Further provided are Cas9 proteins modified to reduce immunogenicity.

(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 CRISPR 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.
Published:
— with international search report (Art. 21(3))
ENGINEERING CRISPR CAS9 IMMUNE STEALTH

CROSS-REFERENCE TO RELATED APPLICATIONS


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 CRISPR/Cas9 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 CRISPR 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 protein 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 affinities. 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), mucopolysaccharidosis (I-IX), multiple sclerosis, muscular dystrophy, myotonic dystrophy, neurofibromatosis 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-orbitally 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 k-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.
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).

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.

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.

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.

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

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

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.

Definitions

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.

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.

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.

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

MAADGYLPDWDLEDNLSSEGIREWWDLKPGAPKPKANQQKQDDDRGLVLPGYKYLG
PFNGLDKGEPVNAADAAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQRLQEDTS
FGGNLGRAVFQAKKRVLDEPGLVEEGAKTAPGKRPLESPQEDPSGGIGKGGQPA
RKRLNFEEDTGAAGDPGPESDTSAAMSSDIEMRRAAPGGNAVDAGQGSDGVSNGASGD
WHCDSTWSEGKVTITSTRTWVLYTPNNHLRLGTSSSSNTNGFSTPWGYFDNFR
FHCHFSRPDQRLNWWGLRPKAlvRVIKIFNQVEVTSSQGETTVANNLTVSTVQIF
ADSSYELPYVMDAGQGSLPPFNDPVMPQYYGICIVTGQENQNTDRNAYCLEY
FPSQMLRTGNFEAYNFKEVPFHSMYAHSQSLDPXIVn[PLLQYIWHQLSTTSGET
LQGNAATTFFGRKIRGDFAFYRKNWLPGCVKQQRFSKTAQSNYKIPASGGNAL
YDTHYTLNRWSIAPGPMMATAGPSDGSNAQLIFPGPSVTGNTTSANNLLEFTSE
EEIAATPRDTRMFGQIADNNQNATTAPITGNVTAMGVLMGVQWQRDIYYQGPIW
AKIPHADGHPSPPLGGFLKHPFQIFIKNTPVPAANPATTFTAARVDSTFIQYSTGQ
VAVQIEWEIEKERSKRNPEVQFSTNYGNQSSMLWAPDTGKYTEPRVIGSRYLTNHL

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

MAADGYLPDWDLEDNLSSEGIREWWDLKPGAPKPKANQQKQDDDRGLVLPGYKYLG
PFNGLDKGEPVNAADAAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQRLQEDTS
FGGNLGRAVFQAKKRVLDEPGLVEEGAKTAPGKRPLESPQEDPSGGIGKGGQPA
RKRLNFEEDTGAAGDPGPESDTSAAMSSDIEMRRAAPGGNAVDAGQGSDGVSNGASGD
WHCDSTWSEGKVTITSTRTWVLYTPNNHLRLGTSSSSNTNGFSTPWGYFDNFR
FHCHFSRPDQRLNWWGLRPKAlvRVIKIFNQVEVTSSQGETTVANNLTVSTVQIF
ADSSYELPYVMDAGQGSLPPFNDPVMPQYYGICIVTGQENQNTDRNAYCLEY
FPSQMLRTGNFEAYNFKEVPFHSMYAHSQSLDPXIVn[PLLQYIWHQLSTTSGET
LQGNAATTFFGRKIRGDFAFYRKNWLPGCVKQQRFSKTAQSNYKIPASGGNAL
YDTHYTLNRWSIAPGPMMATAGPSDGSNAQLIFPGPSVTGNTTSANNLLEFTSE
EEIAATPRDTRMFGQIADNNQNATTAPITGNVTAMGVLMGVQWQRDIYYQGPIW

-9-
AKIPHADGHFHPSPLIGGFGLKHPQQFIFKNTPVTPNAPTFTFARVDSFITQYSTGQ
VAVQIEWEIEKERSKR\WEVEQFTSNYGNQSSMLWAPDTGKYTEPRVIGSRYLTN
HL

[0035] ABI16639.1 AAV-12

MAADGYLPDWLDELNLSEGIREWWALKPGAPQPKANQHQDNQRNGLVLPGYKYLGP
PFNLDKGEPVNEAADAALLEHDKAYDKQLEQGDNPYLKYNHADAEFQQRALTATDS
FGNLGGRVFQAKKRLEPGVLEEGVKTAPGKKRPLEKTPRTPNPDGKAPAXXK
QKDGEPAADSARRLDFEDSGAGDPEGSSEGMSHDAEMRAPPGGNAVEAGQQA
DGVGNASGDWHCDSTWSEGRVVTTSTRTWLPYNNHLRLIGTANSNTYNGFST
PWYGDFNPP\HCIFISPRDWQRLLNNWGLRPSMRKIFNIVQVKEVTSSNGETTV
NNLSTTVQIFADSTYELPYVMADAGQEGSFPPFDVFMPVQYGCGVVTGKNQNQT
DRNAFYCLEYFPSQMLRGT\FEVSYQFKVPKHFMYAHSQSLDRMNPNLLDYQL
WHLQSTTTGNSLQGTATTTTGYKITTGFAYYORKNWLPGCIQKQQKSNANQNY
KIPASGGDALLKYDTHTLNGRSWNSMAPPGPMATAGAGDSDFSNSQLIFAGPNPG
NTTSSNT\LFTSEEIATTSTPP\TDMEFGQIADNNQNAATTAPFIANLDMGIVPGMV
WQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFLKHPQFIFKNTPVAPNNTTFSA
ARINSLQTQYSTGQAVQIDWEIQKEHSKJIIWNPEVQFTSNYGTQNSMLWAPDANGN
YHELRAIGSERFLTHHL

[0036] NP_044927.1 AAV-4

MTDGYLPDWLDELNLSEGIREWWALQPGAPPKANQHQDNARGLVLPGYKYLGP
GNGLDKGEPVNEAADAALLEHDKAYDKQLEQGDNPYLKYNHADAEFQQRALTATDS
FGNLGGRVFQAKKRLEPGVLEEGVKTAPGKKRPLEKTPRTPNPDGKAPAXXK
QKDGEPAADSARRLDFEDSGAGDPEGSSEGMSHDAEMRAPPGGNAVEAGQQA
DGVGNASGDWHCDSTWSEGRVVTTSTRTWLPYNNHLRLIGTANSNTYNGFST
PWYGDFNPP\HCIFISPRDWQRLLNNWGLRPSMRKIFNIVQVKEVTSSNGETTV
NNLSTTVQIFADSTYELPYVMADAGQEGSFPPFDVFMPVQYGCGVVTGKNQNQT
DRNAFYCLEYFPSQMLRGT\FEVSYQFKVPKHFMYAHSQSLDRMNPNLLDYQL
WHLQSTTTGNSLQGTATTTTGYKITTGFAYYORKNWLPGCIQKQQKSNANQNY
KIPASGGDALLKYDTHTLNGRSWNSMAPPGPMATAGAGDSDFSNSQLIFAGPNPG
NTTSSNT\LFTSEEIATTSTPP\TDMEFGQIADNNQNAATTAPFIANLDMGIVPGMV
WQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFLKHPQFIFKNTPVAPNNTTFSA
ARINSLQTQYSTGQAVQIDWEIQKEHSKJIIWNPEVQFTSNYGTQNSMLWAPDANGN
YHELRAIGSERFLTHHL
[0037] **YP_077178.1** AAV-7

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQQKQDDDNGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAADFQERLQEDTS
FGGNLGRAVFQAKKRVEPLGLVEEAGTAPAKKRVPVEPSPQSPDSSTGIGKKGQQ
PARKRLNGFTQTDSESVDPQPLGEPAPAAPSSVGSTVAAGGAPMADNNEGADV
GNASGNWHCSTWLGDVFIVTSTTRWALTYPNNHLYQISSSETAGSTNDNTYFGYS
TPWGYDFRNFHCHPSRPDWQLRINNNWGFRPKRLFKLFNILQVFKEVTINDGVTIA
NNLSTIQVFSDEYQPLYVLGAHQGCLPPPFPADVFMIPQYGYLTLLNGSQQGRRS
FYCLEYFPSQMLRTGNNFESEFSEDVPHSSYHESQSLRMLNPLIDQLYYLART
QSNPGGTAGNRELQFYQGGPSTMAEQAQNWLPGCFCRQVRSVKTLDQNVNTNSNFLA
TGATKYHLNMGSLWPGVAMATHKIDEDRFFSGVILIGKTVATNKTLENVLM
T^EEIRPTNPVATEEYGVSSQAANTAAQTQVNNQGALPGMVWQRNDVYLQ
GPIWAKPHTDGHNPSLMMGFLKHPPQLILEKTNPVAPANPVEFTPAKFAFSFITQYS
TGQVSVIEWELQKENSKRWNPEIQYTSNEFKQTGDFADSQGVYSEPRPIGTRYL
TRNL

[0038] **YP_077180.1** AAV-8

MAADGYLPLDPWLEDNLSEGIREWWALKPGAPKPKANQQQKQDDNGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAALEHDKAYDQQLQAGDNPYLRYNHADAADFQERLQEDTS
FGGNLGRAVFQAKKRVEPLGLVEEAGTAPAKKRVPVEPSPQSPDSSTGIGKKGQQ
PARKRLNGFTQTDSESVDPQPLGEPAPAAPSSVGSTVAAGGAPMADNNEGADV
GNASGNWHCSTWLGDVFIVTSTTRWALTYPNNHLYQISSSETAGSTNDNTYFGYS
TPWGYDFRNFHCHPSRPDWQLRINNNWGFRPKRLFKLFNILQVFKEVTINDGVTIA
NNLSTIQVFSDEYQPLYVLGAHQGCLPPPFPADVFMIPQYGYLTLLNGSQQGRRS
FYCLEYFPSQMLRTGNNFESEFSEDVPHSSYHESQSLRMLNPLIDQLYYLART
QSNPGGTAGNRELQFYQGGPSTMAEQAQNWLPGCFCRQVRSVKTLDQNVNTNSNFLA
TGATKYHLNMGSLWPGVAMATHKIDEDRFFSGVILIGKTVATNKTLENVLM
T^EEIRPTNPVATEEYGVSSQAANTAAQTQVNNQGALPGMVWQRNDVYLQ
GPIWAKPHTDGHNPSLMMGFLKHPPQLILEKTNPVAPANPVEFTPAKFAFSFITQYS
TGQVSVIEWELQKENSKRWNPEIQYTSNEFKQTGDFADSQGVYSEPRPIGTRYL
TRNL
TAGTKYLNGRNSLANPGIAMATHKDEERFFPSNGILIFGKQNAARDNADYSDVM
LTSEEIEKTTPVATEEYGYIVAL^QQQNTAPQIGTWGALPGIVLVWQNRDVLQ
GPIWAKIPHTDGNFHPSLMGGFLKHHPPPQLIKNTVPADPPTTFQNSKLNSFITQY
STQVSVVEJEWELQKENSKRWNPE1QYTSNYYKSTSVDFAVNTEGVYSEPRPIGTRYL
TRNL

[0039] AAT46337.1 AAV-10
MAADGYLPDWEQEEWLDKPGAPPKPANQQQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAYLEHDKAYDQQLKAGDNPLYRYNHADEFQERLQEDTS
FGGLNLGRAVQAKKRLQPLEQLVGLVEEAAKATAPGKRPVEPSQRSPDSTGIGKKGQQ
PAKKRNLNGQQTGESESVDPQPQEPGAPPAGSGLGSOTMAAGGAPMADNNEGADGV
GSSGNNWHCDSTWLVGDRVITTTSTRTWALPYNHLYKQINSGTGSGSTDNTFYFGY
STPWGYFDFNPJHCHFSRDWQRLNNNWGRFRPKRLSFKLHNQVFKEVTQNEGTKI
ANNLSTIQQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPOYGYLTNLNGSQAQVGR
SSFYCYLPFSQMRLTGNFSEYTFDFPEDVHSSYAHQSLDRLMNPLIDQYLYLRSR
TQSTGQTPGTFQLFQAGPAANMSAQAKNWLPGLCPYQRQVRSTTLQSNNSNFAW
TGATKYNHLNRDSVWGVAMATHKDEERFFPSGVLMGFKQGAGRDNVDYSSV
MLTSEEIEKTTPVATEQYGVVAD^QQANTGPIVGNWSQGALPGMYWQQRDVY
LQGPIWAKIPHTDGNFHPSLMGGFLKHHPPPQILIKNTVPADPPTTFQAKLASFIT
QYSTQVSVVEJEWELQKENSKRWWEIQTNSNYYKSTVDFAVNTEGVYSEPRPIGTRYL
TRNL

[0040] AAS99264.1 AAV-9
MAADGYLPDWEQEEWLDKPGAPPKPANQQQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAYLEHDKAYDQQLKAGDNPLYRYNHADEFQERLQEDTS
FGGLNLGRAVQAKKRLQPLEQLVGLVEEAAKATAPGKRPVEPSQRSPDSTGIGKKGQQ
PAKKRNLNGQQTGESESVDPQPQEPGAPPAGSGLGSOTMAAGGAPMADNNEGADGV
GSSGNNWHCDSTWLVGDRVITTTSTRTWALPYNHLYKQINSGTGSGSTDNTFYFGY
STPWGYFDFNPJHCHFSRDWQRLNNNWGRFRPKRLSFKLHNQVFKEVTQNEGTKI
ANNLSTIQQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPOYGYLTNLNGSQAQVGR
SSFYCYLPFSQMRLTGNFSEYTFDFPEDVHSSYAHQSLDRLMNPLIDQYLYLRSR
TQSTGQTPGTFQLFQAGPAANMSAQAKNWLPGLCPYQRQVRSTTLQSNNSNFAW
TGATKYNHLNRDSVWGVAMATHKDEERFFPSGVLMGFKQGAGRDNVDYSSV
MLTSEEIEKTTPVATEQYGVVAD^QQANTGPIVGNWSQGALPGMYWQQRDVY
LQGPIWAKIPHTDGNFHPSLMGGFLKHHPPPQILIKNTVPADPPTTFQAKLASFIT
QYSTQVSVVEJEWELQKENSKRWWEIQTNSNYYKSTVDFAVNTEGVYSEPRPIGTRYL
TRNL
KTINGSGQNQQQTFLKSVAGPSNMAVQGRNYIPGQSYRQQVRSTTVTQNNNSEFAWP
GASSWALNGRNSLMPGAPAMASHKEGEDRFPPLLGSGLFGKQGTGRDNDADKVMIT
TNEEEIHTTPVATESYQVATNHQSAQAQAQQTGWWQNHQGISLPGMVWQDRDYLQ
GPIWAKIPHTDGOTHPSLMMGFGMKHPQPQLIKNTVPADPPTAFNKDKNNSFITQ
YSTGQVSYEIEWELQKENSKRWNPEIQYTSNYYKSNNEFAVNTEGVYSEPPIPGR
YLTRNL

[0041] NP_049542.1 AAV-1

MAADGYPDWNLEDNLSSEGIREWWDLKPAGPKPANQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAAAEHDKAYDQQLKAGDNPYLRYNHADAEEQERLQEDTS
FGGNGLGRAVFQAKKRVLEPLGLVEEGAKTAPGKRKPVEQPSQPEDSSSGIKGTKQQP
AKKRLNGQFTGQDSEVDPQPLGEPPATPAAVGPTTMSAGGGAPMADNNEAGDV
GANSG>WHCDSTWLGDRTTSTRTWLPTYNNHLYQISSASTGASDNHYFGYS
TPWGYDFRNFHCHFSRDPWQRLLLNNWGFRPKLNKLFNIQKEVTTNDVTTIA
NLLSTVQVFSDSEYQLPYVLSAHHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRS
SFYCLEYFPSQMLRTGNNTFFSYTFEEVFHSSYAHSSQSLRMLNPLIQYLYYNRT
QNSGSAQNKDULLFSRPGMSVQPKNWLPGLCPRQRQVSKTLDNNNSFTWT
GASKYNLNGRESIIINPGTAMASHKEDDKFPMGVMIFGKESAGSNTALDNMIT
DEEIKATNPVATERFQTAVNVQSSSTDPATGDVHAMGALPGMVWQDRDVYLQG
PIWAKIPHTDGHFPSLGGFGKLNPQPLIKNTVPANPAAFSATKFASTQYQST
GQVSVEIEWELQKENSKRWNPEIQYTSNYAASNVDFTVDDNGLYTEPRPIGRYL
TRPL

[0042] AAB95450.1 AAV-6

MAADGYPDWNLEDNLSSEGIREWWDLKPAGPKPANQKQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAAAEHDKAYDQQLKAGDNPYLRYNHADAEEQERLQEDTS
FGGNGLGRAVFQAKKRVLEPLGLVEEGAKTAPGKRKPVEQPSQPEDSSSGIKGTKQQP
AKKRLNGQFTGQDSEVDPQPLGEPPATPAAVGPTTMSAGGGAPMADNNEAGDV
GANSG>WHCDSTWLGDRTTSTRTWLPTYNNHLYQISSASTGASDNHYFGYS
TPWGYDFRNFHCHFSRDPWQRLLLNNWGFRPKLNKLFNIQKEVTTNDVTTIA
NLLSTVQVFSDSEYQLPYVLSAHHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRS
SFYCLEWPSQMRTGKFTFSYTFEDVPFHSSYAHQSLSDPvLMNPLIDQYLYYLNLRTQNMNGSAQKDLFFSRGPLGMSVQPKNLPGPCYQRQVRVSKTKTDDNNSFTWTGASKYNLNGRESINPGTAMASHKDDDKFFPMSGVMIFGKESAGASNTALDVMI
TDEEEIKATNPVATERFGTVAVNLQSSSTDPATGDHVHVGALPGMVWQDRDVLQ
GPIWAKIPHTDGFHSPPLMGGFGLKHPPQPILIKNTPVAPANPFAFSATKFAFISITQYS
TGQSVSEIEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRY
LTRPL

[0043] NP_043941.1 AAV-3

MAADGYLPDWWLEDNLSEGIREWWALKPGVQPQPANQHQHDNRRGLVLPGYKYLGPNGNLKGEPVNEADAALAEHKAYDQQLKAGDNLPYKYNHADAEFQERLQEDTSGGNGLGRAVFQAKKRILEPLGLEIEAKATPGKGAVDQSPQEDSSVGGKQKPARKRNFGQTGDESEPVDPQPLGEPAAPTSLSNTMASGGGAPMADNNEGADGVNNSGNNWHDSCQWLGDRTITTSTRTWALPTYNNLYKQISSQSGASNDNYFGYST
PWGWDFNRFHCFSRHPDWRQLNPWJWGFDPKMLSFKLFMQVGRGTQNDTITIAN
NLSTVQVFTDSEYQLPVLSAHQGCLLPPFADVPMPQYGYTYLNLNSGQAVGRSFYCYLEWPSQMRLTGNNFQYTFEDVPFHSSYAHQSIIDRLMNPLIDQYLYLNRTQGTTSGTTQSRLLFSQAGPQSMNLQARNWLPGPCYQRQLSKTA>TONNNNSFPW
TAAKYHLNDRDLSVNPAMASHKDEEKFPMHGNLJFEGTASNAELDNVMITDDEEIRRTOPVATEQGYTVNLNQSSNTAPTTGTVNHQALPGMVWQDRDLYLQGPIWAKIPHTDGFHSPPLMGGFGLKHPPQPILIKNTPVAPANPFAFSATKFAFISITQYSTGQVSEIEWELQKENSKRWNPEIYTSNYAKSNVDFTVDNGVYSEPRPIGTRY
YLTRNL

[0044] ABZ10812.1 AAV-13

MTDGYLPDWDNLSEGIREWALKPGVQPQPANQHQDNARGLVLPGYKYLGPNGNLKGEPVNAADAAALEHKAYDQQLKAGDNLPYKYNHADAEFQERLQEDTSGGNGLGRAVFQAKKRILEPLGLEIEAKATPGKGAVDQSPQEDSSVGGKQKPARKRNFGQTGDESEPVDPQPLGEPAAPTSLSNTMASGGGAPMADNNEGADGVNNSGNNWHDSCQWLGDRTITTSTRTWALPTYNNLYKQISSQSGASNDNYFGYST
PWGWDFNRFHCFSRHPDWRQLNPWJWGFDPKMLSFKLFMQVGRGTQNDTITIAN
NLSTVQVFTDSEYQLPVLSAHQGCLLPPFADVPMPQYGYTYLNLNSGQAVGRSFYCYLEWPSQMRLTGNNFQYTFEDVPFHSSYAHQSIIDRLMNPLIDQYLYLNRTQGTTSGTTQSRLLFSQAGPQSMNLQARNWLPGPCYQRQLSKTA>TONNNNSFPW
TAAKYHLNDRDLSVNPAMASHKDEEKFPMHGNLJFEGTASNAELDNVMITDDEEIRRTOPVATEQGYTVNLNQSSNTAPTTGTVNHQALPGMVWQDRDLYLQGPIWAKIPHTDGFHSPPLMGGFGLKHPPQPILIKNTPVAPANPFAFSATKFAFISITQYSTGQVSEIEWELQKENSKRWNPEIYTSNYAKSNVDFTVDNGVYSEPRPIGTRY
YLTRNL
NLSTVQVFDSEYQLPYVLGSAHQGCLPPPPADVFMVPQYGylTLNNGSQAvgRs SfYcLEYFPSQMnRTGNNFQFSYTFEDVPFHSSYAHQSLSRLMnPLIDQYLYlnR TQTASGTQQRLLFSQAGPTSMLSQAKNWLGPCYRQRLSQANDNNSNFPWtG ATKYHLNDRSLVNPAMASUDDKEKFPMHGTLIFGEGTNNANADLENvMID DmEEIrrTwvAETEQGYTsvnnlQnSnAgpTtGTVNHQGALPGMVwQDvYlQG PjWAKjPHTDgFIFHPSLMMGGFGLKHPPQIMIKNTPVPANPTFNsAAKfASfTqYs TgQQvSVEvEWELQKEnSKRwNPEiqYTSNvNvDFTvDngVYSEPRgTRyL TrNL

[0045] YP_680426.1 AAV-2

MAADGyLDPWdLdLsegriQwWKLPGPPPPKPAERhKddSrgLVLPGyKvLGF NGLdkGEPvNeadAaAAlEHDkADYQDQLDSGnPvLykYHnCAddfQERkldFTSG GlNGRAvFqAkkvLEPLmL vEEPVkTAPGGkRPEVhSpvPEd SSSTGkGkAgQqPA RkRLNFQqTGDADSpDPQGPLQQPApsGLGTnMATGsGAMADnNGAdGvG NssGNWcDSTwMGrDvINTSTRTwALPTYnHLYKissQGASnDHyFPYSTP WGYFDnFRFHCfSPRDrwqRLINNWGFrPKRLwKLfMQVKeVTQNDTtT^ N NLSTVQVFDSEYQLPYVLGSAHQGCLPPPPADVFMVPQYGylTLNNGSQAvgRs SfYcLEYFPSQMnRTGNNFQFSYTFEDVPFHSSYAHQSLSRLMnPLIDQYLYlnR TQPSTTTQSRLQSAGASDRDQSRSNWLPGPcYRqQrKSVkTsADnGnSESwTj GAKYHLNDRSLVNPAMASHKDEEKFFQSPGVLIFKQGSEKTNvDIEKVMTD EEEIrrTNpvAtEQgySVStnLQGRnqQaATADVnTQGvLPGMvWQDrDvYlQGPI WAKjPhtDGFHPSLMMGGFGLKHPPQIMIKNTPVPAnPSTFSAAAKFSfTqYStG QvSVEvEWELQKEnSKRwNPEiqYTSNvNvDFTvDngVYSEPRgTRyL TrNL

[0046] YP_068409.1 AAV-5

MsFDUPDwLDeveGEGlRFGLAGpPKPKNQQHQDQRgLVLPGYnLYGPN GldrgEPvNRADEvAREhDySNQLeAGdNpYLYKnHADAEFQekLADdTsfGGn LgKAfQAKKvLEPFLvEEGAkAtPTgKRDdHFPRKKarTeDskPsTsSdAE AGPSQgqQlqipaqPPSSLGDArtMSAAGGGplDnNQgADGvGnASGDWHCdstW MGDvVtKsTRTWLPSvNHqyreIKSGvDGNANAYFGySTPwGFDFNRFH
SHWSPPJ DWQPvLINNYWGFPvPRSLRVKIFMQVKEVTQDSTTTIANSLTSTVQVF
DDYQLPYVVGNTEGCPLAPFPPQVFITUQYGYATLNRDNTENPTERSSSFCLEYFPS
KMLRTGNPEFTYNFEEVFHSSFAPSQNLKLAW LVDQYLRFVSTNTTGGVQFN
K^AGRYANTYKNWFPGPMTGRTQGWNLGSVGVRASVSFAFTTNRMELEGASYQV
PPQPNGMTW^QGSNTYALENTMIFNSQPANPGTTATYLEGNMLITSESETQPVPN
R AYNVGGQMATNNQSSTTAPATGTYNQLQEIVPGSVWMDRDVYLQGIWAKIPETGAH
FHPSPAMGGFGLKHPMPMLIKNTVPVGNITSFSDVPVSSSTQYSTGQVTVEMEWE
L KKENSKRNPEIQYTONEYNDPQFVDFAPDSTGEYRTTRPIGTRYLTRL
P[0047] 3J1Q_A AA-DJ

MAADGYLPDWELETSLSEGIRQQWKLKPQPPPPKPAERHKDDSRVLPGYKYLGP
NGLDNGPEVNAEADAAAALDKADYQRLQDSGDNPYKLKHADAEEFQERLKEEDTFG
GNLGRAVQAKXRLLEPLGLVEAAAKTAGKKRPVEHSPVPEPSSGTGKAGQHPA
RKRLWGQTDADSVPDPQPIGEPPAAPSGVGLSTMAAGGAPMADNNEDAGVG
NSSNGWHCDSTWMGDRTVITSTRTWALPTYTsnSTHYKQISNSTSGSSNDNAYFGYS
TPWGYFDFNPA^HCHFRSPDRWQLINNNWGRPKRLSFKLFNIQVEVTQNEGKTIA
NNTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLNLNNGSAQAVGRS
SFYCLEYFPSQMLRTGNNFQFTTYTFEDVPHSSYASHSLSRLMNPLIDQYLYLSRT
QTGGTTTNQTLGFSQGGPNTMANQAKNWLPGCPYRQQRVSNTSADDNNSEYWT
GATKYHNLGRDLSVNPGAMASHKDEEKFQPSQVLFGKQGSEKTNVDEIKVMT
DEEERITTNPVATQYGSVSTNLQRGNRQAATADVNTQGVLPQVMVQDRVDYLQG
PIWAKIPHTDGHFHPSLMGGFGLKHPMPQIILKNTVPVADPPTTFNQSKLNSFITQYST
GQVSVIIEWELQKENSIRWNPEIQYTSNYKYSTSDVAFNTEGVYSEPPIGTRYL
TRNL

[0048] AKU89595.1 Anc80

MAADGYLPDWELETSLSEGIREWWDLKPGAPPKANQQQQDDGRGLVLPGYKYLG
PFNGLDKGEPVNAADAAAALEHDKAYQQLKAGDNPYLYRHNADAEEFQERLQEDTS
FGGNLGRAVQAKXRLEPLGLVEEGAKTAGKRPVEQPSQPQPQDSSSIGKGGQOP
ARKRLNGGQTDSHSDVPQPLGEPAPGSAVGSNTMAAGGAPMADNNEDAGV
GNASGNWHCDSTWLGDRVITSTRTWALPTYNHLYKQISSQSSGGSTNDNTYGYS
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").

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.

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 acidoterrestris* and Cpf1 (which performs cutting functions analogous to Cas9) from various bacterial species including *Acidaminococcus spp.* and *Francisella novicida* U112. 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 B1; 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
Further examples of Cas9 are provided in the table below:

<table>
<thead>
<tr>
<th>Name</th>
<th>Protein Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. pyogenes Cas9</td>
<td>MDKLYSIGLDIGTSVNGAVITDEYKIVSSKFKVLAGTDRHISIKKPNLIGALFDQ SGETAEATRLKRTARRKYYTKRKNRYCILQEQIFNSNEMKVDSDFFHRLEESFLVE</td>
</tr>
</tbody>
</table>
Staphylococcus aureus Cas9

MKRNYILGLDIGVSYGIIIDYETRDVIDAGVRLKLFDYNLLTDHSELGYNPYEARVKGLSQKLSEEEFSAALLHLKAKHRHNVNEVNEGAGNLSTKEQISNNSKALEEKYVAELQLERLKDGEVRSGISNRTFDYVEKQAKLLVQKAYHQLDSQFDYTIDLLETIRRTEYEGGESPSPGFWGKDKEWYMELMGCHTYPFEELRSVAYANADYNALNDLNLVITRDENELEYYKFOIENVFKQQKPLKIQASEILVNEEDIGYVRTGSTGPFNTNLKVYHDITARKHEIENALLDQIAKILYTIYQSEDQEELTMLSNETLQEEI

S. thermophilus CRISPR 1 Cas9

MSDLVLGLDIGVSYGVLNKLNVGTEIHKNSRIFFPAQAANENLVRTRNTGQGRRALKRKKHRVRVRLNRLFEESGLITIDFTKSNLNLNPYRLVQLGTDLESNRELHALKNNVKHRGSLYDDASDGGNSSVGYAQVIEKSLKLETQPIQILERYQTGYQLRGDPTEVEKDGKHHRLEREDINRFSVQKDFINNLVDTRYATRGQLMNLKSYFRVNLNLDVKEEINGTSFLRKKWFKFERKNGKYGKHAEDALIIANADFITKEKWKLDLAKKVLKYLHMENQMFEEKQAESMPEIETEQYEKIEFITPHQIKHLKDFKDQYSHVREDKPNRELINDTLSTKRDDGNTLIVNGLYDKDNKLKLLINKSEPKLLMYHHDQPTYQKLLKIMEPYGDEKNPLKYXEETGNYLKSYKDNPGVIIKKIKYQGNKNHALILDTPSNRKNVYKLSPYRFVDYLDNVOYKVFVTNKLVDLVKKEKEYENVNSKCYYEAAKLLKILLSQNAEFISFYNHNLKLELYRIVGVNNDLNLRIEVMIDTREYYLENMDKRPRRIKTIAKTSQISIKYSTDLGNYLEVSKKHPQIIK*K

-19-
N. meningitidis Cas9

MAAKFPNPINYLGLDIGIASVGWAMVEIDEDENPICLIDLGLGVRVRFEAEPVKTG
DSLAMARRLRSVRLRRARRHRRLLRARRLLKREGVQLAAFDENGLIKSLPN
TPWQLRRAARDLKRTPLEASVLLHLLHHGVLQRKNEGETADKEGLALLKG
VADNAHALQTGDFRTPAELNKLEKESGHIRNQRGQDGSDHTSFRKDLQAEILLL
FEKQKEFNPVHSVGGKLGIEGLTMLMTQRPLASGDAVQKLMLGHCTFPAAEPKAA
KNTITYAERWTLLKLNRLIEQGSRPLTDLTATLMDHPKSKLYTQAQRK
LLLLDDTAFFKGLRYGDNAAEASTLMEMKAYHISAALKEKLKKSPNLS
PELDQEIASEFSLKDTEDIGLRKDRQPEIALLKHKSEDFKVQLSKALARRVT
PLMEQCKGKYDEACAEIYGDHYGKMTTEKYLPPPADEIRNPVVLALSQARK
VINGVRRYGPSARHIETAREVGKSFDRKEIEKRFQEEQNRKDREKAAKFREY
FPFNVPGFKPSKDIKLRLYEQHQKCLYSQKEINLQRLNEKGYVEIDHALFSSRT
WDDSFSNFVLYLGSNSENQKGNQMTPYEEFNGKDNSREWQEFKAVRVEFPRRS
KKRRILLQKFDGKFLRNLLNTYTVLNLQCFQVADRMRLLGKGGKVRFSAN
GQITNNLLGFWGFLRVAENDRHALDVAACSTVAMQKTRFVRYKEMN
AFDKGTKIDETGEVLHKTHFPQPPWEFAAQEMIRVFKGPDKEFEEADTEPK
LRTLLEALEKSSPEAEVHEYTPFLVFSRAPNKSMQGHBVIEVTQKVKRDLGV
VLRVPLTQLKLKDLEKMNVREREKLYEALKARLEAHKDDPAKAFAEFPYKY
DNGANTRQQVKAHRVEQVKGTVGVRHNGIADNTMVRYDVFEGKDYY
LPVIPYSWQVAKGIPDRAVQVGKDEEDWQLIDDSDNKFSLHPLNLDVEVTGK
ARMGFYFAASCRHTGNTNINIRHDLDKHKNGILEGIVKTAFLSFKQYIDELGKEI
RPCRLKRRPPVR*

Parvibaculum lavamentivorans Cas9

MERIFGDIDGTTSGFSVIDYSTSQAQNIQRGLVRIFPEARPDGBPITPLNQQRQRK
RMRRQCLRRRIRRRKLANTHELAEQFLPGVYGDWVMADEYPYLGLDGLG
EGLSAYEFGRAYILHLAQHRHFKGRELESDPTPDFVDEKEAAANERATLMA
KNEQETTLGAWLKRPSSDRLKRGHIAHRNVEAEFEFLWEYVSFRKPILSEEEM
RARSIDTFIAQFQPVFRNTGECRDFMPGEPGLPKGWSLQSMRRQMLKKNLAI
AGGNRPLDADAERDAILSKKLQQASMSWPGVRSALAKYKQRGPAEKLSKL
FNLELGKESLLGNLAEALKADMFPDPWAHPKQERIHAVHERLWADYGE
TPDKKRVIILSEKDRKAHREAANFSVAFDGIQEAALQALKLPTGWEPYSL
PALNLFLAELKGERFGLAVNGDPWDEWGRWRTNFHQRNFQTEILDKLPSASKE
EREISQLRNPTVVRVTQRNLKVVNNLGILYUGPKDRIEVGRGVGSKERRIEI
QSGRIRNKEQRRKATELIDKINGNAPIRDSRDEVKWLWKEQGQCPYTDQGIFN
ALFREGRILEYHEVIWSRSFDPNKRNTLCKDRNIEKNGNMRPEAFHGHDOR
WSAIQRLQGMSVASKGTTGMSPGVKRFALKTPMDPEAFAQRQNDTRYAAQKQI
LAIQKLRLCLDFMPAEAPVKVEAQTQGVTAGLKLKLWTLNNLADDEKTRAHD
RHHAIDALTVACHPMNKLSTRYQLRQDPRAEKPALTPWDIRADAEKAV
SEIVVSHHRVKVSGHPLKETKSYGTGTDKTKSCTYQFVTRKIESLSEGLE
DEIRDPRKEIJAHAVYEVGDPKKPFAFPYCPVSGPGLPGEIRKVRLTSKQLNLM
AQTNGNYADGSLNNAIYRLPDGKADFIEVSLFDASRRLAQHRNPVTQRTADG
ASFVMSLAAEAGAMIMEPSKKGKIVQVWASGQVQVLREDTADHSTTRRMP
NIPLKDDAKKVSIDPIGRVRSND*
Corynebacter
diphtheria Cas9

MKYHVIGDVTFSVGGLAIEVDDAGMPIKTLSVLVSHHDSGLDPEIAKSVTRL
ASSGIAARTRLRRYRKRRLQLQLQDFQIRQGPVPWIELDSDPYLPWVKRAELA
ASYAIADERKGEKLSVLRHIAHRGWPNRPYAKVSVLLDPGSPDAFKAIREEI
KRIASGQVPETAVGQMTCLGTLRLGEGGVLSLARLOQSYDAREIQEICR
MQEIQGELYRKYIDDVFAAESPQGSASSRVRKDPQPGKRNKLASKADAFQYRI
AAALSGNLVRVDEGEKLSVVEEKNLVDFHLVNLTPKEPWFVTAELLIGDRGQL
IGATMTDDEGARAGPHTDHTNSIVRPLVDWVKTASALEQHAMVKAL
SNAEVDFFDSPEGAKQAFAALDDVDHAKLDSLHPGRAAYSIDTTLRVLTR
RMLSDGVDLYTARLRQGEIPSWTPTPRIGEPVGPNAPAVRVLKTVSRWLESAT
KTWGPAPREVIEHVREGFTKEKRAREMDGDMRRAAANAKLQFEMKLNVQ
GKPRSAFDLWRYSVQVRQCNQCAYCGSPTRFNSMDEMIVHPRAQGFSTNTREN
VAVCHRCNO5SKGNTFPAWAKNTSIEGVSVKEAVEVRTHRHWYDTGMRSTDFK
KFTKAVPERDDEIADRSMESVAMWARELSRSAVQHPSHTGTVVY
RAGSITAEARRASIGSKLGFDDGVSRLDRHHAADAVIAFTSDYVAETLAV
RSNLQKOSQAHROAPQWREFTGKDAAHAAWRVCQWKMSSLALLTEDLRD
DrVVVMSVRLRGLNGSAHKETIGKLKVLSQLSVSDIDKASEALWCALT
RFPDGFDKELPANPERHRVNGTVHYAGDINGLFPVSAGSIALRGGYAEGLSF
HHARVYKITSGKPAFAMLRYVTDLPPYRNQDLSFLKQPTSMRSAEKL
RDAALATGAEYLGVWLVDDELVDTSKIDTDQVKAVEAELGTIRWRWVDGF
SPSKLRLRP1_QMSKEGIKKEAPSLIDRGWLPAYNKLFSIDGNVTYVRRDSDL
GRVRELSTAHLPVTKVQ*

Streptococcus
pasteurianus Cas9

MTNGKILGLDIGIASVVGIGIEAKTGKVVHANSRLFSANAAENAAERGFGRGSR
RLNRKRRKHKVRKVRDFFEKQIGTDFRNLNLNPYELRVLKGLTEQKNNLEEFLA
LRTISKRGGISYLDAAEDDSTGSTDYAKSIDENERRLLKNKTPQIJLRLKGYQ
LRGNKAVTQDENGHELRSVVINFSEYKALRDKHELQTVETKIREKQDCEKSK
LTPQKRYHYHGPNEKSRDTYGRFRTGSTGNLIFGILGKCNFYDEYKASCAS
YTAQEYNFLDNLNVLKSTETKGLSTEQKESLVFAKANTATLPGACKLKEIAKI
LDCVQIEIKYGREDDKDGPKDLHTEFYKLKFLQINLINIDLRESVIDKADILT
LNTREGIEDAIKRNLPQFITEEQSEIKYRKSQQAFTKGWSHPSAKLMMENEL
EYATSDEQMTIILTRELKFKVNKSSKNTKIDEKVDIETFISNQVVAKSVQRTI
KINAAVKCYGDFKIVIEMPRDKNADEDDKIFDDKFRNENKKEKIDDALKRAAYL
YNSSDLKLPDEVHIKNGQLEKIRLKLWQYQGERCILYSGPKISIQELVHNSNNEID
HILPSLSFDSSDLKVLVYAWTNQEKGQKTPYQVVMDSMDAAWOFREMDVY
LQKQGGLKKCRDYYLLTTENIDKEVKKFKIERNLDNTRYASRVLVLSQALSRE
LGQTKSVSQRGQFTQLRKRWDKSKRTRHEYVHHAADDILIIASQKLWKEQ
DNMFVVDYMGKQVDDKQTEGILSVDSEYELKVLQFQYPQFYGFVTISSEKGFEEI
LFSYQVDSKYNRRKVSATSTRKAIJKGDKEETVVLGKIDYSQMPFDTFIK
KYKNKDTQFLMYQKDQLTWENVIEVLDJYTPKTKSSEDGKDVKNCPFEYRR
ENGLICKYSKGKGTPIKLSDVRKSLKGLGNCITDDEIERSNKVNLSINPDWAD
FNPETKLYELMGLKSYLDSFEKGTVNYHISQEKYDAIKEEKIGKKSFEKFTLY
RNDLGIKDIASDGEQIEYRLSRTMPNVHVELOWKSYDEKFDNVQELVAEALGE
ADKVGRC1KGLNKPNISYKVRDTVLGKQVFVKKGLDKPFLFKNYK
Neisseria cinerea Cas9
MAAFKPNi+sNLYGLDIGIASVGWAIEDEEENPIRLILDGLGRFVFAEYPKTCPGDLAAARLRVSRKLRARRHLLARLLLLRREGLQADFDENGILKSLPN
TPWQRLAALAKLRLTPLWLEASVLLLHILHRGLYSQKNGETADKELAQLGK
VADNTHALQTFDRTPAELALNKFKEQSHIREQRGDSYHTFRKDLQAEILNL
LFEKQKFEGPHVSDGLKIEGIEQMTLRQPAALSGAVQMLGLHCZTEPTEPKA
AKNTYAERFVLWTLLNRLQIEGSEPLTDRTATLMDPERYKSKLLHTAQYA
RKCLDLDDTAFFKGLRYGKDNROLEMAYHSAIRELEKLGKDLP.
NLSEPLQEAFLSKFTKDEATRGLKDREVPQLEEPALLLKHSDFKQFVSILKAL
RKQJLQNNEQGNYQDEACTIEYGHDHYKGNTEEKYYLPPIPADEIRNPVVLRLASQ
ARKVNYKRYOSIPKARIEFVYKGSFDRKREIEKREQUENKRDREKSAKAF
REYFPHVOFPKVPSDLKILKLYEQQHQKCLYSGKIEKLNGLNKEGYEIDHALP
FSRTWDDSNFKLVKALSQENQNQKGNQTFPEYNYGKDNNREWQEOFKARVETSR
FPRSKQIRILLQKFDDEGFKERNLNDTRYNRFQFCVADHMLLGTKRRGFVR
ASNGQITNLQGFWGLRKVRAENRDHALADVVACSTIAMQQKITRTFRVYK
ENANAFLQDIETKGEVNLHQAHPQFWEAFAQVIRFGKPDGKFEEADT
PEKLRLLLAEKLSSRPAPAVYKVTPTLSFRAPRKMSGQHGMETVYKSAKL
RDGSLVRLPRLQKLKLDEKIVMNREPKLEYALKARLEKDDPAKFAYEFY
KYDKGANTQRQVKVAREVQKQTGKVWVHNHIGADNATIRYVDFEVKGGK
YLPIYLPWSQVAKGILPDRAVVQKGDDEEDTVMDSDFEKFFVLYANDL
TKRNKEFNGYSVSLNLRTAGAIDIRTHTDTSTKGGNIFQSVGVKALTFSQKYIQIDEL
GKEIPRCLLKKPRPVR^*

