26:01	High	9	0	120-128	Yes	Y128N	162	0	NIVDEVA-N	16237.6	6.5	Streptococcus_N	HLA-A2601	No	S
7	3215	EVVKKMKNY	Streptococcus_P	HLA-A26:01	High	9	0	873-882	Yes	Y882N	162	0	EVVKKMKNN	11513.6	4	Streptococcus_E	HLA-A2601	No	S
8	2141	IPYYVGPL	Streptococcus_P	HLA-B07:02	High	8	0.063905	447-455	No	P448C	24	0	ICYYVGPL	34157.5	48	Streptococcus_I	HLA-B0702	No	E
9	40	PSKKFKVL	Streptococcus_P	HLA-B07:02	High	9	0	27-35	Yes	P27D	23	0	-DSKKFKVL	26712.9	25	Streptococcus_-	HLA-B0702	No	S
10	40	PSKKFKVL	Streptococcus_P	HLA-B08:01	High	9	0	27-35	Yes	P27D	23	0	-DSKKFKVL	26712.9	25	Streptococcus_-	HLA-B0702	No	S
11	2613	LKRRRYTG	Streptococcus_P	HLA-B08:01	High	9	0	650-658	No	R653P	94	0	-LKRPRYTG	3079.4	3.5	Streptococcus_-	HLA-B0801	Low	S
12	2617	RRYTGWG-	Streptococcus_P	HLA-B27:05	High	8	0	653-660	No	R653P	14	0	PRYTGWG-	7382.2	7.5	Streptococcus_P	HLA-B2705	No	E
13	2165	SRFAWMTRK	Streptococcus_P	HLA-B27:05	High	9	0	459-468	Yes	R460Q	23	0	SDFAWMTRK	13809.5	12	Streptococcus_S	HLA-B2705	No	S
14	1688	HHQDLTLL	Streptococcus_P	HLA-B39:01	High	8	0	327-335	Yes	H328D	23	0	HDQDLTLL	14138.9	6	Streptococcus_H	HLA-B3901	No	L
15	1689	HQDLTLLKAL	Streptococcus_P	HLA-B39:01	High	10	0	328-338	Yes	H328D	23	0	HDQDLTLKAI	22112.3	12	Streptococcus_H	HLA-B3901	No	LI
16	1688	HHQDLTLKAL	Streptococcus_P	HLA-B39:01	High	11	0	327-338	Yes	H328D	3	0	DQDLTLKAL	17106.3	8	Streptococcus_D	HLA-B3901	No	LD
17	2524	LEDVLTL	Streptococcus_P	HLA-B40:01	High	8	0	615-623	No	L615D and E616V	39	0	LVDIVLTL	29463.8	34	Streptococcus_L	HLA-B4001	No	L
18	2556	REMEERL	Streptococcus_P	HLA-B40:01	High	8	0.020848	628-638	No	E629P	34	0	RPMIEERL	35930.5	65	Streptococcus_R	HLA-B4001	No	L
19	2445	EDRFNASL	Streptococcus_P	HLA-B40:01	High	9	0	583-591	No	E583G	27	0	-GDRFNASL	20438.3	13	Streptococcus_-	HLA-B4001	No	L
20	3045	KELGS-QIL	Streptococcus_P	HLA-B40:01	High	9	0	788-796	Yes	E788G	27	0	KGLGS-QIL	17094.3	10	Streptococcus_K	HLA-B4001	No	L
21	2524	LEDVLTLTL	Streptococcus_P	HLA-B40:01	High	10	0	615-625	No	L615D and E616V	39	0	LVDIVLTLTL	19888.8	13	Streptococcus_L	HLA-B4001	No	LD
22	2323	KAIV-DLLF	Streptococcus_P	HLA-B58:01	High	9	0	545-553	Yes	F553R	161	0	KAIV-DLLR	4677.9	4.5	Streptococcus_K	HLA-B5801	Low	L

Contd.; same rows:

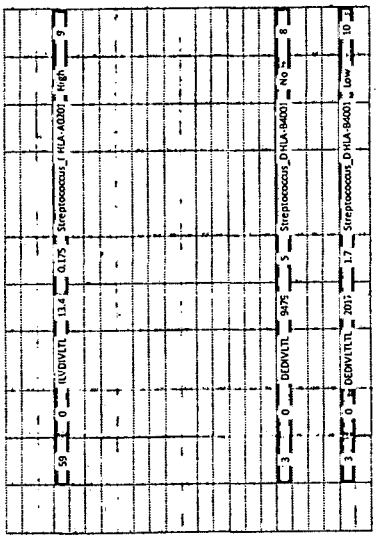


Table 2

sgID [gene_strandtargeted_PAMcoordinate.sgRNAlength-transcript]	gene [gene targeted by the sgRNA, or "negative_control"]	transcript [TSS targeted by the sgRNA]	protospacer sequence [protospacer sequence; 5'G is included whether or not it is present in the genome]
SCN3A_+_166060543.23-P1P2	SCN3A	P1P2	GATCTCAGAACAGGAAGCGG
SCN3A_+_166060199.23-P1P2	SCN3A	P1P2	GTGTAAATTACAGGAACCAA
SCN3A_-_166060301.23-P1P2	SCN3A	P1P2	GACCTGGTAGCTAGGTTCTA
SCN3A_+_166060552.23-P1P2	SCN3A	P1P2	GATAGAGTGAATCTCAGAAC
SCN3A_+_166060129.23-P1P2	SCN3A	P1P2	GAATAGAGCCTGTCTGGAAA
SCN3A_+_166060346.23-P1P2	SCN3A	P1P2	GTGTTATGCTGTAATTCATA
SCN3A_+_166060119.23-P1P2	SCN3A	P1P2	GGTCTGGAAATGGTGATTTA
SCN3A_+_166060135.23-P1P2	SCN3A	P1P2	GAAAGAAAATAGAGCCTGTC
SCN3A_+_166060371.23-P1P2	SCN3A	P1P2	GCCTAACCATCTTGGATGCT
SCN3A_+_166060281.23-P1P2	SCN3A	P1P2	GACCATAGAACCTAGCTACC
SCN9A_+_167232419.23-P1P2	SCN9A	P1P2	GGCGGTCGCCAGCGCTCCAG
SCN9A_+_167232052.23-P1P2	SCN9A	P1P2	GCCACCTGGAAAGAAGAGAG
SCN9A_+_167232416.23-P1P2	SCN9A	P1P2	GGTCGCCAGCGCTCCAGCGG
SCN9A_+_167232010.23-P1P2	SCN9A	P1P2	GCCAGCAATGGGAGGAAGAA
SCN9A_-_167232085.23-P1P2	SCN9A	P1P2	GTTCCAGGTGGCGTAATACA
SCN9A_+_167232476.23-P1P2	SCN9A	P1P2	GGCGGGGCTGCTACCTCCAC
SCN9A_+_167232437.23-P1P2	SCN9A	P1P2	GGGCGCAGTCTGCTTGCAGG
SCN9A_+_167232409.2	SCN9A	P1P2	GGCGCTCCAGCGGCGGCTG

3-P1P2			T
SCN9A_+_167232021.23-P1P2	SCN9A	P1P2	GACCGGGTGGTTCCAGCAAT
SCN9A_+_167232018.23-P1P2	SCN9A	P1P2	GGGGTGGTTCCAGCAATGGG
SCN10A_-_38835462.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GTGACTCCGGAGTAAAGCGA
SCN10A_-_38835311.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GGGAGCTCACCATAGAACTT
SCN10A_-_38835269.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GACGGATCTAGATCCTCCAG
SCN10A_+_38835213.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GCCGGGTAAGAGCTACTAGT
SCN10A_-_38835251.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GCCCGGTGTGTGCTGTAGAA
SCN10A_+_38835434.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GTTTACTCCGGAGTCACTGG
SCN10A_-_38835449.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GCTATCTCCACCAGTGACTC
SCN10A_-_38835156.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GACATCACCCAGGGCCAAGG
SCN10A_-_38835491.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GTAGTTTCGAGGGATCCAAT
SCN10A_+_38835272.23-ENST00000449082.2	SCN10A	ENST00000449082.2	GCTCCCAGCAGAACTGATCG
SCN11A_-_38991624.23-ENST00000302328.3,ENST00000450244.1	SCN11A	ENST00000302328.3,ENST00000450244.1	GATGGGTCCAAGTCTTCCAG
SCNllA_+_38992032.23-ENST00000302328.3,ENST00000450244.1	SCN11A	ENST00000302328.3,ENST00000450244.1	GGTTCCTGCTATACCCACAG
SCN11A_-_38991801.23-ENST00000302328.3,EN	SCN11A	ENST00000302328.3,ENST00000450244.1	GCCAGAGAGTCGGAAGTGAA