Campylobacter lari Cas9
MRJLGFDCIGSNGWAVENELDKCDGVRTFTKAENPNKESILALPPRNARSSRR
RLKRRKARLLAIKRAIKARLKLNYKVDVAADELPAKAYESILVYRLKALT
QNLETKLDDLARVHIHAKHRGYMNKNEKSDKNAGKIKSLKANNALKLENYQS
VGEVYFKEFFQKYKNTKNFKNKJRTKDNYNVCVLSSDIKEKLJLLEKQKEFG
YYNEDSFIEINEILKRFAPQFLPKDSHLGYACTFFEEEEEKACKNYSWAWEFVA
LTKINIEKSLIEGIEVPFTQTNITENVLNLIDKIGSITEYKFRSFCINEHSIFKSLK
YDKENAEKIKLDRFVLKEGALVHSLQSELQIDSTHFLSDKVLTVKLYNS
LSNEQINNNLIEEFNYDNIILSFKALGMILPMRELKQREDEACEIANLKPITVDEK
KDIFLPACDSIFAEHSLNPYVNAISRYKVNLKYGVKHILERARVDG
LSKKAIREKIEEKQENQAANAVALCNEIGNKLASSIKLKLWKELKEICIEY
SNGNISIEHLLKDEKALCEVDHHPYRSFDSDFINKVLYFTKENGKELNKTPEFAE
GKIKWIEKSIQTLAQNLPYKKNKILDENFDFKQDQEDPSLNLNTRYATLIA
KTYKELNYPLLENNISNFLKSGKSHQVHTISGMSTLSVLHRTWGDFFKDRN
NLHILHALDAIVYESTINSIIKAFSDFRNQELKARKFYAKELTSNHYQKVFKEF
PFKFSDREILSKIDEFVSKPFFPRKARKNIHKDTHESNIIKDCSNNYKEGLIAL
SCVRGKREIKFYEGNTINDTVRFDFKENQKFAYPIYAMDFALGIPKNIVITKG
KDNKPNQPKQWTIDESYEPCFSLYKNIDLQKMKMPFQEPFAYYNDSSSTSSCIVC
KHNDNKFENTNSQKLSSNAKEEGSVKVESGQLKVFKEYHITPLGDKIDAFQ
PRENSILKISKYGLR^*

T. denticola Cas9
MKKEKIDFYLFGLGDTGTSGVWAVTDYDKLKLANKRDLGWRMCRCFETAEATAE
VRHLRHRGARRVIRSRKRIKJKLQKILQFQELSIEAIKTDEGFFQRMKESPFAVEADKTLQ
ENTLFNDKDFADKTHYKAYPTINHLKAIWENIKVPKDPLRLLYACLHIKKRH
FLFRGDFSDENQFDTSQALFELREDVIDADSDQVKEILKDSSLKSNEKQS
RLNKGLKPSDKQKAIINTLSINGKNIFADLYNPDLKAEKNSFFSDFFDA
LDSDLASSGCDSFELLKAKAVNSVLSKVIGDEQLYSFAKVKYIKEHTDIL
KLKNVIHHPKFDKYKKYVFGRAPHKNNNKNYSGYQVGCYKSKKLIINNVQQ
EDFYKLFLTLSSAISEIKEVNDLITEEIGFTTYLPOISKSNAAEPYQLMKEMEK
SLNAAEHPSFLQKDEKGISHEKIMILTFKIPYY1GPINDNHKKEFFPDDRCWVVK
KOKPSGPKSPPTWNNFHDIDKEATAEFTSRNFTCNYLGVESLPGKSSLSSYETT
VLNLEHNLQIIDIQGKICDLCQKQ1YELDFKYYKITQKQSTSIKHEICNG^11
UVEILGDIKETSSLKSYELKINIFQKVEIDSTKMMLEIEIRWIYATJDEGKTLK
TKIAKAYGKYSECDEQKILKILKFSGWGRSLRKFLETVSEMPGFSEPNVTIAT
RETQNNMLLESSFETNOKKINSGFEDAEEKQSFYDGLVKLPLSPVKKML
S. mutans Cas9

MKKYPYSIGLDINTSNGWAVVTVTDKYKPAKMKVLGTNTDKSHEKNNLLGALFDSGNTAEDRRLKRTARRYTRRRRNLYLQIEETFSEMGVQDSSFPHRLEDLSFTVETDKRGERHP1IFNLEEVEVYKENHPFTHYHRQLYADNPEKVDLRLYALALAIKIKFRGHEFLEKGFDTRNVDVQLQREFQFLAVDYNFENSSLQAEQNVQEELILTDKIKSSAKAKVRKLQLPNKESNRFAGELKLLIVNGQADFKKHFEEKEAPALQFSKIDSTEEVEOLEVAAIGQNDNYAEFLSLAKKYLDSISSLGISLTDVGTKAPLSAMQIRDNEHQMLAQKLFQFRKLSDKYVEIVFSVDSDKYAYGIDQKTNQAEFYKLGLKLKIEGGYFLDLDIERFLKQRTFDNGISPHIQHLMERAIIRQAEAYFFPLADNQDIERKLILFTRIPYYVGLAICRDSFALWLSKADKTPWFDIEVDESSXEAFAPIRMTNYIDLYLNPQKVLPHSLLYYEKFVYYTLEVTKYKTEQGKTAFFAMDQKQIEFDGVQVFVYKVKDLEKEFDFRVIDTLDGKLKVNAYSTYHDLCKLKDIFDNSKEKLEYLDEVIDLTLTSTFREDRIKKLRENYSDLLTEQVVKLKRHYTGWQGSAELIHGIRNKESKRTLDYLDDGNSNRNFMLQINDDA

S. thermophilus CRISPR 3 Cas9

MTKYPYSIGLDINTSNGWAVVTVTDKYKPSMKMKVLGTNTSKYIKKKNLLGVEFLDSGTAIEERKLRTARRYTRRRRNLYLQIEETFSEMATLDAFFQRFDLSLFDVDPQDKRDSKYPIFGNLVEEKAYHIDEFTYHRKLRYLDSTKADLLLRLYALAHMKARYHGLEIFESNKSNDIQKNGFQDFLTDYNAIFSLEDSLSENSQLEELIKKDISKLEKLDRIKLDFEGKNSGIFELSFLKLIVNGQADFKCFNLEKASHLFSKESYDLEDDEFTLQKFKIRRYTWDYKLYGFTGDFPSVAYSVIAIEDKGSXKLKTLKALGVTIMEKMTFERDPVAFLERKGRYNNQOEIPIKYLPSKQFLKGNRKLASSARELQKONIEVLPNLQLTLHAYKNHMDKFPDLHDVYKHDKEKELELVDSNSFIKDLYLAEINLEKLYAQNNGELKKEALASSFINLFTTAIGAPATFKKDFNDRKDSTYRNLATMLQSIISTQIGLETYRDLNLKLGDG
ETRQITKHVARLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVR
EINDFHHAHDAYLNAVVASALLKKYPKLEPEFYGDYPKYNSEFRJEKATEK
VYFNSINMFFKSLIDGRVIERPLVEETNUTGEVWNKLELATVRVSLYPPQ
VNVGKKEVEQNHGLRDKPRKGLFNAALSHPKPSNENLVGAEYLDPKXYG
YAGYSPSIFTVLOKGTIEKGAUKKTNVLEFGISLDRINRKYDRLKFLILEGK
KHDLIELPKLYSLFELDGSSRALSLSTNNKKRMEHNGQNIFLSQFVKELYH
AKRSITNISRENHKYVENNHKEFEELFYIYLENVYGAAGKNNLSSAFQSW
QHNSIDCCLSFUGPSTGERSKRLFELTSRSGAADFEFLGVKIPRHYDTETTSSLKD
ATLHIQSVTGLYETRIDLAKLLEG

C. jejuni Cas9

MARILAFDGISSGWAFSENDELDKGVRIFKTVENPGTKESLALPRRLARSK
KLARKLARLHLNKLHIANEFLKNYEDYQSPFSLKAKAYGKSPPLELFRAL
NELLLSQDFVARLIHIARKRGGDYDDKNSDDKGEKAILKAQNEEKLANYQVSG
EYLVEKHFETAEFLKKFVSNKKEFVRYCQATSLKDHELIFKQOREGFSS
SKKFEFEVLSVAFYKRALKDPSHLVGNCSFFTDKRAPKNSPLAFMFVALTRIN
LLNLNNKEXITGTYKDDNLNYNEVLKNNGTLYTKQTKKLGLSDDYEFKQGEG
TYIEEFKYYKEFIKALGEHNLSSQDDNLNEIAKDTLIEKIIDKALAKYLDLNQNG
IDSLLSKLEFKDLHNSFALKLVPTLMELGKDYDCEACNECNLKVAIENDKFEL
PFNATYYVYDEVTNPVLRHAYRKYVLNALLKKGVHKIINELAREVQANKHN
SRAKRIEKEQENNYAKKKDAECEKLGKNSNKLKLRLFKEQKEFCASYGE
SIKSKLQDEKLMIEHYPPRSFDSDMYMKVLFTKQNEKQLNQTFEAGNS
DASKWQKIEVLAKNLPTKQKRDLKNDYKDEEQNKFRDNLNDTRYARIKLV
NYTCKYDYLDFPLSDDENTKNDTQGKSHVEAKSGMALSALHTWGSAKD
RNNHHLHAIAVHIANYSINCKSFQDKFQNWIGENASLKYDKKKKFFEEPGF
FRQKVLEDKIDIEFVSPERKPPGSHLHEIFKKEEEFYQYSQGGKEGVL
KALELGKIRKVKNGVKNQDMFVRDEFHKTHNKFYAPIYDSFMDKLVLPNK
AVARSKKGEIKEKDWILMDENYEFCFLYKDSLCLIQTDYMDQEPFVVYNAFTST
VSLVSKVDNFKETLSKNNKILFKANEKVEIKSIGNQLKVEYLYSVET
KAERFQRELSDK

P. multocida Cas9

MQTTNLSTYGLDLGASVWAVVEINENEDPQIGLYGVDFERAEVPKTESL
ALSRSLARRSTRRRIARHLLKLAKRFKREGSTIDLEKGPNQAWELRVAQL
ERRLSAEGWAVLLHLHKLHGYLSKKKESQNNKELLGALLSGLANAQLLOQ
DSYDRTPAELFKFKAKEGHNQFRGAYTHTPFRLDALAEMLNFQAOLQGFNG
PHCEHEIQQYMTELLMQWPKASGEAILMLGKTCHKEFNAKAAKHTYAER
FVWLTLKNNLRILEDGEARLALNEERQLHPELYSHKSLTAYQRKGLSLQEA
IFKHFLRYSKANESAFTMELKAWAIRESQNLGKDLTQDLAKKPDDEIDG
TALSFLYKTEPDIOQLYLTVKNPSENVLSNFDKIELSLKRLKIPLMEQG
KRYQDACREIHYGAYHEGAQNKTSQOLIPAIPEAIERNPVLRTLSQARKVINAIR
QYSGSPARWHEQREGLSKFREREQQKQDDNRTKRESAQVEKFLFDSFSEP
SKDSDLKFRFLYEQOHGKCLYSGEIKHNRKLYGEYIDHAPLSRTWDSNFF
KVLVLASENQNKGNQTPQYEYWQGKINSERWKNFVALVLSQCSAAKKQRLLT
VQIDNDDLRNLNTRYARISFYNQINELLVGNKNNKVTFPPNGQIIALLRSS
WRGKLAKARNMEMHJHALAVIACATPSMQQKTRIRFKVEEHPHTEEKR
VQDSQSEGIIHPFEPWAWYRQEVYRIRVFDNHPDTPVLKLMPLRDQPNAHQQQFVQPL
FVSAVTPTKMSGQGHMETHAKSKLAEGLISVRLPLTQLNPLEMNKRN
ERALYAGILKRLAEFQDPAKAFATFQYQKGGQVAIREFQVQKSGGLVRENN
GVADNASDRTVDFKNNKFPVFVNYTPWOAKGILPKAINHKEAENEMEDM
EGASKKFLSFLPDNLVELKTKEFFYGGYIGLDRATGNLSKHEDEISGKGDKG
YRVEGKVLALSFEKYQVDELKNQRCIQPRQORPRV

F. novicida Cas9

MNFKILPIAIDLGVKNTGYSFAYQFKGTSLERDNKNGVYELSKSYTLLMN
RTARRHRQGRGDRKKLVRKLFKLIWTEQLNLNLEWDKTDQQAIFSLRNRGFFSFT
DQGSYPELYDWQPKVEKAILMDHPFNEEDLDLSYLKALEQSEKIEYKLM
QKILEFKNLKLCTDIDKDKSTKLKELSIYFELLADYLNYSLELKTQKFSYT
DRCQGNKLKSYHHDHYKIYQIEFLKHKATINDRILTDLTDLSMFNPKEKDF
**Lactobacillus buchneri Cas9**

MKVNYNHIGLIDQGTSQGIGVAGIKDDKPLVRKLGTAQARFQEGNPAADRRMFRTRRTSRQKWRLKLEIEPYITPDVTSTTARLLKQSNLSKPSKFKGFKLFLPDLTDQYHVHNYPTIYHLRHAMTDQDKFIDRMYVLAIIHVYKVRNGFLNSPTVPSDFASVVDVDFQPFKLKFNELAYAANPEENSKFINKLANSEIDHOFQFESPKFDFKKQKIPKIVPVMNKDVTDRNQKASEIHAILGYAKLVDVLQCTLVPSDSPWALKKFDDEDAKIKEPQEMENQDSVQISAVILQLSHYQLTVLQVINPMGLESMIEKYNHIDHLKLYKDKLQADPKKAKLVCYKYASQYGVDGKVIKQVEAEQFSSVSKKNLDDLSQKQIMDLDAEKFMPKQRTSQNVGPIQHLQRDEIQEHEQSKYYPWLVIEINPNKFLALKHYKIIEQFLVFVYVYVVPITMPDQKESAETVFSWMERKGTETQIPTPWNPDFTEAFKKASKARASIFBNRKMRTDDTLYGLDVPDESLVYEFKFKVNLNMLVYRNVQKKEKLVWDAKQAIQDFLFFQMLNYKHVSKLKLNQYIKAKTGPLDFSEIGSLDPEHFNSFLYTDFFKLGSKVDEPDIQDFEFKVEWSTFEDKPKLIREKLNEITWSDLQOKDVLLESSRQYOGWLLSKLTLTGINVDEGRIJDKLWQTNKMFQIQUQSFQDVKER1HNEAMDQVAMQVDVVLADAYTSPQNNKAIRQVQKVVDNDIQQKAMGGVAPKISIESFTRSDRNPRTISQRQOLENTLKDASTAELKSIENPELSDNALAKSMQDFTRLKLYXLQFQLMQDLTQIYFEGINPDELNLYIDHILPQAFIAKDNSLNRVLTVATVNNKDPROJECTLRMFAGAKMHFWPKQALEAGLISRKLKLNQTDPTKSIYMHGFIRFRQETSQVIKLVALNKQYNRDFTKJEITARMHQMHRDENGFKREINRYDAHFRDATLAFGLYLYHRYKLRPYFVGVYDGFKFREDKVTMLNFLHNFLDHTDQQIKIAIETAEVGEWDRESIQLKDYVHYIKFMLISHENYTGRAMFMQNTVQVDPSADGRKFLPKVPADKVPVVNYGGYSAGADMAVIRHNNKKGDKVYVGRPMARLRLAIIAMKDNFADFRALKDVLAPQLTGTKKSRKETGEIQFEDPEILYKMCYQMLMDGDKFMGSSSITYQYANKLQLLSQSVKTLASKRLQPLESNQNYNTVEILDKNQYFSLYDMNFKHILNLGFJKFISPWFHNVLQGNTKVSSGRELQELNGLHANPFTGQNLKVGETTPQGQLQOPQNIGLDTSEKRYQPSGETFKEVLSDLK

**Listeria innocua**

MKKPYTIGLIDGTSQGIGVAGIKDDKPLVRKLGTAQARFQEGNPAADRRMFDEQATADRRMATARRRIERNRSRLQGFAEMSDKTADFNCFLRSLDSFYVDNEKRNSRFHHATIEEVEYHKNYPTIYHLRELVNSSEKADLRLVYALAHI

-25-
Cas9

IKYRGFLJEGALTDQTNSVDGQYKQFIQTYNQVAFASGIEGDSKLKLEDKVA
KLYEVRKTEKLERILGLYPGKSAAGMFAQFLLVSIGSKGQRFQDKPLLEKSDIE
CADDKSYEEDSLNALIGDEYALFVAACKAYSAVLLISHTVAETETNKLASS
MIEFDTHEEDLGLKAFIKLHLPKHYEEIFNEKTHGAYGIDGKTQADFYK
YMKMTCLEIGADYFIAKIEFNLRLRQRTFDGAIHPQHLLEELAILHQAK
YYPPFLKNENYDKLSTVPYFIQPVPGLMGQSFALWLTKRAGEIRPTMEEK
DFGKSAFLIEKMNTKDTYLPKHLPCYQKYLYVYNELTKVRYINDGQK
TSYPFSQGEKIEQFNLDFQKRRKVDLRELNMSLYQEHQSSTIEGLDSNYSSY
TYHDDKLGVIQEILDNPVNETLEMELV1KTLTVFMDEKLQMIKELQIQSFSDVLGD
VLLKLREHRHTGWRGLASAKLLMGRDKQSHLTILYLMNDGGNLNRMLQLIN
DSNLFSKIIIEEFUQATDADKIQSIVADLAGSPAIKG11QSLIKVDELVSMGYP
PQTVIVEMARENQITGKGKNNSRPYSLEKAIKPGQILKHEHPDQNLRRN
RLYLQYQNGKDMTYQQDLDHNLNSNLYDHDHIVQPSFTDSKIDNLVLTSSAG
KEDGDPPVPLEVQKRFIVWEKLQQGNNMSKRFKIDLTQGELLEADEKAR
FIHRQVETQITYTNVANILHQRNVEKDKHHGTMQVRIVLTLSALSQFRKQQ
FQOLKVRDNVQLEQHSAHYADAYLNGVAVANTLQVYQLPEFVGYDHQFDWFK
ANKATAKQFFQTY1MLFAQKDRIDQENGEILLWKKYLQVXVSMQYRMNIV
KKTQEGFLQSKTIPQNSNSKPIQTPKNDPQMKAYGLDPSNMAVIEYAS
GKKNKLVFEEKIERIVT1MERAKEDKAFLEEQGYRQPKVLAKLPKTLYECE
EGRRMASANAEAGKQGQQLPVNLHVTLLLHAANCEVSDGKSDLIESNRMFR
FEAHLHAYSEFARKTYALEANLNKINQLEFFQKEQDKIAQSVFDLMAFNAM
GAPASFPKEFETTIERKRYNNKLKELNNSTYIQSYSGLYSERKLD

L. pneumophilia
Cas9

MESSQILSPGDLQGKGKFTGVCSCSHLEAFELPNHANTKYSVILIDHNNFQLOSA
QQRATRHRVKNKRNQFQVRKJLQQLHSLRDLNAKEETALCHLYLNNRGT
YVYTDQDLLDEYIKDETINLIIKELLSEHNFIDWPLQKMQISERKILVSKVEEK
KKDKELKNAVKNKJPEIFGQFNSVEQVHKYVENIKSDITKDNQLSDKIKK
PSVLCSNLLGQHLQWQKNLHRKLYANKPKQDEQTFQGELMRKLNFHRLKGS
QESLAVRNLQOLEQOSDYISILEKTPEITIPPYARNTMGEKQDOSSLNLPEK
NNLYNPWNRLLPIGIHAIHPFLKDEHTLKLDRKRIPKSQKDERSYILQRYLD
LNKKIDFKQIKQLQSLFQGKQPLANLITEQKEMETHFLSSLSVLQIASSAYN
KEDAAQGIFWNFDNASLCELISNPPKRQKILPLLQGAILSEDINLNKDQWAFK
(IFWNTKIKRGTSLSKSSCEIEEARKNSNAGFQDIYEEALNHPESSNKALKIQIQT
IPDIIAQQISGHLDNSQSLYIHFNPSLQLYLTILEKTRDFQHNCVACNTNYW
RSQKTEIDPSVYASRPADSPVFDPGVLARMMQRLAYEAIMAKEWQIHIHDN
SSLIPILYQNERFEESFKEGKIGSKSDQKTLEQAIKEQHNOI7EIQFRQINASMINI
CPYKGASGQGQGFIDHHYPRSLKKHKGQVFSNVELYCSQGNNRKEEHYHLL
EHLSVLKHKQFGTDNVSDFKIQSNNVANIKYKSFELLTEPOQKARHAFLED
YDDEAFTKPFMLQSQQAROKVTNGKLFQKJMELSTLADSKQLLEFSEIQKT
AEEVDHREKLSSKQEPSKVLKSRQQFSFHAIDALTMSIGLKEFPQFOSQELDNS
WIFINHMFDVPLNPSVRKSEPKYNNISITPSLFKDSLYMRFVQPPVWVKGETFAIG
FSEKDLEIFKPSNKEEKLFTLTLYKTSPNGELQZAKSKAKWLYPKNTKLAL
EPLFHYKIEKIVPTDDTVCHFINSLLYTYKESITVILKLEPMVLSKVEFSSK
NVLGSFQHTILAPKTDWELNFHPLASKAPNPKENEFIRKYLSDN
PNSDIPNNGHNIKQQKHKAVKVFPLVPIGAGTMRRKDNKKQPLYQLQ
TIDDPSTMGIQINERDLVLQEVFLMDAYKTRNLSTIDGNNQNGSAYATDWNWL
LPVSTSFTPEIKEMLKPHSSTRYRIRQTSLIADFIKTIDEAIKMPSDSDPPLMN
NEIVCKNLFGNKLPRDGKMIIVSTGKVITYEFESDSTPQWQIQTLYVTQLKQK
P
N. lactamica Cas9

MAAFKPNMNYILGIDIGASGVWAMVDEEENPRLDLGVRVFERAEVPTKGDLSLAMARRLARSRVRLTRRRRAHRLARRRLKREGVLQADADFENGKLPSNTPWQLRAAALDRKLCTCLEWSAVLHLVKHRYLSQKNEGETADKELGALKGVADNAHALQTGDFRPTEAELANKFKEKESHIRNQRGQDHYSHFSRSDLQAELNLLEFKQKGEFPGNHPVSDGLKEDDITLMLAQRPLASGDAVQKMLHGFCTFEAPEPKAAKNTYAERFWLTKLNNLRLIKEGASVHRDPQKHSENGFKQKDESFQGKMKLGTLNLGKQQQAIKQFPNFVPEGKPSKDMILIQLYLIEQOHGKCLSYGKINLNEKGVIPDHALSRTWDSSFNNKVLGSENQNGKNTQTYEEFNGKDNRSWQEFKARVETSFRPRSKQRIQIKQDFEDEGMKNNLNDTRVNYNFLQFVADHILTGLKGRVFESANGITNLHGFGLKVRVYETNDRHDLADVVACSTVMASPQKTRFYRKEKMNADFGKTIKDTTEVNLHHQKFQHQPWFEAQQVMIRVFPGKDPQFEEFEADETEKLRTRLAEKSSPQEHYVRTPLFSRANPKMSQGHMETVKSARKLDEGSVLRVPLQTKLKLGKEKVMNREERPKLVDALQAQLKETKRDDPAPFAEFYKDYKSGRTSQVQAVRQEQQKVTGVWVVRNHGIAQNTVMVRDVDFEEKGKYYLYPIVSVWQAVKGGFDRARAFDVDEWDVMDSIFEFVFLLYANDILKLATTKEFLEGYFVLSNARTAGIADIRDTHDTSTKXKGIFQSVQVKTALSFQJNIQDELGKEIRPRCLLKRPVPR

N. meningitides Cas9

1VLAAFKPNPNYILGIDIGASGVWAMVDEEENPRLDLGVRVFERAEVPTKGDLSLANARRLSRVRLLRARRRLKREGVLQADADFENGKLPSNPWQLRAARALDRKLCTCLEWSAVLHLVKHRYLSQKNEGETADKELGALKGVADNAHALQTGDFRPTEAELANKFKEKESHIRNQRGQDHYSHFSRSDLQAELNLLEFKQKGEFPGNHPVSDGLKEDDITLMLAQRPLASGDAVQKMLHGFCTFEAPEPKAAKNYTAYEWFILWTLKNNLRLIKEGASVHRDPQKHSENGFKQKDESFQGKMKLGTLNLGKQQQAIKQFPNFVPEGKPSKDMILIQLYLIEQOHGKCLSYGKINLNEKGVIPDHALSRTWDSSFNNKVLGSENQNGKNTQTYEEFNGKDNRSWQEFKARVETSFRPRSKQRIQIKQDFEDEGMKNNLNDTRVNYNFLQFVADHILTGLKGRVFESANGITNLHGFGLKVRVYETNDRHDLADVVACSTVMASPQKTRFYRKEKMNADFGKTIKDTTEVNLHHQKFQHQPWFEAQQVMIRVFPGKDPQFEEFEADETEKLRTRLAEKSSPQEHYVRTPLFSRANPKMSQGHMETVKSARKLDEGSVLRVPLQTKLKLGKEKVMNREERPKLVDALQAQLKETKRDDPAPFAEFYKDYKSGRTSQVQAVRQEQQKVTGVWVVRNHGIAQNTVMVRDVDFEEKGKYYLYPIVSVWQAVKGGFDRARAFDVDEWDVMDSIFEFVFLLYANDILKLATTKEFLEGYFVLSNARTAGIADIRDTHDTSTKXKGIFQSVQVKTALSFQJNIQDELGKEIRPRCLLKRPVPR

B. longum Cas9

MLSRQOLLGASHLARPSVNRDNQHVDNCSYGERCFMRGKRYRIGDVGVLNSVGLAAEVSDENPSRVLQNNASVLHVDGVDVQKNNKEIRTKMSGVARRTRMRRRRRRLKHLDMILGKFYGIEQVEPSEDLDKPEEFWHYRAELATRYIEDERRESIASLHNNAIHRGWRNPYVQVDSLSIDNPQKSYGKELKEKAKAYNDDAAEEESTPAQLVVAMLDDAYEAAPRLWRGTGSKKPDEAGYPLVRMLQEDNALSEIFHRVRQRPVPAEDWKLFRSVFYASFKGSAEQRQGKDLAEPQQARALASLFQERYRANVTNLIKDSAESLRLTVLEDQSYIDQLVPSEDTWSLDCDFGKLRSQKLGSTGVQGSLTEDIGEESRRPLLSTQVSQYEDSNKIRPLLAVWKNASDNEHEAIMRLLNSTVIDKVRDADVYASAIIEFGDLDLDAKTLUSVULPSRAYAEEVLTLQTRQMLTDDHIJEARTLKFNTDSWRPPADPIEGPLNPSDVRLKNVNRYLMLCQWRNGPVSNVIEHRSSSSVAFARKDREYKNEKRSIFRSKLSEQLRADEQEMKVESDLRRLEAIQRNGQCLYCRITITPRCMEHDHPR
A. muciniphila Cas9

```plaintext
MRSLTFSFDIGYASIGWAVIASASHADDAPSVCGCTVLFKDCQCAFKKREY
RRRLNIRSRVRIERIRGLLLQAVIQUITPEMKGSHPAPYLAESALKGHTRLAP
IELLWHVRLWYAHNRGYNNASWNSWSNLSDLGGSNGEDTRVKAHQDMLDCHGT
ATMAETICRELKLEEGKADAPMEVSTPAYKNLNTAFPPRLIVEKERIRLIELAP
ILGKTAIEILIAQAHIPLTTSEORQVLLQHIKLARRYRSLLLFGLIQIPRFDNJRIS
CPVTWQAVYLLCSVKSNQSNARERALKSKVTPCNEFEEYRIMARLINCN
ADGEPLSARERLMNCAEQGKLTAKLSEAIAISRLGKGTDTSNVSYFTLHDP
SEEALYNAPLNVQREGGIQLPSVYRIAANLRKGKSVTPNLYNLKSSRG
SGEALEKCIKESKKEADYADLPIKRPYATCPRAYRTVLLKVVEEILGEDDP
TRPARGHAEHPGDLKAHDGCYCLDTSVSNQPKERRLDTMNNLVRH
MLILDRKLILQIDFAEQGDKRISVCVERELTIFTSAMDSKKIQRELTLQFK
SHTDAVNRLKLPLKALKSLIRKCKIRAMDMWPTFTGAYTDHIHELLELD
EHIVPHERQFSNLASSLVLTYWPGYNVRMRGKRGRYDFVEQEQENPVPPKLNLHI
CSLNYRELDKDCDKGHREDRRRKLALLMVRLSHKQSONHEAMK
EIGHTEMGMTQSSHALMKLACKSITSLPDADHMPGAVTEAEVRKAWDVFVG
KELCPERASDPGSKILKIRNLSTLHIDAVLCLGYPILIPAHINGLNRRLVRA
MRRIPEKLPQVPSANQRYLVNLNDGMRMLRLDSLKALEQMEQVRVIQ
HVPADMGALLKETMQRVLDSGDAMVLSKKGDKGKQVAKSKLGV
GFVPEPSKLLALKAKAEIDGNYGVALKPVPIRVHKIVFMRKIALEKQNGGKP
VKRLKGMKLILHTSSKKDHAGVWRIEISQDSKKGVVKDLQRAHVACPKNNKTH
ECNWRVEFLSLLLLKQMKRPTSYTGTTP
```

0. laneus Cas9

```plaintext
METTEGIDLGTSNISGLALVDQEEHQILYSGVRIFPEQGKNTDGILTGEKEESRNATR
RKRQMRQYRFKRLKAKLLKLILYDMPCLKPEVRKWNWDQKQSTTV
ROQFTDFPAFKREWLRKQNPCPQRSFTEDTRPELRLGYIQLMQRGFLRSSKGG
KEEIVPHERQFSNLASSLVLTYWPGYNVRMRGKRGRYDFVEQEQENPVPPKLNLHI
TLMYRDEEFHWEHILQAGLHAEQTARTRKNNFLGESATNVRKLHQLA
KYGRHGHLDEITRIVTFQLPKCELVGIEIEEEOLFKSNESFLWRQPLRSQ
KSSLKCCEFVEFVRQPKWHPQVIIAGPTAPSHPEEFQREYQFIINNYGKH
LTATQREAEEVFMCESIDKnFKPIPKHLKLFKENFDFTTVPACRTASIQRRKL
FPHVPWEEKREEIWHCFCPFYDNTLLFELKQWDALQTNDLKEKIRLLESYGF
NVLSLAIKRPDEKLGCTVYLSALGGIRNSFGKRFYFYFKEYEPEIEKAVCRIL
KEKNAEVEGIRKIKDYLHINRFAGKDNRAFQKLYHHSQAITTAQKERRLPT
GLNRNPVIQGQLNLQVTRMLLATCREKYPSPKFHDHVMGERLSSKTER
EKOQSKRJIRENEKKNIAKVKLAEYGLKARYDNIQKYLKYEIIEKGGTCCPY
TGKTLNISHTLSDSNSVQIEHIPIYSILSDSANKTCDATFRENKELTPYDFY
QUKDPSPEKWSGWAEEIDRALLPFAKAQQFIRKQPFSENPISFIRQLNDTRYI
SKKAVELYAISCDVAKFPGNLTAERLWGNLNLQASPDITFPLPSATENHR
EYEVITNEQDNVERLPFQGETQRTPEKEGLLTVGEVERKVFRCGGMQEQFTDTYS
DKGYWKRKLSSVTSWPLSKAPPSADQGVLILGKRIEVCVQNCOKLKT
LPDQGYSWISLPVLSQTFKKEGESVNNKTSQQVLGFRVREGIRFCHNYOCPAG
ADGNFWCCTLDTATQAPFTPKANPQGGVQQILQTLVDGDKGGDFADDLOLYE
LIPLSKPGKYGTIFVESCDPTLIELAPSKTSKGENGLIENIWYDEHTEGVFRDF
PKKNREDQHHAIADLSISQLSFLQRLSTYNARRENKKRLDSTEHFPSPWP
```

-28-
Those Cas9 sequences used in the examples disclosed herein are provided below.

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

MNFKILPAILDGLVKTNGYAFQKGTSLERLDNKNQGKVYELSJKDSYTLMLMNNTAR RHQRRGIDRKLQVKKLXITEQLNLEWDKDTQQAISSFLFNLNRGSFITGDYSEPE LMVEPQVAKLMDIFDDYNGEDDLSYKLATEQESKJSEYNKLQMKEFLKMKL CTDIKDDKYSTKLETIESFELLADYLANESLKTQFKSTDYKDQNLKELSYYH HDKYMQEFLKJIHATNNDWLDLLLQDIWNFENFEKFDKNEEKLQNQEDKDIH QAHLHHTFVAVNKIKSEMASGHRHSQWQEIWTLDENNQEGYLKNFCENLHMC KYSNLSVKNLVLGNLNLKPLRKYFNDKIHAKADHWDEQFTETYUCHWILGE WRVGVKDQDKKIGAKYSYKDLCNELKQKVTKAGLVDFLLELDCRTIPPLYLDNNN RKPPKCPSSLNPFLDNSQPNQWQQLQELKQLQISIQNLDSFTDLKVLKSSKDQP YFVEYKSSQNIAQGQRYDKDLARILQFIDRVSASDEDLLNIEYFQAKKLQKASS ELEKLESSKLDIELAISLSQILKSDQSTHNGIEQFTGLHVLCKYYKQRQARDLSL YIMPEYDCKLHYKWTGRFDDDNQLLTYCNHKPRQKRYQLNDLAGVLQVSPNF LKDKIGSDDLFSKWLVEIRIGFKCAREDLSKIQKDNRLLNHKINATTNGKCEK EIFNLICKIEGSDKKGGNYKHLAYELGVLFGEPNEASKPEFDKIKKFNSIYSFAQI QQIAFAERKGNANTCAVCSDNNAHRMQQITKITEPNDKDKILSAKQRLPAIPTRI VDGAVKKMATILAKNIVDNWNQIKQVLSAHQFLFIIIESTNANFEPFADAVJGK SLKDRRKALERSIFKDPNNRIKEFAKGISASYGANKTGDGDFAKEELDIHPR SHKKYTLNDEANLICVTRGDNKGNIFCLRLADNYLKLQFETTDDLEIEKIA DTIWDANKKDFKFGNYRSFINLTPEQKAFRRHALFADNPIQAVIRAINNRRNFV NGTQRYFAEVLANNYLRKKEKLNTDKISFDYFGPIPTIGNRGLAEIRQLYEKVDSDI
QAYAKGDKPQASYSYLIDAMLFCAIADEHRNDGSIGLEIDKNYSLYPLDKNTGEVF
TKDIQIKITNDEFSDKLLVRKKAIFGNTHRQMTRDGIAEWLPLIHKLENEVPvK
GYTWK^SEEIKIFKGGKDYIQDLNVLVYCLKFVDKPISIDIQISTLEELRLMTNNIAA
TAEYYYINLTQKLMHEYYIENCYNTALGYKKYSKEMEFLRSLAYRSERVKIKSIDDVK
QVLDKDSIIAGHGKJLPPFXEWQRLYREWQNTHQTKDDYEFKLSSFNVKSITKLHKKVR
KDFSPLISTNEKFLVRKRTWDNRFYQILNDSDSRADGTKFIPPADFSKNEIVEAID
SFTSKNIFWPKNIELQKVDNKNIAIDTSKWFETPSDLRDIGIAIQYQIDNNSRPK
VRVKLDYVDDDSKINYMHNLSLLKSRYPDVKVILEIQKSTIIIEFSEGPNFKTIKEMLG
MKLAGIYNETSNN

[0055] ZP_05061364.1 CRISPR-associated large protein (provisional), putative
[gamma proteobacterium HTCC5015J]
MTKNYISPIAIDLGKFTGVALYQLEGADCTEVAKGLLVDGRGNTWSQREGGRG
KRHQVRGYJKIRMAKRLWLILDSEYGIKREEVTEPLKFINGLLNRGGTYISEEV
DEESMm^*SPLFPSEMMPDYFNSSAPLLEQLAKLLSDKNNLVRFAEKGKPSNKNEFK
KLDTADGKYKDEKKELESEAWGMLIASENVLKSTVDGHKRSYELANIKEDKSN
EELKQISSKEIDGFYNLVGHLSNFLQRLRLKRYNDPMSGYWYDEKRLFYFYQ
WVQGWHKTGGTDEAEKMKMLTKGPPLLKLTLKSLADLTIPYEDQNNRPPKCSQV
VLSDEKLTMYHPKWEWVQVLKQNDNAYLNENVTLANALHRIVERSRSIDPYQ
LRLISITDAEKRNPLAEEKIKLSLPSGEVDFELLLVKNIVDETEKEARLEGFETENK
LFKCGKTPPRKEKКKSTLLSAVLGKNLSDDEQSSFIEEFKSGTPKIERRNVRGWCRLASQ
QVKTYGVLKEYGLQQHLKLEAGKLLDDKPLLALYKNSGLIASKIGEALNIEP
DEVSPJ^*ASPHSLAQIFNIIEGVDAGFNTCRACTYENIWMQEEKVESLTLQNLSEIH
GERKVPLKSAMCTRLSADSTRPFDQGMASIIEfllApvKIAQHKIAQINDVPEKSFIDIPIII
ESNQFSFATAELEIIKRGSGAKLAIAKALKGEKSKAGWVSKTERIKTSSEGICPETYGAP
LGGSGEIDHIIPSRTGTRKKTWFNSANLICSSKGNHDGKNRYVIEQLNDKLLKK
QFSTSDVLNIKKIKIITTIQRFTEGEGKLRSELSREDQKAFRHALFVPEKLSVESTSLL
AVKNITRVNGTQAWLAKKIALAEHLDKQGRDYTLSAHQIDPSVSDKRKLMLASAEPIWAKKDPQPAASHVVDACVTFLEALEQPHTASRLKTSTSFETGWSALPDLIK
VDALDRRKYRNYGTSFLKDGIIYAEKFLPILDENGNLAMYDIGDNSLKKAGADV
VFESLSPFLFEGGEVQASLSDWQERIDGRLYMSIDKVKAFDLYGKEVKEIDIA
ELLNSHTQXRTELRAKFSDDSKKMKMTLDAIRSLKLTLYTVNIEGKRKEEKGCGGFTIGIPAKSAWENLLDEPLLEYWTMKMPPQEIWEKVYRFttPRMPNQAHRRKVRKDFS
LPVVDSVGGFRVKPvKTPNGYNYQLAIDGYSAVGFKKEGD A /... VGPRVIFNYI VGG A A S SLKEIF S E AGKERS
Y P 122507.1 hypothetical protein lpp0160 [Legionella pneumophila str. Paris]

MESSQILSPIGIDLGGKFTGVCLSHLEAFELPNHANTKYSVILIDHDHNNFQLSQAQRRA
TRFmVRNKKRNQFVJKIVALQLFQHILSRDLNAKEETALCHLYNRRGYTYVDIDLE
YIKOETTINLLEKPLESEHHM^IDWFLQKMQSFREXRILVSKVEEKKDDELKNAVK
MKNFITGFENKSEVGRHRKVKYMKSITDKNQLSIIKKSIPSCLSNLGLGHNLSL
QWKNLHRY1AKNPQFDEQTFGNEFLRMLKM^RHLKGSQESLAVRNLQIQLEQSD
YISILEKTPEITIPPEARTOTGMEKDSLQLNEKLLNLYPNWRNLIPGIDHFPLE
KDELTEKLP)RKRISPSKQDEKRDSTYPEIRQYLDLNKIDIFKIKKQLSFLGQGGQLQP
AM.IETQKEMETHFNSLVSVLQIASAYNKERDAAQGIWFDNAFLSCELSN1NPPRK
QKILPLLYGAILSEDFDsINKDKWAKKFIFWNTKIGRTSLSKCKCIEEARKNSGNAF
KIDYEEALNHPESSNKLAIKIQTIPDIQIAIQSHLGHNDSQALIYHNPSFSLQYTILE
TKJRDFGHKNCVAVCTENYRSQKTEIDEPIESYSARPLADSVRFPDQGVLARMMQRLA
YEIAMAKWQIKHIPDNSSLLIPYLEQNREFEIESFKIKGSSSKTLEQIAEIQKNQR
EEKJQWmASMCMPYKGASIGGQQGEIDfIYPRSSKKHFGVIFNSEVNLICYYCSSQGR
EKJ<eEHYLPHELSPYLYKHPQFGTDNVSDIKFISQVANIKKYISFHLTLPEQQKAAR
HALFLDYDEAFKTITKLMSQQKARVNGTQKFLGKIMEFLSTLDKQLQLEFSI
KQITAEVHDLRELSLKQEPKLVRKRSQFSFPSHADALTMSIGKKEEPQFSQELDNS
WFnNiHLMPDEYHLNPVRSKEYNKPNISSTPLKFDSLYAEFIPVWVKGDAPFSE
KJDLFEIKPSNKEKFLTLKTYSTKSNPGEISLQALEAASKAKWLYFPINKTALALELHY
FHKEIVTPDDTTVCHFrNLSRYTYTKEISITVKLKEPMVLSVKFESSKKNVLGSFKHT
IALPATKDWERLF>^NFLA}

NP_907747.1 hypothetical protein WS1613 [Wolinella succinogenes DSM 1740]
MLVSPSVDLLGGKNTGFSSFTDSLDNSQGTIVYDESFVLSQVGRSKRHSKRNLRN
KLVKRLFLILILQeiiHGLSIDVLPDEIRGLFNRKGYTYAGFELDEKKDAESDTLKEF
LSEKLQSISRDSDVEDFLINQIANSASEFSDKYKKGFEAFAVSAIHSNKKLELKDLEKS
EGYENAKEILLSGLRVTKEILDEFDQFNQGLPLRKYFEELGEYATNEIKVGSFDFS
NSLKLTDMTKilGNISNYQLKELR'RFM)KEMEGDIWIPKNLHKITERFVSRWHPK
hTOADRQRRAELMKDLKSKEIMELLTTTEPVMTIPPPYDDMNNRAGAVKCQTLRLNEEY
LDKHPNWRDIakLHNKHGFNDLADSTVKGYEDSTLLHRLDLTSEIDYELRGK
KPNELLVKTLGQSDARNLGYFGAQNYELRQVKVRAGIWPVKNKDDSLNLSDNSN
MLKRCNHAPPKKNQIHNLVAGILGVKLEAKAFAEFKEKLSAWAKVGNNKLASYCK
N1EELRTHGNTFKIEDLRLKDPAESLSKEAKAKRLTDVILNEWSQKIANFFDDID
KHRQFRNNLSMAQLHVTIDPTPSGFSSTCRCTAESNRFSETAFYNDTFEGHKK
TATCQRPLADTQRPSGKIQEYIDKLGYELAKIKAKELEGMEAEEKVPIILEQNAFEY
EESLRKSKTGSJRVINSKIDDGKGLAKAKENAEPLKDLDKRAIKAFSSGCPYC
DTSIGCVGACDILPVEYQDFKLQDPPNADVPVACLACLWPSANPELTKGLSTNQ
KIEKMIKSGDYGQJKMRVEFVGSISGIEAGERYKIVPVIQEGGYIGPATAVKGYE
LKNCKVVTStK0IAKLEJKQDQLSLENQYIKISKQNKTSISELRWNNMVKNL
VERDKEIVGLLLEFIVENCRRYT KKVDVFAPKYIHEGTKPYIDDWRFDREWRYLQ
ENQNTSSKDRFVIDKSSLNEYYQPDKNEYKLDVTQPIWDDFCRWYFILDYKTAN
DKSIRIKARKTVSLLAESGVQGKVRASKRGIPTGYAYALQALPMDNVIAGYANILL
EANKTSLSLVPSGSIIEKQDLKDLVVIKKTVDRLAIDNNSFFNADFDTHGIRLIDENT
TSVKVGNFPISAIDSKSARKMIFRALFEKEKGI0CKKTSIFKESGPVQDYLVKFLKKI
VKIQLRTDGSINSNVRVKAADFRLSFRESHIQKLLK

[0059] ADX75954.1 CRISPR-associated protein, Csnl family [Staphylococcus pseudintermedius Ed99]

MGKPYILSDLDIGTSVGYACMDKGFNVLKHYDKDALKVYLDFGALTALKQERRQFRT
SRRRKNRRIKJRLQQLLELPQLVQPNNFYQFQRQFAWKNDNDFKNSLSEVSLFL
GYESKPYKTIYHQLALLKDKEFKDPEIMALYHLVKYGRHFLDHLKELNTOND
NMHDFVLEIETYENLNMKRLYDEETKIVIEILKDNETMKNKDRAKRVKNEMKLE
QFSIIMLLGLKFNEGKLFNHADNAAELKGANQSHTFADNYETMTPFLTVEQSEFIERA
NIYLSLTQILGKKSAMSKVAAYDKFRNELKQKIVDAKSTRTQFFKIFVS
SKSKLQYDATPQDQTSSLCLDFQYLRPKKQYSLLISKELKIKIPQDSELYFIEAENT
LLKVLNNTDNSAPIPMQLNYEATLIRNQQKHYAEITEKEMIEKVLISLPFRIPYYVGPL
VM)

HTASKGFWMEPVKSNSEIKPWPESPNFDVEVDRSKSATSQFIRRMTOKCYSL

DVLPC
KNSLLYQEMEVLNENLNAQTQIRLTQDPPKRNKRYRMMPQIKLFAVEHFIKKYKTVSHSKF

-33-
LEIMLNSNHRENFMNHGEKL SIFGTQDDKKF A SKL SSY QDMT KIFGDIEGKRA QIEII QWITIFEDKKILVQLKE CYPELTS KIQINQLKLKNYS GWRL SEKL LTHAYQGHSIE LLRHS DENFMEILTNDVGYQFQNIKEENQVQSNKIQHQDIANLTSPAL KGIWI STIK LVRELTSIFGEPEKII MEFATEDQQQGKKQKSRKQLWDDNIKKNKLKSVDEYKIIIDV ANKLNEQLQQEKLWL YLSQNGKCMYSQISIDLDALLSPATKHYEVDIFIFPRSFIFIK DDSIDNKLVLKMNQTGDQVPQFIQQQPYERIAYWKLNLKAGLISDSLKLFLIKLMKP EFTAMDKEGIFIRQL VETRQISVHYRDFLKEEYPNTKVIPM AKMV SEFRRKF DIPIK IQWITIFEDKKILVQ KLIYPQMYTWAIAKSVNKKGKE MEYQMIDHYVFDFYKF QNGNEKELALYLAQRENKDEVDLADQIVYSLNKGDLLYNHPCYFVSR KEVNAKQ FELTVEQQLSLYNMN KETNVKLLIEYDIFIAEKVINEYHYHKLNSKLEK ERTFTFS ESNQTHEDFIKAlDELFKVTASATRS DKGSRKNSMTHRAFLGKGKDVKIAYTSISG LT KTPKLKF KLAESREL N [0060] ZP_10206685.1 CRISPR-associated protein, Csn family [Planococcus antarcticus DSM 14505]

MKNYTI GDLGIVASVGWVCIDEN YKLNY NTRHA FG VHEFAESAAS AAGRLKRGM RRRYNNR RKRLQLQLQLSDYITDSQFSKTDSQHFHW NNEFENRSLTEWL SSLRISS RKYPTIYHLRS DLIESNKKMDLRLVYLALHNLKVYRGHFLQEGN WSEAASAE GMDD QLELVT AEEENLPLDLSESQWKA AETTTLNRLNKTQSKEL TAMFGKEYEPF CKLVAGLGVSLHQLFPSSEQALAYKETKTVQSLNENVEVMELLEESALLEAVQ PFYQVQVLYELLKGETYVAKVA SFKQYQKDMASKLNLDTKTFGEK VYRSFISD KNSQREYQKSHKVEVLCKIDQFNKEAKFAETFYKDLKLLEDSKTSIGTDEDEM LRIIAKADSNQFKQKGIQNAAI PHQNSLYEA EKILRNQCAHYPFITTE WIEKVQIL AFRIPPYYIGPLV KD T TQP SPSW VER KGDAPITPWFQ IDKAAASEAFISMRKCTC YLKGQEVLPKSSLLYERFVLSENQIQLR T G TGAESDFR FL SYEMKCI WDNVFQK YKTVSTK RLLQ ELKSPY ADELYDEHTGEKIEVFTGQK ENSFATSLSGYISMKSI LGA VVDDNP AMTEELYWI A VFEDREILHLKI QEKYSITDVRQQLALVKLPGWGRFSRL LIDGLPLDEQQQS VLDHEQYSSVFME VLNKNGFLEK QKMNQHQV DGT KIRY E D I EELA GSPAL KRI GWRSVKE VE ELVSI FGE PAN I VE L VAREDEK KRTK SR KD QWE ELT KTTL KND PDLKS FGEIKQSKQGDFQRNQF QRF LW YVTQ QGK CL YTG KALDIQNL SM YEVDmL PQNFVK DDSL DNLALVMP EAN Q KRNQGVQNK MPLEIE AN Q Q Y AMR TL
WERLHELKLISSGKLPvLXKPSFDEVDKFIARQVLVETRQIIKHVRDLLDERPSKSDI
HLVKAGIVSKFPvRFSEIPKIRDYNKHHAMDALFAALIQSILGKYGKNFALFDSLKK
DRQKQWRVSVKGSNKaffLFKNFGLNLRLQSPVGTGEVSGVEYMHVYFELPWQTTRK
MTQTGDGMFYKESSFSPKVQAKVMVPSKTEKVFHYDVKNHSCLVEFTFMKKEKEV
QETKFIDLKVEHHQFLKEPESQLAKFLAEKETNSPPIHARIIRTIPKYQKIWEHFPYYFYI
STRELHARQFESYELMEVKQLSERSSEVEELKIVFGLLIDQMNDNYPIYTKSSIQD
RVQKFVDTQLYDFKSFEIGFELEKKAANAAQRSDTGSRSRISSKPKPEEVAIGYESIT
GLKYRKRPRSVGVTRK

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

MTKFKNKYSIGLIDGVSVGGYAVVETEDDYRVPAPFIFKVLGNTEKEKJJKNLJGSTTFFS
AQPAKGRVFRVNRJRIRDRNNHRITYLRLDFIQKEIEKVDKNFYRRLDESFRVLGDKSEN
DLQIKQPPFGTKELTAHYHKYPT1YHLKRLKHADADKNSPVA1REVYMAISHILKY
RGHFLLDKINPNMNMQSNWIFIESCQEVDFKEDSESMADIFSSESQKSEQKKI
LPYFQQELLLKDKSIFQKQLLLFFGLKTDFKDFCEELEEPELNFSKENYDENLENFLGS
LEDIFSDVFKAKLVLRTDILLHJLTYTGGTHARPSATMVREYEHKDLQRFFICK
IKQNLSEQYLDIFGRKTQNGFDVFDKETKGYVGTYTNKMLTNSPQKKTQIQNFYDI
YISGKITGIEGEAYFLNLKSIDGTFLRKLRTSDNGAIQPQIYAELEKIKERQKDYFPOLL
ENKDKLLSLITFKIPPYVGPLAGKSNPSJ^AWIKRATSSDILDDNEKDTRNGKRPWNY
QKLINMDDTDAFITNLIQGTOIILLNEKVLPRKSLYIEEVMLQNELTRVKYKDYGKA
HFFDSELQRMQINGLFKNNSKRVNAKSLIKYSLDHKDLNAIEIVSGVEKGKSFNSLKL
TYNDLKTIFSEEFDSEIYQKELEEnKVTIFDDKKSINLYKTFFGHELIDDEEKINQL
SKLRYSGWGRYSAKLDDIRDDETFGNNLQFLLNDEENRNLTLDNLSFEPKIKDIDI
QSKSTIEDDIFDEIKKLGSPAISKRLNSIKIVDELVQIGYPPHNIVEMARENMTTE
GQKKAKTRKTLKESALKMENSLLENGKVPVSDEQILQSEQKLYLYLQNGKDMYTLID
KTGSPAPYLQDLQDYEVDDHIYPFSLPSIDNKLTHRENNQQKLNNIPDKETVAN
MKPFWALKLYNAKLISQTKYQRLTTSERTPDGVLTESMKAGFIERQVLVETRQIIKHVA
RILDNRFSTDDKTTILKSQLTINFRNFhAIKIRELNDYHAFIDAYLAVVVGQTTLKVPY
KLAPELJYGHHAOTNRmENKATLRFKLYSMRMRFFh^DSKVSKDIWDCNRDLPIK
DVIIYNSQINFVKRTMIKKGAFYQNVPNPGFKNQALANNYRPLKTKALCDLTSIYG
YGPMNSALSHIIAAERFNEKXGKIKETVKEFHDIFnDYEKFNNPQFLDNDTSENGFLKK
NNiiÌ³4VLFYWPXYLMKIDGTRMLFESKSNLHKAQFTKLTQNLFFHMKRLL

-35-
TKSW:MDLKSASKAIKESQOTILKHKEEDMSQNLSAFSQKMKLGNNTSLKNIKGYNE
RKIKEIDIDETIKFYDNFIKMSFVKSGAPKIDNDFDKCTVARMRKPDKKLLN
ATLHIQSITGELYTRIDLKSGLGED

[0062] AAK33936.1 conserved hypothetical protein [Streptococcus pyogenes M1 GAS]
MDKKYSIGLDGTSVGAVIDTEYKVPSKFKVLGNTDHSKKNLIGALLFDSE
TAEATPvLJKITARRYTRRKNPJCQLQIFSNSMAKVSDFSHHRLIESFLVEEDKHE
RHPIFGMVDEVAYFiEKYPIHLRLKVDSTDKADLRILYLAHMKFKGRHFILIEG
DNLPDNDSDVDFQILFTQTYNQLFENINASGVDAKIALSRLKSRLRLNIALQPL
GEKKNFLGZNLAALSLGLTPSKNSFNLADAKLQSKDTYDDLDNLLAQIGDYA
DLMFRAKLNSDAILLSDILRVTNTEITKAPLSAMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQFSKNGANYIDGASQQEYYKYFKIPILEMKMDTGELVLKLNREDDLRKKQR
TFDNSPIHQHQLGELHAILRRQEDFYFPLKDNREKIEKLTFRIPYYGVLARGNSRFA
WMTRKSETITTPWWEVEVDKGASAGQSFIERMTNFDSKLNPNEKLPKHSLLEYFTV
YNETLTKYVTGMRXPAFLSEQDKKAIYDLLFKNRKTQVEQKLFKIDFYKIECFD
SVEISGVEDRFNASLGTYHDLLIKDKDFLDEENLEDIVLTLTLFDREMIEERL
KTYAm FDDBKMQLKRRRYTGWGRSLRKLINGIRDQGKTILDFLKSDFGANRN
FMQLIHDDSTLFKEDIQKAAQVSGQGDLSHEHIANLSPA1KSGILQTVKVVDL
VEMGRHKPPEMIEMARENQTTKQKQKNSRERMKREEGEIKESQILKEHPVENTQL
QNEKLYLYLQNGRDVMQELDINRLSDYDVFIVPQSLKDDSIDNKVTLRSDE
NRGMKDSVPSEEEVKKMKNYWRLNWAKLITRQFNDLTKAERGLSLELDKAFIK
RQLVETRQITKHVQAQLDSRMNTKYE>n3KLIREVKVTLLSKSLVDSDKFQFYKV
REINNYYFIAYHAYLNAVGTLALIKYKPLESEFVYGYDVYDKRIAMKESEQIEGK
ATAKYFFSYMMPNFTPKTEITLANEIRKRPLIETNGETGIVWDGKRDFAVTQVL
SM PQVMVKTEVQTTGFSKESILPRKNSDKLARIKWDPPKYGPDFISPTVAYSVLV
VAKVEKGSKLLKSVELLGIMITERSSEFKNPDIFLEAKYKEVKKDDKILKPASFLE
LENBRKRMLASAGELQKGNELALPSKVYVNFLYASHYELKKGSPEDNQKQLFVEQ
HKJT^LDEIIIEQISESKRVLADANLDDKVLSSYNKHDKPREQAEMHFLTLTNLG
APAFFYFTTDITYKRTSTKEVLDATLHIQSITGELYTRIDLKSGLGD

-36-
YP_820832.1 CRISPR-system-like protein [Streptococcus thermophilus LMD-9]

MTKPYSIGLDGTSVGWAVTDDWVPSKMKVLGNTSKXYIKKNLLGVLFDSGI TAEGRRKRTARRYYYTRRNRILYQIEFSTEMATLDDAFFQRLDNSLFLPPPvDS KYPIFG$VEEKAHYDEFTIYHLPvKYLADSTKXADRLVLYALAMIKYRCHFLIE GEFNSKNNDIQKNFQDFLDTYNAIFESDLSENKQLEEIVDKISLKEKDRILKLP GEKNSGIFSEFLKLIVGNQADFRKCFNDEKASLHFSKESYDDELETLLGYIGDDYSDF VFLKAKKLYDAIALLSFGLTVDNETEAPLSSAMIKRYNEHKDLALLKEYIRMALKTYN EVNFKDDTNGYAGYIDGKTNQEDFYVVLKKLAEFEGADYFLEKIDREDFLRKQ RTFDNGSIPYQIHLMARAILDKQAKFYFLAKNERIETKLFTRIPYYVGPLARGNSDF AWSIRKRNEKITPWNFEDVIDKESSEAAFINRMTSFDELPLEEKLPSLLEYETFV YNELTKVFIAESMRDYQFDSKQKDIVRLYFKDKRVTVDKIDIELYHAIYGIDG IELKIEKQFNSLSLTYHDLLNINDKEFLLDSSNEAIIEIHTILTFDEMIRKQLSKF EMFDKSVLKKLSRRHYTGWGLKLSALKINGINDKSGNTILDLYLDDGISNRFNMQLI HDDLALFKKKIQKAQIIGDEDKG1KEVKSLPGSPAIIKGILOTSIKVEDELVVMGGRKPEsrVVEVHARENQYTNQGKSNSQQRKLRELSKLKELSKILKENIPAKLSKIDNA LQNDRLYLYLQNGKIDMYTGGDDLDDITISMLYIDHIQAPLKDNDINDNLVSSAS NRGKSSDDVPSLEVKKKRTFWYQLKSLISQRKFDNLTKAERGGLSPEDKAGFQQR QTVETRQITKHVARLLEKFNNKKDNNRATTNVKIIKSTLJVSQFIRKDFLKVR EINDFHHAHADAYLNAVVASALLKLYKPCPEFVGYDYPKNGSFREKSAEKEVYYF SMJVINHFKKSISLADGRVIERPLIEX^ETGESVW^ KESDLATVRRVLSYPOQNVMVKK VEEQNHGDLRGGPKGFLFA^SSKPNXNLENSLGVAGKEYLDPKYGAGISNSFT VLVKGTIEKAKKIKTWLEFQGISLDRDNTNRKDNLFLKEGYKIDLIEIPLKSLF ELSDGSpvRMLASILSTONKRGEIHKGNQIFLSQKFVK^LYHAKRISENTINENRHKYVE NHKKEFEEFYYEILEFNYVAGAKNGKLLNSAFSWSQWNHSEIDCSSFIPTGSERK GLEFETSRGSAADFEFLGVKIPRYDTPSSLLKDATLHIQSVTGLYETRIIDLAKLGE

NP_721764.1 hypothetical protein SMU_1405c [Streptococcus mutans UA159]

MKKPYSIGLDGTSVGWAVTDDYKVPACKMKVLGNTDSIIEKNNLGLALLFDSG NTAEEDRRLKRTARRYYYTRRRNRLYQIEFSEMKGKVDSSFHRLEDLSFLVTEDKRG ERHPIFGNNLEEYKHYENFPTIYHLRQYLANPKEVDIRLVLVYALASHIIKFRGHIFLIEG KFDTRNNDVQRLFQFLAVYDNTFENSSLLEQNVQVEEILTDKISKAKDRVLKLF

-37-
PNEKSNGPvFAELKLIVGNQADFHHFELEEAKPLQFSDKTYYEELEEVLILLAQIGDN
AELFLSALKLYDSILLSGILTVTDVGTKAFLASMSMIQRNYEHQMDLALQLQFIRQKL
DKYEVEFSDVSVDGAGYIDGKTNQAEFYKVLYKGLLLNKIESGSYFLDKIEREDFLRK
QRTFDNGSPHQIHLMERAIIRQAEFYPFALDNSQDREKLLTFRIPYYYGPIARGKS
DFAWLSRSDAKITWPNFDEIVKKESSAEAFINRMTWDLYPQKVLPKJISLLYEK
FTVYNEFLKAYTEPQGKTAFFA-MMQEFIDFGVFKYRKVDKLMDFLKEFEDE
FRIVDLTVGLDENKVFNASYGTHYDLCKIDLKDLDNSKNEKILEDIVLTTLTLDRE
MIRKRELWSDLITKEQVKKLERRHYTGWGRSAELIHGINRKNESROTKITLYDIDG
NSNRNFMQLINDALLSFKEEIKAQVIGETDNLNQVVSIDAGSPAIIKGQLQLVKEID
LVKIMGHQPENIVVEMARENQFTNQRNSQRLGCLKLTSIKEFSGQILKEHPVENS
QLQNDRLFLQLYQGRDYMTEGELDDLYQYYDIIDHIPAQFMDNIDNVRSLSSKE
NRGKSDDVPSKDKVRMKSYSWKLSSAKLITQRKFDNLTKAERGGLTDDDKAGFIK
RQLVETRQITHXVARLDERFNTETDENNKIRQKVITKLKNVLNSFRKEFELYKVR
E1NDYHHAHDAYLNAVIGKALLGYPQLEPEFYGYDYPFHGKENGKATAKKFFY
MMNFFKDDVDRDEKNGEnWKKDEFISNIKKVLSPYQVVMKVEEQTPGFSKESIL
PKGNSDILPRKTKKFYWTDTKKYGGFDSPIVAYSILVIAKIEKSKKLLKTVALKLG
VTIMEKMTFERDPVAFLERKGYRNVQEEEIKLPSLYKLENGKRLASARELQK
GNEIVLPNHLGTVLYHAKMKHVDEPKHLKYVDKHKDEKFELLDVSNSFKKTYLTA
EG*EIKIKLEYAQNNGEDLKLKASSFNLTLTFATAGAPATKFFKDNIDRKRSTSETI
LNATLHQSITGTYETRIDLNLKLGGD

[0065] YP_004373648.1 CRISPR-associated protein, Csnl family [Coriobacterium
glomerans PW2]
MKLRGBEDDDYSIGLDMTGSVVGAIVTDERGTLAHFKRKTWSRGLFREAOQTAAVAR
MRGPQRRRRWRRWRWLDDLQKLFEQQMQADPDFFIRLRSRLRDDRRAEHHADDY
RWPFLFNDCKTHERDYQRFPIYHYVRSWLMETDEAIDRLYLLNHIKCHRGNFLR
EGQSLSAKSRDEALNHLRETDLVWSSERGFCISDNGNSLALMTHPDSLPSDRR
KIAFLDVKSDAAAADKKLGIALAGAVILKTEFKNTGDPCEDSISLYNSDEAVDA
VRSACPDDCAELFDRLCVEVYSAYVLQGLSLYAPQFTISANMVKRYRGEDLLK
KLKVIIYAPDQYRMFFSGATYPGTGIYDAQQARYGTKYNLPGPKSEKYPSEMDSQY
DDFRKAKEKLFAKTDARADERYRMMMDRDQFKQFFLRLKRTSDNISYHQLHLLELKAI
VENQGRPYFYFLKIDAKLQVLVSFRIPYVVGLPLSTRNARDQHGENRFAWSERKPG
MQDEPIFVPNWESIISRKSLEKIKLRMTGMCTYLLQQEPVLKSSLLYEEFCVLNELN

-38-
GAHWSIDGDDEHRFDAADREGIIEELFRRKRTVSYGDVAGWMEPvERNQIGAHVC
QGEKGFESKLGSYIFFCDKFVKVPvLEQSDYMPIEPJILWNTLFEDPvKILSQPvLKEEG
SPvLSAEQIKTICKXRFTGWRLSEKFLTGTIVQVDVESVSIIvDNVLREGCPVSGBKGR
AMVMMEILRDEELGFQKKVDDFNRAFFAENAQALGVNELPGSPAVERSSLNQSI
VEIASIAGKAPAMFIEVTREDPKKKGRJRTKRRYNDLKDALAEFKKEDPELWRELCTA
PNDMDEPvLSLYFMQRGKCLYSQPvAIDIHQSNAGIYEVDHPRTYKDDSLNKLAL
VYREENQJPCTDMLLIDPEIRRRMSGYWRMLHEAKLIGDKFRNLRSRIDDKALKG
FIAQQLVETQMVKLVRSLLERYPETNISVKASISHDLRTAAELVCREANDFHHA
HDAFLACRVLFIQKRRPCVYENPIGLSQRVVRNYVRQQADIFKRCTRIPGSGFIVNS
FMTSGFDKETGEIFKDDWDAAEAEGIRRLNFRQCFISRMPFEDHGVFWDATIYSPR
AKKTAALPLKQGLNPSYGFSGFEQFAYFFYKARNPRKEQLTEFAQVQPVRLSAQIR
QDENALERYARELAQDFLEFIRERSKILKNQLEIDGDRLCITGKEEVNACELAFA
QDEMVRIRMLVEKPVSRECVISLFNRIILGHDQASRRLSQKLALLSEASEADN
VQRNVVLGLIAIFNGSTOMV^SDIGGSKFAGNVRIKYKELASPVKNYHLIDQSV
GMFERTKIGL

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

MTKKEQPYMGLDYGSSGVAWTO^TOLLMK^ NLWGVRFLFEEAQTAKEETRLN
RSTRRRYRRKNPNWNLNIEFSEEAKDPSFLIRLQNSWVSDDPDRDKRDYNLFDI
GYPYDKKEYYREFPTIFHLKELNOKADIRLIYLALFINILKRYGNFTYEHQKFNSN
LNNNLKELIENQLKYDISPDDCWNHSIDILIGRNATQKSSNLIKDF LTDKET
KKLKEVINILGNAVHLLTFTKSLTDKEEKLNFSGKSKDILLLDLDSDLDQFTVL
DAAMUYSTILNENGESYFSMAKVNQYENHAIDLCLKLRDMMWHTTKNEEAVEQSR
QAYDDYFiv^KYGTKELLYSIXKFLKVALPTNLAEAVEEISKLTLYLKVPRSENENGV
VPYQLNKJEMKIDNQSQYYYPFLKENKEKLSSILSFRIPYYGPQLQSAEKNFAMGE
RKSNGHARPWNF EIVDREKSSNKFRIRMVTDSYLVGEVLPKNSLIYQRYEVLNE
LNMRITENLKTNPIGSPXVTETKVQMNELFKKYKKTETVKKLTWLLAQGYKPNILI
GLSQKDENGSTLTYLMKIKFGSFMEDNKYNQIEEELWTELIFEDEKVQKE
KSYTYPDQIKKISNMYKKGWRLSKKILMDITETTNPQLQLQSNYSILDLMWATNN
NFISIMSNDKVKWENHNLKNQDMSLVDNHVASPLKRGITQSKIVQIEIVK
FMGHAPKiiFIIETVPvETKKEITTSREKnMPvLQSKLPKANDKFPQLPvELVPNKIQ
EELKKHK>TOLSSERIMLYFLQNGKSLYSESLNUNKLSDYQVDHLPRTYPDSDLN-39-
KALVLAK^NQRKADDLLLNS^IDRNLERWTYIVO.NNNMIGLKKFKNLTRRVITDK
DKLGFHQLVQTSQMVKGVANILDNMKNGQNQTCTCQAARANLSTAFRKLGSQDND
YHFKIIPELVK>mNV>roFHHAQDAYLASFLGTYRLRRFPTNEMMLMNGEYNKFYGGV
KVELYSKKKLPLSRKNGFIISPLGVNGTGYDRTNGEIWINVGFDRKILKIFNYHQN
VTPICTEITGQFYDQTIYSPKRPKQILIAQKKMDPNIYGGFSGDNKSSITVKIDN
NKJKPVUPIPLNLKDCTLQNWLEEEm^KHKSIQIQKNNVPIQIIYSKKVGLSLN
SDREVANRQLLPILPEHASSLLRLLQIPDELDQILAFYDKNILVEIQLETIKMKKFYPF
YKGERIEFLIANENFNQATTSEKVNSLEELITLLHANSTSAHLFINNIEKKAIFGRKTHG
LTLNNTDFIQYQSVTGLEYTRIIE

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

IVDYCIGLDLGTGVGWAVDMNHLMKRNKHLWGSRLFSNAETAANRRASRSI
RRrynKJRERIRLLRAILQDMVLEKDPTFIRLHEATSFLDEEDKAKYLGTDYKDYN
LFIDEDFNDTYYHKYPIYHLRKLACESTEKADPRIYLALHHIVKRYGFLYEGQK
FNMDASMEDKLSDFITQFISTSNFMPYEDDEKKNEILEILKPLSKKAKAVDEVMTLIA
PEKDYKSAKELVTGIAGNKMVTKMCEPKIQQDEISKLFSDNSYDDQFSEVEK
DLGVEYVEFVADLNHVYSWVELQTIMGATHTDNASIEAMVSVYNKHHDDLLKLKD
CIKNNVPNKFDMFRNDESEKSKGYYNYINRPSKAPVDEFYKYKKCIEKVDTPEAK
QILDIELENFLKQNSRTNGSVPYMQMLDEMIKIIDNQAEYYYPILEKREQLLLSTLF
RIPYWGPLNETSEHAHWIKRELGENQRIIPWNYQDIVDVDAEGFIRKMRSYCTY
FPDEEVLPKNSLIVSKYEYVNEL>¾IRVDDKLLLEVDDKNDIYVNLFMKKNKTVEK
KNNWLVNNQCCSKDAEIKGFKENQFSTSLTPWDNFTIFGKIDQSFNDLIEYIIDYLT
FEDKIKKMKRKLKLYALPDDKVQILKLKYKDYWSRLSKKLLDGIVADNRFGSSVTV
LDVLEMSRLNLMEINDDKDLGYAQMIEATSCPDGKFTYEEVERLAGSPALKRGIW
QSLQIVEEITKVMKCRPKYIYIEREEASEEAKERTESIKKLENYKDKLDQETKKEEKS
VLEELKGFDNTKIKSSDLSFLYFTQLGKCMYSGKLDIDSLDKYCIDflIPVPSLSVDD
SFDNRLVLPSENQRKLDLVLVPDFIDRKMRFWKLLFDHELISPKKFYSLIKYTE
DREERFINRQLVETRQITKNVTQnEDHYSSLTKVAAAIRLNSHEFRVKNHIYKNRDIND
YHAHADYIIVALIGGFMRDYPNMMDSKAUSYEMKMKFRRK>TOQKRWKDGFI
NSIVnWPIVEVDGKLJWNPDLNIEIKXCFYKDCYCTTLQDKQSGLFNLTVLSNDAII
ADKGVTKAVPVPNKRSVDHYKGFGSLQYTIIVAIEGQKKKKGKTELVKKISGVPL
iiLKAASINEKINYEEKEGLSDVPIJKOMPVNQIMIEMDGGEYLLTSPETYVNARQLVL

-40-
NEKQCALIADIYNAIYKQDYDNLDD1LMIQLYIELTNKMKVLYPAYRGIAEKFESMN
ENYVVISKEEAMIKQMLIVMHRGPQNGMVYDDFKISDWGRLKTKNHNLNMVFIS
QSPTGIYTKKYKL

[0068] YP_003171950.1 CRISPR-associated protein Csnl [Lactobacillus rhamnosus GG]
MTKLNQPYGIGLDSNSIGFAVVDANSHLLRLKGETAIGARLFREGQSAADRRGSR
TRRRLSRTWRWLSFLDFFAPHTIKIDPDFFLRQKYSEISP KDREFKYKERLFNDRTD
AEFYEDMPSYHLRHLHMTTHKADPREIFLAHIHILSGHRHLTGPAAKDFNRTDK
VLDLDIAPALTEAYAQVYDPDLETFDLAKADDFKAKLLDEQATPSDTKALVNLLS
DGEKEIYKKRQVLTETFANAITGLKTFNLALGTEVEDADSNWQFSMGQLDDKW
SNIETSMTDQGEIIEFQIELYRARLLNGIVFAGMSLQAKVADYQGHKEDLELKTY
LKKLNDHELAKTIRGKYDINGDDAKPLFREDVFKALTKETVAHPNVESEQLLNR
MGQANFMLKQRTKANGAIPI2L0QQLQRELDQIAANQSKYYDWAANPVEAHWRKMP
YQLDELLNFIHPIYWPGLIPKQQAEGSFENVFAWMVRKDPGNITPNFEMEKVARED
SANTFQIRMKTFTDYLIGEDVLKPSLLYKQYEVNLNLRINNECLTDQKQLLE
REVFERHSSVTIKQVADNLVAHGDFAAPRPEIRGLADEKRFSSLSTYHQLKEILHEAI
DDPTKLLDIEMTSTVFEDHTIFETKLAIEIEWLDPKINELSGIRGYRWGQFSRKL
DGLKLGNGHTVIQELMSNHNLMQILACETLKTMTELNQDKLKTDDEDVINDAY
TSPSNKKALRQVLRVVEDIKAHAANGQDPSWLFIETADGTTGAKRQSKIQTVY
ANAAQELDSAVRGELEKIDAKASFTDRLVLYFMQGRDIYTGAPLDNQSHYDI
DfiiLPQSLIKDDSNDNLVNLVANINREKNNVFASTLFAGMKATWRKWHEAGLISGR
KRRLNLRPDEIDKFAKGFVARGQLVETRQIIKTLAQIAAAQYPNTKJIAVAGLHQL
REELDFPKNpSVNYHHAFAFCLAARIGTLYLKRYKPLAPFFTYGEFAKVDVKKFR
EFOTIGALTHAKKMMIAKTGDI EIVWDERLDRILDRIYNFKRMLITHEVVYFETADLFK
QTIYAAKDSKERRGSKQILPKQGYQTQVYGGYTQESG SYN LA VRVAAEDDTATVQ
IKISAQNASKIASNLKSRKQMKQGILNIE YVKLKRAKKNKPSANSF KIPRFQMG
TLFQNAKYGFLMVNDS TYRRNYQELW SRENQKLLK LFSIYKETQMN HDALQV
YKAaDQVEKFPKYDINQFRKLSDAIERFEKLPI NT DGN KIK G TEL RQILG LQANG
TRSNNKLNGLIKTD LLQQVG SGIKLKD KT DQ IY QYPS GLFK RRI PL ADL

[0069] YP_003937986.1 CRISPR associated protein [Bifidobacterium bifidum S17]
MSRKNYYVDDAYAISLDIGNASVGS WAFTPYNLVR AGHELIGVRLFDPADTAESRR
MARTTRRRYSRRWRLLDALFDQALSEIDPSFLARRKYSWVHPDDENNADCWY

-41-
GSVLFD_SQ_DKRFYEKYPTIYHLRKLALMEDDSQHDIREIYLAIHHMVKYRGNFLVE
GTLESSNAFKEDELLBCLLGRITRYEMSEGEPONSDEQDDENKLVAPANGQLADALCA
TRGSRSRMVDNALEALSAVNDLSREQRAIVKAIAFALEGNXDLAKIVSKEFFSEN
KKILGIYFNKSYEEKCVQIVDSSGLLDEDEREFLDRMQGQYNIAJALKQLLGRSTSVS
DSCASKADYAHANWNLKILKLRTKEDKINENEYGLVWKGQIDSRQKVRGESA
ENMRRKANVFFKMIETSDLSETDKRNLIHDEEDKLFIQRSNVSNGVIPHQLQ
KQIYKKQGKYPPFDLAFEKDGGKQINKIEGLLTFRKYFVGGLVVPEDLQKSDNSEN
HWMVRXKKGETIPWFSFEVM
DIAASGRKFIERLVTDSYLLGEPTLPKSNLLYQEYE
VLNELNNVRLSVRTGNHNWNDKRRMLGREEEKTLCQLRFMKGQVTGTKRATENL
LREYGRTEYSGLSDELSKTSSLSTYGKMCRIFEGKYVNEHRMLKEIVELQTFEDK
ETLHLQRLQLEGISEDACLLYVTHYTGGWRLSRKLLTTKAGECKSDDFFAPKHSI
EIMRAEDRNLMEIITTQDLGFSWIEQENLGAENGSSMLMEVVDDLVRVSPKVKRGIIQ
SRLDDISKAVGKRPSRIFLELADDIPQSPGRTSRKSLQDLRYNANLGEKFKGIADEL
NACSDKDLQDDRLFLYYTQLKGMDYTEEGLDLRLSSAYLIDHIIQAVTQNDSDIN
RLVARAENARKTDSFTYPEQAPDIARMRNFQIILLDNGLSREVKFERLTRQNEFSEREK
ERFVQPSLSVLQTRQIMKNVATLMRQRYGNSAAVGLNAELTKEMHRYLGF
HNDYHQAQDALCVGAIAQQAFAANRGFFADGEVSDGAQNSYNQLRDLRYLGRYREKLS
AEDRQGRAFVGSMRSQDEQKRVNPRTEVWSEEDKYLKVMNYRMLV
QKVGDGFALYDETRYAATDPPKIQGIPFGDAKGQDSTLYGSGSAKAPVSLIESKG
KTRLNVNTMQEYSLLGDREPSDDLRXVLAKKSEYAKAMLRLTJVPVQMLIRYGG
LMVIKSAEGLNNAQQLWLPHYEEYCFFDLQGPGSKLEKDJDLKLLDSLGSQVCLY
PWRHFTEECLDLFTVAFDKLPEDEKKNCITGIVSALHADAKTANLSVGMGTGSWRR
MNNSGYTFSEDESEFIQPSPSLGFKEKRTVGEKJRKAKKEWSKYRTOTKRLTSLG
ASQP

[0070] EHN59352.1 CRISPR-associated protein [Oenococcus kitaharai DSM 17330]
MARDYSVGLDIGTSSSVGWAIDNXYHLIRAKSKSNLIGVRLFDASVTAEKRRGYRTTR
RRLSRRHWRRLRLNDIFAGPLTDGFEDNFLARLKYSWVHPQDNSAQAHFAAGLLFD
SKEQDKDFYRKYPTIYHLRLALMrTODQKHDRLREVYIAIHVLKVRYHFLIEGDV
KABAFAVTFADFADAIQRYAES>WSDE>n.LGKIDEKLLSAALTJHKGSQRAETEATAF
DILLQSKKQIQAILKSVGQANLMAIFGLDSSAIKDEQKNYKFSFDADIDEKIA
DSEALLSDTEFPLCDLKAADFGLTLMILLGDKTSAAMVRFRNEHQKDWEYIKS
fIIRNAKNAAGNGLYEKSKKFDPINAAYLALQSDNEDDRKKAKKIFQDEISSADIPDDV

-42-
KADFLKKIDDQFLPIQRTKNNTGHRNELEQIEKQGIYYPFYKLDKSYQENSHEL
>mTALINFRVPYYVGQLEEEQKIAADDGKMPDP'TNHWjvVRXSDNTITPWNLSQVV
DLDSGGGRRFIERLGTDTYLGECPTLPSNLVLQKFDVQLQELNNRVSGRRSLDIRAKQ
DAFEHLFKVQKTVESATNLKDFLVQAGYISEDQIQIEGLADVNGKFNANALTNYLV
SVLGRFVENPSNEELLIEETELTQVFEDKKVLRQQLDLQGDHLREKLSKHYT
GWGRISKKLLTTKIVQADKIDNQTFDVPQRMNQSIITDLNYTLMNMEIINNAEDDF
GVRAWIDQNTTDDDEQDVYSLIDELAGPKEIIRGIVQSFRLDITDKAVGYPKR
YLEFARKQESHLTSRQQLSTTLNKAGLSLETQVSQYDAALQDRLYLIFQ
QGKDMSGKLEDNLNPNIDKIPQ AYTKDNSLDRNYVL VSjTNRRKSDSSNYLP
ALIDKMRPSWLSKQGLSSKHKFANLTRTRDFDDMEKIERLAVSLVEFRQIKNYVS
LIDSHFGQTSNVAIRSSLTADMRRYVIDPKNRINDYHYHAFDALLFSTVQYFENS
GLMKGQLDSLADGQNYRIKEWHYARLNAQSRVNPFGFVYGYMRNLAPAPGKLN
PETGEITPEENADWSIADLSLAHLKVMFRTKVTRRLDKQDGQLYDSRYPVSVLHDAD
SKSASIWDKHPDLGYGGFSSAIPAYALIKFKNFKLVRNLQWTTYSDKNSDYI
LEQIRGKYPKAEKLVHSLPQYGQLVKKDGALVTVTALEHFNQQLWLADYKILINT
LKTDENLDILHNLDDLPEMTIESAFYKAFDSLNSAFNRYALHQNALVKLQAHARD
DFNALNYEDKQQTSLRJDLALHA+SASDLKINLSFNGRFLSFPSUFTLADTFDFI
QSVTGFEETQKTVAQLYQETK

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

MVYDVGLDGTSGVGWVADENGLKARAKGKNLQVGLFLDFDQTAADRRGFRTT
RRRLSRKRWp'lRLDELISAINEIDSSFFQRLKYSVHIPKDEEKAHYGGYLYFFTE
EETKFIIRSYTIYHLRQELMAQPVtCRIEDIREYAIiILVYKRGHFLSSQEKITIGST
YNPDELANAIEVIADKGLSWELNNEQLEISGEGAYGLNKSMKAEALKLFDEF
NNQDVKAIKLQGLTGNQDFAKLFGKIDSKDEALWLKLLDDEEIKSQTLS
QLTDEEIELFHAYVQAYDFVGLIGLNGADSVSAAMVQLYDQHREDKLLKSLAQK
AGLHKHRFSEIIYEQLALATDEAIKNGIARLVEESNLSEVEKLEDTRLDENEFLP
KQRTKANVHPHLAELQKIQGQYOYFPFLDTFEDGQDNKIEELLRFRIYY
VGPLVTKVDHEAGDGADNHVVERNEGFEKSRVTPWNVFDKVFNRDKAARDFIERL
TGlvroTYLIGEKTLPQNSLRYQLFTVLNENNYVRNGKFKDSKI KADLINDLFKARKT
VSLALKDYLKQGKDGVTITGLADESKFNSLSSYNDKKTDAEYLENEDNQETL
EKIEIQTVFEDSKJASRELKPLQDDQVFKLSQTHYTGWGRLSEKLDSKIIDERGQ
KDFTPJ'PJCSKGKVIWTPEKGRKLIVDLTKPSVLISNESHVKKGELFNAIAGKKDY
KJCGKIYLPLKKDDRQLDVSKYGGYKAmGAFFLVEHTKSKKRIRSIELFPLHLSSKF
YEDKNTVLDDYAINVLQLQDPKIIIDKINRYTEIIDDNFSLYLISTKSNDGTVKPNEQMY
WRRVDEISNLKJaENKYKKDAILTEEDRIMESYIDKIKYQQFKAGKYNKRRTTDTIIEK
YEIIIDLDTLDNQLLYQLVAFISLYSTSNNAVDFTVIGLGETGCKPRITNLPDNTYLV
YSITIGIYERIKR