ST00000450244.1			
SCN llA_+_38992029.23-ENST00000302328.3, ENST00000450244.1	SCN11A	ENST00000302328.3, ENST00000450244.1	GCCTGCTATACCCACAGTGG
SCN 11A_+_38991609.23-ENST00000302328.3,ENST00000450244.1	SCN 11A	ENST00000302328.3,ENST00000450244.1	GGGAAAGCCTCTGGAAGACT
SCN11A_-_38992040.23-ENST00000302328.3,ENST00000450244.1	SCN 11A	ENST00000302328.3,ENST00000450244.1	GGAAGAGATGACCACCACTG
SCN 11A_-_38991666.23-ENST00000302328.3,ENST00000450244. 1	SCN11A	ENST00000302328.3,ENST00000450244.1	GGAATGTCGCCATAGAGCTT
SCN 11A_+_38991618.23-ENST00000302328.3 ,ENST00000450244.1	SCN11A	ENST00000302328.3,ENST00000450244. 1	GGAGCTCATAGGAAAGCCTC
SCN 11A_+_38991924.23-ENST00000302328.3, ENST00000450244.1	SCN 11A	ENST00000302328.3,ENST00000450244.1	GCTTTAAGACTGGAATCCTA
SCN 11A_+_38991653.23-ENST00000302328.3,ENST00000450244.1	SCN 11A	ENST00000302328.3, ENST00000450244.1	GGGAAGTTGCCCAAGCTCTA
SHAN K3_+_51135959.23-P1P2	SHANK3	P1P2	GGAATTCGAATACAGCTCCT
SHANK3_+_51136404.23-P1P2	SHANK3	P1P2	GCTTCAGGCAGAGACCCCCG
SHAN K3_+_51136356.23-P1P2	SHANK3	P1P2	GGAGCCTCCGTGGTGACACA
SHAN K3_+_51136302.23-P1P2	SHANK3	P1P2	GCACGGCAGGAACCTTCCCC
SHANK3_+_5H36319.23-P1P2	SHANK3	P1P2	GAGCACCGGAGGGACCCGCA
SHANK3_+_51136333.23-P1P2	SHANK3	P1P2	GGCCCGGAACGACAGAGCAC
SHANK3_+_51136329.23-P1P2	SHANK3	P1P2	GGGAACGACAGAGCACCGGA

SHANK3_-_51136143.23-P1P2	SHANK3	P1P2	GACcgcggcgaggccgtgaa
SHANK3_-_51136336.23-P1P2	SHANK3	P1P2	GCCTGCCGTGCGGGTCCCTC
SHANK3_+_51135950.23-P1P2	SHANK3	P1P2	GTACAGCTCCTGGGCGCGCC
TRPVl_+_3500355.23-P1P2	TRPV1	P1P2	GAGCGACTCCTGCTAGTGCA
TRPV1_+_3500317.23-P1P2	TRPV1	P1P2	GCGGGCCCGGGACCCCACGG
TRPVl_+_3499964.23-P1P2	TRPV1	P1P2	GCTCCTTGGAAGCACCTGGG
TRPV1_-_3500391.23-P1P2	TRPV1	P1P2	GAGTCGCTGTGGACGCCCTT
TRPVl_-_3500224.23-P1P2	TRPV1	P1P2	GGGACTCACCAGCTAGACGC
TRPVl_-_3500327.23-P1P2	TRPV1	P1P2	GTGGTCTCCCCGCCTCCGTG
TRPVl_-_3500298.23-P1P2	TRPV1	P1P2	GGGGAGAGCTGGGCTCGTGT
TRPV1_+_3500017.23-P1P2	TRPV1	P1P2	Gtgcctcaaaggtggtcgtg
TRPVl_+_3499899.23-P1P2	TRPV1	P1P2	GCTGCATCAGCCGTCCTCGG
TRPVl_-_3500400.23-P1P2	TRPV1	P1P2	GGGACGCCCTTCGGCACTCA
GRIN2B_-_14133341.23-P1P2	GRIN2B	P1P2	GGATTCGCGTGTCCCCCGGA
GRIN2B_+_14132929.23-P1P2	GRIN2B	P1P2	GGATATGCAAGCGAGAAGAA
GRIN2B_-_14132903. 23-P1P2	GRIN2B	P1P2	GCTCTAGACGGACAGATTAA
GRIN2B_-_14133316.23-P1P2	GRIN2B	P1P2	GGGGGAAAAAGAGGCGGTCA
GRIN2B_+_14132924.23-P1P2	GRIN2B	P1P2	GGCAAGCGAGAAGAAGGGAC
GRIN2B_-_14133295.23-P1P2	GRIN2B	P1P2	GCCAAAGCGTCCCCTTCCTA
GRIN2B_-_14133298.23-P1P2	GRIN2B	P1P2	GAAGCGTCCCCTTCCTAAGG
GRIN2B_+_14132855.23-P1P2	GRIN2B	P1P2	GGCTTCTACAAACCAAGGTA
GRIN2B_+_14133247.23	GRIN2B	P1P2	GACCATGCTCCACCGAGGG

-P1P2			A
GRIN2B_+_14133252.23-P1P2	GRIN2B	P1P2	GGAATGACCATGCTCCACCG
PRDM12_-_133540047.23-P1P2	PRDM12	P1P2	GgctccgggccgcccATGAT
PRDM12_+_133540034.23-P1P2	PRDM12	P1P2	GGCACGGAGCCCATCATggg
PRDM12_+_133540230.23-P1P2	PRDM12	P1P2	GGACTGCGCCAGCACCTCGG
PRDM12_+_133539846.23-P1P2	PRDM12	P1P2	Gctgggaggaaagcgaacga
PRDM12_-_133540263.23-P1P2	PRDM12	P1P2	GTGGCGCAGTCCTTCTCCGG
PRDM12_-_133540260.23-P1P2	PRDM12	P1P2	GTGCTGGCGCAGTCCTTCTC
PRDM12_+_133540257.23-P1P2	PRDM12	P1P2	GCGACGGCTGGACTCACCGC
PRDM12_+_133540233.23-P1P2	PRDM12	P1P2	GAAGGACTGCGCCAGCACCT
PRDM12_-_133540304.23-P1P2	PRDM12	P1P2	GCCGGCGCAATCCCTCCTCC
PRDM12_+_133539961.23-P1P2	PRDM12	P1P2	Ggggcgagaggggagcccaa
HCN2_+_589972.23-P1P2	HCN2	P1P2	Gtcgcgcccgggctctcccc
HCN2_+_590106.23-P1P2	HCN2	P1P2	GCAACGCCTcggcccggggc
HCN2_+_589880.23-P1P2	HCN2	P1P2	GgccgccggccggAGCCCGA
HCN2_+_590306.23-P1P2	HCN2	P1P2	GcggcACGAGAACGACACCT
HCN2_-_590253.23-P1P2	HCN2	P1P2	GCAGCCCGAACGGCGAGTGC
HCN2_+_590235.23-P1P2	HCN2	P1P2	GGCGCCCGCACTCGCCGTTC
HCN2_-_590335.23-P1P2	HCN2	P1P2	GTCGTTCTCGTgccgcgggg
HCN2_+_590407.23-P1P2	HCN2	P1P2	GAGCTGGCCTGGCTgccgcg
HCN2_-_590332.23-P1P2	HCN2	P1P2	GGTGTCGTTCTCGTgccgcg
HCN2_+_590204.23-P1P2	HCN2	P1P2	GGCCGTGCTcgccgcgcccg

Table 3

sgID [gene_strandtargeted_PAMcoordinate.sgRNAlength-transcript]	gene [gene targeted by the sgRNA, or "negative_control"]	transcript [TSS targeted by the sgRNA]	protospacer sequence [protospacer sequence; 5'G is included whether or not it is present in the genome]
Scn3a_+_65567459.23-P1P2	Scn3a	P1P2	GTGAATCTCAGAACAGGAAG
Scn3a_+_65567442.23-P1P2	Scn3a	P1P2	GAGCGGAGGCATAAGCAGAA
Scn3a_-_65567234.23-P1P2	Scn3a	P1P2	GATCTGGTGGCTAGATTCTA
Scn3a_-_65567301.23-P1P2	Scn3a	P1P2	GAGGAATCACAGCTCAACAA
Scn3a_-_65567522.23-P1P2	Scn3a	P1P2	GATCAGAAAACGGCCCTGGA
Scn3a_-_65567271.23-P1P2	Scn3a	P1P2	GGTTTTGTCAGCTTACCTGA
Scn3a_-_65567326.23-P1P2	Scn3a	P1P2	GGCATCCAAGATGGTTAGAA
Scn3a_+_65567264.23-P1P2	Scn3a	P1P2	GATTCCTAAGGCTCTCCATC
Scn3a_+_65567031.23-P1P2	Scn3a	P1P2	GCAATACAGACTAGGAATTA
Scn9a_+_66634758.23-P1P2	Scn9a	P1P2	GAGCTCAGGGAGCATCGAGG
Scn9a_-_66634675.23-P1P2	Scn9a	P1P2	GAGAGTCGCAATTGGAGCGC
Scn9a_-_66634637.23-P1P2	Scn9a	P1P2	GCCAGACCAGCCTGCACAGT
Scn9a_-_66634689.23-P1P2	Scn9a	P1P2	GAGCGCAGGCTAGGCCTGCA
Scn9a_-_66634610.23-P1P2	Scn9a	P1P2	GCTAGGAGTCCGGGATACCC
Scn9a_+_66634478.23-P1P2	Scn9a	P1P2	GAATCCGCAGGTGCACTCAC
Scn9a_-_66634641.23-P1P2	Scn9a	P1P2	GACCAGCCTGCACAGTGGGC
Scn9a_+_66634731.23-	Scn9a	P1P2	GCGACGCGGTTGGCAGCCG