[0073] ZP_07316256.1 CRISPR-associated protein, Csnl family [Veillonella atypica
ACS-134-V-Coi7a]
METQTSNQLITSII1LKDYPQDWDVGDLGNTSVPWATSTSELLLKFHSHHKWMGSR
LFEEGESAVTRRGFRSMRRRLERRKLRLKLELFADAMAQVDSTFFIRLHESKYiiY
EDKTGTGHSSiillIIIFIDETYDQFYTFFEFYIIHFLRKLLYMGNTMENTDIRKLFLAVHHLK
YRGNFLYEAGATFDNEAFFDLQKALVNTIFNCFTSAISSINILMESGKTSDK
AKAIERLVDTVTYPDNPDKPQKQQVKEDDDKLAFANLVLGLSALDLFLGSVE
DIIDDDLKLQIVGDTEQKDELAQKVGDEIFiiIDDCSYVDAILLMSIKFEGPILTSQS
KVKAIDFKIQELVGLSLKLDRNVYNEFMKSDKKGLFLYNVNYIYIKQGRTEETSCSR
EDFYKYTkTTTIVEGLADSKLEYIENELQTLLPLQRIKDNGVIPYQYLHLELKVILDK
CGPKFPFLHTVDGFSVEKLIKMLEFPYVYGPLHNTHHIDNGGFSWAVRQAGQVR
VTWPWEEKJDREKSSAAAFIKLNT>¾CTYLFGEDVLPKSSLLYSEFMLNLNLYRID
GKALAQGVQKHODSIFQKDQXKRMKTLDNITWTTKHKEITGLDIEKND
LTYSDMVLRJGNWDSMAMEDTITDTIFGESKKMLRQTLRNSQDQNLNDTITKKLS
KLRYREDWGRJSLJIO.LKGIDGCDKAGNGAPKTIIELMRDNSYNLMELGDKSFMECI
EEENAI1AQQGVV1SIFFIIDEALSPAVKRAVWQALRIVDEVAKHKLPSRIFFEVEV
ARTNKEKICICKDSRQKRLSLEYAIKDQDVLQSLQDKEFLAKSGLANYDDAAAR
SKXXLYYTYQMRGCAYTGNIDLQNLNTNDNYIDiYYPRSLTKDDSFDNLVLCERTA
NAAKSDKIYPIRDIQTQKPKFPLMTQKHLISERKYERLTRIAPLTADDLSGFHARQLV
ETOQSVKATTTTTLRLYPDIDVVFKAENYSDFRRHHNNIKVRSLLNHHHHHAKDAYL
MVVG>TVYHEKFRNFRLLFKKNGANRTYNLAKMFINYDVICTNAQDGKAWDKTS
MNTVKi<MMASNDVRVTTRRLEEQGALSADATYIKASVAAKBGAGYGMKLYKS
FADVTKYGGMTKIKNAYSIIVQYVTGKGEIKEIEVPLIYLRNATDIEILDIDYVKSVP
KAKDISIKYRKLNCINVLKWGFYYLYGGKTODKIIIDNAIELVVPHDATYIKLLDK
YDLRRKENKTLKASSITTSIYINTSTVVSLSNKVIGIDVFYFMSKLRTPYMKMKGN
KVDELSSTGRSKFIKMTLEEQQSYLLEVNLNLNTSKTTFVKPLGITGSRSTIGVHI
DEFKIINESITGLYSNEVTIV

[0074] ZP_08029929.1 CRISPR-associated protein, Csnl family [Solobacterium moorei F0204]
MEGQMKNNGLNQLQQGNYYLGLDVGTSSGVWAVTDTYNYVLKFRGKSMWGRALF
DEASTAEERRRTHGRNRRRLRRKYRLLLLEQLFEKEIRKIDDNNFVRLHESNLWADD
KSKPSKPLLMF)TNFTDKYLKKYPFYLHGLSDLHNSTEHDRLVFLALHHLIKYG
fifiYDNSANGVTKLETDAEVSDFEEYNENIEFNEMKKErNTVLSDKFLITKKEKIS
LKKLYGIDITSENINISVLEMLSSISLSNFLKDIEFDGKQNLDSLSDIEETLNDVVID
LGDonDLIIHAKVVDIATLSSLGKFIKYLCDAKVELFEKNNKOLMLKXYIKKNNHP
EDYKIKFSSPETEKKNAYAQSNTSINVCSQEEFCFLKYPYRDVMKSENEDEVRIAKE
VEDKSLTTLKGTA^SVVPQIQRELHNLQILKMVAYLPMNDEQEDISVDVKIKLIFK
FKIPIXVGPLNTKSTRSWYRSDEKIYPSV\WS>rVIDLDKTAHEFMNRLGECTYTTW
PVLPMDSLVLKYNtNEIPIKVKNGAIPEVKQAIYTDLFSNKKVTTRKSIYILIYLL
KNGYIEKEDIVSGIDIEIKSLSKHSHDF2TQVIQKENKCTPEEIERIKILVYSDSKSMLRR
WLMKMNKLGSENDVKYLALK>T*KIWGRSLKTLTIDTIYNPDGEDACESILDMWNTN
ATLMEILSNKYQFKQnENYKAEYDEKQNHLHEELDDMYISPAARRSWQALJRVD
EIVDIKKSAPKKIFIEMAREKXSAMKKKRTESRKDTLLELYKSCSSEQADGFYDEELFE
KLSNESNRLRRDRQLYLYTQMSRYTGKRIQDFDKLINDKNTYIDf\fYPRSKIKD
DSITIVLVEKDmGEKTDIPSEDRQCMQPWFkLKEKGLINEEKYKpvrTNYElt
DEELSSFVARQLVETQQTSTKALTLKXEYPSAKIVYVSAGNVSEFRNKDKELPKF
REINDLHAKDAYLNLVNGVNYDFTKIFENNNYENYSLRKFDFDFSVGAWDAK
GSTFNTIKXAKMNPIIIAFAYEVKGELFDQVIVPKGQKQPKQIYKKYGGYNK
LSSAFILFVEYKMKARERTLEYYKVDVLDAYQDIKYSYESVLGLKEPQIIKPKILMG
SLFSINMCLEVTGRSGQYVCHf\fYQLSnedcoutsQLKNIAYLQEEPNGQERQNI
LMTSVMKLFVDVLTCKFNSNTEYIILNSLKDNYVNEGREKFSELDILEQCNILLQLKA
FKCNRESSENLKNNKQAGVIIPFILLFTKCSFVKVHQSITGLFEKEMDLK

[0075] ZP_03989815.1 crispr-associated protein [Acidaminococcus sp. D21]
MGKMYLGLDIGTNSVGYATDPSYHLLKFKEPMWGAWHFAANGQSAERRSFR
SRRRLDRRQQVRKLVQEFAPISPIDPRFFIRLHESALWRRDVAE\DDFKFNIIFNPDPTYT
DKEYYSDYPTIHLLIVDLMESKFDPRRLYLVAVAWLAVHRGFIIFLNEVZKNIGDV
LSFDFYYPEFLAFLSNDGVSVPWVCESKLQALTLRSNSVDKYYALKSLIFGSOQKPE
DNFDAMSEDGLIQLLAGKKVKVNKLFPQESNDASFTLNDKEDAIEEILGTLTPDECE
WIAfIP-PXFDFWAIMKHALKDGRTISESKVLYEQHHHDLTQLKLYFVKTYLAKERDD
IFRNVDSSTTNKYNVAYSIVKEVGTKLTPKNKATQEEFCYVLGKVENECSEADKV
DFDEMIQLTDNSFMPKVSQGENRVIPYQLYYYELKTIENKAAASLYPLFRQCGDIAIS
NQDKLSMTFRIPYFVGLRKLKDSEHAWERLKAGKIYPWNNDKVDLDKSEEAFIR
RMTNTCTYYPEDVPLLDSLIEKFMILNEINNIRDGYPSVDVKQQVFLFEKRR
VTVDQIQLLLSSLGALDKGVKHLGIDTTILHISNYTYYHKFSLMERGVLTRDDVERIV
ERTMYSDDTKVRWLNSNYTTLADDVKHISRLKHDGFLKSMFLTLKGVHK
ETGERASILDMFMWNTODMLQLLSECYTIFSDTEITKQLQAYAYAKAQLSDLFDLSMYI
SNAVQRPIYRTLAVNDIRKACGTAPKRIFITVAREMADGSKSSKRSVTRQEKIKNLYRSI
RKFQFQEVIDFLNENKDSQGQLSQDALLYYFAQLGRDMYTDIPKLEHKQDSFYN
IDPfIPQSMVKODSLNDKVLVQVSEINGKSSRYPLDAIRNKKMPLWADAYNHGLI
SLKKYQRLTRSTPFDKEKWFENDQRLVETRQQSTKAILLLKRKFPDEIVYSKAGLS
SDFRHEFGLVKSRTNDNLHAKDAILVTNVYFIERFRNRFMRNMQNPYSVKT
FTHSHKNGOTVAVNGEELGRVKMLKQNKINTFIFTRFSDFKREGLFDIQPIKLSTGL
VPRKAGLDVVKGYGGYDKSTAAYYLLVRFTLEDKKTQHLMMIPVEGLEYKARIDHD
KEFLTDYAQTTISEILQDKOQVmiMFPMPTRHIKLNSIMIDGFYLSIGKKSSKGS
VLCHAMVPLIVPHKIECYIKAMESFARKFKENNKLRIVEKFDKITVEDNLNYELFLQ
KLQHNPYNKFSTQFVDLNTGRSTFTKLSCPEEQVQTTNLNSIKFTRCSSDGLKSING
SAQAARIMISADLTGLSKKSYSDILEVQASGGLFSKSRKSNLLEYL

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

MENKQYYIGLDVGTNHSVWGTDSYNYLRLAGKDDMKWARGFLFEKANTAERRTK
RTSRRSREKARKEMKELFADEINRDSPFFIRLEESKFLDDRSENQRQRTYTLFN
DATFTDKDYKYEKYTKFTHRLSAINDEKFVDRLVFLAILNLFSHRGFLNALSKDG
DIQGMDVFYYNDLVDSECEYFEIETRINDFNEKLSQKGRSTKIDELSEISIKKDS
KSKYINLKLISGLEAIVELYMEDIQDENKKIKGFRESIDYEESSLKVEEIIDFD
VERAKSVHDMGLLMSGNSKYLCEARVEAYSmmHKLKKIKELKGYDKNAYNMD
FRKMTDKNYSAYUGSNVSNIKERRSVDKRKIEDLYKYIEDTALKNPDDDKRIEIL
EKIKLGFLKQQLTASNGVIPNQLQSLRELKIAENLYPFLKKEKGEKLTVSEMIIQ
LFEFQIPYYVGPLPKDPKDKDANSWAKIKQGGRPWNFEDKVDVKGSRKKEFIEK
IVrVRKCTYSDEHTLPQSLLEKFMVLANEINNTKIDGEKISVEAKQIKYNDLFLVKGKK
VSQKDIKKELISLNIMDKDSVLSGTDTVCNAYLSSIGKFTGVFKEEINKQSIVDMIEDII
FLKTVYGDDEKRFKEEIVELYGDEIDDKDIKRLGFKFNSWGNLSKFLSELEAGDVGT
GEVRSHQLWETNFNLMELLLSSRTYMMDELEKRVKICLEKPLSEWTDIELDDMYLSSP
VKRMIWQSMKIVDEIQTIVYGAPKRIFVEMTRSEGKEVRTSKRDKRLKELYNGIked
SKQWVKEDESKDEYSFKSMMYLYLYQKGRCIVrYSGVIELDKLMLDDNLYDDFLYP
RSFVKDSDLNVLVVKKeqNRRKQNDPITPQIQASCQGFWKILHDQGFMSNEKYSR
TRKTQEFSEELKSFINRQIVETGQAIKCMAQILKSMGEDVDVFSKARLVSEFRH
KJELFQRSRLINDFFiIANDAYLNIVVGNSYFVKFTRPNAPFIDARKNPDNPVVYKH
MDRFFERDVKSKEAVAWGQSEGNSGTIVVKTMAKNSPLITTKVEEGHGSITKETI
VGVEIKEIFGKTVKADKTPKPNLQAYRPIKTSDELRLNILRYGGRTSISISGYCLV
EYVVKRRKTSRESAPVYLGRKDSLSEEKLLNYFNYLNGDDKSVSDIRLCPLFISTN
SLVKIDGYLLGYGGKNDPPQLYTSiaAQLLKMKEEVEYIRKIAKAVSMKDFEIDREK
NPVL.TEEMELYNKQDFKEFNTVFSKRMWLVYKNDLSSFGDFLKNNKSKFREDDEL
KQCKVLVYIFLNSNLKEVDDIGSKSTGCKCKKNITNYEKFQLQIQSITGLYCE
KDLMTI
[0077] CBK78998.1 CRISPR-associated endonuclease, Csn1 family [Coprococcus
catus GD/7]
MKQEYFLGLDMGTGSLGWAVTDSTYQVMRKHGKALWGLFRLFESATAEERRMF
TARRRDLRRNWRQIVLQEIFSEEISKVDGPFFLPRMKESKYYPEDKRAEGNCPelpY
ALFVDDWTDKWHKOYPTIYHLRKMLMETTEIPDIRLVYLVHMMKFRGAFNLLS
GDISQIKEFSTFEQLIQNIDLEELWfiISLDDAAIQFVEHLKDRNLTRSTKSRKIL
QLNASKACEAILNLSSGTVKSDIFNNKLDERSPKVSDASGYYDDYIGIVEAE
LAEQYIIASAKAVYDWSVLVEILGNSVISEAKIVYQKHQADLKTLKIVRQYMTK
EDYKRVFVDTEEEKLNNASYIYGMTPKKNKGDVLDKSSKQCTQADFYDFLKNNVIKVID
HKEITQEIESEIEKENFLPKQVTKDNIVIPYQVHDYELKXLDNLTRMFPIKENAEK
IQLLFEPJOPYVVGLNLNVRDDKGDKGKFPMWSVRSARJYPWNFTEDITVEASEAكف
RMTNKCTYLGDVLPKDSLYSVKSMVelnIyRLNEXESVELKQRIYEIFEL CYK
RKVTREKLYLVEIGAIKKGEVEITIDGDFAKSLAYHDFKERLTDVSQRAKEAI
VLNVVLFDDKKLLQKQLSCMKYPNLTTQLGKICLSLYQGWLRSLKTFLSEITVPAP
GTGEVWMATLWQTTvTDLNLQMLLSRNYGFTNEVEENFLKKETDLSYKTVDELYV
SPAVKIQWTLKVKVEIQQKVMAPKRFVEMAREKGGRSKSRKQLVELYR
ACKNEERDWTELNAQSDLQLRSKLFLYYIQKGRCMYSGETIQDLDELWDNTKYDI
DHIYPQSKTMDDSNRNLVRVXX>rWAISDTSYPLSDLIQKKMSFWKMLQQQFGIT
KEKYVRLVRSDELSADLEAGFIERQVETRQSTKAVATILKEALPDTEIVYVKAGNV5
WRQTYELEKVRNMNDLHHAKDAYLNNVGNAYVFKFTKNAAWFIRNNPGRSYNL
KEMFEDFIERSGEIAWKAGNKSIVTVKVMQKNMLVTMXAYEVKGGFLDQQIMK
KGKGVQPIKGMDEpLADIEKYGGYNAAGTYFMLVKSLDKKGKEIRFVEFVPLYLKN
QIENHESAIQYLAQERGLNPEILLSIKIDTLFKVGDGWMWLGRTNGQILFKGANQ
LILSHQEAAILKGKWNKRNENKDAKLIERDGMTEEKLLLQLYDFTLKDLSNTVY
SIRLSAQIKTLTEKRAKFIGLSNEDQICVILNEILHMFQCQSGSANLKLIGPGPSAGILV
MNNNITACKQISVINQSPTGIEKIDILKL
[0078] ZP_00143587.1 hypothetical protein [Fusobacterium nucleatum subsp.
vincentii ATCC 49256]

MKKQKFSDDYLYGFIDGTNSGVCVTLDYNYLRFNKKDMWGSRLFDEAKTAER
RVQRNRRRLKJIRKWRNLLEEEIFSDEIMKIDSNFFRLKESSLWLEDKNSEKFTLF
NDDNYKDYDFYKQYPTIFHLRDELIKNEPKKDRLYALHSIFKSRGHLFEGQNLK
EIKNFETLYNNLISFLDNGINSKIDMKELKIIICDSGKGLDKEKEFKGIFNSDKQ
LVAIFKLSVGSVSLNDLFDTDEYKKEEVEKEKISFREQIYEDDKPYYISILGELLD
IASKFYDFMV LNNILSINSEAYAKVLKYYEEHKKLKLNKYIRKYNENYKDLKFD
KNNENWPAIYGLNKEKDKKEVVEKSRKJDDLKVIGYLPKPERIEEKTIFNEILN
KIELKTLPKQCRSDNGTHLPIHIHVEEKEILKENQSKYDDFLNVEENGVSTSDKLLKTF
KFPJPPYYVGPLNSYHKDKGGNSWIVPvKEEGKLPWENFEQKVDieKSAEEFIKRMTNK
CTYNGEDVIPKDSFLYSEYIIILNELNKVQVNEF>ffENKPvKIDELFKEM<XVSEKK
FKEYLLVQIANRTVELKGKDENSNSYVSYIKFDIGFEGKLDIYKEIJEKSIWLWKC
LYGDDKKEFKEIYKINDEIKNKISFKFNTWGRSLKELTTGIEFINLETGECY
SSVMEALRRTNYNLMELSSKFTLQESIDNENKEMENVYSYRDLIESYYVSPSLKRAIL
QTLKJYEEIKITGVPXXVFIEMARGDESMKKKIPARQEOQKLKYDSCGNDIANF
SIDEMKNSLSSYDNSLRLQKXXLYLQLYQFGKCMYTGREIDLDRLLQNNDTYIDH
IYPYRPSKVIKDDSFDNLVLVKNENAEKSNENYPVKKEIJEKEMKSFWRFLKEKVFISDEK
YKRLTGKDDELRGMARQLVNRQRTTKEVKGILQIEPIKYSKAEIAFSSREFMF
DFIKVREIPNDTHHAKDAYNVATV>GWYNTKFEKPYRYLEIQHENYVDVKKIYNYDIK
NAWDEKNSLEIVKKNMEKNTV0*TRFKKEEGKELFNLIPKKGKGETSNIISIKPKLYDYG
KDNKLEKYGTYTSLKAYFYIEVEHEKNNKVKTFERITRIDSTLKNKMEKLYLVS
QXXLLNPKIIKIKIYEKQTLLIDSYTPYFTGVDWNKVEKLNKKQLYEKKYEQILKNA
LKFVEDNQGETEEWKFIYLKKRhWNEKNETIDAVKERYMEFNEIyTiKFLEKLSK
DYKNYNNKLYTNFLNSEEKFKKKLKLWEKLSSLRFLKIFNNTYGYKYEIKDSQTKE
KLFSFPEDTGRIRLGQSSLGNKELLEESVTFLVKKIKL

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

MTKEYYLGLDVTSQNYLNCLKKDDMGWGFESANTAKDRLQR
GNRRLERKKQRRIDILQIEISPEICKDPTFFIRL.NESRLHELDKSNDFKYPLFIEKDYS
DIEYYKEFTIFHLKHLIESIEEKSDFIRLGQSSLGNNKELLEESVTGLFVKKIKL

YP_005054169.1
CRISPR-associated
protein,
Csnl family

[Filifactor alocis
ATCC 35896]

MTKEYYLGLDVTSQNYLNCLKKDDMGWGFESANTAKDRLQR
GNRRLERKKQRRIDILQIEISPEICKDPTFFIRL.NESRLHELDKSNDFKYPLFIEKDYS
DIEYYKEFTIFHLKHLIESIEEKSDFIRLGQSSLGNNKELLEESVTGLFVKKIKL

YP_005054169.1
CRISPR-associated
protein,
Csnl family

[Filifactor alocis
ATCC 35896]
[0080] ZP_07398877.1 csnl family CRISPR-associated protein [Peptoniphilus duerdenii ATCC BAA-1640]
MKNLKEYYIGLDIGTASVGAWVTDESYPKFGNKGCMWGVPvLFDDAKTAEEERTQ
RGSPJIRLNRPVKERINLLQDLFATEISKVDPNNFLRDLNSDLRYREDKDEKLKSYTLFN
DKDFKIJRHYHKKYPTIIHLMMLIEDEGGKIDIRLLYLACHYLLKRGNHGFIFEGQKFD
TNSFDKSI>TOLKJFIJRDEYMDLEFWretDLIEIITDTTLNKTNKK^ ELKNIVGDTKFL
KAINAIRMGGISQKLVDFEGFEETTVKSVDFSTTAFFDKYSEEEAEALGDTISLLNIL
KSYIDSSILN[LLKI]ADKSDKDNGYISKAFFFFKKNHGDKLTLKRIKJLYLPESEYAN
IFRNSrhTONYVAYTKSMTSNKRTKASKFTKQEDFYKFIKKFILDIETKTNLESSNED
LKLIDEMLTDIEFKTIPKLKDSSNGV1PYQLKLMLKKLNDLQLSKYDFLINESDEYGT
VKOKVESIMEFRIYIGVPLNPSKYAWIKRENTKIPPWMDVDLDSSREEFIRLI
GRCTYLKEEKVLKASLYNEFMVLNLENNLKNEFLITEEMKAIFEEFLFTKIKKT
LKAIVSNLLKKEFNLTDILLSLGTDGDFQKGNLNSYDFKNIIGDKVDDIRYKIEEIIK
LIVLYEDDKTYLLKXKJSAYK>TOFTDDEIKXXIAALNYKDWRGSLKRFLTGIEEVDKT
TGEKGSIIYFMREYNLNLMLMELSMGHYFTFEEVEKLNPVENELCYEMVDELSPSV
KRMLWQLRVDEIKRIIGKDPKKIFIEMARAKEKNSRSXERKNXLLEFKFGKKA
FINEIGEERYNLLEINSEESKFRWDLNLFLYYTLGYRGCMSLEYPDLADLKSNNMY
DQdfiiYPKSIYDSDLRNVLKVKNLHEKNQYPKEPVKLKNAYFGWKILFDKGL
IQK닥KYTRLTRTTPFEERELAEIERQIVETRQATKETANLLKNCQDSEIVYSKAENA
SRFRQEFDDIKCRCTVNDLHLMHDAVLMVGNVYNTFCFTKNPLNFIKDKDNVSYNL
ENMFKDYVRGSTAWDIADDSEGNVKAATIKVRELEKNGYFRMYSYGTYTGL
YDQNLMRKGKQIQPKKENTKSNIESKYGGYNKASSAYFALIESDGKAGREGLTERIP
MVYNQEKYNTJndoVDKYKDNLELDQPKIKLKDGIINSLNKIDGFNYIKGKTGDSL
SIAGSVLIQNEVKKEQQLIKKMDQKKVXXKNDKICKVTSFDMKEELKLYKTLSDKL
NNYGYSNKRNAQAKMSELADKFKEISIEEKJVDVLNQHLLFQSYNNGCNLKSIGLASKT
GVVFIPKKLYKECKLINQSITGLFENEDLNL

[0081] NP_970941.1 CRISPR-associated Cas9e [Treponema denticola ATCC 35405]
MKKEIKDYFLGLTDVGTSGVGATVTDYKLLKANPVKLGLGMRCFETAETAEVRR
LRGRARRRIERRKKRIKLLEQFSQEQIAKDEGFFQRMKESFYAEKDKTQLENF
DKDFADKTYHKAATtmHLLUKWENKVKDPPJ_1LYLACHNIIKKRGFLFEGDFDSE
NQFDTSIQALFEYLREDMEVDADADSQKVKEILKDSSLKNEKQSRLNKILGLKPSDK
-51-
QP KKAITNLISGNKINFADLYDNPDLKDAEKNSIFDKDDFDALSDDLASILGDSFELL
KAKAVYNCVLSKVIGDEQYLSFRAKVIYEHKHTDLLKLNVIKHFKPKDYKVKVG
YMCNEKNMWYSYGVGVCYKSSKAKINNSVQEDFYYKFLKILTAIKEKSEIKEVNDILT
EIETGTFLPQISKSNASEIPYQLRKIMELEKLSNAEKHFSLKQKDEKGLSHSEKIML
TFKIPYIGPINDNHKKFFPDRCVCWVKEKSPSGKTTPWNFDFFIIIDKEKTAEEAFITSR
TNFCCTVLGESSVLPKSLLYSEYTVLNFRLNLQIIIDGKINCIDIKLQKIVYDKLKYYK
KITFKQUISTFIKHEGICNKTDENVnLGIDKECTSSLKSYELKNIGFGQVDEISTKNMLEEI
IRWATIYDEGEKTLKTIKAEGYKICSDEQIQIKLNLKFSGWGLRSLKIFLETIVTSE
MPGFSEPVMITAMRETOQNLMLMELLSEFTFTENIIKINSFEDAEKQFSYDGLVKPPLF
LSPSVKMLWQTLKLKEISHITQAPPKKIFIEMAKGAELEPARTKTRKLIQLDLYNN
CKNDAAOFSEKSEDGSKJE>¾%DIKNLRSDKLYYTLQGKCMYCGKPIEIGHVFDTD
NYIDHICYPQSKIKDDDSISNRVLVCCSCKNKEDKYPKSEQSKQRGFWNFQNNF
ISLEKLNRLTRATPISSDETAKFIARQLVETRQATKVAVKVEKMFPEITKVYSKAE
TMVSMRFNKFVKECINDDHHADLTYLNVGANYNKTFNPNWIFIKEKRDNPKIA
DTRYNYKVFYDFYVKNMTAWEKGTITVVDMLRNPITYTARAACKKEGFQNT
IMKKGGLQHPLXEGPSMSKYGGYNVSAAYLIEEYKKGINSLETILPYVK
DIQKDQDLVSKYLTDLLGKEFKTKVPKIKINSSLKINGFPCHTIGKTNDSSLRTPNAVQ
FCCSNVEVLFYKKLRFSEIRSREKJGKTNPFYEDLSFRLSYIKNKLWKTIXNIDEIKEE
FYDLLQQKNLEIYDMLTKFmCDITYKKRPNSATIDLKVGEFKSKLIIENQFIVILEIL
KLFSATRNVSDDQLHIGSKSYGVAKIGNKISLDNCILYQSTIgfEKIRDDLKV
[0082] ZP_07912707.1 conserved hypothetical protein [Staphylococcus lugdunensis
M23590]
MNQKEILGLGIDTVSYGLIETYKMIADGVRLFPEANVENNEGRRSRKSGRRLKR
RRIIHRLEVKLLEDNAQQPSQIQPSITPMMAYRKPGEALSKDELVIAL1f1AKRGG
IHKJDVIDSNDDVQGSLNTEKQENLKNNSKLKDQKFCVQIQMrNERMGNQFrKFT
ADIKEIQQLLNQKNQFHQLDENFINKYIELMVERMLEYFEGPGBKPGKSPYGWEQPKA
WETLMGHCTYFPDELSVYAYSADLFNALNDLNLVIQRDGLSLESHEYKYNiEN
VFQKKQKPTLQIANEINVPNQPKYRTKSGKPFTEFKLYHDLKVLFDSILENE
DVLDQIAEILTIIQDKSISKLTLEDILIINEKNAQLTGTYGTHRLSLKCIRLVE
EQWYSSRQMEIFTHLNIKJNNXTAANIKPKAMIDEFILSPVVKRTFGQAÅINLIni
EAYGVQEDHIELARENSSKDQKQFINEMQKKNENTRKINEIYGKGNQNAKRLEVEK
IRLFIQESEQKLYSLESIPLEDLLNENPHYEVHIDIIPIRSVSFDSNYSYHKVLEK
QKSEQNSK
KSNLTPYQYFNSGKLSYNQFKQfHNLKSQDmSKKKEYLLEERINKFEVQKE
FINW^VDTRYATRELTNYLKAWSANnKWKTNGSFTDYLKRVW^K
KERNH
GYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIESKQILQVQVDSEDNYSEMIIP
KQVQDIKDFWRKYSYRDXXPRNQLnNTOTLYSTRXKDNSTYIVQTIDYAKDNTT
LKQFQDKSPEKFLMYQHDPRTFEKLEVMQYANEKPLAKYHEETGELYLTKYSSK
^GPIVKSLSKYGNKLGSMLDVTHQFSSYKLKLVLKSKIPYRFDVVLTDGKYKFTIS
YLDVLKXDNYYYIPEQKYDKKLKLGKAIDKNAKFIASYFKYNLDKLIDGIEYKIIGVNSD
TRNMIELDLPDIRYKEYCELNNIKGEPRIKKTIGKKVNSIEKLTDVLGNSFTNQTY
KPQLLFKRGNE
[0083] ZP_02077990.1 hypothetical protein EUBDOL_01797 [Eubacterium dolichum DSM 3991]
MMEVFMGRLVGLDIGITSVGFIIDLDESEIVDYGVRLFKEGTAANENETRRTKRGGR
RLKRMRVTRREDMLHLLKQAGIISTSFHPLNNPYDRVKGLNRELNGEELTALLHL
CKHRGSSVETIEDDEAKAEAGETKVKLSMDQIQLLSKYVCEIQKERLRTNGfHfRG
HENNFKTPvAYVDEAFQIISQDSLNELSAHTnSRKRMYDGPLGSPTPGRYTY
FGQKPIDLIEKMRGKCSLFPNERPAPKLYSAELF>n _LNDLNNSIEGEKLTYSEQKA
MILKIVHEKGINEPSQKLAEGVSELEQIRGRGIDTKGSPLLSLETYGKMIERVLEKSND
EHLKDHFVYDEIAEIIKTDFGLIRNLSMLLNSLNEESTHQLAGDLKNATYHSLSFK
ALRINEEMLLNLQMSITFLGKLQVfHfLSVGMKMQADDTALISPVAKRAQE
TFKVYNRLREIYGEFSISVEMAREKNSSEQRQKAIERRQFFEMRNKQVADIGDDR
KINAKLREKIVLYQEDQGKTAISTPIDLKLDDPNAYEVDHIGIISLDSITNKL
VTHPvENQEGNLTIPAFVKGRTFKGSLAQYKAYCLKLKEIVNKTNGYRKQVEQY
LL>fHfINIDYQDIQETQINNNLVDTSYASRVL,NTLTTYFKQNEIPTKVFQVSGSTNA
FRRKINLKLKDREIYGHHAIDALIAMPSKRMRLSTIFRKYIEDIYDESTGEVFGSGD
DSMYYDDRYAFIAKSLFRKFSKIDTPRNSVADETIYSTRVIDGKEVVKK
KDIYDPKFTALADILNNAYQEKYLMALHDPQTFDQIVKVYVNVYFEMSKSEKYFT
KBKKGRIKISGMPNLSLYVDEHGMKLKYYSGPGAPTQMKYFDGVIGNHIDSAH
YQVRDOKVVLQISSPYRTDFYYSKEWYKFTPRTYKDVRSKQWSEEKKKYVQDQA
AMKKAEEKKIDDYTEFQFSMHRDELIGITKAEGALIPDTQFINTNFHHAGETPEILK
FTATONDNSKIEVKPIHYCCKMVPLOMTPISKIVPDJKLMNOPVATDVGVNLKVVK^ TLKF
EFD
YP_820161.1 CRISPR-system-like protein [Streptococcus thermophilus LMD-9]

MSDLVGLDGISSGVGVGLAVTGEEHKHNSRFPAAQAEINLLVRRTORQGRPxLARRKXHRRVRLNRLFEESGLIDFTKISINLPYQLRVKGLTDESLNEELFLALKNMVHRGISYLDASDGNSSVGDYAQIVKENSQKLETKTPGQIQLERYTGYGQLRQDFTVVEKDGGKHRLNVFTPSYRSEALRLQTPQEFNPQITDEFINRYLEILTGKRKYHGPNESKRTDYGRTSGETLDFAGILGKTCTFYPDEFRAAKASYTAQEFLNLNDLNLNLTVPTETKSLSEQKIQNYVKNEAMGPALFKYIAKLSSCDVADIKYRGKDSGKAESISHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLTNEREGIQEALHEFADGGSFKQQVDELVQFRKANSSIFGKWHSVKLMMELIPELETSEEQMTLTRLKGKQKTTSSSNKTKYIDEKLLEETEINPVKASVQRQAIKmSTAAIKEYGDFDNIVIEMARETENDDEKKAIIQIKNKDEKDAALMILAAANQNYNGAEFLPHSVFGHKQLATKRLWHQGGERCLTYGKTISHDLINNSNEVESFDHILPLSETFFDSDLANKLVYATANQEGQRTQYQALDSMDDAWSFRELKAFVRESKTLSSKKKEYLLTEEDISKFDVRKINFLVDTRYARSVVLNALQEHRFAHIDKIDTQSVSVRQFQSRLQRHRWGIEKTRDYTEHHHAVDALIIASSQNLWKKQKNTLVSYSEDQDLLDIETGEISDDEYKESVFAPYQHFDVLTKSKEFEDSILFSYQVDSKFKNRSIKDATTIYATRQAKVGKDADETYVILGKITDIYTDQGDYDAFMKYYKDKSKFLMYRHDQTFEKVIEPILENYNPKNEKGEKLPCWFLKYKEEHYIYRKYSSKGNGPEIKSLKYYDSKGLNHIDTPKDSNNKVLSVSPWRADVNYKTTHKYEILGLKYADLQFEKGTGTYKISQEKKNDIYKKEVDSDSEFKFTLYKNDLLLKVDDELTKEQQLFRLSRTMPKQKHVEKYPDKQKFEAGAEALIKVLGNVANSGQCKKGLGKSNISYKVRTDLGVNQHIKNEGDKPDLDF

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

MYSIGDLGISSGVWSVIDERTGNVLDLVFRSAKNSEKINRLLETNRRGRRRLRRKTNRLKDAKKILAAVGKYEDKLSNCPYQLRVKGLTEPLSRGEIKYVTLHLIKKRGISYLDEVDTEAAKESQDYKEQVRKNQALLTKitTPGQIQOLRLKRNVRKGTQNAQNYQLNVFKSAYANELATILKTTQQAFYPNELLTDWIALFVQPGIAEAEAGILYRKRPPYYHGPNEANNSPYGRWSDFQKGTGEPATOIFDKLIGKDQFGELRASGLSIAQQYNLLNDLTNLKIDGVEPLSSEQKEYLTEMLMTKETTRFVGNDVVKLGSVKKRLYSGWRDLKGGKPEIHTKGYRNWKIAEAGIDLAPTEDLCLKLTVLTNTEREGIETLAFELPELSVKLLVLDRYKELQSIQTQSWHRFLKLTLHILLPELMATSEQNTTLEQFQQLKSDV
Ki/YSEYKKLPTKDVLAEIYNPTVNVKTSQAFAKVIDALLVKYGKEQIYRTIEMPRDDN
EEDEEKRIKELHAKNSQRXTOSQSYFMQKSGQWSQEFQFTTQKNRIFLAKLLYYE
QDGI/CAYTGLPISPPELLVSDSTEIDHIIPISLDDSSINNKVLVLSKANQVKGQQTYPDA
WITOGSFXXTGKSNWDDYEQKVWESRHSYKHKENMLERTRNFDSEQVEKFLARNL
M3TRYASRLVNLTIQLSFFTNQETKVRVNVGTFTHFLRKKGADLDKTRETFIHHHA
VDATLCAVTSFKVSRHYAVKEETGEKVMREIDFETGEIVNEMSYWEFKKSKKYE
RKTYQVKWPNFREQLPX^FIPRPKFISHQVDRAKANRKLSDATIYSVREKTEVKTLKS
GKQKITTDEYTIGKIDYTLDGEWAEFFKQDKDLKMLDKLDEKTIERLLSFDPQ
EVEEKNKGVKVRKRSFPAVYCEENIDPAIQKYAKKNNPLIRSLKYDYGKLNKHINI
TKDSQGRPVEKTNGRKTQLQSLKPYRDIYQRDLTEKAYTVQLYSDLRFVEGKY
GITEKEYMKVAAEQTGGQVRFCSLQKNGDLGIELWDSQRHVDVRFYNFQANSIN
FKGLEGQEMPMAENQFKQPYNNAGINLNAKYKYGKEGK/GLKPFCHNIDILGKHYLYFE
KEPKNIK

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

MWTEKEKLFMKYILALDGIASVGWAILEDKESSETVIEAGSNIFPEASAADNQLRRDM
RGAKR^NRRKLKTRINDFIKLWENNNLSIPQKFSTEIVGLKVRAITEITLDELYLILSY
LKHRRGISELDALDDTVSGSSAYANGLKLNAKELETHYPCEIQQERLNTIGKYRQS
QIINENGELDLGNVTFIGAYXQIEQRVEQFiQQYXHPELTDIFGGLYMIFKRRKY
ECPGNEK5RTDYGRFHTKLDANGNITYEDNIFELKLGKCSVYPDELRAAASYTAQE
YTWLNDLNmTINGRKLEENEKHEIVERIKSSNT>hVIRKIIDCMGEMDDFAGARIDK
SGK^IFHKFEVYNKRMALLNEIDISINGYSNREELDEIGYIMINTDEAKMEAFQKSW
IDLSDDVKQCLFmriRTKOGAL^WQSFSLKIMNELIPEMYQAQKA^QMTLTLTEMGV
TKGTQEEFAGLKYIYPDVSEDINPVVRSSVRISFKILNAVLYLKKYALDTIVEMPRD
RNSEEQKRRINDSLQKLKEMEYIEKXLAQTVGIKLSPDFSSQKQLSLLKWNTEQ
DGCICLYSGKTIDPNISriPQLFEIDHIIPRSISFDARNSKVLVYRSENQKGGNQTTPY
YLTSHSHESWSFQYKATVMNLSKXXKEYAIRKKIQNLLYSEDITKMDVKLGFINRNI
>TOTYSARLVLNTINQNFMANEDTKVKVIKGSHYTHQMRCLNLKDNRESYSHFIA
VDAMLIGYSEGHLYHKQEGRETFETEILRKMWDENSMDEVYADLYGKKW
AMRNEVKAEX>TVKYWHYVMKRS^4GLCNQTRGTRGKQYKINKLDIRTK
GIKVFALKFSSKDDREDLLVLNLDRRTFDDLCKIYEYSDAAAPPFVQQYKEKTDGII
RKYSKKGHPJDKLLKYKDGEVGACIDISFQCGYGFEGKSJKVILESVPYPxMDVYYKE

-55-
[0087] YP_015730.1 hypothetical protein MMOB0330 [Mycoplasma mobile 163K]
MYFYKNKENKLKKVVLGDLGLIASVGLCTDISQKEDNKFPILHGVRFLFETVDDS
DDKLLNETTRKRGRQBnRnRRLFTKRDFIKYUDNNIELEDFKNPKILVRNPIEYl
NPFSKNLELYKSYTVNLPiGHNLRKAAINEKYKLKSELVLLLHYFSLRGAFFDNP
EDTKSEMKNKNEIEIFDKNESIKNAEFPDKIEFYKISGKIRSTLNLFHQGDYTEIQK
VEFKQMDMNYEKFAMEEKSFSRIRNYSEPGNEKSKFYGLYANENGNPELIINE
KGQKJYTIFKTILVESTGKCSYDKLKYRACKNSFSAKVDIOTKLTDWKhKNEYlS
ERLKRKILLISELKSDKSAVEKILKEWKEFAEIAYNKDDhraPMnNAYNLSSTT
IFKKHlNENYlSNEDSLKLMFYKQQSEKLFVPNEKGSYEINQNNYLVHIFDAIS
MLNKFSTIQDRILEGEYFESFNLKDVKSSlJEIylKREFSGTSSLSFAGAYKYFIPN
LISGSKYNIStsYEEKAlQNQKKNWHSNLFEKTWEDLIASTPVKSRLQTMNLLK
EIKYSEkwLElEIKVEVTRSSNNKFlRIKlEgINKYRlEKYElKlKYYLYDPNENT
TLLKKWLRLQQQGYDAYSLRlKlEwAEOINKPNWDYDHIlVPNISFDDFSNSlLVN
KLDNAXXSNDLSAKQFIEKlYlGIEKlKEAKENWGNWYLRNANGKAFNDKgKFBIlY
TIDNlDEFDlNSDFINRNLSDTSYITNALVNHlTSNSKYYSVSVNGKlQlSNlRNQL
AFVGIKNlKTEREWKRPENGSFNNSDFNFIREEGKNDVKDDVlKDRSFNGHHAED
AYF1IISQlYlRSKFRIELNlNYlRKEtRELDDlELKNNIKFKEkASFDNFLINALDELN
EKLNQMRFSRMVlTkkNTQLFNlTeYSKYDKGKlNTIKVEKLlLDNRDTKlKKIE
EFFDDlKlKEnELTlKHfNNDKnLYETLKlWNEVKElKlNlNEKYNFKYMNK
LQEGKISlNEWVPILDNDKFIRKlYlKFSESEKETDEIFSQSFLKlDIQRQNFShNT
LYWVQlWVYlKQnKQnCQFISlDJARlNSKFkKElKlNYEKlKlTQlKEtLQlIEnELPKIN
KGDLFENIEELFYlVRDEKPKqLEIKYlLGKlKIDQlQlQIKPKVlKFNPWKnKVNl
TYMElIFKK

[0088] ZP_09312133.1 hypothetical protein MoviSl_00710 [Mycoplasma ovipneumoniae ScoI]
MFINKXMTlGFDLGlASIGWAIIDSTSkJKlLDWGTnTRTFEElXTANERRAFRSTRRMRR
KAYRNQRFINLlKDYKDlFKNTSDIQRANKKTEN YEKISSFTEIYKCAAKHSNlL
EVKvKALDSKIEKlDLWIHDYLlElnRGFFYDFLEEEVAdKtEyGIElEPsIlLYDFIKK
NGFFKSNSSIPKDLGGYFSNSlQWVNElEkCFLFvEINPElFEKlFLNlFTsvRDIYAKGP
ADC31648.1 Csnl family CRISPR-associated protein [Mycoplasma gallisepticum str. F]

MNNS1KSKPEVTIGDLVGSVGVWAIWNEMTHHLGSRFLSQAATAEDRSSFGRV
mRRRKYKLFVPLIKYNSYFPGFKNKEDILNNYEQQKHLNVLKSEALNA
KIDPKALSLWIHLKYNHRGFEYEDNRFNYPTKELAKYFDEYKYYKGIIDSKEDN
DNKLEELTKYFNSKHLEELVKVLSQNTGLPEFKKEYESLFSYVRNYSEPGSI
NSVSPYGITYLHKDEEGKVVQKYYNNIDTIGCNIFPDEYAPKNPIAMIFNEINELS
TIRSYSIYLTGWINQFEKFAYNLKLDLILIKTNQEKPIDARQFKLKLREETIAESIGET
LKDVENEEKLEiCDEHKWKLKGLKNTQKIQLNLSLAKFHVKLQHKLDFLLE
DQMAYTLDFINLQVFLYLGKHLRYSNVDARSNLKEFSDSNKLFE1QKQKGLFLK
LFEQTDKDEKILAQTHSLSTKAMALALTIPzMNLNDENQKNFKNDG\VNFEAIKNF
DQKFIDIT<K<NNw_5LQKNSKRYLDDRIFNDAILSPGVKRLREATKVFNAILKQFSEEY
DVTKVVEIARESEELSEEKNTKNKLLKNGDKEISEKALGISEDEIKDILKSPTK
SYKFLLLQVQDfifDPYSLKEIAFDDFTKTEKFEIDHIIPYSFDDSSSNKLLVLAESEQN

[0089]
YP_278700:1 hypothetical protein MS53_0582 [Mycoplasma synoviae 53]
KLLNEINLKPENKKFNSINEIEFVKFVKPGALLLNFENQQIYYISTLESSSLRAKIKLL
N\text{KMDKGKAVSMKKI}^{14}\text{TPDEYKII}^{14}\text{EHVNPLGINLNWTKKLENNN}

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

MEKPKJCVTGLFDLGIASVGWAIVDSETNQVYKLGSRLFDADPTOLEERRTQRGTRRL
LRp_{\text{VK}}\text{RKYRNQKPYNLVKRTVEFGGLSSREAIIENRFRELISIKYPNIEELKTALQSE}
QVECPDEIAWILHDYLKNRGGFYDEKETKEDFDQQTVESMPSYKLNEFYKKYGFKGAQLSQPTE
ESENIKDNKLKEAFFDFSNKEKWLEINYF\text{Fm}^{\text{a}}\text{QKMSETFIEEFKJaFSFTRDISKG}
PGSDMNPSPYIGFEGFDNGQQQGGRYEHIWDKNIGKCSIIFTNEQRAPKLYLPSAEIFNFL
NELAMP_{\text{VL}}\text{LSTDDKKMQPLWKSLSSVDKLLNFLFMPISEKKKKLSTTNODVKKE}
ISISIMISEDINDKDEAGKENPNYVGVGLSGLNIEEASKENKFQFDKLHILTVLITLL
D>WG:\text{IKFEMKDR}^{10}LNLLELEL\text{N}^{10}YLYFLYQKESNNKKDDISSLIAKNESLMENKLK
LKEFLLLGAGNEFENHSKTHSLSKKAIDEILPKLLE^{10}GW\text{NLEAIK}NYDEIQSIE
DNSSLMAOQDKKLYLNDLFKDAIPPNVKVTFQQAILIF\text{N}^{10}IKQKFSKDFEIDKVKVIEL
AREMTQDQENDALKGIAKAQAQKSXSLVEERLEANMDKSVFDNYKLELYKIFLWIS
QDFKDPYTGAQISVNEIVNKVEDHIIPYSLCFDSSANXVLVHKNQSNKESNLSLPY
EYIKQGHSGWNWDEFTKYKRVFVM^{\text{a}}\text{DSILSKKERLKKSENNLLTASYDGYD}^{10}KLG
LARNLNDTRYATILFRDQLNWAEEHLIDNKKMKFVIAMNAGATSFIRKNMSYDNK
LRLKCHRDSFISHHADAAIALFSNKTGTKTLYNLIDPSLNGISKRSEGYWDRIYTGIE
KELKKEWTSIKNNVQARKIAKEIEEYLIDDLDEVFFSRKTKRTKNRQLNYETIYIGA
TKTDEDGITNYKKEKFSILDDKDIYLLREREKFINQSNPEVIDQJIEIESYGKEN
hnPSRDEAIKMKTIGNLYLQYMRSLTKSLQDSEIFQMNIAKTNFTVLYNPT
KNTTRIKKFLRLVNDVNNCINDIRKQNYVTKNFGNKNPEAKFENINSLGAIVFKNSAN
FKTLTIQAIIFGDKNWDIEFDTYNMEKEKYKEIYGDKYTFSHFIFPFIGTTLDLKQ
N^{4}E\text{FY}^{4}\text{I}^{4}\text{SI}^{4}\text{EQTVRD}^{14}E1\text{KFLLKKEFKFJENKNQDTSKRPKL}^{14}\text{MFGIKSIM}^{4}\text{^Y}^{4}\text{YEQVDIS}
PFGINKK\text{IF}\text{E}

[0092] \text{NP}_{\text{907605.1}} \text{ hypothetical protein WS1445} [\text{Wolinella succinogenes DSM 1740}]

MIERILGVDGLISSLGWAIVEYDKDEAAAPJIDCGVLFTAAETPKKESPNNKARRE
ARGIRVRVLNRRRVRMMNKIXLFLRAGLQVDLDGEGGMYFSKANRADWVELWHD
GLYRLKGDELARVLIFLAKHRGYKFGDDEADEESGKVKKAGVVLQRNFEAAGCR
TVGEWLVWREGANGKKNKHKHDYESIHHRDLIVEEVAIFVQEMRSTIATDAL
KAYREIAFFVRPMQRIEKMVGHCTYFPEERRAPKSAPTAEEKFIAISKFFSTVIIIDNEGW

-99-
EQKIERKTEELLDFAVSREKVEFVHLRLKFDLSDNEIFKGLHYKGKPKTAKKREATLFDPNPETELEFKVEAEKAWISLGRAAKLPvEALGNEFYGRFVALGKHADEATKILTYYKDEGQKJIRELTKPLAEAMVERLVKIGFSDFLKLSLKAIRDILPAMESGARYDEAVMLGVPHEKESAILPLLNLTDILNPTVIRAFAQFKVANALVRKYGAFDREVHFELARENTKGEIJEDEKESQRKNEKERKERAADWIAETSFSQVPLTRKNLRLYYIQQDGRCAYTGDVIELRFDGECIDHILPERSADDSFANVCLCLARANQQQKTDRTPYEWFGHDAARWNAFETRSTSAPSNRVTGKGDIRLKLKNFDENSEMAFKDRNLNDTRYMARAICYTEQYWWFKNSHTKAPVQRSGKLTSVLRYQWGLSEKSHRTTHEADHAIAFSTGQIVvQKLSYYFYRFKETHREKERPFLAVPLANFRDAVEEATRIENTETVKEGVEVKRLLESIPPRPARVTGQAHEQTAKPYPRIKQVNKKKWRALPIDEEKFESFKADRVASANQKOTYETSTIPRDVYHKGFHFVYPLHLVMELPLNLSGLTNPEAMDENFFKFSIFKDLISITQGTPKPKAIIMGYFKNMGANMVLLSSINSPECGFTCTPVSMDDKHDKCKLCPEENRAGRLQGLFDLYWSQEGLRRPKEFEDCQGVFKALDVKKYQIDPLGYYVQKQLGRTIPQMRSAKKLVK

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

MARILAFDIGSISSIGWAIFSENDLKDGCVRFITKVENPKTGESLALPPRLSARKRLARRJCARLhfllKHLIANEFKLNYEDYQSFDESLAKAYKGSILSPYELRFRALNELLSKQDFARVIIflAKRGGYDDKNSDDKEGKAIKQNEEKLANYQSQVEGYLYKHEYFKKFKENSEKFTVRKNNKESYERCAIQSFLCOELmFKQREFGFSSFKKFEFEEVLSVAFYKRALKDHSFLVGNCSFDTDEKRAPKNSPLAFMFVALTRIINLLNNLKTTEGILYTKDNLALLNEVLIKNTLTYQKTKLGLLSDDYEFEKIGKEGTYFIEFKKYKEFIKALGHEHLSQDDLNEIAKDTILIKEILKAAKYDLNQIDLSKLFKDIILNLISFALKLVTPLMLEGKYYDEACNENL^KVAINEDDKDFLPAFNETYKDEVTNPVLRAIEYRKVLNALLKKYGKVeMARELVQGHNSQARKJEKQNEYAKKAADACEKLGKLKWSMLKLRLFKEQKEFCAYSGEKIKISDLODEKMLEIDYIYPYRSFDSSYMNKVLVTFTQDNQEKLNQTPFEAFGNSDAKWKIEVLAKNLPTKKQKRIDKNYKDKEFKNFKDR^>TOTRYIARLVLNYTKI)YLDLFLPSDENTXX>roTQKGSKVFRVEAKSGMTSALP^TWGFSAKDRRNHLLHHAIDAVnAYANNSIVKAFSDKFKEQESNSAELYAKKISELYKKNNKRJFEFSFGKVRKVLKIDIEIVSKPEPICKPSGLHEEIRKEEFFEQSYSGGKEVLKAELGKIKYNGKIVKNMDFRVDIFKIFKNTKNFYAVPIYTMDFALKVLPKAVARSKKGEIKDWILMNDENVFCFLYLKDSLILIQTJKDMQPEFQVYNYAFTSSST
VSLIVSKHDNKETFLSKFNQKILFKNANEKEVIAKSIGIQNLKVFEEKYIVSALGVEVTKAEFRQREDFKK

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

MRTLGDIGIASIGWAVIEGYTDKGLENKEMVAVGVRVFTKAENPKNKESLALPRTL
ARSARRRNARKKRQIQKQVYKHDLSKLDLECFVQGEKINATLQFTSKDFLPWELE
ERALYLVLDKEELARVLHKAIKRGGYDINTYGEDNSGKKKAESNKRKEEQCK
TIGEMMYYKLFQKSLNVRKESYNRCVGRSELREELKTFIQIQQELKSWVEENL
YKLLGNPAQSKQEREGLIFYRQRPKGFGBKGCISHKKGKNSP YRACKHAPSAEE
FVALTKSINFLKTNLTRNHRGLCFSQEDMCVYLGKILQEQAKNEKGLTYSLKLLLLDP
SDFEFLGLDYSGNKPEKAVFLSLPSTFKLNKITQDRKTQDKIANILGANKDWEAILKE
LESGLQKSEQIQTDLNKLFSKHINLLEALYHILLPLMEQRYDEGVEILQERGIFS
KPQPKNRQLLPLSELAKSEEFDPNPVLRALSEFRKVVCNALLEKYYGGFHYFHIEL
TRDVCKAKSAMQLEKVINNNKSENDAAASQLLVLGLPLLASYNNLKLFK
CITQDEEYSVQKVKLWQCEE
YCLYSGKSKITIDLDQKALQIDAHAPLSRLSDQNSKVCLTSSNQKSKNPSNCTPEW
LGSDEKXXWDIVTVGVRVYSSNSFSPPPVRKRTLKQKW^RNEEFLARNLVDGTYGIRVT
KEYIKHLSFLPLDGKXEfiRIJSSMTSTMRSFVQKNEKRDHLHILHAQDAIIACI
EPSMIQKYTTYLKL)KETHHLKSHQKAQILREGDFIKLSLRWPMNFKDKQPIEQS MP
SHHVSHEKVTGELHQTERTKKEFYYQAEGGEEGVKKALKFGKIKREINQQGIVDNGAMV
RVDIFKSCKDKGFYAVEITYDFAIGKSPNKAIVQGKNGKIDWLENIDENYECFSL
FKNDCIKIQTKEMQEAVLAIYKSTNSAKATIELEHLSYALKNDEEEKMFDTDDEKE
KNTMTRESCGIIQGLVFQKV KLVS LGEVLEHKPRN QRQIALKTTPKH

[0095] ZP_06887976.1 CRISPR-associated protein, Csn1 family [Methylosinus trichosporum OB3b]

MRVLGLDAGIASLGWALIEIEESNRMGELSQGTTIIGAGTWMDAPEEKKTQAGAKLKE
QRRTFRRGQRVRVRRRRQRMMNVERRRLSHGHPSSDRALKQPLGDPWRIAEALD
RLLGPVELAVALGHIARHGRFKSNSRGAKT>roPADDTSKMKRAVENETRKLARFGS
AAKMLVEDESFVLRQTPKNGASEIVRRFRNREGDYSRSSLLRDLAEMRALFTAQ
ARFQSAIAATADLQTAFFKKAFFQRPLQDIKSVLPGLPCPFEVDEKRAPKRQGYSFELRFR
SRLNHVTLDKQERTLRTDELALAAADFAGAAKVSFTALRKLPLPETTVFVGVK
ADECLVDVVARSGKAAEGTARLSVIVDALGELAWGALLCSPEKLDKIAEVISFRS
DIGRISELAGQACNAPLVDALTAASDRDFPFTGAGHISAKARNLSGLRQGMT
YDKACCAADYDHTASREGERGAFDVGGHHGREALKRILQEREISRELVGSPTARKALIESI

-61-
KQVKAIERRYGVGDRIHVELARDVGKSIEEREERGIEKRNRQKDKLRGLFEKEVGR
PPQDGARGKEEMLRFELWSEQMRCLTYTDY1SPSQLVATDAAVQVDHILPWSRFA
DSSYANKTLCAKANQDKKGRTPEWFKAEKTDTEWDIFVRVEALADMKGFKKR
R>Rm.RNAEAAAARKFRRNMDTWRACRLALAEALQQLYPKGEKDGDGKERRRVP
RPGALTDRLLRAWGLQWMKSTKGRIIPDDRRHHALDAIVIAATTESSLQRATREVQ
EIEDKGLHYDLVKVNVTTPWPGFREQAAVEAVEKVVARARRARKAHDATIRHIA
VREGEQRVYERRKVALKLADLRVKDAERNARLIEKLNWIEAGSDKPPLPSPK
GDPFIEWRLVPDKSMALDTPNQKPRFGEVTDREGMARVDVFGRASKGKYYLYLVP
IYPHIATMKTPIRAQQAYKPEDEWPEMDSSYECWNLPMYIQVSSKGEIFEGY
YRGMNRSVGAIQLSAHSNNSDDVQVQGIGARTLTEFKFNVVDRFGRXHEVERELRTWR
GETWRGKAYI

[0096] YP_003968716.1 CRISPR-associated protein, Csn1 family (plasmid)
[Ilyobacter polytropus DSM 2926]

IVKYSIGLDIGIASVGWSVINKDKERIEDMGVFQKAENPKDSLLASSPJIEKRGR
RMIRIKKHRDLRKKMCLCESGLKVKNIEKIKYKNAYLSPWELRAKSKLEAKINSK
LLfflAKRRGFKSFRTDNRADDGTGLSSGIEKNKIMENEGYLTGDMSVAKDPKFNT
HVRNAGSYLFSRSRLLEDREVKIKQKAKQLGANTHFTDVLEKYIEVFNSQRFNE
GSPKPSPPYSEIQAIKMIGNCTFESSEKTRAKNTWSGERFVFLQKLMNFIVGLSGK
RPLTEETEDIEKEVYKKEVYKLKLRLKILEERFGLNLYSKDEQDKTEKTK
FISLIGWTTKDLNILLSKEIIEEDEKSKGLDKIKEILTFTKDKSITLKLLEISDIEIL
LSEEFSGLNLSLKAIKILPYLEKGLSYEACEKADYDYKNNGGKFRGELLPVVDK
DLIANPVLRAISQTRKVNAIIRYKGTPHTHIHEVARLAKSYDDQTIKENKKRE
LENEKTKKFISEEFGKVKGLLKYKLYQESEQRCAYRSKELSLSEVILDESMTDI
DffIIIPYRSMDSSDKVLVLSGENRKSNNLPKEYFDRQGRDWDFTVLNVKAMK
HIPRKSNNLKEEFTREDNKLWRAADTRY1SRFVANYALEYIRDSPKPRVF
MIPGQLTAQLARWRNLKVRENGDLHALDAAVVVATDQKAINMSNISRYELKNN
CKDVIPSIEHYADEETGEYFEVKDTRFMPSGFDLQKLESENPREEFYNLLS
DKRYLGWFWEGRFIEKLRPVFVSRMPNRRGVQGAHQETIRSSKSIASQIAVPLKPL
NSIKLKDLEKMQRGDSRDKLYEALKRNLEEEYDDPKAFKAEFYKPTNSKGRGPL
VRGIKVEEKQNVGYYVVNGQASNSMGVRIDVFRKNGKFKYTVPIYVHQTLLKELPNR
AINGKPYKDWDLIDGSEFELYSFYPNDLIEEFIKSKSISNKLTKTEIPVNLSEVLG
YYRGMNRSVGAIQLSAHSNNSDDVQVQGIGARTLTEFKFNVVDRFGRXHEVERELRTWR
GETWRGKAYI

-62-
[0097] ZP_09352959.1 CRISPR-associated protein cas9/csnl, subtype II/nmemi

[Bacillus smithii 7_3_47FAA]

MNYKMGLDGIASVGWAVINLDLKPvIEDLGVRIFDKAEHFPQNGESLALPRRIARSAR
RRLRRKKHRLERIRRLVSENLKEEMNNLFKQKKQIDVWQLRVDALERKLNNDE
LARVLHLAKRRGFRKSNRKSERNSESKSFLKNEENQSIAPYRVSVMIVKDSKFA
YHKRh^DSYSNM1ARDLREIKLIFEQREFNNPVCTERLEEKYNLWSSRPFAS
KEDIEKVKVGCFTFEPKEKRAPKATYTFQSFIVWEHIKLRLYSPDETRALTEIERNLLY
KQAFSKNKMRTYDRLKLNLSDDIIHKCGLLYDPKSSLKQENIRLFLEDSYHKIRKCIE
NVYGYKGDIRMFNETIDDFGTGYALTIFKDDEDIVYALQNEYITKNGKRVSNLANKVYD
KSLIDELLNLFSKFAHLMSMKAIRNILPYMEQGEIKYSAECALAGNYFENPGPKEKFLAN
LPVPMPANPVMRALTQSRYVYNAnKGYGSPVSIffELARDLSHFEDKRKIQKDQT
ENRKANNETAIQLIEYELTKNPTGLDIVFKLWSEQQGRCMYSLPIELRLELLEPGYV
EVDHILPSRLDDSANYKVLTKENREKGNHTPEYVLGLGERWKKEFKFVLAN
KFQSKKKQNLRLRYYEETEKEFKERNLNDTRYTSKFanFIEHKLFDADGGGQK
VYTINGKITAHLRSRWDFNKNREESDLHHAADVADIVACATQGMIKIKITEFYKAREQY
KESAKKKKEPIFPQPWPFFADEKARLSDKFQPQESIEAFALTNYDRKKELESRLPFPFVSF
PKRSTVAHQA HitELRRCVIGDEQSKEQTKVATKIKLDKDKGFIIPMYQKESDPR
TEYAIRQRLLEHNA^DPKKAFQEPLYKPKKNGEPGVIRTVKIDTNNKVVHLUGGKTV
AYNSNWRTDVFEDKGYCVVPYTMNGMKTLPNKAEKPKGSEYKEMTTEYTF
QFSLFPNLVRIVLPREKTIKTSTNEIJKDIFAYKTDATSAGLGLSLHDNRFLGTV
GSKTLKRFKEKYQVDVLGNIUKVKGEKRVGLAAAPTQKKGKTVDLSLQSVSD

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

MKYTLGLDVGIASVGWAVIDKDNKIIDGLGVRFCDKAESKTGESLATARRIARGM
RRRIRSRQRLRLVKKLFVQYIEIKDSSEFNRIIFDTSRDFGWKPWELRYNAIISRLKPY
ELVQVŁTHITKRRGFKSNNKREDLSTTKEVGVTISKNNSEMLRTKNYRTIGEMIFMET
PENSNNKRNKVEYHTITAREKLNLIEKYSIQRKLGSFVTEKLEHDFTLNIWEFQRPFA
SGIDSLSVKGVKCTLLKEElRAPTSCYTYEYFGSSLQSNLVLVENDNTLTDNQRAK
nEYAHFKNEIKYSEIRKDLDDIEPEILFKAFINLTHKNPSGNNESKFFYEMKSYHKLKST
LPiDiWGKLHS^SLDNLFYCFTVYKNDEIKDLYQA^WLDLIEYIAKLPTFNKF
KHLSLVAMKUIUFPMEKGYKYSACNMAELDFGSSKLEKCh^TVEPn^EVTNVP
VIRALTQARKVINAIQKYGLPYIVm^ELAREAGMTRQDRDNLKEHEHNRRKAREKI

-63-
SDLIRQNGRVASGLDILKWRLWEDQGGRCAYSKGIPIVCDLLNDSLTDHITPYPSRS
MDDSYMNKVLTVLDENQKNRSTYTPYEVGSTEKEDFDEARIYSMMLPSKEKQLL
NRNFTKDLDSIFSRNLNDTRYISRFLKNYIESYLQFSNDSPKSCWCWNGQCTALQRS
RWGLNKNREEDLHHALDAVIAACRDIKEITNTNYNENHENHNYKVKYPLPHWSF
RQDLMETLAVCFRADQPARRKTGPFIDETRSPKHNGLTSVKPLTVTLKLETMV
KNTKGGSIDKAYIWLKNRiEHNKNPLKFAEKIYKPLKNGTNGAIIRSIRVETPSY
TGVRHNEGKGSNDNLVRVDFVFKKDKYYLVPYYVAMIKKLPSKAIVPLKPSQ
WELDSTHEFLSLYQNDVLVTTKKGITEGYRSCmGTGLSLMPHFANNKnKID
IGVRTAISIEYKYNVDSIKSVGEPPRGMEKYNFSKSN

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

TPSY

MAQHVFGLDGIASVWAILGEQRIIDLGVRCDKAETAKEGDPLNLTRQRANLRLRR
RLYRRAWRLTLQRLRKRKKGLIADALKFAKAPSYGDSAWELRQGQLDLLTLTPLEWAR
VIYHQCCHKHRFWHTSKAEEAKADSAEGGRVKQGLHTKALMQAKNYRSAEMV
LAEFPDAQRNKRQGYDKALSRLVGLGEELALLFatQRRLGPHASDFEFKLILGDGDR
KSGLFWQQQKALSGADLLKMLGKCKTFEGYRAPHKSAFSLTHKLRIVLV
DGRNPLNEAERQAALLLPYQTETSKYKTLKNAFIKAGLWGDGVFGGALAYPSQAQI
DAEKTKDPEDQFLVKLPWHELRLKAFKAAGHEALWQQISTPAIGDPTLDQATV
LVSYYKDGAEVQVQLRQLALPAAASIAVLEKISFKDFSSLKALRRIVPLMQSLRY
DEAVAQIPEYGHSSHQIEPGAALAKHYLYPPFYEAQRKYAGKDIIGMSMQFIDDAPR
NPWLRAKNQARKVNA#EYGSIAVMEMARDLSPDLERNKVRAQEEFRDRN
DRARSEFDGFKPCA#AAIEKWM#YLEQQLGQC#YSQQPLD#RVLDDHYAQV
HAPLSRSYDDSKNKNKLVLTQENQ>IKGNTA#YLSFPDGERWRTFVAWV
QGNKAYMRAKNRLRLKNNYGDESQGFDRLN>TOTRYICKFFKNWEEFDQLAAR
ADGDTARRVCVYQVQLRFLARAWGLTRVQGDSRDHHALDAVAAACTHGMVK
ALADYSRREISFLQEGFDPETGEILNPA#ADRARQHFPEPWHFAHELKARLTDD
LAALREDMQRLGYSYTEDLGLRLTFVSRADFRRSEGAVFIKETYAQPESLQKQGG
VIEKJILTSLQLQDFDKLNPJESNDHVFEPHRNERLYAARIQRELQFGRADKAFGPD
NFLHKPDKNQPTGPVVRSLKVRGKTQGIPRGLAKNDSMLNRVI#FTKAGKFHHLV
PVYVRHR  VI#LPRNRAIVAKDEEDWTLIDE#SFALFSVPNDYVKVTLKKEQQSY
SGADRSTGAMNICALWADRAASVGDGLIRGIGKTVLALSVEKFNVDLVGRILAPPET
RSGLA[0100] YP_002342100.1 hypothetical protein NMA0631 [Neisseria meningitidis Z2491]

MAAFKPNPINYLGLDIGIASVGVAMVEIDEDENPICLDGVRVFERAEVPKTGDSSL
AMARRLARSVRLTRRRAHRLRARRILKREGVLQAAFDENGLIKSLPNTPWQLR
AAALDRKLTLPETWSALHLHIGLYRQKNEGETADKELGALKVADNAHAL
QTGDFRTPAELALNKFEKESGfIRNQRGDYSHTSRKLQAEILLFKEKFGNHV
SGGLKEGICTLLMTQRPALSVDHAVQKMLGHCTFEPAEPKAAKNTYAERFIWLTKLN
NLRILEQGSERPLTDATERATLMDEPYSKSLTYAQARKLLGLEDTAFKGLRYKGD
AEASTLMEMKAYHAIASALEKEGLKKSKPLNLSPLEQDGIEATAFSLFTKTEDITGRL
KDRIQPEIEALLKfIRFSDKFIQISQLKLALRRIVPMLMEQGKRYDEACAEIYGDHYGKKN
TEEKILIPPADEIRNPVVLARLSQARKVINGVRVVRGSPARIHETAREVGSFKDR
KEIEKRPQENRKRDEKAAMAFREYFPNFGEPSKDIKLRLYEQQHGKCLLYSGKEI
NLGRLNEKGYVEIDHALPSRTWDDSFNKVLVLGSENQNKNGTQYFENFKDNS
REWQEFKARVETSFRPSKKQIRLLQKFEDDFKERNLNDTRYVNFLCQFVADRME
RLTGKGGRRVFASNGQITNLNRGFWGLRKVRAENDRHHALDAAVVACSTVAMQQ
KITRFVRYKEMNAFDGKTIDKETGEVLHOKTHFQPWEFFAQEVMIKFGPKDGKPE
FEEDATPEKLRTLAEKLLSRPEAVHEYVTPLFVSRAPNRKMSGGHMETVSKAKRL
DEGVSVLRVPLTQLKLKDLEKMNREEREPLKLYEALKARLEAHKDDPAKAFAEFFYK
YDKAGNRTPQQVAKREVEQVKTGYVVRHNPHGJADNATMVRVDVFDGKDKYLYLV
PIYSWQVAKGILPDpAvVVQGKDEEEDWLDDSFNFKSLHPNDLVEVITKXPAPvMFYG
FASCHRGRTGNINIRIHDLDHKIGKNGILEIGVKTALSFQKYQIDELGKEIRPCRLKKR
PPVR

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

MQTNLSYIYLGLDIGIASVGVAMVEIDENPEGLIDGVRVFERAEVPKTGEDSLALSR
RLARRSTRRLRRRAHRLLRRAKFLKREGLSTIDLEKGLPQAWELRVAGLERLLEAE
WGAVLLHLIKmGYLSKRNESQTNKELGALLSGVAQNHLQSDDYRTPAELAL
KKFAKEEGfIRNQRGAYTHTFNRLDLLAELNLFAQHQFGPMCHKEDIHQYMTEL
LMWQKPALSGEILKMLGKCTHEKNEFKAARHTYSARFVVLTKLNNLRILEDGAER
RALNEEERQLLINHPYESKSLTYAQVRKLGLSEQAIKFHKRYSKENAESATFMELK
AWHAIRKALENQGLKDTWQLAKKDPDLDEIGTAFLYKTDDIQYLTNKPNSVI
NALLVSLNFDKFIELSLKSLRKLPLMEQGKRYDQACREIYGHHYGEANKTSQPLL
AIPAQEIRNPVVLRTLSQARKVINAIIRQYGSPARVHIETGPvELGKSFKERREIQKQQEDNRTKRESAVQKFELFSDFSSEPKSDKDLKFRLYEQQHGKCLYSYGENINKHLNEKGEWIDHALPFSRTWDDSFNKVLVLASENQNKGQTYPYELQGKINSERWKNFVALVLGSQCSAAKKQRLLTQVIDDNKFIRDRLNDTRYIARFLSNYIQENLLLIVGKNNKNVFTPNGQTALLRSWLGKIAENNNRHHALDAIVVACATPSMQQKITRFIRFKEVHPYKIE>mYEMVDQESGEEnSPHFPEPWAYFRQEWIRVFDNHPDTVLKEMLPDPQANHQFVQPLFSRAPTRKMSGQGHMETIKSAKRLAEGISVLRPLTQLKP^LENMVNKEREPAylaGLKARLAEFQDPAKAFATPFYKQGQGQVKAIRVEQVKSQGVLVRENNGVADNASIVRTDVFIFKNNKFLLFPIYTQVAKGILPNKAIHAVKENDEWEEMDEGAKFKFSLFPNDLVELKTTKKEYFGGYGIGLDRATGNISLKEHDGEISKGDGVYRVGVKLALSFEKYQVDELGKRNQICRPQORPV

[0102] ZP_07738815.1 CRISPR-associated protein, Csn1 family [Aminomonas paucivorans DSM 12260]

MIGEHVRGCCFLFDHHTPWGAFRPLNTVRTFTAENPKDGSSLAEPRRQARGLRRRLRRKQRLEDLLRLRAKEGVELESDLETFTRETAPDKPYLRAGELDRLPSFPEWVRLVYeiTKHRGFQSNSNNPVQEDGERSQEEEGKLLSSLSVGSURRENTLREGGYRTAGEMLARDPKFQDHHRNRRAGYSHTLRRSLLLEEARLRFQSRQRTLGNPHASSEKLEELFHLVAFQNPFAASGEDLRNKAGHCSLEPDQIRAPRSASAETFMLQKGTKNLRIHRRTGEERPLTDKEREQIQHLLLAWKQKTVHTKLRHLIEPEELWFTGLPYFIRSDKAAELKLFVHLAGIHEIRKALDKGPDPAVWDTLRSRRDLDISDIALTFYKNEDEILPRLESGLSPENARALPLSFTGTAHLTSLALGKPLLHEEGKSYTQRADAGYAAPPPDRHPKLPPEADWRNPVFVRALTQTRKVVALVRYPWQCIHLETARLSQPAKYRRRIETQQAENKKQPAEFLDIVGTAPPGDLLKMRLLWEQQQGGFCPYCEEYNPTRLAEPGYAEMDHLPSRSLDNGWHNRVLVHFKDNDRKGNRTPEAEFGGTARWDRLVAWVAQFILSAPKRNREDFGEEAERLKDRLTVTDFRITKTAATLDRDFTHEPAKDPVMTLNGRLTAFLRKQWGLHKNRKNGDLPNHIDALADDALAVASRSFVYRRSSHAAWGELPRGREAENGFLPYPAFRSEVLARLCPTRIERLLRLDQGQGVGYDEAFRNGLRPVFVSRAPSRRLRGKAHMETLRSKPKWDHPEDGPRTRASRIPKDLNLEKREMVKGDRDRKLYEAIRLARLAEFFNGKAKFVAPFRKPCRSGEGPLVRSRIFDSGYSVGRVLDGGEVYAVADHESMRVDVYAKKNRFYLVPVYADVARIGKNNARvAHKSEEEDWLVDSFDFRSFLFGDLVEIEKDGAYLGYKSCRHGDRGRLLDRHMRPRESDCGTFYVSTRKDVLSMCKYQVDPLGEIRLVGSEKPPFVL

-66-
[0103] ZP_08574780.1 CRISPR-associated protein, Csn1 family [Lactobacillus coryniformis subsp. torquens KCTC 3535]

MGYRIGLVDVGTSTGYYAVLKTDKNLGLPYKILTDLSDVYIPRAENPQTGASLAEPRIKR
GLPvPXRTVRTKFPvKQRTQQLFIHGSLLKSEGEQILATTPQAKYSVYELRAGVLDRRTL
SELFRLVYFIFGHGFKSNRKAELPNEADKKQMGGQLNSIEERKAIAEKGYRTVG
ELYLKDPKYNJDHKNKGYIDGVLYSTPNQRQMLVDEIKQILDKQRELGNEKLTDEFYA
TYLLGDENRAGIFQAPAQRDFDGPGAPGAYGDQIKKMKVDIFPEPTEDRAAKATYTF
QYFNLLQKMTSLNYQNTTGDWTLNGLDRQAIDAVFAKAEPKTYKPFITGEL
PJO.LKLPPDARFNLVNYGLQTSQEIETVEKKTRVFDKAYHDLKVLPWEMWQSR
QLLDfGTLATLTYSDKJURRRYFAELNLPAELIKEKLPLWSKFGHSISKMQNIPYL
EMGQVYSEATMtYGDFPvKQQISKDTIREETNPVRRVAVKTIKIVEQIRRGGKPDG
INIELARELNGFRAQORQKDJOQNTDKIAELTELGIPNVQNIYRLKHLKQE
NGVDPTGQDFPFAFARSEQYEVHDHPJSWINCSNVTLSAKCNREKGNRMPV
YLAMEAGQNLALTNNADMRNSRKKQRLKQLSDEELKDQKRMDFTFLRL
YNYQRFQAIENPELEKQVRPLNGLGEVTSSRWSGKVRGDLKGDHDAITVIAA
TPKFQFQVTKYSQHQVEKNNQLWLHDAEIKDAYSAAEAQRMADLFNKFNGFPG
WPEFLDELLAPJSDPVMEMKRSWNYTIPIEIAKLRPPFPVRINHKISGPAPHLTDI
RSAKLFDGERGIVLSRVSITKLKINKKGGQVATGDGIYDPENSNNGKVVYASQALEA
HNGSGELAFPDGLEYVHDGTKLWKLKRVAKVSLPRLKAADNMSVRID
VFNPDKKQKVFPVPIYIKDTEQVLPNCQAROXKLWQITEQDECQFCSLYPGDMHIES
KTGKIPKYSNKENNTSVVPIKNFYGYFDAGIAITASILVRAHDSYRTASIGAILKF
EKYQVYDVSFYRYKHEKXRQLFVRDE

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

MNAEHKGKLLIMEENFYRIGLDIGITSVGWAVLQNNSSQPEPVRDLDLGVRFIDVA
ENPKNGDALAPJIDARTTRRRLRRHHRLRERIKFLLQENGLEMDSFMEYYKGN
LPDVYQLRVEGLDRKLKEDELAQVLHIAKFIRGFRSTKAEKKEGAVLKTATTEN
QKIMGEGKYRTYSGEMLYEAFHECLWNEKGYVLRTNRPPDQKYHTLRSMLV
VEVHAIFAAQRHGNQKATGELAAYVEIMTSQRSFMDGPLQPQDPKSPYAMEGFGD
RVGKCTFKEKDEYRAPKATYTAELFVALQKINHNTKIDFETGGRFFSEEERK
111GLLLS
SKEKLYGTRIKLNIKIPSKFKNLNYSAKKEGETEEERVLDETKAKFASMFWTYEYS
KCLKDTEREMPVGEKADLFDRIEILDAYKNNDSRSSRLKELGLSGEIDGLDLSPA
KYQRVSLKAMRKMQPYLEDGLIYDKACEAAGYDFRALNDGNKHKLLGEE1NAIV
NDIT^VVKRSVQTICNVAHQCQYGSQPAVNLAREMSKFNQDRTNLEKEMMKRQ
QE>ffiRACKQIIELGKQNPTGQDLKJRLLWNDQQYCYLSGKIPLEELFDGGYIDHID
LPYSITFDSSYRKNVLVTAQENRQPQGRNRPYEYFGAEEKRVEDYESVRLLRVRDYK
KQQKLKKNNFTEREERFEKJERNLNDTTTRVYNNIRQNLELFNHHPEKKKQVV
AVNGAVTSYLRKRWGLMQKDFSDRHHAMDAWACCDTGMIKISRYMQGRELA
YSRNFKFPDFETGELNRDNTFREQWDFEGKVPLWNFSRDELJIRLLLNEPDKN
FLLTHADVQELDYPGWYMGEESPIEEOGRNYIRPLFVSRPMNPKVTGSAHDATI
RSARDYETRGGVITKPLTDLKLKDNEIEGYYDKDSRLLYQALVRQLLLHNGG
NKKAFAEDFHKPKDAGTEGPVVRKVIKIEKKQTSGVMMRGGTIAANGEMVRIDVFR
ENGKYVFVPTADVVRKLPNAATHTPSEWVRMDANVFLSRLYSDRILHVK
KDDIKTNLVNGGLLQLKEIFAYYTGADIATASIAGFANDSNFKRGLQISLEIFKECQ
VDILGNISVVRHENQF

[0105] ZP_10953934.1 HNH endonuclease [Alicyclobacillus hesperidum URH17-3-68]
MAYRLGLDIGITSVGAVVALEKDESGLKPVRIQDLGVRIFDKAEDSKTGASLALPR
REARSARRRTRRRRHRRLVRKRLLEEKGILSMQIEALYAQRTSSPDVYALVAGL
DRCLIAEEIIAVLHIIHRRGFQSNKRSEIKSDAGKLKAVQENLENMQSKGKRT
AEMLVSEATKDAEGKLKVHGKKHGYVSNVRKAGEYRHTSRQAIQDEVRKIFAA
QRALGNDVMSEELEDYSYKILCSQRNFDGDPGDSGYHGSVPDGVQSIYERMV
GSCTFETGKEGKRAPRSSYFQQLTTRVNLRIYQQEDGGRYPELCTQTTERAVIDC
AYEQKTITYGKLRLDDMDKDTESFAGLTYGLNRSNKTEDTFTVEMKFYDFDEVRKAL
QRAQVFQIQLDIELDQIGWLSVWKSDDNRRKLSTLGLSNVEELPLNGSKFCH
LSLKAIRKJLPFLEDGYSVDACEALAGYQFGKTEYVKQRLPLGEGEVTNPVVRR
ALSAQIKVNLAVIRKHGSEPESIIPELARSKNLDERRKIEAQKENQKnQIkiDkE1R
ELIGSAHTGDRIVKLFKQQEFCMYSGEKLDVTRLEPGYAEVDIHPGIYSFDD
SYDNVAKLVTQENRQKGNRPYLERLDRKPEQAKFIALVESPLSQQKKNHLMIDK
RAIDLEQEGFRENLSDDTYTRALMNHQAWLLBETASTRKRVCVNGAVTAY
MRARWGLKDRDAGKHDHAADAVVACGDSLlIQRTKYDFKKNALADNRXYV
QQVSXSKSEGITQVYDKETGEGVFTWESFDERKLPNPELWVFPFRDELLARLSDDPSK
KIRAIGLNTYSETEQIPDFIVSRMPTRKVTGAAFEKETIRSPRIVKVDNNKGE1EIQWVSK
VALTEKLTKEDKGYDFRPPEDDRLYNTLRELVQFGDDAKAAFKEPYVKISKDGS
VRTPVRKVIQKELTLGVPVHGGRIAENGGMVIRDFVAKGKYYFYVPIVADVVK
PvELPNPJATAHKPYSEWRVVDSDYQFKFSLYPEVDNWAVMIKPSREVDITYICDPvKEPVG
CRIYMVFVSAASISASILRTHDOSGELEGGLIQGLEVFELVYVVGVPLGGDTUPVYKERRM
PFRVERKMN

[0106] ADI9058.1 uncharacterized protein conserved in bacteria [uncultured delta
proteobacterium HF0070_07E19]

MSSKAIDSQEQLDFKQyteYLTGLDLGIKSIGWAILSGERIANAGYVLFETAELNSTG
NKLI SKA AERGRKJRRJRRMLDRKARRGRII R YLLEERGLPTDELEEV V VHQ SNRTL W
DVRAEAVERKLTKQELAAAVLFLYRvmGYFPNTEKXLLPPDDEDSADDEEQGKINTIATS
PvLREELKASDCDTGIQQFLAQNRDRQRNREGDNSLMLCKVFYFKEALQIFQRKQG
HESLDFEKTYLQDMQGRSRSPLGLCNGLSIPSELRAAPPSTIIIVFKLQNLGLNQ
ISNAYREESWSDAPRAQIADDACIIQRSTSSSYQIQRIFQIDPVEFNLNVYERRDLPDV
DLQELYQQEQERXTLA>FRNWKQLEKIIGTGHPIQTLDEAARLIITLKDDEKLDSDLAD
LLPEASDQAIQQLCLEDFTTAAKISLEAMYRILPHMNQGMGFDDACQQESLPEIGVPP
AGDRVPPFDEMYPNVNVLQSRKLNIVADEYGMPAKIRVELARLDLKGRELRE
RIKLDQLDKSKQNDQRAEDFRAEFQAOQRGDSLQRLWKEQNCTCPYSGRMPVPN
SVLSEDQTQIDHILPSIQFDNLSNKLCEFTEEANQKSNRTPFEYTLADAFQRLAEISG
NWPEAKJINLKLHKSFGKVAEEWKSRLhTOTRYLTSALADHLRJHHLPSDKIQTVNGR
ITGYRLQWGLEKDRDKHHTHHAVIDIAIVACTTPAIQVQTVLYHDJIRRYKKLGEKR
PTPPETFRQDVLDEVEEIFTRQPKVSQGIQTKDLRLHRSKPRQDPRVALTVKLA
DLERLVEKADSNRNLYEHLKQCLEESGDQPTKAKAPFYMPGPEAKQRPILSKTL
LEKPEPPKQLTTELGGPJJYDSMAQGRDLJYRYKPGKJRDEYRWLQRMIDLRG
EENVYHFQKGVPYDQPEIQNYTFLESLYFDDLVFQFQSADESVIRGGYRTFNIANG
QLKISTYLEGRQDFDFGANRLAHFAKVQVNLLGKVIK

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

MGNYLGLDVGIGSIGWAVINIEKRIEDFNVFRIFKSIQEKNRNSASQCRSCRSEG
LRRLYRRKSFIPvKLRRKN тысLGLTSEKIdYYETADNVIQLRKNKLSEKJLTRPEIA
ACLIffICNNRGYK)JFYEVNVEDEDIPDERNEYKEEHDISVLISNLNMEGGYCTPAEMI
CNCREFDEPSVYRFHNSASAKNYLITRMLVKEVDLILEQSKYIGILDDKTI
KJKDIIAQVPvDFIEIGPGKNEPIYRRFTYLDSIQCQFFKDQERGSRTPVIADIYAFPVNV
LSQTYTTONGKESVTDTSFANDLINSALKNGSMDKRELKAIASKYffDISSKNSDTS
TKCFKYYKVKLFEKYGYDWSDKLJENYTDNDNVLNPJGIVLSPQAQPFRREKLK
ALNIGLDDGLINEKTKLSSLGTANVSYKYMQGSlEAFCEGDLYGKYQAKFNKEID
ENAKPQKLPPFNEDDEFFKVPVFRSINETKLILNIDKGYPPAAVNIETADENK
TFEDRAITKRN>JDNNKIIIDRIVEIECICKICDHEVHARLEKYLWLEAEPQGKCLYS
ETITKEDMLPJJDKIK.FDVDMIPYSILDLNANTINMCALYAEENQKKGQRTPLRARYMNE
AQADYVRVNTMFSSKXCSKKEYQYLYMLPDLNQELLLGWRKSNLDDTRYICK
YLVM^LRRNLPRDSYESSEDDEELKLIRDFHTYRVFPVKSRFTSMFRRWLNKETWGR
YDKAEKLKLTLDHAADDIALIANCRPEYVVLAGEKLKLNKMYQAGKTPPEYEQS
KKACID>hYKLFPvMDRTRAETKLLSGHGLTPnP ^SEEDVRLKDLKNIYEQFWKD
KDKNCEELYRENVSLYKGDPKFASLSMPVSLKPDH2GYRTITGEEAIRVKEDI
GLIKLKRKSIEITAESINSYIITDDKILIDSIFLTFE2QADYKVDGLYKKTQHFFTT
S GGKRNKTVIEKVPWSRLKEIIEDDNNFSSLLNDSSYYCIELYKDSKGDNNLQGIA
SVIVFTORKTCJLYKldFWDYYTHVMYIFPGYDLRIKSTSSQSEQLKFEGYFISV
K NVNENSFmSDNPKCAKDRYSITKDGKLADLVLSMGVKVQGNNQGKISCGEPLSL
LKEKN