P1P2			A
Scnl0a_+_119719110.23-P1P2	ScnlOa	P1P2	GGCAGGGTGGAACTCGTGAC
Scnl0a_+_U9719123.23-P1P2	ScnlOa	P1P2	GCACCATCCAGCAAGCAGGG
Scnl0a_-_119719078.23-P1P2	ScnlOa	P1P2	GCGTCACTCAAGGATCTACA
Scnl0a_+_119719086.23-P1P2	ScnlOa	P1P2	GATGGGAATGGCACCCACGA
Scnl0a_+_119718921.23-P1P2	ScnlOa	P1P2	GCCTTTAGACGGAGAACAGA
Scnl0a_+_H9719051.23-P1P2	ScnlOa	P1P2	GAGATCCTTGAGTGACGGAC
Scnl0a_-_119719025.23-P1P2	ScnlOa	P1P2	GCGGGGCTCCTCCACGAAGG
Scnl0a_-_119719095.23-P1P2	ScnlOa	P1P2	GCAAGGAATCACGCCTTCGT
Scnl0a_+_119718881.23-P1P2	ScnlOa	P1P2	GGCCATGCGCGAATGCTGAG
Scnl0a_+_119719014.23-P1P2	ScnlOa	P1P2	GGCAAGCCCAGCCACCTTCG
Scnlla_+_119825404.23-P1P2	Scnlla	P1P2	GAGGTAAGCCATCCAGGCTG
Scnlla_-_119825450.23-P1P2	Scnlla	P1P2	GTTCCTGCTAGGGAGGCTCA
Scnlla_-_119825400.23-P1P2	Scnlla	P1P2	GCCTGAAACGACAGAGGATG
Scnlla_+_119825277.23-P1P2	Scnlla	P1P2	GTCAGAGGTGGAGACCAGGT
Scnlla_-_119825394.23-P1P2	Scnlla	P1P2	GCCCCAGCCTGAAACGACAG
Scnlla_+_119825463.23-P1P2	Scnlla	P1P2	GGCCAAGAGCGAGAATCTCC
Scnlla_+_119825246.23-P1P2	Scnlla	P1P2	GGTCAGGTGTCAGAGCCCAT
Scnlla_+_119825242.23-P1P2	Scnlla	P1P2	GGGTGTCAGAGCCCATCGGT
Scnlla_+_119825431.23-P1P2	Scnlla	P1P2	GTGCCCTGAGCCTCCCTAGC
Scnlla_-_119825253.23-P1P2	Scnlla	P1P2	GTCTGTGAGAACCGACCGAT
Shank3_+_89499659.23-P1P2	Shank3	P1P2	GGGCTCCGCAGGCGCAGCGG

Shank3_+_89499688.23-P1P2	Shank3	P1P2	GgggccagcgcgggggACAG
Shank3_+_89499943.23-P1P2	Shank3	P1P2	GCCGCTAGCGGGCCACACAG
Shank3_+_89499679.23-P1P2	Shank3	P1P2	GcgggggACAGCGGCTCCGG
Shank3_+_89499612.23-P1P2	Shank3	P1P2	GCATCGGCCCCGGCTTCGAG
Shank3_+_89499924.23-P1P2	Shank3	P1P2	GGGGTACGGCGAGATCGCAA
Shank3_+_89499878.23-P1P2	Shank3	P1P2	GATGCCGACGCGCACGACCA
Shank3_-_89499676.23-P1P2	Shank3	P1P2	GGCCGCCGCCGCTGCGCCTG
Shank3_+_89499818.23-P1P2	Shank3	P1P2	GGGGCCCGGACTGTTCCCGG
Shank3_+_89499938.23-P1P2	Shank3	P1P2	GAGCGGGCCACACAGGGGTA
Trpvl_+_73234353.23-P1P2	Trpvl	P1P2	GGGACTTACCAGCTAGGTGC
Trpvl_-_73234330.23-P1P2	Trpvl	P1P2	GCCCACAAAGAACAGCTCCA
Trpvl_-_73234384.23-P1P2	Trpvl	P1P2	GGCTGGTAAGTCCTTCTCAT
Trpvl_+_73234339.23-P1P2	Trpvl	P1P2	GGGTGCAGGCACACTCCAAA
Trpvl_-_73234537.23-P1P2	Trpvl	P1P2	GACTTAACTTGGCTGACTGT
Trpvl_+_73234478.23-P1P2	Trpvl	P1P2	GTCAGCCTCCCAGAAGTCCA
Trpvl_-_73234495.23-P1P2	Trpvl	P1P2	GGCTGCCTTGGACTTCTGGG
Trpvl_+_73234635.23-P1P2	Trpvl	P1P2	GCCACGGAAGGCCTCCAGAT
Trpvl_-_73234346.23-P1P2	Trpvl	P1P2	GCCAAGGCACTTGCTCCATT
Trpvl_+_73234280.23-P1P2	Trpvl	P1P2	GGGCTGCTGTGTGGTAAGAG
Grin2b_-_136172154.23-P1P2	Grin2b	P1P2	GCCAACCTGAATGGAAGAGA
Grin2b_-_136172179.23-P1P2	Grin2b	P1P2	GAGGGAAGTGGAAAGCAAGG
Grin2b_-	Grin2b	P1P2	GTGGGACAGGCATGGATGA

_136172123.23-P1P2			A
Grin2b>_136172089.23-P1P2	Grin2b	P1P2	GCCTGTCCCAGGAACGGCAT
Grin2b_-_136172145.23-P1P2	Grin2b	P1P2	GTGAGAAAAGCCAACCTGAA
Grin2b_-_136171934.23-P1P2	Grin2b	P1P2	GGATTCGAGTGTCTCCCGGA
Grin2b_-_136171999.23-P1P2	Grin2b	P1P2	GACCAAGTCGTTATAAGGAA
Grin2b_-_136172002. 23-P1P2	Grin2b	P1P2	GAAGTCGTTATAAGGAAAGG
Grin2b_+_136171844.23-P1P2	Grin2b	P1P2	GGAATGACCACGCTCCACGG
Grin2b_+_136172019.23-P1P2	Grin2b	P1P2	GCCTCTG GTGTGTACTCTGT

WHAT IS CLAIMED

1. A method of generating a protein comprising:

identifying one or more regions of a protein with affinity for a major histocompatibility complex (MHC), and

modifying the one or more regions of the protien with affinity for the MHC through one or more amino acid substitutions, such that the modified region has no affinity for the MHC,

wherein the resulting modified protein is immunosilent upon administration of the modified protein or a polynucleotide encoding the modified protein to a subject.

2. The method of claim 1, wherein the affinity for the MHC is high affinity.

3. The method of claims 1 or 2, wherein at least one substituted amino acid is an amino acid which does not serve as an MHC protein core residue.

4. The method of any one of claims 1 to 3, wherein the protein is selected from the group of a cytidine deaminase, an adenosine deaminase, a zinc finger nuclease, a transcriptional activator-like effector nuclease, a Cas9, or an AAV capsid protein.

5. The method of claim 4, wherein the protein is Cas9.

6. The method of claim 5, wherein the Cas9 is SpCas9.

7. A modified Cas9 protein produced according to the method of any one of claims 1 to 6.

8. A modified Cas9 protein comprising one or more, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, fifteen or more, or twenty or more of the amino acid modifications provided in **Table 1**.

9. An isolated polynucleotide encoding the modified Cas9 protein of claim 7 or 8.

10. A vector comprising the isolated polynucleotide of claim 9.

11. The vector of claim 10, wherein the vector is an AAV vector, optionally wherein the AAV vector is AAV5.

12. An AAV capsid comprising the vector of claim 11.

13. The AAV capsid of claim 12, wherein one or more of the AAV capsid proteins has been modified according to the method of any one of claims 1 to 4.

14. A method of avoiding an immune response in a subject being administered a regimen requiring Cas9, the method comprising: administering to the subject, in sequence, two or more Cas9 proteins that are immune orthogonal.

15. A method of gene editing or gene regulation in a subject, the method comprising: administering to the subject, in sequence, two or more Cas9 proteins that are immune orthogonal.

16. A method of treating a subject in need of gene editing or gene regulation, the method comprising: administering to the subject, in sequence, two or more Cas9 proteins that are immune orthogonal or polynucleotides encoding said Cas9 proteins.

17. The method of any one of claims 14 to 16, in which the Cas9 proteins that are immune orthogonal do not share an amino acid sequence of greater than 5 consecutive amino acids.

18. The method of any one of claims 14 to 17, in which the Cas9 proteins that are immune orthogonal do not share affinity for a major histocompatibility complex (MHC).

19. The method of any one of claims 14 to 18, in which three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more Cas9 proteins that are immune orthogonal are administered in sequence.

20. The method of any one of claims 14 to 19, in which each Cas9 protein that is immune orthogonal is a Cas9 derived from a distinct species of bacteria.

21. The method of claim 20, in which the Cas9 proteins that are immune orthogonal are selected from *S. pyogenes* Cas9 (spCas9), *S. aureus* Cas9 (saCas9), *B. longum* Cas9, *A. muiciniphilia* Cas9, or *O. laneus* Cas9.

22. The method of claim 21, in which the Cas9 proteins that are immune orthogonal comprise spCas9 and saCas9.

23. The method of any one of claims 14 to 22, in which at least one of the two or more Cas9 proteins is modified to reduce immunogenicity upon administration to the subject.

24. The method of claim 23, wherein the at least one of the two or more Cas9 proteins is modified according to the method of any one of claims 1 to 6.

25. The method of any one of claims 14 to 24, wherein at least one of the two or more Cas9 proteins or polynucleotides encoding said Cas9 proteins is comprised in an AAV vector.