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

MIMKLEKWRGLDLGTSIGGSVFSLKDNSVQLIDMGVRIFSDGRDPKTKEPLA
VARTARSQKRIYRRKLRRQKFV4LDREQQLDPKTEECMTLKSLNPYELRIKALDE
KLEPYELGRAFNLAVRGRFGS>niKDGSREEVEKSPDEIKTQAMQTHLEKAI
ENGCRCTITEFLYKQENGGRFAPGRMTYYPTRKMYEEMFNLIRSQQEKYYPQVDW
DDIYKAIFYQRPQLPQQRGICYIYENDKERTFKAPCSQKLRLQDIGNLAYYELO
KRVELNDQKVLIELLYSNDKKIVTDQMRKALCALSNSFNEENLRDFLIGNPTAV
KMRSNKRFQKLGWDEIPLEQDLIIETIITADDEDAVVEYIKYDLTQERQDFIVN
TIL QSGTSMCKEVEKSWLVRLEELDKYHEAVSLGYKFADQTVEKYDLPPYGKVL
PGSTMEILDSAPETNPKEHYGKSNQPETVHVALNQTRVVNALEYGKPSQIA1LSRD
LK>nWEKKAIEARKQNKQMRAKEMAJTIASLYHTAFPGKSFYMINDRMKYLRLWSE
LGLGNKCICYCGKIGSAELFTKEIEEJHELPSRTLDAESNLTVAHSSCVNKAEPSF
EAFGTNPGSWSQIEIQQRANQLKNTSKKKNFKSPAMDFEKSSFIARQLSDNYAZK
AALYRLKCLVENPSDVTGNTMSMTKLLRDKWMDSILCRKFEKEVALLGLKPEQGI
GYKNNRFDHHRHAIDAVVIGLTDRMSVQKLATKNSFCGNRJEIFPEPFLRSIDLIEK
VKMVSFKPDHGAEGKLKSKETLLGKIILHGTKETFVCRENIVSLSEKNLDDIVDEIKS
KVKDYVAKHKCGQKIEAVLSDFSKEGIKVRvjVQRVPCEITSKISRYLSPEYFA

-70-
AVIWEIPGKETKFAQYIRRNEVEKNSGKLNVKPAVLENGKPHPAAKQVCLLHKD
DYLEFSDKGMYFCRIAGYATNNKLDIPVYAYSVCDWINSTNETMLTGTYWKPT
PTQNWVSVNLFDKQKARLVTVSPIGRVFPvK

[0109] ZP_11150502.1 CRISPR-associated protein, Csnl family [Alcanivorax pacificus W1-5]
NmYRVGLDLGTASVGAAYFVSEMDEQPIMEYWHERYLFLSEPLVPMGQLKPKAA
RRLARQQQRIDRSSRLRIAIYSRRLGAPRNDGSVHGNDVPTRLRAMAVERIEL
GQLRAVLLRMGKKGRTYGFTKAVRVGEAGEVAGSARLEEEMVLSAVQNKS
VTGYEYLAARVEHLPSKLKVAANNEYAYAPEALFRQYLGLPAIKGRPDCLPNMYA
LRHQIEHEFERIWATQSFHDVMKDHSVKEYIRNAIFFQRLPKPDADKVGRCSLQTN
LPRAPIRAQIAAQNRIEKMQADLRWGMGRRAMEINHDBAKAVIRELLNNQQKELSRFK
IYKELERAGCPGEPGKGLNMDRAALGGRDDLSQNTLAARWKLGLERDQWQELDEV
TQIQVINFLADLSPEQLDDWDDSCFRMGKNGPRNSDFVAFMELRMNTDFRDL
LSKMFDGGRSSSYISIKALATEWMAIPHRETETHRVDEEAIRECYSEPATPA
QGGQSKLEPPPLTGEVDVNRQVRHHTINMMIDLGSVPQAIWEMAREMKGGV
TRRNDIEKQNKFASERKKAASIESENGTTPPARILYQLWIEQGHQCPYCESNISL
EQALSNGTNYHILPRTLQIQGRKRSHELVLARHCNEDEKGENETPYQAFGHDDRRWR
IVEQRANALPKKSSRKLRLIKLDGEAGALTDESIDFADRQLHESSWLAKVTTQWLSL
SSLGSDYVSRSGLTAEILRRWGLDTVIPQVRFESGMPPVDEGAEAITEFEEKFRLQ
WEGRHVTREMRTDRDPKDRHHELVLDAITVALTSRSLYYQQYAFAWKVADEKQR
HGVDVKVVELMPMILTDIALEAIVRVSIRHPDPDGRFEDAYGAIAQRRLDERS
GEKVDWLRVRSLTDLAEPKSIDVVKVLANISRVIQEAIRLHISNIFEKRVSKGMTQP
QQAALREPQEFQGMLRKVCYADDDCVRIEHSSHRGHHMcYNDFAYMEVPCKE
EGILYGVPNLRVSEAYGVKRAPESGDFIRFYKGDTPKNKTRGQYETIKQILGDGGG
KLILTPETKPADLLASAKWGRKLGVGRNNIFILLRLCAE

[0110] ZP_18919511.1 hypothetical protein C882_0672 [Caenispirillum salinarum AK4]
MPVLSPLSPNAAQGRRRWSLALDIGEGSISGWAAYEVAEGRVLQLTGTVTLPSPA
WSNENGTVAHGAADRAVRSEQQQRHDSRRRLAGLRLCDPVELSPEEDLKLILTR
TPPKADPRAIFLRAARPLDGPELFRVLHMAAHGRGRLAEQVDEPPESDAD
DAAPAAEDEDTGTPRAAADADFAFRRMLMEFMHRHGHTQICGEMAGRLRETPAGA
QPVTBAMLGRLVGGGVAVPTRALIEQFEDAIRIAQAHPDLPWDSLRLVLQAPI
-71-
AVPPATPCLFLEELRRRGETFQGRTITREAIDRGLTVDPLIQALRIRETVGNLRLHERIT
EPDGQRQRYVPRAMPELGLSHGELTAPERDLVRLMDPDGLAAKDGPRIPYTRLRK
LYYDNPVCQFERDTSGGITVNPONDMLARWIDGWVLPKARSILYVRDVVAR
GADSALARRLAEAGAHGVPVAAAAPAAATAAILESID1MQGGRYSVCPWAAAILD
AWANAPTEGFYDVTGLGFGAPGEIVLEDLLRRAGALLALHPRTMAAARTPNRAA
QRGPLPAYEVSVIPSLTSLLRAADWSAADPEERPFLRTWTGNAATDHILN
QVVRTANEVITYKNRRGWDPDLSRITVELAEREAKHQGVRRENEKAIKENRENGRRKK
ESAAADTFQDNTVSWQAGGLPKERALRLALQRQEPFPCPACYARPKRLATDFFSP
AETEIDHVIERRMGGDGPDLVLAHHDWNNAKGKKTPEHAGDLDLSPALALWQ
GWRKENADRLKKGKHKARTPPxEDKDFMDRVGVRFEEDARAKAENQERRGRRML
HDTARATRALLRLYAAAAVMPEDPAEIGAPVETPSPEDPTGYTIAIYRTISVRQPVNG
SVTHMLRQLRQLRQKRNDYQTHHAEDACLLLALAGPVQQAFNTEAAQHGADAPDD
RPVDMPTSADYHQQRARALGRVLPATVDAALADIVMPESDRQPETGRVHWRL
TRAGRGLKRIDDLLTRNCVLSRPRPSETGTPGALHNNATHYGRREITDGRDTILVT
QRMNARLLVALDLNAKIPAALALDAAPGDTILKEICIEADHRDVRVDPVTHARR
WISARLAALVPAHAEAVARDIAELALDIALDADRTPEQEARSSRLQPSYLGRAIS
AKXADGRARAREQELTRALLDHPWGPRLRFILIMREALAPSLLVRIRANKTDAFGRP
VPAAVVWKTDGNAVSQLWLRSTVTDGRIPLKIEKRIEINSLEYARLNGLDE
GAGYTGNNAPPRLQRIDDLTPLWRDHHTAPGGYLTAVGELSEKARSALRGKA
MRQTLTDAGITAEAGWRLDSEGAVCDLEVAKGDVTKKGKTYKVVGVTIQGIGMP
VDAAGSAPRTPEDCEKFEQYGIKPKWAKGIPLA

YP_425545.1 CRISPR-associated endonuclease Csnl family protein
[Rhodospirilhim rubrum ATCC 11170]
MRPIPEWILGIDTDGLGWAVFSCEEKGPPTAKELLGGGVRLDFSDRDAKDHTRQ
AERGAFRRARRQTRTPWPPRDRLLALFQAAGLTIPAAETQIALARREAVSRPCL
DALWAAALLHLHHRGFRSNDKRERRAAAKALAKAKPAKATAKATAPAKEADDEA
GFWEGAAALRQRRMAASGPATVGALLADDLDRGQPVRMRYNQSSTDGTVVAPTRA
LIAELAEIVARQSSAYPGDLWPAVTRLVLDQRPRLSKGAGPACFLPGEDRALRalph
tVQDFIIQRTLNLRLPSADEPRLPTDEEHAKALALLSTAFVEWPAVRRLRALGLKR
GVKFTAETERNGAKQAARGTANLTLATAPLIPGWSGDLDKRDRVFSDLWAAR
QDRSALLALIGDPGRPTVTEDETAEADAIQIVLPTGRASLAKAAAARAIQAMAP
GIGYDEAVTLALGLHHSRRPRQERLARLPYYAAALPDVGLGDGPVGPPAEDDGAA

-72-
AEAYYGWGMSvfifALNETRKIVNALLHRHGPIRLVMVETTRELKAGADEPvKRMIA
EQAERERENAEIDVELRKSRWRMANAERRQRVRRLARRQNNLCPTYSTPIGADLL
GDAYDIDHVILARGGRDSLNMVLCQSDANKTGKDKTPWEAFHDKPGWIAQRDD
FLARLPQTAKALAWRFADDAGEVRVKAEDAEDQGFLPRQLTDGTIVARVALRYL
SLVTNEPNAVTNGRLTGLLRLAWDITPGPAPRDLTPLPRDLRSDTAAARRFLDGL
TPPPLAKAVEGAVQARLAALAGRSVRDADAGLADGLSTLALGGGGKRRHARDHRHHF
DAAMIAVTTRGLINQINQASAGRIRDLRWRPRTNFPFPPYPFTAEMVKMQWDfifIHSI
RPAPIHRDGGLHAATFGVRNRPDLARVLRQRPVEKFLFDANAKPLPADKIAEIDDG
ASPRMAKRFKALLARYQAHAHPVPPALLAALAARPDAPFPRGMPANTVIAGRSDG
DGEDAQLITPFRANPKAARVTMGNAYEVEWIEQVKGPRWSRRVLTTRFDRTQPAPP
PPENARLVMRGLRGLVYYPLESGDRLFLVKKMADVGLRLWAPARLTOGTAKATL
YAQLSCPNLNGDQGYCVQSAEGIRKEEKTTSCTALGRRLSKKAT

[C0112] CCA84553.1 conserved hypothetical protein [Ralstonia syzygii R24]
MAEKQHRWGLDITNSGAVIALIEGRPAGLTVATGRSFSDGNPDKDGSSLAERR
GPRQMRPvJUUDRLRRRDFMQALrNVGLMPGDAARALKLVENTPYVLRQRLGDQA
LTLPEFGRALFHNLNQRGFQSNRKTRDRATASEGSKKNMRKAARAGMNGNARTVGEA
LARRLEDGRPVRARMVGQGKDEHYELYIAREWIAQEFDALWASQRFQHAELVADA
ARDRLRAIIFFQRRKLPPVVPVGKFLEPNQPRVAAALPSAQRFRLMQLNHLRLVMTLA
DKREPLSFSQENLLAQVLPARPKCGFDMLRKTFVGANKEAYRTIESERKELKGC
DTAAKLAKYNALGTRWWQLSLEDQDRLLCQLLDGDENDAVLADALREHGLTDQI
DTLLGNSFDGHMRGLRSALLRLVDAESGRDEQGLPLSYDKAVVAAAGYAHTADL
ENGERALPYEGYELLWRYTQDAPATKTIOTAERKFGKIANTPVHIGNQLRKLVNALI
QRYGKPAQIVVELAPvNLKAGLEEEKIRKQQQTANLERNRIRQKLQDAGVDPNREN
RLRMRLFEELGQGNGLGTCPICYSQRQISLQLRFSNDVQVDHILPSTKLDSDFANV
AQHDANRYKGNRPFEAGFANRGDYAWDDIRARAALVPvNKRNFATAMQDWL
HNEDFLARQLDLTAYLSRVARQYLTAICSKDDVYVSPGRLTAMRLRAKWKGLRNVL
DGMMEQGRPAVKMRDDHHHAIDADAVVGATDRAMLQVATLAARAREQDAERLI
GDMPTPVNFLEDVRQVARCJVSHKPDHPGPEGLHNDTAAYGIVGFPDGRYR
RHRVSLF DLKPGDLSNRCDAPLQAELEPIFEQDDARAREVALTAERYRQRKW
LEELMVLPIRPRGEDGKTLPSADYPAKYKGDSNYCYLEFINERGRWDGELISTFRAN
QAAAYRRFRNDPAPFRFYTAGRPLLRLCINDYIAVGTAAERTFRVVKMSENKITL
AHEFEGTLKQRADKDPFKYLTKSGALRDLGARRIFDLGIRVLDPGIKGD
CRISPR-associated protein, Csnl family [Rhodovulum sp. PH10]

MGIRFAFDLGTNSIGWAVWRTGPVGFEDTAASLDGSXVLIFKDGSRNPDKDGQSLATMPJIVPRQSRKPUFVFVLRRDDLALLARKAFLPVDVEEGRRLAATDPYHRLRAKALDESLTPHEMGVRFLNRQGRFSNKRKADRDQREGKIAEGRSKLAEATLACTGTFSLWSRHRGTPRTRPSRTRIMEGEGAKALAYFYPTREMVAEERLWTQSALPAPLTPERHEEIAIGILFRQDLPKIGCTFEPSERRPRLPALPSVEARGIYERLRAHLRIT TGPFSDRGLTRPDEVLSALLAGKSLTRKAVRTKLKILPHALVNEEPSNEAEKGLDGAL TAKLLSKPDPHYGAAWHGLSFAEKDTFVGGKLLEDADEELRIRLVTENRLSEADAR RCASICPLADGYRGRTANTEILAALVEETDETGTVYTAEEAARAGERTGRNHWHSDERGDrLRLPDYGYEG1QRHVVPGSSEPKEENEAARWGRANPTVfJGLNQLRKV VNRLIAAHGPRDPQIVVRELAREKLNREQKERLDRENKRKNEENERTAILAEHGQRD TAENKIRLRLFEEQARANAGIALCYPHTGRIAELFTSEVIDHLPVSLTDSDLANR VLCKREANSREKJRQFTQAFGATPANAIDVARAAKLPPNKRWDFPAALFEREFGDFLRQQLNKTCKLAKYLGICDPDRYVVTTFGLTLG23RARWRGLNSILSDSNFKNNRSDQHLHAVDAVVGXLTRGMIQRIAHDAARAEDQDRLVFRDVPVFEDFRDHVRERVSTITTAVKPEHKGALHEDESTSGLYVPDTPNAALNGNVRKPRISLTAGEVDRVRDRLRARLGLAAAPFRDESGVRVRAKGLAQALEFGAENGIRRVRLKPDASVV TIAHDRTGPYRVAPVNGHIVQMRGDSWGRGFAASVFENPGRWPWEWEVKK LGGKLVMRLHGDMVSLKDQGQRVKVQVQIEISANRNLSPHNDGKKLQDRHA DADDFPRWDLATIPLLKDRCGAVRDPIGVVLRRSNV

CRISPR-associated protein, Csnl family [Alicyciphilus denitrificans K601]

MRSRLYRLALDLGTSGLWALFRDLACNRPATAVKAGVRIFSDGRNPDKDGSSAVTR RAARAMRRRRDRLKLRKMMAKLVEHGGFPPADAGKRKALEQLNYPYALRAGLQEAALLPGFARALFFnNQRRGFKSNRKTDKDNDGSVXLKXAIQQLRQMAEQQGSRRTVGEYLWTRLQQGGVRARYREKPYTEGKKRSDKYDLY1DRAMEQEFDALWAQA AFNPRTLHEAARADLKDTLLHQRPCRPVKPPRRCRCLLPNEERAPLALPSQFRHQEV ^^RLLDENLREVALTLAQDADVTTALETAKLCSFEQUIKLLKLSGVSQFNLEDRAK TELKGNATSAALARLFAALAGSFDEALQDEIVWQLVTEEGALJAWLQTHTGVDEARAQIADVSLPPEYGNLASRKLARIVPALRAAVITYDKAVQAAGFDHHSGLGF YEADASEVELVHPETGEIRSVFKQLPYYGKALQRHVAFSGKPDDEKRYGKIANPT -74
YP_003552871.1 CRISPR-associated protein, Csnl family (Candidatus Puniceispirillum marinum marinc CMM1322)

MRRRLGLDLGNTSIGCWCLLDDLGDGDGEQPSVIFRTGARIFSGDGRDPSKSLGSGLKATRREALR
TRRRRDRFIQRQKKNLINALVKYLEMLPADEQIRQQALAYKDPYPIRQKKADEAQPYEM
GAIIFfiINQRGFKSNRSADNEAGVQKQSIADLMKLGEAGARTIGEFLADRQATN
DTVRARRLSGTNAYEFYPDMLQEDFTLWAKQAFNPSTLYEAAERLKEIVFF
QRKLKQEVGRCICIFLSERISKALPSFQFRYIQEQLNWLAFIDGVAHRITASLALR
DHLFDELEHKKKFLTFKAMRAILRQGVXVPGFNSLESH^HLLIGINLSCIMRADA
KKMIGSAWDRLLEEDEQSFILMQLDDQKGGDEVSILTQQYGLSDVADCLDLVR
PDGHGSLSKKAIARRLPLRLDQPGLYYDAVEAGLGEANYDPAALSQDLDDYQK
ALAGHVMGASGKFDSDEKRSTQISNPVHIALNQVRANVNLRLHGHKPDDEVIEI
GPJLPMGADGKJIELEFQGEIKRRAKEREDELKGLfISFRARQKFKQHELWELA
PKVDRCCCFTRGKMSISDLFSDKVIEEHLPPSLTLDMSANAVCVRQANRDKGNRA
PFDAGFNSPGAYDQVELGRSQRNPYAKRWRFLPDAMKRFEDGGFLERQLNNDTRY
ISRTYTEYTIPIPNKVVTGGRLTSLLRFGWGNLSILRGLHTDGDTPAKRSDHHRH
HAIDAIYVGMTSRGLLQQARKVSKAARRSDELDDLTRLFEGRIPWDFGRDEVKfHIIAIV
SHPRKKSQGALHNDTAYGIEVAENIAGASTTVHRVPITSLGKQSDIEKVRDPLKSAL
LNETAGLSGKSFENAVQKWACDNSIKLRIETVSSIIPDKEGAYQYKGDGNAY
MDIYQDPTSSKWKGEIVSRFDANQKGFIPSWQSQFPTARLIMRLRINDLKLQDGIEE
EIYRVQRLSGSFLMAPHTEANYDARDRDKNDFKLTKSKPGKLQASARVfHISPT
GLIREG
YP_003448082.1  CRISPR-associated protein, Csnl family [Azospirillum sp. B510]

MARPAFRAPRUJEHVNGWTPDPPHPJSKPFILLVSWHLLSRVVIDSSSGCFPGTSRDHTD
KFAEWECAVQPYLSDLGTSNISGWLLLLDRQGKKPREIRALGSRIFSDGRDPQDKA
SLAVARRLQRMRDRLTRRTRRLVRFLMADPAARKRLAVADPYLA
RERATRERLEPFIEGRALFHLNQRRGKYKPVRTATKPDDEEAGKVKEAVERLEAAIAAA
GAPTILGAWFAWPvKTRGETLRAPvLAGKGEAAAYFPFAPRMLEAEFIDLWAEQARH
HPDLLTAEAREILRHRIFHQRPLKPPPVRCYCTLPDDGRAPRLAPAQRQLRFLQELAS
LRIVHLNLSERPITPAERDRIVAFVQGRPPKAGRPGKVKQSVFKEKLRLGLLELPPGT
GFSLEDKRPELGDTGARIAPAFGPGWTLALPLEEQDALVELLLTEAEPEAIAALT
ARWALDEATAAKLAGATLPDFHGYGRRAVELPVLENEDPGDRVGRMPRLDE
AVKLRRGGKDHSFDSREGALLDALPYYGAVLHERHFAVGTNPDPEEKRVGRVAN
PTVfIAALNQLRHLVNAILARHRGPEEVIARELKLKSAEDRRRERDKQRADNQKRNE
EPJCRLILSLGERPTRPNLLKRLWEEQGPRVENPJCYPYSGETISMRMLSEQVIDHLP
FVSLLDDASAhpVVCLREANPPKvMISPWEAFGHDSERWAGILARAAELPKNKRWR
FADALEKLEGEGLRHLNDDRTLRVEYLRCVPKVRVSPGRTALLRRRW
GIDAILAEDGPEPPEVPAETLDPSPAEKNRADHRHALDADVIGCIRSMMQVRQVLA
AAAEREAARAEDNIRRVLEGFKEEPWDGFRAELERRARTIVSHPvPEHGIGGALHK
ETAYGPVDPPEEGFNLVVR1PVIDGLSDKEmSVRDPRLVPRALIDRLAIRANDPAT
ALAKAAEDLAAQPASRGIRRVRLKKESENPRVEHGGNPSGPRSGPFHKLILAGEV
HFIVDVALRADGRRRVGWVHTLFREAHGGGRAGDAAPRPLGDERFLMLRHKGDC
LKLEFDCGRVRMQTVKLEPSNSVVVVEPHQVKTRDSKHVKICDQLRARGARRV
TVDPGLVVRVHAPGARVGGGDAGRMTAMEPDIS

YP_571550.1  hypothetical protein Nham_4054 (plasmid) [Nitrobacter hamburgensis X14]

MHVEIDFPFFSRGDHLAMMCEIIRLRGSSVLYRLGDLGSNSLWGFTVHLEKRGDR
HEPVALGPGVRIFPDRDPQLSGTSNAVDRMARGARKRRDFVERKELLAALIKY
NLLPDCERRALEVLDPYALRAIKTALTDTDPAHAVGRFALHNLQQRGQSNRKTD
KQSEDGAKQASRLATDKZGNETLGVFFADMHLRKSYEDQTAIRAEVLRLGKDH
TGNARKKIAKVRKRLFDEVLPRADAPHVRARAg1TGA4MKASYDYYPTRDMLRD
EFNAIWAQQSAHATTDEARTEIEHIHYQPLKPAIVGKCTLDMPATPFKDEPEGYR
APWAPFARQFRJLSEARNLEIRDTKGSRRKTLQKQALVVAALLANREVKFDKLR

-76-
YP_001239928.1 hypothetical protein BBta_3952 [Bradyrhizobium sp. BTAii]
MRVFKGDMIPJDHAGAEKFVKJVRLSPSNNLLYLVEHHQAGVFQTRHDDPEDSFRWLFASFDKiREWNAELVRIIDLQWPVMrKRGLETGSEDATRIGWTPPKWP

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

MRGLDGTSSIGWWLYETDGAGSDARITGVVDGGVRIFSDGRDPKSGASLAQVDRR
AARAMRRDRLYRRATLKVATETGLMPADPAEAKALEALDPALRAAGLDEPLPLPHLPvALFHLNQRGRFGKSNRTKDRGDVESGKIKDATARLDMEMMANGARTYG
EFLHKRKQKATDPRHVPSVRTRLSIANRGGPGKEEAGYDFYPDPJJHLEEEFHLKW
AAQGAHHPETELTLDLSFEKIFFQRPKLEPEVGLCLFSGHHGVPPKDPRPLKAHPLETQRRVLQETVNLRTADGREAPRLTRERDDQVIHALDKPSTKSLSMVLKLPAMA
KVLKLDRGFTLETGVDAIACDPLRAPHAHPDFGPRWSILDADAQWEVISRRR
VQSDAESHELLVDWLTEAHLRAHAEATAHAAPLDYGRGLTLATTRILYQLTAD
VVTYADAVKACGWHSDDGTGECFDRLPYEVELRHIVPGSYHPDDDITTRFGR
TWTVfIGNLRPvLVPNIJETHKQPQWHDELKKEEEKQKAA CRIPRI
KKRSEKLEEEIEIDNGRNRMLLLRLWEDLNPPAMRRFCPYTGTRISAAMIFIGSMDV
DFhILPYSRTLDDSPFNPRTLCLREANRQKRNQTPQWAGDTPWHIAAANLKLNLPN
KRWRFAPDAMTRFEGENFLDLRAKDTQYLARIKSYLDLFTKGKHGVWVVPGRFT
EMLRHHWGLNSLSSDAGRGAVKAKNRTHHRHAIDAAAVIAATDPOLLNRAISRAAGQ
GEAAGQSAELIARDTPPWEFIGRDLVLRLIDIVSHRADHGRDAHKQGRDSTA
GQLHQETAYSVVIDDHVAASRTDLSSLKPAQLLDDEGRSGQVRDPQLRLKARLVTAGGK
TGKDFENALRYFASKPGPYQAIRRVRIVIKLPQAQRVPVPAQDPIYAOGGSGNHLEI
WRLPDGIEAIQFITSFAYHTELEKEKRPHPAAKRLLRHVGDMVALERDGRFVGHV
QKMDIANGLFIVPHNEANADTRGNKDSDPKFWIQIGARPAAISGIRRVSVDEIGLRDD
GGTRPI

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

MEWFGFDIGTSIGFVDSYSTQSAGYINQRLGVRIFPEARDPPLNQQRQKRMM
RRQLRRIRRKKALNETLHEAGFLPAGYSGADVWPMADEPYELRRGLEEGLSAYEFGRALYHQAHRFGRLEEDSTDTPDDEKEANRERATLKLKNEQQTTLGAW
LARRPPSDRKRGIHARHNVVAEEEFERLWEVQSKFHPALKSEMRARSITFAQRPVPV
WPvKNTLGECPFMGEPPLCPKGSLSQQRRLMKLNNLAIAGGNARPLDAEERDAIL
SKLQQQASMSWPGVRSALKALYKQRGBEPAEKLKFNLELGGESKLLGNALEAKLA
DMFPGPDWPAAHPvKQEIRHAVHEPVlWAADYGETPDKKRVIISEKDRKAKHREADAAN
FVADFGITGEQAAQLKLPTGWEPSIPLLALFALEKGERFGALVNGPDWEG
WPJITWPHRNQPTGEILKLPSASPKEEPvEPJSQGPLTWRTQNELRKKVNNLGLY
GKPIDRIRIEVGRDYKSSREEEEEQSGIRRENEKRRKATEDLIKINGIANPSRDDVEK
ILWKEQGERCPYTGDQIGFNALFREGRYEVEfiWPRSRSDSPRNDKLIRKDNIEK
GNRMPFEAFGHEDRWSAIQRLQGMVSAKGGTGMSPGVKVRFLAKTMPEDFAAAR
QLNDTRYAACKQILAQLRLWPDMPGEPAPKVVEAVTGQVTAQLRLWTNNLADD
GEKTRADHRfiAIALTVACTIFGMPTNKLSRYWQLRDPPRAEKALTIIPTWDIRAD
AEKAVSEIVVSHVRKKVSQGPLHKETTGYDGTFTDIKTSGQYRQFVTRKIESLKGGE
LDEIRDPRKEIVAHAIRAGGDPKKAFFPYPVCVSPGGEPEIRKVRLTSQQLNLMAAT
GNYADLGSNFHfiAIIYRLPDKGAFDIVSLFDSASRLAQRNQVTRADGASFVMS
LAAGEAIMPEGSKGIWIVQGVWASQVQVLERDADHDSTTRPMPNPILKDAK
VSIIDPIGRVRPSND

[0121] ZP_17295095.1 CRISPR-associated protein cas9/csnl, subtype II/nmemi
[Bergeyella zoohelcum ATCC 43767]
IVIKffILGLDLGTNSNGWALIERNIEEYYGKIIIMGSMRSIVPMGAMESKFEQQAQT
KNAD
RRTNQRARRLNKRYQFRRNLIYILQKLMLPSQIKLKDIEDPSNPNKIKTIIPISKQ
EQLTADFVLSRVLKALTEKVGLEDLGKIIYKYNLRGYAGGSLPEKEDIFDEEQSKD
KKNSKFIFSVIIVLGFPEQEEIFKNKKLNRAlIVETEEGNGFSTFKVFLNSBELL
MSASKSGDITTIIKLPKNWRXQENQLEKRESMGREYISEFELLEKLKERNW
AKJRNNTLRARSEFEAIW>ffQVKHYFPLENLDKKTLEISIFPGEQESQKRYRE
LQLEKGLKYIIKNQVFYQRELKDQSHLHISDCRYPNEKAIASKPHPFQYEKWEQIN
KLIVNTKJEAGTRKGEKXYIDRPIPTALKWEFEELQNKFEITSAIFKILK^EFD
LREGIDFLNGMKDPDK1KQGNETKLQLQSLGELWDLVLGDSINRQIELWNLYNEKG
NEYDLTSRDKSVLEFIYKGNINVIDDAETAIISKIKFAARYSILSKVERILPL
VRAGKYFNNDFSSQLQSVKILLLNENVEDFPPAAQQTQYLDNQSVLSEGCGVNSIAT
ILVYDKMTAKEYSHDELKSYEINLLKQGDLRNPVEQQINEALVLRDIWKNYGIK
PNEIRVELAPJDLKNSAKERYTIHKKNKDNQTTNQKETLIVKNKELSLANJKEKVLK
WEAQHRHSPTYGQPPILPDFDKKVIDVHIIPIRSYFDDSTFKKBEKSVQENKANR
TamseyfevGLKSYSIFTEKFQIAIWENEYFSGVKRNLLATSIPDWQQRQKDFYIAI
RVKEELNWKGNENVTTGTGTSYDLRNHWGLTDKFKLLKERYADEELEKEFLEAE
YDNYKICDFDSRRKKEYEKEEVLFFEQETREEFKKEYKENIRYKKNKLKIKGWSKRID
-79-
HPjHAIDALIVACTEPAFFIKRLNDLNKVLQDWLVEHKSEFMNFEGSNSLLEEILSL
PENERTEIFTQIEKFRAEMPWKGFEQVEQKLKEIIISHKPKDKLLLQYNKAGDRQIK
LRGQLHETGLYGISQGKEAYRIPLTKFGGSKFATEKNTQIVSPLSGFIAHNLKEYNN
KKEEAFSAEGIMDL>WKLAQYRNEKELKPHTPISTVKIYYKDPSKNKKKKDEELSL
LQKLDREKAOFNELYVTGDNLFAVLGAEKTTKTSQIKRLYDIISFDATNFLKEE
FRNAPDKKTDFKDLRLQYFEERNKAKLLFTLQGDFVYLPNEENEEVILDKESPLYN
QYWGLDLKERGKNIVYVQKFSKQKIFIYKFIHIADIIXCDVEFGSQNCYETVEGRSIKEN
CFKLEIDRLGNIVKIKR

[0122] ZP_07217791.1 conserved hypothetical protein [Bacteroides sp. 20_3]
MKKIVGLDLGTNSIGWALÍNAYINKEfILYGIEACGSRIPMDAAILGNFDKGNISQTA
DRTSYRGRRLRERHLRLRLHLDDLGFLPHYSDSLNRYGFLNDEECKLPWVK
DETGSYKFIIFESFKLAMFTEffIPILIANNKKVYDWTIYYLRLKALTQISKEEL
AWILLNFKQRGYYQMLREEETPNKLVEYYSLKVEKVEDGERGKDWTYNYHVL
ENGMIYRSTNPILDEWEGKTFIVTTDDAEISPKDKEHKISFRAPKDDWTI
KKKTEADIKIKMTVGVAYIYD TTLKQPDQKIRGKLVTIERKYKNELYQILKTQSEFF
HEELRDKQLYIACLNELYNPEPRNSISTRDFCHLLFIEDIIFYQRPLSKSKSLIDNPCY
EENRYIDKESGEIKHASIKCIASHPLYQEFRLWQIVNLRIYKETDVQELLPE
ADYVTLEWLINEKKEIDQKAFFKYPPFKKTTShTliRW^YVEDKPYPCNETHAQIIAEA
RLGKaffIPAKLSKEKEETLWfIILYSLIKYQRGEETPNKLVEYYSLKVEKVEDGERGKDWTYNYHVL
ENGMIYRSTNPILDEWEGKTFIVTTDDAEISPKDKEHKISFRAPKDDWTI
KKKTEADIKIKMTVGVAYIYD TTLKQPDQKIRGKLVTIERKYKNELYQILKTQSEFF
HEELRDKQLYIACLNELYNPEPRNSISTRDFCHLLFIEDIIFYQRPLSKSKSLIDNPCY
EENRYIDKESGEIKHASIKCIASHPLYQEFRLWQIVNLRIYKETDVQELLPE
ADYVTLEWLINEKKEIDQKAFFKYPPFKKTTShTliRW^YVEDKPYPCNE qualitative assessment is not supported in this context.
YP_005848005.1 hypothetical protein IALB_3034 [Ignavibacterium album JCM 16511]

LDGDINEEIFRGAVFETDKGSNVYFVMYENNQTKDREFLPKPNPSISVLKAIEHKNKI
DFFAPNRLGFSmiLSPGDLYVYVPTNDQVYLKDNNSSNETIIINWDDNEFISNRiyQVKK
FTGNSCYFLKNDIASLILSYASNGVGEFGSQN1SEYSVDDPPIRRJKDVCIKIRVDRGLGN
VRPL

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

MKPJLGLDLGTNSIGWALWEAENKDERSSIVKLGVRVNLPTVDELTEWEGKSI^NADRTLKRGMRRNLQRYKLRRLETLEVLKEHKLEDSTILSENQRTTFETYRLRAK
AVTEEISLEEFARVLMLMINKKRYKGSSRKAKGVEGETLDGMDIARELYNNNLTPGEL
CLQLLDAGKKFLPDFYRSDQNEILEWEKQKEYPEILTDVLKKEELRGGKRDAVWC
AICAKYFYVWKeHT/TEWKN^KGKTQEQEREHKLEGYSKRKRDAKRENLQQRWVNG
LKEKLSEQLLYVQFEMTNQINSSGSLGAIARDSKELYFVQKTVGQYQMEMLDNKP
NASLRMNVFYRQDLDEFNMLWEKQAVYKHELTEELKEIRDI11FYQRRRKSQKGL
IGFCEFERSQIETVDDGGKVJKTGVNVRISRSSPLFEQFIWQILhIEWTVEVGGKRRR
KLKENYSALFEELNDAEQLELNGSRRLCQEEKELAQELFIRDKMTKSEVLKLLFDN
PQELDLNFKTIDNGNTGAYLFQAYSKMIEMSGHEPVDFKKPVEKVYIKAVFDLLN
WNTDILGFNSNEELDNQPPYKLWHLLYSFEGDNTPTGNGRLKIQKMTELYGFEKEYA
Tn_ANVSFFQDYGSLSAKAIHKILPHLEGNYRDVACVYAGYRHSESSLTREEIANKV
LKDRLMILLPKNSLHM'VEEKLINQIVvirTV^INVnDIYGPDEIRVELARELKKNAKER
ELTKSIAQTTKAHEEYKTLQTQGFTGLNVTIDLRKLYKEESCYSQLTLYSYTIS
REKLFSK£FDIEfIIIPQPAFDDSSFNKTLEARSVMEKGNKAYTDFVEKEFGEADN
SLEYHYNEDLFKSKGKISDTKYKMLKMAEQDPDFGIERDLNRTQYIAKAKSMLNLE
ISHRVAYSATGSVTDLREDQWQLVMDKNELWYKYALGLEYEDRQGGRKIKD
WTJKmDIIRHMADALTVAFKTDQFVFYQFNNKNASLDPNANEHAIKNKYFQNGRAI
APMPLPvEFRAEAKKHLENTLISIAKAKNKVTGEMNTRKKGVNTKMQQTPRQLHL
ETIYGSGQYLTKEEKNASDFMRKGTIVSKSAYRDALLKLRYENNDPKKAFAGK
NSLDKQPIWLDEKQMRKVPEKVIVTCLEAYTIKERISPLDKVDKVIVGVRKILIDRL
NEYGDAAKAKFCNLDNFWNLKKEGKSIIKVTISGSIANSQSLFTVVKKDKGKIPLEDN
GRNIPVDFVNTGNNHUVAYYRPVIEKQGQLVVDAGNPKYEELVWESFEAVTR
ANLGLPIIDKDYKTTETGWQFLFSMKQNYEVFPNEKTGFNPKEIDLLDVENYGLISP
LFRVQKFSLKNYVFRHRZETIKDTSILRGITWIDFRSISKGLDITKVRVNHQIGYVSV
VEY

-82-
[0125] ZP_10895610.1 CRISPR-associated protein Cas9/Csnl, subtype II/NMEMI

Porphyromonas sp. oral taxon 279 str. F0450

MLMSKHVLGLDLGVSGISWGCLALDAQGDPAEILGMGSRVVPNNTAKIAIEAFNA
AAFTASQERTARTRMRGGFARYQRRLRELEKVMGLPDAALIQPLLELWELR
EERATAGRRTLPELPRLCHINQKRGRYKVSDAAAVGDGEKEDKDNSAYLAG
IRANDEKLLAQEHTVQFYAEQRLRQQSESPTGGSIRKHDQISFRQCYIDEYDQIMA
VQRVHYPDILTDEFIRMLRDEVIFMQPRKLCKHVLSCIFEKEQVRMRVQQDDGK
GGWQDLLERVVFKGPVPKSSPLFQLCICIEYAVNIRLTPGNCIDIPTEWARVA
HLQSSASLSFAALKKLEKALDIADQKSGKLGNSTRVALASALQCPQYPPYHHLD
IVfLETRMMMTQLTDEETGEVTERAVVTDYSVRKPLYRLWHILYSEEREMARRA
LITQLGKMEDELDGGLLQYRLDFVKPGYGNKSAFICKLLQPLQQGLYGSEACA
AVGYRHSNPSPTSEEITERTLLEPIQNLRRQNLVKEILQNMINLVAAYGIDE
VRVELARELKMNSREERERMANNKDREERNKGAIAKIRECGLYPTKPIQKYMILW
KEAGRQCLCYSIEEEEQCLREGMVEHIIPKSVLYDYSNGKNCRANCNKEKGN
RTALEYIRAKGREAEYMKRNDLLEKKISKYSKHQRLKEDIPSDFLERQLRLTQ
YISQRAMAILQGIRVSAEGGGVTARLRSLWYGKHTLNLDRYDSMGETYSR
EGATEELmTNWSKRMDHRHAIDALVVACTRQSYIQRLNLSEFGREDKKEDQ
EAQEQQATETGRLSNLERWLTQRPQFSVRTVSDKVAEILISYRPQWTRGRNIYR
KKMADGREVSCVQRLVPQGRVSGRJVRKYPHYLDKEVDP
HLRELITTYNQELKSKREGAPIPPLCLEDKDKQVEQSVRCYAKTSLDKAIPMCFDEK
GEPTAFVKASNFLHLALYRTPKGLVESIVTFWDAVDRAYGIPLTHPREVMEQV
LQRGDIPEQVLSSLPSDWSFVDLSQQDEMVIGLDEELQRALEAQNYRKISEHLY
RVQKMSSSYYVFRYHLETSVADDKNTSGRPFKHRVSQKAYEERMKVRVDLGL
RISLL

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

Barnesiella intestinihominis YIT 11860

MKN1LGLDLGLSSIGWVIRENSEEQELVAMGSRVSLTAAELSSFTQGNGVSINSQR
TQKRTQKGYDRLRRTTLRNLKDLTLGMLPDDSLSYLPKQLWGLRALATQRIE
LNELGRLVHLHNQKRGYKSDFSDDKKTNYVKTVKTCRYDELMKERTGELFFR
RLTENAFFRCKQEVQPRQAYVEEFDCTMCNQRKFPDILTDIECRTIDEIIYQPRKL
SCKYLVSRCFEKFRFLNAAGKTEAHPKVSPRTSLFQVCRLWEISINIVKVDRRN
EIVFISAEQRAALFDLNTHEKLGSDLKLGSLKSYGURLGEQFTGIQKNKTRVE

-83-
IERYALGNYPDXXRLLQFNLQESSSSMV\^N\text{TETGE\textregistered}P\text{MI}LSF\text{EQEPLYRL\text{WHVLYSIDD}}
\text{REQLQSVLRQKFIDDDDEV\text{ELRSAIDLVKAGFNGKSSKAI}RRLPFLQLGMNYAEAC\text{EAAGYN\text{HSNNYT}KA\text{ENAP\text{v}}ALL\text{DLRPAIKKNERLQP}VE\text{KILNQM\text{VNVNALARMEK}}
\text{YGRFDEIRVELarelQKSVEERSNTYKSKNQRENEQ1AKRIVEYGV\text{PT}R\text{SI}KQKYK
M\text{WEESKHC}CIYCGQPVDF\text{GLRGFDEV\text{EHIIPKSLYF}DSD\text{FAN}KVC\text{SCR}C\text{R}\text{SCNEK}
\text{NNTAYDYM\text{SKGKEKALS}DYVER\text{WTMY}TNQNQ\text{S}KTKQ\text{WNLTPVK\text{ISIDFIDRQ}
L\text{RESQYIARKAKEILTSICYNVTATSGVS\text{TSFLRHVGWD}TvLHDL\text{NF}D\text{RYKVGLT}
EVIE\text{VNmGS\text{VIRRE}QKDWSKJIFHRH\text{HAIDALTIACTQKAYIQ}P\text{v}LNNLRAE\text{EEDPF}
\text{NKMSLERYIQSQPHFSV\text{AQVREAVDRILVSFRAGKRA}TVP\text{KGRYIRKNKR\text{K}}\text{RISVQS\text{V}}\text{LJP\text{R}}\text{GALSEESVYGVIIHVWEKDEQGHV}I\text{Q}QRAVM\text{KPYT}S\text{INRELMDKEKVVDKRI}
\text{HRILSGRLAQYNPNKEAFAKPYVYIDKECRIPTRVCF\text{AKPAINTLV}PLK\text{D}DGNP
\text{V\text{AWVNPGNNNHVAIYR}DEDGK\text{Y}KERT\text{TVF}WEAV\text{DRCRVGPIA}TQP}\text{DIW\text{DNLQ}
\text{RhnDISEhTVL}ESLPD\text{VKVQFLV}LQQ\text{NM}F\text{ILG\text{MNEIDRYAM}DQDYALLNKLYV}
\text{RVQKL\text{SKSDYSF}R\text{Y}HE\text{TITESV}\text{E}\text{D}KGD\text{KP\text{N}}\text{LKiSMQ}MGKL\text{KRVS}K\text{SLLGLNPFIKVH}
IS\text{VLGEI\text{KEIS}}[0127]
\text{ZP\text{0}9642280.1 CRISPR-associated protein cas9/csnl, subtype II/nmemi [Odoribacter Ianeus YIT 12061]}
\text{METTLGIDLG\text{TOSI}GLALVDQEEHQLYS}G\text{VmFPEGrNfKD\text{T}IGL\text{GEKEESRNAT}}RAK\text{RQMRRQYFRK}\text{KLAKLLELIAYD}\text{MCPLKDPEVDV\text{RRWKWN\text{DQ}K}KST\text{VRQ\text{F}}D\text{TP}
\text{AFREWL}\text{KQNPYELRKQAVT}ED\text{VTREP}LEGR\text{ILYQM}\text{IQ}R\text{FRGFLSS}KG\text{KEEGKIFTGKD}
\text{RMV\text{GI}DETRK}L\text{NQK}Q\text{T}L\text{GAYL}D\text{IPKN}G\text{EKGKRFTRERVRARYTRLM}YIRE\text{FEI}W
\text{QR\text{QAGHLG}LAH\text{EQAT}RAKKNIFLEGES\text{ATNYRNSK}LK\text{LTHLAQKYG}RG\text{FTV}L\text{E}D\text{TRIVT}
\text{FQLP}\text{LKEVLL}\text{GKGIEIEEQLKF\text{K}SNE}L\text{F}WQ\text{RQLRSQKSLKSVCFEG}RN\text{FYD}PV\text{HQ}
\text{KWIA}G\text{PTAPLSHPEEFEFRAYQFINIYI}G\text{KNE}\text{HLTAQR}E\text{AVFELMC}TE\text{SKD}FN\text{FE}
\text{KIPKHKLKF}E\text{FDDTTTKVP}ACT\text{TISQLKLFPHPVWKEE}E\text{WHCFYF}DNT\text{LL}
\text{FEKLQKD}YALQT\text{NDE}K\text{KXIRLS}ES\text{GYN}VS\text{LKA}\text{RIRWLY}KL\text{KG}AY\text{STAVL}L\text{G}GIR
\text{NSFGKRPEWKEYEIEKA}VC\text{PJLKEKNAGE}G\text{EVRK\text{KDYLVH}NRF}GF\text{AKNDRAF}QK
\text{LYH}H\text{SQAIT}TQAQKERL\text{PETGNLRNP}IV\text{QQGLNELRITVN}K\text{LLATCREKYGPSFKD}
\text{fNHVEMGR\text{LSSKTER}K\text{QS}R\text{QIRENEEK}NAAKV\text{KLAEYG}K\text{A}R\text{DINIQKYLY}
\text{KEI}E\text{KEG}GT\text{VCC}PYT\text{GTKL}N\text{I}S\text{TGLD}S\text{NSVQ}IE\text{H}P\text{ISL}D\text{L}S\text{LANKTLCDATFNRE}
\text{KGELTPYDFYQKD}P\text{KEWGASS}W\text{EIE}DRAF\text{RLLPYAKAQFRFRKPS}E\text{SNEFISRFQ}
LM)\text{TRYISK\text{KAVEYLSAICSDVKAFPGQLT}AE}LRH\text{LWGLNNLQAPDITF}F\text{LPS\text{ATES}}
NHPvEYYVITNEQNEVIRLFPKQGETPRTEKGELLLTGEVERKVFRCKGMQEFQTDVS
DGKYWRRKLSSSVTSWLFLAPKPIADGQIVLKGRIEKGFVVCNQLKQKLTGLPD
GSYWISLVPISQTFKEGSESVNSKLTSQVQLFGRVEGIFRCHNYQCPASGADNF
WCTLDTDTAQPAPFTPKNNAPPVGGQIILTGDDKGFHADDLDHYELPASLPKG
KYYGIFTECDPILPSAKTSKGENLIEGNIWDEHTGEVRDFPKNREDQRRH
AIDAIVIALSQQSLFQRSLTYNNARRENKKGRLDSTEHPSPWPQFAQDVRQSVVPLL
SYKQNKTLCKISKTLYKDGGKIHSCGNAVQHLKETVYGGRTAPGATEKSYHIRC
DIRELTKSFHGVVDITRQMLLKHLQENyDMITQEFMPSNAFFKEGVYRFJLNPKNH
GEPVPIKKRMRKKEGNAERLDMQNYYVNPNNH1WMIMYQADGKNGKEIEVSFSWSV
IERQNQGQPYQLPREGRMSLQINDTLFLGKEEEEPEVYRNLSTSLKHLRYRKLS
GMYYTFRHHLASTNTVEIREEFRIQSLAEWKRJANPVPKQIDEIGRITFLNGPLC
[0128] YP_004843922.1 putative CRISPR-associated (Cas) protein [Flavobacterium branchiophilum FL-15]
MAKILGLDLGTNSIGWAVVEREN1DFSILDKGVRIFSEGKVEKSEGGIESRARRGTYRS
ARKIKYRXLRXETLKVLRLMCMPSLIEEVEEWKKSGFKDYLNPFLKWLSTDE
ESNVNYFFRRDRASKHKVLSLFELGRAFYHIAQRRFGLSNRLDQSAEGILEEHPKIEAI
VEDLISIDEISTNDITFYFETGIDSNNEKNGYAKDLDEGDKIKLVSYKLSSLAILKKNES
DFENCKSEIIERLNNKDDVLKVYGBKIKDISQAMLKDNNYKTLGQFYFSLYSKEKIRNQ
YTSREEFIYLSFETICKVGQIQINEEKEKFKDGLAKDIYKAIFFQPRLSQKGLIG
KCSFEKSKSRCAISHPFDEEEYvIWTLYLNTIKGTQSDKLLRFQDEKLMVPKFYRK
NDFNFDVLAKELIEKGGSSFGYKSSKKNDFFYWFFNYKPTDTRAVACQVAASLNAIGE
DWKTKSFKYQTRNSNKEQVSRTVDYKDLWHLLTVSVDVYLFELAIDKLGDEKNA
KAFSKTKLJCDFSALSLSAINKILPYLKEGLLYSHAVFVANIENIVDENTWKEDEKORD
YIKTQISEIIENYTLEKSRFEIINGLLEKEYKSENGKJJIVYYSKEAEEQSFENDLKKLV
LFYKSNEIEKNEQGETINFNLPIFIQQLKDYEFIKIQRDLQKVFLKKGNETGQIFCTE
EGKTAEEKFKKTNRLKLKLYHPDIEKFKKIIEDEGNEKVLGSLTPLSKPIMPAMR
ALHQLRKVLNALILEGQIDKFIIFLHEMARENLDANKRGиковыDQYQKNNKKFREDAIK
IKKLYFEDCKVEVETDILYRQLWMEQNRSEIYEEGKNISSCIDIISNPAYPDIHITP
RSRSOQDNSQMNKLCTCQFRNPVEVKKQMPIELNNHLEIPRIAHWKEEADNLTREIEII
SRSIKAATKEIKDKXIRRHYLTJKRDYLQGKYDFWEEPVGFKNQIPTDGIIKTY
YAQAYLKSYFKKVESVKGGVvVAEPFtKIWIQIQLQINSEQIDEMKFIYKVKDRSKHTHHTI
DAITIACMTKEKYDVLAHWATLTDQONKKEARSIIEASKPWKTFKEDLKIEEEILVS

-85-
HYTPDNVKKQAKKIVRVRGKKQFVAEVERDVNGKAVPKKAASGKTIYKLDGEGKK
LPLQQGDTIRGSLHQDSYGAIKNPLNTDEIKYVIRKDLIESIKGDSVESIVDEVKEKI
KEAIANKVLLSSNAQQKNKLVTGVWNEEKRAINVKRIYANSVKNLHIKEHSSL
SKSKHVHKQKVYGG>TOENYAMAIYELDGKJDFELIN1FNLAKLIKQGGFYPILHKK
KEIKGKVFVIPERKRNRDVVLKRGQQVVFYDEVENPKDISE1VDFKGRYIIIESGSIQ
RIVRPSGVDEYGVIMLRFYKEARDDDIQDQTPKGVFLKENKTPvKMNF^QF
TAFVEGIDFKVLPSGKFKEKI

[0129] ZP_08837074.1 hypothetical protein HMPREF0666_03250 [Prevotella sp.
C561]
MTQKLGLDLGTNSIGSAVRNLDSLDDLSQWQLEFFSSDIFFRSVSBNESNGREYSLAA
QRSAHRRSRLNEVRRLRLWATLNLKIHGFCPMSSELPRWCTYDKKRKLGFREYP
IDDKDFNANWLDFTGDPDGDDYSSPYQLRRELVRTQDFEQFPIERYKLGALYfIAQH
RGFKSSKGETLSQETNSKPSSTDEIPDVAGAMKASEEKLKGLSTYMKEHNLTVG
AAFAQLEDEGVRVRNNNDYAIASQRFQHEIETFQQGLQVESYELIERSEKNVG
TIFYKRPLLSQRSQNYGKCTLERSKPRCAIGHPLFKEFRAWTLMKVRMSVDTLEQ
LPMKLRLDDLNYNCEFALAVRETEKFDSDKRLGYLHGFSYNDKTYKDSSTVSAGCP
ITARFRKMLGEWEESFRVEGQKERQAHSNNSFHRVSYENIWFIFCYDAEEPEAVL
AFAQETLRLERKKAEELVRIWSAMPSQGAIMALSQAIRNINKILMLGLKYSDAVLAK
VPELVDVSEDLLSIAKDYYLVEAQVWDKJKINSIVNGLIAKYKSEFYEDHNY
EYLLDESDEKIDRIENSLGARRWSLMMDANEQTDLQVRDRYQDFRSHFKVEVES
PKLGESEFWLTKKFPIVrVEREQWKKLYHPHQITIYRPSVWVKLRLGPNPITA
NPTVLRVLNTRLRRSVNQLLDGVIISPDETRVVEVETARELNDARKWAIDTNYRIRH
DEhffIIKKILIEEFYPKRGDSTDDYKARYIDQREVYDFSGYKTNYKDIKXYKFWLE
QGGQCMYTGTR^SNLFDPNAPFDEHTIPESLSFDSDMNILTCDAnim jRFIKKHNHP
TDMPNYDCAITDGKEPAITSAQRLQWVERLVRNLNEVWYKQARRNQDRK
QCMREMLHKMELEYWKKLRFVTEVTDGFKNQVSDLVTRVTRHAYLVLKLSIFP
HVDVQRGDVTAIFKILGIQSVDEKKDRSLHSHHADATTLTIPVSAKRDRMLLEFA
KmEINKMLLSFGSEDRTGILQELEGLKNKLMVEKVCRIGHNVSEIGTFI NDNJIVNH
mKNAALTPVPxRRLKRKGIVGVDNPWQTGDALRGEIHAKASYGAIQTFQAKDDE
GKVLMLKEGRPQVNPITIKFIVIRRELKYYKSAADSGFASWDDLGKAIADKELFAKMK
QFPAETSFKDACEQGIYMMGKKNMPDIAKHLHURLVRCAPQGLKИEQTYKSEKE
YKIJYFYAAVGDLYAMCCYTOGKIREFIYSLYDVSCHKRSDIEDIPFIDDKGNRL
MLDYKLRTGDMILLYKDNPAELYDLDNVNLSRRLYKINRFESQSNLVLMTTHLSTS
KERGRSLGKTVDYQNLPESRSSVKSNLFLIMGENDFVIKNKGKIFNHR

[0130] ZP_0628774.1 CRISPR-associated protein, Csnl family [Prevotella timonensis CRIS_5C-B1]

MNRKLGLDTGTNSLGAWAVVDWEHAQSYELIKYGDVFQEQGVKIEKGIESSSKAER
SGYKAKRQYFRRRLRKLQVVLKVLKYHCYPYLSDDLRQWHLQPKSDELML
WQRTSDEEGKNPYYDHRCLHEKLDLTEADRYTLGRALYHLTQRRGFLSNRLDTS
ADNKEDWKSQISQSTMEEGCEYLGDYFKLYDAQGKVIRQRQYTRMCH
YQHEDFAICEQKELSSELEDLQRFFQPLKSRQHGVGRCFERGKPRCADSHPDY
EEFRMLCFVNMQVKPGHDLRPLTYEEREKIELFFRKSQNPDFEIAKALAGK
NYAWIHDKERAKYNRMXMQVGPTQPIAQLKSIGFDDWKTIAYEYTLIQKNGS
KSLQEMVDDDV?A^LYSFSVEKIE.KEFA^IHKLQLDEESAEKFAIKLSHSAASLSKA
IRKFLPFLRKGVr^THASFFAMPTIVGKEIWNKQNP^YIMEWGVELVFNQPKHR
EVQGTIEMLIKDFLANNFLPAGATDLKYHPSMIETYPAQRNEFGLQLGSPRTNAl
RNPMAMRSLnllRRVVNQLLKESIIDENTEIVHVEARELNDANKRAIADRFQEQD
KQHKKYDEIRKLKEETGKIEPTQTDVLKFLQWEQNHHCYTGEQIQITDFIGSN
PKFDIEHTIPQSVGSTQMLTLCDNRFVREVKKAKLPTELANHEELTRYEPWKNK
YEQLVRERDKDQRTFAGMDKAVKDRIRQKRRHKLQMEIDYWRGKYERFTMTEVPEGFS
RRQGTGIGLISYAGLYLKLHFAQSRSNKNVYVVKVATEFRMKWGLQSEYK
KCRDNHSCHCMDAIACIGKREYDLMAEYREMTKQKRGRSKFSPKWATFTE
DVLMYKNNLWHDTDPNNMPKHTKYYQTSIGKVLAQDTARGSLHLDTYYGAIER
DGEIRYYRPLLSFTKPEELEMVDETVKRTIEIAKDKNFKAIAEPIYMNEEKGLI
KKVRCAKSVQIPMRQHRDSLKKEYKQYQFTVMNENNYLLAIYEGLVKNNKVREF
EIVSYIEAAKYYKRSQDRNFISSIVPHTSTKYGLPKTLMLQMLVMLFEEPNDEIQV
DNTKDLVRKYCSVIEKDGRIFKFHQEARKEGLPIFSTPYKNNDYAPIFRQSINN
INILVGDFTIDILGKVTLKE

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

MQKNINTKQNFnYIKQAQKEKLIKGDPYRIGLDGVSIGFAIVSMEENDGNNLPPK
EIIMVGSWKASAGAADRKLRSQGRNNHRTRERMRYWLKVLEQKLALPVPADL
DREKNDSEGTSARFLGDVLQKDIYELVKSLEDRLSLEQLYGYVLiffAGHRGSSAI
RTFENDSSEEAKENTENKKIAGNlKRLMAKKNRTYGEILYKEFFENKKEHKREKIS

-87-
NAANNHKSFSPTDLVEIKEAEAILKQAGKDGFKHELTEEYIEKLTKAIGYESEKLPES
GFCPYLKDKEKLPSHHKLNEERRLWELNNARYSDPVIVDTGTEIGTYEYKEQFTKEQ
KQKLFYYLTTGSELPATQTKKLLGLKNTWDEnLQGRDKAQQKGYKLIKLESMPF
WARLSEAQQDFSLYDWSCPDEKLLTEKLSNYHLTEEEIDNAFNEIVLSSSYAPLK
SAMLIILEKKNLDLSYTEAVEEALKTKEKQAIKDLRPYGYAVLQESTQKIIAKG
FSPQFKI)KGYKTPHT>¾YELEYGRIAVVWHQTLNELRKLVEIDILGKKPCEIGLET
ARELKSÅEDRSLSPvEQNDNESNRRRIYEIYIRPQQVITTRÊNPRNYILKELFEE
QQSKQCPFCGGQISPDIINQADIEHLFPIAESEDNGRNVLVISHSACNADAKRSPW
AAFASAASKDSKDYDNRILS>TVKEMPKHAWRNFQGAFEKFIENPKMAARFKTDNSIY
SKVAHKYALCLFEKPNIICVKGSLTAQRLMAWGLQGLMIPFAKQLITEKESESFNKD
VNSNKIRLdHhmHHALDAIVAYASRYGhmXNMAKGDKY_INYSEWNLWISKLLPP
NNIVWENIDISLLESFESSVKTLANKAFISVKHDSDNELVKGTMKIFYSERGTYL
TTYYKLossalDTPQKkkKTPFDLETAL^KFKGRESEMKNIEIKSAENNKRFLDVIQ
DM.EKAKKLLLEE^KSKAEK*KNINDASIQKAISSLGDQYVQLSKKEPKAFFAI
SKPTPTTTGYGYDGTDLSCVLIDYDNKGKLCEIRKIDAOQPKNPLKYKEQGFTLFE
RJJYGIDILEVFDFHDSKDNFSRhhWTGSAEmFIFKVTFTITEITONMQIWFNHIKSTG
GQDSFTFNSMQQYNPRKLILSSCGRFICYRSPlKNKEG
[0132] YP_004248194.1 CRISPR-associated protein, Csnl family [Sphaerochaeta
globosa str. Buddy]
MSKKVSRRRYEEAQAEICQRLGRSPYSIGLDLGVSIGVAVAAYDPIKQKPSDLTVFS
RIFIPSTGAAERRQKRGQRNRSLRHRANLKLFLKLLAERNMLSYSEQDVDPDARLR
FEDAVVRANPYELRKLQNEQLTSELGYALYID ANHRGSSSVRTFLDEEKS SDDKK
LEEQQAMTEQLAEKGGISTFIEVLTAFTNGLIGYRNSESVSKGVPVPTRDIISNEID
VLLQTQKQFQYELPSDEYCDRISAILFENEKIEVFPEGCCPYFPDEKLLPRCHFLNEER
RLWEAINNARIKMPMQECAKKYQASFSDEQRFILFHARSGTDITPKVLQKEFPA
KTSnVLQGKEK*IQKJAGRFRRRLIEEKFWSKRLSEEOQJDDFSAWTNPDDKRLSKY
LMKHLLLTENEDALKTVSLIGDYP1GKTATQLMKHLEDGLTYTEALERGM
GEPQELSVWEQSSLLPYYGILTGSTQALMGKLYWSFAKKEKRDSEGGFFPKNTSDE
EKYGWANPVVHQLNRELKLMLNITILGAKPQETITVRELARLKVGAEKREDIIRQQ
TKQEKEAVLASYSKYCEPNNLKRY1ERFRLLEDQAFVCPYCLEHSIAAAGRADV
DfHFPDDTA DNSYGNKV AHRQCNDDIKGRTPAYAAFSNTSAWGPIMHYLDETPGM
WRKRKRFETNEEEYAKYLQSFGFVSFEDSNYIAAKAKEYLRCFLPNVNTAVGS
-88-
YWKRCDTHVRLPNAKPGRTFVIITFTEMGSGYQYFSNLAKSKKGQDTSTFLTTKIN
YDVRKVQLSSAGLYVRYVSPLLVDKIEKDEVALCGE

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

MGGSEVGTVPVTWRLGVGERSGILAAVSYEEDKPEILAAVSVWIDGGGVGERS

[0134] ZP_07880770.1 conserved hypothetical protein [Actinomyces sp. oral taxon 180 str. F0310]
MLHCIAVIRVPPSEEPGFETFADSCALCHHGCMTYAADKAIYRVGVGIDVGLRSGF
CAVEVDDEHPIRILNSVHVHDAGTGGGGETESLRRKSGVAARARRRGRAEKQRL
KKLDVLLEELGWGVSSNELLDHAPWfIRKRLVSEYIEDETERQCLSVAMASfIARH
RGWRSFSKVDLLEEQAPSMDQMGLKREVDRTGLQFSEEVTQGELVATLLEHDG
DVTIRGFVKGGATKVGVLEGYMQSDLLVaelRQICRTQRVSETTFKLVLSIFH
SKEPAPAARQRERVGLDPALDPAAKQPRAERAHPAFQKFKWATLANMIRER
QSAGERLSTEELRVARVARYLNHTESEPWDDVARKLEVRHRLGSSRASLETGG
GLTYPPVDDTTVRVMSAEVDWLADWDCANDESRGHMIDASINGCGSEPDDVEDE
EVNELISSATAEDMLKLELAAKLPSGRVAYSGLTVRTAAILETDGLSQAIRTYL
GVPDWQWPTPAPIEAPVGNSVDRVLQVARWLFASKRKGVPQTVN1EHTREGLK
SASLLEERERWERFREIREKMYRLISGPGFRRSDQYVERIEILDLLQDCAELCYG
NEINFQTFEVHDIIPRVDASSDSRRTNLAACVCHSCNSAKGGLAFGQWVKRGDCPSGV
SLNAIKRVRSWSDKDLGLTEKAMGKRKSEIRVLKTEMPYEEDGRSMESVAWMA
IEEKKRiegynDSRPGCAAVQVNVYSGRTACARRAHHVKVRKIDLKGDDGH
HKNRFDRRNHAMDALVIALMTPAIARTAVREDREAAAQTLRATESWKNFGLSSEER
MQDRWESWIGDEVEACDRNLEIDADKIPVTENLRNLSGKHLADQPESLKKARRG
SKRPRQPQYVLDLADPMADVNHVTDPGLWTALVRAVGFDSQGLLPADLRNLKLRG
KRISADFPIDYFPTDSPALAVQGGYVGLEFHARLYRI11GPKKVKVYALLRVCADLC
GIDCDDLFEVELKKSISMRTADAKLEAMNGSAKIQWGLVLGDEIQIDPTKFKQ5S
IGKFLKECGPPSSWRVSALDTPSKITLPRLLNSPELKLTSRVRGHHFSLVVAECEV5
IMKKTGWWVEINALCQSGIRVIRRNALGEVRTPSKGLPISLNLR
[0135] ZP_03925169.1 conserved hypothetical protein [Actinomyces coleocanis DSM 15436]
MDNKNYRIGIDVGLNSIGCFCAVEVDQHDTPGLFNLVYRHADGIDPNGKKTNTTRL
AMSGVARRTRRLFRKVpKPvRLAALDRFIEAQGWTLPDHADYKDPYPWTLPVLRRAELAQ
TPIRDE>iDLHEKLAIVRIfIRAHHRGWRSWPVPPVRLHEQQPSPQYLYALKERVEAKT
LLQMPEGATPAEMVVALDSLVDWPpPKNREKTDTRPENKKPGFLGGKLMQSDNA
NELRKIAKIQGLDALLREIELVFAADSPKGASGELVGYDVLPQHGKRRAEKAPH
AFQRYRIAISVNSLRIRHLGSGADERLDTVETQKRVFELYNKAKPTADITWSDVAEEIG
VERNLLMGTTAQTDAGERASAKPPVDV INVAFATCKKLKEWWLNAHYEARCMVM
VSALSHAELLTEGTAEEVEAFELQNLSDEDNEKLDSFSLPIGRAAAYSDVSLERLTKR
MIENGEDLFEARVNEFGVSDWRPPAEPICARVGNPAVDRVLKAVNRXLMAAEAE
-90-
VGAPLSVNIEHYREGISKRQAIEEEPVENQRKVRQIVAVRSGIADfNATSGVRGSD
VTRYLAIQRQVNGECLYCGTAITFVNESEIVrofIVPRAGLSTGRDNLVATCERCNKSK
SNKPAVWAECGIPGVSAEALERVDFIADGFASSKEHELQKGVKGDKLKRKS
DPEIDRSMESVAWMARELAVHRVQYYFEDEKHTGKVRVRGSLTSAARKASGFESR
VNFIGNGKTRLDRHHAMDAATVAMLRTSNVSATLVLQGIRASERAINAEWT
SFRCENVADQQFESWENMRVLVEKFLNALGYDEVSIFSSLRLQLGNGKAEHDDIT
KLQMHKVGDAWSLTEDRASTPPALCALTQPDDFTWKGGLPANEDRT11VNGTHY
PLDKVIGFKAASLLVRGSVDIGSAIHARHHRAGKKPTYGMVRVFDPPLYR
NEDLFNVELPQVSVMRYAEKPKVREAIREDCAKVYLVGDLVLDLDSSETSGP
YQQPDFTTHTWAGFFSPRLRLRPVYLAQEGL
GEDVSEGSKAIIIAGGGRPANVVKGFSAMPEVIRRDGLGRKRRFYSGLPVWQG
[Y036]YP_001955845.1restrictionendonuclease[Bifidobacteriumlongum
DJO10A]
MLSRQLLGASHLRPSYNSVQDNDNHCSYGERCFMRGKRIGIDVGLNSVGLA
AVEVDENSPVRLLNAASQSVIHGDGVGQPKQNEAKITKRNSMGVARRRMKRRRER
LFIOiDMLLGKFGYPVIEPESLDKPFEEWHVRAELATRYIEDDELRESISIALHRM
HRGWRNPYRQVDSLIDNPYSKQYGELKEKAKAYNDDATAAEEESTPQLVAML
DAGYEAEAPRLRWTGKPKDEYLPYVQFEDQNLANEEKFGVRQVYRPADEKWPL
FRSVFYAVSPKGSASEQVRQVQDPQLEPEQARALKASLAFQIEYRIANVITNLRIKDA
RKLTVDEKQIYDQVSPSSEITWSDDLDCFLGFKRSQLKVGLTEDGEERISSRPR
LTSVQRIYESDNKIRKPLVAWWSASDNEHEAMIRLLSVTDIDKVREDVAYASIAE
FIDGLDDDAIATKLDSVDLPSGRAAYSVESLQKLTRQMLTEDDLHEARKTLFNVTDS
WRPPADIEPGLNPSVDRLKVNVRYLMLNCQQWSGPNVPSVEMHYRSSFSSVAFA
RKDKREYKNEKNEKSIFRSSSEQLRADEQMMEKVRESLRLLEAIQRQVNGQCLCGR
TIFRTCEMDfIVPRkgVGYSTNTRTFNAAVCAECNRMSNTPFAPARSEDAAQTRGV
SLAEAKKRTVMFNPAYQAVFQVIPQAVIAQDAUNDDAIDNRSIEVAVAA
DELHRiaDWYFNASQVYNNASODAEAMMTKMTTTSFQGVRVTASARAAAGIEKIH
IGQQSKTRLDEHHADSVIAMMNTAAATMLMRELRESQIRISSLMPGERSWEK
YPEGTSRYESFHMLDNMVLEIENMLDNDRIAVMQSQRVLGNSIAHDRIH
PLEKVPGLSAMSADLRIRASTPPALCALTLPDYDEKEGLPEHREIYHDTRYSA
DDEMGFFASQAQIAVQEGSIADGSAIHARVRCWKTNAKGVRKYFYGMIRVQFQT
DLLRACHDDLFTVPLPPQSISMRYGEPRVQALQSGNAQYLGSLWGDEIEMDFSSL
DVGDQI GEYLQFFSQFSGGNLAWKHVVVDGFFNQTQLRIRPYLAEGLAKAFSDD
VVPDGQKIVTKQGWLPVNTASKTAVRIVRRNAFGEPRLSSAHHPMSWQWRHE
[0137] YP_001878601.1 hypothetical protein Amuc_2010 [Akkermansia muciniphila ATCC BAA-835]
MRSRLTFSDIGYASIGWAVIASHHADDAPSVCGCGTVLFPPKDDCQAFAKRREYRRR
PvRMRSRRVRJERIGRLLVQAqTPEMKTSGHPFPYLASEALKGHRTLAPIELWFTV
LRWYAH≥¾GYDNNASWSNSLSEDGGNGEDTERKHAQDLMDKHTATAMTICR
ELKLEEGKADMEVSTPAYKLNNTAFPRPLIVEKRRILELSAPLIGLTAEIILIAQ
HHPLTTETQRGGLQHGIKLARRYGSSLFGQLIPRFDNR11SRCPVTWAQIVYEAEKK
GNSEQSARERAEKLSDKVPTANCPEFYERMARILCNIRADGEPLSAEIRRELMNQAR
QEGKLTKASLEKAISRSSRLGKETETFNSYFTHPDEEEALYNPAGELQRSGIGQILS
PSVYmAAALRRRRGKSVTVPNLYKLKRSGESGAELEKIEKESKKEADYADTPLK
PKYATGRAPYARTVLKKVQVEILGEDPTTRPAREGAHPGEDLAKADHGCLYLCLTD
SSVNQHQKERRLDTMTNNLHRHRMLIDLRLKLDIQDFADQKDIR SRVCEVGGK
ELTTFAIVTOSKKIQRELTLRQSHDAVRNLRRKLPGKALSANLRKCRIAMDMNWTCPFTGATYGDHELELEFIVPVSFRQNSALQLVLTWPGVNRMKQRTYGDFV
QEQQPVPDKJNLfIICSLNNYRELVEKLDDDGGREDRRRKKKRRALLMVRGLSH
KHQSQNiiEAMKEIGMTEGMMTQSSHLMKLACKSKITSLDPFAEfIDMIPGAVTAEVRK
AWDVFQFVKECPLCAADPDSGKILKENLRLTHLHALDACVSLGLIFYPAHNGL
RRVLAMRRRIPEKLIPQVRPVANQRHVYNDDGRMMRLDSLKSENIREQLMEVQRS
IQHVPADMGGALLKTQMRVLSVDGSGEDAMVLSKKKDGKEKNQVKASKLVG
VFPEGPSKLALAAIEDGNYGALDPKVIRHIVKFRMALKEQNGKPKVRILK
KGMLIHHTSSDKPHAVGWRIESIQDSKGGVKLQLQAHCAPA1KTCNWCREE
LISLKLQKMKYRPTSYGTTPR
[0138] YP_004168469.1 CRISPR-associated protein, csnl family [Nitratifactor salsuginis DSM 16511]
MKKILGVDLGI5FYGAYLQETGKDLRYRCLDNSVVMRNNPYDEKSGESSQSIRSTQKS
MRRLIEKRRKRRICVQAQTMERYGILDYSFTMKINDPKNNPKRNQRQGLRAVDAWKRPL
LSPQELFAIFAHMAKHRGYKIALEDJLIELELGLNDPEKESEKKADERRQVYNAL
RHLEERKKNYGETIAQTIHRAVEAGDLRSYRNHDUYEKMRREDIEEIEKVLRRQA
ELGALKGPLFQVSELIDELKAITQDQEMTIDESLFKCTFYKDELAAAPASYLYDLY
RLYKLADLNIDGYEVTQEDREKVWEVKEKIAQGKNLKTKTFCDLRRKLGLAPEQK
-92-
IFGVEDERIVKGKKEPRTFVPFFFLADIAKFKELFASIQKHPDALQIFRELAEILQRSKT
PQEA
DGWL
MGKD
TIDDDRELLFLK
NK
RSGTREL
SHRYLEA
LPLFLE
QYD
VQmLGF
DREDSYPK
SRLRLHLH
REGNLFEKEEN
P
NNHAVKVSLAS
WA
LGLIAL
WRYGPFDHEHETRD
LPEKIPCEIDKAM
REPv
EKALDK
GKYKKEEPSIDKRLARKI
QLWERQKG
GLRYSGLK
VINLSQ
LLDG
SADIEHVFPQSL
GL
STDYNTIVTL
KS
VA
AK
G
RLPDGWLA
GNDYREPJM
SEKGLD
WKKRNL
LQA
QLD
DEY
twen
S
GRAT
SYEALVA
QVL
RYF
PPDP
LKRNG
GIGV
MIPG
VKTSKR
LLGK
KS
RETF
FH
HA
EDALILSTLTRGWQ
RHLR
NDY
KEASEELK
LW
W
KKYMPHI
EGL
LADYIDEF
RFRMSKGEESLF
Pv
DMFDTIRSI
YSYWDDKK
P
ASSF
KETVYS
H
EVP
TRL
NL
EA
FD
S
V
K
D
F
K
L
T
E
E
F
M
K
R
Y
D
E
R
Q
K
W
L
Im
G
T
D
S
E
Y
R
A
VEERAT
QIAQLT
RYQLMDAQNDKDEIDKJQQLKELITSPIEV
GK
LR
M
R
P
Y
D
KL
NAMQIDRGLV
ETDK>ML
G
f
f

f
i
M
L
F
l

L
G
I
F

G
L
R
S
Q
E
G
K
Q
G
SK
Y
I
A
F
P
R
F
P
A
N
P
K
A
Q
P
S
K
F
T
S
D
S
K
I
K
Q
V
G
S
A
T
G
I
I
K
A
H
L
L
D
L
G
H
V
R
S
Y
E
V
F
G
T
L
P
E
G
S
I
E
W
K
E
S
G
Y
R
G
V
D
P
H
[0139] ZP_08015909.1 hypothetical protein HMPREF9464_01128 [Sutterella wadsworthensis 3_1_45B]
MTQSERFSCSIGIDMGAKYTGVFYALFDREETPNLNSKAMTLVMPETGPRYVQA
QRTA
VRHRL
RGQKR
YTLARK
LAF
L
V
V
DM
MKQ
K
EKR
KTDEE
WKR
G
RE
ALS
GL
R
RGYSRPNADGEDLTLPI
NV
RD
AVFAH
PAFYSEVRS
LAQ
WEEF
TANISVEKF
LGDPN1PADKEFIEFAVAEGLIDKTEKAYQ
ALST
RANANVLTGLRQM
GHPFRSE
W
K
A
I
E
A
D
L
K
K
D
S
R
L
AKINE
AFGGAERLAR
LL
G
N
L
SL
QCLAER
WYFNAP
D
M
K
D
R
GW
P
FRFKXTLVRAFKFFHPAKDQNKQ
HLEIL
QKENSIDEI
E
TCLTD
PNRTPYED
QNNRPPFLDQTLLSPEK
L
TRQGIEGI
WKTS
AR
L
SAEP
L
A
A
E
I
L
S
D
R
K
S
R
V
AVNGHEPLT
PL
SY
ALQRA
F
DRS
K
AL
D
PY
A
L
R
L
A
A
GSKSNKLTSARTA
L
C
I
GGQNVKTFLDCARRYYREADDA
V
GL
WDF
A
DGL
L
ER
L
HPMK
K
KL
L
V
AN
I
LQTD
ETGQK
F
L
D
I
W
R
K
QI
G
R
ETV
A
R
C
I
E
T
V
R
K
S
F
G
F
NIA
YTA
QYREVN
KLPRNAQD
ELLTI
RDRVAETAD
FIAANL
G
L
D
E
Q
KRKR
FANP
FSLAQFYTLIE
T
V
SG
FSATTLAVF
ENA
AWRM
TK
DA
V
E
V
R
A
Q
CR
LS
PAETARPD
FGL
VR
LDRQA
WE
AKR
VSTDI
Q
V
D
S
N
G
I
V
D
S
I
F
V
EENKFEF
S
A
D
L
K
K
K
R
V
D
K
M
L
S
EA
KLETRWL
K
N
R
K
K
ASRG
CT
P
G
T
D
R
L
A
E
G
G
H
P
R
S
L
K
D
AR
G
IV
F
NAP
L
I
Y
ASSRNQLK
K
K
R
Y
LS
L
K
A
Y
N
E
I
F
K
T
S
N
IA
AI
E
I
E
D
D
V
T
K
L
Q
Q
H
L
K
F
D
L
LY
H
E
Q
DCV
R
H
A
L
F
D
G
S
E
A
R
D
A
V
L
L
L
A
T
QRR
TV
N
G
T
W
M
K
K
L
K
I
R
E
E
L
-93-
J7RUA5.1 CRISPR-associated endonuclease Cas9 [Staphylococcus aureus]
[0141] AEX66236.1 CRISPR-associated endonuclease [Corynebacterium diphtheriae C7 (beta)]
MKYHVGI

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

[Streptococcus pasteurianus]
MTNGKILGLD SIGASVGVIIE AKTGKV RLGNAEHTIGKLSKVKLGS QLSVSD IDKASSEALWC ALTREPDPDKPDGLPA NPFRIRVNGTVHYAGDMGLFPVAGSIALRGGYAEGLSSF HHARVYKITS GIXPAFAI

-95-
EEZ17976.1 CRISPR-associated protein, Csnl family [Neisseria cinerea ATCC 14685]

MAAFKPNPMNYLGLDGIASVGWAIVEIDEEENPIRLDLGVRVFERAEVP KTGDSLAAARRLARSVRiiLlTRRRAHRLRARRLKREGVLQAADFENGLIKSLPNTPWQLRAAALDRKL TPLEWSAVLHILKHRGYSQRKNEGETADKELGA LKVGVADNTHALQTGDFRTPAELALNKFEKESGffip^QRGDSYHTFNKRDLQAE LNLFEKQPHVSDGLKEGIETLMLMTQPALS GDVAQKMLGHCTFEPTEPKAAKNTYTAERFVW LTLNNLRLIEQGSRPLTDTERA LTMDEPYRKS KLYAQARKLLLDLDDTAFKKGLR YGDN AESTLMEMKAYHISRALEKEGLK DKKSSPNSLPQDEIGTAFSLFKTDE ITGRLKDRVQPEILEALKii SDFKFVQISIKLA RRIVPLMEQGNRYDEACTE IYGDYGGKN TEKIYLPPPAIDEIRNPVVL RALSQARKVING VVRGSPARIS TAREVGKSKFKR KEIEKRIIE KNDREKSAAKFREYFNPFW GEPSKDILKR LRYEQHGKLCSGKEINLGRLNEKGY VIEIDHALPSRTWDSDSNNKVLALGSENQNKGNQTPYEYFGNKDNSREWQEFKARVETSRFPRSKKQm LLQKFDEDGFKJERNhtOTRYINRFLCQFVADHM LLLGKGRRVFASNQITNLLRGFWGLRKVRAENRDHALD AVVACSTIAMQQ KTRFVR YKEMNAFDGKTIKD KETGEVLHQKAHFPQPWEFFA QEVVMIRVF GKPDEEADTPEKLR LL AEKLSSR PEAHVKYVTPLFISRAPNKR KMSGQGF RMTVEKSAKR LDEG ISVLVRPLTQLKL DLEMKMVNREEPKLYEALKARLEAHKDDPAKAFAEPFYKDY KAGNRTQQ VKA VREQQVQ KTGVVWV NHN GIADNATIVRVDVF EKGGKY LVIYPS-96-
WQVAKGILPDRAVVQGKDEEDWTVMDDSFEFKFVLYANDLIKLTAKKNEFLGYFV
SLNRATGAIDIRTHDSTDSTKGNGIQFSGVGTKALSFQKYQIDELGKEIRPCRLKRPPVR

[0144] BAK69486.1 putative CRISPR associated protein [Campylobacter lari]
MRILGFDIGINSIGWFVENDELKDCGVRIFTKAENPNKESLALPRRNARSSRRRLKRKRKARLIAIKRLAELKLNYKDYVAADGELPKAYEGLASIYELRYKALTQNLTKDDLRVALHIAKFTRGMNKNLKNDACKKGKLALSKNNALKLENYQSVGGEYFYKFFQKYKNTKNIKIRNTKDNYNVNCVLDSDELKLLIEKQKEFGYNOSFIEINELKVAFFQRPDLLDFSVFILGACTFEEEEKRACKNSYAWEFVAALTKEINEKISLEKISGEIVPQTINENEVLNLILDGKSITYKFKRSCINLHESIFKSLKYDKEAPAENALKIDFRKLVEFK
KALGVHSLRQELDQISTHITLKDNLKVTLEKYNLSNQPNNLLEIEFNDYINLSFKALGMILPLMREGKYRDEACEIANLKPKTVDFTKDFPACDSIFAEELESNPWRAISEYRKVLNALLKGYKVKHIHELARDVGLSKKAREKIEEQKFNQAVNAWALKECEMGLKASAKMLKLIWKEQKEICIYGSKIEHLKDEKALEVDFFYNYSRSFDDSFINKLVLFTKINQEK]->TPFEAFGKMKWSKIQTQLAPNYPKNKNKILDENFKDKQQEDISRNLNDTRYIATLIAYKYTEYNLFFLLSENENANLSGEKSKHGQVTISGM
LTSVLRTWHGFDKKDRNNHLHHALDAIIVAYSTNSIKAIFDSFKNQELKARFYAKELSTDNKHYQFKKEPFKSFREIKLSKIDEIFVSKPPRKRARRAYLHKDTFHSXIIDKCSYNSEGKLQIALSCGRVRKIGT kayakvendtivrdvifkkQNKFYAIPIYAMDFAGLGPKNKVTGDKCNNPKQWQTI^EYEFCSFLYKNDLILLQQKNMQEPEFAYYNDFSISTSSICVEKDNKPNLTSQNLFSNAKEGVSQVSLGIQNLKVEKYIIITPLGDKIKADFQPRENISLKTSSKYGLR