26. The method of claim 25, wherein the AAV vector is an AAV5 vector.

27. The method of claim 25 or 26, wherein the AAV vector is comprised in an AAV capsid.

28. The method of any one of claims 25 to 27, wherein two or more Cas9 proteins or polynucleotides encoding said Cas9 proteins are comprised in AAV vectors.

29. The method of claim 28, wherein each AAV vector is comprised in an AAV capsid, optionally wherein the AAV capsids are immune orthogonal to one another.

30. The method of any one of claims 14 to 29, further comprising administering one or more guide RNAs to the subject.

31. The method of claim 30, wherein the guide RNA is selected to treat a disease, disorder, or condition selected from the group of achromatopsia, adenosine deaminase (ADA) deficiency, alpha-1-antitrypsin deficiency, Alzheimer's disease, amyotrophic lateral sclerosis, aromatic amino acid decarboxylase deficiency, Batten disease, choroideremia, Crigler Najjar syndrome, cystic fibrosis, fragile X syndrome, hemophilia, hepatitis B, hepatitis C,

homozygous familial hypercholesteremia, Huntington's Disease, Leber congenital amaurosis, macular degeneration, maple syrup urine disease (MSUD), mucopolysarccharidosis (I-LX), multiple sclerosis, muscular dystrophy, myotonic dystrophy, neurofibramotosis type 1, ornithine transcarbamylase deficiency, pachyonychia congenita, Parkinson's disease, phenylketonuria, polycystic kidney disease, Pompe disease, retinal degeneration, Rett's syndrome, rickets, spinal muscular atrophy, severe combined immunodeficiency, sickle cell disease, Smith-Lemli-Opitz syndrome, Y-linked nonobstructive spermatogenic failure, thalassemia, Tay-Sachs disease, Wilson's disease, cardiovascular disease, metabolic syndrome, pain management, and X-linked retinoschisis.

FIGURE 1

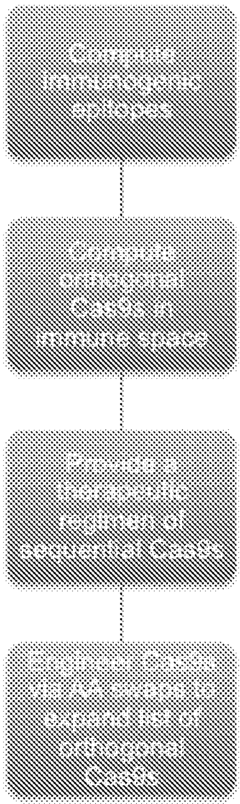
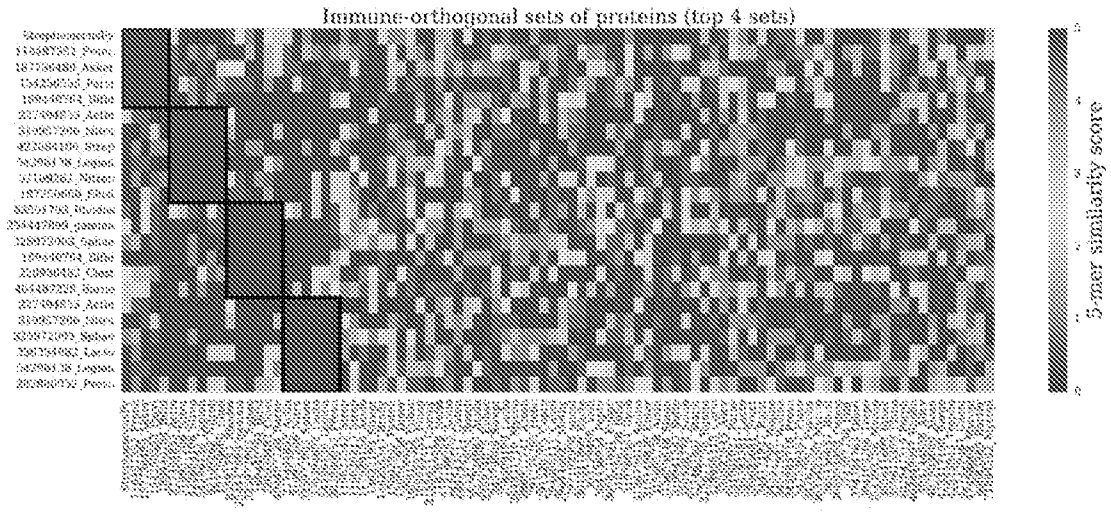

Compute immunogenic epitopes
Compute orthogonal Cas9s in immune space
Provide a therapeutic regimen of sequential Cas9s
Engineer Cas9s via AA swaps to expand list of orthogonal Cas9s

FIGURE 2

A


Immune-orthogonal sets of proteins (top 4 sets)
5-mer similarity score

B

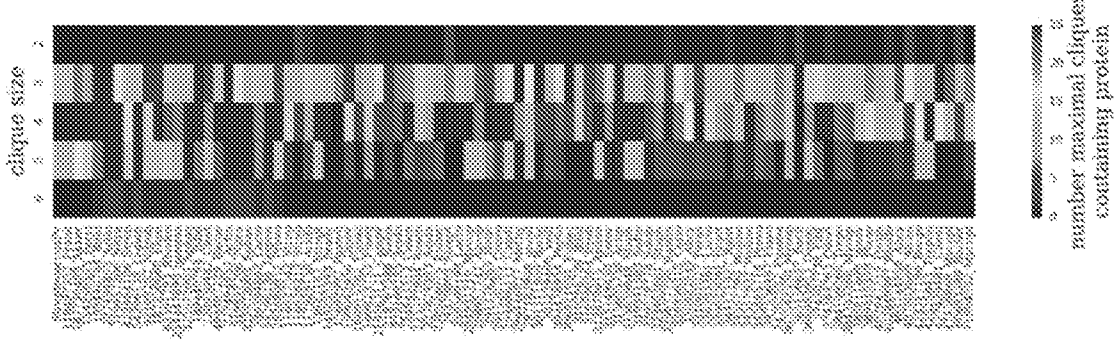
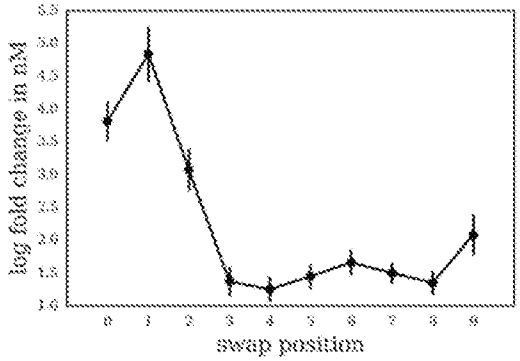
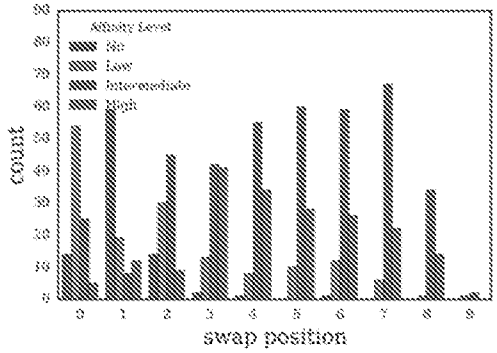

clique size
number maximal cliques containing protein

FIGURE 3

A


log fold change in nM
swap position

B


Affinity Level
no
low
intermediate
high
count
swap position

C

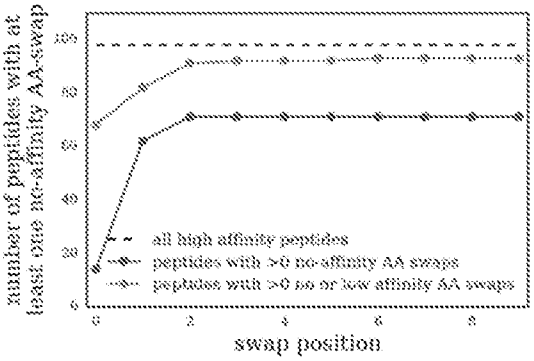

number of peptides with at least one no-affinity AA-swap
all high affinity peptides
peptides with >0 no-affinity AA swaps
peptides with >0 no or low affinity AA swaps
swap position