[0145] OJI07263.1 hypothetical protein BK997_03320 [Candidatus Micrarchaeum acidiphium ARMAN-1]
MRDSITAPPRSSALAARIEKFNSAFKLGIDLGTTKGVALVKNVLKAKTFLDHYHKTQLEERRHRNRRSRALRKRIARLRSWLIRQKGYKGLPDPYKKMQLPNGVRKGENWIDLVVSGRDLSPFVRAITLIFQGFRGNYEVAVEIEEMSYKEFSTfflKALTSVTF.FEALAAEIERQDDVVDYTDKEAERTQUELSELLSKVSESKESSKDQARQREDLGKVNVAFCSAHRIEKDKWCEMLKLLDRPVRHARFLNKVLRCINCIDRATPKKSRPDVRELLFYDFTVRNFLKAGRVEQNPVDVISYKYYKIMDAEVIRVLKNEKLTDDEKKQKRKLASELNYNKEWTDQKMMQEQLKTLLMKTLGRSRYCAMHLKERAAGKDVEEGLHGVQKRHRRNIAQRNHDLRVINLIESLLFDQNKSLDRAINNGLMYVTIE

-97-
APEPKTKHAKKGAAVVRDPRKLKEKLFDDQNGVCIYTGLQLDKLEISKYEKDfflFPD
SPJ^GPSIRDNLVTrKEINSGBKGDRTPWEWMHDNPEKWKAFERRVAEFYHKGRINE
RKRELLNKGTEYGDPNPTELARGGARVNFITEFNDRLKTHGQVQELQTIFERNKPIV
QVVRGEETQRLLRQWNALNQFIPLKDRAMSFNHAEDAAIAASMPKFWREQYRT
AWHFGPSNERPDFALAEQPQNVNDFMTKGPIAQLVGTKYSWKHISIDTDYKIP
FSKSAIYVGYIKPKNAITSNIAKVLRPKLLNGEHTMSKNAKYYHQSIGNERFLMKSQ
KGGSnTVKPHDGPEQISPTYECAVLTKHDGKIIVKFPKPIKLPDMYARGVIKAM
DKELETSLSSMK^AKYKELHTHnYLPATKKHDGYFIITKLASKHIGKALPESMV
KVKYTQIGSENSEVKLTKPKPEITLDSEDITNIYNPR

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

MLGSSRYLRY^TSFEGKEIPFLIMGYYKEYNKELSQAQKEFNDQISEFNSYKGLID
LGDKTGIAVKGKNIAILAKTLIDLHSQKLDKRJIEARJINRRTRLSXKRDLRWSVM
RQKVGNQRLPDYKIIMidDKNYSIYNKNSANKKNWIDLLIHSNSLADDHFVRGLTI
IRKRGLAFKYLSDLSDKEFKEYIDNLKPISKYYEYDEDLEDSESSRVERAGEEEKKEF
GLKNLKDIDKESDFQVKQREEVKKELEDVLDFALKSDKNDKARWKRELLNLL
DKKKRKIRFDNRFLKCKIKGCNKNTPKXEVPIJFELKMVLNARSDYQISDDELNS
FRNEVINIFQKENLKKGLKGVTIEDLRQKLNTFNKAIKKGIREQIRSIIVFEKISGR
SKICKEHLKFEKAPSDPVIGGVNSAREQHDFRLVWIDKIKFDKDLIPSKRLYITI
ESPEPETEKLEKGQISESFETKLEKLAKETGGIDITGKEKLKDFEIEHFPRARMGPS
IRENEVASNLETTSTKEKADRTPWETFQGEKRWSEFEKRVSLSKSKISSERKREILL
NKSNEYPGLNPTELSRIPSTDLDFVESIRKMFVKYGYEFPQTLVQKGKIPQYVRGRDT
QALRWWRHLDNSIPEKDRRSSHFNAEDAVIAACMPPYXLRQKIFREEAKIKRKVS
MCEKEVTTRDIVfTJKJAPNWSEFMKTRNEPVIEVIGKVPWSNIMDQTFLFYYLLK
PFKDNLIKIPNVKNTGVMGTQDSLPSKVLISNKVDSTTVLVHDKKGGKP
xNWVPSIGGLLVIYTPKDGPKRIVQKVPAQTQGILLIPxNEDGRDVAVFENPVIEMY
NNGLAFVEKENEEELKYFNLLEGKMQKFERIRRYDMITYSNKFYYVTKINKHRVT
IQEESKIAESDKVSSGKEYTRKETEELSQLKLALISI

[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
Ga0054994_10813 Geobacillus stearothermophilus Cas9

MRYKIGLDIGITSVGAWVMNDLIPJEDLVPJFDRAEWQTGESLALPRLRARSAPvR
RLRPvRXHRLERIRRVLIPJEGILTKKEEKLFEKIFiiDIDVQQLRVLPEALRLKLNNDLVAR
VLLHIALKRGKFKNPvKRSRNSKENSTMLFEIIEENRAILSSRTGVEMIVKDPKFALH
KRNKGENYNTIARDDLEIRELISFKQREFGNSCTEFEENYITTAWSQRPVASKD
DIEKKVGFCTEPKEKRAPKATYFQSFIAWEHINKLLRLISPARGARLTDERRLLYEQ
AFQKNITYHDIRTTLLHLPPDTYFKGIVYDGRGESPQKKNENRIFLEDAYHQPvKAVDK
VYGKGGKSSSLPFDFTFGYALTFLKDDADIHYSYRENEEQNGKRMPNKLANKVYDN
ELIEELLNLSTKFGiisLALKALSLPLYPEMEOQEVYSSACERAGYTFTPGPKKKQKTMRL
PNIPPIANPVVMALTQKVPNIAIJKKYGSPVSiifELlardLsQTFDERRKTKEQDE
NRKKNETAIRQLEMYGLTLNPTGHDKIEKLWSEQNGRCAYSLQPIEIERLLEPGEYVE
VDHVIPYSRSLDDSNTKNVLVTRENREKGNRIPAELYLGTERWQQFETFVLNTKNQ
FSKKXRDLRLLRLFYDENEETEFKNRNLDTRYISRFANFIREHLFAESDKKQVKVY
TunsfGRVTALRSLRWEFNNKREESDLHHAVDAVIVACTTSDIAKVTAFYQRERQNK
ELAKKTEPHFPQPWpiFADELRAKSLKHPKESIKALNLGNVDDQQKLESQPVFVSMR
PKRSVTGAHQTLLRRYVGDERSGKIQTVDKVKTLSEIKLDASHFMYGKEDPRT
YEAIRQLLEHNNDPKKAFAQEPLYKPKNGEPGPVIRTKIITKQVIPLNDGKTVAMYN
MRVVDVFKEKDGKYVCVPYTGTM1MKGLPNKAIENPKPSEWEMTEDYTFR
FSLYPNDLIRIELPREKTVTAAGgeeINVDVFVYYKTDSANGGLEISHDHFRSFLRG
VGSRTLKRFKQYQVDGNIYKVRGEKRVGLASSAHSKPGKTIRPLQSTRD

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

MAAFKPNNPMPNYILGLDGIASVGAWAMVEDEENPIRIDLGLGRVFERAEPKTDGS
LAMARPXARSVRLTRJIRAHRLRRARRLKRREGVLQADFDENLAVKSLPNTPWQ
LRAAALDKRLTCLEWSAIVLLHLVKHRGYSRQRKGNETADKELGALKGVADNAHALQ
TGDFRTPAEALNLKEKESGifiIRNQRGDYHTSRSKDLQAEQLNLLEFKQKEFGN
PHVSGDLKEDIELLLMAQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLT
KLNNLRILEQSERLPDLTDERATLMDEPYRKSRLTYAQARLGLLEDATAFFKGLRY
GKDNEASTLMMKAYHJSSRALEKEGLDKKKSPNLSTLEQDIEQTAFSLFKTDKD
ITGRLKDRVQPEIEALLKiiisFDKVFQISLKLARRIVPLMEQKGYDEACAEICYGDH
YCKKNAEEKIYLPPPADEIRNPVVRALSRQARKVINCVVRRYGSPARiffETAREVGYK

-100-
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.

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.

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. 2015; 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 II). 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, non-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 β1 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. pneumophila* Cas9, *N. lactamica* Cas9, *N. meningitides* Cas9, *B. longum* Cas9, *A. muciniphilu* Cas9, and *O. lunatus* 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 present technology, 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.
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 10x 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 1x SSC to about 0.1x SSC; formamide concentrations of about 55% to about 75%; and wash solutions of about 1x SSC, 0.1x 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.

"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.

Modes of Carrying out the Disclosure

Methods of Generating Immunosilent Proteins and Identifying Immune Orthogonal Proteins

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.

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.

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) matrices.

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/mhcei/ (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 can 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 capsules, 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 protein 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 CRISPR-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 embodiments, 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 orthogonal orthologs based on non-overlapping sets of affinities. 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. muciniphila 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 repertoire of pre-existing immunity that could result in cross-reactivity against proteins or their orthologs. Thus, there exists some risk of confounding in sequential administration of proteins that are immune orthogonal. Non-limiting exemplary proteins 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.
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. muciniphilia Cas9, or O. laevis 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.

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), mucopolysaccharidosis (I-IX), multiple sclerosis, muscular dystrophy, myotonic dystrophy, neurofibromatosis 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.
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 (SCN1 1A gene) and 1.3 (SCN3A gene); transient receptor potential cation channel subfamily V member 1 (TrpVI), also known as the capsaicin receptor and the vanilloid receptor 1; PRDM12; or HCN2.

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.

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, intraperitoneal, intramuscular, intracardiac, intrathecal, subventricular, epidural, intracerebral, intracerebroventricular, sub-retinal, intravitreal, intraarticular, intraocular, intraperitoneal, intruterine, intradermal, subcutaneous, transdermal, transmucosal, and inhalation.

Methods of determining the most effective means and dosage of administration are known to those 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.
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.

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.

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 CRISPR-Cas systems which can be used herein. In general, dosing considerations are well understood by those in the art.

Compositions and Kits

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.

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**.
Example 2 - Effect of single amino acid swaps in immunogenic peptides in SpCas9

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.

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. 'HHQDLTLLK', 'HHQDLTLN', 'HHQD LTLD'). 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. 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.

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.

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

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.

<table>
<thead>
<tr>
<th>Week 0</th>
<th>Week 3</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>B2</td>
<td>Assay (baseline and role of AAVs)</td>
</tr>
<tr>
<td>A2</td>
<td>A1</td>
<td>Assay (baseline and role of AAVs)</td>
</tr>
<tr>
<td>B1</td>
<td>B2</td>
<td>Assay (baseline and role of AAVs)</td>
</tr>
<tr>
<td>B2</td>
<td>B1</td>
<td>Assay (baseline and role of AAVs)</td>
</tr>
<tr>
<td>A1</td>
<td>B2</td>
<td>Assay (Cas9 orthogonality)</td>
</tr>
<tr>
<td>B2</td>
<td>A1</td>
<td>Assay (Cas9 orthogonality)</td>
</tr>
<tr>
<td>A2</td>
<td>B1</td>
<td>Assay (Cas9 orthogonality)</td>
</tr>
<tr>
<td>B1</td>
<td>A2</td>
<td>Assay (Cas9 orthogonality)</td>
</tr>
</tbody>
</table>

[0237] Legend:
A1: AAV8 SpCas9 CD81; A2: AAVDJ SpCas9 Scarbl; B1: 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 91 Type II CRISPR-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\textsuperscript{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\textsuperscript{11-16}.

\textbf{[0243]} A common approach to circumventing these issues has been to utilize human proteins, or to humanize proteins by substitution of non-human components\textsuperscript{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\textsuperscript{18}. Another way to circumvent an immune response to protein therapeutics is the removal of immunogenic T cell epitopes.\textsuperscript{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,\textsuperscript{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\textsuperscript{22-25}.

\textbf{[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\textsuperscript{26,27}. This provides an even greater challenge as repeated exposure to the same antigen can elicit a more robust secondary immune response\textsuperscript{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. Comparative genomics has demonstrated that Cas9 proteins are widely distributed across bacterial species and have diversified over an extensive evolutionary history. 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, which promotes long-term transgene expression and treatment efficacy. Despite the relatively low immunogenicity of AAV vectors, antibodies against both the capsid and transgene may still be elicited. Additionally, the prevalence of neutralizing antibodies (NAB) against AAVs in the human population and cross-reactivity between serotypes remains a hurdle for efficacious AAV therapy. Although AAVs were initially considered non-immunogenic due to their poor transduction of antigen-presenting cells (APCs), it is now known that they can transduce dendritic cells (DCs) and trigger innate immune responses through Toll-like receptor (TLR) signaling pathways. The ability to transduce DCs is dependent on AAV serotype and genome, and may be predictive of overall immunogenicity.

[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.\textsuperscript{53,54} Finally, Applicants experimentally
confirmed these immunological predictions by assaying treated mice for induction of protein-
targeting antibodies.

\begin{itemize}
\item[\textbf{[0247]}] \textbf{Humoral immune response to AAV and Cas9}
\item[\textbf{[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\textsuperscript{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.
\item[\textbf{[0249]}] \textbf{Identifying immune-orthogonal proteins}
\item[\textbf{[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 91 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

-127-
(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 AAVs56. This analysis suggests, consistent with previous observations5758, 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, 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, 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. Similar to what has been done in cancer antibody therapeutics, 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.
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.

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. 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 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.

Previous work has identified that MHC affinity is highly dependent on anchor residues at either end of the binding pocket. 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, 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-CRISPRs (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. 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. 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. 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 91 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 81 HLA-1 alleles using netMHC 4.074 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 3.175 for class II (alleles can be found at http://www.cbs.dtu.dk/services/NetMHCIIpan-3. l/alleles_name.list). 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
AAV Production

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 1x 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 1x PBS (pH 7.2) supplemented with 50 mM NaCl and 0.0001% of Pluronic F68 (Thermo Fisher) using 100kDA 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′-GGAACCCCTAGTGATGGAGTT-3′.

Animal studies

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 1x10^12 vg/mouse.

ELISA

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 1x 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, 10⁹ vg of AAV-2, -5, -8 or -DJ in 1x coating buffer was used to coat each well of 96-well Nuc MaxiSorp Plates. Plates were washed three times for 5 min with 350 µl of 1x Wash Buffer (Bethyl) and blocked with 300 µl of 1x 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 µl 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 1hr at RT, wells were washed four times for 5 min, and 100 µl 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


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


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).


Table 1

<table>
<thead>
<tr>
<th>Peptide Pos</th>
<th>Peptide ID</th>
<th>Affinity</th>
<th>Affinity u.m.m.</th>
<th>Score</th>
<th>Annual per Surface</th>
<th>Maturation</th>
<th>Pos</th>
<th>Peptide</th>
<th>mat.</th>
<th>Rank</th>
<th>ID</th>
<th>Affinity</th>
<th>ID</th>
<th>Affinity</th>
<th>ID</th>
<th>n.m.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3011-3023</td>
<td>1-78</td>
<td>Streptococcus sp.</td>
<td>HA-BSI (03)</td>
<td>High 9</td>
<td>0-280-280</td>
<td>Yes</td>
<td></td>
<td>236-244 No</td>
<td>816</td>
<td>675-555</td>
<td>80</td>
<td>Streptococcus sp.</td>
<td></td>
<td>M-A0201</td>
<td>No</td>
<td>E</td>
</tr>
<tr>
<td>3024-3036</td>
<td>2-79</td>
<td>Streptococcus sp.</td>
<td>HA-EPD (02)</td>
<td>High 9</td>
<td>0-280-280</td>
<td>Yes</td>
<td></td>
<td>236-244 No</td>
<td>816</td>
<td>675-555</td>
<td>80</td>
<td>Streptococcus sp.</td>
<td></td>
<td>M-A0201</td>
<td>No</td>
<td>E</td>
</tr>
<tr>
<td>3037-3049</td>
<td>3-80</td>
<td>Streptococcus sp.</td>
<td>HA-EPD (02)</td>
<td>High 9</td>
<td>0-280-280</td>
<td>Yes</td>
<td></td>
<td>236-244 No</td>
<td>816</td>
<td>675-555</td>
<td>80</td>
<td>Streptococcus sp.</td>
<td></td>
<td>M-A0201</td>
<td>No</td>
<td>E</td>
</tr>
<tr>
<td>3050-3062</td>
<td>4-81</td>
<td>Streptococcus sp.</td>
<td>HA-EPD (02)</td>
<td>High 9</td>
<td>0-280-280</td>
<td>Yes</td>
<td></td>
<td>236-244 No</td>
<td>816</td>
<td>675-555</td>
<td>80</td>
<td>Streptococcus sp.</td>
<td></td>
<td>M-A0201</td>
<td>No</td>
<td>E</td>
</tr>
</tbody>
</table>

Contd.; same rows:
<table>
<thead>
<tr>
<th>sgID</th>
<th>gene</th>
<th>transcript</th>
<th>protospacer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCN3A_+(_.166060543.2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GATCTCAGAACAGGAAAGG</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>G</td>
</tr>
<tr>
<td>SCN3A_+(_.166060199.2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GTGTAATTACAGGAAAC</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>SCN3A_+(_.166060301.23-P1P2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GACCTGGTACGCTAGTCT</td>
</tr>
<tr>
<td>SCN3A_+(_.166060552.2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GATAGAGTGAATCTCAC</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>SCN3A_+(_.166060129.2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GAAAAAGCGCTGCAGGAAAG</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>SCN3A_+(_.166060346.2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GTTATGCTGTAATTCATA</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>SCN3A_+(_.166060119.2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GCTTGGAAATGGTATTG</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>SCN3A_+(_.166060135.2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GAAAGAAAATAGAGCCTGCLA</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>SCN3A_+(_.166060371.2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GCCAAATCTTGGATGCT</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>SCN3A_+(_.166060281.2</td>
<td>SCN3A</td>
<td>P1P2</td>
<td>GACCATAAGACCTAGTACC</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>SCN9A_+_.167232419.2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GCGGTGCGCCAGCCTCCA</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>G</td>
</tr>
<tr>
<td>SCN9A_+_.167232052.2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GCCACCTGGAAGAAAGAGA</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>G</td>
</tr>
<tr>
<td>SCN9A_+_.167232416.2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GCTCGCCAGGCGTCCAGG</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>G</td>
</tr>
<tr>
<td>SCN9A_+_.167232010.2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GCCAGCAATGGAGGAGA</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>SCN9A_+_.167232085.23-P1P2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GCTCGGTCGCTACCTCCA</td>
</tr>
<tr>
<td>SCN9A_+_.167232476.2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GGGCGGCGTCTCCAGG</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>SCN9A_+_.167232437.2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GGGCGCAGTCTGGTCAGG</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td>G</td>
</tr>
<tr>
<td>SCN9A_+_.167232409.2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GCCGCTCCAGGCGGCGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G</td>
</tr>
<tr>
<td>3-P1P2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>T</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
<td>------</td>
<td>----------------</td>
</tr>
<tr>
<td>SCN9A_+._167232021.2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GACGGGTGGTTCCAGCAA T</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN9A_+._167232018.2</td>
<td>SCN9A</td>
<td>P1P2</td>
<td>GGGTGTTCCAGCAATGG G</td>
</tr>
<tr>
<td>3-P1P2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_._._38835462.23-</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GGTACTCCGAGTAAGCG A</td>
</tr>
<tr>
<td>ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_._._38835311.23-</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GGGAGCTACCATAGAACTT G</td>
</tr>
<tr>
<td>ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_._._38835269.23-</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GACGGATCTAGATCCTCA G</td>
</tr>
<tr>
<td>ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_+_38835213.2</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GCCGGGTAAGAAGCTACTAG T</td>
</tr>
<tr>
<td>3-ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_._._38835251.23-</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GCCCCGGTGTTGCTGTAGA A</td>
</tr>
<tr>
<td>ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_+_38835434.2</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GTTTACTCCGAGTCACTG G</td>
</tr>
<tr>
<td>3-ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_._._38835449.23-</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GCTATCCACCAGTGACTC</td>
</tr>
<tr>
<td>ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_._._38835156.23-</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GACATCACCACGGCCAAG G</td>
</tr>
<tr>
<td>ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_._._38835491.23-</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GTAGTTTGCAGGGATCCAA T</td>
</tr>
<tr>
<td>ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN10A_+_38835272.2</td>
<td>SCN10A</td>
<td>ENST0000044908 2.2</td>
<td>GCTCCACGACAGACTGATC G</td>
</tr>
<tr>
<td>3-ENST0000044908.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN11A_._._38991624.23-</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3</td>
<td>GATGGGTCAGATCTTCCA G</td>
</tr>
<tr>
<td>ENST0000030232.3,ENST00000450244.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCNIIA_+_38992032.2</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3</td>
<td>GTTCCTGCTATACCCACAG G</td>
</tr>
<tr>
<td>3-ENST0000030232.3,ENST00000450244.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN11A_._._38991801.23-</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3</td>
<td>GCCAGAGATCGGAAAGTGA A</td>
</tr>
<tr>
<td>ENST0000030232.3,EN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST00000450244.1</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3, ENST00000045 0244.1</td>
<td>GCCTGCTATACCCACAGTG</td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
<td>--------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>SCN11A_+ _38992029.2 3-</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3, ENST00000045 0244.1</td>
<td>GGGAAAGCCTCTGGAAGAC</td>
</tr>
<tr>
<td>SCN11A_+ _38991609.2 3-</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3, ENST00000045 0244.1</td>
<td>GGAAGAGATGACCACCAC</td>
</tr>
<tr>
<td>SCN11A_+ _38991666.23-</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3, ENST00000045 0244.1</td>
<td>GGAATGTCGCCATAGAGCT</td>
</tr>
<tr>
<td>SCN11A_+ _38991666.23-</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3, ENST00000045 0244.1</td>
<td>GGAGCTCATAGGAAAGCCT</td>
</tr>
<tr>
<td>SCN11A_+ _38991942.2 3-</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3, ENST00000045 0244.1</td>
<td>GCTTTAAGACTGGAATCCTA</td>
</tr>
<tr>
<td>SCN11A_+ _38991653.2 3-</td>
<td>SCN11A</td>
<td>ENST0000030232 8.3, ENST00000045 0244.1</td>
<td>GGGAAGTTCGAATACAGCT</td>
</tr>
<tr>
<td>SHANK3_+ _51135959.2 3-P1P2</td>
<td>SHANK3</td>
<td>P1P2</td>
<td>GGAATTCGAATACAGCTCCT</td>
</tr>
<tr>
<td>SHANK3_+ _51136404.2 3-P1P2</td>
<td>SHANK3</td>
<td>P1P2</td>
<td>GCTTCAGGCAGAGACCCG</td>
</tr>
<tr>
<td>SHANK3_+ _51136356.2 3-P1P2</td>
<td>SHANK3</td>
<td>P1P2</td>
<td>GGAGCTCCCGTGATTGACAC</td>
</tr>
<tr>
<td>SHANK3_+ _51136302.2 3-P1P2</td>
<td>SHANK3</td>
<td>P1P2</td>
<td>GCACGGCAGAAGCTCCT</td>
</tr>
<tr>
<td>SHANK3_+ _51136333.2 3-P1P2</td>
<td>SHANK3</td>
<td>P1P2</td>
<td>GAGCACCAGAGGACCCGC</td>
</tr>
<tr>
<td>SHANK3_+ _51136329.2 3-P1P2</td>
<td>SHANK3</td>
<td>P1P2</td>
<td>GGGGCGGAAAGACAGACAG</td>
</tr>
<tr>
<td>SHANK3_+ _51136329.2 3-P1P2</td>
<td>SHANK3</td>
<td>P1P2</td>
<td>GGGACAGACAGACAGCGA</td>
</tr>
<tr>
<td>Gene</td>
<td>Start Position</td>
<td>End Position</td>
<td>Nucleotides</td>
</tr>
<tr>
<td>------------</td>
<td>----------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>SHANK3</td>
<td>51136143.23</td>
<td>P1P2</td>
<td>GACcgcggcgaggccgtaa</td>
</tr>
<tr>
<td>TRPV1</td>
<td>3500355.23</td>
<td>P1P2</td>
<td>GAGcgACTCCTGGCTAGTC</td>
</tr>
<tr>
<td>TRPV1</td>
<td>3499964.23</td>
<td>P1P2</td>
<td>GCTCCTTGGAAACCTGG</td>
</tr>
<tr>
<td>TRPV1</td>
<td>3500391.23</td>
<td>P1P2</td>
<td>GAGTGCGTGGAGGCCTCT</td>
</tr>
<tr>
<td>TRPV1</td>
<td>3500224.23</td>
<td>P1P2</td>
<td>GGGACTCACCAGCTAGAC</td>
</tr>
<tr>
<td>TRPV1</td>
<td>3500327.23</td>
<td>P1P2</td>
<td>GTGGTCTCCCAGCTCCGTG</td>
</tr>
<tr>
<td>TRPV1</td>
<td>3500298.23</td>
<td>P1P2</td>
<td>GGGGAGAGCTGGGCTCGT</td>
</tr>
<tr>
<td>TRPV1</td>
<td>3500017.23</td>
<td>P1P2</td>
<td>Gtgcctaaaggtggtcg</td>
</tr>
<tr>
<td>TRPV1</td>
<td>3499899.23</td>
<td>P1P2</td>
<td>GCTGACATCAACGGTCTCG</td>
</tr>
<tr>
<td>TRPV1</td>
<td>3500400.23</td>
<td>P1P2</td>
<td>GGGAGGCCTCTCGACCT</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>14133341.23</td>
<td>P1P2</td>
<td>GGATTCGCGTCTCCCAGG</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>14132929.23</td>
<td>P1P2</td>
<td>GGATATGCAACGGAGAAGA</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>14132903.23</td>
<td>P1P2</td>
<td>GCTCTAGACGGCAGGATTA</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>14133316.23</td>
<td>P1P2</td>
<td>GGGGAAAAAGGAGCCGTCA</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>14132924.23</td>
<td>P1P2</td>
<td>GCAAGCGGAAGAAGGAC</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>1413295.23</td>
<td>P1P2</td>
<td>GCCAAAGCGTCCCCTCCTA</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>1413298.23</td>
<td>P1P2</td>
<td>GAAGCGTCCCCCTTCTAG</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>14132855.23</td>
<td>P1P2</td>
<td>GCCCTTCAAACGACAGGTA</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>14133247.23</td>
<td>P1P2</td>
<td>GACCATGCTCCACGGAGGG</td>
</tr>
<tr>
<td>Location</td>
<td>Gene/Protein</td>
<td>Region</td>
<td>Sequence</td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>--------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>-P1P2</td>
<td>GRIN2B</td>
<td>P1P2</td>
<td>GGAATGACCATGCTCCACC</td>
</tr>
<tr>
<td>14133252.23</td>
<td>PRDM12</td>
<td>P1P2</td>
<td>GgtcgggcccgcgATGAT</td>
</tr>
<tr>
<td>133540034.23-P1P2</td>
<td>PRDM12</td>
<td>P1P2</td>
<td>GCCACGGAGCCCACCATGgg</td>
</tr>
<tr>
<td>133540230.23-P1P2</td>
<td>PRDM12</td>
<td>P1P2</td>
<td>GGACTGCAGCGACACTCGG</td>
</tr>
<tr>
<td>133539846.23-P1P2</td>
<td>PRDM12</td>
<td>P1P2</td>
<td>Gctggaggaagcgaacga</td>
</tr>
<tr>
<td>133540263.23-P1P2</td>
<td>PRDM12</td>
<td>P1P2</td>
<td>GTGGCGCAGTCCTTCTCCG</td>
</tr>
<tr>
<td>133540260.23-P1P2</td>
<td>PRDM12</td>
<td>P1P2</td>
<td>GTGCTGGCGCAGTCCTTCTC</td>
</tr>
<tr>
<td>133540257.23-P1P2</td>
<td>PRDM12</td>
<td>P1P2</td>
<td>GCGACGCTGACTCACCAGG</td>
</tr>
<tr>
<td>133540233.23-P1P2</td>
<td>PRDM12</td>
<td>P1P2</td>
<td>GAGGGACTGCGCCAGCACC</td>
</tr>
<tr>
<td>133539961.23-P1P2</td>
<td>PRDM12</td>
<td>P1P2</td>
<td>GCCGCGCAATCCCTCTCC</td>
</tr>
<tr>
<td>589972.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>Gtggcggaggagggagccaa</td>
</tr>
<tr>
<td>590106.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>GCAACGCCTcggccgggggc</td>
</tr>
<tr>
<td>589880.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>GgcccgccgaggAGCCCGA</td>
</tr>
<tr>
<td>590306.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>GcggcACGAGAACGACACCT</td>
</tr>
<tr>
<td>590253.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>GCAGCCCCAGCGCGAGTG</td>
</tr>
<tr>
<td>590235.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>GGGCGCGCAGCTCGCGGTTC</td>
</tr>
<tr>
<td>590335.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>GTCGTTCTCGTggccggggg</td>
</tr>
<tr>
<td>590407.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>GAGCTGCCCTGCGTggccg</td>
</tr>
<tr>
<td>590332.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>GTGTCGTTCTCGTggccg</td>
</tr>
<tr>
<td>590204.23-P1P2</td>
<td>HCN2</td>
<td>P1P2</td>
<td>GCGCGTGCTggccggcggccg</td>
</tr>
<tr>
<td>sgID</td>
<td>gene</td>
<td>transcript</td>
<td>protospacer sequence</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Scn3a+_65567459.23-P1P2</td>
<td>Scn3a</td>
<td>P1P2</td>
<td>GTGAATTCAGAACAGGAA</td>
</tr>
<tr>
<td>Scn3a+_65567442.23-P1P2</td>
<td>Scn3a</td>
<td>P1P2</td>
<td>GAGCGGAGGCATAAGCAG</td>
</tr>
<tr>
<td>Scn3a-_65567234.23-P1P2</td>
<td>Scn3a</td>
<td>P1P2</td>
<td>GATCTGGTGAGGCTAGATTCT</td>
</tr>
<tr>
<td>Scn3a-_65567301.23-P1P2</td>
<td>Scn3a</td>
<td>P1P2</td>
<td>GAGGAATCACAGCTCAACA</td>
</tr>
<tr>
<td>Scn3a-_65567522.23-P1P2</td>
<td>Scn3a</td>
<td>P1P2</td>
<td>GATCAGAAACGCGCTGG</td>
</tr>
<tr>
<td>Scn3a-_65567271.23-P1P2</td>
<td>Scn3a</td>
<td>P1P2</td>
<td>GGTGTGTCAGCCTACCTGA</td>
</tr>
<tr>
<td>Scn3a-_65567326.23-P1P2</td>
<td>Scn3a</td>
<td>P1P2</td>
<td>GCCATCCAAGATGGTTAGA</td>
</tr>
<tr>
<td>Scn3a+_65567264.23-P1P2</td>
<td>Scn3a</td>
<td>P1P2</td>
<td>GATTCTAAGGCTCTCACT</td>
</tr>
<tr>
<td>Scn3a+_65567031.23-P1P2</td>
<td>Scn3a</td>
<td>P1P2</td>
<td>GCAATAACAGACTAGGAATT</td>
</tr>
<tr>
<td>Scn9a+_66634758.23-P1P2</td>
<td>Scn9a</td>
<td>P1P2</td>
<td>GAGCTCAGGAGCATCGAG</td>
</tr>
<tr>
<td>Scn9a-_66634675.23-P1P2</td>
<td>Scn9a</td>
<td>P1P2</td>
<td>GAGAGTCGCAATTGGAGCG</td>
</tr>
<tr>
<td>Scn9a-_66634637.23-P1P2</td>
<td>Scn9a</td>
<td>P1P2</td>
<td>GCCAGACCAGCCTGACAG</td>
</tr>
<tr>
<td>Scn9a-_66634689.23-P1P2</td>
<td>Scn9a</td>
<td>P1P2</td>
<td>GAGCGCGAGCTAGGCGCT</td>
</tr>
<tr>
<td>Scn9a-_66634610.23-P1P2</td>
<td>Scn9a</td>
<td>P1P2</td>
<td>GCTAGGAGTCCGGGATACC</td>
</tr>
<tr>
<td>Scn9a+_66634478.23-P1P2</td>
<td>Scn9a</td>
<td>P1P2</td>
<td>GAATCAGCGAGGCTCACT</td>
</tr>
<tr>
<td>Scn9a-_66634461.23-P1P2</td>
<td>Scn9a</td>
<td>P1P2</td>
<td>GACACGGCTGACAGTGGG</td>
</tr>
<tr>
<td>Scn9a+_66634731.23-P1P2</td>
<td>Scn9a</td>
<td>P1P2</td>
<td>GCGACGCGGTTGGCAGCG</td>
</tr>
<tr>
<td>P1P2</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_+119719110.2 3-P1P2</td>
<td>Scn10α P1P2 GCCAGGTTGGAACTCGTGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_+U9719123.2 3-P1P2</td>
<td>Scn10α P1P2 GCACCATCCAGCAAGCAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_-119719078.23-P1P2</td>
<td>Scn10α P1P2 GCGTCACTCAAGGATCTAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_+119719086.2 3-P1P2</td>
<td>Scn10α P1P2 GATGGGAATGGCCACCCAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_+119718921.2 3-P1P2</td>
<td>Scn10α P1P2 GCCTTGAACGGGAGAACAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_+H9719051.2 3-P1P2</td>
<td>Scn10α P1P2 GAGATTTTGAGTGACCGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_-119719025.23-P1P2</td>
<td>Scn10α P1P2 GCGGGGCTCTCCACGAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_-119719095.23-P1P2</td>
<td>Scn10α P1P2 GCAAGGAATCAGCCCTCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_+119718881.2 3-P1P2</td>
<td>Scn10α P1P2 GCACCAGCGGAATGCTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scn10α_+119719014.2 3-P1P2</td>
<td>Scn10α P1P2 GGCAAGCAGGACCCTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_+119825404.2 3-P1P2</td>
<td>Scnlla P1P2 GAGGTAAGCCCATCCGGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_-119825450.23-P1P2</td>
<td>Scnlla P1P2 GTTCGCTAGGGAGGCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_-119825400.23-P1P2</td>
<td>Scnlla P1P2 GCCTGAACGACAGAGGAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_+119825277.2 3-P1P2</td>
<td>Scnlla P1P2 GTCAGAGGTGGGAGCCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_-119825394.23-P1P2</td>
<td>Scnlla P1P2 GCCCGACCTGAAAGCAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_+119825463.2 3-P1P2</td>
<td>Scnlla P1P2 GGCCAAGAGCGAGAATCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_+119825246.2 3-P1P2</td>
<td>Scnlla P1P2 GGTCTAGGTGCAGAGCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_+119825242.2 3-P1P2</td>
<td>Scnlla P1P2 GGGTGTCAGAGGATCCTGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_+119825431.2 3-P1P2</td>
<td>Scnlla P1P2 GTGCCCTGAGCCTCCCTAGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scnlla_-119825253.23-P1P2</td>
<td>Scnlla P1P2 GTCTGTGAGAACCGACCGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shank3_+89499659.23-P1P2</td>
<td>Shank3 P1P2 GGGCTCCGCAGGCAGCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shank3_+_89499688.23</td>
<td>Shank3</td>
<td>P1P2</td>
<td>GgggccagcgcgggggACAG</td>
</tr>
<tr>
<td>Shank3_+_89499943.23</td>
<td>Shank3</td>
<td>P1P2</td>
<td>GCCGCTAGCGGGCCACACA G</td>
</tr>
<tr>
<td>Shank3_+_89499679.23</td>
<td>Shank3</td>
<td>P1P2</td>
<td>GcggggACAGCGGCTCCGG</td>
</tr>
<tr>
<td>Shank3_+_89499612.23</td>
<td>Shank3</td>
<td>P1P2</td>
<td>GCATCGGCCCGGCTTCA</td>
</tr>
<tr>
<td>Shank3_+_89499924.23</td>
<td>Shank3</td>
<td>P1P2</td>
<td>GGGTACCGGAGATCGCA</td>
</tr>
<tr>
<td>Shank3_+_89499878.23</td>
<td>Shank3</td>
<td>P1P2</td>
<td>GTGCCGACGCACGACC</td>
</tr>
<tr>
<td>Shank3_+_89499676.23</td>
<td>Shank3</td>
<td>P1P2</td>
<td>GCACGCCGCTGGCCCT</td>
</tr>
<tr>
<td>Shank3_+_89499818.23</td>
<td>Shank3</td>
<td>P1P2</td>
<td>GGGGCCGGACTGTCCCG</td>
</tr>
<tr>
<td>Shank3_+_8949938.23</td>
<td>Shank3</td>
<td>P1P2</td>
<td>GAGGCGGACACAGGG</td>
</tr>
<tr>
<td>Trpvl_+_7323453.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GGGACTTACCAGCTAGGTG</td>
</tr>
<tr>
<td>Trpvl_+_73234330.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GCCCAAAAAGACAGCTCC</td>
</tr>
<tr>
<td>Trpvl_+_73234384.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GGCTGTAAGTCTTCTCAT</td>
</tr>
<tr>
<td>Trpvl_+_73234339.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GGGTCAGGGCACACTGCA</td>
</tr>
<tr>
<td>Trpvl_+_73234537.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GACTTACTTGCGTGACTG</td>
</tr>
<tr>
<td>Trpvl_+_73234478.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GTAGGCCCTCCAGAAGTCC</td>
</tr>
<tr>
<td>Trpvl_+_73234495.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GGCTGCTTGACCTGTGG</td>
</tr>
<tr>
<td>Trpvl_+_73234635.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GCCACGGAAGGCCTCCAG</td>
</tr>
<tr>
<td>Trpvl_+_73234346.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GCAAGGGACCTTGCTCCAT</td>
</tr>
<tr>
<td>Trpvl_+_73234280.23</td>
<td>Trpvl</td>
<td>P1P2</td>
<td>GGCTGCTGCTGGAAGAAGG</td>
</tr>
<tr>
<td>Grin2b_+_136172154.23</td>
<td>Grin2b</td>
<td>P1P2</td>
<td>GCCAACCTGATGGAAAGAG</td>
</tr>
<tr>
<td>Grin2b_+_136172179.23</td>
<td>Grin2b</td>
<td>P1P2</td>
<td>GAGGGAAGTGGAAAGCAAG</td>
</tr>
<tr>
<td>Grin2b_-</td>
<td>Grin2b</td>
<td>P1P2</td>
<td>GTGGGACAGGCGATGAG</td>
</tr>
<tr>
<td>Genome Position</td>
<td>Referenced Position</td>
<td>Strain</td>
<td>Annotation</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td><em>136172123.23-P1P2</em></td>
<td>Grin2b&gt;_136172089.2</td>
<td>P1P2</td>
<td></td>
</tr>
<tr>
<td><em>136172145.23-P1P2</em></td>
<td>Grin2b-</td>
<td>P1P2</td>
<td></td>
</tr>
<tr>
<td><em>136171934.23-P1P2</em></td>
<td>Grin2b-</td>
<td>P1P2</td>
<td></td>
</tr>
<tr>
<td><em>136171999.23-P1P2</em></td>
<td>Grin2b-</td>
<td>P1P2</td>
<td></td>
</tr>
<tr>
<td><em>136172002.23-P1P2</em></td>
<td>Grin2b-</td>
<td>P1P2</td>
<td></td>
</tr>
<tr>
<td><em>136171844.2_3-P1P2</em></td>
<td>Grin2b+</td>
<td>P1P2</td>
<td></td>
</tr>
<tr>
<td><em>136172019.2_3-P1P2</em></td>
<td>Grin2b+</td>
<td>P1P2</td>
<td></td>
</tr>
</tbody>
</table>
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 protein 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. muciniphilia* 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), mucopolysaccharidosis (I-LX), multiple sclerosis, muscular dystrophy, myotonic dystrophy, neurofibromatosis 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)

B

clique size

number of cliques containing proteins

5-mer similarity
FIGURE 4

Immune-orthogonal sets of proteins high and medium affinity overlaps
FIGURE 5A

Administration of protein therapeutic

---

Immune response to therapeutic cargo & delivery vehicle

FIGURE 5B

Cas9
gRNA

cargo: CRISPR-Cas9

+

vehicle: AAV

FIGURE 5C

in vivo genome targeting
AAV-SaCas9 serum ELISA vs. PCSK9 protein

Absorbance

empty_SaCas9

gPCS9_SaCas9

Week

0 1 2 3 4 5
**FIGURE 5D**

α cargo (CRISPR-Cas9)

AAV-SaCas9 serum ELISA vs. SaCas9 protein

**FIGURE 5E**

α vehicle (AAV)

AAV-SaCas9 serum ELISA vs. AAV virus
FIGURE 5F

Protein orthologs

k-mer analyses

Sequence comparison

Comprehensive peptide overlap

All possible immune overlap

Immune orthogonal cliques

Immunogenicity predictions

MHC Class I & II peptide overlap

Predicted immune overlap

MHC analyses
**FIGURE 7**

Cas9 cliques: 6-mer peptide overlaps

Log peptide overlap

- S. pyogenes
- S. aureus
- B. longum
- A. mucilaginosa
- O. tannae
FIGURE 8C

Cas9
MHC Class 2

\[ R^2 = 0.119 \]

FIGURE 8D

AAV
MHC Class 2

\[ R^2 = 0.392 \]
INTERNATIONAL SEARCH REPORT

International application No. PCT/US20 18/022258

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - A61 K 38/00; A61 K 39/395; A61 P 35/00; C07K 16/28; C12N 5/10; C12N 15/09 (201 8.01 )
CPC - A61 K 38/00; A61 K 39/395; A61 K 2039/505; C07K 231 7/24; C07K 231 9/00; C07K 231 9/30 (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

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>US 7,615,217 B2 (GILLIES et al) 10 November 2009 (10.11.2009) entire document</td>
<td>1, 2</td>
</tr>
<tr>
<td>X</td>
<td>WO 2015/153789 A1 (EDITAS MEDICINE, INC.) 08 October 2015 (08.10.2015) entire document</td>
<td>14-17</td>
</tr>
</tbody>
</table>

P. X


A


A


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

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: 23 May 2018

Date of mailing of the international search report: 25 Jun 2018

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450
Facsimile No. 571-273-8300

Authorized officer
Blaine R. Copenheaver
PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

Form PCT/ISA/2 10 (second sheet) (January 2015)
## INTERNATIONAL SEARCH REPORT

Box No. 1  Nucleotide and/or amino acid sequence(s) (Continuation of item l.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 31er. 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 13fr. 1(a)).
      □ on paper or in the form of an image file (Rule 13fr. 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.

---

Form PCT/ISA/210 (continuation of first sheet (1)) (January 2015)
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. □ 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 4.4(a).

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.