FIGURE 4

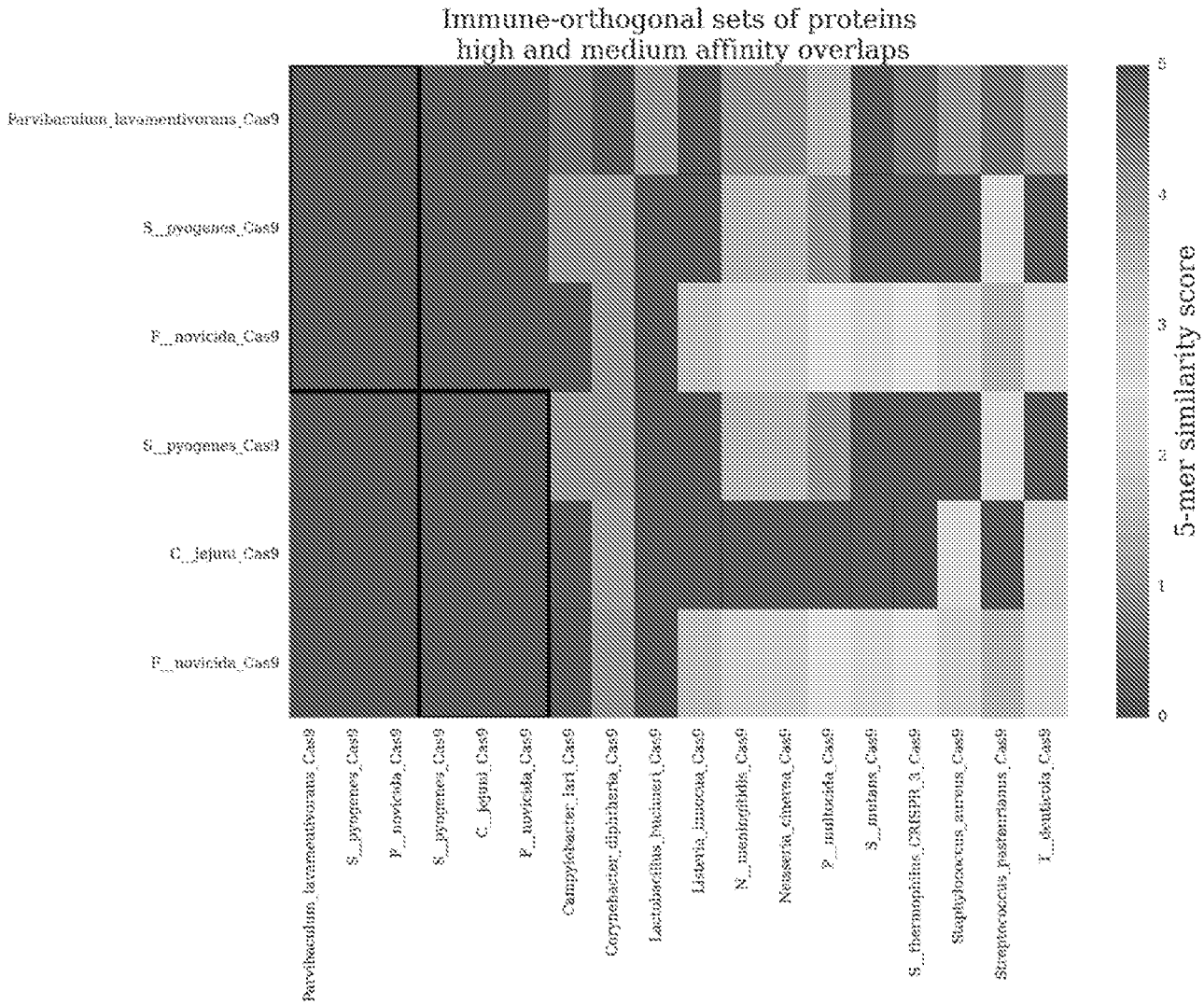

Immune-orthogonal sets of proteins
high and medium affinity overlaps
Parvibaculum_lavamentivorans_Cas9
S__pyogenes_Cas9
F__novicida_Cas9
S__pyogenes_Cas9
C__jejuni_Cas9
F__novicida_Cas9
Parvibaculum_lavamentivorans_Cas9
S__pyogenes_Cas9
F__novicida_Cas9
S__pyogenes_Cas9
C__jejuni_Cas9
F__novicida_Cas9
Campylobacter_lari_Cas9
Corynebacter_diphtheria_Cas9
Lactobacillus_paracasei_Cas9
Listeria_innocua_Cas9
N__meningitidis_Cas9
Neisseria_cinerea_Cas9
P__multocida_Cas9
S__mutans_Cas9
S__thermophilus_CRISPR_3_Cas9
Staphylococcus_aureus_Cas9
Streptococcus_pasteurianus_Cas9
T__denticola_Cas9
5-mer similarity score

FIGURE 5A

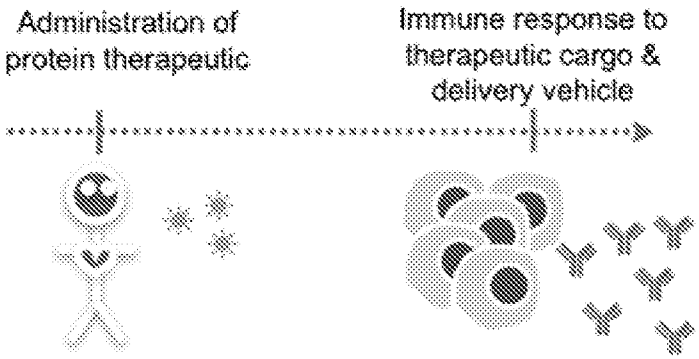


FIGURE 5B

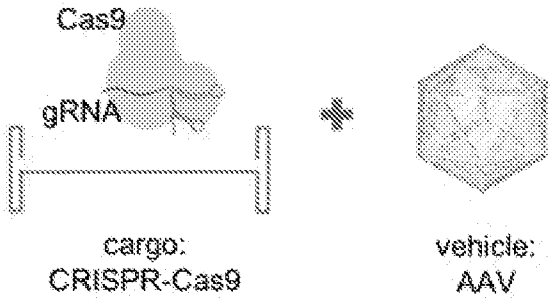


FIGURE 5C

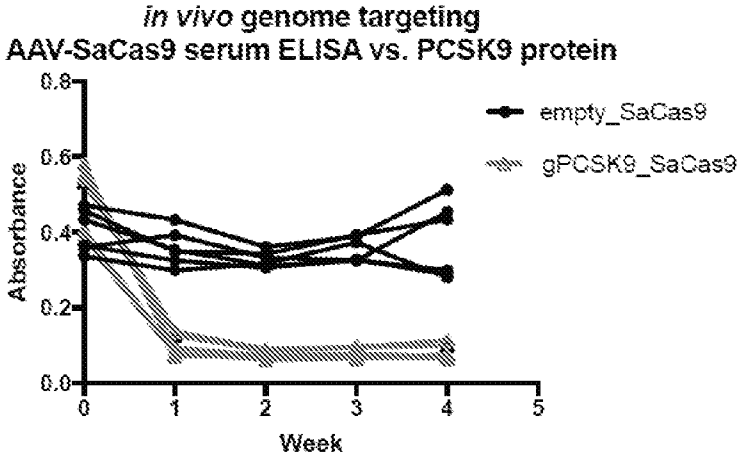


FIGURE 5D

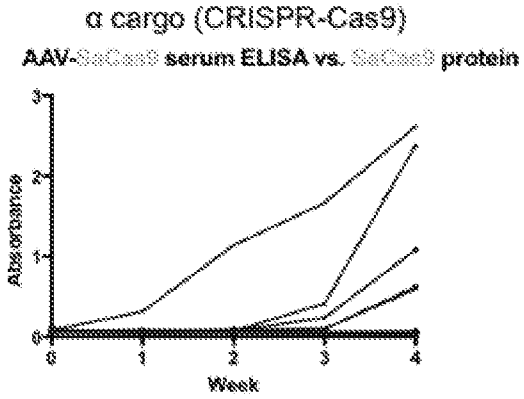


FIGURE 5E

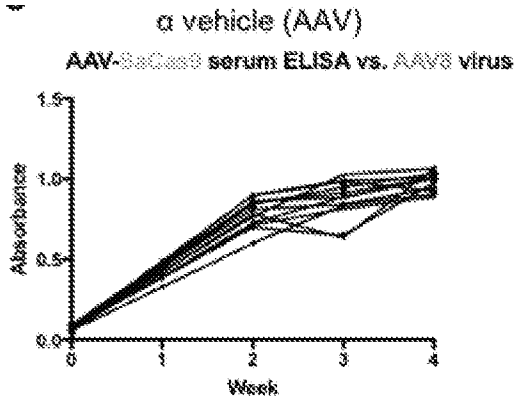


FIGURE 5F

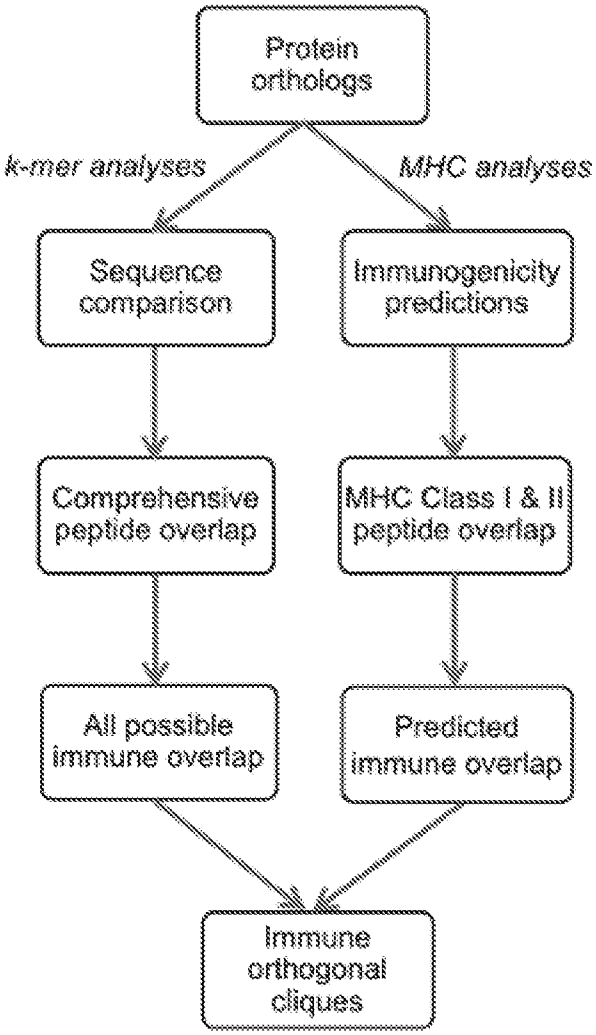

Protein orthologs
k-mer analyses
MHC analyses
Sequence comparison
Immunogenicity predictions
Comprehensive peptide overlap
MHC Class I & II peptide overlap
All possible immune overlap
Predicted immune overlap
Immune orthogonal cliques

FIGURE 5G

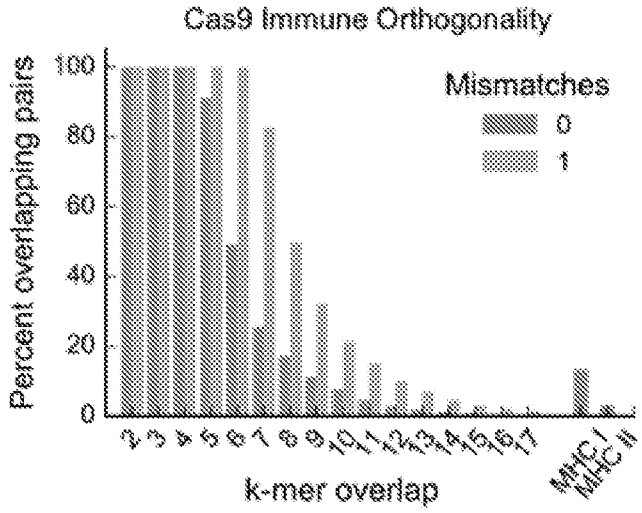


FIGURE 5H

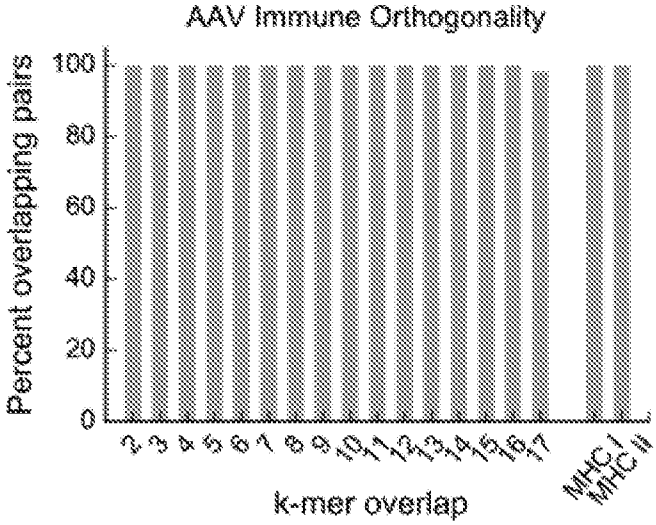


FIGURE 6A

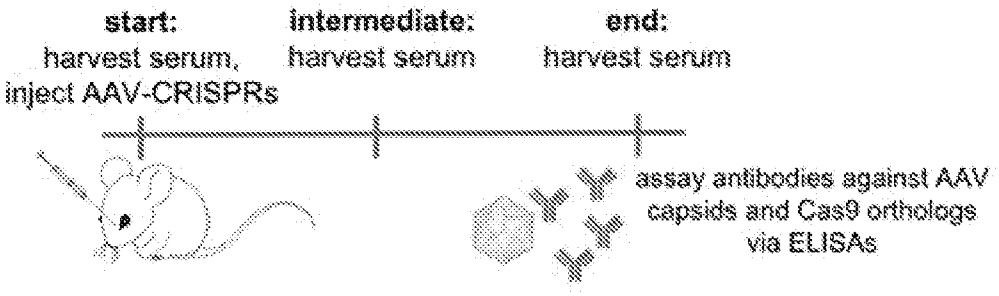


FIGURE 6B

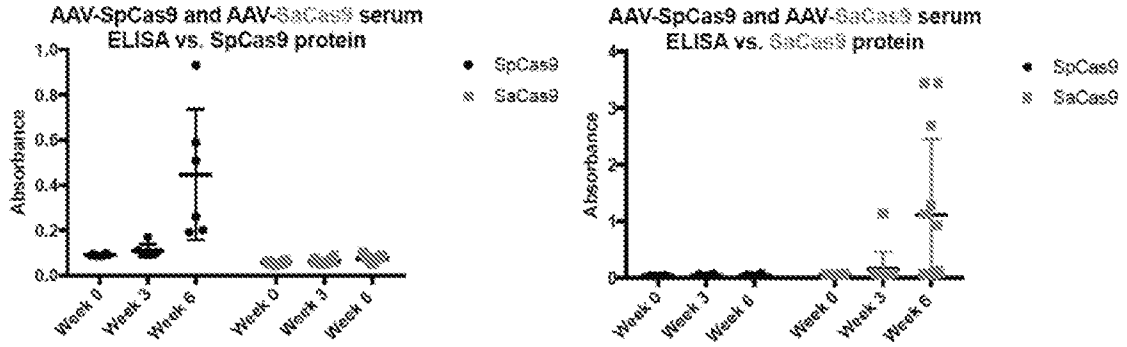


FIGURE 6C

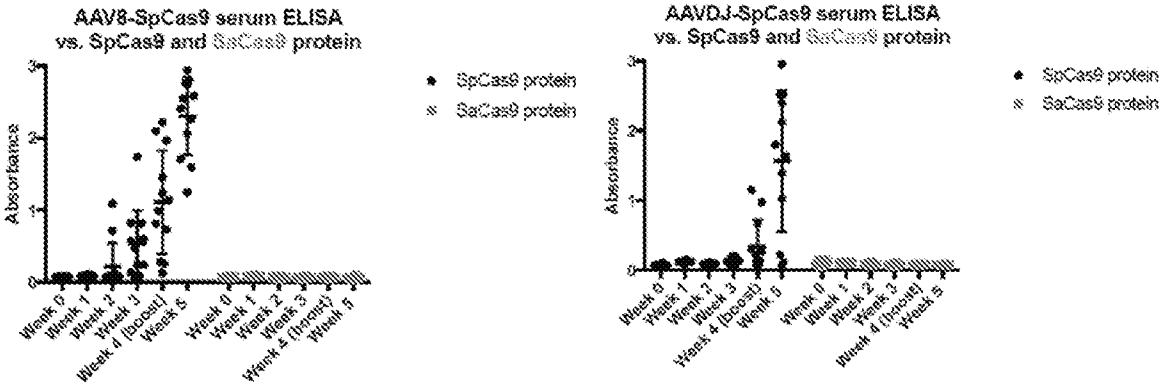


FIGURE 6D

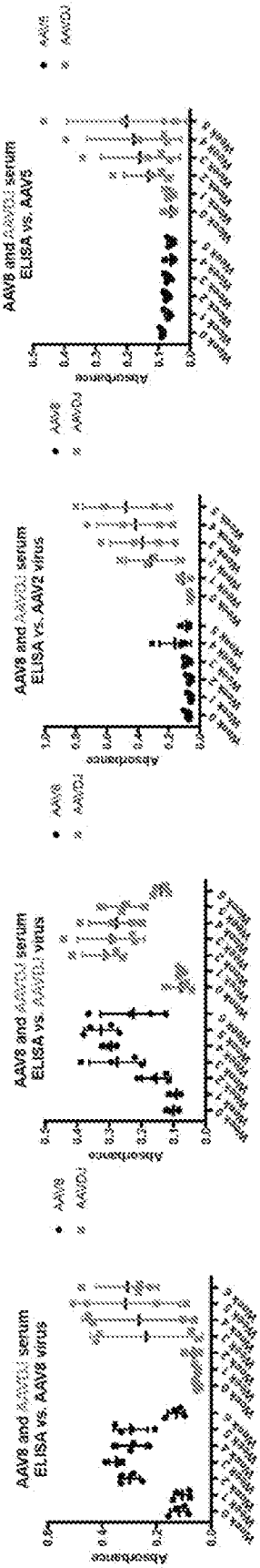


FIGURE 6E

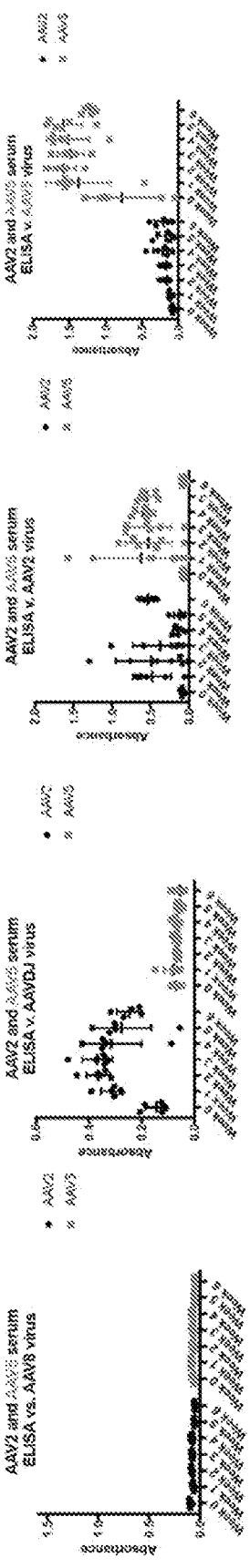


FIGURE 7

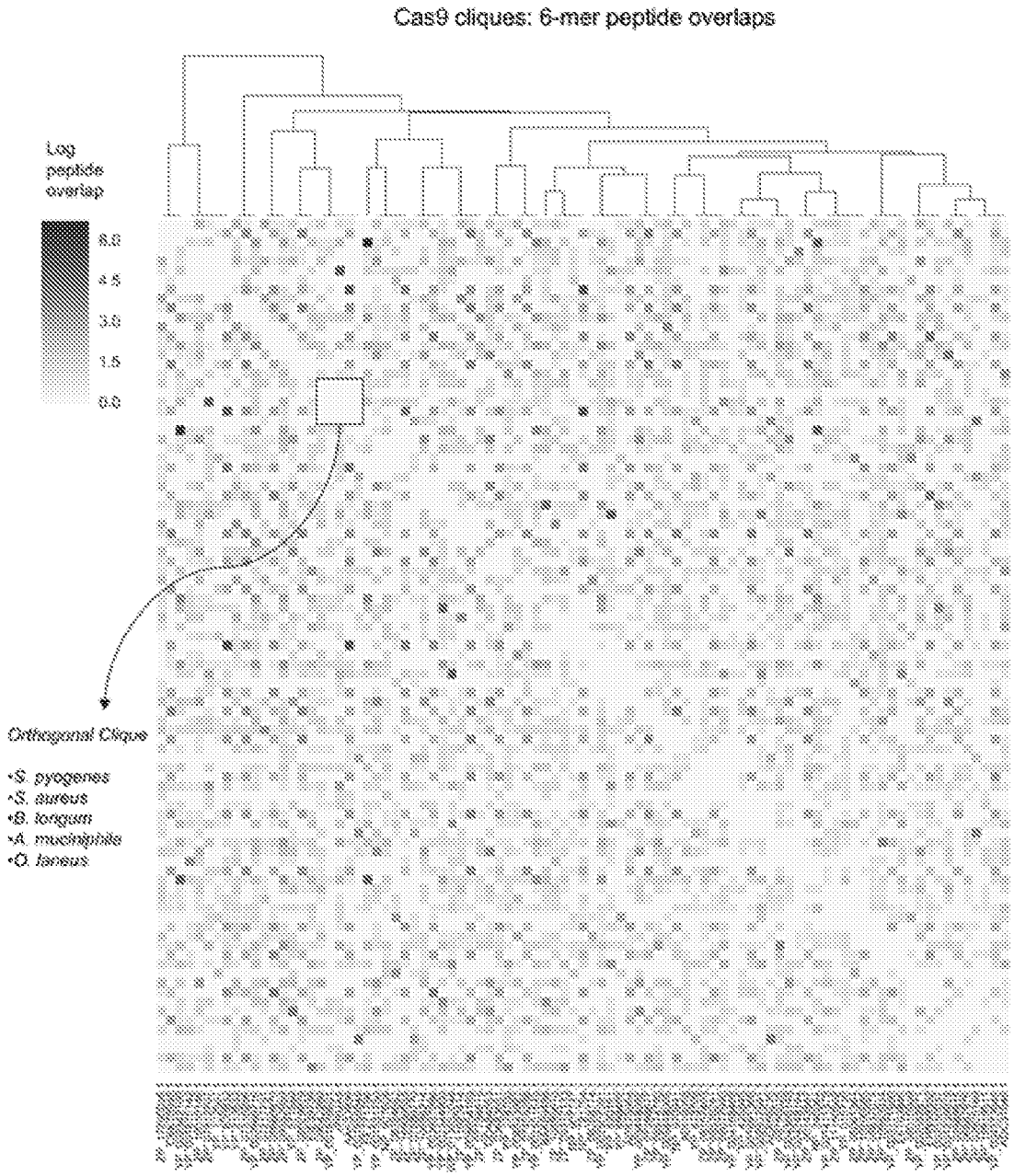

Cas9 cliques: 6-mer peptide overlaps
Log peptide overlap
6.0
4.5
3.0
1.5
0.0
Orthogonal Clique
•S. pyogenes
•S. aureus
•B. longum
•A. muciniphila
•O. laneus

FIGURE 8A

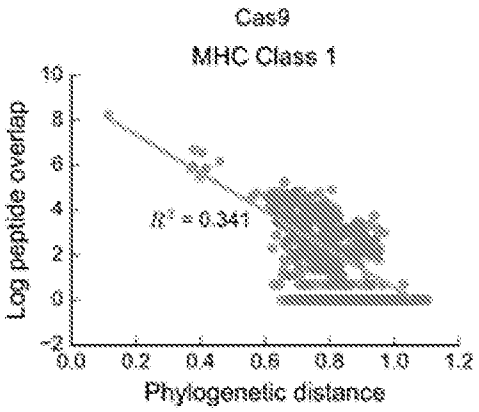


FIGURE 8B

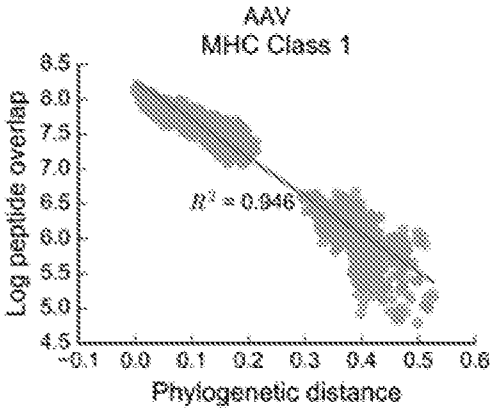


FIGURE 8C

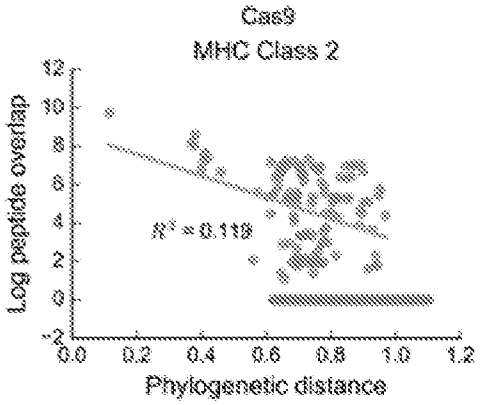


FIGURE 8D

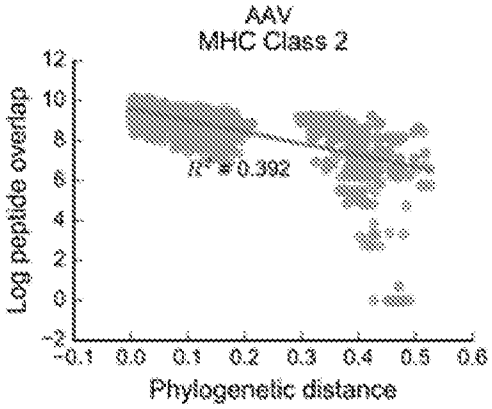


FIGURE 9A

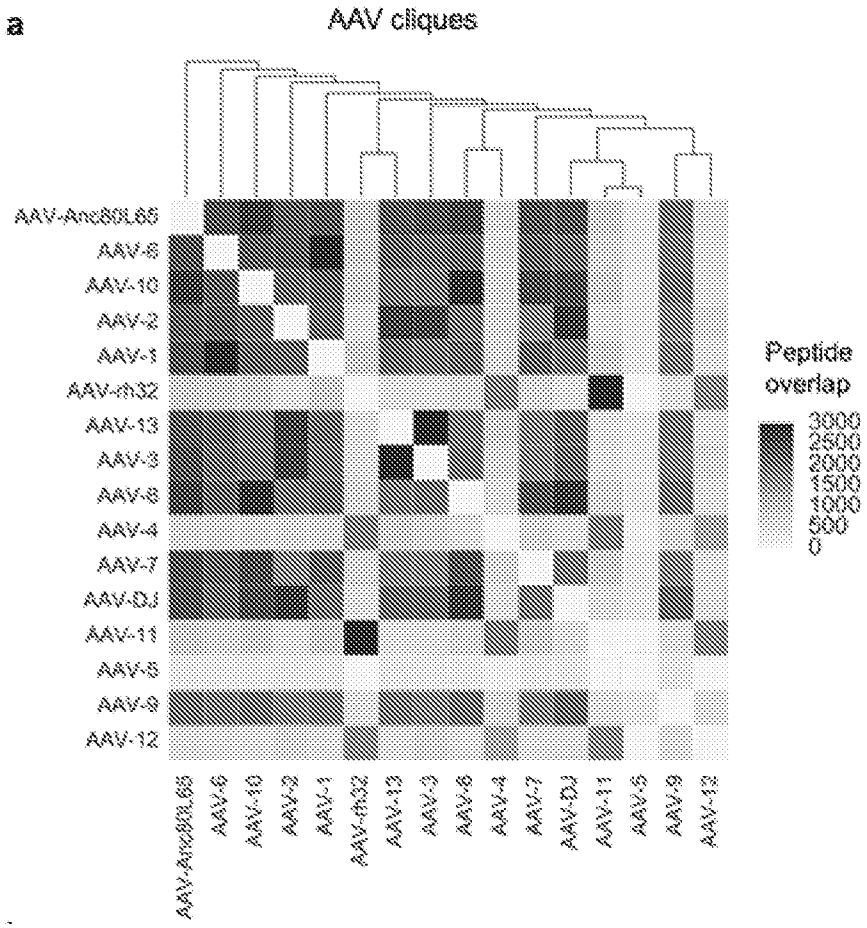

a
AAV cliques
AAV-Anc80L65
AAV-6
AAV-10
AAV-2
AAV-1
AAV-rh32
AAV-13
AAV-3
AAV-8
AAV-4
AAV-7
AAV-DJ
AAV-11
AAV-5
AAV-9
AAV-12
Peptide overlap
3000
2500
2000
1500
1000
500
0

FIGURE 9B

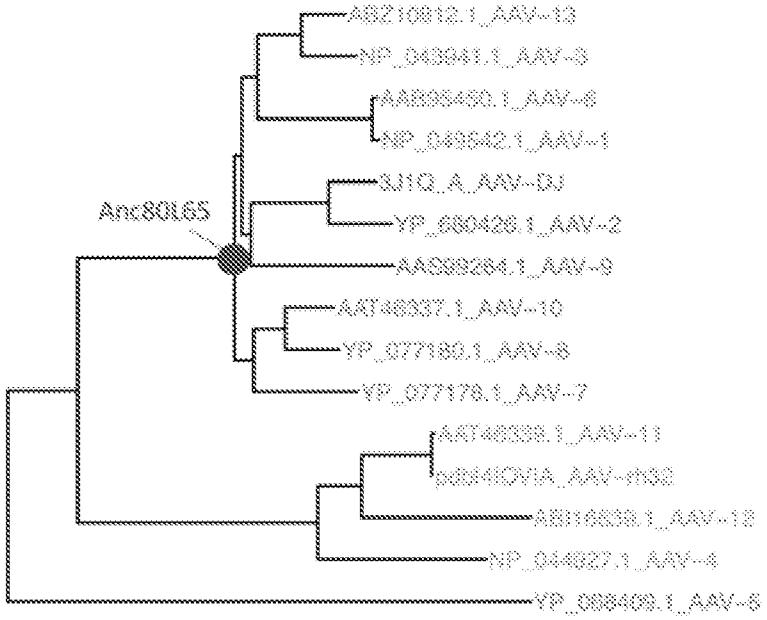

ABZ10812.1_AAV-13
NP_043941.1_AAV-3
AAB95450.1_AAV-6
NP_049542.1_AAV-1
SJ1Q_A_AAV-DJ
YP_680426.1_AAV-2
AAQ89284.1_AAV-9
AAT46337.1_AAV-10
YP_077180.1_AAV-8
YP_077178.1_AAV-7
AAT46338.1_AAV-11
5e83HOV1A_AAV-rh32
ABI16639.1_AAV-12
NP_044927.1_AAV-4
YP_068409.1_AAV-5
Anc80L65

FIGURE 10


Capsid : AAV8
Cargo : SaCas9
Harvest: 4 weeks

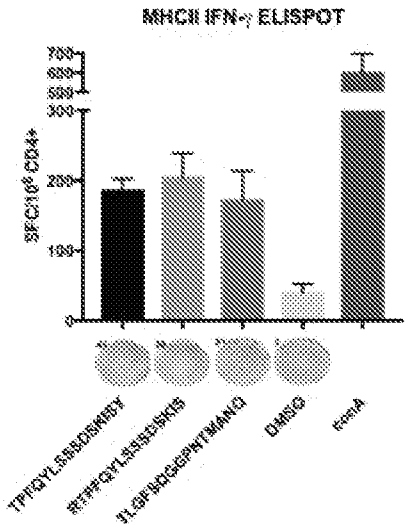
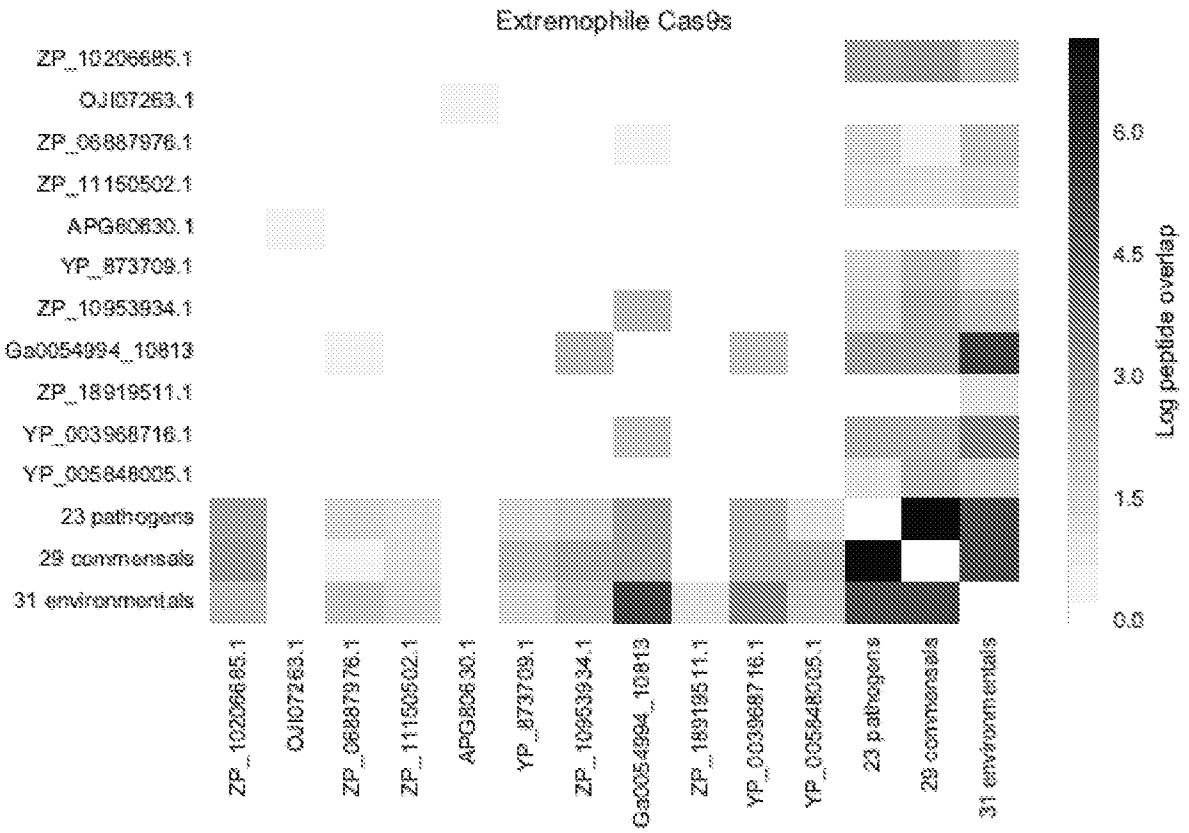

MHCII IFN-γ ELISPOT
SFC/10^6 CD4+
700
600
500
300
200
100
0
TPFQYLSSSDSKISY
RTPFQYLSSSDSKIS
TLGPSQSGQPQTMAAQ
DMSO
ConA

FIGURE 11


Extremophile Cas9s
ZP_10206685.1
OJI07263.1
ZP_06887976.1
ZP_11150502.1
APG80630.1
YP_873709.1
ZP_10953934.1
Ga0054994_10813
ZP_18919511.1
YP_003968716.1
YP_005848035.1
23 pathogens
29 commensals
31 environmentals
Log peptide overlap
6.0
4.5
3.0
1.5
0.0

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20 18/022258

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A61K 38/00; A61K 39/395; A61P 35/00; C07K 16/28; C12N 5/10; C12N 15/09 (2018.01)

CPC - A61K 38/00; A61K 39/395; A61K 2039/505; C07K 2317/24; C07K 2319/00; C07K 2319/30 (2018.05)

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)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 424/133.1; 424/134.1; 424/136.1; 435/462; 530/350; 530/402 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/0185038 A1 (CARR et al) 23 September 2004 (23.09.2004) entire document	1-3
X	US 7,615,217 B2 (GILLIES et al) 10 November 2009 (10.11.2009) entire document	1, 2
X	WO 2015/153789 A1 (EDITAS MEDICINE, INC.) 08 October 2015 (08.10.2015) entire document	14-17
P, X	MORENO et al. "Exploring protein orthogonality in immune space: a case study with AAV and Cas9," bioRxiv, 10 January 2018 (10.01.2018), Pgs. 1-24. entire document	1-3, 14-17
A	DEGROOT et al. "Prediction of immunogenicity for therapeutic proteins: State of the art," Current Oppinion in Drug Discovery & Development, 31 December 2007 (31.12.2007), Vol. 10, Iss. 3, Pgs. 1-9. entire document.	1-3, 14-17
A	MOISE et al. "Effect of HLA DR epitope de-immunization of Factor VIII in vitro and in vivo," Clinical Immunology, 31 March 2012 (31.03.2012), Vol 142, Iss. 3, Pgs. 320-31. entire document	1-3, 14-17
A	SANTANGELO et al. "Recognition of core and flanking amino acids of MHC class II-bound peptides by the T cell receptor," European Journal of Immunology, 22 August 2002 (22.08.2002), Vol. 32, Iss. 9, Pgs. 2510-2520. entire document	1-3, 14-17

☐ Further documents are listed in the continuation of Box C. ☐ 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
23 May 2018	25 JUN 2018

Name and mailing address of the ISA/US	Authorized officer
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/022258

Box No. 1 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. ☐ forming part of the international application as filed:
      ☒ in the form of an Annex C/ST.25 text file.
      ☐ on paper or in the form of an image file.

   b. ☒ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☒ furnished subsequent to the international filing date for the purposes of international search only:
      ☒ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
      ☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

ISA/225 mailed on 28 March 2018. No approved electronic sequence listing was submitted in response to the ISA/225.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US201 8/022258

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ¾ Claims Nos.: 8
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

Claim 8 is held unsearchable as a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit, furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

3. ☒ Claims Nos.: 4-7, 9-13, 18-31
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.